The EMBO Journal (2012) 31, 3588–3595 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

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

Filipp Frank[1,2,3], Jesse Hauver[1], Nahum Sonenberg[1,3] and Bhushan Nagar[1,2,*]

[1]Department of Biochemistry, McGill University, Montreal, Quebec, Canada, [2]Groupe de Recherche Axe sur la Structure des Proteines, McGill University, Montreal, Quebec, Canada and [3]Goodman Cancer Center, McGill University, Montreal, Quebec, Canada

The 5′-nucleotide of small RNAs associates directly with the MID domain of Argonaute (AGO) proteins. In humans, the identity of the 5′-base is sensed by the MID domain nucleotide specificity loop and regulates the integrity of miRNAs. In *Arabidopsis thaliana,* the 5′-nucleotide also controls sorting of small RNAs into the appropriate member of the AGO family; however, the structural basis for this mechanism is unknown. Here, we present crystal structures of the MID domain from three *Arabidopsis* AGOs, AtAGO1, AtAGO2 and AtAGO5, and characterize their interactions with nucleoside monophosphates (NMPs). In AtAGOs, the nucleotide specificity loop also senses the identity of the 5′-nucleotide but uses more diverse modes of recognition owing to the greater complexity of small RNAs found in plants. Binding analyses of these interactions reveal a strong correlation between their affinities and evolutionary conservation.

The EMBO Journal (2012) **31,** 3588–3595. doi:10.1038/emboj.2012.204; Published online 31 July 2012
Subject Categories: RNA; plant biology
Keywords: Argonaute MID domains; Argonaute sorting; crystallography; NMR; small RNAs

Introduction

Endogenous small RNA molecules (~20–30 nucleotides) help regulate virtually every cellular process in eukaryotes by repressing the expression of target genes (Ghildiyal and Zamore, 2009). Gene silencing can occur either transcriptionally through chromosomal modifications or post-transcriptionally through mRNA cleavage, mRNA destabilization and/or translational repression. Small RNAs contain a signature 5′-phosphate and 3′-hydroxyl, and associate with Argonaute (AGO) family members, which are the core protein component of the RNA-induced silencing complex (RISC). AGOs consist of four distinct domains, which make extensive contacts with the associated small RNA: the PAZ domain recognizes the 3′-end (Lingel *et al*, 2004; 2003; Ma *et al*, 2004); the PIWI domain adopts an RNase-H fold and confers target cleavage activity to certain AGOs (Song *et al*, 2003; Yuan *et al*, 2005; Kawamura *et al*, 2008); the MID domain interacts with the 5′-end (Ma *et al*, 2005; Parker *et al*, 2005; Frank *et al*, 2010); the function of the N-terminal domain is not known, but was recently proposed to be involved in duplex unwinding (Kwak and Tomari, 2012).

Micro (mi)RNAs and short-interfering (si)RNAs are two major classes of small RNAs that are initially processed from longer, double-stranded RNA precursors by one of the Dicer enzymes. AGO is then loaded to form the RISC (Kawamata *et al*, 2009). During loading, two important decisions are made with respect to the fate of the small RNA and its targets: strand selection and AGO sorting (Czech and Hannon, 2011). AGO sorting is the process by which a particular class of small RNA becomes associated with a specific AGO. Strand selection determines which strand of the precursor duplex remains associated with AGO (guide strand) and which strand is removed (passenger strand). The complex of AGO with the guide strand is called the mature RISC, as it is now competent to guide the AGO to complementary nucleic acid targets via base pairing. For miRNAs, this occurs mainly through interactions with nucleotides 2–8, termed as the seed sequence (Lewis *et al*, 2003). The 5′-nucleotide (position 1) of the small RNA flips out of the duplex to interact with the MID domain (Ma *et al*, 2005; Parker *et al*, 2005; Wang *et al*, 2009). Upon binding to a cognate target, different AGOs dictate specific modes of repression. It is therefore critical that individual small RNAs associate with their appropriate AGO partner.

Early studies found that the strand with the 5′-terminus at the thermodynamically less stable end of the duplex, known as the asymmetry rule, is selected as the guide (Khvorova *et al*, 2003; Schwarz *et al*, 2003). However, in many cases the passenger strand could also be functionally loaded into AGOs (Mi *et al*, 2008; Takeda *et al*, 2008; Czech *et al*, 2009; Ghildiyal *et al*, 2010), suggesting that the asymmetry rule is not a complete explanation for strand selection. The subsequent discovery that swapping the 5′-nucleotides of a small RNA duplex made it possible to reverse their association with the intended AGOs (Mi *et al*, 2008; Takeda *et al*, 2008) indicated a critical role for the identity of the 5′-nucleotide in controlling strand selection and concomitantly AGO sorting. Indeed, analysis of the distribution of AGO-associated small RNAs in *Arabidopsis thaliana* (At) confirmed this notion by revealing that different AGOs associate with small RNAs based primarily on the identity of the 5′-nucleotide (Mi *et al*, 2008).

The evolutionary signature for this function of the 5′-nucleotide is imprinted in sequence conservation profiles of small RNAs, which reveal that each class of endogenous small RNAs displays a bias towards certain 5′-nucleotides. miRNAs in plants and animals (Lau *et al*, 2001; Czech *et al*, 2009; Ghildiyal *et al*, 2010), as well as miRNA-like RNAs in *Neurospora crassa* display a strong bias for uridine (U) at their 5′-ends (Lee *et al*, 2010); endogenous siRNAs in flies (Ghildiyal *et al*, 2008; Kawamura *et al*, 2008) and intergenic siRNAs from plants (Mi *et al*, 2008) tend to start with cytidine

*Corresponding author. Department of Biochemistry, McGill University, 3649 Promenade Sir William Osler, Bellini Building, Room 464, Montreal, Quebec, Canada H3G 0B1. Tel.: +1 514 398 7272; Fax: +1 514 398 2983; E-mail: bhushan.nagar@mcgill.ca

Received: 11 March 2012; accepted: 3 July 2012; published online: 31 July 2012

(C); endogenous triphosphorylated-22G and monophosphorylated-26G siRNAs in *Caenorhabditis elegans* begin with guanosine (G) (Ruby *et al*, 2006; Gu *et al*, 2009); piRNAs have a 5′-U (Brennecke *et al*, 2007; Lau *et al*, 2009), and the 5′-nucleotide of tasiRNAs and rasiRNAs in plants is usually adenosine (A) (Mi *et al*, 2008). These biases are reflected in the selective recruitment of different small RNA classes by different AGOs and, therefore, it is likely that both the asymmetry rule and the identity of the 5′-nucleotide determine which strand of a precursor duplex becomes associated with a particular AGO.

