AGO-unbound cytosolic pool of mature miRNAs in plant cells reveals a novel regulatory step at AGO1 loading

Ágnes Dalmadi [ID][†], **Péter Gyula** [ID][†], **Jeannette Bálint, György Szittya** [ID] **and Zoltán Havelda** [ID]*

Department of Plant Biotechnology, Agricultural Biotechnology Institute, National Agricultural Research and Innovation Centre, Szent-Györgyi Albert Str. 4., Gödöllő 2100, Hungary

ABSTRACT

RNA interference (RNAi) is mediated by small, 20-24-nt-long, non-coding regulatory (s)RNAs such as micro (mi) and small interfering (si) RNAs via the action of ARGONAUTE (AGO) proteins. High-throughput sequencing of size-separated sRNA pools of plant crude extracts revealed that the majority of the canonical miRNAs were associated with high molecular weight RNA-induced silencing complexes co-migrating with AGO1 (HMW RISC). In contrast, the majority of 24-nt-long siRNAs were found in association with low molecular weight complexes co-migrating with AGO4 (LMW RISC). Intriguingly, we identified a large set of cytoplasmic sRNAs, including mature miRNA sequences, in the low molecular size range corresponding to protein-unbound sR-NAs. By comparing the RISC-loaded and protein-unbound pools of miRNAs, we identified miRNAs with highly different loading efficiencies. Expression of selected miRNAs in transient and transgenic systems validated their altered loading abilities implying that this process is controlled by information associated with the diverse miRNA precursors. We also showed that the availability of AGO proteins is a limiting factor determining the loading efficiency of miRNAs. Our data reveal the existence of a regulatory checkpoint determining the RISC-loading efficiencies of various miRNAs by sorting only a subset of the produced miRNAs into the biologically active RISCs.

INTRODUCTION

RNA interference (RNAi) is a fundamental and widespread regulatory mechanism, which has been in the focus of extensive basic and applied researches. RNAi mediates sequence-specific regulation of target RNAs by the action of small, 20–24 nucleotide (nt) long, non-coding regulatory (s)RNAs such as micro (mi) and small interfering (si) RNAs (1,2). In plants, various RNAi pathways play indispensable roles in diverse developmental processes and in responses to biotic and abiotic stress factors (3). MiRNAs are typically encoded by independent transcription units transcribed by RNA polymerase II producing primary miRNA transcripts (pri-miRNAs) (4). In the nucleus, pri-miRNAs are targets of extensive post-transcriptional processing to produce mature miRNAs. The DICER-LIKE1 (DCL1), an RNase III enzyme, cleaves the hairpin out of the pri-miRNA producing an 80–300-nt-long miRNA precursors (pre-miRNAs). The same DCL1 enzyme cleaves the precursor at sites determined by structural features to generate small double-stranded miRNA intermediates, the miRNA:miRNA* duplexes (5). These duplexes are $2'$-O-methylated at their $3'$ termini by the HUA ENHANCER 1 (HEN1) methyltransferase and then transported to the cytoplasm probably in complex with ARGONAUTE1 (AGO1), the central executor component of miRNA mediated RNAi (6). During this process, one of the strands of miRNA duplexes (miRNA*) is displaced while the mature strand (miRNA) remains bound to AGO1. In the cytoplasm, the miRNA loaded AGO1 containing RNA-induced silencing complex (RISC) mediates the sequence-specific repression of target RNAs via mRNA-cleavage or translational repression. Polysome fractionation indicated that AGO1 can be associated with polysomes (7) and it was also revealed that AGO1 linked miRNAs are enriched in membrane-bound polysomes (8). Since miRNAs are key regulators of essential developmental processes and stress responses, their action must be tightly controlled. There are several transcriptional and post-transcriptional regulatory mechanisms which control the production level and function of miRNAs. The $2'$-O-methylation of miRNA intermediate duplexes at their $3'$ ends protects them from oligo-uridylation and degradation (9) while exoribonucleases of the SDN (SMALL RNA DEGRADING) family degrade mature

*To whom correspondence should be addressed. Tel: +36 28 526106; Fax: +36 28 526101; Email: havelda.zoltan@abc.naik.hu
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

miRNAs influencing their accumulation level (10). AGO1 can also stabilize the levels of some miRNAs in the cytoplasm as a result of stress response (11). In addition to AGO1, other nine AGO proteins are encoded in the *Arabidopsis thaliana* genome and they have specialized functions for various RNAi pathways but often show functional redundancies (12). Regulation of miRNA sorting into various AGO proteins is strongly influenced by the 5′ nucleotide of small RNAs, i.e. AGO1 shows a preference for 5′ U (13) and also by structural features on the miRNA duplexes (14). In contrast to miRNAs, different siRNA classes are generated from perfect double-stranded (ds)RNA intermediates produced by RNA-dependent RNA polymerases (RDRPs) (1). The 21-nt-long phased (pha)siRNAs are originated from Pol II-dependent mRNAs cleaved by a 22-nt long class of miRNAs (15). Trans-acting (tasi) RNAs are a subset of phasiRNAs and act at the post-transcriptional level by negatively regulating disease resistance and transcription factor genes (16). DCL3 produces the 24-nt-long heterochromatic siRNAs (hetsiRNA) by processing the Pol IV/RDR2-dependent long dsRNA products originating from intergenic or repetitive regions of the genome. HetsiRNAs are predominantly loaded into the RNA-induced transcriptional silencing (RITS) complex playing fundamental roles in maintaining genome integrity by regulating the *de novo* DNA methylation of transposable and repetitive elements. It was shown that hetsiRNA species are predominantly loaded into the AGO4 in the cytoplasm and subsequently imported into the nucleus (17). Moreover, it was also demonstrated that some hetsiRNAs are also present as double-stranded siRNAs in low molecular weight fractions which correspond to the AGO-unbound sRNAs. Until now, studies mainly focused on the molecular details of sRNA biogenesis and sorting into the specific AGO proteins, but relatively limited attention has been devoted to the identification of functional pools of the full sRNA population of the cells. Analysis of total RNA extracts cannot distinguish between RISC incorporated and unincorporated sRNAs while immuno-precipitation experiments detect only sRNAs loaded into a given AGO protein. By applying a gel-filtration mediated size separation method, the association of sRNAs with various protein complexes, such as AGO containing RISCs, can be determined (18–21).

In this work, using gel-filtration assays we identified three pools of sRNAs at a genome-wide level based on the molecular size-dependent mobility of nucleoprotein complexes. High-throughput sequencing (HTS) analyses of these pools revealed that their molecular composition is markedly different confirming that they are distinct biological entities. We show that canonical miRNAs predominantly associate with a high molecular weight (HMW) RISC which co-migrates with AGO1. By providing HTS data we show that the 24-nt siRNAs are mainly associated with a low molecular weight (LMW) RISC co-migrating with AGO4. More intriguingly, we identified a large pool of protein-unbound sRNAs containing 2′-*O*-methylated mature and star sequences of annotated miRNAs potentially in a double-stranded form and also unloaded 23- and 24-nt siRNAs. We show that miRNAs differ in their RISC-loading efficiencies indicated by their different distribution between the AGO-bound and -unbound pools. Moreover, we demonstrate that

the RISC-loading efficiency of some miRNAs can be different between tissues. This can be partly attributed to the different level of AGO proteins because the loading level of miRNAs can be enhanced by overexpressing AGO proteins. We also show in transient and stable transgenic expression systems that the RISC-loading efficiencies of distinct miRNAs are predominantly controlled by their diverse precursor RNAs.

MATERIALS AND METHODS

Plant materials and growth conditions

Arabidopsis thaliana plants were grown in Jiffy peat blocks and kept under 8 h light/16 h darkness cycle at 21°C for 4 weeks. Next, these plants were planted in pots and were moved to a light room under 16 h light 8 h darkness at 21°C. Young leaves were harvested from six-weeks-old plants being in the vegetative phase. Young flowers and flower buds were collected from inflorescences which already have 2–6 immature siliques. *Nicotiana benthamiana* plants were grown under the lightroom conditions described above. The *hen1-1* mutant was ordered from the Arabidopsis Biological Resource Center. Plasmid construct of miR168, miR171, miR159 and miR390 hairpin structures (described below) were introduced into Col-0 plants with the floral dip method (22), transformants were selected according to their kanamycin resistance and homozygous lines were established through subsequent self-mating and monitoring of transgene expression.

Gel filtration assay

Size separation gel-filtration experiments were carried out using Superdex 200 10/300 column (Akta-FPLC, GE Healthcare) or the same size Sephacryl S-300 High Resolution (Pharmacia LKB) columns. The gel-filtration experiments were carried out as it was described previously (20,23). Briefly, the plant tissues were homogenized in liquid nitrogen using 200 μl elution buffer (50 mM Tris–HCl pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$ and 4 mM DTT) per 0.1 g plant material. The homogenized crude extracts were kept continuously in ice and centrifuged 3 times for 10 min (12 000 g) at 4°C to eliminate tissue debris. 200 μl of the cleared crude extract was immediately applied on the cold buffer equilibrated gel-filtration column and fractions of 325 μl were collected immediately after sample injection. The size-separation chromatography was always carried out in a cold room at 4°C. Altogether, we collected 68 fractions but discarded the first 20 and from the remaining 48 fractions we either purified total RNA to detect sRNAs or precipitated the proteins to detect endogenous or overexpressed AGO proteins. Markers of known sizes (Carbonic Anhydrase, 29 kDa; Albumin, 66 kDa; β-amylase, 200 kDa; apoferritin, 443 kDa; blue dextran, 2000 kDa) were used to calibrate the columns. Crude extracts were prepared from 0.3 g plant material collected from leaves of *N. benthamiana* plants three days post-infiltration, young rosette leaves of six weeks old *Arabidopsis thaliana* plants or *A. thaliana* flower buds. Protein-free small RNA extract for gel-filtration (Figure 1A) was prepared using the mirVana miRNA Isolation Kit (Thermo Fisher Scientific).

RNA and protein extraction

For miRNA analyses (Figures 5A, C and 6C) 0.1 g plant material were homogenized, in an ice-cold mortar, using 650 μl extraction buffer (0.1 M glycine–NaOH, pH 9.0, 100 mM NaCl, 10 mM EDTA, 2% sodium dodecyl sulfate, and 1% sodium lauroyl sarcosinate), for miRNA and protein analyses (Additional file 1: Supplementary Figure S3B) 0.05–0.2 g of plant material was collected and homogenized in 355 μl extraction buffer and divided into two aliquots. To one part (60 μl) one volume of 2× Laemmli buffer was added, boiled for five minutes, and was centrifuged at full speed for five minutes, and used as a protein sample. The remaining part was supplemented with 355 μl extraction buffer and was used for RNA extraction with the standard phenol-chloroform method. For analyses of gel-filtration fractions, odd number fractions were used for RNA while the even number fractions were used for protein extractions. For RNA extraction, 350 μl of phenol:chloroform (1:1) mixture was added to an equal volume of fraction, the aqueous phase was precipitated with ethanol, resuspended in 10 μl sterile water, and used for small RNA northern blotting. To investigate proteins 1200 μl cold acetone was added to each fraction, left at −70°C to precipitate, centrifuged, washed with 70% EtOH, dried and resuspended in 10 μl 2× Laemmli buffer and used for western blot analyses.

sRNA detection

For small RNA northern blot analyses (24), 4 μg of total RNA or samples of gel-filtration were separated on denaturing 12% polyacrylamide gels containing 8 M urea and transferred to Hybond-NX membrane (GE Healthcare) with semi-dry blotting (Bio-Rad). Membranes were chemically cross-linked (25) and probed with radiolabeled locked nucleic acid (LNA) oligonucleotide probes (Exiqon, Vedbaek, Denmark) or DNA probes, complementary to the mature miRNAs, as described previously (26). Signal was detected using X-ray film or Phosphorimager screen (Amersham). Loading rates of miR390 were calculated from the summarized volume intensity of the strongest four signals of fractions representing HMW RISC-bound and -unbound miR390 using ImageLab 1.1 (Bio-Rad).

