A bacterial Argonaute with noncanonical guide RNA specificity

Emine Kaya,a,b,1 Kevin W. Doxzen,a,1 Kilian R. Knoll,a,2 Ross C. Wilson,b Steven C. Strutt,c Philip J. Kranzusch,a,d and Jennifer A. Doudna,a,b,c,d,e,f,g,3

*Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; 1Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; 2Biophysics Graduate Group, University of California, Berkeley, CA 94720; 3Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; 4Center for RNA Systems Biology, University of California, Berkeley, CA 94720; 2Innovative Genomics Initiative, University of California, Berkeley, CA 94720; and 3Department of Chemistry, University of California, Berkeley, CA 94720

Edited by David P. Bartel, Massachusetts Institute of Technology, Cambridge, MA, and approved March 3, 2016 (received for review December 10, 2015)

Eukaryotic Argonaute proteins induce gene silencing by small RNA-guided recognition and cleavage of mRNA targets. Although structural similarities between human and prokaryotic Argonautes are consistent with shared mechanistic properties, sequence and structure-based alignments suggested that Argonautes encoded within CRISPR-cas [clustered regularly interspaced short palindromic repeats (CRISPR)-associated] bacterial immunity operons have divergent activities. We show here that the CRISPR-associated Marinitoga piezophila Argonaute (MpAgo) protein cleaves single-stranded target sequences using 5′-hydroxylated guide RNAs rather than the 5′-phosphorylated guides used by all known Argonautes. The 2.0-Å resolution crystal structure of an MpAgo–RNA complex reveals a guide strand binding site comprising residues that block 5′ phosphate interactions. Using structure-based sequence alignment, we were able to identify other putative MpAgo-like proteins, all of which are encoded within CRISPR-cas loci. Taken together, our data suggest the evolution of an Argonaute subclass with noncanonical specificity for a 5′-hydroxylated guide.

Argonaute | small noncoding RNA | RNA interference

Argonaute (Ago) proteins bind small RNA or DNA guides, which provide base-pairing specificity for recognition and cleavage of complementary nucleic acid targets. Members of this protein family are present in all three domains of life (1). In eukaryotes, Argonautes are the key effectors of RNA interference (RNAi) pathways that regulate posttranscriptional gene expression (2–4). However, the role of Argonaute proteins in bacteria and archaea, which lack RNAi pathways, remains poorly understood (5).

Recent studies suggested that DNA-guided bacterial and archaeal Argonaute proteins are directly involved in host defense by cleaving foreign DNA elements, such as DNA viruses and plasmids (6, 7). In addition, a catalytically inactive Argonaute protein in Rhodobacter sphaeroides (RsAgo) was demonstrated to use RNA guides and possibly recruits an associated nuclease for subsequent target cleavage (8). Despite these divergent modes of action, bacterial and archaeal Argonaute proteins adopt a highly conserved biloled architecture. Herein, an N-terminal and a PIWI-Argonaute-Zwille (PAZ) domain constitute one lobe, whereas the other lobe consists of the middle (MID) domain and the catalytic RNAse H-like P element–induced wimpy testis (PIWI) domain (9–15).

Molecular structures of a eukaryotic Argonaute MID domain and an Archaeoglobus fulgidus Piwi (AIPiwi) enzyme bound to a guide RNA showed the importance of the 5′-terminal base identity, as well as the 5′ phosphate in guide strand binding, to Ago (10, 13–17). Notably, recognition of the 5′ end of the guide in the MID domain and guide strand preorganization for target interaction are conserved across the entire Argonaute superfamily (1).

The nucleic acid-guided binding and cleavage activities of Argonaute proteins are reminiscent of the activities of RNA-guided proteins within CRISPR-Cas systems (clustered regularly interspaced short palindromic repeats (CRISPR)-associated). CRISPR-Cas systems use the Cas1-Cas2 integrase complex to acquire viral sequences into the CRISPR locus (18, 19). On CRISPR RNA (crRNA) maturation (20–22), the resulting crRNAs assemble with one or more Cas targeting proteins capable of binding and cleaving foreign nucleic acids bearing a sequence complementary to the guide RNA (23, 24). Notably, the majority of CRISPR-cas loci in bacteria and archaea lack an Argonaute gene, and there is no evidence of Argonaute participation in any CRISPR system to date (25). However, in the genomes of several bacterial species, there is an Ago protein encoded within a cas gene operon (Fig. 1A). Due to their genomic context, we wondered whether these newly identified Ago proteins might have evolved distinct functions within the prokaryotic genome defense (26).

