A bacterial Argonaute with noncanonical guide RNA specificity

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

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [b]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [c]Biophysics Graduate Group, University of California, Berkeley, CA 94720; [d]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; [e]Center for RNA Systems Biology, University of California, Berkeley, CA 94720; [f]Innovative Genomics Initiative, University of California, Berkeley, CA 94720; and [g]Department 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 bilobed 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 (AfPiwi) 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. 1*A*). 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 RNAs 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 profunda*, indeed has a preference for 5′ hydroxylated RNA guides.

Significance

Argonaute proteins are key effectors of eukaryotic RNA 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.

Author contributions: E.K. and J.A.D. designed research; E.K., K.W.D., K.R.K., and S.C.S. performed research; E.K., K.W.D., R.C.W., S.C.S., and P.J.K. analyzed data; and E.K., K.W.D., and J.A.D. wrote the paper.

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).

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

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

[3]To 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.

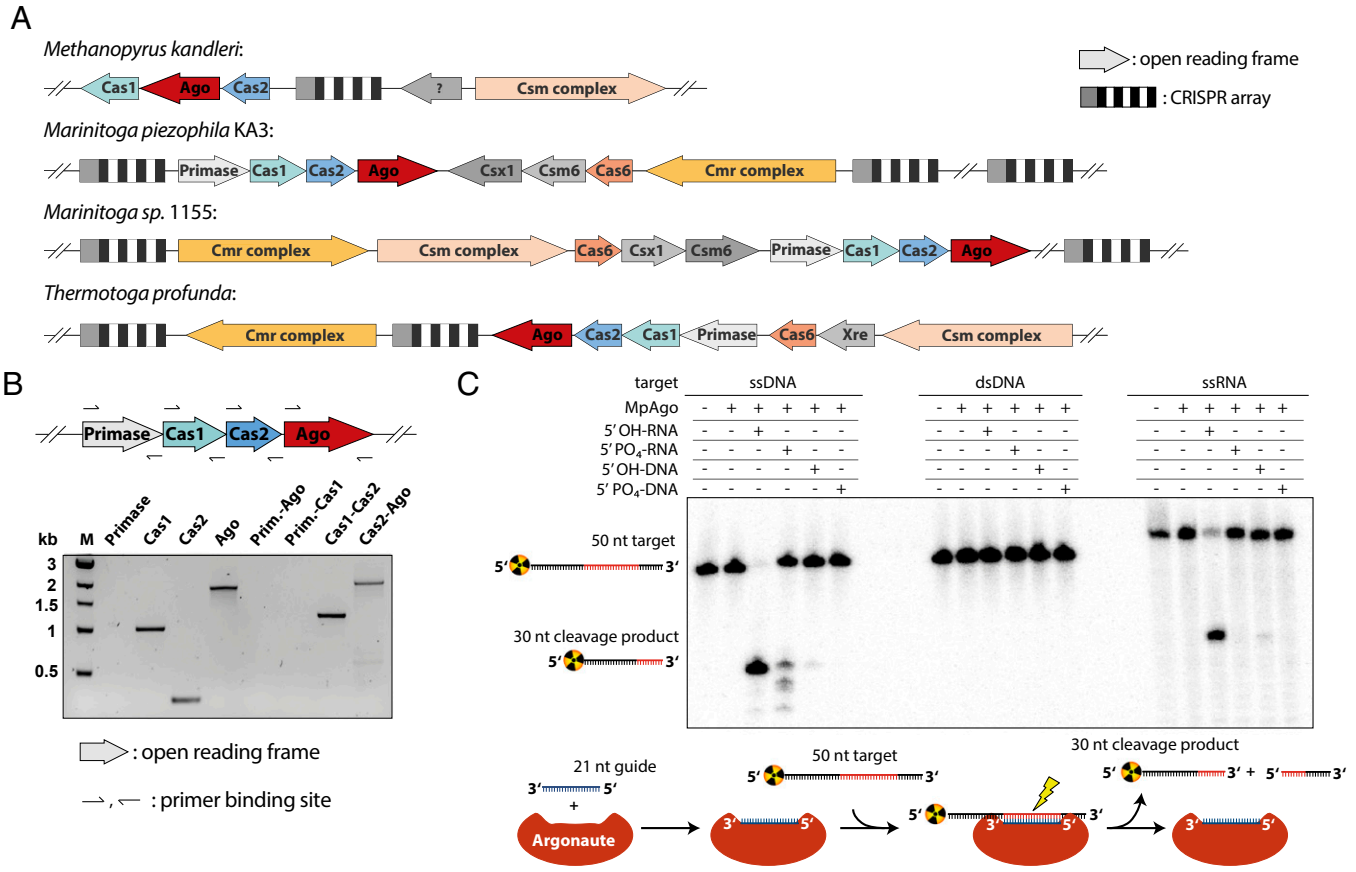


Fig. 1. MpAgo is a CRISPR-associated endonuclease with a 5′ hydroxyl RNA guide specificity. (*A*) Nonredundant (nr) protein–protein BLAST search revealed four prokarotic species with *ago* genes encoded within CRISPR-*cas* operons. These Ago homologs colocalize with *cas1* and *cas2* in diverse CRISPR subtype III loci. (*B*) RT-PCR of the acquisition operon using *M. piezophila* cDNA. (*Upper*) Forward and reverse primers annealing to the ends of the individual genes of *primase*, *cas1*, *cas2*, and *ago* were used to amplify cDNA products. (*Lower*) cDNA products were subsequently separated on a TAE Agarose gel and visualized via SYBR safe staining. (*C*) (*Upper*) In vitro assays with MpAgo in the presence of different 21-nt guides and 50-nt radiolabeled ssDNA, dsDNA, or ssRNA targets. Successful cleavage reactions are expected to yield ∼30-nt radiolabeled cleavage products. (*Lower*) Depiction of the workflow during in vitro cleavage analysis.

Results

Genomic Colocalization and Coexpression of Mp*ago* 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. 1*A*). 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 the *cas* 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 of Mp*ago* 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* gene products (Fig. 1*B* and Fig. S1*A*). 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. S1 *B* and *C*). Using Western blot analysis, low levels of soluble MpAgo were detected in *M. piezophila* cell lysate (Fig. S1*D*). 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. 1*C*). 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. 1*C*).

