HUNGARIAN UNIVERSITY OF AGRICULTURE AND LIFE SCIENCES

Doctoral School of Biological Sciences

ANALYSIS OF ARGONAUTE PROTEINS IN BARLEY: INSIGHTS INTO miRNA LOADING EFFICIENCY AND AGO4 FUNCTIONALITY IN HETEROLOGOUS COMPLEMENTATION

DOI: 10.54598/004940

The Thesis of the PhD dissertation

Fabio Miloro

Gödöllő

2024

The PhD program

Name: Doctoral School of Biological Sciences

Discipline: Plant Biotechnology

Leader of the school: Prof. Dr. Nagy Zoltán, DSc

Head of Biology Sciences Doctoral School

Hungarian University of Agriculture and Life Sciences, Department of Plant Physiology and Plant Ecology, Agroecology Research Group

Supervisors: Dr. Zoltán Havelda (DSc) and Dr. Ágnes Dalmadi

Hungarian University of Agriculture and Life Sciences, Genetic and Biotechnology Institute, Plant Physiology and Developmental Biology Group

……..……

Approval of the PhD School leader

……..……

Approval of the Supervisor

……..…

Approval of the Co-supervisor

1 INTRODUCTION

Barley (*Hordeum vulgare* L.) is the fourth most widely grown cereal in the world, behind wheat, rice and corn. Its nutritious grain is a valuable human and animal food resource. Its close relationship with wheat and its diploid ancestral genome makes it an ideal crop to study cereal genetics. The importance of barley for epigenetic studies lies in its ability to exhibit remarkable phenotypic plasticity and high epigenetic diversity in response to environmental changes such as increased temperature and decreased precipitation. In addition, the potential role of epigenetic regulation as a mechanism of adaptation to novel environmental conditions can be elucidated by studying epigenetic responses in barley. And its genome is full of transposable elements (TEs), so epigenetic control of these elements helps maintain genome stability.

In eukaryotes, small RNA (sRNA)-mediated gene silencing is a central mechanism for developmental regulation, responses to environmental cues, and epigenetic control of transposable elements (TEs). This nucleotide sequence-specific gene regulatory mechanism plays a critical role in diverse biological processes. The sRNAs, including a subclass known as microRNAs (miRNAs), are the hallmark molecules of RNA silencing and are typically 20-24 nucleotides in length. Of the two major subgroups, miRNAs primarily control endogenous gene expression to coordinate developmental processes and stress responses, whereas small interfering RNAs (siRNAs) are also involved in maintaining genome integrity and in biotic stress responses.

All types of sRNAs described are known to associate with Argonaute (AGO) proteins to form RNA-induced silencing complexes (RISCs). Each RISC is directed by the bound sRNAs to achieve specific interactions with target transcripts based on sequence complementarity, resulting in either mRNA cleavage, translational repression, or chromatin modification. Eukaryotic AGO proteins, which are divided into the AGO and PIWI subfamilies, have specific associations with different types of sRNAs, enabling them to operate diverse regulatory pathways. While PIWIs and their interacting sRNAs (piRNAs) are predominantly found in animal germlines, plant genomes encode several AGO proteins, all of which belong to the AGO subfamily. This diversity of functions has been made possible by the expansion of gene families encoding RNA silencing components.

The *Arabidopsis* genome encodes ten different AGO proteins, each with unique roles and some with overlapping functions. The functionality of these proteins is generally reflected by the sRNA content associated with them. These AGO proteins are composed of multiple domains including N-terminal, PIWI/Argonaute/Zwille (PAZ), MID, and P-element-induced WImpy testis (PIWI)

domains; the latter plays a critical role in recognizing their 5′ end of sRNA. sRNAs can be classified into different AGO proteins based on their length and the nucleotide at the 5′ end. For example, AGO1, a key regulator of the miRNA pathway, preferentially loads 21-nt sRNAs starting with a uridine (U) residue, while AGO4 and AGO6, involved in RNA-directed DNA methylation (RdDM), show a preference for 24-nt sRNAs starting with an adenine (A) residue.

In contrary with Arabidopsis which has a relatively small genome, the number of AGO proteins increases in other flowering plants. As an example, soybean contains 22 AGO proteins (Glycine max, a paleopolyploid), in rice 19 have been identified (*Oryza sativa*), in maize 17 (*Zea mays*), and in barley 11 have been described so far (*Hordeum vulgare*). However, further analysis suggests that a reassessment of both the quantity and nomenclature of barley AGO proteins may be required. All flowering AGOs are grouped into three major clades: AGO1/5/10, AGO2/3/7, and AGO4/6/9. This expansion of the plant AGO family suggests functional diversification of AGO proteins, presumably reflecting the expansion of sRNA-directed regulatory pathways.

An important pathway of sRNA is post-transcriptional gene silencing mediated by the action of miRNAs. In miRNA biogenesis, a genome-encoded primary miRNA (pri-miRNA) transcript with a specific stem-loop structure is cleaved to produce 20-24 nucleotide long miRNA duplexes. These miRNA precursors, which exhibit considerable variability in length and structure, are initially cleaved by DICER-LIKE1 (DCL1) in plants. The resulted miRNA intermediate, consisting of a guide (miRNA) and a passenger (miRNA*) strand and either proximal, or distal part of the precursor, are further cleaved based on structural signals in the pri-miRNA and require DCL1 co-factors for accurate processing to a miRNA/miRNA* duplex. These duplexes are then methylated for protection against exonucleases and are thought to be exported to the cytoplasm. The mature strand of the duplexes are then incorporated into RISC, the effector of RNA silencing. During RISC assembly, the miRNA guide strand is loaded onto AGO1, the central component of RISC, while the miRNA* is ejected and degraded. It's been shown that AGO1 containing RISC is mainly assembled in the nucleus and exported as a complex to the cytosol. AGO-RISC loaded with a miRNA identifies its RNA targets via complementary base pairing and mediates their repression. The level of functional AGO1 protein is finely controlled by mechanisms involving both transcriptional and post-transcriptional feedback loops: AGO1 mRNA expression is controlled by miR168-programmed AGO1 RISC, and the expression of AGO1 and MIR168 genes are co-transcriptionally regulated. This complex, multi-layered regulation of AGO1 ensures robust and precise control of this RNA silencing pathway.

