

Manuscript EMBO-2012-81355

## Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs

Filipp Frank, Jesse Hauver, Nahum Sonenberg and Bhushan Nagar

*Corresponding author: Bhushan Nagar, McGill University*

---

### Review timeline:

|                     |               |
|---------------------|---------------|
| Submission date:    | 11 March 2012 |
| Editorial Decision: | 16 April 2012 |
| Revision received:  | 06 June 2012  |
| Accepted:           | 03 July 2012  |

---

### Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision

16 April 2012

---

Thank you for submitting your manuscript to the EMBO Journal. Your study has now been seen by three referees and their comments are provided below.

As you can see the referees appreciate the analysis, but also find that some further work is needed to consider publication here. In particular further mutational analysis is needed to support the proposed binding modes. Should you be able to address the criticisms raised in full then we would consider a revised manuscript. I should add that it is EMBO Journal policy to allow only a single round of revision, and that it is therefore important to address the raised concerns at this stage.

When preparing your letter of response to the referees' comments, please bear in mind that this will form part of the Review Process File, and will therefore be available online to the community. For more details on our Transparent Editorial Process, please visit our website:  
<http://www.nature.com/emboj/about/process.html>

Thank you for the opportunity to consider your work for publication. I look forward to your revision.

Yours sincerely,

Editor  
The EMBO Journal

### REFeree REPORTS

Referee #1

In this manuscript, Frank et al. show crystal structures of the MID domain of AtAGO1, 2, and 5, and assessed their affinity to NMPs. The authors further determined crystal structures of AtAGO1 MID domain in complex with NMPs. Collectively, the presented data provide a structural explanation for how plant small RNAs are sorted into different AGO proteins according to the 5' nucleotide identity. A couple of concerns need to be addressed before publication in The EMBO Journal.

Major points:

1. Throughout the text, citations are surprisingly inappropriate. I almost suspect that there was an error in the reference management software. The authors should cover the references as precisely and completely as possible.
2. In Fig. 3D and Supplementary Fig. S7, the authors compare the structures of the nucleotide specificity loops of AtAGO1, 2 and 5. However, it is unclear how flexible the loop conformations are and by what interactions they are anchored. The authors should 1) show the electron density maps of the loop regions, and 2) show the stick models as in Supplementary Fig. 10 and specify the interactions surrounding the loop regions, for all three AGOs.
3. The authors failed to determine the structures of AtAgo2 and 5 in complex with NTPs, and resorted to manually docking nucleotides into the binding sites. It is important to experimentally validate the proposed model by mutational analysis at least for AtAgo2.

Minor points:

1. Page 3, "Dicer loads the small RNA duplex into an AGO in an ATP-dependent manner to form the pre-RISC". This is accurate only for fly Ago2, the best characterized and yet unique RISC assembly pathway. It was suggested that Dicer is not required for RISC loading in mammals and also for fly Ago1.
2. Page 18, "Interestingly, in Arabidopsis miRNAs are unusual in that they are generally devoid of mismatches". While plant miRNAs generally bind to highly complementary targets, many plant miRNA/miRNA\* duplexes do contain internal mismatches. If the authors wish to discuss this point, they should statistically show mismatch positions and frequency of miRNA/miRNA\* duplexes in plants and animals.
3. Definition of the nucleotide specificity loop region in Fig. 1B is different from (1 residue longer than) that in Supplementary Fig. S1 and that in a previous paper by the same authors.
4. Page 10, line 3, "Figure 6, Supplementary Figure S3" should presumably read "Figure 3, Supplementary Figure S6".

Referee #2

#### Structural basis for small RNA sorting by Arabidopsis Argonautes

In this manuscript, Frank et al. have analyzed the structures of the MID domain of three different Arabidopsis Argonaute proteins. Using single quantum coherence NMR titration, the authors measure the affinities of AtAgo1, AtAgo2 and AtAgo5 to the individual NMPs and find that AtAgo1 prefers UMP, AtAgo2 AMP and AtAgo5 CMP. To further characterize the nucleotide specificity in more detail, Frank et al. solved the structures of the three MID domains. NMPs were soaked into the crystals to draw conclusions regarding the nucleotide specificity of the analyzed Ago proteins. The obtained structures only allowed conclusions on the binding of UMP to AtAgo1. Here, they find that the asparagine side chain Asn687 is important for binding. Further modeling revealed that this Asn forces purines into the syn conformation thereby reducing binding affinity. Based on the hAgo2 structure of the specificity loop, the authors present a model for AtAgo2 specificity.