We verified that alanine substitution of residue D516 (underline) within the predicted DEDX motif in MpAgo (13), abolished detectable cleavage activity (Fig. 2*A*). 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. S2*A*). In the presence of different divalent metal ions (Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}), we found that the 5′ hydroxyl guide RNA preference remained but overall substrate cleavage rates increased ∼10-fold in the presence of $MnCl_2$ compared with $MgCl_2$ (Fig. S2*A*). 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. 2*B* and Fig. S2*B*). 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 nts in length) (Fig. 2*C* and Fig. S2*C*). 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. 2*D* and Fig. S2*D*).

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. 3*A* 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. S3*A*). 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 (I383, V387, P398, L635, Y636), which surround the 5′ hydroxyl group of the guide strand (Fig. 3*B*, *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. 3*B*, *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. 3*B*, *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. 2*D* and Fig. S2*D*), the adjacent amide and carbonyl of I380 are positioned close enough to interact with the first base (Fig. S3*B*). 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. 3 *B* and *C*). A surface representation of the MpAgo MID domain binding pocket shows inadequate space to accommodate a 5′ phosphate (Fig. 3*D*). By contrast, the binding pocket of TtAgo is both larger and solvent exposed (Fig. S3*C*). 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

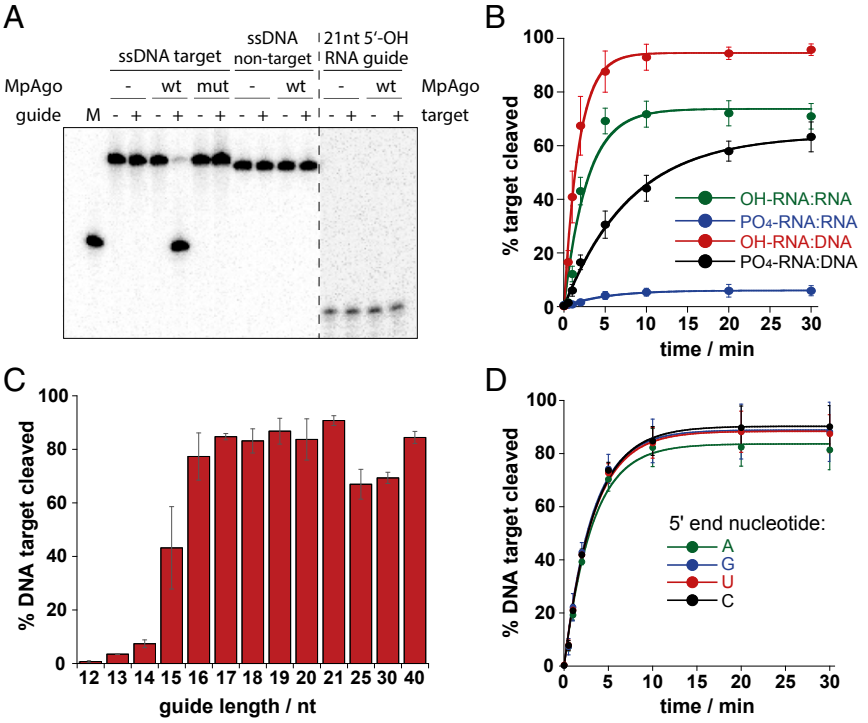


Fig. 2. MpAgo cleaves ssDNA and ssRNA targets and has no 5′ end nucleotide specificity. (*A*) In vitro cleavage assay using MpAgo and ssDNA targets. The first line above the gel indicates the radiolabeled component of the reaction shown below. The dotted line separates areas of the gel with different contrast and brightness settings. The ssDNA target is perfectly complementary to the 5′ hydroxyl-RNA guide used, whereas nontarget 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.

