Sorting Out Small RNAs

V. Narry Kim[1,*]
[1]School of Biological Sciences and National Creative Research Initiatives Center, Seoul National University, Seoul, 151-742, Korea
*Correspondence: narrykim@snu.ac.kr
DOI 10.1016/j.cell.2008.03.015

Small RNAs carry out their functions by guiding Argonaute (AGO) proteins to their targets. Diverse types of small RNAs and multiple AGO proteins exist in most eukaryotic species, but how small RNAs are sorted into specific AGO complexes remains unclear. Two papers in this issue (Mi et al., 2008; Montgomery et al., 2008) now reveal the importance of the 5′ terminal nucleotide of the small RNA in the sorting process in *Arabidopsis*.

Initially described as developmental regulators in plants, Argonaute (AGO) proteins were later found to be core molecules in RNA silencing (Hutvagner and Simard, 2008). Small RNAs bind directly to AGO proteins and guide the AGO complex to their target molecules (DNA or RNA). In these complexes, it is the small RNA that confers the specificity of targeting, whereas the AGO protein determines the effect of RNA silencing. Two new studies in *Arabidopsis* by Mi et al. (2008) and Montgomery et al. (2008) now explore features of the small RNA that determine its targeting to specific AGO complexes. The choice of AGO complexes has important functional consequences given that different AGO proteins mediate diverse effects on RNA and chromatin.

Endogenous small RNAs are diverse and can be categorized into three classes: microRNAs (miRNAs), small interfering RNAs (siRNAs), and piwi-interacting RNAs (piRNAs). *Arabidopsis* has a particularly complex small RNA system (Chapman and Carrington, 2007). Apart from miRNAs, *Arabidopsis* contains three subclasses of endogenous siRNAs: *trans*-acting siRNAs (tasiRNAs), natural antisense transcript-derived siRNAs (natsiRNAs), and repeat-associated siRNAs (rasiRNAs). Each of these small RNA subclasses is preferentially incorporated into one or more of the ten AGO complexes in *Arabidopsis*. miRNAs are generated from imperfect hairpin structures, and cleavage of the hairpin by DICER-LIKE 1 (DCL1) releases a small RNA that is then usually incorporated into AGO1. Biogenesis of tasiRNAs involves miRNA-mediated cleavage of noncoding transcripts. After cleavage, RNA-dependent RNA polymerase 6 (RDR6) synthesizes double-stranded RNA (dsRNA) that is then processed by DCL4 into ~21 nucleotide small RNAs, which are then bound to AGO2. natsiRNAs are produced from natural antisense transcripts that are transcribed by convergent transcription of two overlapping genes. The dsRNA is processed by DCL2 and DCL1. The most abundant small RNAs in *Arabidopsis* are rasiRNAs that are ~24 nucleotides in length. They are derived from transposons and other repetitive elements and are associated with several AGO proteins, most significantly with AGO4.

Each type of small RNA somehow finds its way to a particular AGO protein. Failure to properly sort small RNAs could result in unwanted silencing reactions that could be potentially harmful to the organism. So how are these small RNAs sorted into specific AGO complexes? To address this issue, Mi and colleagues (2008) isolated the small RNAs that are associated with AGO1, AGO2, AGO4, and AGO5 complexes. The cDNAs made from these small RNAs were analyzed by high-throughput sequencing to retrieve 140,000 to 310,000 reads for each AGO-associated RNA pool. Although all categories of small RNAs are associated with each AGO to some extent, there is a clear preference for specific small RNA classes (Figure 1A). As expected, miRNAs are the major class of small RNA in AGO1 complexes (comprising ~82%), whereas AGO2 complexes are highly enriched for tasiRNAs. AGO2 complexes also contain other classes of small RNAs with a relatively high proportion of rasiRNAs. Consistent with previous reports, AGO4 complexes are predominantly