Insights into the structural basis for the selective recognition of the 5′-nucleotide came by way of crystallographic and binding analysis of the human AGO2 (hAGO2) MID domain with nucleoside monophosphates (NMPs; Frank *et al*, 2010). This study revealed that a rigid loop in the MID domain, termed as the *nucleotide specificity loop*, makes specific hydrogen bonds through peptide backbone atoms only with either uridine monophosphate (UMP) or adenosine monophosphate (AMP) but not with cytidine monophosphate (CMP) or guanosine monophosphate (GMP). However, in humans the two major known classes of endogenous small RNAs, miRNAs and piRNAs, have strong biases towards 5′-U. Thus, the characteristic of small RNA sorting is not readily apparent in the human RNA silencing system. *Arabidopsis*, however, presents a particularly striking example of small RNA sorting in that it contains 10 AGO family members among which are sorted a variety of different classes of small RNAs (miRNAs, siRNAs, tasiRNAs and rasiRNAs) based on the identity of the 5′-nucleotide (Mi *et al*, 2008). Whether AtAGOs use a similar structural mechanism to distinguish different 5′-ends as human AGOs do is unknown.

To address this question, we determined crystal structures of the MID domains from three different AtAGOs with distinct nucleotide biases: AtAGO1, AtAGO2 and AtAGO5. The structures of AtAGO1 in complex with NMPs revealed that, as with hAGO2, the *nucleotide specificity loop* plays a critical role in nucleotide selectivity, but in this case, by way of specific side-chain interactions. Dissociation constants determined using NMR confirm that the 5′-nucleotide biases observed in *Arabidopsis* small RNAs can be explained by 5′-selectivity by the AGO MID domain. 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.

Results

Of the 10 AtAGOs, only AtAGO1, AtAGO2, AtAGO4 and AtAGO5 have verified 5′-nucleotide biases, and AtAGO7 has no detectable bias (Montgomery *et al*, 2008; Mi *et al*, 2008; Supplementary Figure 1A). Sequence analysis of the region encompassing the potential nucleotide specificity loop in these proteins based on alignment with the hAGO2 MID domain reveals a divergent amino-acid composition that may underlie their distinct nucleotide biases (Supplementary Figures 1 and 2). We obtained diffraction quality crystals from the MID domains of AtAGO1, AtAGO2 and AtAGO5.

NMP binding to AtAGO MID domains correlates with known 5′-nucleotide biases

To understand if sequence biases at the 5′-position of small RNAs can be traced to their interaction with the MID domain, we measured the binding affinities of the isolated MID domain with NMPs, which mimic the 5′-end of the small RNA. Note, we expect to measure weak absolute affinities since we are working with only a single nucleotide from a physiological small RNA that otherwise makes extensive contact with multiple domains in AGO. Affinities for full-length AGOs with small RNAs have been reported to be in the low μM to nM range (Lima *et al*, 2009; Tan *et al*, 2009). To determine affinities, we used ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) NMR titration experiments, which are ideal for measuring low affinity (μM to mM range) interactions.

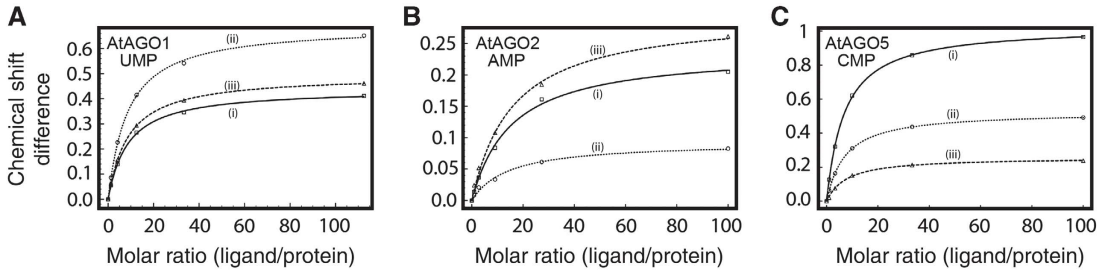
The binding analysis between NMPs and the MID domains of AtAGO1, AtAGO2 and AtAGO5 is shown in Figure 1. With AtAGO1 we observed the expected highest affinity for UMP (1 mM), while CMP (3 mM), GMP (6 mM) and AMP (5 mM) have much lower affinities (Figure 1A, D and E; Supplementary Figures 3 and 4). In accordance with the nucleotide bias profile, CMP is the second best binder, suggesting that AtAGO1 prefers pyrimidine nucleotides to purines. This is in contrast to hAGO2, which has the strongest bias for 5′-U, but its second most preferred 5′-nucleotide is A, suggesting that a different mechanism of nucleotide selection may be operating in AtAGO1.

For AtAGO2 we observed the following binding affinities for NMPs: AMP (3 mM) interacted with the greatest affinity, whereas CMP (79 mM), GMP (27 mM) and UMP (25 mM) bound more weakly (Figure 1B, D and E; Supplementary Figures 3 and 5). The extent of discrimination in binding between AMP and the other NMPs ranges from 8- to 25-fold, which is in agreement with AtAGO2 being the most stringent in its selectivity for only 5′-A. Thus, AtAGO2 contrasts with AtAGO1 and hAGO2, which both have a preference for 5′-U, but also permit a second type of nucleotide to bind. This observation may indicate tighter regulation of small RNA interactions in AtAGO2.

AtAGO5 displays the least pronounced selection bias for NMPs (Figure 1C–E; Supplementary Figures 3 and 6). It interacts most strongly with CMP (2.0 mM). However, AMP (2.3 mM) and GMP (3.0 mM) also bind with similar affinities, while only UMP (5.9 mM) has significantly weaker affinity. Although the fold differences here are less, the results are consistent with previous reports where the bias in AtAGO5 for associating with small RNAs containing a 5′-C was the least stringent among AtAGOs (Mi *et al*, 2008). Thus, it appears that the observed binding affinities of NMPs for the MID domains correlate well with the conservation of the 5′-nucleotide in small RNA sequences.

