Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute

Tomohiro Miyoshi1, Kosuke Ito2, Ryo Murakami2 & Toshio Uchiumi2

Argonaute proteins are key players in the gene silencing mechanisms mediated by small nucleic acids in all domains of life from bacteria to eukaryotes. However, little is known about the Argonaute protein that recognizes guide RNA/target DNA. Here, we determine the 2 Å crystal structure of *Rhodobacter sphaeroides* Argonaute (RsAgo) in a complex with 18-nucleotide guide RNA and its complementary target DNA. The heteroduplex maintains Watson–Crick base-pairing even in the 3'-region of the guide RNA between the N-terminal and PIWI domains, suggesting a recognition mode by RsAgo for stable interaction with the target strand. In addition, the MID/PIWI interface of RsAgo has a system that specifically recognizes the 5` base-U of the guide RNA, and the duplex-recognition loop of the PAZ domain is important for the DNA silencing activity. Furthermore, we show that Argonaute discriminates the nucleic acid type (RNA/DNA) by recognition of the duplex structure of the seed region.

1 Center for Transdisciplinary Research, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan. 2 Department of Biology, Faculty of Science, Niigata University, 8050 Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan. Correspondence and requests for materials should be addressed to T.M. (email: miyoshi@bio.sc.niigata-u.ac.jp) or K.I. (email: k-ito@bio.sc.niigata-u.ac.jp).
Argonaute-family proteins have crucial roles in gene regulation in collaboration with small guide strands, which recognize complementary sequences in the target. This recognition occurs mainly through the seed region, a 2–8 nucleotide segment of the guide strands. The crystal structures of full-length Argonaute proteins have been investigated to help understand the mechanism of their action; the first structures were reported in the prokaryotes Pyrococcus furiosus and Aquifex aeolicus. Argonaute proteins belong to the PIWI superfamily, and contain a PIWI (P element-induced wimpy testis) domain, a MID (middle) domain, a PAZ (Piwi–Argonaute–Zwille) domain and an N-terminal domain. The proteins form a bilobal scaffold composed of MID–PIWI and N–PAZ lobes connected with linkers. This bilobal structure binds to guide and target molecules along a nucleic acid-binding channel. The PIWI domain of Argonaute is a structural homologue of the RNase H catalytic domain, and contributes to endonuclease (slicing) activity for the target strands. However, the slicing activity is not a property of all Argonaute proteins, and interacts directly with the 5′-end of the guide strand by recognition of the phosphorylated first nucleotide and base features. The PAZ domain contains a hydrophobic pocket of an oligosaccharide-binding-like fold and is involved in binding of a single oligonucleotide; the PAZ domain associates with the 3′-end of the guide strand. This domain is required to unwind the small RNA duplex during RNA-induced silencing complex formation in non-cleavage Argonautes and to regulate slicer activity. The N-terminal domain of Argonaute is the least well characterized of the four domains. Crystallographic analysis of the Argonaute complex in Thermus thermophilus showed that the N-terminal domain prevents extension of base pairing of the guide and target strands at the 3′-region of the guide strand. In addition, biochemical studies of human Argonaute (HsAgo) showed that the N-terminal domain facilitates the unwinding of the guide/target duplex during RNA-induced silencing complex assembly.

Argonaute proteins are evolutionarily diverse. In eukaryotes, Argonaute proteins are key players in the small ‘guide RNA’-mediated silencing systems for ‘target RNA’, such as RNA interference, microRNA gene silencing and piRNA transposon silencing. In contrast, prokaryotic Argonaute proteins from T. thermophilus and P. furiosus associate with DNAs as guide and target strands, and interfere with the propagation of foreign DNA. This latter system is also called DNA-guided DNA and target strands, and interfere with the propagation of foreign DNA. To address this question, we first performed an electrophoretic mobility shift assay using 18-base single strand RNA and DNA. In this assay, 5′-phosphorylated (5′P) strands were used because 5′-phosphorylation is known to be required for the recognition of the guide strand by Argonaute proteins. Our assay showed that the binding ability of RsAgo to the 18-base single strand RNA was higher than to the 18-base single strand DNA (Fig. 1a,b and Supplementary Table 1). Measurement of the equilibrium dissociation constant (Kd) values by a fluorescence polarization assay indicated that the 5′P 18-base RNA bound 46-fold more tightly to RsAgo than to the 5′P 18-base single strand DNA (Kd: 5′P-RNA, 0.91 ± 0.06 nM; 5′P-DNA, 42.6 ± 3.4 nM) (Table 1 and Supplementary Fig. 1a,b). This result is consistent with those of the electrophoretic mobility shift assay (Fig. 1a,b).

To examine the specificity of RsAgo for the nucleic acid duplexes, we carried out similar electrophoretic mobility shift assays using four types of complementary 18-base double strands, namely, 5′P-RNA/RNA, 5′P-RNA/DNA, 5′P-DNA/RNA, 5′P-DNA/DNA. Formation of an RsAgo–duplex complex and the significant disappearance of free strands were unambiguously observed only when 5′P-RNA and non-phosphorylated DNA were used (Fig. 1c–f and Supplementary Table 1). Considering that 5′-phosphorylation is necessary for the recognition of the guide strand, these results suggested that RsAgo specifically recognized RNA as the guide strand and DNA as the target strand. To confirm the specificity of RsAgo for the guide and target strands, we carried out the electrophoretic mobility shift assays by sequentially adding the guide and target strands to
containing a 5′P 18-base guide RNA strand and an 18-base complementary DNA target strand, and solved the structure at 2 Å resolution using the single-wavelength anomalous dispersion method with SeMet-labelled Rd proteins (Fig. 2a and Table 2). RsAgo was composed of N-terminal, PAZ, MID and PIWI domains, and L1 and L2 linkers (Fig. 2a). Overall, the structure of RsAgo was similar to other reported Argonaute proteins39. The fully solved heteroduplex is bound within a positively charged channel between N–PAZ-containing and MID–PIWI-containing lobes by an extensive hydrogen-bonding network (Fig. 2b,c and Supplementary Fig. 3). The nucleotide at the 5′-end of the guide RNA strand and the nucleotide at the 3′-end of the target DNA strand were unpaired in RsAgo (Fig. 2b,c), as in previously determined Argonaute complexes8,9,15,16. Surprisingly, the 3′-region of the guide RNA strand was base-paired with the 5′-region of the target DNA strand; this behaviour has not been reported previously for other Argonaute complexes8,9,16. We discuss the implications of this structure in more detail below in conjunction with the results of biochemical analyses.