Here we show that the CRISPR-cas associated Marinitoga piezophila Ago (MpAgo) uses 5′-hydroxylated guide RNA guides that are chemically distinct from the 5′-phosphorylated guide strands used by all other Argonautes studied to date. The crystal structure of MpAgo bound to a guide RNA reveals the distinct coordination of the terminal 5′-hydroxyl group and how the canonical Argonaute fold has evolved to include a MID domain with unique 5′ end binding specificity. Based on structure-based sequence alignments, we were able to predict other CRISPR-cas associated argonaute genes and show that one of these candidates, the Argonaute protein from Thermotoga profundus, indeed has a preference for 5′-hydroxylated RNA guides.

Significance

Argonaute proteins are key effectors of RNAi interference and, in prokaryotes, function in host genome defense. We show here that a noncanonical clustered regularly interspaced short palindromic repeats (CRISPR)-associated Marinitoga piezophila Argonaute uses 5′-hydroxylated guide RNAs to recognize and cleave substrates rather than using the 5′-phosphorylated guides used by all other known Argonautes. By identifying other prokaryotic Argonautes with a 5′-hydroxyl RNA preference, our data suggest the evolution of an Argonaute subfamily with unique specificity for a 5′-hydroxylated guide.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 5I4A).

1E.K. and K.W.D. contributed equally to this work.

2Present address: Gene Center, Department of Biochemistry, Ludwig-Maximilians-Universität, 81377 Munich, Germany.

3To whom correspondence should be addressed. Email: doudna@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1524385113/-/DCSupplemental.
Results
Genomic Colocalization and Coexpression ofMpago with cas1 and cas2. Bioinformatic analysis revealed that prokaryotic Argonaute homologs colocalize within CRISPR-cas gene loci (5). Our recent sequence analysis of available prokaryotic genomes identified four examples of these putative operons (Fig. 1A). One of these CRISPR-cas adjacent argonaute genes is found in the genome of M. piezophila KA3, hereafter referred to as Mpago. Mpago is encoded downstream of cas1 and cas2 and is separated from cas2 by only 16 bp. Within this locus, a fourth gene is located 12 bp upstream of cas1 and is predicted to contain a primase domain similar to that found in the small subunit of archaeal and eukaryotic DNA primases (Pfam accession no. PF01896). Based on the composition of the neighboring cas interference operon, this locus belongs to CRISPR subtype III-B system (Table S1) (27). Furthermore, three CRISPR arrays flanking thecas genes share a common ~470-bp conserved leader sequence followed by an array of 36-bp conserved repeats and varying numbers of ~41-bp unique spacers, suggesting an active CRISPR system.

As an initial test ofMpago gene expression, we used RT-PCR using total RNA obtained from M. piezophila cell lysate and PCR primers to detect full-length gene products. We observed robust amplification of the cas1, cas2, and ago genes products (Fig. 1B and Fig. S1A). Furthermore, transcripts encoding cas1–cas2, as well as cas2–ago, could be amplified, suggesting that cas1, cas2, and ago constitute a single transcript. To test whether Mpago is expressed in its native host, we purified recombinant Mpago from Escherichia coli and generated an Mpago-specific polyclonal antibody (Fig. S1B and C). Using Western blot analysis, low levels of soluble Mpago were detected in M. piezophila cell lysate (Fig. S1D). However, attempts to purify native protein using immunoprecipitation were not successful, consistent with previous reports of low Ago concentrations in bacteria and archaea (6–8).

5′ Hydroxyl RNA-Guided Cleavage of Single-Stranded DNA and RNA. We next tested whether purified recombinant Mpago is an active enzyme using in vitro nucleic acid cleavage experiments. Because the biological guides and targets for Mpago are unknown, we tested canonical 21-nucleotide (nt) RNA or DNA guides containing a 5′ phosphate or 5′ hydroxyl group for sequence-specific cleavage of RNA or DNA target strands (Table S2). Strikingly, Mpago uses a 5′-hydroxylated RNA guide to cleave both complementary single-stranded RNA and DNA targets (Fig. 1C). This 5′ hydroxyl guide preference has not been observed in other eukaryotic or prokaryotic Argonaute homologs, all of which use 5′-phosphate–containing RNA or DNA guides for target interference.