Crystal structures of the MID domains from AtAGO1, AtAGO2 and AtAGO5

It was previously shown that the position of the NMPs crystallized within the MID domain mimics precisely the position of the 5′-end of a small RNA bound to a full-length AGO (Frank *et al*, 2010), and was used to unravel the mechanism of 5′-nucleotide bias in human miRNAs by hAGO2. In hAGO2, the nucleotide specificity loop makes specific interactions with only U or A, but not with C or G, and therefore contributes to a nucleotide bias at the 5′-position of miRNAs. Importantly, the sequence of this loop is highly conserved in animal AGOs of the miRNA pathway (e.g., human AGO1-4, *C. elegans* Alg-1 and Alg-2, *Drosophila*



D

	AMP (mM)	CMP (mM)	GMP (mM)	UMP (mM)
AtAGO1	4.99 ± 0.67	3.00 ± 0.48	6.13 ± 1.03	0.98 ± 0.01
AtAGO2	3.22 ± 0.17	79.28 ± 3.99	26.86 ± 1.13	25.14 ± 2.41
AtAGO5	2.30 ± 0.37	2.00 ± 0.09	2.98 ± 0.35	5.90 ± 0.27

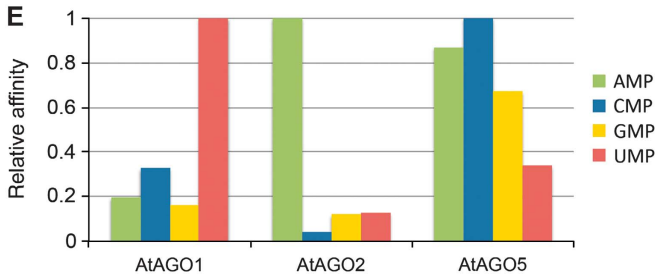


Figure 1 Binding affinities of NMPs with the MID domains of AtAGO1, AtAGO2 and AtAGO5 probed by NMR titration spectroscopy. (**A–C**) Chemical shift differences were calculated from ^{1}H-^{15}N HSQC NMR titration experiments and plotted as a function of the molar ratio (nucleotide/protein). Representative binding curves for AtAGO1, AtAGO2 and AtAGO5, with their preferred 5′-nucleotides, UMP, AMP and CMP, respectively, are shown. The data for multiple peaks were fit using the maximum shift and dissociation constant (K_D) as adjustable parameters. (**D**) Summary of dissociation constants of NMPs with MID domains from AtAGO1, AtAGO2 and AtAGO5. Values are means of curve fits from three different peaks with indicated standard deviations. The complete set of NMR spectra and curve fits for all interactions are shown in Supplementary Figures 3–5. (**E**) Graphical representation of data shown in (**D**). Vertical bars represent the association constant of the interactions ($K_A = K_D^{-1}$), so that a higher bar represents stronger binding. Data were normalized to one for the highest affinity association constant for each protein. This representation makes apparent the stringency of binding selectivity for each AtAGO.

AGO1), all of which display the same 5′-nucleotide bias. Thus, the nucleotide specificity loop is likely working in an analogous manner in these systems to enforce a 5′-bias. Conversely, in AtAGOs there is considerable sequence variability in the region corresponding to the nucleotide specificity loop (Supplementary Figure 2), which led us to question whether AtAGOs determine nucleotide bias in the same manner as animal AGOs.

To investigate the structural basis for nucleotide specificity in AtAGOs, we determined crystal structures of their MID domains in the presence and absence of NMPs. The structures of the MID domains from AtAGO1, AtAGO2 and AtAGO5 resemble a Rossmann-like fold as seen previously in structures of this domain from other AGOs (Figure 2; Supplementary Figure 7; Supplementary Tables 1 and 2). The fold consists of a central parallel β-sheet straddled on either side by two α-helices. The importance of this fold for binding the 5′-end of small RNAs is reflected in the AtAGO MID domains and the hAGO2 MID domain having low pairwise RMSD values between 1.2 and 2.2 Å, despite sequence identities ranging from 26 to 52% (Supplementary Table 3). Each structure contains a sulphate ion from the solvent marking the position of the 5′-phosphate-binding site normally occupied by the 5′-end of a small RNA in a loaded AGO. Highly conserved residues Tyr, Lys, Gln, Lys (Tyr691, Lys695, Gln707, Lys732 in AtAGO1; Tyr678, Lys682, Gln694, Lys718 in AtAGO2; Tyr650, Lys654, Gln666, Lys691 in AtAGO5) contact this ion, which are all in very similar conformations to those previously seen in other AGO structures (Ma *et al*, 2005; Parker *et al*, 2005; Boland *et al*, 2010; Frank *et al*, 2010), reflecting a conserved mode of interaction with the 5′-phosphate of small RNAs across all AGO MID domains.

Adjacent to the 5′-phosphate-binding site is the region corresponding to the nucleotide specificity loop identified in hAGO2. In contrast to the conserved conformations of residues directly contacting the 5′-phosphate, the nucleotide specificity loops in AtAGO1, AtAGO2 and AtAGO5 all have distinct, rigid conformations (Figure 2D; Supplementary Table 4; Supplementary Figure 8), which is in accordance with the lack of conservation observed in their sequences (Supplementary Figure 2). 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) (Boland *et al*, 2010), demonstrates that this loop can assume variable conformations between species (Supplementary Figure 9). Fittingly, with the exception of QDE-2 for which no data are available on 5′-bias, each of these AGOs displays distinct 5′-nucleotide biases in their associated small RNAs. Thus, we suggest that the nucleotide specificity loop of AGO MID domains is an adjustable element whose conformational variability can be tuned to select which nucleotide(s) at the 5′-position are permitted and which are discriminated against, and in doing so, impacts on the evolutionary conservation of the corresponding small RNAs.