This is an extension of a recent paper by the same group on the sequence specificity of human Ago proteins. The authors have clearly presented their data and come up with a plausible model for AtAgo1 specificity for UMP. However, other points are less clear and should be further validated. A number of other specific points need to be addressed.

1. The novel finding in this manuscript is that AtAgo1 uses Asn687 to specifically contact UMP. The manuscript leaves the impression that the authors try hard to fill the space of a full article. There is a lot of text and discussion and the manuscript could clearly be more concise. More specifically, Figure 1 does not contribute much and could easily be merged with one of the other Figures.

2. All three MID domain structures look quite similar with the exception of the specificity loop. The authors state that the AtAgo5 loop is involved in crystal contacts and therefore distorted. How defined are the AtAgo1 and AtAgo2 specificity loops? Are they also involved in crystal contacts? The authors should add B factors for these critical residues.

3. Figure 4: the binding of CMP and UMP to AtAgo1 appears rather identical. The authors should explain why they observe a 3-fold difference in their binding affinities.

4. Figure 5: first, why did the authors use human Ago2 as basis for modeling nucleotide binding to AtAgo2? Isn't the AtAgo1 domain more similar to AtAgo2? Second, in their model on AtAgo2 AMP binding the authors claim that stacking of the base to tyr678 is inefficient. However, the determined binding efficiency (Figure 2) is similar to the other AtAgo MID domains. Therefore, the model seems not fully conclusive. The authors should calculate binding energies for the observed UMP and CMP binding to AtAgo1 and their models for AtAgo nucleotide binding.

4. The authors should perform many more point mutations to verify both the observed effects for AtAgo1 in the crystal structure and also their models on AtAgo2 and AtAgo5 nucleotide binding. For example, N687 could be mutated to Ala to show that it contributes to binding and more importantly, whether or not it contributes to purine repulsion. In addition, L697, which clashes with UMP in the AtAgo2 model should be mutated as well. For AtAgo2 and AtAgo5 the authors could systematically mutate the specificity loop in order to obtain a detailed verification of the proposed binding models.

#### Minor points

1. Introduction: some of the references are doubled in the first part of the introduction.

2. First part of the results section. Please add a reference after the sentence "...miR390 to any other nucleotide had no effect on its association."

#### Referee #3

This is a very interesting and technically well-executed study aimed at understanding the structural basis for small RNA sorting in Arabidopsis. Previously, the authors showed how the MID domain of human Ago2 recognizes 5' U and A bases in small RNAs. Here, the authors dive deeper into this area by exploring the structural basis for 5' nucleotide recognition by plant AGOs, which have a diverse range of preferences. Overall, the work is a significant advance in understanding AGO proteins. The authors present the following major results and conclusions:

1. Binding studies, using HSQC NMR titration experiments, reveal the 5'-nucleotide binding affinities for the MID domains of AtAGO1, AtAGO2, and AtAGO5. AtAGO1 prefers UMP, while AtAGO2 prefers AMP and AtAGO5 is the most promiscuous of the three examined, with a slight preference for CMP. These affinities are reminiscent of the bias seen in the small RNAs associated with each AGO in vivo, suggesting that the MID domain could be the main determinant of small RNA sorting.

2. Crystal structures of the AtAGO1, AtAGO2 and AtAGO5 MID domains reveal differences in a rigid nucleotide specificity loop that is proposed to confer specificity in recognition of 5'-nucleotides.

3. AtAGO1 crystals soaked with NMPs reveal the structural basis for 5'-nucleotide recognition by AtAGO1 and explain the preference for 5'-U or C. This is proposed to be largely due to the specificity loop residue Asn 687, which forms positive interactions with pyrimidines (hydrogen bonds) and negative interactions with purines (sterically forcing the less favorable syn conformation).

4. NMPs modeled into the AtAGO2 MID domain crystal structure provide a model for the major preference for 5'-C over all other nucleotides.