Another prominent molecular pathway is RNA-directed DNA methylation, which ensures the transcriptional control of numerous endogenous genes and TEs. In *Arabidopsis*, AGO4 is primarily

responsible for binding 24-nt siRNAs associated with repeats and heterochromatin and plays a critical role in RdDM. During this process siRNAs are originated from DNA-DEPENDENT RNA POLYMERASE IV (Pol IV) transcripts, which are converted into double-stranded RNAs (dsRNAs) by RNA-DEPENDENT RNA POLYMERASE 2 (RDR2). These dsRNAs are then processed by DICER-LIKE 3 (DCL3) into 24-nt siRNAs, which are finally integrated into AGO4. The AGO4-bound siRNAs guide the targeting of nascent scaffold transcripts of DNA-DEPENDENT RNA POLYMERASE V (Pol V) based on sequence complementarity and recruit DNA methyltransferase activity to facilitate *de novo* methylation. The proteins effector of the process, AGO4 and its paralogs, have been extensively studied in *Arabidopsis* using mutant models. In monocots, however, experimental studies have been restricted to orthologs in rice and maize. Four different members of the AGO4 clade have been identified in rice: AGO4A, AGO4B, AGO15, and AGO16. Notably, both AGO4A and AGO4B were found to load 24 nt long miRNAs (lmiRNAs) and siRNAs, but they exhibited unique specificity for their loaded sRNAs. AGO4 also plays an important role in maintaining genome stability by silencing TEs. Small TEs and TE fragments near genes are targeted by RdDM and are typically found in open and accessible euchromatic regions of the genome that promote gene expression. TEs significantly influence the epigenetic modifications that regulate plant development and stress adaptation. In addition to its role in maintaining TE silencing, RdDM can also induce transcriptional silencing of foreign DNA, including novel TE insertions, virus-derived sequences, and transgenes.

The aim of this study was to identify and analyze AGO proteins in barley using bioinformatics approaches and to investigate the function of Arabidopsis AGO1 and barley AGO4 in vivo. By searching the barley genome and transcriptome, AGO protein sequences were identified and subjected to phylogenetic analysis alongside rice and *Arabidopsis* AGO proteins, revealing orthologous relationships. Further investigation focused on the impact of miR168/miR168* duplex structure on AGO1 loading efficiency, highlighting how structural features influence loading efficiency, which in turn affects AGO1 protein levels and phenotypic outcomes. Manipulation of the cellular miRNA pool demonstrated the establishment of a new steady-state AGO1 protein level, revealing a regulatory mechanism in which the loading rate into the RISC determines the biological activity of miRNAs. In addition, our study identified two active paralogous barley genes and a putative pseudogene within the AGO4 clade that exhibited distinct properties in sRNA binding and functionality when introduced into an *Arabidopsis ago4* mutant. This approach confirmed the functionality of the active barley AGO4 genes in regulating RdDM targets and revealed their specific preferences for sRNAs from different TE regions.

2 OBJECTIVES

Barley, an important cereal crop, has a complex and intriguing set of Argonaute (AGO) proteins. These AGO proteins are key components of the RNA silencing machinery in plants and play critical roles in gene regulation, transposon control, and defense against viruses.

The objectives of this study were:

- Genome-wide identification and expression analysis of AGO proteins in barley using *in silico* analysis.
- *In silico* analysis of barley AGO1 expression levels and miR168 target site.
- Investigation of the factors that influence the role of miR168 on the loading efficiency of RISC complexes.
- Exploring the potential of miR168 precursor structure modification to alter the AGO1 loading efficiency.
- Identification and bioinformatic analysis of putative *AGO4* genes in barley, including determining their expression levels and phylogenetic relationships to AGO4 proteins of other plants.
- Assessment of the functionality of the identified barley AGO4 proteins through complementation of the *Arabidopsis ago4-3* mutant.
- Analyze barley AGO4-associated small RNAs in heterologous complementation.

3 MATERIALS AND METHODS

3.1 Plant material and growth conditions

In this study, *Arabidopsis thaliana* plants were incubated for three days at 4°C, then germinated at 21°C on MS agar medium with 1% sucrose, with or without antibiotics. Seedlings were grown under an 8-hour light/16-hour dark cycle at 21°C for three weeks, then switched to a 16-hour light/8-hour dark cycle. The *ago4-3* mutant, homozygous from the WiscDSLox338A06.0 T0 line, was checked using phosphinothricin, anti-AtAGO4 antibody, and PCR. Various plant tissues were collected for DNA and RNA extraction. For heat stress, 1-week-old seedlings were moved to 37°C for 24 hours under the same light conditions.

Nicotiana benthamiana plants were grown at 21°C with a 16-hour photoperiod and used at the 3-leaf stage for agroinfiltration.

Hordeum vulgare cv. Golden Promise was grown with a 16-hour light/8-hour dark cycle at 20°C/16°C. Developing inflorescences (15-25 mm) were collected for RNA extraction.

3.2 Phylogenetic analysis and *in silico* predictions

AGO protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, and *Hordeum vulgare* were retrieved from Ensembl Plants and Uniprot, aligned with ClustalW, and used to create a phylogenetic tree using the neighbor-joining method. Bootstrap values (1000 replicates) are shown on branches. Evolutionary distances were calculated with the Poisson correction method in MEGA11.