Western blotting

Gel-filtration samples or 20 μl of *Arabidopsis* protein extracts were separated on 10 or 8% sodium dodecyl sulphate–polyacrylamide gel, blotted overnight to PVDF Transfer Membrane (Hybond-P; GE Healthcare, Freiburg, Germany) using wet tank transfer and subjected to western blot analysis. Membranes were blocked using 5% nonfat dry milk in phosphate-buffered saline (PBS) containing 0.05% Tween 20 (PBST) for 60 min. Blots were incubated with Actin, BiP or Histone H3 antibody (Agrisera, AS13 2640, AS09 481 and AS10 710) one hour or with anti-AGO1, anti-AGO2 or anti-AGO4 (Agrisera, AS09 527, AS13 2682, and AS09 617) 2.5 h at a dilution of 1:7500 in 1% milk powder (1× PBST). After washing in PBST, the membrane was incubated with secondary goat anti-rabbit IgG HRP conjugated antibody (Agrisera, AS09 602) 1 h at a dilution of 1:10 000 in 1× PBST with agitation. Blots were developed with High Clarity Western ECL (Bio-Rad), exposure was made using ChemiDoc (Bio-Rad) equipment in signal accumulation mode.

Plasmid constructs

All constructs were built using pGreen binary vector system (pGreen0029) and 35S cassette according to the manufacturer's instruction (http://www.pgreen.ac.uk). cDNA was produced with RevertAid First-strand cDNA Synthesis Kit (Thermo Fisher Scientific). Constructs of *ath-MIR168a*, *ath-MIR171a*, *ath-MIR159a* and *ath-MIR390* contained the hairpin region plus 10–10 bp upstream and downstream of the respective pre-miRNA. The *4m-Ago1* coding sequence was amplified and cloned from *4m-Ago1* transgenic plants, respectively. Primers used to create constructs are presented in Additional file 3: Supplementary Table S2. All constructs were introduced into *Agrobacterium tumefaciens* AGL1 strain with electroporation (360 Ω, 25 μF, 2.5 kV; Bio-Rad) in the presence of pSoup helper plasmid.

Transient assay

Young leaves of six weeks old *N. benthamiana* plants were infiltrated with a mixture of *Agrobacterium tumefaciens* (AGL1) suspensions at 1.0 optical density of 600 nm [OD_{600}] as described previously (23). P14 was included in the experiments as a suppressor of siRNA pathway (27). In the case of Figure 5A and B, full leaves were infiltrated with a mixture of constructs harbouring p14, pre-miRNA and empty pGreen0029 construct in a portion of 0.1:0.4:0.5 OD, respectively. In Figure 6A, the leaves were treated with a suspension containing pre-miR168, p14 and *4m-AGO1* constructs in 0.2:0.2:0.6 OD proportion or pre-miR168, p14 and empty pGreen0029 constructs in 0.2:0.2:0.6 OD proportion. Samples were harvested on the third day post-infiltration and pooled from patches of 3–4 separate leaves.

Purification of plant cell nuclei

Four gram young leaves of 6 weeks old *A. thaliana* Col-0 plants or 2 g mixed *A. thaliana* flowers were homogenized in an ice-cold mortar, and 40 ml of extraction buffer (10 mM Tris pH 7.5, 1.14 M sucrose, 5 mM $MgCl_2$, 7 mM mercaptoethanol) was added. This extract was carefully filtered through four layers of Miracloth (Merck, 475855-1R), and input samples were taken from the flow-through. Following 10 min centrifugation at 4°C with 900 g, samples were taken from the supernatant. The pellet was resuspended in extraction buffer supplemented with 0.15% Triton-X100 and centrifuged with the same parameters. This washing step was repeated twice, and the pellet was resuspended in 20 ml solution of 50 mM Tris pH 7.5, 5 mM $MgCl_2$ and 5 mM KCl. To this point, 200 and 100 μl of supernatants were used for RNA and protein extraction in every case, respectively. Finally, the purified nuclei were resuspended in 400 μl of this later solution. 100 μl and 50 μl of this was used for RNA and protein extraction, respectively.

High-throughput sequencing of sRNAs

RNA content of the gel-filtration fractions was extracted with Trizolate Reagent (UD-GenoMed Ltd., Debrecen,

Hungary) and fractions representing the identified sRNA pools were combined. The extracted RNA samples and also 30 μg of total RNA extracted directly from the crude extracts (input) were separated on an 8% denaturing polyacrylamide/urea gel along with RNA size markers, and then the sRNA range was isolated. Pellets were dissolved in 20 μl RNase-free ultrapure water, of which 2.5 μl was used for library preparation. The sRNA libraries were constructed using the TruSeq Small RNA Sample Prep Kit (Illumina, CA, USA), according to the manufacturer's instructions. The PCR-amplified, gel-purified and bar-coded cDNAs were submitted to UD-GenoMed Ltd (Debrecen, Hungary) for sequencing on Illumina HiSeq 2000 platform.

Bioinformatic analysis

The quality of the sequences was checked with FastQC 0.11.3 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Illumina TruSeq Small RNA adapters were trimmed from the raw reads using cutadapt v1.9.1 (28), with a minimum phred score of 20 and a size selection filter that allowed only the 20–25-nt-long sequences to pass. Untrimmed sequences were discarded. Sequences were mapped to the *A. thaliana* TAIR10 genome (without the mitochondrial and the chloroplastic sequences) using ShortStack v3.4 (29) with the default setting except that the '–bowtie_m' parameter was set to 1000. After the alignment, sequences mapped to the *Arabidopsis* rRNA and tRNA genes were removed from the alignment file using SAMtools v1.3.1 (30). The number and quality of the reads before and after the filtering steps were analyzed with FastQC. The read statistics are summarized in Additional file 2: Supplementary Table S1. Next, non-redundant, genome-matching reads were collected and their normalized abundances (read per million, RPM) were calculated for every library with an in-house script. Heat maps were generated with the pheatmap R package v1.0.12 (31) using the log$_2$-transformed, mean normalized expression values of the two biological replicates. Z-scores were calculated separately for the two tissues because the goal was to illustrate the distribution of a sequence within the fractions of the same gel-filtration event, yet be able to compare the distribution patterns between the tissues. For the miRNA heat maps, the sequences matching the *A. thaliana* mature miRNA sequences in the miRBase v21 database were extracted from the abundance table containing all the mapped, filtered sequences (it can be downloaded from GEO database) and filtered to retain only miRNAs having a mean abundance of at least 1 RPM across the FPLC-fractions in both tissues (input samples excluded) (Additional file 4: Supplementary Table S3A). The clustered abundance tables that contain sequences in the order as they appear in heat maps are in Additional file 4: Supplementary Table S3B. Sequence logos in Figure 3 were created with WebLogo (32). For siRNA analysis (Additional file 1: Supplementary Figure S5A), only the top 5000 abundant sRNAs were considered. This abundance table was annotated using patman v1.2 (33). MiRNAs were annotated using the known *A. thaliana* mature miRNAs in miRBase v21 (34), tasiRNAs were annotated using the tasiRNAdb database (35). All

sequences were annotated using the TAIR10 annotations and intergenic regions (Additional file 4: Supplementary Table S3C). For the siRNA heat maps, the top 5000 most abundant sequences were filtered to remove sequences that matched the mature miRNAs in miRBase and TAIR10 miRNA genes. Clustered abundance tables that contain the sequences in the order as they appear in heat maps are in Additional file 4: Supplementary Table S3d (21-nt siRNAs), and Additional file 4: Supplementary Table S3e (24-nt siRNAs). The principal component analysis was performed using the prcomp R package. For this, normalized expression values of the top 5000 most abundant sRNAs were log2-transformed and centered before the analysis. The plot was created using the ggfortify R package v0.4.6 (36). Genome browser tracks were created with BEDtools v2.26.0 (37) and visualized with the Integrated Genome Browser v9.0.0 (38). The raw sequences, the normalized abundance table, and the genome browser tracks (21- and 24-nt sRNAs) can be downloaded from the GEO database. Scripts used for the analysis are available at GitHub.