### MID/PIWI interface for guide RNA 5′-end recognition.

In the MID domain, the 5′-phosphate of the guide RNA strand forms hydrogen bonds with the side chains of Y463, K467, K506 and the main chains of Q479 (Fig. 3a). These amino acid residues, with the exception of Q479, are largely conserved in RNA-guided eukaryote Argonaute proteins as well as DNA-guided prokaryote Argonaute proteins (Supplementary Fig. 4)36, suggesting that the 5′-phosphate recognition mode is conserved irrespective of the nucleic acid type of the guide strand. In addition, the MID domain interacted with the 2′-hydroxy group of the nucleotide 1 and the phosphate group of nucleotide 2 of the guide RNA strand via R481 and T484 (Fig. 3a). Hydrogen bonds to the 2′-hydroxy group of the guide strand 5′-terminal nucleotide have also been observed in RNA-guided eukaryotic Argonaute proteins11,13,14, suggesting that this bond is important for the recognition of RNA as a guide strand.

To examine the contributions of the 5′-phosphate of the guide strand to the interaction with the MID domain, we measured the Kd values using a fluorescence polarization system. The elimination of the 5′-phosphate of the guide strand RNA severely impaired the binding activity to RsAgo, as compared to the 5′P-RNA strand (Table 1 and Supplementary Fig. 1a,c). This result indicates that the 5′-phosphate of the guide strand RNA is an essential component for guide strand recognition by RsAgo.

In addition to the interaction described above, the MID domain of RsAgo forms base-specific interactions with the 3′-hydroxy group at the 3′-end of the target DNA strand. These interactions are mediated by the PAZ domain (Supplementary Fig. 3), which is a conserved domain in Argonaute family proteins.

### Overall structure of the RsAgo–RNA/DNA ternary complex.

On the basis of the results of the nucleic acid recognition assay (Fig. 1), we crystallized RsAgo complexed with a hybrid duplex containing a 5′P 18-base guide RNA strand and an 18-base complementary DNA target strand. In the structure, the nucleotide at the 5′-end of the guide RNA strand and the nucleotide at the 3′-end of the target DNA strand were unpaired, and the 5′P-RNA and 3′DNA ends were base-paired at the 5′- and 3′-end positions, respectively. The structure of the nucleic acid type of the guide strand. In addition, the MID domain interacted with the 2′-hydroxy group of the nucleotide 1 and the phosphate group of nucleotide 2 of the guide RNA strand via R481 and T484 (Fig. 3a). Hydrogen bonds to the 2′-hydroxy group of the guide strand 5′-terminal nucleotide have also been observed in RNA-guided eukaryotic Argonaute proteins11,13,14, suggesting that this bond is important for the recognition of RNA as a guide strand.

To examine the contributions of the 5′-phosphate of the guide strand to the interaction with the MID domain, we measured the Kd values using a fluorescence polarization system. The elimination of the 5′-phosphate of the guide strand RNA severely impaired the binding activity to RsAgo, as compared to the 5′P-RNA strand (Table 1 and Supplementary Fig. 1a,c). This result indicates that the 5′-phosphate of the guide strand RNA is an essential component for guide strand recognition by RsAgo.

In addition to the interaction described above, the MID domain of RsAgo forms base-specific interactions with the 3′-hydroxy group at the 3′-end of the target DNA strand. These interactions are mediated by the PAZ domain (Supplementary Fig. 3), which is a conserved domain in Argonaute family proteins.

### Table 1 | Equilibrium dissociation constant (Kd) values of RsAgo for 18-base single strand RNA and DNA

| RsAgo     | Nucleic acid       | Kd (nM) |
|-----------|--------------------|---------|
| WT        | 5′P-RNA (5′U-RNA)  | 0.91±0.06|
| WT        | 5′P-DNA            | 42.6±3.4 |
| WT        | 5′OH-RNA           | 1,144±54 |
| WT        | 5′P-RNA (with EDTA)| 7.71±0.36|
| WT        | 5′A-RNA            | 24.8±1.7 |
| WT        | 5′C-RNA            | 67.0±3.5 |
| WT        | 5′G-RNA            | 181.3±13.3|
| Y463A/K467A | 5′P-RNA         | 1,697±123 |
| R481A/T484A | 5′P-RNA         | 718±36   |
| K506A     | 5′P-RNA            | 1,034±47 |
| ΔL777     | 5′P-RNA            | 38.7±2.1 |

The Kd values were measured by a fluorescence polarization assay using nucleic acids modified with 6-carboxyfluorescein (6-FAM) at the 3′-end position. The nucleic acid sequences are shown in Supplementary Table 1.

---

RsAgo (Supplementary Fig. 2). The results showed that the DNA strand only bound to RsAgo that had been pre-incubated with 5′P-RNA. This observation further supports our interpretation that RsAgo specifically recognizes RNA as the guide strand and DNA as the target strand.