The miRNA sequences were used to predict target sites on AGO1 cDNAs using psRNATarget. Sequences from *Arabidopsis thaliana* and *Hordeum vulgare* were retrieved from TAIR and EnsemblPlants and uploaded in FASTA format. Default scoring settings (Schema V2) were used.

3.3 Plasmid construction and plant transformation

All plant expression plasmids were constructed using the pGreen 0029 vector (kanamycin resistance) (www.pgreen.ac.uk) and all the amplifications were performed using Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific) following the manufacturer's instructions.

The miRNA constructs for *MIR168a*, *MIR168b*, *MIR156a*, and *MIR171a* included a 10 bp flanking

region on both sides. For amiRNA constructs, a modified *hvu-MIR171* stem-loop was used. Specific nucleotide changes were introduced in the *MIR168* stem-loop via PCR mutagenesis. All constructs were inserted into pGreen0029 with a 35S expression cassette. The AGO1 sensor used a 558 bp cDNA fragment with the miR168 target site linked to GFP.

For barley AGO4 heterologous complementation, vectors were made with *AtAGO4* promoter and terminator sequences, *HvAGO4a* and *HvAGO4b* cDNAs, modified with HA epitope tags and restriction sites, cloned into pGreen0029. Plasmids were transformed into *E. coli*, sequenced, and then into *Agrobacterium tumefaciens* AGL1. *Arabidopsis* plants were transformed using the floral dip method and screened on selective media.

3.4 Transient assay

Six-week-old *Nicotiana benthamiana* leaves were infiltrated with *Agrobacterium tumefaciens* (AGL1) suspensions containing sensor, miRNA-producing, and P14 constructs. P14 was included to inhibit the siRNA pathway without affecting signal monitoring. Normalized amounts of constructs were supplemented with AGL1 containing the empty pGreen0029 vector.

Samples were collected three days post-infiltration; four 1 cm discs were pooled from different leaf areas for each construct. Sampling was from both sides of the same leaf. Each construct combination was tested on four to five plants, with experiments repeated at least three times.

3.5 RNA isolation and RT-qPCR

Total RNA was extracted from *Arabidopsis* seedlings, leaves, flowers, and barley inflorescences using the phenol-chloroform method. Samples were homogenized in extraction buffer and RNA was precipitated with ethanol and Na-acetate at -80°C.

For RT-qPCR, 4 µg total RNA was DNaseI-treated, re-isolated, and resuspended in water. First-strand cDNA was synthesized from 2 µg RNA using random hexamers and oligo$(dT)_{18}$ primers. qPCR was performed with Luminaris Color HiGreen qPCR Master Mix on a LightCycler 96, normalizing data to *AtUBC9*, *AtACT2*, and *AtPP2AA3*. Results were analyzed using GraphPad Prism 8 from three biological replicates.

3.6 Gel filtration assay

Size separation gel-filtration experiments were performed using Superdex 200 10/300 or Sephacryl S-300 High Resolution columns. Plant tissues were homogenized in liquid nitrogen with elution buffer (50 mM Tris-HCl pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 4 mM DTT) and centrifuged to remove debris. 200 µl of the extract was injected into the equilibrated gel-filtration column at 4°C. 48 fractions were collected: odd samples for RNA extraction (phenol-chloroform) and even samples for protein purification (acetone precipitation).

3.7 miRNA detection and quantification

Small RNA northern blot analyses were performed with 4 µg total RNA or gel-filtration samples. Samples were separated on 12% polyacrylamide gels with 8 M urea, transferred to Hybond NX membrane, and chemically cross-linked. Membranes were probed with biotinylated LNA or DNA probes, and signals were detected using a Chemiluminescent Nucleic Acid Detection Kit and visualized with a ChemiDoc™ MP Imaging System.
For gel-filtration blots, the volume intensity of the four main RISC-loaded and unbound fractions were measured. Loading efficiency (LE) was calculated as the percentage of RISC-loaded intensity over the total intensity of RISC-loaded plus unbound fractions.

3.8 Protein extraction and western blotting

Arabidopsis seedlings, leaves, and flowers were homogenized in extraction buffer with Laemmli buffer. Samples were boiled, centrifuged, and loaded on 8% or 10% SDS-PAGE gels. Proteins were transferred to PVDF membranes and subjected to western blot analysis.
Antibodies used: anti-AGO1 (1:7500), anti-HA-peroxidase (1:2000), anti-actin (1:2000), anti-AtAGO4 (1:5000), and anti-BiP (1:10000). Secondary antibodies: goat anti-rabbit HRP (1:10000) and goat anti-mouse HRP (1:10000). Blocking was done with 5% milk powder in PBST for 1 hour, primary antibodies in 1% milk PBST for 1-2.5 hours, and secondary antibodies for 1 hour. Blots were washed with PBST and developed with High Clarity Western ECL on ChemiDoc™ MP Imaging System. Signal intensity was quantified using Image Lab 6.1, normalized to actin or BiP.

3.9 Chop-qPCR analysis

The Chop-PCR assay used the ZenoGene Plant DNA Purification Kit to extract genomic DNA from Arabidopsis T3 inflorescence, following the manufacturer's guidelines. DNA amount was assessed with a Nanodrop ND-1000 spectrophotometer. MspJI enzyme was employed for DNA digestion, sensitive to methylation, targeting cytosine modifications like 5-mC and 5-hmC. The reaction mix contained 10× rCutSmart™ Buffer, MspJI enzyme, and Enzyme Activator Solution, incubated with 600 ng genomic DNA for 4 hours at 37 °C. Controls lacked MspJI. Chop-qPCR used digested or undigested DNA as templates, with measurements performed using Luminaris Color HiGreen qPCR Master Mix on a LightCycler 96 Instrument. Data, normalized to undigested *AtSN1*, were collected from three biological replicates and analyzed with GraphPad Prism 8.