RESULTS

sRNAs are partitioned into three distinct pools in plant crude extracts

To identify the localization of sRNA in plant cells, total crude extracts were prepared from *Arabidopsis* leaves and young flowers with flower buds. These total crude extracts were applied on size separating gel-filtration column and the collected fractions were investigated for their AGO1 and AGO4 or small RNA content. In line with our previous findings (23), we identified the position of HMW RISC which co-fractionates with the majority of AGO1, the main executor of the miRNA pathway (39) (between 2000 and 669 kDa; Figure 1A). Small RNA northern blot hybridization of RNA extracts from the corresponding fractions also showed that the majority of miR159 co-localize with the HMW RISC. In addition, miR168 is also detected in the HMW RISC suggesting that HMW RISC is the main executor component of the miRNA pathway. We also found that ribosomal RNAs co-fractionate with AGO1 and miR159 indicating the polysomal association of HMW RISC (Additional file 1: Supplementary Figure S1). However, as we previously showed (23), the majority of miR168 is detected in low molecular weight fractions (at 29 kDa) possibly representing a protein-unbound miR168 pool (Figure 1A). In contrast to miR168, no high amount of miR159 was detected in the unbound sRNA fractions suggesting that different miRNAs exhibit various loading properties into the HMW RISCs. To reveal that these fractions represent *bona fide* protein-unbound sRNA species, the small RNA content of the leaf sample was extracted, deproteinized and loaded onto the gel-filtration column. Hybridization of the collected fractions with a probe detecting the highly conservative miR159 showed that the deproteinized miRNAs co-migrated with the suspected unbound miR168 of previous blots indicating that these fractions indeed represent AGO-unbound miRNA species (Figure 1A). To test the accumulation of AGO4, the main executor protein of the 24-nt-long siRNA mediated pathway, the fractions were analyzed for their AGO4 content. Similarly to
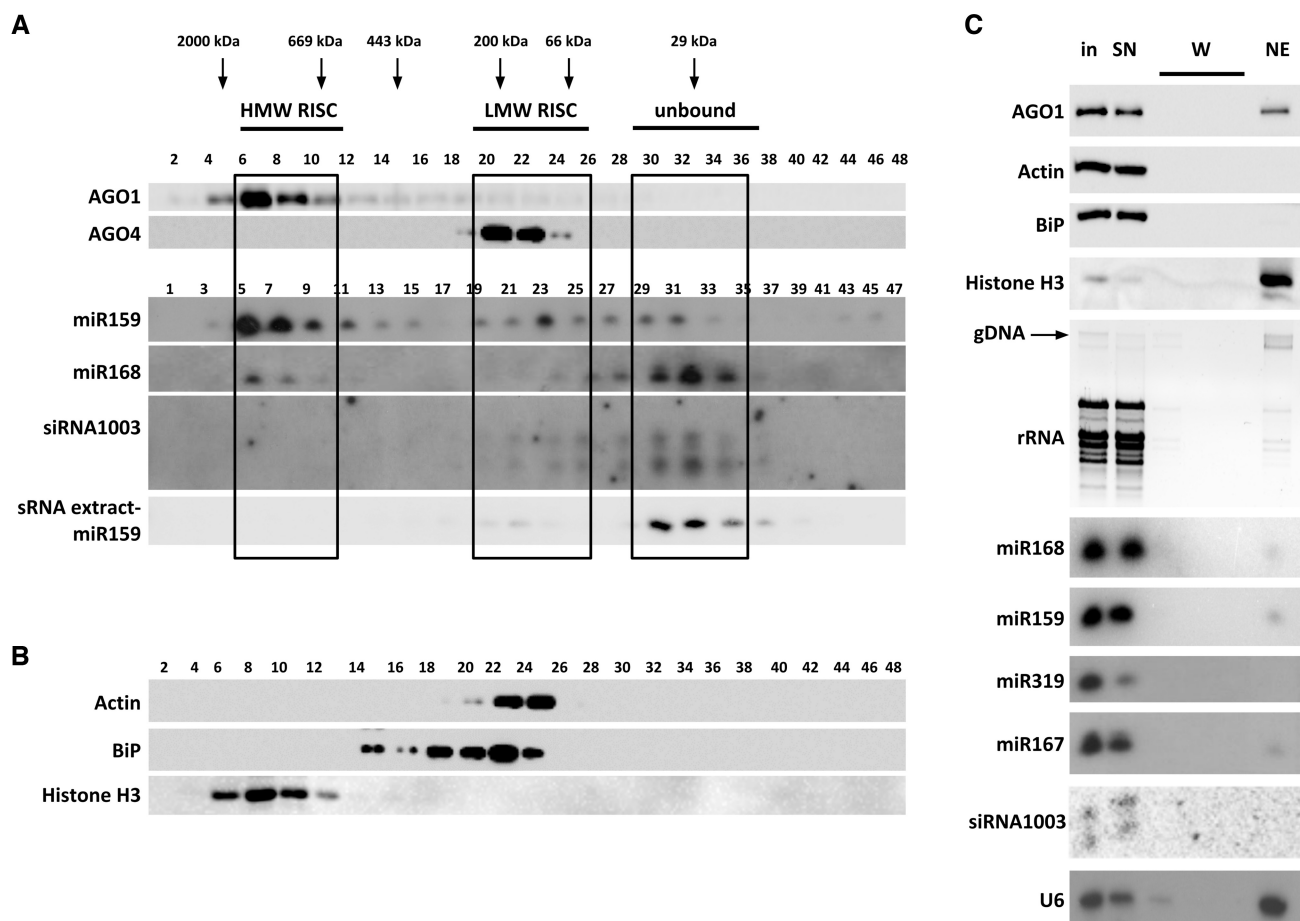
Figure 1. Identification of sRNA pools in the crude extract of *Arabidopsis thaliana* Col-0. (**A**) Distribution of AGO1 and AGO4 proteins (upper panels) and miR159, miR168 and siRNA1003 sRNAs (lower panels) in the high and low molecular weight complexes (designated as HMW and LMW RISC). sRNA extract represents a purified total, deproteinized sRNA content of crude extract loaded onto a gel-filtration column and hybridized for the presence of miR159. Black arrows: positions of known size markers; even numbers: protein fractions; odd numbers: RNA fractions; black frame: fractions representing HMW RISC, LMW RISC bound and protein-unbound sRNAs. Experiments were carried out using leaf samples, except for AGO4, which was detected in young flowers. (**B**) Western blots of gel-filtration fractions prepared from crude extract of Col-0 leaves to cytoplasmic (Actin), endoplasmic reticulum-localized (BiP) and nuclear (Histone H3) proteins. (**C**) AGO1, miR168, miR159, miR319, siRNA1003 and miR167 content of nuclear extract of leaf crude extract (NE). in, input; SN, supernatant after first centrifugation; W, washes. Cytoplasmic and endoplasmic reticulum-derived contamination was checked with Actin and BiP at the protein, while ribosomal RNA at the RNA level. Histone H3 and U6 were used to demonstrate nuclear content at the protein and RNA level, respectively. Black arrow, genomic DNA (gDNA); TNE, total nucleic acid extract.

a previous work (17) we detected AGO4 protein predominantly in fractions corresponding approximately to the size of single AGO proteins eluted between 200 and 66 KDa (designated as low molecular weight or LMW RISC; Figure 1A). In addition, in line with a previous work (17), we detected heterochromatic-siRNA 1003 in our experimental system in fractions co-migrating with AGO4 and representing protein unbound sRNAs (Figure 1A). We were not able to detect AGO4-specific signals in samples representing HMW RISC demonstrating that AGO4 is dominantly present in LMW RISC. In addition to the two experimentally investigated AGO proteins (AGO1 and AGO4), there are further eight in *Arabidopsis* which can also associate with sRNAs influencing their mobility properties during gel-filtration experiments. Based on these observations we identified three dominant pools of sRNAs in plant crude extracts: HMW RISC co-fractionating with AGO1, LMW

RISC co-fractionating with AGO4 and AGO-unbound sRNAs.

Cytoplasmic component of crude extracts is predominantly responsible for miRNA signals detected in gel-filtration experiments

To reveal the cellular composition of the crude extract used in our experiments, cytoplasmic and nuclear markers were also detected on gel-filtered samples (Figure 1B). As for cytoplasmic markers, BINDING IMMUNOGLOBULIN PROTEIN (BiP) localized in the endoplasmic reticulum lumen and ACTIN, the essential component of cell cytoskeleton were used. Both of them were present at a high level demonstrating the cytoplasmic content of the crude extract (Figure 1B). The 17 kDa nuclear specific marker HISTONE

3 (H3) also accumulated in high molecular weight complexes indicating that the crude extract also contains nuclear components and protein complexes remained intact during gel-filtration. To reveal the cellular origin of the signals associated with miRNAs, nuclei were purified from *Arabidopsis* leaves and the protein and sRNA contents were investigated (Figure 1C). We found that the purified nuclei did not exhibit strong signals of the investigated miRNAs. The majority of the observed low signals could be associated with cytosolic contamination of the sample indicated by the faint signals of ribosomal RNAs. We also found that in addition to its dominant accumulation in the cytoplasm, AGO1 is also present in the nuclei at higher amounts than expected from cytosolic contamination indicating that this nuclear pool of AGO1 is not loaded with the tested mature miRNAs. We found, in line with previous results (17), that the genome associated siRNA1003 was present predominantly in samples representing cytoplasmic components (Figure 1C; Additional file1: Supplementary Figure S2). Moreover, investigation of purified nuclei of *Arabidopsis* flowers indicated that AGO4 is also enriched in the cytoplasm (Additional file1: Supplementary Figure S2). Altogether, these data show that the crude extract used for gel-filtration contains both cytoplasmic and nuclear protein complexes, but the cytoplasmic component of the plant crude extracts is predominantly responsible for the miRNA signals detected during gel-filtration experiments.

HTS analysis reveals three sRNA pools with distinct sequence-length distribution profiles

To gain genome-wide information on the sRNA contents of the different pools, young *Arabidopsis thaliana* Col-0 leaves and flower samples containing mainly flower buds (Additional file 1: Supplementary Figure S3A) were subjected to size separating gel-filtration in two independent biological replicates. The purified RNA content of the fractions representing the three identified pools was separately collected, fractions 6–10 were combined for HMW RISC, 19–25 were combined for LMW RISC and 29–35 were combined for unbound sRNAs (Figure 1A). These combined RNA samples and an input total RNA sample originating directly from the crude extract were used for small RNA purification and subsequent HTS. Raw sequences were trimmed to remove adapters and low quality reads, and then filtered to keep only the 20–25-nt-long sequences. Filtered reads were mapped to the *A. thaliana* reference genome and sequences mapped to rRNA and tRNA genes were removed. The number of sequences obtained during the different steps of the process is summarized in Additional file 2: Supplementary Table S1. Abundances of the mapped sequences were calculated and normalized to make them comparable across the samples. The normalized abundance table containing all the filtered sRNA sequences have been deposited in the GEO database (for details, see the Data Availability section).

To get an initial assessment of the quality of our samples and biological replicates, we performed a principal component analysis (Additional file 1: Supplementary Figure S4). According to this, the PC1 clearly separates the samples by the tissue of origin (>58% of the variance between the sam-
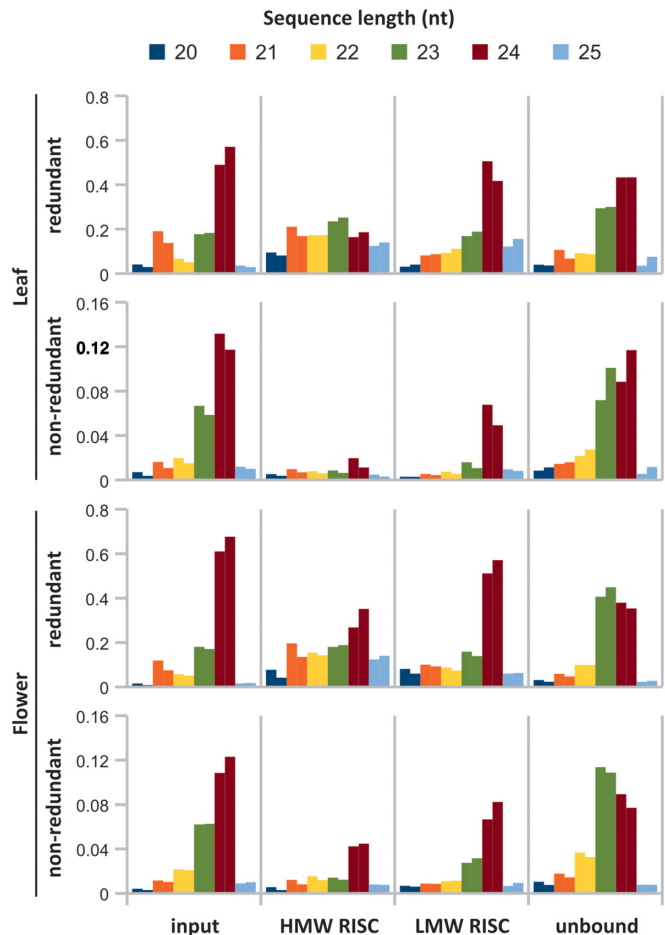