**Figure 1 | Nucleic acid types of 18-base guide and target strands bound to RsAgo.** (a,b) Binding affinities of 5′P-RNA (5′P-RNA) (a) and DNA (5′P-DNA) (b) to the WT recombinant RsAgo protein were examined using an electrophoresis mobility shift assay. The oligonucleotide sequences of the 32P-labelled guide strands are shown in Supplementary Table 1. Single strand RNA and DNA (5 mM each) were incubated with various concentrations (0, 1, 10, 100 and 1000 nM) of WT RsAgo. (c–f) Binding affinities of 5′P-RNA/non-phosphorylated RNA (5′P-RNA/RNA) (c), 5′P-RNA/non-phosphorylated DNA (5′P-RNA/DNA) (d), 5′P-DNA/non-phosphorylated RNA (5′P-DNA/RNA) (e), 5′P-DNA/non-phosphorylated DNA (5′P-DNA/DNA) (f) duplexes to the WT recombinant RsAgo protein were examined using an electrophoresis mobility shift assay. The oligonucleotide sequences of the 32P-labelled guide and target strands are shown in Supplementary Table 1. The duplexes were incubated in a same way as a,b. The small RNA and DNA strands are shown as red and blue lines, respectively.
5′-terminal uracil of the guide RNA strand. The aromatic residue Y463 stacked against the 5′-terminal uracil and the main chain of A454 formed a hydrogen bond with N3 of the uracil (Fig. 3a). Y463 is widely conserved across Argonaute proteins (Supplementary Fig. 4) and this conserved tyrosine has also been shown to stack against the 5′-terminal base of the guide strand in other Argonaute proteins\(^6,36,38,41,42\). However, other Argonaute proteins do not possess a residue corresponding to A454. These proteins contain a rigid loop corresponding to that containing A454; this rigid loop has been demonstrated to function as a critical determinant for the 5′-nucleotide preference of the guide strand\(^20,21\). In addition to A454, the side chain of R754 of the PIWI domain also formed hydrogen bonds with O4 of the 5′-terminal uracil of the guide RNA strand (Fig. 3a). A similar hydrogen-bonding pattern with A454 and R754 was also possible if the 5′-terminal base was guanine. However, if the guanine formed hydrogen bonds with A454 and R754 like uracil, then the 5′-phosphate could no longer form the proper interaction with RsAgo (Fig. 3b). Indeed, the measurement of Kd values using a fluorescence polarization system showed that the substitution of the 5′-nucleotide with guanosine severely impaired the binding activity to RsAgo compared to 5′U-RNA (Kd: 5′U-RNA, 0.91 ± 0.06 nM; 5′G-RNA, 181.3 ± 13.3 nM) (Table 1 and Supplementary Fig. 1a,g). Furthermore, the substitution of the 5′-nucleotide with adenosine or cytosine also severely impaired the binding activity to RsAgo (Kd: 5′A-RNA, 24.8 ± 1.7 nM; 5′C-RNA, 67.0 ± 3.5 nM) (Table 1 and Supplementary Fig. 1e,f).

These structural and biochemical observations are in agreement with the fact that RsAgo prefers 5′U-RNA\(^37\).

To investigate the functional role of the MID/PIWI interface of RsAgo in the RNA-guided DNA silencing mechanism, we measured plasmid DNA silencing activity in E. coli cells\(^37\) using RsAgo variants (Fig. 3c). On the basis of our present structural data, we selected the amino acid residues that interacted with the duplex via side and main chains are indicated in normal and italic types, respectively.
single strand RNA (Table 1 and Supplementary Fig. 1h–j). The silencing assay (Y463A/K467A, R481A/T484A and K506A) the three most effective mutants from the plasmid DNA replaced by a conserved lysine side-chain (K566 in Hs

However, in eukaryotic Argonaute, the magnesium ion is indicating the conservation of the magnesium coordination. 

Table 2 | Data collection, phasing and refinement statistics

|                           | Native       | SeMet        |
|---------------------------|--------------|--------------|
| **Data collection**        |              |              |
| Space group               | P2₁          | P2₁          |
| Cell dimensions           |              |              |
| a, b, c (Å)               | 68.2, 118.3, 118.5 | 67.6, 116.7, 117.7 |
| α, β, γ (°)               | 90.0, 95.8, 90.0 | 90.0, 95.6, 90.0 |
| Wavelength                | 1.0          | 0.97921      |
| Resolution (Å)            | 50–2.00      | 50–2.10      |
| Rmerge (°)                | 6.3 (40.0)   | 7.4 (43.7)   |
| completeness (%)          | 97.7 (100)   | 99.1 (100)   |
| Redundancy (%)            | 3.8 (3.8)    | 7.5 (7.3)    |
| **Refinement**            |              |              |
| Resolution (Å)            | 35.6–2.00    |              |
| No. reflections           | 117,604      |              |
| Rwork/Rfree (°)           | 18.4/23.3    |              |
| No. atoms                 | 11,669       | 5,104        |
| Protein                   | 11,669       | 5,104        |
| RNA/DNA/Mg²⁺              | 748/750/2    |              |
| Water                     | 646          |              |
| B-factors                 |              |              |
| Protein                   | 59.5         |              |
| RNA/DNA/Mg²⁺              | 50.1/55.2/40.1 |              |
| Water                     | 55.2         |              |
| R.m.s deviations          |              |              |
| Bond lengths (Å)          | 0.016        |              |
| Bond angles (°)           | 1.758        |              |

Values in parentheses are for the highest resolution shell.

*Rmerge = Σhkl(|I(hkl)−<I(hkl)>)/Σhkl|I(hkl)|, where <I(hkl)> is its average.

Contribution of the PAZ domain to DNA silencing. The PAZ domain of RsAgo (R204, R209, G210, L211, E242 and G243) mainly interacted with the middle region (nucleotides 8, 9, 11 and 12) of the guide RNA strand in the heteroduplex (Fig. 2c and Supplementary Fig. 3d). To evaluate the functional significance of this interaction, we performed a plasmid DNA silencing assay using RsAgo mutants of amino acid residues R204, R209 and E242, which interact with the heteroduplex via the side chains. The analysis showed that substitution of E242 with alanine had no marked effect on plasmid DNA silencing (Fig. 4a). However, the substitution of R204 or R209 with alanine, both of which are present in the same loop (which we named the PAZ loop, Fig. 4b), resulted in a significant reduction of the plasmid DNA silencing activity; the double mutation of these residues to alanines further reduced the activity (Fig. 4a). These results indicate the importance of the interaction between the PAZ loop and the guide RNA strand. In addition, the PAZ domain interacted with the guide RNA and target DNA together with L1 and L2 linkers. The details are described in the Supplementary Note (Supplementary Figs 3c and 6).