Major concerns:

1) There is an apparent disconnect between the affinities of the MID domains for NMPs in solution and the ability of the proteins to bind NMPs while in crystal form that is never properly addressed in the manuscript. For example, AtAGO1 has a strong preference for UMP in solution, but appears to bind UMP and CMP equally well in the crystal. In contrast, the affinity for AMP is only slightly worse than that of CMP ( $5 \pm 0.7$  mM vs.  $3 \pm 0.5$  mM), but the density for AMP is extremely weak. Moreover, AtAGO1 has a lower affinity for GMP than AMP in solution, yet the density for GMP is much clearer than the AMP density in Supplementary Fig. 9. The authors should address these discrepancies. As written, the idea that the NMPs bind to AGO in solution in the same fashion as in crystallized form is an assumption critical to most of the conclusions of the paper.

2) Similarly, the AtAGO1 and AtAGO2 MID domains appear to have crystallized in the same condition. The affinity of AtAGO1 for CMP is essentially the same as the affinity of AtAGO2 for AMP. And yet, CMP can be soaked into AtAGO1 crystals, but AMP cannot be soaked into AtAGO2 crystals. How do the authors explain this discrepancy? Does this not speak to the possibility that the AtAGO2 MID domain crystallized in a conformation incompetent for nucleotide binding? If so, that could complicate the modeling studies. Was this taken into account? What exactly do the authors mean, on page 13, when they say that determination of AtAGO2 and AtAGO5 in complex with nucleotides failed?

3) Page 7, the authors say that affinities for NMPs are weak because the authors are working with a single nucleotide from a physiological small RNA that otherwise makes extensive contact with multiple domains in AGO. But if the affinities for the 5' nucleotides are weak compared to the overall binding of the small RNA, how then does the MID domain provide any power for discrimination?

4) What was the basis of selecting only AtAGO1, AtAGO2, and AtAGO5? Why not also analyze AtAGO4 to compare the nucleotide specificity loops of AtAGO4 and AtAGO5, since both have bias for A and C, but very different selectivity loop sequences? Maybe there are some interesting differences between nucleotide specificity loops of highly selective AtAGOs versus less selective relatives? Why not look at AtAGO7? Even without structures, binding studies with NMPs would allow the authors to correlate the relative affinities of the MID domain with known biological biases (as they did so well for AtAGOs 1,2 and 5). Without any analysis of the MID domains from these AtAGOs the title of the paper is a bit of an over-statement.

5) In order to understand the exquisite selectivity of the AtAGO2 MID domain the authors rely on modeling the apo structure. Modeling suggests that selectivity is mediated by Asp676, which could clash with GMP, and Leu697, which could clash with purines. The modeling makes clear predictions that can be tested using the binding assay established in the paper. The authors should make MID domains with each of these residues mutated (for example, D676A should bind GMP because it would not clash, and L697A or L697V should be able to better bind purines than the wild type) and test the mutants in NMP binding assays. Without this analysis (and the issue of being unable to soak AMP into the structure) the insights into AtAGO2 selectivity are weak. With AtAGO2 weak and AtAGO5 insights weaker still, the title of the paper feels very misleading.

Minor comments:

6) page 10-11, "A comparison of this loop in all available structures..... demonstrates that this loop can assume variable conformations." This statement is a little confusing because the term "variable conformations" conjures the idea of a single flexible loop taking on variable conformations, but in fact, the authors argue that the loop in any given AGO is rigid. Adding the clause "between species" would clarify this point.

7) Page 13, "AMP and GMP do not have electron density" but, there is clearly some electron density for GMP in figure 4 and figure S9. This point is confusing.

3) Page 6, "These results show that the MID domains of AtAGOs can direct the sorting of different

classes of small RNAs into the appropriate AGO family member." These feels like a very strong statement to me, which I am not sure is fully supported by the data. The authors have not actually examined the process of sorting between AtAGO family members. Therefore, it has not yet been shown that the MID domain specificity alone is sufficient for small RNA sorting. It seems entirely possible to me that there is another sorting mechanism that acts upstream of MID domain binding, and the affinities of the MID domains for various NMPs reflect that each AGO was evolved to bind to different pre-selected small RNAs. This would actually be more consistent with the finding that the MID domain has relatively low affinity for NMPs compared to the overall affinity of AGO for small RNAs. Have the authors considered/excluded this possibility?