Cleavage of the target strand by Mpago occurs after the 10th nucleotide counting from the 5′ end of the guide strand, consistent with previously characterized Ago homologs (9, 12, 28, 29). We also observed weak cleavage activity when Mpago was programmed with a 5′-phosphorylated RNA guide, resulting in products 1–2 nts longer than for canonical cleavage counting from the 3′ end of the target. This cleavage pattern is similar to that catalyzed by human Ago2 (hAgo2) when loaded with a 5′-hydroxylated RNA guide, which
caused a shift of the cleavage site by one base in the 5′ direction relative to its normal position (28). In vitro cleavage of a double-stranded DNA target could not be detected (Fig. 1C).

We verified that alanine substitution of residue D516 (underline) within the predicted DEDXX motif in MpAgo (13), abolished detectable cleavage activity (Fig. 2D). Furthermore, cleavage occurred only with guide RNAs bearing sequence complementarity to the substrate strand. We next tested substrate cleavage kinetics in the presence of different divalent metal ions, which are required for Ago activity (9, 29), to explore whether divalent metal ion identity contributes to the discrimination between 5′ hydroxyl and 5′ phosphate guide RNA termini (30) (Fig. S2A). In the presence of different divalent metal ions (Mg²⁺, Ca²⁺, Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cu²⁺, and Zn²⁺), we found that the 5′ hydroxyl guide RNA preference remained but overall substrate cleavage rates increased ~10-fold in the presence of MnCl₂ compared with MgCl₂ (Fig. S2A). With these optimized cleavage conditions in hand, cleavage kinetics were measured using either 5′-hydroxylated or 5′-phosphorylated guide RNAs targeting perfectly complementary ssDNA and ssRNA targets (Fig. 2B and Fig. S2B). The fastest reaction rates were observed for 5′-hydroxyl-RNA-guided ssDNA cleavage followed by 5′-hydroxyl-RNA-guided ssRNA cleavage. We tested MpAgo cleavage efficiency using guide RNAs of different lengths, revealing that RNAs between 17 and 21 nt long supported similar rates of target strand cleavage, and this cleavage efficiency was maintained in the presence of longer guides (up to 40 nt in length) (Fig. 2C and Fig. S2C). In contrast to observations for some other Argonaute proteins (31, 32), we did not observe a preference for a 5′ end nucleotide as all guides tested resulted in efficient target cleavage (Fig. 2D and Fig. S2D).

Crystal Structure of an MpAgo–Guide RNA Complex Explains 5′ Hydroxyl Preference. To determine how MpAgo has evolved to use 5′ hydroxyl guide RNAs, we determined a 2.0-Å resolution crystal structure of MpAgo bound to a 5′-hydroxylated guide RNA (Fig. 3A and Table S3). Although the overall fold of MpAgo is similar to previous Ago structures, significant divergence in both the N and PAZ domains is consistent with bioinformatic analysis showing that the N/PAZ lobe is the least conserved region of Argonaute proteins (33) (Fig. S3A). In particular, the MpAgo PAZ domain is smaller than that from Thermus thermophilus Argonaute (TtAgo), and the N domain differs in both its secondary structure and its orientation relative to the MID/PIWI lobe.

The 5′- and 3′-terminal nucleotides of the guide strand are anchored in the MID and PAZ domains, respectively. The first nucleotide is flipped into a compact pocket formed at the interface of the MID and PIWI domains, precluding the base from any potential interaction with a target strand. Interestingly, the MID binding pocket of MpAgo is lined with hydrophobic residues (I380, V387, P398, L635, Y636), which surround the 5′ hydroxyl group of the guide strand (Fig. 3B, Left). This unique hydrophobic pocket is distinct from all previously determined Ago–guide complex structures, which contain MID domains with several highly conserved charged residues that coordinate a metal ion or interact directly with the 5′ phosphate terminus (12–15) (Fig. 3B, Right).

In the absence of terminal phosphate interactions, the first nucleotide of the guide strand remains anchored within the MID binding pocket by two contacts (Fig. 3B, Left). The 5′ hydroxyl is hydrogen bonded to the second phosphate group of the guide backbone, and the inverted first base is coordinated by pi-stacking with Y379, a residue in the canonical “nucleotide preference loop” of the MID domain (16, 34). Although MpAgo did not exhibit a significant preference for a specific 5′ guide nucleotide (Fig. 2D and Fig. S2D), the adjacent amide and carbonyl of I380 are positioned close enough to interact with the first base (Fig. S3B). This observation suggests that a majority of the stabilization energy is generated from the stacking with Y379 rather than base-specific interactions.