In the TtAgo complex, the interaction of the PAZ loop with the guide strand is also observed, although the nucleic acid type in the guide strand is different to that of RsAgo (Fig. 4b,c). On the other hand, analogous interactions involving the PAZ loop have not been observed for eukaryotic Argonaute proteins, because the structure of the eukaryotic Argonaute-RNA/RNA duplex complex containing base pairs beyond the seed region is still unresolved (Fig. 4d). However, in view of the conservation of the interaction of the PAZ loop with the guide strand in RsAgo and TtAgo, and considering the importance of the PAZ loop for plasmid DNA silencing in RsAgo, we suggest that the PAZ loop in the eukaryotic Argonaute protein will have an important role in the silencing activity.

In general, the PAZ domain of the Argonaute protein has two subdomains: an oligosaccharide-binding-fold like structure with one or two helices on one side; and an α-helix, a β-hairpin or loop structure followed by another α-helix (Supplementary Fig. 8a,b). These two subdomains are orientated to form a pocket (hereafter the PAZ pocket), and it has been demonstrated that the 3′-end of the guide strand is anchored within the PAZ pocket. However, the PAZ domain of RsAgo lacks the second subdomain, and therefore does not possess the PAZ pocket (Supplementary Fig. 8c). Thus, it seems that the binding mode of the 3′-end of the guide strand to the PAZ domain in RsAgo is different from that of other Argonaute proteins, or that the 3′-end of the guide strand might not bind to the PAZ domain in RsAgo. Further experiments are needed to clarify this point.

simultaneously. We confirmed that all mutants used in this study were expressed to the same level in E. coli cells (Supplementary Fig. 5). Mutants of the MID domain, Y463A/K467A, Q478A, R481A/T484A and K506A, strongly impaired plasmid DNA silencing activity (Fig. 3c). Subsequently, we measured the activity of single mutants, Y463A, K467A, R481A and T484A, and R481A/T484A and K506A, strongly impaired plasmid DNA silencing activity (9.3-fold reduction) (Fig. 3c). We next examined the contributions of the carboxyl group of L777 and the magnesium ion to the interaction with 5′P guide RNA by measuring Kd values. The ΔL777 RsAgo mutant exhibited low affinity for 5′P 18-base single strand RNA compared to wild-type RsAgo (42-fold reduction) (Table 1 and Supplementary Fig. 1k). Removal of the magnesium by EDTA caused the binding ability of the 5′P guide strand RNA to wild-type RsAgo to be reduced more than eightfold compared to that in the presence of magnesium (Table 1 and Supplementary Fig. 1d). These results indicate that the binding of 5′P guide strand RNA to RsAgo is supported by the magnesium ion and the carboxyl group of L777 in the PIWI domain. In addition to the above-mentioned amino acid residues R754 and L777, the non-slicing PIWI domain of RsAgo widely interacted with the heteroduplex. The details are described in the Supplementary Note (Supplementary Figs 3 and 6).
Functional N-terminal domain for target strand association.

Our crystal structure of the RsAgo complex showed that the 5’-region (nucleotides 14’–18’) of the target DNA strand in the heteroduplex was located on the positively charged surface of the N-terminal domain (Fig. 2b). On this surface, the target DNA strand (nucleotides 14’–16’) formed hydrogen bonds with the side chains of K49, R52, H62 and R97, and the main chain of W63 (Fig. 5a and Supplementary Fig. 3j). In order to investigate the functional role of the N-terminal domain, we performed a plasmid DNA silencing assay using the mutants K49A/R52A, H62A and R97A, and found that these mutations significantly reduced DNA silencing activity compared to wild-type RsAgo (Fig. 5b). To confirm the importance of the N-terminal domain, we truncated an α-helix-β-stand segment on the edge of the N-terminal domain (P45–W63, Fig. 5a), which contacts nucleotides 15’ and 16’ of the target DNA strand, and performed a plasmid DNA silencing assay. This assay showed that the truncation mutation reduced plasmid DNA silencing activity to a greater extent than the point mutations (Fig. 5b).

On the basis of the results of the DNA silencing assay, we carried out an electrophoretic mobility shift assay using the RsAgo-guide RNA (18-base) complex and [32P]-radiolabelled target DNA strand (70-base, Supplementary Table 1) to analyze the binding activity of the truncation mutant ΔP45–W63 to target DNA. The ΔP45–W63 truncation mutant showed no binding to its target DNA (Fig. 5c), despite the mutant having the same...
The heteroduplex segment required for recognition by RsAgo.

To identify the structural feature of the RNA/DNA heteroduplex that binds to RsAgo, we analyzed the minor groove width using CURVES. In general, RNA/DNA duplexes adopt an A-form similar to double-stranded DNA. However, the minor groove width of the RNA/DNA duplex structure bound to RsAgo was mid-way between the standard values for the A- (11 Å) and B-forms (7.4 Å) (Supplementary Fig. 9a,b). A notable feature was that the width of the minor groove at nucleotides 12:12' was clearly narrower than the standard value for the B-form. In general, sugar puckers in the A-form and B-form are predominately in C3'-endo and C2'-endo, respectively. In the RNA/DNA heteroduplex bound to RsAgo, the guide RNA strand maintained a strong preference for C3'-endo sugar puckering, whereas the target DNA strand showed various sugar puckering conformations (C2'-endo, C1'-exo and O4'-endo), probably due to interaction with RsAgo (Supplementary Table 2). Other geometrical parameters are shown in Supplementary Table 2. The most notable feature was that helical parameters of Tip and Buckle at nucleotides 12:12' have large negative values. In addition, the Roll angle from the 11:11' to 12:12' step was large because of the bend of the duplex structure between nucleotides 11:11' and 12:12' (Supplementary Table 2 and Supplementary Fig. 9c). This characteristic bending may be promoted by the interaction of nucleotides beyond 12:12' with the N-terminal and the PIW1 domains.