8) In some places the references seem inappropriate. For example, on page 16, "Our previous structural analysis of the hAGO2 MID domain in complex with NMPs implicated the nucleotide specificity loop as a critical structural determinant for 5'-bias (Kawaoka et al, 2011; Frank et al, 2010)." Are the authors suggesting that their structural analysis extended into the Kawaoka paper?? The references should be checked to make sure they make sense.

1st Revision - authors' response

06 June 2012

### **Reviewer 1:**

#### **Major points:**

*1. Throughout the text, citations are surprisingly inappropriate. I almost suspect that there was an error in the reference management software. The authors should cover the references as precisely and completely as possible.*

There was indeed an issue with the reference managing software. All references have now been corrected and double-checked.

*2. In Fig. 3D and Supplementary Fig. S7, the authors compare the structures of the nucleotide specificity loops of AtAGO1, 2 and 5. However, it is unclear how flexible the loop conformations are and by what interactions they are anchored. The authors should 1) show the electron density maps of the loop regions, and 2) show the stick models as in Supplementary Fig. 10 and specify the interactions surrounding the loop regions, for all three AGOs.*

We have included a supplementary figure (Supplementary Figure 8) showing stick models and the surrounding electron density of the loops. We also included B-factors of the loop backbone residues compared to the whole protein backbone to show that the loops are structured and rigid (Supplementary Table 4).

*3. The authors failed to determine the structures of AtAgo2 and 5 in complex with NTPs, and resorted to manually docking nucleotides into the bindings sites. It is important to experimentally validate the proposed model by mutational analysis at least for AtAgo2.*

We carried out additional experiments with mutants of all three proteins:

- 1) AtAGO1 N687A was generated (as proposed by reviewer 2) to provide further evidence that the long side chain of N687 blocks binding of purines.
- 2) AtAGO2 D676A was generated to confirm that D676 blocks binding of GMP. For AtAGO2 L697 we were unable to generate a mutant producing soluble protein, which may reflect its role as a structurally important residue. As a result, we have removed the section addressing this residue from the manuscript.
- 3) For AtAGO5 the loop was extended by an additional residue, which changed the nucleotide bias and further implicates the nucleotide specificity loop in nucleotide selectivity.

#### **Minor points:**

*1. Page 3, "Dicer loads the small RNA duplex into an AGO in an ATP-dependent manner to form the pre-RISC". This is accurate only for fly Ago2, the best characterized and yet unique RISC*

assembly pathway. It was suggested that Dicer is not required for RISC loading in mammals and also for fly Ago1.

This sentence was replaced by "AGO is then loaded in to form the RISC".

2. Page 18, "Interestingly, in *Arabidopsis* miRNAs are unusual in that they are generally devoid of mismatches". While plant miRNAs generally bind to highly complementary targets, many plant miRNA/miRNA\* duplexes do contain internal mismatches. If the authors wish to discuss this point, they should statistically show mismatch positions and frequency of miRNA/miRNA\* duplexes in plants and animals.

We removed the following paragraph from the discussion:

"Interestingly, in *Arabidopsis* miRNAs are unusual in that they are generally devoid of mismatches, which would increase duplex stability. Hence, the identity of the 5'-nucleotide in *Arabidopsis* RISC formation is probably more important for stabilizing the mature RISC rather than duplex loading."

3. Definition of the nucleotide specificity loop region in Fig. 1B is different from (1 residue longer than) that in Supplementary Fig. S1 and that in a previous paper by the same authors.

Figure 1, which is now Supplementary Figure 1, was corrected to include the proper loop boundaries.

4. Page 10, line 3, "Figure 6, Supplementary Figure S3" should presumably read "Figure 3, Supplementary Figure S6".

This mistake was corrected according to the changes we made to our figures during revision of the manuscript.

## **Reviewer 2:**

### **Major points:**

1. The novel finding in this manuscript is that AtAgo1 uses Asn687 to specifically contact UMP. The manuscript leaves the impression that the authors try hard to fill the space of a full article. There is a lot of text and discussion and the manuscript could clearly be more concise. More specifically, Figure 1 does not contribute much and could easily be merged with one of the other Figures.

Our manuscript confirms that the MID domains of AtAGO1, AtAGO2, and AtAGO5 selectively interact with small RNA 5' nucleotides, which was previously proposed based on the analysis of small RNAs associated with these proteins. Furthermore, we solved the crystal structures of these three domains and provide evidence about the structural basis of selectivity including the detailed characterization of 5' nucleotide interactions with the AtAGO1 MID domain.