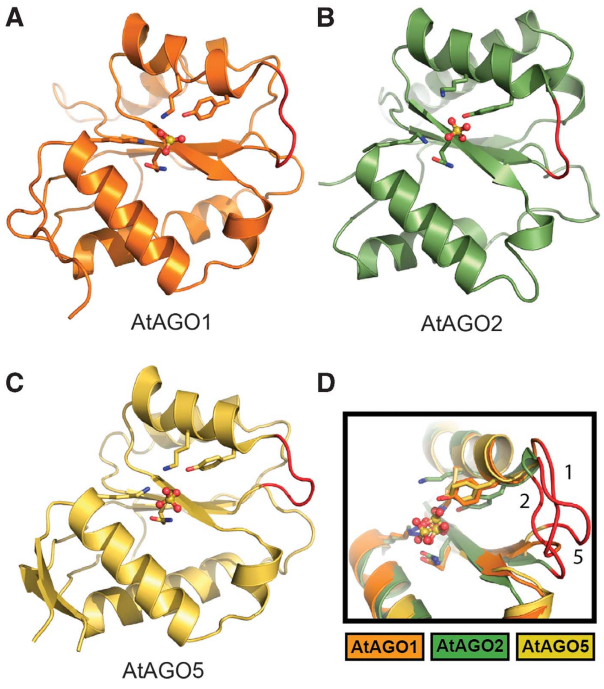


Figure 2 Crystal structures of the MID domains of AtAGO1 (**A**), AtAGO2 (**B**) and AtAGO5 (**C**). Structures are represented as ribbon diagrams and the sulphate ions bound to the 5′-phosphate-binding sites are shown as balls and sticks. Highly conserved phosphate interacting residues (Tyr, Lys, Gln, Lys) are shown as sticks. The loop region corresponding to the nucleotide specificity loop is coloured red. (**D**) Structural alignment of MID domains from AtAGO1, AtAGO2 and AtAGO5 highlighting the structural variability of the nucleotide specificity loops. Loops are coloured red and numbered according to the AtAGO from which they originate.

The identity of the 5′-nucleotide in AtAGO1 is probed by an Asn side chain

AtAGO1 is the founding member of the AGO family and is the predominant AGO of the *Arabidopsis* miRNA pathway. Small RNAs co-immunopurifying with AtAGO1 are mostly miRNAs containing 5′-U (Mi *et al*, 2008), and as our binding analysis demonstrated, the AtAGO1 MID domain preferentially binds U, but C is also permitted. To gain insight into the structural mechanism that enforces this 5′-bias, we determined crystal structures of the AtAGO1 MID domain in complex with NMPs (Figure 3).

The structure of the UMP complex reveals strong electron density for the entire nucleotide. As with hAGO2, the base of UMP stacks up against Tyr691 (Tyr529 in hAGO2) of the phosphate-binding pocket, which contributes non-specific recognition of the 5′-nucleotide, and extends towards the nucleotide specificity loop. However, the nucleotide specificity loop in AtAGO1 and the mode of base recognition are markedly different from that observed in hAGO2. In AtAGO1, the nucleotide specificity loop has a different sequence and is one residue longer than in hAGO2, resulting in a distinct conformation (Supplementary Figures 2 and 9). In hAGO2, the MID domain distinguishes 5′-nucleotide bases using atoms only in the peptide backbone of the nucleotide specificity loop. Strikingly, the specificity determining interactions in AtAGO1 invoke the use of an asparagine side chain, Asn687 located at the centre of the nucleotide specificity loop. Asn687 extends inwards from beneath the loop and makes hydrogen bonds with the O2 and N3 (via a water molecule) atoms of uracil. As a consequence of these distinct modes of recognition of UMP by AtAGO1 and hAGO2, the position of UMP beyond the phosphate group is quite different between the two proteins (Supplementary Figure 10). Beginning at the 3′-hydroxyl of UMP, there is an $\sim 30^{\circ}$ tilt between the two binding modes such that the uracil base in the AtAGO1 structure is much further away from the nucleotide specificity loop owing to the reach of the Asn687 side chain.

The manner by which Asn687 reaches out towards the base suggests that it is able to probe the identity of the 5′-nucleotide primarily through size, presumably allowing it to distinguish pyrimidines from purines. This notion is confirmed by the structures of the AtAGO1 MID domain in complex with the other NMPs. CMP also shows strong electron density for the entire nucleotide, and Asn687 as with UMP, makes hydrogen bonds with the CMP base. Interestingly, in hAGO2, although UMP is permitted to bind, the similarly sized CMP is not. This discrimination in hAGO2 is accomplished by hydrogen bonds from the peptide backbone of the nucleotide specificity loop with the C4 substituent of the pyrimidine ring, which is different in UMP and CMP (hydrogen bond acceptor O4 in UMP and hydrogen bond donor N4 in CMP). Thus, in hAGO2 UMP and CMP are distinguished by charge repulsion due to the alternating hydrogen bond forming ability of the C4 substituent. Conversely, in AtAGO1 the alternation of the pyrimidine C4 substituent is bypassed by instead making a side chain mediated hydrogen bond with the carbonyl group at position 2 of the pyrimidine ring, which is identical for both UMP and CMP.

In stark contrast to the two pyrimidines, both purines, AMP and GMP, do not have comparable electron density for the nucleotide sugar or the base in the binding pocket and only the phosphate is visible, presumably because the presence of the Asn687 side chain sterically prevents proper accommodation of a purine containing nucleotide. 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 Figure 11). In the observed *syn* conformations, both GMP and AMP maintain the hydrogen bond with Asn687 through their N7 atoms. However, in solution, but GMP and AMP have an energetic preference for the *anti* conformation about the glycosidic bond, which tends to maximize the distance between the base and the ribose sugar (Lee *et al*, 1975). Thus, the presence of Asn687 appears to force the purine nucleotides into the energetically less stable *syn* conformation, which concomitantly reduces their binding affinity for the protein.