Figure 2. Sequence length distribution of sRNAs in the samples. Redundant sequences are the cumulative abundances of the unique sequences in the libraries with the given length. Abundances are normalized to the library size (read per million, RPM). Scales were modified for better visualization. For example, 0.5 of redundant 24-nt sRNAs means 500 000 24-nt sequences in a library with a size of one million total sequences of 20–25-nt length. The non-redundant sequences reflect the number of unique sequences of the given size class normalized to the redundant library size. If the redundant abundance is high but the non-redundant one is low, it means that there are only a few sequences present with a high abundance. On the other hand, if both abundances are high, it means that there are many different sequences with low expression levels. Columns with the same color represent the two biological replicates. Please note the different scales of the redundant and non-redundant abundances on the y-axes.

ples can be explained by this), while a further 26% variance can be explained by PC2, which separates the various combined fractions (PC2). The replicates are close to each other which reflects the good quality of our data. This analysis also suggests that the combined fractions represent different biological entities. Analyses of redundant reads revealed that the majority of the sRNAs in both leaf and flower input samples were 24-nt-long (Figure 2) which is a typical characteristic of plant small RNA profiles (40). We identified the 23-nt-long sRNA species as the second most abundant class in the input samples. The 21-nt-long and the 22-nt-long sRNAs accumulated at comparable levels. Analyses of the unique read distributions of leaf and flower input samples revealed similar sRNA size distribution profiles al-

though the proportion of 21-nt-long reads was clearly reduced indicating that this sRNA population contains less individual sRNA sequence with higher abundances (Figure 2). In contrast to input, the gel-filtrated samples show markedly different size distribution profiles. Analyses of redundant reads revealed that in HMW RISC the proportion of 21-nt-long reads increased markedly, while the presence of 24-nt-long sRNA species was drastically reduced (Figure 2). The relative ratio of 22- and 23-nt-long sRNA species also increased in HMW RISC sRNAs. Analyses of the HMW RISC nonredundant reads show a drastic reduction, especially in case of leaf sRNA classes, which again suggests that HMW RISC associated sRNA population consist of abundant reads with limited sequence diversities (Figure 2). In contrast, LMW RISCs were dominantly associated with 24-nt-long sRNAs but also the 21, 22 and 23-nt-long sRNA species were detectable in these samples. Similarly to the HMW RISC, the investigation of unique sequences revealed the profound reduction of reads. More intriguingly, sRNA species were detected with high abundance in the protein-unbound fraction of sRNAs (Figure 2). The 24 and 23-nt-long sRNAs were the most abundant classes in this sample but extensive amounts of 21- and 22-nt-long sRNAs were also present. The especially dominant presence of 23-nt long sRNAs indicates that this size class of sRNAs is mainly sorted into the biologically inactive unbound sRNA pool. The moderate reduction of nonredundant reads of this sample indicates that this sRNA population mainly contains diverse reads with low abundance.

Distribution of 21- and 24-nt-long siRNAs in the identified pools

RNA-directed DNA methylation is regulated by 24-nt-long heterochromatin-associated siRNAs loaded into their effector, the AGO4 protein (41). Previous works demonstrate that selected 24-nt-long siRNAs are associated with AGO4 in the cytoplasm and accumulate in the fractions corresponding to LMW RISC and also as unbound siRNA duplexes (17). In agreement with this finding, we observed that the majority of the 24-nt-long siRNAs are indeed associated with AGO4 containing LMW RISC and unbound sRNA pools (Additional file 1: Supplementary Figure S5A) both in the leaf and flower samples. To confirm the distribution of this class of siRNAs, we selected two representatives and investigated their distribution in gel-filtration fractions by northern blot analyses. In line with HTS data, we observed the presence of these 24-nt-long siRNAs mainly in the LMW RISC and unbound sRNA containing fractions (Additional file 1: Supplementary Figure S5B). In contrast to the 24-nt-long siRNAs, the predicted 21-nt-long siRNAs exhibited an enhanced association with the HMW RISC both in the leaf and flower samples, and they were also present in the protein-unbound fraction of sRNAs (Additional file 1: Supplementary Figure S5A).

Genome-wide quantitative distribution of the 21- and 24-nt-long sRNAs in the different sRNA pools were calculated and visualized in a genome browser (Additional file 1: Supplementary Figures S6 and S7, respectively). This allows the investigation of spatial correlation between the sRNAs and the different genomic features, and at the same time, an assessment of the RISC-loading efficiency of certain sRNA specimens.

miRNAs have various RISC-loading efficiencies

Next, we investigated the presence of the known *Arabidopsis* miRNAs in the identified pools. Heat maps were prepared to display the distribution of miRNAs in the different pools. For the convenient display, only miRNAs with a mean abundance above the threshold of 1 RPM were included in the heat map (Figure 3A). Data of less abundant miRNAs can also be found in Additional file 4: Supplementary Table S3A, while miRNA data displayed in the order as they appear in the heat maps are in Additional file 4: Supplementary Table S3B.

By hierarchical clustering analysis, we identified six clusters including two that represent tissue-specific miRNAs (Figure 3A). However, since we wanted to compare the RISC-loading efficiencies between the two tissues, we removed these two clusters from further analysis and investigated only those miRNAs that are present in both tissues. In Cluster A, miRNAs have moderately, while in Cluster C miRNAs have strongly limited ability to incorporate into the HMW RISC in the leaf, while a large portion of these miRNAs is present in the unbound sRNA pool. However, both in Cluster A and C, miRNAs show enhanced loading ability in the flower where AGO1, AGO2 and AGO4 proteins are represented in higher amount (Additional file 1: Supplementary Figure S3B). Cluster B contains miRNAs that are able to incorporate into the HMW RISC very efficiently in both tissues. Finally, miRNA sequences that accumulate mainly in the AGO-unbound pool both in the leaf and flower samples were sorted into Cluster D. These are almost exclusively miRNA* (passenger) strands. It is also apparent, that miRNAs are predominantly associated with the HMW RISC rather than the LMW RISC. This is in line with the dominant presence of AGO1, the main executor protein of the miRNA pathway, in the HMW RISC. Comparison of leaf and flower samples shows that the loading abilities of some miRNAs can be changed depending on the tissue of expression. The sequence logos of the Clusters (Figure 3B) do not indicate any common sequence motifs within the miRNA sequence itself which can be characteristic for the particular miRNA clusters. This observation suggests the existence of other, structural features on the miRNA precursors determining the loading efficiency of a particular miRNA. The majority of Cluster A–C miRNAs possess U at their 5′ end (Figure 3B) predicting the AGO1-dependence of their action. Moreover, heat map representation of 5′-U miRNAs without an abundance filter (Additional file 1: Supplementary Figure S8A) reveals various loading rates of miRNAs further supporting the regulated loading of miRNAs into AGO1 and suggesting that this phenomenon is not the consequence of sorting into different AGOs. We also could not detect a correlation between the absolute abundance of the miRNAs and their loading efficiencies (Additional file 1: Supplementary Figure S8B). Based on these data we revealed the abundant accumulation of unbound miRNAs which postulates a regulatory checkpoint determining RISC-loading efficiencies of particular miRNAs subsequently of their production.
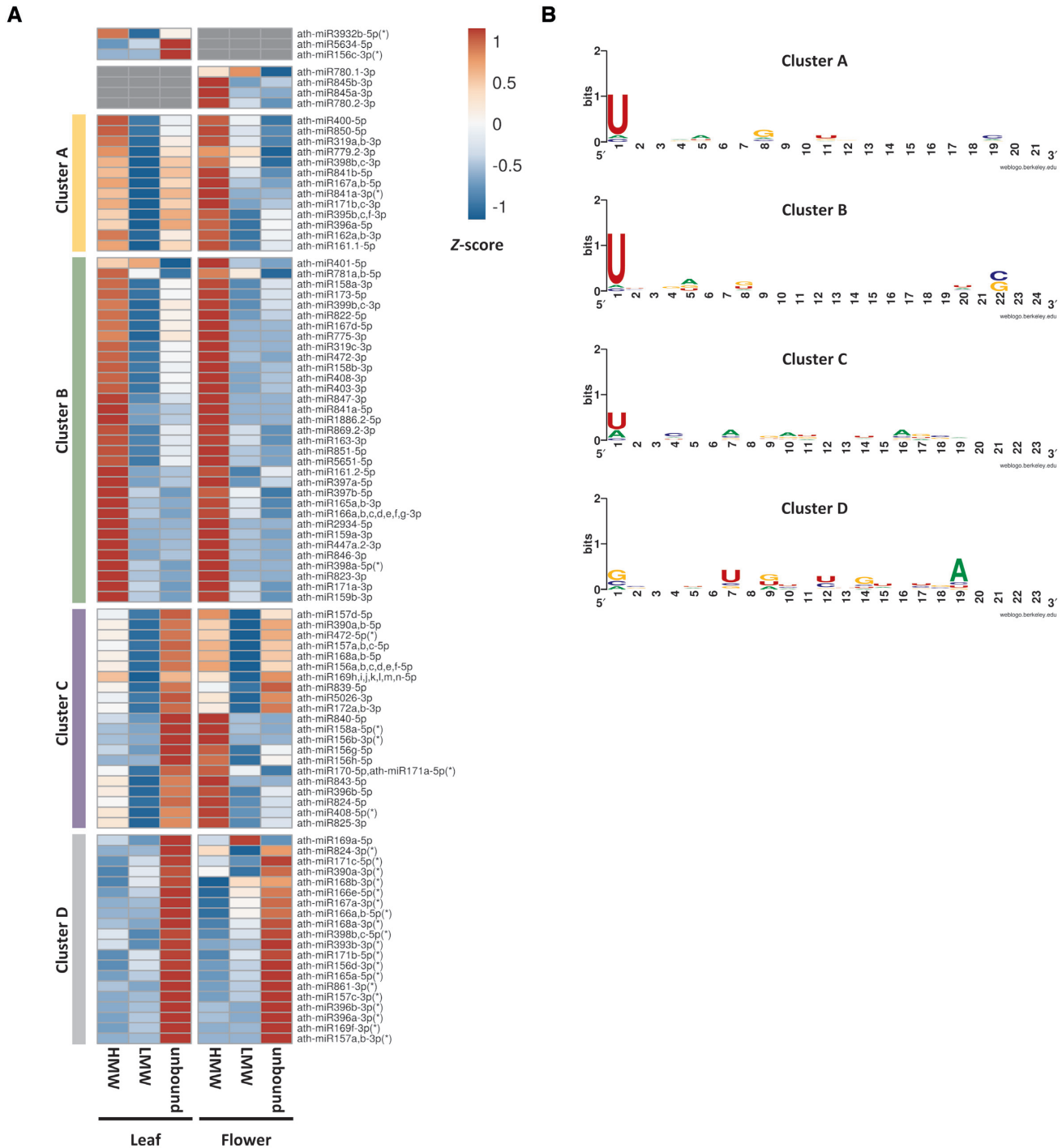
Figure 3. Relative miRNA abundances in the FPLC-fractions. (**A**) Only the *Arabidopsis thaliana* miRNAs in the miRBase v21 with a mean abundance of at least 1 RPM across the indicated samples are shown. The normalized expression values were log$_2$-transformed and *Z*-scores were calculated for the two tissues separately to indicate the relative abundance of a sequence across the related fractions. The *Z*-scores show how many standard deviations the given value is above (red) or below (blue) the mean (white) of all the values in the row in one tissue. The grey colour indicates 0 expression values. The miRNAs were clustered by their distribution patterns. For further analysis, we considered only the miRNAs that were present in both tissues. Cluster A contains the miRNAs which are moderately loaded into the HMW RISC in the leaf but efficiently loaded in the flower, Cluster B represents the miRNAs which efficiently load in both tissues, Cluster C contains miRNAs which poorly load in the leaf but have increased loading efficiency in the flower, while Cluster D contains the poorly loading miRNA* strands. (**B**) Sequence logos of the clusters indicate if there is an enrichment of a nucleotide in a given position along with the sequences.