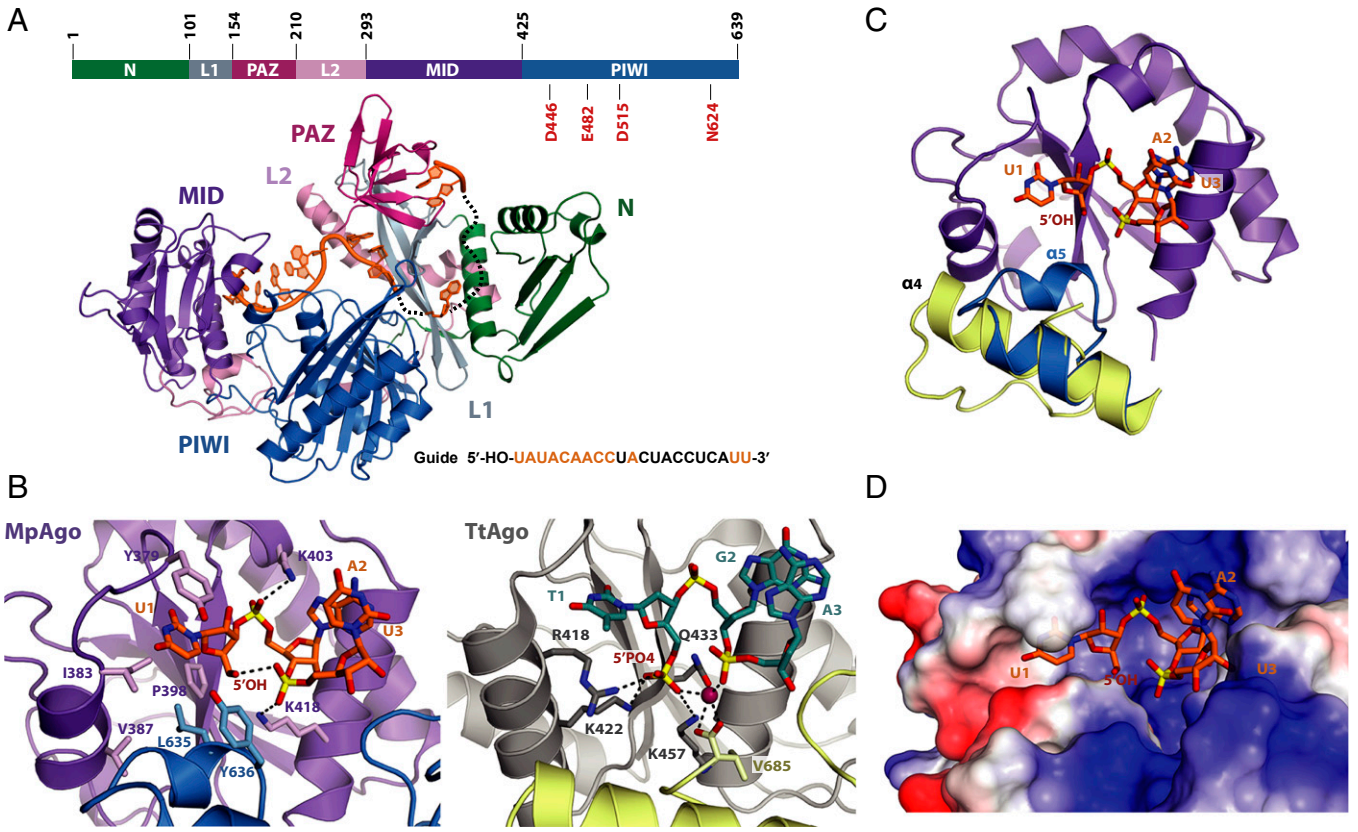


Fig. 3. Crystal structure of MpAgo and its implications for 5′-hydroxyl RNA guide binding. (*A*) (*Top*) Domain architecture of MpAgo with residues numbered. The DEDN catalytic tetrad is labeled in red. (*Middle*) MpAgo structure with N (green), Linker L1 (gray), PAZ (magenta), Linker L2 (pink), MID (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 Mg^{2+} ion (purple) along with other charged residues from the MID domain form interactions with the 5′ phosphate. (*C*) In MpAgo (blue), an additional helix (α5) 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.

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*A*). 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. 4*B*). 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*A*). Flipped nucleotide A11 is stabilized through pi-stacking with Y89 within the N domain (Fig. 4*C*). 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. 4*A*).

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. 4*D* and Fig. S4*B*). 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. S4*C*) and it remains to be determined how the kink is formed.

Catalytically Competent DEDN 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

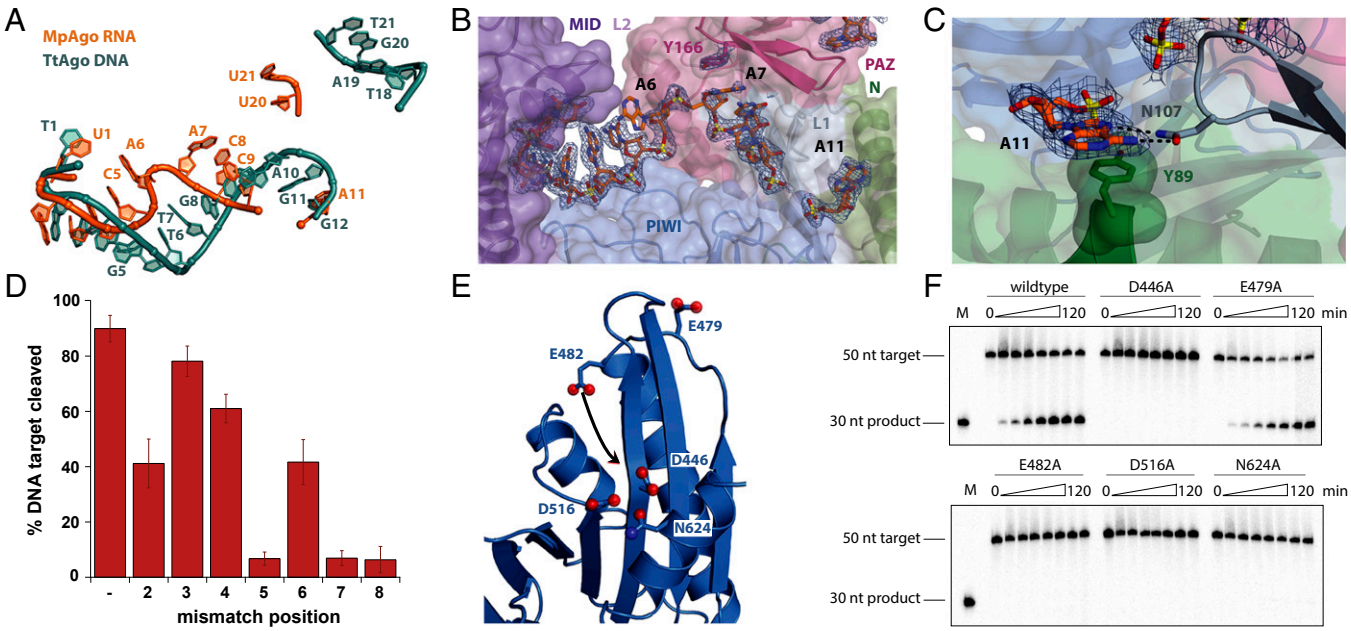


Fig. 4. Unusual guide conformation and noncanonical active site are revealed by MpAgo's crystal structure. (*A*) The structure of TtAgo (3DLH) and MpAgo were aligned using their PIWI domains. The first three nucleotides of the TtAgo DNA (sea green) and MpAgo RNA (orange) align, but then the guides begin to drastically diverge. (*B*) ($2F_{obs}$–F_{calc}) 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 pi-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. S4*C*). 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.

(13). Three of the four catalytic residues, D446, D516, and N624, are apparent in the MpAgo structure, providing the first example, to our knowledge, of an asparagine residue within the catalytic tetrad of an active Argonaute protein (Fig. 4*E*). Two residues (E479 and E482) are located close together in the linker of interest and are thus both candidates to be the glutamate finger. We introduced single alanine mutations at these five positions and tested the resulting MpAgo variants in cleavage assays to verify the catalytic residues. A single mutation at positions D446, E482, D516, or N624 abolished cleavage activity, whereas mutation of E479 had no effect, suggesting that E482 is the catalytic glutamate (Fig. 4*F*).

An Argonaute Subfamily with 5′ Hydroxyl Guide Specificity. Comparison of guide strand-containing Argonaute protein structures including hAgo2, TtAgo, and *Archaeglobus fulgidus* Piwi (AfPiwi) 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. S5*A*). 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 *Marinitoga* sp. 1155 (MsAgo) that share greater than 40% sequence identity with MpAgo (Fig. 1*A* and Fig. S5*A*). 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. S5*B*). 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 PfAgo 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 distinct from the guide preference of previously studied Ago proteins. 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 RNA. These properties hint at a previously unidentified mechanism of guide generation and MpAgo loading. To date, the biogenesis of guide RNAs or DNAs used by prokaryotic Agos is unknown.