The inability to bind a 5′ phosphate is also explained by compression of the binding pocket due to the presence of an ordered α-helix (α5) at the C terminus of the PIWI domain (Fig. 3B and C). A surface representation of the MpAgo MID domain binding pocket shows inadequate space to accommodate a 5′ phosphate (Fig. 3D). By contrast, the binding pocket of TtAgo is both larger and solvent exposed (Fig. S3C). Previously, α5 of the PIWI domain had only been observed in eukaryotic Argonautes and showed no sequence homology to MpAgo. These eukaryotic helices are positioned below α4 and further away from the

![Fig. 2.](image-url) MpAgo cleaves ssDNA and ssRNA targets and has no 5′ end nucleotide specificity. (A) In vitro cleavage assay using MpAgo and ssDNA targets. The first line above the gel indicates the radiolabeled component of the reaction shown below. The dotted line separates areas of the gel with different contrast and brightness settings. The ssDNA target is perfectly complementary to the 5′ hydroxyl-RNA guide used, whereas non-target ssDNA shares no sequence complementarity; wt, wildtype MpAgo; mut, MpAgo D516A variant. (B) Cleavage kinetics of ssDNA and ssRNA targets using RNA-guided MpAgo. Results from three independent experiments were quantified. Error bars represent SD of three independent experiments. (C) RNA guides of 12–21, 25, 30, and 40 nts in lengths were tested for cleavage efficiency in vitro. Error bars represent SD of three independent experiments. (D) Permutations of the first nucleotide on the 5′ end of the guide are tolerated by MpAgo and lead to efficient cleavage of an ssDNA target. Assays were performed in three independent experiments and plotted using nonlinear regression. SDs are represented by error bars.
binding pocket, leaving room for 5′ phosphate binding (14, 15). Thus, the unique coordination of the 5′ nucleotides of the guide within the MID domain, along with additional structural constriction by the PIWI domain, explain the preference for MpAgo to use guides bearing a terminal 5′ hydroxyl group.

MpAgo Kinks the Guide RNA. Nucleotides 2–8 of Ago-bound guide strands—the seed region—nucleate target base pairing due to preorganization in an A-form-like geometry (3). However, alignment of the MpAgo and TtAgo PIWI domains bound to their respective guides reveals a unique conformation of the seed region of the MpAgo-bound guide strand (Fig. 4/4). Although the first three 5′ nucleotides of the guides superimpose well, the DNA guide in TtAgo extends into the PIWI domain, whereas the MpAgo-bound guide arcs up toward the PAZ domain. The PAZ residue Y166 disrupts the A-form trajectory of the guide, introducing a sharp kink between nucleotides A6 and A7 that separates the first six nucleotides from the rest of the MpAgo-bound RNA (Fig. 4B). Human Argonaute 2 shows a similar but less pronounced kink between bases 6 and 7, caused by I365, which extends from the L2 linker rather than the PAZ domain (14, 15). The guide strand following A7 is stabilized through phosphate backbone interactions with multiple charged residues extending from the L1 linker (Fig. S4). Flipped nucleotide A11 is stabilized through pi-stacking with Y89 within the N domain (Fig. 4C). Base specific contacts are made with A11 by N107, which extends from the L1 linker. As is typical of guide-bound Ago structures, nucleotides 12 through 19 are poorly ordered, but density reappears for nucleotides 20 and 21 bound to the PAZ domain (12, 14, 15) (Fig. 4A).

Complementarity between the guide strand seed sequence and the target sequence determines Ago cleavage specificity (35–37). Because the MpAgo-bound guide RNA is kinked at the 3′ end of the seed sequence, we tested whether target mismatches within this region affect cleavage efficiency. Surprisingly, single mismatches at nucleotides 5, 7, and 8 of the guide showed the greatest reduction in target cleavage, whereas mismatches at positions 2–4 were less inhibitory (Fig. 4D and Fig. S4B). These results suggest that complementary binding to the nucleotides surrounding the kink could be necessary to release the contorted strand and propagate binding to the 3′ end of the guide. Mutation of Y166 to alanine had no effect on cleavage efficiency (Fig. S4C) and it remains to be determined how the kink is formed.