To address the reviewers concerns about the length of the manuscript we have now reduced lengthy discussions. In particular, we removed the section on modeling of residue L697 in AtAGO2, for which we were unable to generate soluble mutant protein to test our hypothesis. Furthermore, Figure 1 was moved to the supplementary material.

2. All three MID domain structures look quite similar with the exception of the specificity loop. The authors state that the AtAgo5 loop is involved in crystal contacts and therefore distorted. How defined are the AtAgo1 and AtAgo2 specificity loops? Are they also involved in crystal contacts? The authors should add B factors for these critical residues.

AtAGO5 has a negative charge coming from the C-terminus of a symmetry related molecule inserted into the NMP binding site and interacts with the nucleotide specificity loop. Neither AtAGO1 or AtAGO2 have crystal contacts near the nucleotide specificity loop. We have now included a supplementary figure (Supplementary Figure 8) showing stick models and the surrounding electron density of the loops. We also included B-factors of the loop backbone residues compared to the whole protein backbone to show that the loops are structured and rigid

(Supplementary Table 4). Additionally, we show crystal contacts for AtAGO1 and AtAGO2 in Supplementary Figure 12.

*3. Figure 4: the binding of CMP and UMP to AtAgo1 appears rather identical. The authors should explain why they observe a 3-fold difference in their binding affinities.*

We do not expect the electron densities to correlate perfectly with the binding constants determined from NMR. The observed electron density is affected by systematic errors such as the protein molecules being fixed within a crystal lattice and variation in data quality, crystal resolution, data completeness, B-factors, radiation damage, etc. Thus we use the electron density more as a qualitative representation of binding to understand the molecular mechanism behind the interaction. Both CMP and UMP have similar densities in the crystal environment, but the density cannot reflect subtle differences in binding (for example due to electronic effects) that are measured by the NMR titrations in the solution phase.

*4. Figure 5: first, why did the authors use human Ago2 as basis for modeling nucleotide binding to AtAgo2? Isn't the AtAgo1 domain more similar to AtAgo2? Second, in their model on AtAgo2 AMP binding the authors claim that stacking of the base to tyr678 is inefficient. However, the determined binding efficiency (Figure 2) is similar to the other AtAgo MID domains. Therefore, the model seems not fully conclusive. The authors should calculate binding energies for the observed UMP and CMP binding to AtAgo1 and their models for AtAgo nucleotide binding.*

First comment: AtAGO2 is more similar to human AGO2 than to atAGO1 (compare sequence identity or C-alpha RMSD values in Supplementary Table 3 or see Supplementary Figure 13). Furthermore, in atAGO1 the base of AMP adopts the less energetically favored *syn*-conformation, which, when modeled into the structure of atAGO2, does not make contacts with the protein and therefore is unsuitable for modeling.

Second comment: comparing the average affinity of all nucleotides or the affinity of the strongest or weakest binders, AtAGO2 does in fact have clearly the lowest affinity for nucleoside monophosphates. These differences are as high as 13-fold (comparing the lowest affinity binders between the three proteins). A simple explanation could be the unusual (in comparison to all other MID domain structures) conformation of this tyrosine residue. To reflect the concerns of reviewer 2 we changed the sentence addressing this to:

“This reduction in stacking **may** explain why the binding affinities for NMPs to AtAGO2 determined above are weaker than others. “

*4. The authors should perform many more point mutations to verify both the observed effects for AtAgo1 in the crystal structure and also their models on AtAgo2 and AtAgo5 nucleotide binding. For example, N687 could be mutated to Ala to show that it contributes to binding and more importantly, whether or not it contributes to purine repulsion. In addition, L697, which clashes with UMP in the AtAgo2 model should be mutated as well. For AtAgo2 and AtAgo5 the authors could systematically mutate the specificity loop in order to obtain a detailed verification of the proposed binding models.*

We carried out additional experiments with mutants of all three proteins:

- 4) AtAGO1 N687A was generated (as proposed by reviewer 2) to provide further evidence that the long side chain of N687 blocks binding of purines.
- 5) AtAGO2 D676A was generated to confirm that D676 blocks binding of GMP. For AtAGO2 L697 we were unable to generate a mutant producing soluble protein, which may reflect its role as a structurally important residue. As a result, we have removed the section addressing this residue from the manuscript.
- 6) For AtAGO5 the loop was extended by an additional residue, which changed the nucleotide bias and further implicates the nucleotide specificity loop in nucleotide selectivity.