Structural differences within the MID domain of MpAgo compared with other Ago proteins explain the observed guide RNA

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 PAZ domain disrupts the A-form 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 PAZ 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. piezophila* cells; Akshay Tambe, Nicolas Bray, and Megan Hochstrasser for helpful discussions; and J. Holton, G. Meigs, and A. Gonzalez for technical assistance with data collection 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 Program. P.J.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.

1. Swarts DC, et al. (2014) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21(9):743–753.
2. Meister G (2013) Argonaute proteins: Functional insights and emerging roles. *Nat Rev Genet* 14(7):447–459.
3. Bartel DP (2009) MicroRNAs: Target recognition and regulatory functions. *Cell* 136(2): 215–233.
4. Siomi MC, Sato K, Pezic D, Aravin AA (2011) PIWI-interacting small RNAs: The vanguard of genome defence. *Nat Rev Mol Cell Biol* 12(4):246–258.
5. Makarova KS, Wolf YI, van der Oost J, Koonin EV (2009) Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol Direct* 4(1):29.
6. Swarts DC, et al. (2015) Argonaute of the archaeon Pyrococcus furiosus is a DNA-guided nuclease that targets cognate DNA. *Nucleic Acids Res* 43(10):5120–5129.
7. Swarts DC, et al. (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507(7491):258–261.
8. Olovnikov I, Chan K, Sachidanandam R, Newman DK, Aravin AA (2013) Bacterial argonaute samples the transcriptome to identify foreign DNA. *Mol Cell* 51(5):594–605.
9. Song J-J, Smith SK, Hannon GJ, Joshua-Tor L (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305(5689):1434–1437.
10. Parker JS, Roe SM, Barford D (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434(7033):663–666.
11. Rashid UJ, et al. (2007) Structure of Aquifex aeolicus argonaute highlights conformational flexibility of the PAZ domain as a potential regulator of RNA-induced silencing complex function. *J Biol Chem* 282(18):13824–13832.
12. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ (2008) Structure of the guide-strand-containing argonaute silencing complex. *Nature* 456(7219):209–213.
13. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. *Nature* 486(7403):368–374.
14. Elkayam E, et al. (2012) The structure of human argonaute-2 in complex with miR-20a. *Cell* 150(1):100–110.
15. Schirle NT, MacRae IJ (2012) The crystal structure of human Argonaute2. *Science* 336(6084):1037–1040.
16. Ma JB, et al. (2005) Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. *Nature* 434(7033):666–670.
17. Frank F, Sonenberg N, Nagar B (2010) Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465(7299):818–822.
18. Yosef I, Goren MG, Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* 40(12):5569–5576.
19. Nuñez JK, Lee ASY, Engelman A, Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519(7542):193–198.
20. Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* 22(24): 3489–3496.
21. Haurwitz RE, Jinek M, Wiedenheft B, Zhou K, Doudna JA (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* 329(5997): 1355–1358.
22. Deltcheva E, et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471(7340):602–607.
23. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709–1712.
24. Garneau JE, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67–71.
25. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* 1:7.
26. Burroughs AM, Ando Y, Aravind L (2014) New perspectives on the diversification of the RNA interference system: Insights from comparative genomics and small RNA sequencing. *Wiley Interdiscip Rev RNA* 5(2):141–181.
27. Makarova KS, et al. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467–477.
28. Rivas FV, et al. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat Struct Mol Biol* 12(4):340–349.
29. Sheng G, et al. (2014) Structure-based cleavage mechanism of Thermus thermophilus Argonaute DNA guide strand-mediated DNA target cleavage. *Proc Natl Acad Sci USA* 111(2):652–657.
30. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W (2005) Crystal structures of RNase H bound to an RNA/DNA hybrid: Substrate specificity and metal-dependent catalysis. *Cell* 121(7):1005–1016.
31. Wilson RC, Doudna JA (2013) Molecular mechanisms of RNA interference. *Annu Rev Biophys* 42:217–239.
32. Willkomm S, Zander A, Gust A, Grohmann D (2015) A prokaryotic twist on argonaute function. *Life (Basel)* 5(1):538–553.
33. Burroughs AM, Iyer LM, Aravind L (2013) Two novel PIWI families: Roles in inter-genomic conflicts in bacteria and Mediator-dependent modulation of transcription in eukaryotes. *Biol Direct* 8:13.
34. Mi S, et al. (2008) Sorting of small RNAs into Arabidopsis argonaute complexes is directed by the 5′ terminal nucleotide. *Cell* 133(1):116–127.
35. Brennecke J, Stark A, Russell RB, Cohen SM (2005) Principles of microRNA-target recognition. *PLoS Biol* 3(3):e85.
36. Jo MH, et al. (2015) Human Argonaute 2 has diverse reaction pathways on target RNAs. *Mol Cell* 59(1):117–124.
37. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D (2009) Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol Cell* 33(2):204–214.
38. Schirle NT, Sheu-Gruttadauria J, MacRae IJ (2014) Structural basis for microRNA targeting. *Science* 346(6209):608–613.
39. Kranzusch PJ, et al. (2014) Structure-guided reprogramming of human cGAS dinucleotide linkage specificity. *Cell* 158(5):1011–1021.
40. Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
41. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
42. Terwilliger TC (1999) Reciprocal-space solvent flattening. *Acta Crystallogr D Biol Crystallogr* 55(Pt 11):1863–1871.
43. Emsley P, Cowtan K (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
44. Karplus PA, Diederichs K (2012) Linking crystallographic model and data quality. *Science* 336(6084):1030–1033.
45. Tatusova T, et al. (2015) Update on RefSeq microbial genomes resources. *Nucleic Acids Res* 43(Database issue):D599–D605.
46. Pei J, Kim B-H, Grishin NV (2008) PROMALS3D: A tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36(7):2295–2300.
47. Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2256–2268.
48. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25(9):1189–1191.

Supporting Information

Kaya et al. 10.1073/pnas.1524385113

SI Materials and Methods

Cloning and Purification of MpAgo and TpAgo. The sequence encoding MpAgo was codon-optimized for expression in *E. coli* and cloned into a custom pET-based expression vector using ligation-independent cloning. The cloned construct encodes a fusion protein containing an N-terminal His_{10}-tag followed by an Asn_{10}-linker, Maltose Binding Protein (MBP), and a PreScission protease (MpAgo) and TEV (TpAgo) cleavage site. Mutations were introduced using QuikChange site-directed mutagenesis and verified by DNA sequencing.