3.10 Relative copy number assessment of transposable elements

Genomic DNA from Arabidopsis non-treated (NT) and heat-stressed (HS) seedlings was extracted using the ZenoGene Plant DNA Purification Kit per the manufacturer's instructions and quantified with a Nanodrop ND-1000 spectrophotometer. Relative quantification of *Ty1/copia*-like retrotransposon *ATCOPIA78* (*ONSEN*) copies compared to non-treated wild-type DNA, with eight *ONSEN* copies in the Col-0 ecotype genome, was performed using qPCR. Luminaris Color HiGreen qPCR Master Mix and 20 ng of genomic DNA per reaction were utilized with the LightCycler 96 Instrument. Data, normalized to *AtUBC9*, were collected from three biological replicates.

3.11 RNA-seq of developing barley inflorescences

RNA from developing barley inflorescences was quantified using the Qubit RNA HS Assay and assessed for quality with the LabChip GX Touch Nucleic Acid Analyzer, resulting in a score of 10 for all samples. Library preparation for RNA-seq was conducted using the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0. This involved mRNA purification, fragmentation, cDNA synthesis, adapter ligation, and PCR amplification. Library quantification and quality control were performed with the XMark HT chip on the Labchip GX Touch Nucleic Acid Analyzer. Equimolar pooling was done before loading onto the Illumina NovaSeq 6000 for paired-end sequencing using the S4 Reagent Kit v1.5. Analysis of paired-end fastq raw data was done using salmon with the Morex

V3 reference genome, transcripts fasta file, and a GTF file for transcript-to-gene mapping, providing read mapping and transcript abundance in TPM.

3.12 Immunoprecipitation

Crude extracts were prepared by homogenizing 0.4 g of *Arabidopsis* seedlings, rosette leaves, or barley mixed inflorescences in lysis buffer. After three centrifugation steps at 4°C, the supernatant was transferred to new tubes to remove cellular debris. Dynabeads Protein G Immunoprecipitation Kit was then used with anti-AGO1 or anti-HA-peroxidase antibody. RNA purification followed by the phenol-chloroform method and protein extraction with Laemmli buffer were performed. Western blotting using anti-HA-peroxidase antibody verified transgenic AGO4 presence, while total protein contamination was assessed using BioRad gels with TGX technology.

3.13 Small RNA library preparation and analysis

The cDNA libraries for sequencing were constructed using RNA samples from bulked seedlings of transgenic lines overexpressing different miR168 precursors and from immunoprecipitation of complementation lines for HvAGO4A and HvAGO4B. Small RNA fractions enriched in the 21-22 nt range were isolated from polyacrylamide gels and used exclusively for library preparation with the Truseq Small RNA Library Preparation Kit (Illumina), following a modified protocol. Sequencing was conducted on an Illumina NextSeq 500 system with single-end 50 bp reads. Raw data for AtAGO4 sRNA-IP sequencing were retrieved from NCBI, and analysis was performed using the Galaxy platform to control quality, trim, and map reads to the *A. thaliana* reference genome (TAIR10.1) using hisat2. The sRNAPipe pipeline categorized reads into different genomic sequences categories and allowed for the selection of the size range (18–27 nt) of the sRNAs. Visualization of sRNAs mapped to chromosomes or loci was done using IGV, while protein alignments were generated using ESPript 3. GraphPad Prism 8 was utilized for graph generation and statistical analysis

4 RESULTS

4.1 Bioinformatic analysis of putative AGO genes in barley

A detailed in silico analysis was performed to identify the putative *AGO4* gene(s) and to clarify the phylogenetic relationship of barley AGOs with rice and *Arabidopsis* orthologs. The whole barley genome (both Morex V3 and Golden Promise v1) yielded 21 putative candidate genes in the Argonaute clade. Translated protein sequences were subjected to InterPro analysis to confirm characteristic AGO domains. Phylogenetic analysis revealed three major clades: AGO1/5/10, AGO2/3/7, and AGO4/6/9, with AGO18 forming a distinct grass-specific subclade. Barley had 21 AGO proteins compared to 19 in rice and 10 in *Arabidopsis*. Gene duplication and divergence were observed, particularly in the AGO1/5/10 clade, indicating expansion in size, whereas clades 2/3/7 remained more similar across species, suggesting stability of function over time.

4.1.1 Barley *AGO1* Genes

Barley and rice each have multiple *AGO1* orthologous genes, with barley having 5 copies of AGO1 compared to 4 copies in rice. Barley *AGO1* genes expression patterns were analyzed using the BarleyExpDB database, revealing distinct expression profiles in different tissues. *HvAGO1B_1* stood out as the most highly expressed gene in inflorescences, particularly at the 5 mm and 1.5 cm stages. Following expression analysis, psRNATarget was used to predict miR168 target sites among barley *AGO1* genes, revealing variations in expectation values and mismatches compared to Arabidopsis *AGO1*. Alignment of Arabidopsis, rice, and barley AGO1 proteins revealed conserved regions involved in miRNA binding, with differences observed in amino acid sequences affecting binding affinity. These variations could lead to functional divergence among AGO1 proteins, affecting their interaction with miRNAs and overall protein stability and conformation.

4.1.2 Structure and Expression Analysis of Putative AGO4 Genes in Barley

In the AGO4/6/9 clade, *AGO6* did not undergo gene duplication, while *AGO4* did, resulting in three genes: *AGO4a*, *AGO4b*, and *AGO15*. The similarities with rice highlight the importance of further studying this gene duplication in barley. RNA-seq analysis of developing barley

inflorescences revealed that *HvAGO4a* was the most highly expressed, followed by *HvAGO4b*, while *HvAGO15* showed significantly low expression. *HvAGO6* showed significantly lower expression compared to the *AGO4* genes. Putative orthologous AGO4 proteins from barley showed higher identity to rice proteins than to each other. Sequence analysis revealed identical gene structures for HvAGO4A and HvAGO4B, including the position of the PAZ and PIWI domains. PIWI domain analysis revealed conservation in the region anchoring the 5′ end of sRNAs among barley, Arabidopsis, and rice AGO4 proteins, except for a single amino acid change in the monocot proteins that is not present in Arabidopsis, suggesting a similar function for HvAGO4A and HvAGO4B compared to their rice counterparts.