Previous studies showed that various Argonaute proteins bind to distinct types of nucleic acid, such as guide RNA/target RNA (all eukaryotes) and guide DNA/target DNA (most prokaryotes); here, we have identified a guide RNA/target DNA in *R. sphaeroides* (Figs 1 and 2). However, the selection mechanisms for guide/target strand types by each Argonaute protein are unknown. To determine which part of the guide/target strands is responsible for recognition by RsAgo, we performed an electrophoresis mobility shift assay using chimeric nucleic acids with a boundary between positions 11:11' and 12:12' based on the unique helical conformation of the heteroduplex bound to RsAgo, that is, the chimeric duplexes RR/DR and RR/DD, in which parts of 1'–11' and 12'–18' of the target DNA were replaced with RNA, respectively (Supplementary Table 3). Duplex recognition by RsAgo was disturbed by the replacement of the 1'–11' region of the target DNA with RNA (RR/DR), whereas recognition was maintained following replacement of the 12'–18' region with RNA (RR/DD) (Fig. 6a–c). These results suggest that the 1'–11' region of the target DNA is important for selection of strand type. To confirm this conclusion, we tested RsAgo binding to the other chimeric duplexes (Fig. 6d–f and Supplementary Table 3). Although the full-length guide DNA/target DNA duplex showed no binding to RsAgo (Fig. 1f), the RD/DD duplex, but not the RD/DR duplex, could bind to RsAgo (Fig. 6d,e). Moreover, the RD/DR duplex retained the RsAgo-binding ability (Fig. 6f). Overall, it is likely that a heteroduplex structure composed of the 1–11 region of guide RNA and the 1'–11' region of target DNA is required for recognition by RsAgo.

**Discussion**

The binding assays performed in this study showed that RsAgo specifically recognized RNA as a guide strand and DNA as a target strand (Figs 1 and 2). The structure of RsAgo in complex with the heteroduplex provides some insight into the mechanism of this specific guide and target recognition. RsAgo formed hydrogen bonds with the 2'-hydroxyl groups of 8 of the 18 nucleotides of the guide RNA strand (Fig. 2c). This binding to 2'-hydroxyl groups might contribute toward the ability of RsAgo to specifically recognize RNA as a guide strand. Point mutations of almost all the amino acid residues involved in the recognition of 2'-hydroxyl groups did not show marked effects on plasmid DNA silencing activity (Figs 3c and 4a and Supplementary Fig. 7c). Therefore, we infer that the amino acid residues of RsAgo, which recognize 2'-hydroxyl groups of the guide RNA, act
cooperatively and contribute to the selection of the nucleic acid type of the guide strand.

To investigate the mechanism of RsAgo target strand recognition, we examined the interaction between RsAgo and the minor groove of the seed region of the heteroduplex. Schirle et al. suggested that this interaction is crucial for target binding in HsAgo2. The structure determined here showed that RsAgo interacted with the minor groove of the seed region and with its periphery (Fig. 7a), as is the case in HsAgo2 (Fig. 7b). Specifically, the helix 28-turn-helix 29 segment in the L2 linker of RsAgo (S244, K245, E246, T249, Y260, I264, N265 and R275), which corresponds to helix α7 of HsAgo2, and two loops in the PIWI domain (K692, R693, F697 and L734–A737), which correspond to the loop containing I756 and Q757 and the loop containing R795 in the PIWI domain of HsAgo2, respectively, interacted with the minor groove and the backbone of the heteroduplex (nucleotides 4:4–8:8), making several hydrogen bonds and extensive van der Waals interactions (Fig. 7a,b and Supplementary Fig. 3b–d,h).

Considering the similarity of the recognition structures in the seed regions and also the conformational differences of the minor groove between the RNA/DNA duplex (Fig. 7a) and RNA/RNA duplex (Fig. 7b), we suggest that RsAgo (and possibly other Argonaute proteins) might distinguish the nucleic acid type of the target strand through examination of the conformation of the minor groove of the seed region.

Next, we docked an RNA/RNA duplex, which bound to HsAgo2, to RsAgo. We focused on the interaction with the 1:1′–11:11′ region of the duplex because this region is crucial for recognition by RsAgo (Fig. 6). The docking model showed that, unlike the case of target DNA (Fig. 7c), nucleotides 3′ and 4′ of the target RNA clashed with the area around R693, and that nucleotide 5 of the guide RNA clashed with the area around A737 in the PIWI domain (Fig. 7d). In addition, the helix 28-turn-helix 29 segment was separated from the seed region of the duplex (compare Fig. 7d with Fig. 7c). These shape mismatches for the guide RNA/target RNA duplex might also explain why RsAgo specifically recognizes DNA as the target strand.

Two possible mechanisms can be suggested for the selection of 5′U-RNAs at the guide strand loading stage: first, the guide RNA is introduced to RsAgo by an unknown loading machinery that selects 5′U-RNAs, and second, that RsAgo itself selects the 5′U-RNAs from the pool of available sequences as the guide strand. Olovnikov et al. suggested the latter possibility was more likely as the heterologous expression of RsAgo in E. coli results in the specific loading of 5′U-RNAs, as occurs for expression in R. sphaeroides cells. We concur with their conclusion because our structural and biochemical data showed that RsAgo preferably interacted with 5′U-guide RNA (Fig. 3a,b and Table 1). Moreover, the R. sphaeroides gene shows no homologies to Dicer or TRBP/PACT sequences, that is, small RNA-processing proteins. We therefore conclude that RsAgo itself directly selects 5′U-RNAs from the available pool of sequences, without the aid of any loading machinery.

In the present study, we solved the structure of RsAgo in a complex with a guide RNA/target DNA heteroduplex. Surprisingly, the structure showed that base pairing in the duplex is maintained in the 3′-region of the guide strand (nucleotides 14–18) by the packing provided by the N-terminal and the PIWI gene shows no expression for R. sphaeroides cells. We concur with their conclusion because our structural and biochemical data showed that RsAgo preferentially interacted with 5′U-guide RNA (Fig. 3a,b and Table 1). Moreover, the R. sphaeroides gene shows no homologies to Dicer or TRBP/PACT sequences, that is, small RNA-processing proteins. We therefore conclude that RsAgo itself directly selects 5′U-RNAs from the available pool of sequences, without the aid of any loading machinery.