The WT and mutant proteins were expressed in *E. coli* strain BL21(DE3) (New England Biolabs). For protein expression, cells were grown in LB medium to an OD_{600} of 0.8, expression was induced by addition of IPTG (isopropyl β-D-1-thiogalactopyranoside) to 0.4 mM final concentration, and cells were incubated at 16 °C while shaking for 16 h. The cell pellets were resuspended in 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 1 mM TCEP [Tris(2-carboxyethyl)phosphine], 5% (vol/vol) glycerol, 0.5% Triton X-100, and 10 mM imidazole, and supplemented with Complete protease inhibitor mixture (Roche). Cells were lysed via sonication, and clarified lysate was bound in batch to Ni-NTA agarose (Qiagen). The resin was washed extensively with 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 1 mM TCEP, 5% (vol/vol) glycerol, and 10 mM imidazole, and bound protein was eluted in wash buffer containing 300 mM imidazole. The His_{10}-MBP affinity tag was removed by cleavage with PreScission protease while the protein was dialyzed overnight at 4 °C against 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 1 mM TCEP, 5% (vol/vol) glycerol, and 10 mM imidazole. The cleaved MpAgo protein was separated from the fusion tag using a second Ni-NTA affinity step. The protein was dialyzed into 50 mM Tris·HCl, pH 7.5, 150 mM NaCl, 1 mM TCEP, and 5% (vol/vol) glycerol, applied to a 5-mL Heparin HiTrap column (GE Life Sciences), and eluted with a linear gradient of 0.15–1.2 M NaCl. Final purification was achieved by size exclusion chromatography using a HiLoad 16/60 Superdex 200 column (GE Life Sciences) in 50 mM Tris·HCl, pH 7.5, 300 mM NaCl, 1 mM TCEP, and 5% (vol/vol) glycerol. Eluted protein was concentrated, flash-frozen in liquid nitrogen, and stored at −80 °C.

RNA in Vitro Transcription and Oligonucleotide Purification. RNAs longer than 29 nt were in vitro transcribed using T7 RNA polymerase and synthetic DNA templates carrying a T7 promoter sequence. All DNA and short RNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). All synthetic and in vitro transcribed oligonucleotides used for cleavage assays and crystallography experiments were gel-purified and quality checked before use (RNA and DNA oligonucleotides used in this study are listed in Table S2). Double-stranded DNA oligonucleotides were hybridized in 20 mM Tris·HCl, pH 7.5, 100 mM KCl, and 5 mM $MgCl_2$, incubated at 95 °C for 30 s, slow-cooled at room temperature, and gel purified via native PAGE.

In Vitro Cleavage Assays. Purified oligonucleotide (10 pmol) was radiolabeled using T4 polynucleotide kinase (NEB) and [γ-^{32}P] ATP (Perkin-Elmer) in 1× T4 PNK buffer (NEB) at 37 °C for 30 min. The T4 PNK was heat inactivated at 65 °C for 20 min. The labeling reactions were purified with illustra MicroSpin G-25 columns (GE Life Sciences). For single turnover experiments, MpAgo–RNA complexes were reconstituted by mixing 1 nM MpAgo with a 1-nM guide strand in 10 mM Tris·HCl, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$ or $MnSO_4$, 2 mM DTT, and 5% (vol/vol) glycerol and incubating at 37 °C for 30 min. Cleavage reactions were initiated by addition of 0.1 nM radiolabeled DNA or RNA substrates and performed at 60 °C; 10-μL aliquots were removed at various time points and quenched by mixing with an equal volume of formamide gel loading buffer supplemented with 50 mM EDTA. Cleavage products were resolved by 12% denaturing PAGE and visualized by phosphorimaging. Cleavage experiments were tested in three independent experiments. Percentage of cleavage was analyzed by densitometry using ImageQuant (GE Healthcare), and the average of three independent experiments was plotted against time (Kaleidagraph 4.1). The data were fit by nonlinear regression analysis, and the cleavage rates were calculated. Single turnover experiments with TpAgo were performed under the similar conditions, with the difference that 2 nM TpAgo was reconstituted with a 2-nM guide strand followed by addition of 0.1 nM radiolabeled target.

MpAgo Immunoprecipitation. An antibody was generated by immunizing a rabbit with 5 mg recombinant MpAgo (Josman, LLC). Clarified rabbit serum was tested for desired antibody production by both Western blot analysis and immunoprecipitation using *E. coli* BL21(DE3) cells expressing recombinant MpAgo, as well as detection of endogenous MpAgo in the host cell. For Western blot analysis, Zirconia beads were added to *E. coli* cells resuspended in PBS-T and lysed by vortexing for 10 min at 4 °C. After removing cell debris by centrifugation, total protein concentration of the cleared lysate was determined via Bradford. Twenty micrograms total protein was boiled in SDS-sample buffer and resolved by SDS/PAGE before transfer to a PVDF membrane. Protein G purified anti-MpAgo rabbit IgG (50 mg/mL IgG concentration) was diluted in blocking buffer and incubated with the membrane overnight at 4 °C. After washing with PBS-T, anti-rabbit antibody HRP conjugate was diluted in blocking buffer and incubated with the membrane for 1 h at room temperature. The Western blot was developed by adding Amersham ECL substrate (GE Life Sciences) and visualized via chemiluminescence in a gel imager (Bio-Rad Universal Hood II). The purified anti-MpAgo rabbit IgG was further tested for immunoprecipitation of the folded protein. His_{10}-MBP–tagged MpAgo was expressed in a 10 mL BL21(DE3) culture as described above; 50 μg anti-MpAgo rabbit IgG was diluted in PBS and incubated with 20-μL Protein A DynaBeads (Life Technologies) for 15 min at 4 °C while rocking. DynaBeads were washed extensively to remove unbound IgG before addition of His_{10}-MBP–tagged MpAgo-containing lysate. After binding for 1 h at room temperature, Dynabeads were washed with PBS-T, and bound protein was eluted in 50 mM glycine, pH 2.8. Eluted fractions were immediately brought to neutral pH using 1 M Tris·HCl, pH 9, and samples were analyzed via SDS/PAGE.