A potential mechanism of 5′-nucleotide selection in AtAGO2

Structure determination of either AtAGO2 or AtAGO5 in complex with nucleotides failed. For AtAGO2, as with AtAGO1, there are no crystal contacts that impede nucleotide binding (Supplementary Figure 12). However, the high concentration of ammonium sulphate in the AtAGO2 crystallization buffer places a sulphate ion in the binding that presumably prevents NMPs from binding. Thus, we resorted to manually docking nucleotides into the binding site to propose a possible mechanism of selection. AtAGO2 is the

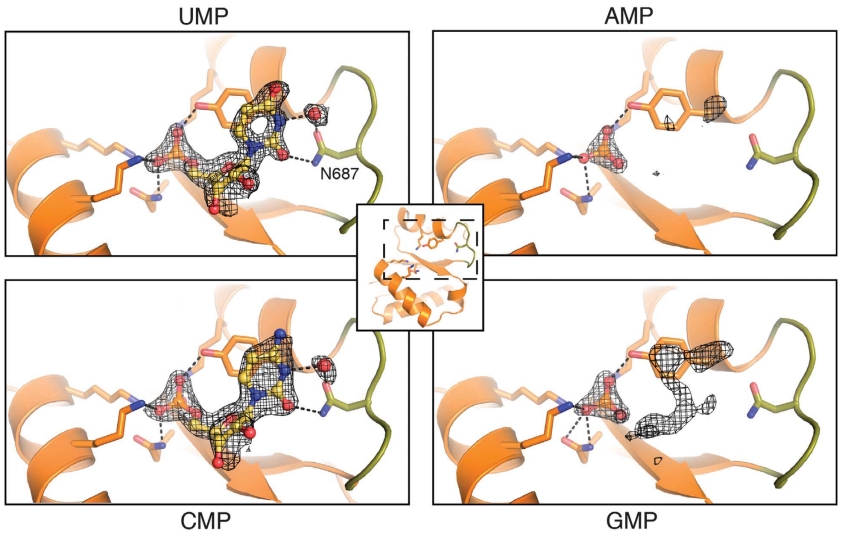


Figure 3 AtAGO1 MID domain in complex with NMPs. UMP and CMP are modelled and shown in ball and stick representation. The nucleotide specificity loop is coloured green. Only the phosphate groups of GMP and AMP were modelled (not shown). Protein side chains interacting with the 5′-phosphate or the base portion of the NMPs are shown as sticks with carbon, oxygen and nitrogen atoms coloured orange, red and blue, respectively. A water molecule that indirectly hydrogen bonds N687 to the NMP is shown as a red sphere. Difference electron density contoured at 2.5σ is shown before inclusion of any nucleotide in the model. Dotted black lines indicate hydrogen bonds.

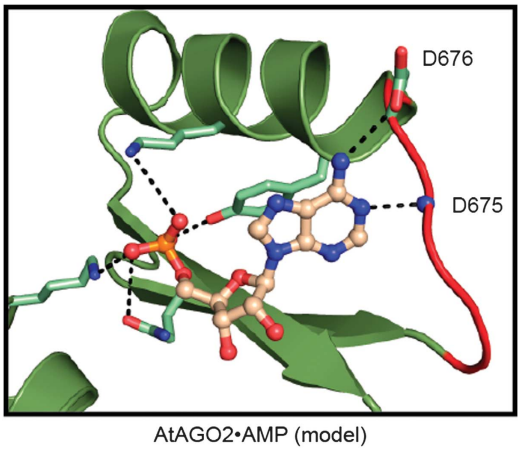


Figure 4 Modelling of AMP into the structure of the AtAGO2 MID domain. The MID domain is shown in ribbon representation coloured green with the nucleotide specificity loop highlighted in red. AMP is shown in ball and stick representation. Shown as sticks, are conserved phosphate-binding residues as well as Asp676, which could potentially hydrogen bond with AMP. The amide backbone nitrogen atom of Asp675, which could also hydrogen bond with AMP, is represented as a sphere.

most stringent selector, interacting only with small RNAs that begin with A and discriminating against all other nucleotides. The path of the nucleotide specificity loop in AtAGO2 somewhat resembles that of hAGO2 with regards to the orientation of the backbone peptide units and most of the side chains (Supplementary Figure 13). But, unlike hAGO2, AtAGO2 must somehow permit binding of A, yet prevent binding of U, as well as G and C.

AMP was docked into the nucleotide binding site of AtAGO2 using the sulphate group from the crystal as a guide to position the phosphate, and superposition with the hAGO2:AMP structure (Figure 4). One difference immediately apparent upon docking is that Tyr678 in AtAGO2 does not stack up as closely against the base as was found in all other structures of the MID domain with bound nucleotides. This reduction in stacking may explain why the binding affinities for NMPs to AtAGO2 determined above are weaker than others. The model also reveals that AMP can potentially interact (via its N1 atom) with the backbone of the nucleotide specificity loop (peptide amino group of Asp675) in a manner similar to hAGO2. However, because of a variation in the nucleotide specificity loop that replaces Pro527 of hAGO2 with an aspartate in AtAGO2, the side chain of Asp676 can hydrogen bond with the amino group of AMP. This interaction would clash with GMP since it contains a carbonyl group at this position.

We also attempted to model nucleotides into the binding site of AtAGO5, but could not come up with a consensus as to how selectivity for C is achieved based on the observed conformation of the nucleotide specificity loop. One possibility is that for AtAGO5, the C-terminus from a symmetry-related molecule directly blocks binding of NMPs and the conformation of the loop may be slightly distorted due to its extensive involvement in crystal contacts (Supplementary Figure 14). Another possibility is that because AtAGO5 is the least stringent at 5′-nucleotide discrimination, the mechanism underlying selection is likely subtler and hence not easily rationalized by simple inspection of the binding site.