Validation of the HTS data by northern blot analyses

To validate the HTS data experimentally, further gel-filtration experiments were carried out using crude extracts of young leaves and flower buds. RNA content of the collected fractions was used for small RNA northern blotting and subsequent hybridization with selected radiolabeled LNA and DNA probes. Coherently with the sequenced data, we experienced great variety in RISC-loading efficiencies of various miRNAs. In line with bioinformatic data, miR156-157 and miR390 showed extremely low RISC-loading rate and high-intensity signals in the pool of unbound miRNAs in leaves (Figure 4). In contrast, miR319 similarly to miR159 showed efficient RISC-loading since strong signals were detected in HMW RISC pool and only a moderate signal was detected in the protein-unbound pool. Other miRNAs such as miR161 and miR167 showed intermediate efficiency of loading since comparable signals of HMW RISC and protein-unbound pools were identified (Figure 4). Next, we investigated the presence of miRNA* strands in the three defined pools of miRNAs by hybridizing with probes specific for miR156*, miR168* and miR390* strands. In line with the bioinformatic results, we found that these miRNA* strands were present in the low molecular weight fractions suggesting that the AGO-unbound pool of miRNAs represents, at least partly, dsRNA species (Additional file 1: Supplementary Figure S9A). To test this hypothesis, a crude extract from *Arabidopsis* leaves was directly separated under non-denaturing conditions to preserve miRNA duplexes. We were able to detect the miRNA duplexes of miR168 and miR390, miRNAs which predominantly accumulate in the AGO-unbound pool but duplexes of miR159, exhibiting efficient HMW RISC-loading ability, were not detectable (Additional file 1: Supplementary Figure S9B). To reveal the methylation status of the unbound miRNAs, a crude extract prepared from *hen1-1* plants (42) was applied onto a gel-filtration column and the collected fractions were subjected to small RNA northern blot analysis using a probe specific to miR167. Since miR167 possesses an intermediate HMW RISC-loading characteristic (Figure 4) it is possible to detect the polyuridylation, characteristic to unmethylated miRNAs, of both the HMW RISC-loaded and protein-unbound miRNAs. In contrary to Col-0, in *hen1-1* both pools show the typical size increase of small RNAs due to the addition of U residues suggesting that in wild type-plants, RISC-loaded and AGO-unbound miRNAs are methylated as well. The unbound pool of miR167 showed slightly less efficient uridylation (Additional file 1: Supplementary Figure S9c). This phenomenon can be explained by the protective nature of the supposed dsRNA intermediate form of miR167 in the AGO-unbound pool. Comparison of gel-filtration experiments of vegetative and generative tissues revealed that many of the miRNAs have a higher level of HMW RISC incorporation in flowers than in leaves, i.e. miR157 or miR167 (Figure 4). These results confirm the accumulation of various miRNAs in the protein-unbound sRNA pool and indicate that loading efficiencies of miRNAs into the HMW RISCs are tightly regulated according to the cellular environment.

Overexpression of precursor RNAs confirms the altered HMW RISC loading abilities of different miRNAs

Our data suggest that there are miRNAs which despite their effective production from their precursors are not able to incorporate with high capacity into HMW RISCs. To experimentally validate the different loading efficiencies of various miRNAs, we selected three miRNAs exhibiting higher (miR159a, miR171a) and lower (miR168a) HMW RISC-loading capacities. The precursors of the selected miRNAs were cloned into binary plasmids and transformed into *Agrobacterium tumefaciens*. *Nicotiana benthamiana* leaves, a heterologous plant system, were infiltrated with *Agrobacterium* suspensions harbouring the individual miRNA precursors or the empty vector as a negative control. Leaf samples were collected at three days-post-infiltration (dpi) and used for RNA extraction or preparation of crude extract suitable for gel-filtration experiments. First, we assessed the overexpression levels of miRNAs in the various infiltrated leaves by small RNA northern blot analyses. We found that all the three miRNAs were massively overexpressed in the infiltrated patches compared to the control infiltrations (Figure 5A). Next, we investigated the HMW RISC-loading abilities of different miRNAs by gel-filtration assays. We found that overexpression of miR168 resulted in an extremely high-level accumulation of protein-unbound miR168 species. In contrast, the overexpression of miR159 was accompanied by efficient HMW RISC-loading and only a smaller portion of miR159 was present in the fractions representing protein-unbound sRNAs (Figure 5B). In the case of miR171 overexpression, we experienced an intermediate HMW RISC-loading ability since, in addition to the high-level accumulation of unbound miR171 species, we observed a marked level of miR171 loaded into the HMW RISC. The availability of functional RISCs in miR171 and miR168 overexpressing samples was investigated with the re-hybridization of gel-filtration blots with probe detecting miR159 while the presence of unbound miRNAs in miR159 overexpressing sample was confirmed by the detection of endogenous miR168 (Additional file 1: Supplementary Figure S10A). To further confirm the results of the transient *N. benthamiana* studies, we stably overexpressed the same miRNA precursors by establishing homozygous transgenic *Arabidopsis* lines. We first assessed the level of overexpression of particular miRNAs in selected lines (Figure 5C) than investigated the HMW RISC loading properties of overproduced miRNAs by gel-filtration assays. In the transgenic plants, the overexpressed miR168 exhibited very poor, miR171 intermediate, while miR159 very efficient HMW RISC-loading (Figure 5D). The availability of functional RISCs in miR168 overexpressing plants was investigated with the re-hybridization of gel-filtration blot with probe detecting miR159 (Additional file1: Supplementary Figure S10B). These findings are very similar to the results of the transient assays further supporting the central role of diverse miRNA precursor RNAs in determining the HMW RISC loading efficiencies of various miRNAs. These observations confirm our results about the altered HMW RISC-loading capacities of individual miRNAs since the exper-
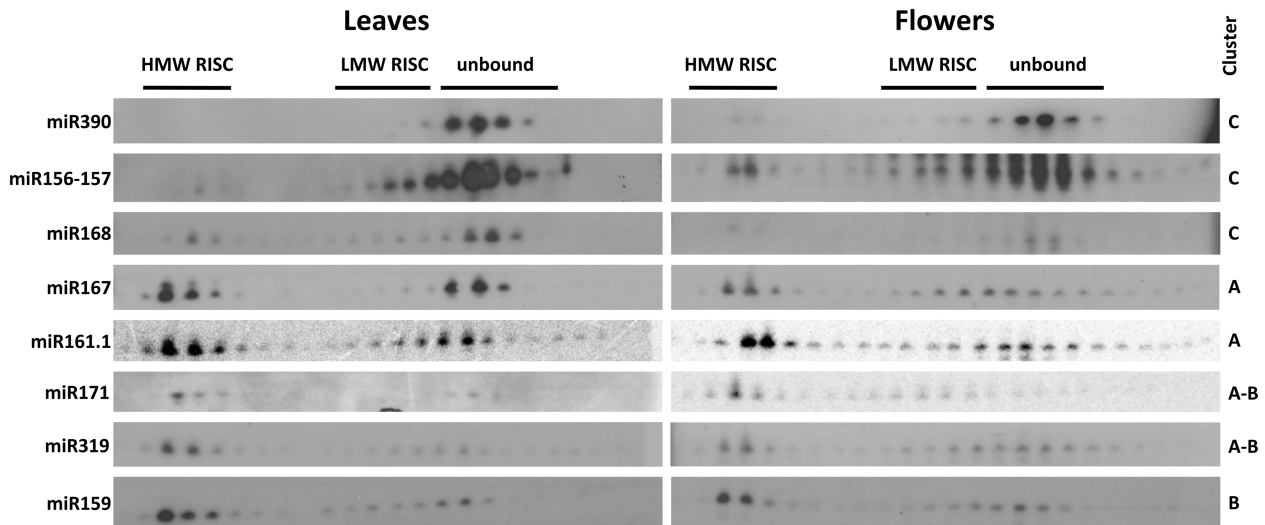
Figure 4. Validation of the HTS results with small RNA northern blots. Blots of gel-filtrated fractions were prepared from leaves and young flowers of *Arabidopsis thaliana*. Blots were sorted according to the RISC-loading efficiency of the miRNAs in leaves. Probes can also detect other miRNA family members. HTS cluster categories were presented on the right side of the panels. Indication of two categories means the miRNA family members are present in different clusters.

imental results are in line with the previous bioinformatic and gel-filtration data. These results also exclude the possibility that the altered tissue-specific expressions of particular miRNAs and AGO proteins are the main cause of the presence of protein-unbound miRNA species since the *Agrobacterium*-mediated transient expression studies were carried out in a similar cellular environment.

The availability of AGO proteins is the limiting factor of miRNA loading

Bioinformatic analysis of our HTS data showed that numerous miRNAs with U at their 5′ end, directing AGO1 sorting, show restricted RISC-loading efficiencies. This suggests a competition mechanism between miRNAs for limiting AGO1 proteins, the main executor component of HMW RISC. To test this hypothesis, we transiently overexpressed miR168 in *N. benthamiana* leaves in the presence or absence of transiently overexpressed *4m-AGO1*, a functional, miR168 regulation resistant version of AGO1 (43) (Figure 6A). Gel-filtration experiments revealed that in the excess of AGO1 protein more miR168 were loaded into the HMW RISC indicating that there is an available active reservoir of miR168. Very similar results were found in transgenic plants overexpressing *4m-AGO1* (43) where the *AGO1* mRNA overexpression resulted in the enhanced HMW RISC-loading of miR167 compared to the wild plant (Figure 6B). These results also imply that miR167 species sorted into the unbound pool in wild type plants can be directed into HMW RISC in *4m-AGO1* plants. To test the hypothesis of AGO limitation of miRNA loading in another system, we examined the specific miR390-AGO7 interaction as well. It was shown that in contrast to the broad expression domains of *MIR390a* and *MIR390b*, *AGO7* mRNA expression was confined to the vascular tissues indicating that miR390 is active only in those cells that co-express AGO7 (44). In line with this theory, we found that miR390 accu-

mulated predominantly in the protein-unbound pool of sRNAs (Figure 4). High level of miR390 overexpression resulted only in a moderate increase in HMW RISC-loading rate and the vast amount of the overexpressed miR390 accumulated in the unbound pool (Figure 6C) probably due to the limited availability of AGO7. In contrary, in AGO7 protein overexpressing plants (44) (Additional file 1: Supplementary Figure S11) we detected the markedly promoted loading of miR390 into the HMW RISC (Figure 6C). These data show that the availability of AGO proteins represents a major bottleneck in determining the levels of biologically active, RISC-loaded miRNAs and that the excess miRNAs can be sorted into RISCs when more AGO proteins are available.