Catalytically Competent DEDD Motif in the MpAgo PIWI Domain. The PIWI domain of Argonaute proteins adopts an RNase H fold and contains a conserved DEDX catalytic tetrad (where X is commonly a His or Asp), which coordinates the two divalent cations required for catalysis (13). However, in the precleavage state of the enzyme, a glutamate residue (called glutamate finger) in the DEDX tetrad is located away from the active site in an invariant linker region and moves into the catalytic pocket only on release of the guide-strand 3′ end from the PAZ domain.

Fig. 3. Crystal structure of MpAgo and its implications for 5′-hydroxyl RNA guide binding. (A) (Top) Domain architecture of MpAgo with residues numbered. The DEDN catalytic tetrad is labeled in red. (Middle) MpAgo structure with N (green), Linker L1 (gray), PAZ (magenta), Linker L2 (pink), MID (purple), and PIWI (blue) domains in a cartoon representation, bound to a guide RNA (orange). (Bottom) Guide RNA sequence with ordered nucleotides in orange and disordered nucleotides in black. (B) Left) The MID (purple) and PIWI (blue) domains of MpAgo form the binding pocket for the 5′ end of the RNA guide. The presence of hydrophobic residues within the MID pocket deters binding of a 5′ phosphate guide. (Right) The interface of the MID (gray) and PIWI (yellow) domains of TtAgo (3DLH) form the binding pocket for the 5′ end of a DNA guide. A Mg2+ ion (purple) along with other charged residues from the MID domain form interactions with the 5′ phosphate. (C) In MpAgo (blue), an additional helix (αs) compresses the MID domain compared with the MID domain in TtAgo (3DLH, yellow). (D) An electrostatic heat map (red = negatively charged, blue = positively charged, white = neutral) shows the first base of the MpAgo guide is securely anchored, and the white surface around the 5′ hydroxyl indicates a lack of charged residues for possible 5′ phosphate binding.
Hydroxyl Guide Specificity.

end recognition, we generated a structure-based comparison of guide strand-containing Argonaute protein structures including hAgo2, TtAgo, and Archaeglobus fulgidus Piwi (A/Piwi) reveals a conserved set of four amino acids in the MID domain, [R/Y]-K-Q-K, that contributes to 5′ phosphate recognition on the guide strand (10, 29, 38). To pinpoint residues important in differential 5′ end recognition, we generated a structure-based MID-domain alignment using the above protein sequences, as well as those of R. sphaeroides Ago (RsAgo) and prokaryotic Argonautes (pAgos) closely related to MpAgo (Fig. S5A). As anticipated, RsAgo aligns well with TtAgo and contains residues involved in guide-strand 5′ phosphate binding, consistent with copurification of 5′-phosphorylated RNAs with RsAgo (8). By contrast, amino acids responsible for 5′ phosphate coordination are substituted with more hydrophobic residues in MpAgo. A BLAST search against the Refseq protein database revealed two putative pAgo sequences from Thermotoga profunda (TpAgo) and Marinimicrospora sp. 1155 (MsAgo) that share greater than 40% sequence identity with MpAgo (Fig. L4 and Fig. S5A). TpAgo and MsAgo, which are also located within a CRISPR locus in an operon similar to that for MpAgo, contain hydrophobic residues at the positions responsible for guide-strand 5′ end binding. The presence of similar hydrophobic residues suggests that TpAgo and MsAgo are also capable of preferential binding to 5′-hydroxylated RNAs, as seen in MpAgo. These two candidates also share a conserved tyrosine (Y166 in MpAgo) in the PAZ domain that kinks the guide RNA, suggesting that the guide strand in these pAgos may also adopt a similar unique conformation. To test the prediction that TpAgo preferentially uses 5′-hydroxylated guide RNAs, we performed in vitro ssDNA cleavage experiments using both 5′-hydroxylated and 5′-phosphorylated guide RNAs (Fig. S5B). We observed a sixfold increase in cleavage efficiency when TpAgo was programmed with a 5′-hydroxylated RNA compared with cleavage with a 5′-phosphorylated guide RNA.