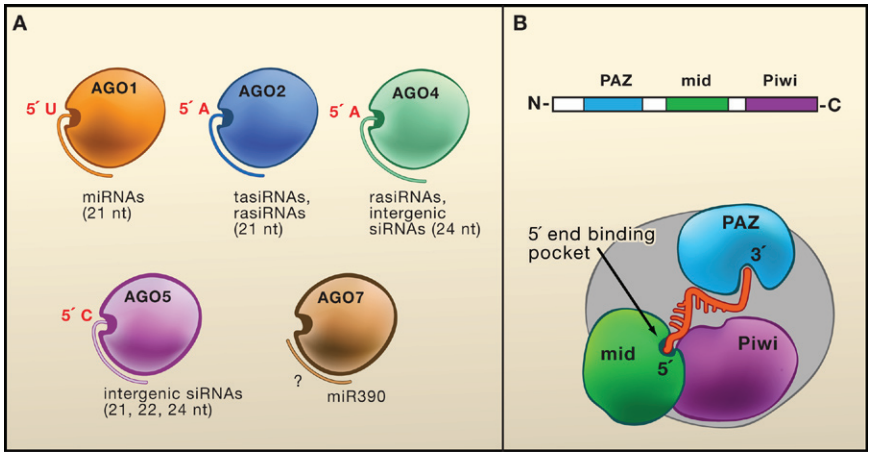

Figure 1. Argonaute Proteins and Small RNAs
(A) The Argonaute (AGO) proteins in *Arabidopsis* and their associated small RNAs. The preferred 5′ terminal nucleotide for each AGO protein is indicated. AGO binds specifically to miR390 in a 5′ end-independent manner. (B) Representative domain structure of an AGO protein (top) and schematic depiction of a model for AGO-small RNA interaction (bottom). The 5′ end nucleotide of the small RNA is inserted into a basic pocket in the mid domain. The 3′ end interacts with the PAZ domain. miRNA, microRNA; rasiRNA, repeat-associated short interfering RNA; tasiRNA, *trans*-acting siRNA.

associated with rasiRNAs. Meanwhile, AGO5 proteins are bound to small RNAs derived from unannotated intergenic sequences.

Bioinformatic analyses demonstrated a strong bias in the 5′ terminal nucleotide: uridine (86%) in the case of AGO1, adenosine (93%) for AGO2, adenosine (79%) for AGO4, and cytosine (83%) for AGO5. Interestingly, the opposite strands of miRNAs (called miRNAs*) with 5′ adenosines and 5′ cytosines are bound to AGO2 and AGO5, respectively. Another recent paper also reports that the 5′ nucleotide of small RNAs associated with *Arabidopsis* AGO2 and AGO5 are predominantly adenosine and cytosine, respectively (Takeda et al., 2008). Furthermore, miR163.2 containing a 5′ adenosine is preferentially associated with AGO2 instead of AGO1. This led Mi et al. (2008) to hypothesize that the binding affinity of AGO proteins for small RNAs is determined by the nucleotide at the 5′ end. Further biochemical and genetic experiments support this initial hypothesis. Most notably, immunoprecipitation experiments using chimeric AGO proteins illustrate the importance of the mid and piwi domains in the recognition of the 5′ terminal nucleotide (Figure 1B).

This observation is consistent with the structural studies on piwi-RNA complexes from Archaea. The 5′ nucleotide was found to be inserted in a pocket in the mid domain (Ma et al., 2005; Parker et al., 2005). The basic and aromatic residues lining the pocket are important for the anchoring of the 5′ end of the RNA. One can envision that the residues in the 5′ end binding pocket may differ between AGO proteins, and the difference may explain their preferences for particular 5′ terminal nucleotides. Thus, it would be of interest to solve the structures of different AGO proteins.

In a thorough analysis of the function of AGO7 and miR390 in TAS3 tasiRNA biogenesis, Montgomery and colleagues (2008) discovered that miR390 is specifically associated with AGO7. They also found that AGO1 and AGO2 can discriminate small RNAs based on the 5′ terminal nucleotide, which is consistent with the related study by Mi and colleagues. Interestingly, miR390, which begins with an adenosine, has a strong affinity for AGO7 and only a moderate affinity for AGO2. The structure and sequences in the miRNA precursor did not affect the selectivity for AGO7. Changing the 5′ end nucleotide also did not change binding to AGO7. Thus, AGO7 must have evolved to specifically recognize the miR390 sequences outside of the 5′ terminal nucleotide. Notably, although miR390 is generated by DCL1, it is not associated with AGO1. This result indicates that Dicer processing may be uncoupled from association with AGO. Similar uncoupling has been described recently in *C. elegans* and *Drosophila* (Hutvagner and Simard, 2008).