#### **Minor points:**

*1. Introduction: some of the references are doubled in the first part of the introduction.*

There was an issue with the reference managing software. All references have now been corrected.

2. First part of the results section. Please add a reference after the sentence "...miR390 to any other nucleotide had no effect on its association."

An additional reference to the paper of Mi *et al.* (Cell, 2008) has been added here.

### **Reviewer 3:**

#### **Major points:**

1) There is an apparent disconnect between the affinities of the MID domains for NMPs in solution and the ability of the proteins to bind NMPs while in crystal form that is never properly addressed in the manuscript. For example, AtAGO1 has a strong preference for UMP in solution, but appears to bind UMP and CMP equally well in the crystal. In contrast, the affinity for AMP is only slightly worse than that of CMP ( $5 \pm 0.7$  mM vs.  $3 \pm 0.5$  mM), but the density for AMP is extremely weak. Moreover, AtAGO1 has a lower affinity for GMP than AMP in solution, yet the density for GMP density is much clearer than the AMP density in Supplementary Fig. 9. The authors should address these discrepancies. As written, the idea that the NMPs bind to AGO in solution in the same fashion as in crystallized form is an assumption critical to most of the conclusions of the paper.

We do not expect the electron densities to correlate perfectly with the binding constants determined from NMR. The observed electron density is affected by systematic errors such as the protein molecules being fixed within a crystal lattice and variation in data quality, crystal resolution, data completeness, B-factors, radiation damage, etc. Thus we use the electron density more as a qualitative representation of binding to understand the molecular mechanism behind the interaction. Both CMP and UMP have similar densities in the crystal environment, but the density cannot reflect subtle differences in binding (for example due to electronic effects) that are measured by the NMR titrations in the solution phase.

2) Similarly, the AtAGO1 and AtAGO2 MID domains appear to have crystallized in the same condition. The affinity of AtAGO1 for CMP is essentially the same as the affinity of AtAGO2 for AMP. And yet, CMP can be soaked into AtAGO1 crystals, but AMP cannot be soaked into AtAGO2 crystals. How do the authors explain this discrepancy? Does this not speak to the possibility that the AtAGO2 MID domain crystallized in a conformation incompetent for nucleotide binding? If so, that could complicate the modeling studies. Was this taken into account? What exactly do the authors mean, on page 13, when they say that determination of AtAGO2 and AtAGO5 in complex with nucleotides failed?

The following sentence was added in the "Materials and methods" section to explain that AtAGO2 could not be soaked with NMPs:

"The MID domains of AtAGO1 and AtAGO2 crystallized in the same condition, which contained a high concentration of sulfate (2M ammonium sulfate). Sulfate occupies the 5' nucleotide binding site in both structures and soaking of nucleotide in this condition does not result in protein-NMP complexes. Therefore, for the purpose of soaking, crystals of AtAGO1 MID domain were transferred to a drop of different composition containing less ammonium sulfate (0.2 M ammonium sulphate, 0.1 M sodium acetate pH 4.6, 25% PEG 4000, 20% glycerol, and 20 mM NMP). AtAGO2 MID domain crystals, however, are very fragile and dissolve when transferred to drops containing less ammonium sulfate and, therefore, soaking was not successful."

In the case of AtAGO5, soaking also did not result in complexes, which is due to crystal contacts blocking the 5' nucleotide binding pocket.

As for the modeling of AtAGO2, there are no crystal contacts in the vicinity of the nucleotide specificity loop (see the new Supplementary Figure 12). Conversely, AtAGO5 has a negative charge coming from the C-terminus of a symmetry related molecule inserted into the NMP binding site and interacts with the nucleotide specificity loop, potentially distorting it and blocking nucleotide binding.

3) Page 7, the authors say that affinities for NMPs are weak because the authors are working with a single nucleotide from a physiological small RNA that otherwise makes extensive contact with multiple domains in AGO. But if the affinities for the 5' nucleotides are weak compared to the



*overall binding of the small RNA, how then does the MID domain provide any power for discrimination?*

It has been shown by a number of studies that the 5' nucleotide affects directly the small RNA loading process, most notably by Kawamata *et al.* (EMBO Reports, 2010) and Mi *et al.* (Cell, 2008). In particular, Kawamata *et al.* use synthetic small RNA duplexes and changing only the 5' nucleotide in a duplex changes the efficiency of generating the pre-RISC and eventually the RISC.