**Discussion**

Prokaryotic Argonaute proteins have been implicated in host genome defense using divergent mechanisms (6–8). Although TtAgo and PiAgo use small DNA guides to direct the cleavage of target DNA strands, the RNA-guided RsAgo induces plasmid degradation by an unknown, indirect mechanism. We present here a third mechanism of target nucleic acid cleavage by a bacterial Argonaute, in which MpAgo uses chemically distinct 5′ hydroxyl guide RNAs to bind and cleave single-stranded DNA and RNA targets. MpAgo reveals a nucleic acid usage that is uncommon for catalytically active pAgos. First, it binds preferentially to small RNA guides, which is uncommon for catalytically active pAgos. Second, MpAgo has a strong preference for a terminal 5′ hydroxyl group on the guide strand (10, 29, 38). To pinpoint residues important in guide-strand 5′ end recognition, we generated a structure-based comparison of guide strand-containing Argonaute protein structures including hAgo2, TtAgo, and Archaeglobus fulgidus Piwi (A/Piwi) reveals a conserved set of four amino acids in the MID domain, [R/Y]-K-Q-K, that contributes to 5′ phosphate recognition on the guide strand (10, 29, 38). To pinpoint residues important in differential 5′ end recognition, we generated a structure-based MID-domain alignment using the above protein sequences, as well as those of R. sphaeroides Ago (RsAgo) and prokaryotic Argonautes (pAgos) closely related to MpAgo (Fig. S5A). As anticipated, RsAgo aligns well with TtAgo and contains residues involved in guide-strand 5′ phosphate binding, consistent with copurification of 5′-phosphorylated RNAs with RsAgo (8). By contrast, amino acids responsible for 5′ phosphate coordination are substituted with more hydrophobic residues in MpAgo. A BLAST search against the Refseq protein database revealed two putative pAgo sequences from Thermotoga profunda (TpAgo) and Marinimicrospora sp. 1155 (MsAgo) that share greater than 40% sequence identity with MpAgo (Fig. L4 and Fig. S5A). TpAgo and MsAgo, which are also located within a CRISPR locus in an operon similar to that for MpAgo, contain hydrophobic residues at the positions responsible for guide-strand 5′ end binding. The presence of similar hydrophobic residues suggests that TpAgo and MsAgo are also capable of preferential binding to 5′-hydroxylated RNAs, as seen in MpAgo. These two candidates also share a conserved tyrosine (Y166 in MpAgo) in the PAZ domain that kinks the guide RNA, suggesting that the guide strand in these pAgos may also adopt a similar unique conformation. To test the prediction that TpAgo preferentially uses 5′-hydroxylated guide RNAs, we performed in vitro ssDNA cleavage experiments using both 5′-hydroxylated and 5′-phosphorylated guide RNAs (Fig. S5B). We observed a sixfold increase in cleavage efficiency when TpAgo was programmed with a 5′-hydroxylated RNA compared with cleavage with a 5′-phosphorylated guide RNA.

An Argonaute Subfamily with 5′ Hydroxyl Guide Specificity. Comparison of guide-strand-containing Argonaute protein structures including hAgo2, TtAgo, and Archaeglobus fulgidus Piwi (A/Piwi) reveals a conserved set of four amino acids in the MID domain, [R/Y]-K-Q-K, that contributes to 5′ phosphate recognition on the guide strand (10, 29, 38). To pinpoint residues important in differential 5′ end recognition, we generated a structure-based MID-domain alignment using the above protein sequences, as well as those of R. sphaeroides Ago (RsAgo) and prokaryotic Argonautes (pAgos) closely related to MpAgo (Fig. S5A). As anticipated, RsAgo aligns well with TtAgo and contains residues involved in guide-strand 5′ phosphate binding, consistent with copurification of 5′-phosphorylated RNAs with RsAgo (8). By contrast, amino acids responsible for 5′ phosphate coordination are substituted with more hydrophobic residues in MpAgo. A BLAST search against the Refseq protein database revealed two putative pAgo sequences from Thermotoga profunda (TpAgo) and Marinimicrospora sp. 1155 (MsAgo) that share greater than 40% sequence identity with MpAgo (Fig. L4 and Fig. S5A). TpAgo and MsAgo, which are also located within a CRISPR locus in an operon similar to that for MpAgo, contain hydrophobic residues at the positions responsible for guide-strand 5′ end binding. The presence of similar hydrophobic residues suggests that TpAgo and MsAgo are also capable of preferential binding to 5′-hydroxylated RNAs, as seen in MpAgo. These two candidates also share a conserved tyrosine (Y166 in MpAgo) in the PAZ domain that kinks the guide RNA, suggesting that the guide strand in these pAgos may also adopt a similar unique conformation. To test the prediction that TpAgo preferentially uses 5′-hydroxylated guide RNAs, we performed in vitro ssDNA cleavage experiments using both 5′-hydroxylated and 5′-phosphorylated guide RNAs (Fig. S5B). We observed a sixfold increase in cleavage efficiency when TpAgo was programmed with a 5′-hydroxylated RNA compared with cleavage with a 5′-phosphorylated guide RNA.