Selenomethionine Protein Expression and Purification. Selenomethionine-substituted protein was prepared as previously described (39). Briefly, cultures were grown in M9 minimal media supplemented with 1 mM $MgSO_4$, 100 μM $CaCl_2$, 1 μg/mL thiamine, and 0.4% glucose. After growth at 37 °C to an OD_{600} of ∼0.5, cultures were supplemented with leucine, isoleucine, and valine at 50 mg/L; phenylalanine, lysine, and threonine at 100 mg/L; and selenomethionine (Acros Organics) at 75 mg/L. Cultures were grown at 37 °C for an additional 20 min, cooled at 4 °C for 15 min, induced with IPTG at a final concentration of 0.4 mM, and incubated while shaking at 16 °C for ∼18 h. Selenomethionine-substituted protein was purified using identical conditions as native protein.

Crystallization and Structure Determination. Crystals were obtained by hanging-drop vapor diffusion at 18 °C. Approximately 4 mg/mL purified MpAgo was incubated with an equimolar amount of RNA

on ice for 30 min. Original crystallization conditions were identified by sparse-matrix screen using 400-nL drops set over 70-μL reservoir solutions in a 96-well format. Final crystals for data collection were grown in Easy-Xtal 15-well trays (QIAGEN) in 2-μL drops with a 1:1 ratio of protein and reservoir solution. Optimal crystallization conditions were 0.1 M Hepes-NaOH, pH 7.5, and 27% (wt/vol) PEG 2000.

Crystals were transferred with a nylon loop to a new drop-containing reservoir solution supplemented with 20% (vol/vol) ethylene glycol as a cryoprotectant and incubated for ∼30 s before flash freezing in liquid nitrogen. Selenomethionine crystals were grown and harvested using similar conditions and optimized by microseeding crushed native crystals. Native and anomalous data were collected under cryogenic conditions at the Lawrence Berkeley National Laboratory Advanced Light Source (Beamline 8.3.1).

X-ray diffraction data were processed with XDS and AIMLESS (40) using the SSRL *autoxds* script (A. Gonzalez, SSRL, Stanford, CA). Indexed crystals belonged to the space group *P* 1 with two copies of MpAgo in the asymmetric unit. A single crystal was used to collect two datasets at an energy midway between the peak/inflection selenium absorbance energy (λ = 0.97964) and at a high-energy remote (λ = 0.97197). Two rounds of collection were used, one at a low dosage and the other at a 4× dosage. Data were collected in inverse beam mode with consecutive collection at each energy before crystal rotation. Initial phase determination was performed using Multiple Anomalous Dispersion methods, and five selenium sites were identified with HySS and PHENIX (41). An initial map was calculated using SOLVE/RESOLVE (42) that allowed placement of a homology PIWI domain model and manual building of partial domains and poly-alanine sequence. This partial structure was used as a search model to extend phases with native data to 2.0 Å, and iterative rounds of manual building in Coot (43) and refinement in PHENIX allowed model completion (Table S3). X-ray data for refinement were extended an additional ∼0.1 Å beyond an I/σ resolution cutoff of ∼2.0, as determined by CC* correlation, R_{pim} parameters, and visual inspection of the resulting map (Fig. S3*D*) (44).

Sequence Alignments of MID and PIWI Domains. Sequences from *Homo sapiens* Ago 2 (hAgo2, NP_036286), *Archaeoglobus fulgidus* PIWI protein (AfPiwi, WP_048064369), *Thermus thermophilus* Ago (TtAgo, WP_011174533), *Rhodobacter sphaeroides* Ago (RsAgo, WP_044249552), *Marinitoga piezophila* Ago (MpAgo, WP_014295921), *Thermotoga profunda* Ago (TpAgo, WP_041081268), and *Marinitoga* sp. 1155 Ago (MsAgo, WP_047265940) were collected from NCBI's Refseq protein database (45) and aligned using Promals3D (46). Structural constraints resultant from a pairwise structural alignment of MpAgo and TtAgo using PDBeFold (47) with the default settings and structures of MpAgo and TtAgo (4N41) were used as input. The alignment was generated in JalView (48) using the ClustalX coloring scheme. Corresponding secondary structural elements were determined from the structures of hAgo2 (4W5N) and MpAgo.