In the present study, we solved the structure of RsAgo in a complex with a guide RNA/target DNA heteroduplex. Surprisingly, the structure showed that base pairing in the duplex is maintained in the 3′-region of the guide strand (nucleotides 14–18) by the packing provided by the N-terminal and the PIWI gene shows no expression for R. sphaeroides cells. We concur with their conclusion because our structural and biochemical data showed that RsAgo preferentially interacted with 5′U-guide RNA (Fig. 3a,b and Table 1). Moreover, the R. sphaeroides gene shows no homologies to Dicer or TRBP/PACT sequences, that is, small RNA-processing proteins. We therefore conclude that RsAgo itself directly selects 5′U-RNAs from the available pool of sequences, without the aid of any loading machinery.

In the present study, we solved the structure of RsAgo in a complex with a guide RNA/target DNA heteroduplex. Surprisingly, the structure showed that base pairing in the duplex is maintained in the 3′-region of the guide strand (nucleotides 14–18) by the packing provided by the N-terminal and the PIWI gene shows no expression for R. sphaeroides cells. We concur with their conclusion because our structural and biochemical data showed that RsAgo preferentially interacted with 5′U-guide RNA (Fig. 3a,b and Table 1). Moreover, the R. sphaeroides gene shows no homologies to Dicer or TRBP/PACT sequences, that is, small RNA-processing proteins. We therefore conclude that RsAgo itself directly selects 5′U-RNAs from the available pool of sequences, without the aid of any loading machinery.
N-terminus and PIWI domains. This model is consistent with the fact that RsAgo-associated RNAs are derived from cellular transcripts that lack distinct secondary structures. That is, in this situation, there is no need to release the passenger strand for target binding; therefore, the N-terminal domain of RsAgo does not have to function as a wedge to unwind the duplex, unlike TtAgo and HsAgo2 (Fig. 8b)9,29. Furthermore, our model can clearly explain the recently described behaviour in which the RsAgo–RNA complex loaded onto a complementary foreign target DNA (such as transposons, phage genes or exogenous plasmids) could inhibit RNA polymerase loading or RNA polymerase elongation, leading to the repression of transcription of the target DNA37. In this situation, it seems likely that, in addition to the base pairing in the seed region, base pairing at the 3′-region of the guide strand (nucleotides 14–18) might contribute to the strong interaction with the target strand. In conclusion, our present study of RsAgo represents a significant advance toward the complete understanding of the molecular mechanism of RNA-guided DNA silencing.

Methods

Plasmid constructions. We constructed expression plasmids encoding full-length RsAgo (residues 1–777). DNA fragments encoding the full-length Argonaute gene were amplified by PCR from the genomic DNA of R. sphaeroides (ATCC No. 17025D-5) using KOD plus NEO DNA polymerase (TOYOBO). The PCR product was cloned between the NdeI and XhoI sites of the pET-28a (Novagen) and thus the construct contained the sequence MGSSHHHHHHSSGLVPAGSH (6x His-tag and thrombin-recognition sequence) in the N terminus. Point and deletion mutants were generated in RsAgo plasmids by PCR site-directed mutagenesis48 using a pET-28a plasmid containing the RsAgo gene as the template (Supplementary Table 4). These modified Argonaute gene sequences in pET-28a were verified by DNA sequencing.

Protein expression. RsAgo wild-type (WT) and mutants were overexpressed in E. coli BL21 (DE3) cells (Novagen). The transformed E. coli cells were grown in LB medium containing 15 mg L⁻¹ kanamycin to an OD₆₀₀ of 0.6 at 37°C and expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside. The cells were further cultured at 25°C for 18 h and collected by centrifugation. To perform the crystal structure analysis, a Selenomethionine (SeMet)-labelled RsAgo NΔ20 mutant was overexpressed in E. coli B834(DE3) cells in SeMet core medium (Wako) supplemented with 20 g L⁻¹ d-glucose, 1 × MEM Vitamin solution (Sigma), 250 mg L⁻¹ MgSO₄, 4.2 mg L⁻¹ FeSO₄, 15 mg L⁻¹ kanamycin and 50 mg L⁻¹ L-SeMet (Wako).

Figure 6 | The guide RNA/target DNA heteroduplex region essential for binding to RsAgo. (a–f) Schematic representation of the guide and target strands in the duplex is given at the top of each gel panel. RNA and DNA components are shown in red and blue, respectively. Binding affinities of the 18-base chimeric duplex to RsAgo were examined using an electrophoresis mobility shift assay. The oligonucleotide sequences of the chimeric hybrid duplexes are shown in Supplementary Table 3. The positive control heteroduplex RR/DD (a) that normally binds to RsAgo protein was incubated with various concentrations (0, 1, 10, 100 and 1,000 nM) of recombinant RsAgo protein. Chimeric hybrid duplexes of RR/RD (b), RR/DR (c), RD/DD (d), RD/RD (e) and RD/DR (f) were incubated in the same way as the positive control RR/DD heteroduplex.
and samples were stored at −80°C. For crystallization, the purified protein was first purified using Ni-NTA agarose (Qiagen) and eluted with buffer A containing 100 mM Tris–HCl, pH 7.5, 1 M ammonium chloride, 5% glycerol and 5 mM β-mercaptoethanol and passed through a Resource Q column (GE Healthcare). The protein was concentrated using an Amicon Ultra 10 K filter and dialyzed against buffer B (20 mM Tris–HCl, pH 7.6, 300 mM NaCl and 5 mM β-mercaptoethanol) and disrupted by sonication. The lysate was clarified by centrifugation at 100,000 g for 30 min. The protein was then dialyzed against buffer C (20 mM Tris–HCl, pH 7.6, 500 mM NaCl and 5 mM cacodylate trihydrolate (pH 7.0 for native RsAgo and pH 7.4 for SeMet-labelled RsAgo) and 2-methyl-2,4-pentanediol (42% (v/v) for native RsAgo and 45% (v/v) for SeMet-labelled RsAgo). These crystals were grown at 20 °C using sitting-drop vapor-diffusion methods in a 24-well VDX plate (Hampton Research).