4.1.3 PIWI Structure of AGO6 in Barley and Rice

An analogous analysis was performed for AGO6, which revealed that the anchor site for the 5′ end of the sRNAs is located at the same position as in AGO4 (α12-β29-α13). Furthermore, this analysis revealed a significant conservation among the sites across different plants, with the exception of one amino acid in the four-amino acid site (QCIx), which represented a variation at a different position compared to that observed in AGO4.

4.1.3 *AGO15* as a Pseudogene

In barley, unlike rice, *AGO15* does not originate from an *AGO4a* tandem duplication, as indicated by its separate chromosomal location and the absence of intronic TEs. However, as in rice, *HvAGO15* has undetectable expression levels, suggesting a pseudogene status or tissue-specific expression. Analysis revealed six tandem repeats in exon 1 with potential start codon variations. Despite attempts, PCR amplification of *HvAGO15* from leaf and inflorescence yielded only fragments, confirming its pseudogene status.

4.2 Investigation of mechanisms behind the AGO1-miR168 feedback regulatory loop

AGO1 is regulated by miR168 in an autoregulatory loop in which miR168 accumulates primarily as unloaded free miRNA in the cytoplasm. Previous studies have suggested that miRNA precursor structure influences the loading efficiency of miR168. Here, we investigated the effect of miR168

overexpression on AGO1 accumulation using transient expression in *N. benthamiana* leaves and transgenic Arabidopsis plants overexpressing *ath-MIR168a*. Despite strong miR168 overexpression, only modest reductions in AGO1 protein levels were observed, as confirmed by western blot analysis. Transgenic plants showed subtle phenotypic changes such as serrated leaves and delayed flowering, which were proportional to miR168 accumulation levels. However, the decrease in AGO1 protein levels and miR159 accumulation was moderate. Size separation gel filtration assays revealed that miR168 is primarily present in a protein-unbound form, with only a small fraction loaded into high molecular weight AGO1-RISC complexes. Even at elevated miR168 levels, loading into AGO1-RISC remained modest, suggesting a tight regulation of miR168 loading efficiency.

4.2.1 Changes in duplex structure can further reduce the AGO1-loading of miR168

Computational analysis of miR168/miR168* duplex structures across plant species revealed a conserved nucleotide mismatch at the fourth position, which may affect AGO1 loading efficiency. To investigate this, a modified construct (*MIR168-4bp*) was designed to introduce the fourth position mismatch exclusively in the miR168* strand. Overexpression of *MIR168-4bp* resulted in increased GFP signal and higher levels of AGO1 sensor protein compared to wild-type *MIR168a*. Transgenic *MIR168-4bp* lines exhibited less flowering delay and higher AGO1 protein levels compared to *MIR168a* lines, suggesting reduced AGO1 downregulation. Gel filtration experiments showed lower miR168 loading efficiency in *MIR168-4bp* plants compared to *MIR168a*. AGO1 immunoprecipitation experiments confirmed reduced miR168 accumulation in *MIR168-4bp* plants. These results suggest that the fourth nucleotide mismatch in the miR168/miR168* duplex reduces AGO1 RISC loading capacity, resulting in different AGO1 protein levels compared to MIR168a overexpression.

4.2.2 Altered duplex structure can enhance AGO1 loading of miR168

To assess the role of duplex structure in regulating loading efficiency, the *hvu-MIR171* precursor, known for its efficient loading, was used. A heterologous barley precursor showed that structure alone regulates loading efficiency. To mimic the three mismatches of *hvu-MIR171* in the miR168 duplex, a modified duplex (*MIR168-3mm*) was designed by altering only the passenger strand.

Transient expression in *N. benthamiana* showed reduced fluorescence of the AGO1-GFP sensor, and protein quantification confirmed this effect. Genetically engineered plants containing *MIR168-3mm* exhibited modest miR168 overexpression, resulting in more pronounced flowering delay and AGO1 downregulation compared to higher miR168 levels from the original MIR168a construct. Gel filtration experiments demonstrated enhanced HMW-RISC incorporation of *MIR168-3mm*-derived miR168, resulting in increased AGO1 protein downregulation. Immunoprecipitation of AGO1 from *MIR168-3mm* transgenic plants further supported these findings, suggesting that the structural features of the miR168 duplex tightly regulate AGO1 loading.

4.2.3 Artificial precursor structure enhanced AGO1 loading of miR168

The *MIR168-3mm* construct, derived from the *MIR168a* backbone but with duplex structural features similar to *hvu-MIR171*, was designed to investigate factors influencing AGO1 loading of miR168. To further investigate these structural influences, artificial miR168 precursor (AMIR) constructs based on modified *hvu-MIR171* fragments were designed. Two AMIR variants, *AMIR-1* and *AMIR-2*, retained the *hvu-MIR171* stem-loop structure while altering the orientation of the miR168 guide strand and modifying the star strand. Both AMIR variants exhibited a greater reduction in AGO1 sensor fluorescence and protein levels compared to *MIR168a* when transiently expressed in N. benthamiana leaves. Small RNA Northern blot analysis confirmed lower miR168 overexpression levels with AMIR constructs. Stable transgenic lines of *AMIR-1* and *AMIR-2* showed pronounced phenotypic changes, such as delayed flowering and reduced rosette diameter, which correlated with miR168 overexpression levels. Despite slightly lower miR168 overexpression, AGO1 protein levels were significantly reduced in AMIR lines compared to *MIR168a* lines. Gel filtration experiments revealed increased miR168 accumulation in HMW AGO-RISC fractions in AMIR lines despite reduced AGO1 protein accumulation. Immunoprecipitation experiments confirmed increased miR168 incorporation into AGO1-RISC in both AMIR lines, suggesting improved loading efficiency of AGO1-RISC with alternative stem-loop structures, potentially enhancing biological activity.