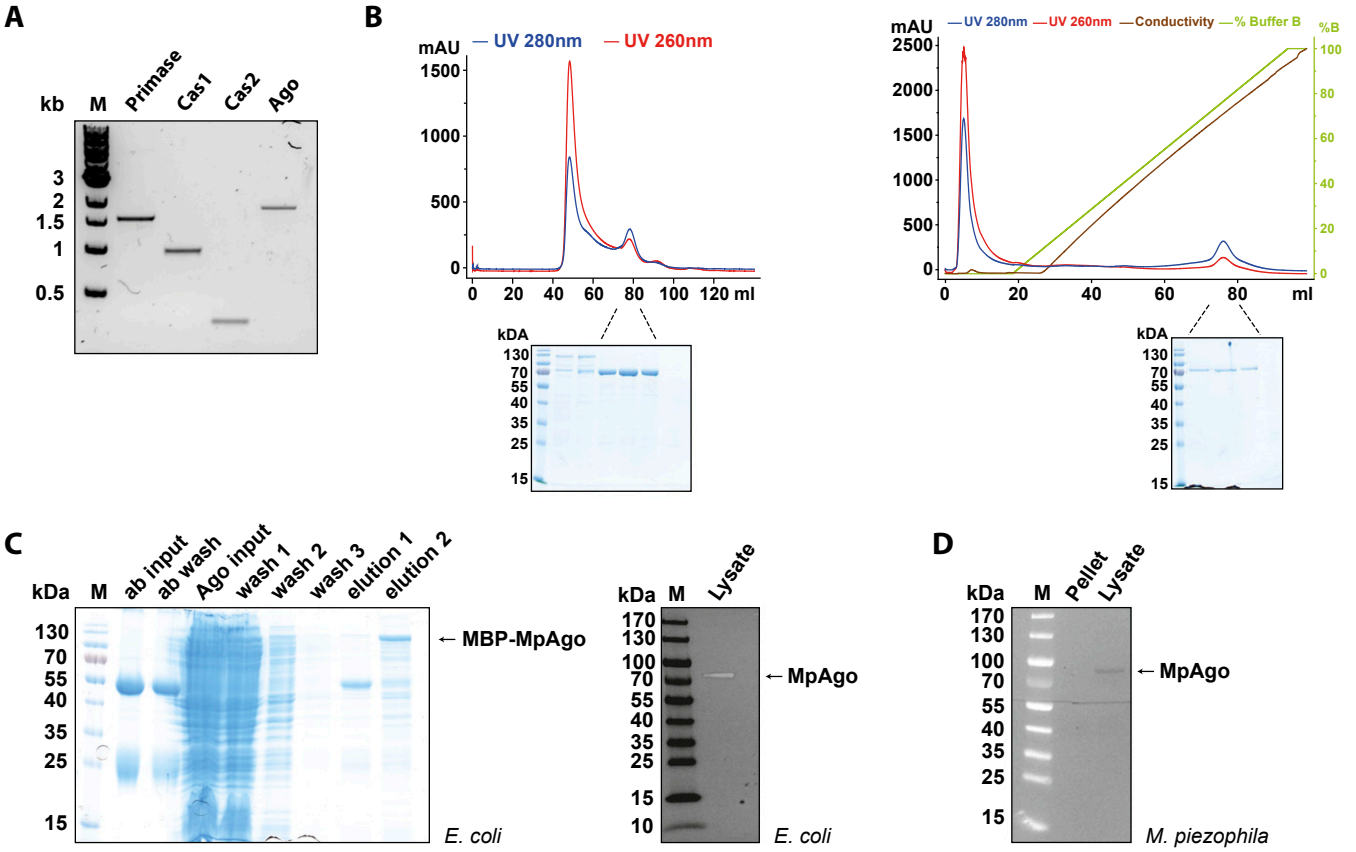


Fig. S1. MpAgo is cotranscribed with Cas1 and Cas2 and is detectable under native conditions. (*A*) RT-PCR of the acquisition operon using *M. piezophila* cDNA. Forward and reverse primers annealing to the ends of the individual genes of *primase*, *cas1*, *cas2*, and *ago* were used to amplify cDNA products. (*B*) (*Left*) Size exclusion chromatogram and SDS/PAGE. WT *MpAgo* was purified using a Superdex 200 16/60 (GE Life Sciences) size exclusion column. (*Right*) Ion exchange chromatogram and SDS/PAGE. To separate recombinant MpAgo from copurifying nucleic acid, the protein was further purified using a HiTrap Heparin (GE Life Sciences) ion exchange column. (*C*) (*Left*) Immunoprecipitation of recombinantly expressed His_{10}-MBP MpAgo in *E. coli* using immobilized polyclonal antibody (ab) against MpAgo. (*Right*) Western blot analysis to test polyclonal antibody for its specificity against heterologously expressed MpAgo in *E. coli* cell lysate. (*D*) Western Blot visualizing the presence of endogenous MpAgo using lysate of *M. piezophila*. Soluble MpAgo could be detected in the native organism.