DISCUSSION

RNAi pathways, mediated by the action of miRNAs, are pivotal regulators of developmental programmes and alleviation of biotic and abiotic stresses. Consequently, the expression and biological activity of miRNAs must be spatiotemporally and quantitatively coordinated. Since MIR genes are transcribed into primary miRNAs (pri-miRNAs) by RNA polymerase II (Pol II), and similarly to protein-encoding genes, *cis*-regulatory elements and *trans*-acting regulators play fundamental roles in their transcriptional regulation (45). Due to this control, miRNAs often exhibit spatially highly coordinated, tissue-specific expression patterns (46–48) and various miRNAs are able to dynamically respond to environmental stresses (49). The transcriptional regulation of MIR genes is followed by multiple levels of post-transcriptional regulatory steps influencing the stability, production rate and biological activity of generated mature miRNAs (50). The 5′ capped and 3′ polyadenylated long pri-miRNAs form an imperfect stem-loop structure and are subsequently processed by DCL1 to precursor miRNA (pre-miRNA) possessing specific stem-
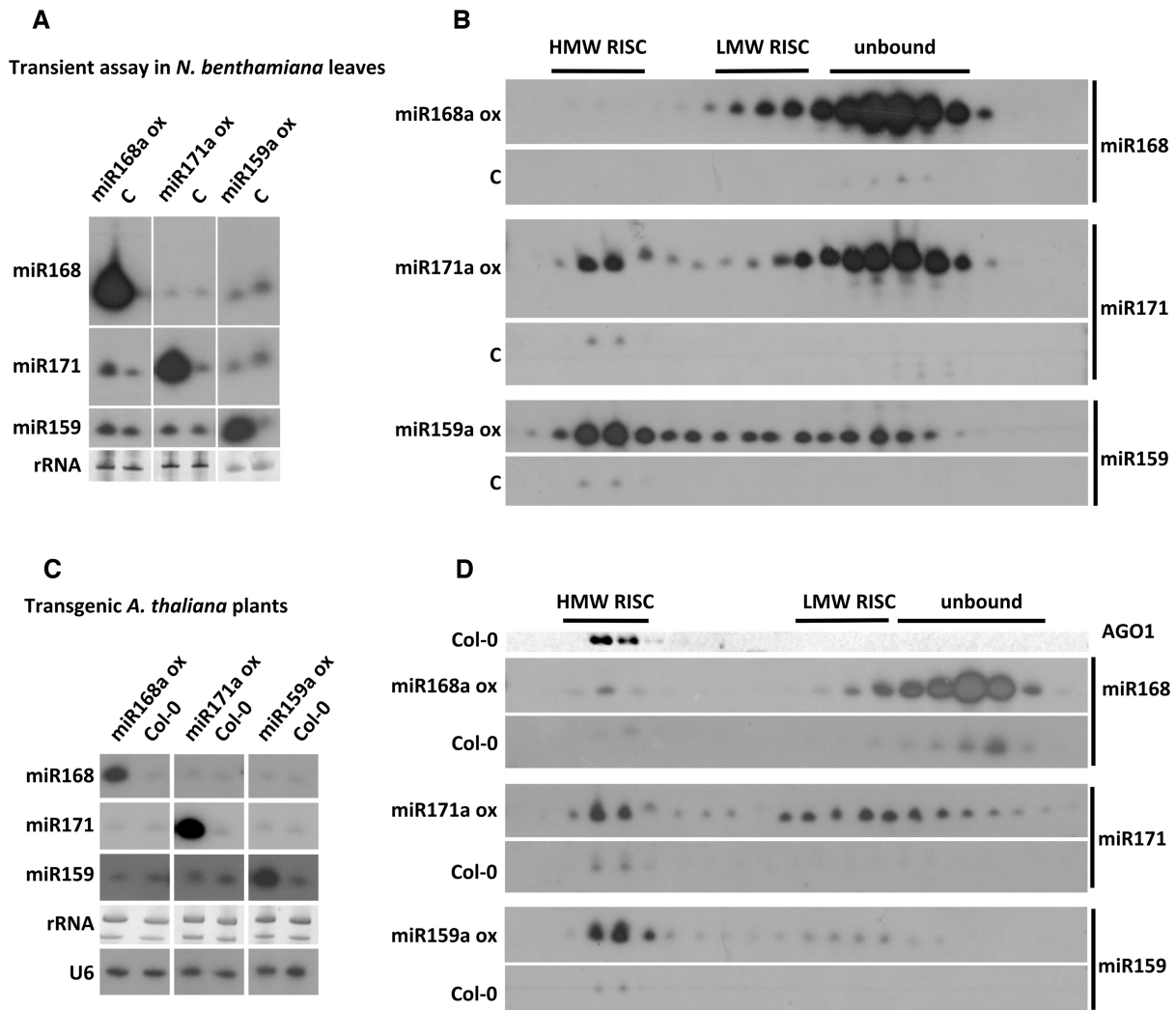
Figure 5. Transient and stable transgenic overexpression of ath-miR168a, ath-miR171a and ath-miR159a. (**A**) Overexpression rates compared to control *Nicotiana benthamiana* leaves infiltrated with empty vector (C). As a loading control, membranes were washed and re-probed for miR168, miR171 or miR159 according to the particular experiment. (**B**) Distribution of miR168, miR171 and miR159 in gel-filtration experiments prepared from respective miRNA overexpressing transient assays and control *Nicotiana benthamiana* leaves infiltrated with empty pGreen0029 vector (C). (**C**) Overexpression rates of transgenic lines compared to control *A. thaliana Columbia* plants (Col-0). As a loading control, membranes were washed and re-probed for miR168, miR171 or miR159 according to the particular experiment. (**D**) Distribution of miR168, miR171 and miR159 in gel-filtration experiments prepared from young leaves of respective miRNA overexpressing transgenic plants and control Col-0 plants. AGO1 was detected using gel-filtration fractions of Col-0.

loop structures. These pre-miRNAs are further processed in the nucleus to miRNA intermediate duplexes consisting of the guide strands (mature miRNA) and the passenger or star strands (miRNA*). In contrast to animal miRNA precursors, which are relatively uniform in size and structure defining the precise cleavage of pre-miRNA stem-loops (51) plant pre-miRNAs are highly diverse in size and structure (5). Maturation of selected plant pre-miRNAs was shown to be strictly coordinated by various structural elements of stem–loop structures (52–54), however, the extremely diverse structure of plant pri- and pre-miRNAs could hide undiscovered biologically important regulatory, structural or sequence features. According to our knowledge gained so far, the produced mature miRNAs are either loaded into AGO containing RISCs or degraded. In this work, we revealed a new regulatory layer acting at post-production of

miRNAs by adjusting the RISC-loading efficiencies of individual miRNAs. Gel-filtration based size separation chromatography is a suitable tool to investigate the large population of sRNAs in interaction with the different RISCs simultaneously. Using this method intact large protein complexes can be separated and investigated. Gel-filtration of plant crude extracts allowed us to separate HMW (co-migrating with AGO1), LMW (co-migrating with AGO4) RISCs and more intriguingly, an unexpected large pool of protein-unbound sRNAs. A previous study using gel-filtration based approach detected AGO1 in a monomer form (18), however, in this experiment FLAG-AGO1 was first immuno-purified from crude extracts and then applied onto the gel-filtration column. It is possible that the HMW RISCs disintegrate under the experimental procedure of immunoprecipitation. Other works using a similar experimen-
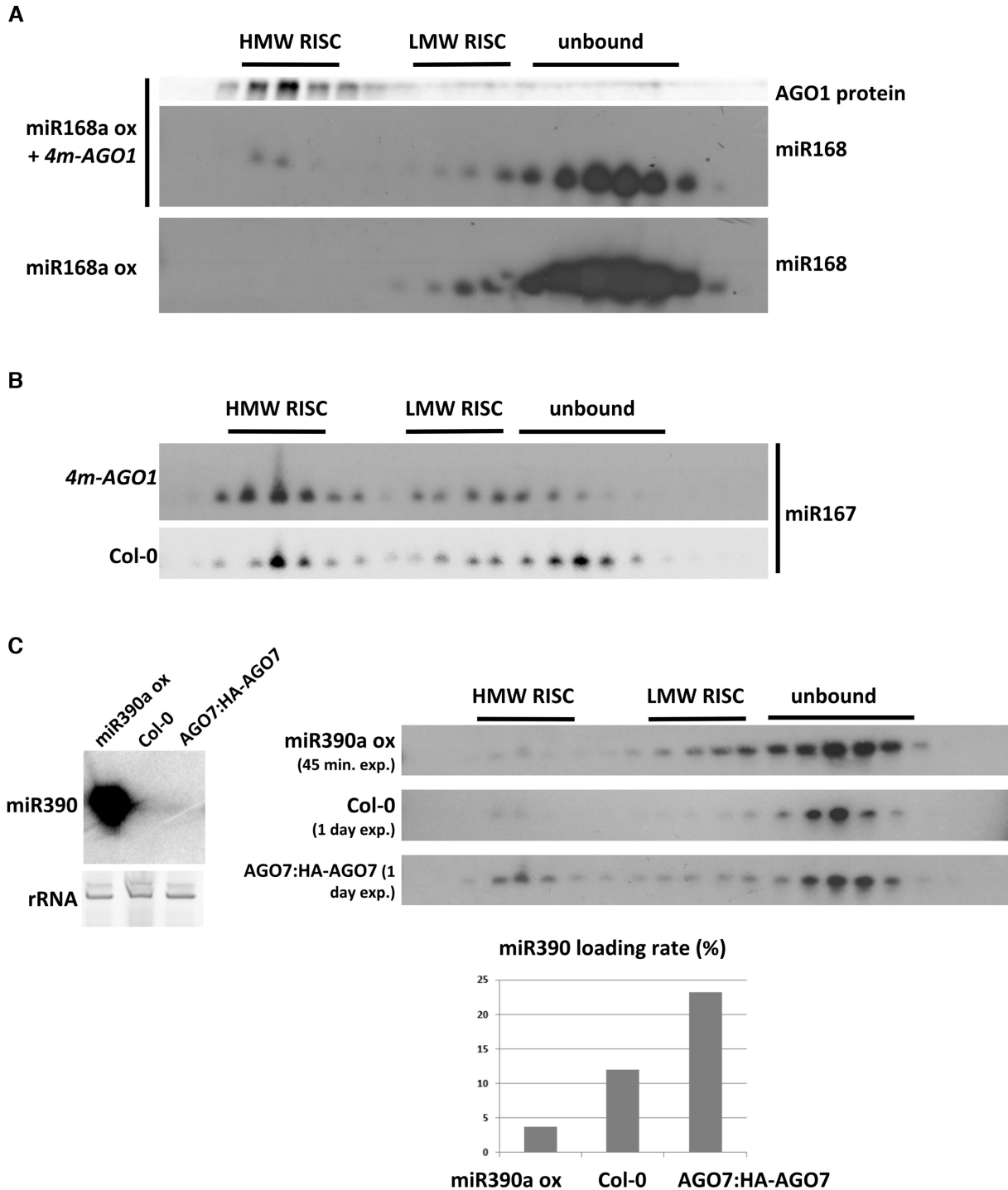
A