4.2.4 Small RNA-Sequencing and miR168 Species

Excessive production of miR168 from altered precursors may lead to differential accumulation of canonical and non-canonical miR168 species (iso-miRs), potentially interfering with RNA silencing autoregulation. High-throughput sequencing (HTS) analyses of sRNA pools from transgenic lines revealed predominantly 21 nt long small RNAs mapped to miR168 precursors, with all precursor fragments overproducing miR168 compared to wild-type plants. Differential accumulation of 5′ U, 5′ C, 5′ G, and 5′ A miR168 species was observed in transgenic plants, suggesting potential miR168 misprocessing that affects AGO1 loading efficiency. However, the relative accumulation of different miR168 species at the 5′ end did not significantly interfere with or contribute to differential AGO1 loading. Comparative analysis with previously published AGO1 immunoprecipitation data suggested that most 5′ C miR168 species are biologically active and follow similar AGO loading rules as 5′ U miR168. In conclusion, while structural features of the miRNA duplex primarily influence miR168 loading efficiency, altered production of miR168 species may also play a minor role.

4.3 Investigation of barley AGO4 gene functionality in Arabidopsis complementation

Our work demonstrates that Arabidopsis thaliana is a suitable organism for testing AGO functionality due to the availability of molecular methods and mutants for rapid functional assays. To validate the functionality of potential barley *AGO4* genes, a heterologous complementation assay using the *AtAGO4* promoter and terminator was developed to introduce 5′ HA-tagged versions of each barley gene into *A. thaliana ago4-3* mutants (WiscDSLox338A06). T0 plants were initially selected for kanamycin resistance, and subsequent generations (T1 onwards) were analyzed for mRNA and protein expression levels in the inflorescence. Transgenic lines resembling the wild type were selected for further study to avoid positional effects of the transgene insertion. The transformants showed different levels of transgene expression at the RNA level, with some lines showing expression levels similar to endogenous AGO4, in particular two HvAGO4B lines (#1 and #17). Western blot analysis using an anti-HA antibody confirmed a correlation between mRNA and protein levels. Three lines with high AGO4 protein levels for both barley genes were selected for further analysis to allow a transgene-level dependent study of the complementation effect.

4.3.1 Functional complementation of barley AGO4 genes in *Arabidopsis ago4* mutant

The functionality of the barley *AGO4* genes was tested in HA-HvAGO4A and HA-HvAGO4B complementation lines. Using the *AtSN1* retrotransposon, known to be regulated by AGO4 in *Arabidopsis*, transgenic HvAGO4A and HvAGO4B achieved a substantial reduction in *AtSN1* expression proportional to their expression levels. Chop-PCR confirmed the restoration of *AtSN1* methylation in the complemented lines, which correlated with the reduced *AtSN1* expression. In addition, RdDM-dependent *AtROS1* expression was restored in the complemented lines, with HvAGO4A plants showing particularly high expression levels, suggesting successful compensation for the *ago4-3* mutation.

4.3.2 Effect of barley AGO4 proteins on TE activation under heat stress conditions

The influence of barley AGO4 proteins on TE activation under heat stress was investigated using Arabidopsis seedlings subjected to 24 h of 37°C heat stress. Both HvAGO4A and HvAGO4B reduced TE activation compared to the ago4-3 mutant, while maintaining *ONSEN* extrachromosomal DNA levels similar to wild type. RT-qPCR analysis showed a significant up-regulation of *ONSEN* under heat stress in the *ago4-3* mutant compared to the wild type, while transformant lines exhibited expression levels closer to the wild type. Lines with higher transgenic protein levels showed significant down-regulation of *ONSEN* under heat stress compared to the wild type.

4.3.3 Differential sRNA-binding preferences of barley AGO4 proteins

The sRNA binding preferences of the putative HvAGO4A and HvAGO4B proteins were elucidated by sRNA-IP sequencing of Arabidopsis complementation lines. Control data from AtAGO4 sRNA-IP sequencing were also analyzed. Both barley AGO4 proteins showed a robust affinity for 24-nt sRNAs, similar to AtAGO4. Reads were mainly derived from TEs, transcripts and unannotated genomic regions. However, a significant difference was observed in the 5′ end nucleotide distribution: HvAGO4B showed a pattern similar to AtAGO4, binding sRNAs with C, G or U residues at the 5′ end, whereas HvAGO4A showed a distinct preference for sRNAs starting

with an A residue.

Sequence conservation analysis of 24nt sRNA pools from sRNA-IP sequencing revealed minor differences between AtAGO4 and HvAGO4 proteins. HvAGO4A selectively loaded 24A sRNAs, whereas HvAGO4B interacted with 24nt sRNAs starting with G, C or U residues at the 5′ end. In addition, variation was observed at the 3′ end: AtAGO4 favored sRNAs with a U residue, HvAGO4B with a C residue at the 23rd nucleotide, while HvAGO4A lacked a conserved position.

4.3.4 Detailed analysis of the role of barley AGO4 proteins in TE regulation

TE-derived sRNAs associated with HvAGO4A and HvAGO4B were compared to those of AtAGO4 to identify differences in sRNA mapping abundance. HvAGO4A was associated with 1877 TEs, with 591 having higher read counts and 1286 having lower read counts compared to AtAGO4. HvAGO4B was associated with 1454 TEs, of which 401 had higher sRNA abundance and 1053 had lower sRNA abundance compared to AtAGO4. Among all TEs with altered sRNA abundance in barley proteins, 128 and 449 TEs had increased and decreased sRNA content, respectively, in both HvAGO4 proteins. In addition, 17 TEs showed higher abundance in HvAGO4B and lower abundance in HvAGO4A compared to AtAGO4, while no TEs showed the opposite trend.