Based on the available literature we believe that the 5' nucleotide's selective recognition in the MID domain plays a role in the handover from Dicer, which is a dynamic process in which the identity of the 5' nucleotide is probed. This likely occurs in combination with the asymmetry rule: once processed into a small RNA duplex by Dicer, this duplex can only be loaded into an AGO protein if the 5' end can be unwound to some extent so that 5' nucleotide can be presented to the AGO MID domain. In addition, only a strong interaction of the 5' nucleotide in the MID domain then results in efficient handover and loading. In such a scenario, the affinity of the 5' nucleotide is most important as an anchor as a first step of loading and once it is bound in the MID domain the remainder of the small RNA is put into place.

*4) What was the basis of selecting only AtAGO1, AtAGO2, and AtAGO5? Why not also analyze AtAGO4 to compare the nucleotide specificity loops of AtAGO4 and AtAGO5, since both have bias for A and C, but very different selectivity loop sequences? Maybe there are some interesting differences between nucleotide specificity loops of highly selective AtAGOs versus less selective relatives? Why not look at AtAGO7? Even without structures, binding studies with NMPs would allow the authors to correlate the relative affinities of the MID domain with known biological biases (as they did so well for AtAGOs 1,2 and 5). Without any analysis of the MID domains from these AtAGOs the title of the paper is a bit of an over-statement.*

We attempted expression of AtAGO4 and AtAGO7 MID domains. Unfortunately, these MID domains did not produce soluble proteins. Furthermore, we believe that AtAGO1, AtAGO2, and AtAGO5 represent a very appropriate selection of possible nucleotide specificities.

*5) In order to understand the exquisite selectivity of the AtAGO2 MID domain the authors rely on modeling the apo structure. Modeling suggests that selectivity is mediated by Asp676, which could clash with GMP, and Leu697, which could clash with purines. The modeling makes clear predictions that can be tested using the binding assay established in the paper. The authors should make MID domains with each of these residues mutated (for example, D676A should bind GMP because it would not clash, and L697A or L697V should be able to better bind purines than the wild type) and test the mutants in NMP binding assays. Without this analysis (and the issue of being unable to soak AMP into the structure) the insights into AtAGO2 selectivity are weak. With AtAGO2 weak and AtAGO5 insights weaker still, the title of the paper feels very misleading.*

We carried out additional experiments with mutants of all three proteins:

- 1) AtAGO1 N687A was generated (as proposed by reviewer 2) to provide further evidence that the long side chain of N687 blocks binding of purines.
- 2) AtAGO2 D676A was generated to confirm that D676 blocks binding of GMP. For AtAGO2 L697 we were unable to generate a mutant producing soluble protein, which may reflect its role as a structurally important residue. As a result, we have removed the section addressing this residue from the manuscript.
- 3) For AtAGO5 the loop was extended by an additional residue, which changed the nucleotide bias and further implicates the nucleotide specificity loop in nucleotide selectivity.

#### **Minor comments:**

*6) page 10-11, "A comparison of this loop in all available structures..... demonstrates that this loop can assume variable conformations." This statement is a little confusing because the term "variable conformations" conjures the idea of a single flexible loop taking on variable conformations, but in fact, the authors argue that the loop in any given AGO is rigid. Adding the clause "between species" would clarify this point.*

This suggestion was adopted and the sentence now reads: "A comparison of this loop in all available structures of eukaryotic MID domains, including the AtAGO structures determined here,

hAGO2 and *Neurospora crassa* AGO (QDE-2), demonstrates that this loop can assume variable conformations **between species** (Supplementary Fig. 9 online). “

7) Page 13, "AMP and GMP do not have electron density" but, there is clearly some electron density for GMP in figure 4 and figure S9. This point is confusing.

To make this clearer we have changed the sentence and it now reads as follows:

“... AMP and GMP, do not have **comparable** electron density for the nucleotide sugar or the base ...”