Fig. 4. Unusual guide conformation and noncanonical active site are revealed by MpAgo’s crystal structure. (A) The structure of TtAgo (3DLH) and MpAgo were aligned using their PIWI domains. The first three nucleotides of the TtAgo DNA (sea green) and MpAgo RNA (orange) align, but then the guides begin to drastically diverge (B) O$_{\text{P}}$-F$_{\text{p}}$ Electron density map, contoured at 1σ, of nucleotides 1–9, 11, and 20–21. The bases for nucleotides 6–9 were modeled in due to incomplete density. Y166 from the PAZ domain induces a kink in the guide strand between A6 and A7. (C) Nucleotide A11 orthogonally flips relative to the preceding nucleotides of the guide strand. This flip is stabilized by π-stacking with Y89 from the N domain (green) and base-specific contacts, depicted as dotted black lines, with N107 from the L1 linker (gray). (D) Single nucleotide mismatches were introduced at positions 2–8 counting from the 5′ end of the RNA guide and tested for their effect on cleavage efficiency (Fig. S4C). Error bars represent SD of three independent experiments. (E) Close-up view of the PIWI domain with potential active site residues highlighted. Black arrow indicates a predicted movement of a glutamate finger toward the catalytic center. (F) WT MpAgo and catalytic site mutants were tested for cleavage activity in time course experiments.
preferences. Although all previously solved Ago structures contain a charged binding pocket for guide-strand 5′ phosphate recognition, MpAgo instead contains a hydrophobic 5′ end binding pocket devoid of charged residues. This pocket is compressed by an additional C-terminal α-helix of the PIWI domain, not found in other pAgo structures, that sterically hinders accommodation of a 5′ phosphate. The kinked guide-strand conformation, in which Y166 from the PIWI domain disrupts the α-helix trajectory of guide nucleotides 2–6, is distinct from the guide conformations observed in other Ago structures. Nucleotides 2–4 remain solvent exposed and thus competent for target scanning (37). Substrate binding to this part of the guide could reposition the PIWI domain, enabling base pairing to the rest of the guide. A similar mechanism has been proposed for hAgo2, where helix 7 of the L1 linker disrupts the guide strand at the same position as Y166 of MpAgo (38).

Structure-based alignments allowed us to identify and experimentally test other pAgo family members predicted to have a 5′ hydroxyl binding preference, and these are also encoded within CRISPR-cas operons. In vivo experimentation will be necessary to determine the biological function of these Ago enzymes. In particular, it will be interesting to identify the biogenesis pathway of the 5′-hydroxylated guide RNAs and determine how they are loaded into the enzyme. Since “guilt by association” suggests that these proteins have been hijacked by the CRISPR-Cas system, MpAgo might provide another layer of complexity in the phage-host arms race. Its unique guide RNA recognition specificity may also be useful for applications including programmed RNA capture and cleavage in heterologous systems.

Materials and Methods
Details of the materials and methods used in this study, including cloning and protein purification, in vitro activity assays, immunoprecipitation, crystallography, and a list of nucleic acids used, are provided in SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Dr. Kenneth Noll (Department of Molecular and Cell Biology, University of Connecticut) for providing M. piezophae cells; Akshay Tambe, Nicolas Bray, and Megan Hochstrasser for helpful discussions; and J. Holton, G. Meigs, and A. Gonzalez for technical assistance with crystallization and processing. We thank the staff at beamline 8.3.1 of the Advanced Light Source at Lawrence Berkeley National Laboratory for crystallographic data collection. E.K. was funded by the German Academic Exchange Program, and K.W.D. acknowledges support from the National Science Foundation Graduate Research Fellowship. P.I.K. is supported as a Howard Hughes Medical Institute (HHMI) Fellow of the Life Sciences Research Foundation. This work was supported in part by a grant from the National Science Foundation (to J.A.D.). J.A.D. is an HHMI Investigator.