B

C

Figure 6. Effect of AGO overexpression on the RISC-loading efficiency. (**A**) Transient overexpression of miR168a with or without *4m-AGO1*. miR168 northern blots of the two infiltrations were handled in parallel. AGO1 Western was prepared from the presented co-infiltration of miR168a and *4m-AGO1*. (**B**) miR167 hybridization of gel-filtrations prepared from leaves of transgenic *4m-AGO1* and wild type Col-0 *Arabidopsis* plants. Experiments and blots were handled in parallel, and the exposure time was the same. (**C**) Distribution of miR390 among RISC-loaded and unbound fractions in miR390a overexpressing, wild type Col-0, and AGO7:HA-AGO7 transgenic plants. The left panel shows miR390 overexpression rate in transgenic leaves with northern hybridization. Gel-filtration was carried out from flowers. Images with similar signal intensity were taken from blots of gel-filtrations. To achieve similar signal intensities, we set exposure time 45 minutes for miR390 overexpressing samples, and one day for Col-0 and AGO7:HA-AGO7. The same blot is displayed for wild type Col-0 as in the case of Figure 4. Calculated RISC-loading efficiencies (%) are presented in the diagram below.
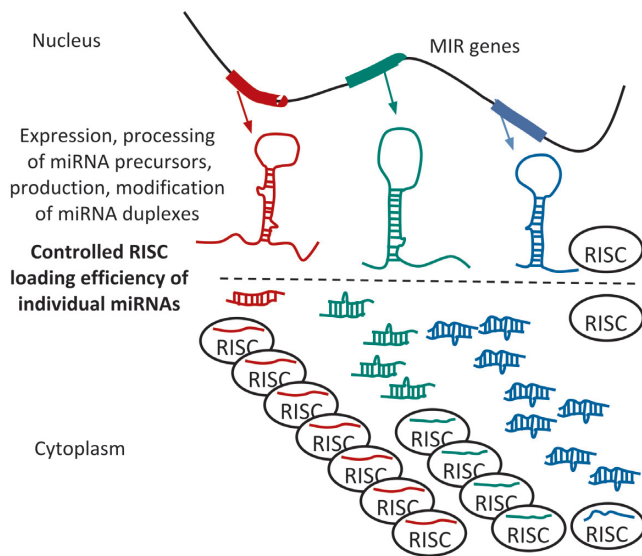
Figure 7. Schematic representation of the proposed model. Genome-encoded MIR genes (red, green and blue boxes) are transcribed to various, extremely diverse miRNA precursors (hairpin structures) according to the regulated activities of their promoters. The different miRNA precursors are processed to miRNA intermediate duplexes containing the mature and passenger (*) miRNA strands. These miRNAs undergo methylation at the 3′ nucleotide. The next step is the selective loading of the mature miRNA stands (red, green and blue curved lines) into the executor complexes, the AGO1 containing RISCs (black circles), while the miRNA* strands are eliminated. However, here, a newly identified regulatory mechanism, likely controlled by information carried on the diverse miRNA precursors, determines the RISC-loading efficiencies of various miRNAs sorting only a subset of the produced miRNAs into the biologically active RISCs. The superfluous miRNA populations accumulate in the cytoplasm as protein-unbound pool very likely as miRNA:miRNA* duplexes. Due to this regulation, miRNAs have high (red), intermediate (green), or low (blue) RISC-loading capacities in the given cellular environment. The availability of unloaded RISCs can influence loading properties of miRNAs suggesting that excess of sRNA population compete for limiting empty RISCs. The place of RISC-loading is not known, it can be either in the cytoplasm or nucleus (empty RISC circles). It is also unknown whether the cytoplasmic pool of protein-unbound miRNAs can be redirected to RISCs or they represent biologically inactive products. Altogether, this model suggests that the production rate of certain miRNAs is not the only factor which determines their biological activity. A highly controlled post-production regulatory mechanism can determine the biologically active portion of the produced miRNA population by adjusting their RISC-loading efficiencies in the given cellular environment.

tal method also detected AGO1 in a monomer form (17,19). The differences in the experimental conditions in these studies may explain the dominant presence of AGO1 in fractions representing the monomer form of AGO1. Moreover, Ye *et al.* did not investigate the fractions corresponding to the HMW complexes (17). In our experiments, we used very young intensively developing tissue samples and tried to apply the mildest experimental treatment of the samples before loading them onto the columns and repeatedly detected the dominant accumulation of AGO1 in high molecular weight complexes likely co-migrating with polysomes consistently with previous results (7,8,55). This does not exclude that some portion of AGO1 can exist in monomer form but it is under the detection level of our system. High-throughput sequencing of the sRNA content of gel-filtrated plant crude extracts made it possible to identify the RISC-

bound and protein-unbound miRNA pools genome-wide in *Arabidopsis*. Moreover, this experimental approach allowed us to assess the ratios of AGO1 co-fractionating HMW RISC-bound versus AGO-unbound pools of individual miRNAs. Based on these observations, we identified miRNA species with U at their 5′ end, characteristic for miRNAs sorting into AGO1, which did not load efficiently (e.g. miR156, miR168) or only with intermediate (e.g. miR161, miR167) efficiency into the HMW RISC. The unloaded portion of these poorly loading miRNAs are present in the cytoplasm, co-migrating with deproteinized sRNAs, probably as methylated miRNA/miRNA* duplexes, indicated by their migration in native agarose gel, the polyuridylation of protein-unbound miRNAs in *hen1-1* plants, and the presence of miRNA* strands in this pool. This is consistent with the previous finding where two miRNAs (miR159 and miR165) were detected at low molecular weight fractions (19) but the origin or importance of the observed signals was not commented. These data show that miRNAs which are abundantly produced are not necessarily fully loaded into executor complexes (RISCs). Similar results were found in the case of 24-nt siRNAs since they were found to be associated mainly with LMW RISC and also with unbound sRNA pool. This result is in line with a previous study where it was shown that 24-nt-siRNA species were associated with AGO4 in the cytoplasm eluting from gel-filtration column between 160 and 66 KDa, at the size of LMW RISC, but they were also present as ds siRNAs in low molecular weight fractions corresponding to protein-unbound sRNAs (17). One hypothetical explanation for the abundant accumulation of these miRNAs in AGO-unbound samples could be the spatially different expression pattern of miRNAs and AGO1 protein. However, this is not likely since AGO1 shows widespread expression in different tissue types throughout the developmental process providing the executor protein continuously (39,56). Another possibility is that different miRNAs have an altered ability to load into the AGO1 containing HMW RISCs. To test this hypothesis, we carried out overexpression of three different selected pre-miRNAs, exhibiting various HMW RISC-loading efficiencies, transiently in *N. benthamiana* leaves and in stable transgenic *Arabidopsis* lines. We revealed that the overexpressed miRNAs behaved differently; the efficiently loading miR159 was detected mainly in HMW RISC, miR171 exhibiting intermediate loading ability was present in HMW RISC and also in the unbound sRNA pool, while miR168 with strongly restricted ability to load into HMW RISC accumulated predominantly in the unbound sRNA pool and only a minority of the overexpressed miR168 was loaded into the HMW RISC. The altered HMW RISC-loading of miRNAs in the same tissue environment indicates that structural, sequence and/or *trans*-factors associated with pre-miRNAs orchestrate the HMW RISC-loading ability of individual miRNAs determining the biologically active, loaded portion of the total miRNAs produced. The elevated loading rate of many miRNAs into HMW RISC in flowers (e.g. miR167, miR157) where AGO1 protein is present at a high level (39) postulates that the accessibility of AGO1 proteins can be a limiting factor. To confirm this assumption we investigated the HMW RISC loading efficiency of the poorly and moderately incor-

porating miR168 and miR167, respectively, in transient and transgenic expression systems, where the miR168 resistant form of AGO1 protein was overproduced. In both cases, we experienced increased loading of the investigated miR-NAs indicating that the availability of AGO proteins can indeed determine the portion of loaded miRNAs utilising the biologically active reservoir of miRNAs. Based on these observations we can hypothesize a regulatory mechanism where abundant production of miRNA species is followed by a competition for unloaded AGO1 pool. Biologically active portions of miRNAs are not solely determined by their production rates but also by the pre-miRNA-driven HMW RISC-loading efficiencies of individual miRNAs. The loading properties of miRNAs are probably further modulated by the competing sRNA population present in the given cellular environment. Previously, it was thought that miRNA duplexes are exported partly by HASTY to the cytoplasm and are loaded into AGO1 here (57). However, a recent finding suggests an alternative mechanism where AGO1 loading takes place in the nucleus (6). AGO-limited RISC-loading was also demonstrated in the miR390 - AGO7 relationship. However, in this case, the spatially different regulation of AGO7 and miR390 expressions can be the reason for the massive accumulation of unbound miR390 pool. This is supported by previous observations demonstrating that while *MIR390a* and *MIR390b* exhibit spatially wide expression, *AGO7* mRNAs express only in the vascular tissues (44). Based on our findings we suggest a model for a new regulatory mechanism which determines the biological activity of miRNAs by controlling their RISC loading efficiencies through the action of diverse precursor RNAs (Figure 7). We also think that our results can be useful from a practical point of view for scientists designing artificial (a)miRNAs (precursor backbone selection, modification for optimal results, enhancing or reducing amiRNA activity). Further experiments are needed to understand the detailed molecular mechanisms and spatial distributions lying behind the co-ordinated HMW RISC-loading of miR-NAs. It will be also important to understand whether the unbound sRNAs retained their biological activities and can be redirected to RISC-loading or they are partly or completely biologically inactive by-products of previous regulatory steps determining the ratio of biologically active miR-NAs.

DATA AVAILABILITY

The raw sequencing data, the normalized expression table of all the genome-mapped small RNAs, and genome browser tracks for the 21- and 24-nt sRNAs have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus [https://www.ncbi.nlm.nih.gov/geo] under accession number GSE130431. All the scripts used for the analysis and to generate figures are available at GitHub [https://github.com/gyulap/sRNA_content_of_FPLC_fractions].