Does the 5′ end-recognition model fully explain the sorting of all small RNAs? Although certainly valid for the AGO1, AGO2, AGO4, and AGO5 proteins of *Arabidopsis*, other selection mechanisms are also likely to be at work. In the fruit fly *Drosophila*, for example, there are two AGO-like proteins, dAGO1 and dAGO2, that are responsible for miRNA and siRNA pathways, respectively (Hutvagner and Simard, 2008). miRNAs are processed by Dicer 1 (DCR1), whereas siRNAs are generated by DCR2. Small RNAs are sorted into either the dAGO1 or dAGO2 complex based on the structure of the dsRNA duplex (Tomari et al., 2007). If the duplex is perfectly matched (which is usually the case for siRNA precursors), the small RNA is routed into dAGO2. If the duplex has a bulge in the middle (as is the case with miRNA precursors) the RNA is shunted into dAGO1. Recent studies in the worm *Caenorhabditis elegans* also support this "precursor structure" model for small RNA sorting (Steiner et al., 2007). In fact, the two models (the 5′ end-recognition model and the precursor structure model) are not mutually exclusive. It is likely that both types of sorting are applicable to most if not all small RNA-AGO associations. In addition, there may be other determinants as yet unidentified. For instance, both AGO2 and AGO4 strongly bind to small RNAs with 5′ adenosines, but they share only a fraction of small RNAs in common (<8%). Perhaps the length of the small RNA may be another determinant. The small RNAs in the AGO2 and AGO4 complexes are 21 and 24 nt, respectively. Because the PAZ domain interacts with the 3′ end of siRNAs, only a fixed length of small RNA may be accommodated by each AGO. It is also plausible that the coupling of each AGO with different Dicer proteins may contribute to RNA sorting. In addition, considering the specific interaction between miR390 and AGO7 reported by Montgomery et al. (2008), specific interactions between small RNAs and AGO proteins may involve nucleotides other than the one at the 5′ end.

A number of small RNAs found in animals contain uridine at the 5′ end. It would be interesting to know whether animal miRNA-associated AGO proteins have a higher affinity for 5′ uridine residues. Piwi-interacting RNAs found in mammalian and insect germline tissues also begin predominantly with uridine (Aravin et al., 2007). The biogenesis mechanism of piRNAs remains largely elusive. Dicer proteins are not involved in the pathway, suggesting that there exist other nucleases that are responsible for the cleavage of long single-stranded RNA precursors. In light of the two reports in this issue, it is tempting to speculate that selective association of 5′ uridine-containing small RNAs with the piwi-like proteins may contribute to the biogenesis of piRNAs.

REFERENCES

Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). Science *318*, 761–764.

Chapman, E.J., and Carrington, J.C. (2007). Nat. Rev. Genet. *8*, 884–896.

Hutvagner, G., and Simard, M.J. (2008). Nat. Rev. Mol. Cell Biol. *9*, 22–32.

Ma, J.B., Yuan, Y.R., Meister, G., Pei, Y., Tuschl, T., and Patel, D.J. (2005). Nature *434*, 666–670.

Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., Wu, L., Li, S., Zhou, H., Long, C., et al. (2008). Cell, this issue.

Montgomery, T.A., Howell, M., Cuperus, J.T., Li, D., Hansen, J., Alexander, A., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008). Cell, this issue.

Parker, J.S., Roe, S.M., and Barford, D. (2005). Nature *434*, 663–666.

Steiner, F.A., Hoogstrate, S.W., Okihara, K.L., Thijssen, K.L., Ketting, R.F., Plasterk, R.H., and Srijen, T. (2007). Nat. Struct. Mol. Biol. *14*, 927–933.

Takeda, A., Iwasaki, S., Watanabe, T., Utsumi, M., and Watanabe, Y. (2008). Plant Cell Physiol. Published online March 14, 2008.

Tomari, Y., Du, T., and Zamore, P.D. (2007). Cell *130*, 299–308.