The electron density shown in Supplementary Figure 11 is contoured at a much lower level (reflecting the weaker binding) than in Figure 3 and was used to model these nucleotides. This is explained in the sentence that followed (“Surprisingly, at lower contour levels residual electron density for the sugar and base of both GMP and AMP are visible, but with the bases oriented in syn conformations (Supplementary Fig. 11 online).”)

3) Page 6, "These results show that the MID domains of AtAGOs can direct the sorting of different classes of small RNAs into the appropriate AGO family member." These feels like a very strong statement to me, which I am not sure is fully supported by the data. The authors have not actually examined the process of sorting between AtAGO family members. Therefore, it has not yet been shown that the MID domain specificity alone is sufficient for small RNA sorting. It seems entirely possible to me that there is another sorting mechanism that acts upstream of MID domain binding, and the affinities of the MID domains for various NMPs reflect that each AGO was evolved to bind to different pre-selected small RNAs. This would actually be more consistent with the finding that the MID domain has relatively low affinity for NMPs compared to the overall affinity of AGO for small RNAs. Have the authors considered/excluded this possibility?

The reviewer is correct in that we do not actually address the sorting mechanism itself. This was done in depth in the paper by Mi *et al* published in 2008 in *Cell* (“Sorting of Small RNAs into *Arabidopsis* Argonaute Complexes Is Directed by the 5' Terminal Nucleotide”), which provides compelling evidence for the model that the 5' nucleotide directs sorting of small RNAs into the different AGO proteins. Among many other experiments, they specifically addressed the involvement of the MID domain by expressing chimeric proteins of AtAGO1 and AtAGO2, in which their respective MID domains are switched. The resulting chimeric proteins associate with miR390a wt (5'-U) and mutant (5'-A) according to their MID domain's selectivity, which is strong evidence for the idea that the MID domain determines small RNA sorting into AGO proteins based on 5' nucleotides.

Our data now confirms the mechanism proposed by the authors of that study, which proposed that there are specific interactions between the small RNA 5' nucleotide and the MID domains of *Arabidopsis* AGOs. To acknowledge the concerns of the reviewer we changed the sentence mentioned by the reviewer to:

“These results show that the effect of 5' nucleotides in different classes of small RNAs on sorting is realized through their base-specific recognition in the Argonaute MID domains.”

Additionally, we do not, of course, rule out the possibility that there are other sorting mechanisms involved. In fact, this must be the case since four nucleotides do not provide enough bias to sort small RNAs specifically between ten different AGOs. These mechanisms are most likely a combination of differential interaction with Dicers, spatiotemporal expression of small RNAs relative to AGOs and Dicers, and other possible mechanisms. This manuscript, however, only investigates the effects of the 5' nucleotides on sorting that had been reported previously and confirms that this effect is based on selectivity in the AGO MID domains.

8) In some places the references seem inappropriate. For example, on page 16, "Our previous structural analysis of the hAGO2 MID domain in complex with NMPs implicated the nucleotide specificity loop as a critical structural determinant for 5'-bias (Kawaoka *et al*, 2011; Frank *et al*, 2010)." Are the authors suggesting that their structural analysis extended into the Kawaoka paper?? The references should be checked to make sure they make sense.

There was an issue with the reference managing software. All references have now been corrected.

Thank you for submitting your revised manuscript to the EMBO Journal. Your revision has now been seen by the referees and their comments are provided below.

As you can see, they appreciate the introduced changes and support publication here. I am therefore pleased to proceed with the acceptance of the paper.

One last point is that we required that structural data is deposited in an appropriate database and that the accession numbers are provided in the main manuscript - please see our guide to authors. Unless I missed it, I didn't see this info in the manuscript. Could you send me a modified version that contains the accession numbers? We can also do this at proof stage, but it would be easier to get it sorted out now.

Yours sincerely

Editor  
The EMBO Journal

#### REFeree REPORTS

Referee #1:

The authors have adequately addressed my previous concerns and, in my opinion, the manuscript is now ready for publication.

Referee #2:

Frank et al. have revised their manuscript entitled "Arabidopsis Argonaute MID domain use their nucleotide specificity loop to sort small RNAs". In the new version, the authors have addressed all points that I had raised and therefore I am satisfied with the new version of the manuscript.

Referee #3:

The authors addressed all concerns to my satisfaction. The addition the mutational analysis adds a lot to the paper and considerably strengthens the proposed models.