Mutations in the nucleotide specificity loop alter NMP selectivity

To assess the functional importance of the nucleotide specificity loop in AtAGOs for conferring 5′-bias, we carried out a mutational analysis. Since the crystallographic analysis of AtAGO1 complexes with UMPs indicates that AtAGO1 makes direct interactions with the base of UMP or CMP via the side chain of Asn687, we mutated this residue to glutamine or alanine. By mutating to glutamine and extending the length of the side chain at this position we expect to sterically hinder binding of both U and C. This is supported by NMR binding

Table I Dissociation constants of AtAGO MID domain mutants with nucleoside monophosphates

	AMP (mM)	*CMP (mM)*	*GMP (mM)*	*UMP (mM)*
AtAGO1 (N687Q)	4.02 ± 0.53	10.54 ± 0.91	4.52 ± 0.66	3.93 ± 0.71
AtAGO1 (N687A)	3.17 ± 0.30	11.58 ± 2.39	3.34 ± 0.64	1.49 ± 0.25
AtAGO2 (D676A)	3.45 ± 0.21	67.71 ± 10.89	9.98 ± 3.19	22.65 ± 3.46
AtAGO5 (loop insertion)	1.58 ± 0.40	4.25 ± 0.13	2.95 ± 0.37	3.08 ± 0.16

Corresponding NMR titration spectra and curve fitting procedures can be found in Supplementary Figures 15–18.

analysis of NMPs, which indicates that the affinity for pyrimidine NMPs is reduced 3- to 4-fold, whereas the affinity for the purines remains essentially unchanged (Table I; Supplementary Figure 15). On the other hand, by mutating N687 to alanine and reducing the length of the side chain we expect to remove sterical hindrance for AMP and GMP binding in the *anti* conformation. Dissociation constants for the N687A mutant changed accordingly with affinities for both purines increasing by ~2-fold and affinities for pyrimidines decreasing (Table I; Supplementary Figure 16). Thus, AtAGO1 selects the small RNA 5′-nucleotide based on base size via the side chain of Asn687.

For AtAGO2, modelling of AMP into the 5′-nucleotide binding site implicated Asp676 as being important for distinguishing AMP. The smaller pyrimidines are not expected to be affected by D676. Mutation of D676 to alanine accordingly increased the affinity for GMP, while binding of the other NMPs remains virtually identical to wild-type levels (Table I; Supplementary Figure 17).

In AtAGO5, for which modelling did not provide a clear rationale for mutation, we resorted to introducing an additional residue into the loop to distort its conformation and change the weak bias of wild-type AtAGO5. Introducing a glycine between V646 and P647 (resulting in the loop sequence…DVGPGS… instead of…DVPGS…) in AtAGO5 changed its nucleotide selectivity (Table I; Supplementary Figure 18). Whereas wild-type AtAGO5 binds strongest to CMP the loop insertion mutant displays the lowest affinity for CMP and instead interacts strongest with AMP.

These results demonstrate that the nucleotide specificity loop in the MID domains of AGO proteins is responsible for providing selectivity for the interaction of these proteins with the 5′-end of small RNAs.

Discussion

Endogenous small silencing RNAs play critical roles in the regulation of an organism's cellular functions and therefore it is no surprise that substantial quality control mechanisms have evolved for their proper biogenesis. The critical importance of the 5′-nucleotide of small RNAs in dictating proper RISC formation, sorting and hence accurate target recognition is now apparent from several studies (Mi *et al*, 2008; Takeda *et al*, 2008; Czech *et al*, 2009; Frank *et al*, 2010; Ghildiyal *et al*, 2010; Seitz *et al*, 2011). Small RNA sorting is particularly relevant in the RNA silencing system of *Arabidopsis,* where many different classes of small RNAs must be sorted among multiple AGOs. 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 (Frank *et al*, 2010). Here, an analogous structural analysis of several AtAGOs in complex with NMPs confirms that the nucleotide specificity loop contributes to proper sorting of small RNAs in plants. This study also emphasizes the functional diversity of the nucleotide specificity loop, which despite consisting of only six or seven residues, is able to coordinate the accommodation of different 5′-nucleotides depending on the particular AGO. In the case of hAGO2, charge repulsion and steric hindrance by backbone atoms in the nucleotide specificity loop are at play to restrict human miRNAs to begin for the most part with a 5′-U. In *Arabidopsis*, owing to the greater diversity of small RNA classes that must be sorted, side chains in the nucleotide specificity loop of AtAGOs play a prominent role. In particular, AtAGO1 utilizes an asparagine residue to sense the size of the nucleotide base, and an aspartate residue appears to be important for distinguishing 5′-A in AtAGO2. Nucleotide bias in AtAGO5 is also affected by the nucleotide specificity loop as extending its length by one residue modulates NMP selection.

Based on the aggregate evidence it is clear that the identity of a small RNA's 5′-nucleotide is sensed by the AGO MID domain. However, the question still remains as to which aspect of RISC function the 5′-nucleotide affects. That is, does it affect small RNA duplex loading into AGO from Dicer, passenger strand removal after loading, or the AGO slicer activity after target recognition? A recent study by Kawamata *et al* (2011) has shed light on these questions by analysing *in-vitro* RISC assembly in *Drosophila* embryo lysates. They found that the identity of the 5′-nucleotide is monitored by the RISC both during loading of the duplex and unwinding of the passenger strand. Proper loading of AGOs depends on the presence of sufficient stabilizing interactions between the small RNA duplex and the protein. This is especially imperative in miRNA loading because of the mismatches often present between the guide and passenger strands, which would work to destabilize their loading. Thus, for miRNAs, the identity of the 5′-nucleotide is critical since the correct base provides an additional anchoring interaction that would contribute to RISC stability. An anchoring role for the 5′-nucleotide would also be important after unwinding of the duplex to form the mature RISC where any stabilizing interactions from the passenger strand are lost. Lastly, it has also been shown that the presence of an incorrect 5′-nucleotide anchor can also perturb the slicing activity of AGOs (Felice *et al*, 2009).

The 5′-nucleotide bias has also been observed in piRNAs, which form complexes with PIWI-like AGOs (Brennecke *et al*, 2007; Kawaoka *et al*, 2011). piRNAs are thought to be generated from a single-stranded precursor in a Dicer-independent manner (Thomson and Lin, 2009). Thus, loading of piRNAs will likely occur via a different mechanism. Nonetheless, it was found that only single-stranded RNAs containing a 5′-U are efficiently incorporated into PIWIs (Kawaoka *et al*, 2011). The sequence of the nucleotide specificity loop of PIWIs is

different from both hAGO2 and AtAGO1, suggesting the 5′-U may be recognized in this system by yet another mechanism. Finally, the worm-specific group-3 AGOs (WAGOs) present a potentially new paradigm in 5′-nucleotide recognition (Faehnle and Joshua-Tor, 2007). These *C. elegans* AGOs associate with secondary siRNAs, which contain di- and triphosphates at their 5′-ends. The phosphate-interacting residues (Y, K, Q, K) are not well conserved in WAGOs, particularly the tyrosine that stacks against the nucleotide base. Understanding the full repertoire of 5′-nucleotide recognition modes that AGOs employ to faithfully regulate small RNA silencing awaits structural analyses from these systems.