**Preparation of oligonucleotides.** The crystallization and biochemical analysis were carried out using synthetic oligonucleotides. The sequences were purchased from GenDesign, Inc. (Osaka, Japan). The sequences are shown in Supplementary Tables 1 and 3.

**Crystallization.** For crystallization, RsAgo-guide RNA:target DNA complexes were formed by mixing RsAgo (AN20) (native or SeMet-labelled), 18-base 5-phosphorylated RNA and 18-base DNA in a 1.1.2:1.2 molar ratio in a buffer containing 10 mM Tris–HCl, pH 7.6, 200 mM NaCl and 5 mM MgCl2 at 20 °C for 30 min. Crystallization drops were prepared by mixing 1 μl of the complex solution (10 mg ml⁻¹ of RsAgo) and 1 μl of a reservoir solution containing 0.05 M sodium cacodylate trihydrate (pH 7.0 for native RsAgo and pH 7.4 for SeMet-labelled RsAgo) and 1 M ammonium chloride, 5% glycerol and 5 mM β-mercaptoethanol and passed through a Resource Q column (GE Healthcare). The protein was concentrated using an Amicon Ultra 10 K filter and dialyzed against buffer C (20 mM Tris–HCl, pH 7.6, 300 mM NaCl and 5 mM β-mercaptoethanol) and then flash-cooled at 100 K. Datasets of native and SeMet derivatives (at the anomalous peak wavelength) were collected at beamline NW12A of KEK PF-AR (Tsukuba, Japan) using an ADSC Quantum 210r CCD detector, and were processed with the program HKL200049. Anomalous dispersion data for SeMet-labelled RsAgo were indexed and integrated with the program XDS50. The location of Se sites was calculated with the program autoSAHRP51. The initial model was built with the program BUCCANEER53, followed by model improvement by iterative cycles of manual model building with the program COOT55. The model was refined using the program REFMAC556 and the CCP4 program suite50. Twenty-eight Se sites were identified. The initial model was improved by iterative cycles of manual model building with the program REFMAC55 and maximum likelihood refinement with the program SHELXL06. Subsequently, density modification was performed using the program SOLOMON52. The initial model was built with the program BUCCANEER53, followed by model improvement against the native dataset using the program ARP/wARP54. The model was further improved by iterative cycles of manual model building with the program COOT55 and maximum likelihood refinement with the program REFMAC556. In the Ramachandran plot, 95.9% of the residues were included in the favored region, 3.4% were in the allowed region and 0.8% were in the outlier region. The statistics are shown as red and blue sticks, respectively, with phosphorous atoms in yellow. (a) Recognition of the seed region in the guide RNA/target DNA heteroduplex by RsAgo. The guide RNA (red) and target DNA (blue) strands are shown in surface representation. The L2 linker (light grey) and the PIWI domain (green) are shown in cartoon representation. Amino acid residues that interact with the heteroduplex are shown as sticks with surrounding dotted regions. Helices α8 (residues K245–L256) and α9 (N259–R275) are indicated with lines. (b) Recognition of the seed region in the guide RNA/target RNA duplex by HsAgo2 (PDB ID: 4W5O). The guide RNA and partial target RNA strands are indicated in red. Helix α7 corresponding to helices α8 and α9 of RsAgo is indicated with a line. (c,d) Steric hindrance between RsAgo and the guide RNA/target RNA duplex in the seed region. (c) Structure of RsAgo in complex with guide RNA/target DNA heteroduplex. RsAgo is shown in surface representation and colour-coded as in Fig. 2a. The guide RNA and target DNA are shown as red and blue sticks, respectively, with phosphorous atoms in yellow. (d) Surface representation of the RsAgo structure docked with the guide RNA/target RNA duplex (red) in the HsAgo2 complex (PDB ID: 4W5O). The regions of RsAgo that clash with the duplex are coloured in purple.

**Diffraction data collection.** For data collection, crystals were cryoprotected by a reservoir solution containing 25% (v/v) glycerol and then flash-cooled at 100 K. Datasets of native and SeMet derivatives (at the anomalous peak wavelength) were collected at beamline NW12A of KEK PF-AR (Tsukuba, Japan) using an ADSC Quantum 210r CCD detector, and were processed with the program HKL200049 and the CCP4 program suite50.

**Structure determination.** The structure was determined by the single-wavelength anomalous dispersion method. The locations of Se sites, refinement and phasing calculations were performed using the program autoSOARHPR51. Twenty-eight Se sites were identified and the amino acid residues were subsequently determined with the program REMACE50. In the Ramachandran plot, 95.9% of the residues were included in the favored region, 3.4% were in the allowed region and 0.8% were in the outlier region. The statistics are shown as red and blue sticks, respectively, with phosphorous atoms in yellow.
for data collection and refinement are summarized in Table 2. A representative 2F_o – F_e electron density map is displayed in Supplementary Fig. 10.

Electrophoretic mobility shift assay. RsAgo and 32P-labelled nucleic acids were incubated in 10 μl of binding buffer containing 20 mM HEPES-KOH, pH 7.5, 200 mM NaCl and 5 mM MgCl₂ for 10 min at 25°C. The concentration of radiolabelled nucleotides was fixed as 5 nM, whereas the concentration of RsAgo varied. After incubation, the reaction samples were mixed with 1 μl dye solution containing 50% glycerol, 0.1% bromophenol blue and 0.1% xylene cyanol. The nucleoprotein complexes were fractionated by electrophoresis (100 V, 1 h) through 6% native polyacrylamide gels (39:1 acrylamide/bisacrylamide) with 5 mM MgCl₂ in Tris–glycine buffer (25 mM Tris, 192 mM glycine) and were detected by autoradiography. Full-size images of the most important blot are presented in Supplementary Fig. 11.