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Borges,F. and Martienssen,R.A. (2015) The expanding world of small RNAs in plants. *Nat. Rev. Mol. Cell Biol.*, **16**, 727–741.
2. Rogers,K. and Chen,X. (2013) Biogenesis, turnover, and mode of action of plant microRNAs. *Plant Cell*, **25**, 2383–2399.
3. Singh,A., Gautam,V., Singh,S., Sarkar Das,S., Verma,S., Mishra,V., Mukherjee,S. and Sarkar,A.K. (2018) Plant small RNAs: advancement in the understanding of biogenesis and role in plant development. *Planta*, **248**, 545–558.
4. Yu,Y., Jia,T. and Chen,X. (2017) The 'how' and 'where' of plant microRNAs. *New Phytol.*, **216**, 1002–1017.
5. Song,L., Axtell,M.J. and Fedoroff,N.V. (2010) RNA secondary structural determinants of miRNA precursor processing in Arabidopsis. *Curr. Biol.*, **20**, 37–41.
6. Bologna,N.G., Iselin,R., Abriata,L.A., Sarazin,A., Pumplin,N., Jay,F., Grentzinger,T., Peraro,M.D. and Voinnet,O. (2018) Nucleo-cytosolic shuttling of ARGONAUTE1 prompts a revised model of the plant microRNA pathway. *Mol. Cell*, **69**, 709–719.
7. Lanet,E., Delannoy,E., Sormani,R., Floris,M., Brodersen,P., Crété,P., Voinnet,O. and Robaglia,C. (2009) Biochemical evidence for translational repression by Arabidopsis microRNAs. *Plant Cell*, **21**, 1762–1768.
8. Li,S., Le,B., Ma,X., Li,S., You,C., Yu,Y., Zhang,B., Liu,L., Gao,L., Shi,T. *et al.* (2016) Biogenesis of phased siRNAs on membrane-bound polysomes in Arabidopsis. *eLife*, **5**, e22750.
9. Park,W., Li,J., Song,R., Messing,J. and Chen,X. (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in Arabidopsis thaliana. *Curr. Biol.*, **12**, 1484–1495.
10. Ramachandran,V. and Chen,X. (2008) Degradation of microRNAs by a family of exoribonucleases in Arabidopsis. *Science*, **321**, 1490–1492.
11. Dolata,J., Bajczyk,M., Bielewicz,D., Niedojadlo,K., Niedojadlo,J., Pietrykowska,H., Walczak,W., Szweykowska-Kulinska,Z. and Jarmolowski,A. (2016) Salt stress reveals a new role for ARGONAUTE1 in miRNA biogenesis at the transcriptional and posttranscriptional levels. *Plant Physiol.*, **172**, 297–312.
12. Ma,Z. and Zhang,X. (2018) Actions of plant Argonautes: predictable or unpredictable? *Curr. Opin. Plant Biol.*, **45**, 59–67.
13. Mi,S., Cai,T., Hu,Y., Chen,Y., Hodges,E., Ni,F., Wu,L., Li,S., Zhou,H., Long,C. *et al.* (2008) Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5′ terminal nucleotide. *Cell*, **133**, 116–127.
14. Zhang,X., Niu,D., Carbonell,A., Wang,A., Lee,A., Tun,V., Wang,Z., Carrington,J.C., Chang,C.-E.A. and Jin,H. (2014) ARGONAUTE PIWI domain and microRNA duplex structure regulate small RNA sorting in Arabidopsis. *Nat. Commun.*, **5**, 5468.
15. Cuperus,J.T., Carbonell,A., Fahlgren,N., Garcia-Ruiz,H., Burke,R.T., Takeda,A., Sullivan,C.M., Gilbert,S.D., Montgomery,T.A. and Carrington,J.C. (2010) Unique functionality of 22 nt miRNAs in triggering RDR6-Dependent siRNA biogenesis from target transcripts in arabidopsis. *Nat. Struct. Mol. Biol.*, **17**, 997–1003.
16. Fei,Q., Xia,R. and Meyers,B.C. (2013) Phased, secondary, small interfering RNAs in posttranscriptional regulatory networks. *Plant Cell*, **25**, 2400–2415.

17. Ye,R., Wang,W., Iki,T., Liu,C., Wu,Y., Ishikawa,M., Zhou,X. and Qi,Y. (2012) Cytoplasmic assembly and selective nuclear import of Arabidopsis ARGONAUTE4/siRNA complexes. *Mol. Cell*, **46**, 859–870.

18. Baumberger,N. and Baulcombe,D.C. (2005) Arabidopsis ARGONAUTE1 is an RNA Slicer that selectively recruits microRNAs and short interfering RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 11928–11933.

19. Wang,H., Zhang,X., Liu,J., Kiba,T., Woo,J., Ojo,T., Hafner,M., Tuschl,T., Chua,N.-H. and Wang,X.-J. (2011) Deep sequencing of small RNAs specifically associated with Arabidopsis AGO1 and AGO4 uncovers new AGO functions. *Plant J.*, **67**, 292–304.

20. Lakatos,L., Szittya,G., Silhavy,D. and Burgyán,J. (2004) Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO J.*, **23**, 876–884.

21. Várallyay,É., Válóczi,A., Ágyi,Á., Burgyán,J. and Havelda,Z. (2017) Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *EMBO J.*, **36**, 1641–1642.

22. Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J. Cell Mol. Biol.*, **16**, 735–743.

23. Várallyay,É., Válóczi,A., Ágyi,Á., Burgyán,J. and Havelda,Z. (2010) Plant virus-mediated induction of miR168 is associated with repression of ARGONAUTE1 accumulation. *EMBO J.*, **29**, 3507–3519.

24. Várallyay,É., Burgyán,J. and Havelda,Z. (2008) MicroRNA detection by northern blotting using locked nucleic acid probes. *Nat. Protoc.*, **3**, 190–196.

25. Pall,G.S. and Hamilton,A.J. (2008) Improved northern blot method for enhanced detection of small RNA. *Nat. Protoc.*, **3**, 1077–1084.

26. Taller,D., Bálint,J., Gyula,P., Nagy,T., Barta,E., Baksa,I., Szittya,G., Taller,J. and Havelda,Z. (2018) Expansion of Capsicum annuum fruit is linked to dynamic tissue-specific differential expression of miRNA and siRNA profiles. *PLoS One*, **13**, e0200207.

27. Szádeczky-Kardoss,I., Csorba,T., Auber,A., Schamberger,A., Nyikó,T., Taller,J., Orbán,T.I., Burgyán,J. and Silhavy,D. (2018) The nonstop decay and the RNA silencing systems operate cooperatively in plants. *Nucleic Acids Res.*, **46**, 4632–4648.

28. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, **17**, 10–12.

29. Johnson,N.R., Yeoh,J.M., Coruh,C. and Axtell,M.J. (2016) Improved placement of multi-mapping small RNAs. *G3: Genes Genomes Genetics*, **6**, 2103–2111.

30. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R. and 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

31. Kolde,R. (2012) Pheatmap: pretty heatmaps. *R Package Version*, **61**, 926.

32. Crooks,G.E., Hon,G., Chandonia,J.-M. and Brenner,S.E. (2004) WebLogo: A sequence logo generator. *Genome Res.*, **14**, 1188–1190.

33. Prüfer,K., Stenzel,U., Dannemann,M., Green,R.E., Lachmann,M. and Kelso,J. (2008) PatMaN: rapid alignment of short sequences to large databases. *Bioinformatics*, **24**, 1530–1531.

34. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.

35. Zhang,C., Li,G., Zhu,S., Zhang,S. and Fang,J. (2014) tasiRNAdb: a database of ta-siRNA regulatory pathways. *Bioinformatics*, **30**, 1045–1046.

36. Tang,Y., Horikoshi,M. and Li,W. (2016) ggfortify: unified interface to visualize statistical results of popular R packages. *R J.*, **8**, 474–485.

37. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

38. Freese,N.H., Norris,D.C. and Loraine,A.E. (2016) Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*, **32**, 2089–2095.

39. Mallory,A. and Vaucheret,H. (2010) Form, function, and regulation of ARGONAUTE proteins. *Plant Cell*, **22**, 3879–3889.

40. Kasschau,K.D., Fahlgren,N., Chapman,E.J., Sullivan,C.M., Cumbie,J.S., Givan,S.A. and Carrington,J.C. (2007) Genome-wide profiling and analysis of Arabidopsis siRNAs. *PLoS Biol.*, **5**, e57.

41. Xie,M. and Yu,B. (2015) siRNA-directed DNA methylation in plants. *Curr. Genomics*, **16**, 23–31.

42. Yu,B., Yang,Z., Li,J., Minakhina,S., Yang,M., Padgett,R.W., Steward,R. and Chen,X. (2005) Methylation as a crucial step in plant microRNA biogenesis. *Science*, **307**, 932–935.

43. Vaucheret,H., Vazquez,F., Crété,P. and Bartel,D.P. (2004) The action of ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are crucial for plant development. *Genes Dev.*, **18**, 1187–1197.

44. Montgomery,T.A., Howell,M.D., Cuperus,J.T., Li,D., Hansen,J.E., Alexander,A.L., Chapman,E.J., Fahlgren,N., Allen,E. and Carrington,J.C. (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell*, **133**, 128–141.

45. Xie,Z., Allen,E., Fahlgren,N., Calamar,A., Givan,S.A. and Carrington,J.C. (2005) Expression of Arabidopsis MIRNA Genes. *Plant Physiol.*, **138**, 2145–2154.

46. Gurjar,A.K.S., Panwar,A.S., Gupta,R. and Mantri,S.S. (2016) PmiRExAt: plant miRNA expression atlas database and web applications. *Database J. Biol. Databases Curation*, **2016**, doi:10.1093/database/baw060.

47. Nag,A. and Jack,T. (2010) Chapter Twelve - Sculpting the flower; the role of microRNAs in flower development. In Timmermans,MCP (ed). *Current Topics in Developmental Biology, Plant Development*. Academic Press, Vol. **91**, pp. 349–378.

48. Válóczi,A., Várallyay,É., Kauppinen,S., Burgyán,J. and Havelda,Z. (2006) Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. *Plant J.*, **47**, 140–151.

49. Zhang,B. (2015) MicroRNA: a new target for improving plant tolerance to abiotic stress. *J. Exp. Bot.*, **66**, 1749–1761.

50. Song,X., Li,Y., Cao,X. and Qi,Y. (2019) MicroRNAs and their regulatory roles in plant–environment interactions. *Annu. Rev. Plant Biol.*, **70**, 489–525.

51. Han,J., Lee,Y., Yeom,K.-H., Nam,J.-W., Heo,I., Rhee,J.-K., Sohn,S.Y., Cho,Y., Zhang,B.-T. and Kim,V.N. (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell*, **125**, 887–901.

52. Kim,W., Kim,H.-E., Jun,A.R., Jung,M.G., Jin,S., Lee,J.-H. and Ahn,J.H. (2016) Structural determinants of miR156a precursor processing in temperature-responsive flowering in Arabidopsis. *J. Exp. Bot.*, **67**, 4659–4670.

53. Werner,S., Wollmann,H., Schneeberger,K. and Weigel,D. (2010) Structure determinants for accurate processing of miR172a in Arabidopsis thaliana. *Curr. Biol.*, **20**, 42–48.

54. Mateos,J.L., Bologna,N.G., Chorostecki,U. and Palatnik,J.F. (2010) Identification of microRNA processing determinants by random mutagenesis of Arabidopsis MIR172a precursor. *Curr. Biol.*, **20**, 49–54.

55. Marchais,A., Chevalier,C. and Voinnet,O. (2019) Extensive profiling in Arabidopsis reveals abundant polysome-associated 24-nt small RNAs including AGO5-dependent pseudogene-derived siRNAs. *RNA N. Y.*, doi:10.1261/rna.069294.118.

56. Vaucheret,H. (2006) Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.*, **20**, 759–771.

57. Park,M.Y., Wu,G., Gonzalez-Sulser,A., Vaucheret,H. and Poethig,R.S. (2005) Nuclear processing and export of microRNAs in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 3691–3696.