To further explore the differential effects of these proteins on TEs, an analysis of sRNAs mapped to specific TE loci was performed. AtSN1 had significantly fewer mapped sRNAs in HA-HvAGO4A compared to both AtAGO1 and HA-HvAGO4B. This was accompanied by a shift in the distribution of reads, with HvAGO4A showing a preference for central regions, reducing the total number of AtSN1-derived sRNAs. Different regions produced sRNAs with different 5′ end nucleotides, with HvAGO4A binding almost exclusively to sRNAs starting with A. AtAGO4 and HvAGO4B showed a stronger affinity for 5′ G but could bind sRNAs with any nucleotide at the 5′ end. Similar changes in the 5′ end nucleotide distribution were observed for other TEs, such as the *RathE3* TE (*AT5TE27090*), where HvAGO4A did not show any significant sRNA presence. This detailed analysis revealed significant differences in the mode of action of the three AGO4 proteins, despite their similar regulatory roles.

4.3.5 Prediction of 3D structure of barley AGO4 proteins bound to sRNAs

The distinct binding affinity of HvAGO4A and HvAGO4B was confirmed by sRNA-IP sequencing, showing a specific affinity for 24 nt 5′ A sRNAs in HvAGO4A and a less strict selection in HvAGO4B. Predictions using AlphaFold3 revealed the conformation of the AGO4-sRNA complex and its interaction, demonstrating the automatic positioning of the sRNA with its 5' end in direct contact with the MID-PIWI protein domains. Intermolecular hydrogen bonding between the PIWI domain and the sRNA was observed, which is crucial for the sequence specificity of the ssRNA. The region involved in 5′ sRNA anchoring, consisting of four amino acids with a single AA difference (QCxA), showed variation between proteins, affecting binding affinity due to differences in the side chain size and properties of individual amino acids.

5 DISCUSSION

This study investigates the role and regulation of Argonaute (AGO) proteins in RNA silencing pathways in barley. AGO proteins are critical for the recognition and binding of small RNAs, which are essential for the regulation of gene expression in plant development and response to stimuli. A key aspect of this research is the regulatory feedback of AGO1 and miR168, which is crucial for maintaining the balance within the plant's RNA silencing mechanisms.

The barley genome analysis identified 21 candidate *AGO* genes and provided insights into their role and evolution. Phylogenetic analysis distinguished three major clades (AGO1/5/10, AGO2/3/7, and AGO4/6/9) and an additional subclade, AGO18, unique to grasses. Notably, barley and rice have more AGO proteins than Arabidopsis, especially within the AGO1/5/10 clade. This suggests an expansion and possible specialization of functions in monocots. The identification of five *AGO1* orthologous genes in barley, similar to rice, highlights the complexity and diversity of the *AGO* gene family. Detailed analysis using the BarleyExpDB database revealed unique expression profiles for each of the five *AGO1* genes in barley. *HvAGO1B_1* was found to be most highly expressed in inflorescences at different developmental stages, suggesting tissue-specific regulatory mechanisms. The conservation of the miR168 target site among these genes suggests its regulatory role in modulating gene expression. The study further examined the structural differences in *AGO1* genes, especially in the miRNA binding sites, suggesting possible functional divergence among AGO1 proteins.

The duplication of the *AGO1* and *AGO4* genes in barley and rice compared to single copies in Arabidopsis suggests a selective advantage in gene regulation and environmental adaptation. For example, the *MEL1* gene in the AGO5 subclade shows germline-specific expression and binds phasiRNAs, suggesting functional specialization and similar localization to *AtAGO9*.

miRNAs play a critical role in regulating growth, development and stress responses by controlling the expression of target transcription factors and proteins. The miRNA pathway is highly adaptive and involves multiple levels of regulation, including transcriptional regulation, tissue-specific expression, and post-translational modifications. The feedback mechanism involving miR168 and AGO1 is particularly important, as an imbalance can lead to developmental defects and plant death.

The study investigated the structural motifs of miR168 precursors and showed that these influence the loading efficiency into the AGO1-RISC complex. Transgenic lines with modified miR168 precursors were used to study the effects of these structural changes. It was found that only a small

fraction of miR168 is loaded into AGO1-RISC, while the majority remains unbound, suggesting a regulated loading mechanism.

The research also focused on the two barley AGO4 proteins, HvAGO4A and HvAGO4B, which were introduced into an *Arabidopsis ago4-3* mutant to study their functionality. Barley AGO4 proteins successfully restored target locus expression and methylation levels, demonstrating the conservation of the RNA-directed DNA methylation (RdDM) pathway across species. The study found that barley AGO4 proteins effectively repressed the heat stress-activated transposable element *ONSEN*, suggesting a role in stress response. The study revealed differences in the binding affinity of the AGO proteins for the 5′ end nucleotide of sRNAs, with HvAGO4B resembling AtAGO4 in binding preferences, while HvAGO4A binds only 5’ A sRNAs. Structural analysis of the AGO proteins, particularly the MID and PIWI domains, revealed key determinants of binding specificity, supported by *in silico* 3D protein models.

In conclusion, this research advances the understanding of the role and regulation of AGO proteins in barley and highlights the evolutionary conservation and specialization of RNA silencing mechanisms. The results provide valuable insights into the molecular basis of plant development, stress responses and gene regulation.

6 CONCLUSIONS AND RECOMMENDATIONS

This study provides a comprehensive analysis of AGO proteins in barley, highlighting their roles in small RNA binding and AGO1 homeostasis. Through genome-wide identification and expression analysis, we revealed the diverse set of AGO proteins, highlighting their evolutionary significance and functional diversification.