Plasmid DNA silencing assay with E. coli cells. E. coli BL21(DE3) cells containing expression plasmids (pET-28a) of RsAgo WT and mutants were grown in LB medium at 37°C with kanamycin (15 μg ml⁻¹) to the log phase (OD₆₅₀ = 0.5). After β-β-1-thiogalactopyranoside (0.5 mM) induction, the cells were further incubated for 5 h at 30°C. The expression levels of all RsAgo variants were checked by SDS-PAGE. After the cell culture, plasmids were isolated using QIAprep Spin Miniprep Kit (Qiagen). The concentration of plasmid DNA was determined with a NanoDrop spectrophotometer (Thermo Scientific). This experiment was based on Olovnikov et al.37

Fluorescence polarization assay. Fluorescence polarization assays57 to measure the interactions of RsAgo and nucleic acids were performed using a Pan Vera Beacon 2000 fluorescence polarization instrument (Invitrogen). Polarization values were measured using a buffer containing 20 mM HEPES-KOH, pH 7.5, 200 mM NaCl, 5 mM MgCl₂, 0.01 mg ml⁻¹ E. coli total-tRNA and 0.01% NP-40. To obtain a titration curve, increasing amounts of RsAgo were added to the buffer with 10 pM of 3′-6-FAM-labelled nucleotide. The mixed samples were incubated for 10 min at 25°C before measurement and the polarization value was measured successively at 25°C. The data were fitted to a nonlinear regression using the one site specific binding function with GraphPad Software Prism 6.

Data availability. The crystal structure of RsAgo in complex with guide RNA and target DNA has been deposited in the RCSB Protein Data Bank (PDB) under PDB ID code 5AWH. All other data that support the findings of this study are available from the corresponding author upon reasonable request.

References
1. Meister, G. Argonaute proteins: functional insights and emerging roles. Nat. Rev. Genet. 14, 447–459 (2013).
2. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. Cell 115, 787–798 (2003).
3. Stark, A., Brennecke, J., Russell, R. B. & Cohen, S. M. Identification of Drosophila MicroRNA targets. PLoS Biol. 1, E60 (2003).
4. Haley, B. & Zamore, P. D. Kinetic analysis of the RNAi enzyme complex. Nat. Struct. Mol. Biol. 11, 599–606 (2004).
5. Song, J. J., Smith, S. K., Hannon, G. J. & Joshua-Tor, L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science 305, 1434–1437 (2004).
6. Yuan, Y. R. et al. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol. Cell 19, 405–419 (2005).
7. Wang, Y., Sheng, G., Jaraniek, S., Tuschi, T. & Patel, D. J. Structure of the guide-strand-containing argonaute silencing complex. Nature 456, 209–213 (2008).
8. Wang, Y. et al. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature 456, 921–926 (2008).
9. Wang, Y. et al. Nuclearia, propagation and cleavage of target RNAs in Ago silencing complexes. Nature 461, 754–761 (2009).
10. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. Science 336, 1037–1040 (2012).
11. Elkayam, E. et al. The structure of human argonaute-2 in complex with miR-20a. Cell 150, 100–110 (2012).
12. Nakanishi, K., Weinberg, D. E., Bartel, D. P. & Patel, D. J. Structure of yeast Argonaute with guide RNA. Nature 486, 368–374 (2012).
13. Frank, F., Hauver, J., Sonenberg, N. & Nagar, B. The crystal structure of the Argonaute2 PAZ domain reveals strand-mediated DNA target cleavage. Proc. Natl Acad. Sci. USA 111, 652–657 (2014).
14. Liu, J. et al. Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol. Cell 15, 185–197 (2004).
15. Joshua-Tor, L. The Argonaute. Cold Spring Harb. Symp. Quant. Biol. 71, 67–72 (2006).
16. Song, J. J. et al. Crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RISC effector complexes. Nat. Struct. Mol. Biol. 10, 1026–1032 (2003).
17. Yan, K. S. et al. Structure and conserved RNA binding of the PAZ domain. Nature 426, 468–474 (2003).
18. Ma, J. B., Ye, K. & Patel, D. J. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature 429, 318–322 (2004).
19. Ge, S., Jin, L., Huang, Y., Zhang, F. & Kay, M. A. Slicing-independent RISC activation requires the argonaute PAZ domain. Curr. Biol. 22, 1536–1542 (2013).
20. Hur, J. K., Zinchenko, M. K., Djuranovic, S. & Green, R. Regulation of Argonaute slicer activity by guide RNA 3’ end interactions with the N-terminal lobe. J. Biol. Chem. 288, 7829–7840 (2013).
21. Kwak, P. B. & Tomari, Y. The N domain of Argonaute drives duplex unwinding during RISC assembly. Nat. Struct. Mol. Biol. 19, 145–151 (2012).
22. Swarts, D. C. et al. The evolutionary journey of Argonaute proteins. Nat. Struct. Mol. Biol. 21, 743–753 (2014).
23. Ghildiyal, M. & Zamore, P. D. Small silencing RNAs: an expanding universe. Nat. Rev. Genet. 10, 94–108 (2009).
24. Swarts, D. C. et al. DNA-guided DNA interference by a prokaryotic Argonaute. Nature 507, 258–261 (2014).
25. Swarts, D. C. et al. Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. Nucleic Acids Res. 43, 5120–5129 (2015).
26. Swarts, D. C., Kohser, J. J., Westra, E. R., Schaap, P. J. & van der Oost, J. Effects of Argonaute on gene expression in Thermus thermophilus. PLoS ONE 10, e0124880 (2015).
27. Zander, A., Holzmeister, P., Kloese, D., Tinnefeld, P. & Grohmann, D. Single-molecule FRET supports the two-state model of Argonaute action. RNA 20, 45–56 (2014).
28. Ma, J. B. et al. Structural basis for 3’-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature 434, 666–670 (2005).
29. Olovnikov, L., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial argonaute samples the transcriptome to identify foreign DNA. Mol. Cell 51, 594–605 (2013).
30. Parker, J. S., Roe, S. M. & Barford, D. Structural insights into mRNA recognition from a PIWI domain–siRNA guide complex. Nature 434, 663–666 (2005).
31. Willkomm, S., Zander, A., Gust, A. & Grohmann, D. A prokaryotic twist on argonaute function. Life 5, 558–553 (2015).