Materials and methods

Protein expression and purification

The MID domains of AtAGO1 (residues 595–740), AtAGO2 (residues 579–721) and AtAGO5 (residues 562–697) were cloned into the *Bam*HI and *Not*I sites of a pSMT3 vector (Mossessova and Lima, 2000), which contains an N-terminal Ulp1 cleavable His_6-Sumo tag. The proteins were bacterially expressed using standard protocols and purified by Ni-affinity chromatography in buffer containing 25 mM Tris pH 8.0, 500 mM NaCl and 10 mM imidazole. This was followed by cleavage of the tag using Ulp1, cation-exchange chromatography, and Superdex-75 size-exclusion chromatography in 25 mM MES pH 6.5, 200 mM NaCl, and 3 mM DTT. Site-directed mutagenesis was performed using the Quickchange kit (Stratagene).

Crystallization, data collection and structure determination

Crystals of native and SeMet-substituted MID domains of AtAGO1 and AtAGO2 and crystals of native MID domain of AtAGO5 were grown by hanging-drop vapour diffusion at 4°C. Crystals of both, AtAGO1 or AtAGO2 MID domains, were grown by mixing protein at 10–15 mg ml^{-1} 1:1 with a well solution containing 0.1 M citric acid and 2 M ammonium sulphate. Crystals of the MID domain of AtAGO5 were grown by mixing protein at 10–15 mg ml^{-1} 1:1 with a well solution containing 0.2 M ammonium sulphate, 0.1 M sodium citrate pH 5.4 and 30% PEG 4000. The MID domains of AtAGO1 and AtAGO2 crystallized in the same condition, which contained a high concentration of sulphate (2 M ammonium sulphate). Sulphate 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 sulphate (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 sulphate and, therefore, soaking was not successful.

Diffraction data for flash-cooled SeMet-substituted crystals (cryoprotected with 15% glycerol) were collected at the Canadian Light Source beamline 08B1-1 at 0.9792 Å and processed using *HKL2000* (Otwinowski and Minor, 1997; Supplementary Tables 1 and 2). Diffraction data for crystals of the AtAGO5 MID domain and nucleotide containing crystals of AtAGO1 were collected on a Rigaku rotating copper-anode generator at 1.5418 Å.

The structures of the MID domains from AtAGO1 and AtAGO2 were solved by single-wavelength anomalous diffraction (SAD) using the *Phenix* program (Adams *et al*, 2002) and refined with *CNS* (Brünger *et al*, 1998) and *Phenix* (Adams *et al*, 2002). The nucleotide complexes of the AtAGO1 MID domain were solved by difference Fourier analysis using the native structure. The structure of the AtAGO5 MID domain was solved by molecular replacement using the program *Phaser* (McCoy *et al*, 2007) and the AtAGO1 MID domain used as a search model.

Modelling

AMP was modelled in the nucleotide binding site of the AtAGO2 MID domain by first superimposing the empty MID domain onto the structure of hAGO2:AMP. From this, the position of AMP from the hAGO2 complex was manually adjusted (~0.8 Å in one direction) until interactions with the nucleotide specificity loop of AtAGO2 were appropriate. Following this, torsion angles connecting the phosphate to the nucleotide were adjusted to match the position of the phosphate onto the experimentally observed sulphate group. UMP was modelled similarly.

NMR titration experiments

All NMR experiments were performed at 293 K using a Bruker 600 MHz spectrometer. NMR titrations were carried out by acquiring ^{1}H-^{15}N HSQC spectra on samples of 0.10–0.25 mM ^{15}N-labelled MID domain with addition of increasing amounts of unlabelled ligand. Chemical shift differences were calculated as $\Delta\delta = [(\Delta\delta_H)^2 + (0.2 \times \Delta\delta_N)^2]^{1/2}$, where $\Delta\delta_H$ and $\Delta\delta_N$ are the observed chemical shift changes for ^{1}H and ^{15}N, respectively. For determination of dissociation constants, $\Delta\delta$ was plotted as a function of the molar ratio (nucleotide/protein) and the data for multiple peaks were fit using the maximum shift and dissociation constant as adjustable parameters.

Accession codes

Coordinates and structure factors have been submitted in the Protein Data Bank under accession codes 4G0M, 4G0O, 4G0P, 4G0Q, 4G0X, 4G0Y and 4G0Z.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Shaun Labiuk and Pawel Grochulski for X-ray data collection performed on beamline 08B1-1 at the Canadian Light Source, which is supported by the Natural Sciences and Engineering Research Council of Canada, the National Research Council Canada, the Canadian Institutes of Health Research, the Province of Saskatchewan, Western Economic Diversification Canada, and the University of Saskatchewan. Dr David Baulcombe for providing plasmids encoding AtAGO1, AtAGO2 and AtAGO5; R Szittner and K Illes for technical support; and Marc Fabian for critical reading of the manuscript. NMR data were collected at the QANUC NMR facility. BN is supported by a Canada Research Chair, a Career Development Award from the Human Frontiers Science Program (CDA 0018/2006-C/1) and an operating grant from the Canadian Institutes of Health Research. NS is funded by a CIHR grant. FF is supported by a Boehringer Ingelheim Fonds PhD Fellowship.