In silico analysis of barley *AGO1* genes revealed distinct expression patterns in different tissues and developmental stages, highlighting the critical interplay between miR168 and AGO1 in maintaining homeostasis. The miR168 duplex structure was shown to play a critical role in determining the loading efficiency of the AGO1 RISC, providing a dynamic regulatory mechanism that adjusts AGO1 protein levels in response to cellular stimuli.

Our study of barley AGO4 proteins revealed their overlapping functionality and distinct small RNA binding properties, illustrating the evolutionary benefits of *AGO4* gene duplication. These proteins play a key role in transposable element regulation, contributing to the complexity and precision of genetic regulatory mechanisms in plants.

Future research should further explore the roles and mechanisms of AGO proteins in barley. Areas of focus include the specific functions of barley AGO1 proteins, their small RNA specificities, and tissue-specific roles. High-throughput sequencing to map small RNA populations associated with individual AGO1 proteins, together with CRISPR/Cas9-generated mutants, would provide deeper insights into their physiological and developmental effects. Comparative studies with rice and Arabidopsis mutants could elucidate conserved and species-specific roles of AGO1.

Furthermore, it is crucial to investigate how different regions of miRNA precursors influence the loading efficiency into AGO1. This involves dissecting precursor structures to identify critical elements for efficient loading and understanding the interactions between miRNA precursors, processing enzymes and AGO1 itself.

Further research on barley AGO4 proteins should include the generation of single and double knockouts to assess their functional redundancy and specific roles in gene regulation. Phenotypic characterization of these mutants under different conditions will help determine the unique and overlapping functions of AGO4 proteins. Investigation of the small RNA populations associated with AGO4 proteins and their target genes and transposable elements, together with transcriptome analysis of *ago4* mutants, will reveal regulatory networks controlled by AGO4 proteins in barley. Addressing these areas will deepen our understanding of RNA silencing mechanisms in barley and contribute to advances in crop improvement and plant biotechnology.

7 NEW SCIENTIFIC RESULTS

- We identified and classified 21 putative candidate genes belonging to the Argonaute family in barley through whole genome analysis.

- *In silico* analysis of five barley *AGO1* genes revealed distinct tissue and developmental stage expression patterns and the conservation of the miR168 target site among these genes.

- We showed that modifying the miRNA duplex structure or expressing artificial precursors can alter the loading efficiency of miR168 into AGO1-RISC.

- We demonstrated that barley AGO4 proteins have overlapping functionality with distinct small RNA binding properties upon heterologous complementation.

- We observed distinct regulatory properties of barley AGO4 proteins on transposable elements, especially on relatively short TEs, affecting cumulative sRNA abundance.

8 PUBLICATIONS LIST

Conference Presentations and Posters:

MBK Napok 30 (Gödöllő, 2020): presentation - "Role of *Arabidopsis* miRNA precursors in RISC-loading efficiency", Előadói Díj III..

XX. Genetikai Műhelyek Magyarországon Minikonferencia (Szeged, 2021): poster - "Functional analysis of barley *AGO4* genes in *Arabidopsis thaliana ago4-3* mutant background".

Microsymposium on RNA biology (Vienna, 2022): poster - "Genetic complementation analysis of barley *AGO4* genes in *Arabidopsis ago4* mutant".

FIKOK 2022 (Gödöllő, 2022): poster - "Identification of barley *AGO4* genes and complementation assay in *Arabidopsis*", abstract published on ISBN: 9789632699998.

GBI Nap (Gödöllő, 2022): presentation - "Genetic complementation analysis of barley *AGO4* genes in *Arabidopsis ago4* mutant", Előadói Díj I.

XXIX. Ifjúsági Tudományos Fórum (Keszthely, 2023): presentation - "Revealing the Diverse Functional Roles of Barley AGO4 Proteins in *Arabidopsis* through Heterologous Complementation Assay", conference paper published on ISBN: 9786156338082.

GBI Napok (Gödöllő, 2023): presentation – "Generation and characterization of barley *dcl3* RNAi mutants", Előadói Díj I.

Microsymposium on RNA biology (Vienna, 2024): poster - "Barley AGO4 proteins show overlapping functionality with distinct small RNA-binding properties in heterologous complementation".

Publications:

Hamar, E., Szaker, H. M., Kis, A., Dalmadi, A., **Miloro, F.**, Szittya, G., Taller, J., Gyula, P., Csorba, T., & Havelda, Z. (2020). Genome-Wide Identification of RNA Silencing-Related Genes and Their Expressional Analysis in Response to Heat Stress in Barley (Hordeum vulgare L.). Biomolecules 2020, Vol. 10, Page 929, 10(6), 929. https://doi.org/10.3390/BIOM10060929 (MDPI - Q2 - 2020).

Dalmadi, Á., **Miloro, F.**, Bálint, J., Várallyay, É., & Havelda, Z. (2021). Controlled RISC loading efficiency of miR168 defined by miRNA duplex structure adjusts ARGONAUTE1 homeostasis.

Nucleic Acids Research, 49(22), 12912–12928. https://doi.org/10.1093/NAR/GKAB1138 (Oxford Academic - D1 - 2021).

Contaldo, N., Zambon, Y., Galbacs, Z. N., **Miloro, F.**, Havelda, Z., Bertaccini, A., & Varallyay, E. (2023). Small RNA Profiling of Aster Yellows Phytoplasma-Infected Catharanthus roseus Plants Showing Different Symptoms. Genes, 14(5), 1114. https://doi.org/10.3390/GENES14051114 (MDPI - Q2 - 2023).

Miloro, F., Kis, A., Havelda, Z., & Dalmadi, Á. (2024). Barley AGO4 proteins show overlapping functionality with distinct small RNA-binding properties in heterologous complementation. Plant Cell Reports 2024 43:4, 43(4), 1–19. https://doi.org/10.1007/S00299-024-03177-Z (Springer Nature - D1 - 2024).