Author contributions: BN, NS and FF designed the project. BN and FF wrote the manuscript. FF performed all the crystallographic work and FF and JH performed NMR titration experiments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr* **58:** 1948–1954

Boland A, Tritschler F, Heimstädt S, Izaurralde E, Weichenrieder O (2010) Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. *EMBO Rep* **11:** 522–527

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ (2007) Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128:** 1089–1103

Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS,

Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54:** 905–921

Czech B, Hannon GJ (2011) Small RNA sorting: matchmaking for Argonautes. *Nat Rev Genet* **12:** 19–31

Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ (2009) Hierarchical rules for Argonaute loading in Drosophila. *Mol Cell* **36:** 445–456

Faehnle CR, Joshua-Tor L (2007) Argonautes confront new small RNAs. *Curr Opin Chem Biol* **11:** 569–577

Felice KM, Salzman DW, Shubert-Coleman J, Jensen KP, Furneaux HM (2009) The 5′ terminal uracil of let-7a is critical for the recruitment of mRNA to Argonaute2. *Biochem J* **422:** 329–341

Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* **465:** 818–822

Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler ELW, Zapp ML, Weng Z, Zamore PD (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science* **320:** 1077–1081

Ghildiyal M, Xu J, Seitz H, Weng Z, Zamore PD (2010) Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. *RNA* **16:** 43–56

Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. *Nat Rev Genet* **10:** 94–108

Gu W, Shirayama M, Conte D, Vasale J, Batista PJ, Claycomb JM, Moresco JJ, Youngman EM, Keys J, Stoltz MJ, Chen C-CG, Chaves DA, Duan S, Kasschau KD, Fahlgren N, Yates JR, Mitani S, Carrington JC, Mello CC (2009) Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the C. elegans germline. *Mol Cell* **36:** 231–244

Kawamata T, Seitz H, Tomari Y (2009) Structural determinants of miRNAs for RISC loading and slicer-independent unwinding. *Nat Struct Mol Biol* **16:** 953–960

Kawamata T, Yoda M, Tomari Y (2011) Multilayer checkpoints for microRNA authenticity during RISC assembly. *EMBO Rep* **12:** 944–949

Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H (2008) Drosophila endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* **453:** 793–797

Kawaoka S, Izumi N, Katsuma S, Tomari Y (2011) 3′ End formation of PIWI-interacting RNAs *in vitro*. *Mol Cell* **43:** 1015–1022

Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115:** 209–216

Kwak PB, Tomari Y (2012) The N domain of Argonaute drives duplex unwinding during RISC assembly. *Nat Struct Mol Biol* **19:** 145–151

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294:** 858–862

Lau NC, Robine N, Martin R, Chung W-J, Niki Y, Berezikov E, Lai EC (2009) Abundant primary piRNAs, endo-siRNAs, and microRNAs in a Drosophila ovary cell line. *Genome Res* **19:** 1776–1785

Lee CH, Evans FE, Sarma RH (1975) Interrelation between glycosidic torsion, sugar pucker, and backbone conformation in 5′-beta-nucleotides. A 1H and 31P fast Fourier transform nuclear magnetic resonance investigation of the conformation of 8-aza-5′-beta-adenosine monophosphate and 8-aza-5′-beta-guanosine monophosphate. *J Biol Chem* **250:** 1290–1296

Lee H-C, Li L, Gu W, Xue Z, Crosthwaite SK, Pertsemlidis A, Lewis ZA, Freitag M, Selker EU, Mello CC, Liu Y (2010) Diverse pathways generate microRNA-like RNAs and Dicer-independent small interfering RNAs in fungi. *Mol Cell* **38:** 803–814

Lewis BP, Shih I-H, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. *Cell* **115:** 787–798

Lima WF, Wu H, Nichols JG, Sun H, Murray HM, Crooke ST (2009) Binding and cleavage specificities of human Argonaute2. *J Biol Chem* **284:** 26017–26028

Lingel A, Simon B, Izaurralde E, Sattler M (2003) Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. *Nature* **426:** 465–469

Lingel A, Simon B, Izaurralde E, Sattler M (2004) Nucleic acid 3′-end recognition by the Argonaute2 PAZ domain. *Nat Struct Mol Biol* **11:** 576–577

Ma J-B, Ye K, Patel DJ (2004) Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. *Nature* **429:** 318–322

Ma J-B, Yuan Y-R, Meister G, Pei Y, Tuschl T, Patel DJ (2005) Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* **434:** 666–670

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40:** 658–674

Mi S, Cai T, Hu Y, Chen Y, Hodges E, Ni F, Wu L, Li S, Zhou H, Long C, Chen S, Hannon GJ, Qi Y (2008) Sorting of small RNAs into Arabidopsis Argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* **133:** 116–127

Montgomery TA, Howell MD, Cuperus JT, Li D, Hansen JE, Alexander AL, Chapman EJ, Fahlgren N, Allen E, Carrington JC (2008) Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. *Cell* **133:** 128–141

Mossessova E, Lima CD (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol Cell* **5:** 865–876

Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology, Volume 276: Macromolecular Crystallography,* Carter Jr CW, Sweet RM (eds), Part A, pp 307–326. New York: Academic Press

Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* **434:** 663–666

Ruby JG, Jan C, Player C, Axtell MJ, Lee W, Nusbaum C, Ge H, Bartel DP (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in C. elegans. *Cell* **127:** 1193–1207

Schwarz DS, Hutvágner G, Du T, Xu Z, Aronin N, Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115:** 199–208

Seitz H, Tushir JS, Zamore PD (2011) A 5′-uridine amplifies miRNA/miRNA* asymmetry in Drosophila by promoting RNA-induced silencing complex formation. *Silence* **2:** 4

Song J-J, Liu J, Tolia NH, Schneiderman J, Smith SK, Martienssen RA, Hannon GJ, Joshua-Tor L (2003) The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes. *Nat Struct Biol* **10:** 1026–1032

Takeda A, Iwasaki S, Watanabe T, Utsumi M, Watanabe Y (2008) The mechanism selecting the guide strand from small RNA duplexes is different among argonaute proteins. *Plant Cell Physiol* **49:** 493–500

Tan G, Garchow B, Liu X, Yeung J, Morris J, Cuellar T, McManus MT, Kiriakidou M (2009) Expanded RNA-binding activities of mammalian Argonaute 2. *Nucl Acids Res* **37:** 7533–7545

Thomson T, Lin H (2009) The biogenesis and function of PIWI proteins and piRNAs: progress and prospect. *Annu Rev Cell Dev Biol* **25:** 355–376

Wang Y, Juranek S, Li H, Sheng G, Wardle GS, Tuschl T, Patel DJ (2009) Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. *Nature* **461:** 754–761

Yuan Y-R, Pei Y, Ma J-B, Kuryavyi V, Zhadina M, Meister G, Chen H-Y, Dauter Z, Tuschl T, Patel DJ (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell* **19:** 405–419