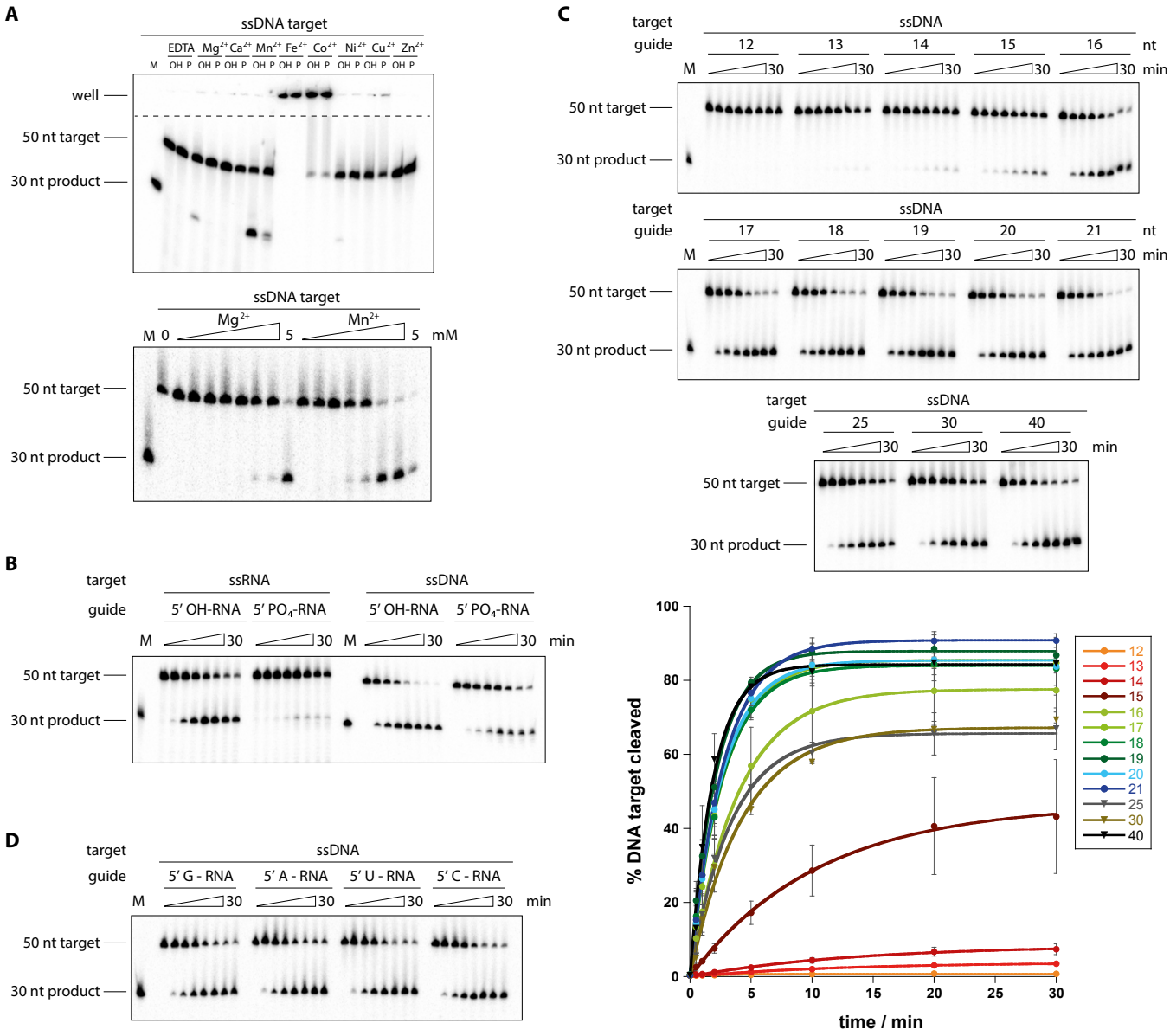


Fig. S2. Metal dependency of MpAgo and time course cleavage experiments. (*A*) (*Upper*) Dependency on divalent metal ions. ssDNA cleavage assays were performed using 5′-hydroxylated (OH) and 5′-phosphorylated (p) RNA guides in the presence of 2 mM divalent cations or EDTA. Addition of Fe^{2+} and Co^{2+} caused aggregation of the reaction mixture and resulted in retention of nucleic acids in the wells. (*Lower*) Comparison of cleavage efficiencies in presence of increasing $MnCl_2$ and $MgCl_2$ concentrations. ssDNA cleavage assays were performed in presence of 0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, and 5 mM $MgCl_2$ or $MnCl_2$. Reactions were incubated at 60 °C for 2 h and resolved on a 12% denaturing PAGE before visualizing on a phosphoimager. (*B*) Representative denaturing PAGE showing the results of DNA and RNA cleavage kinetics in the presence of 2.5 mM $MnCl_2$. Assays were performed in three independent replicates, and time points were taken at 0, 0.5, 1, 2, 5, 10, 20, and 30 min. Cleavage efficiencies were quantified and plotted against time (Fig. 2*B*). (*C*) Representative denaturing PAGEs (one of three independent experiments) showing MpAgo cleavage kinetics using RNA guides of different lengths. Cleavage efficiencies and SDs were obtained from three independent experiments and plotted against time. (*D*) Representative denaturing PAGE (one of three independent experiments) showing MpAgo cleavage kinetics with different nucleotides at the 5′ end of the guide and their respective ssDNA targets. Cleavage efficiencies from three independent experiments were quantified and plotted against time (Fig. 2*D*).

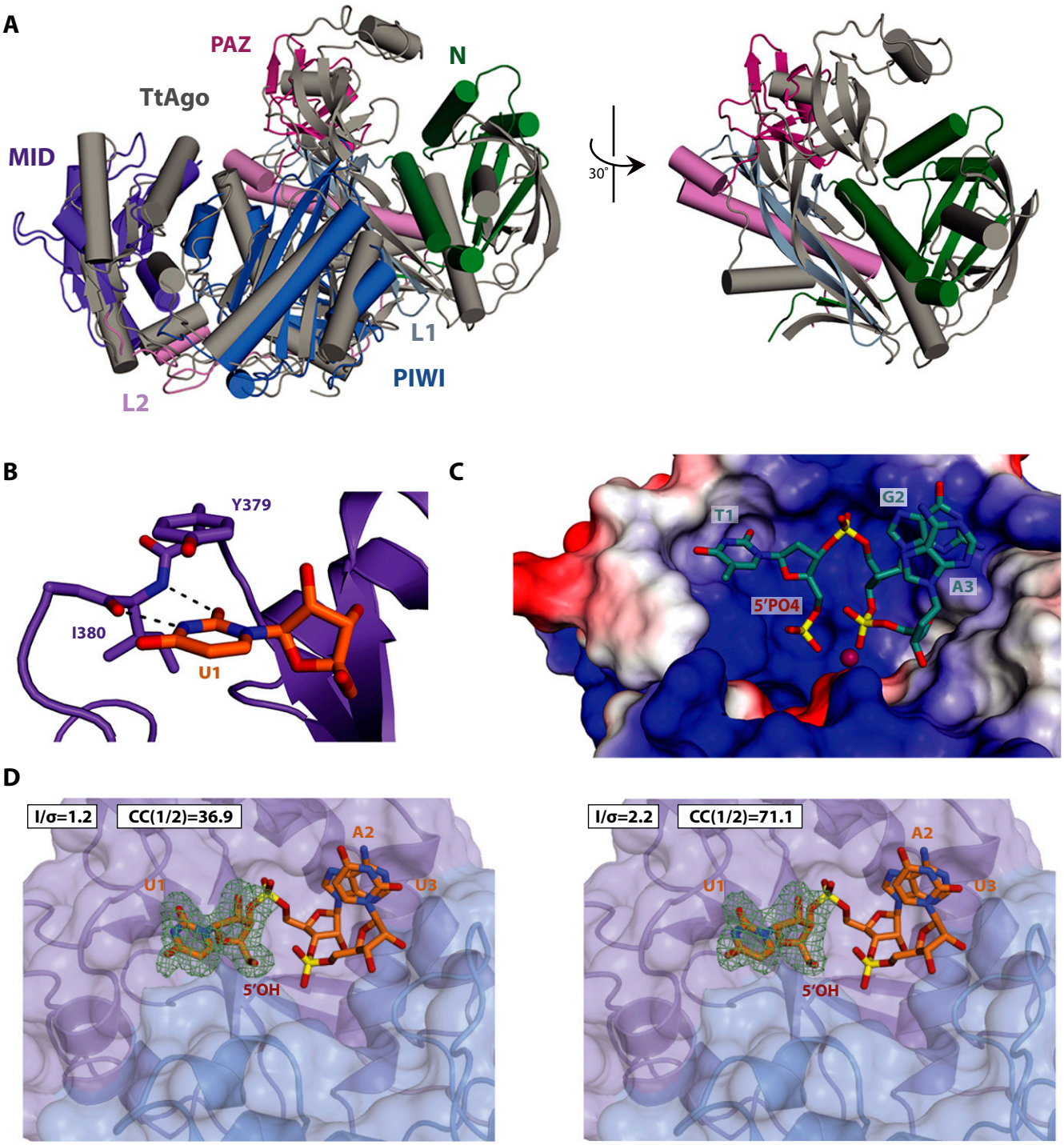


Fig. S3. Structural comparison of MpAgo and TtAgo. (*A*) TtAgo (3DLH, gray) and MpAgo (colored by domains) were aligned using their PIWI domains. The PAZ domains differ in size, whereas the N domains diverge in orientation relative to the MID-PIWI lobe. (*B*) The peptide backbone of the nucleotide preference loop within the MID domain (purple) forms H-bonds, depicted in black dotted lines, with U1 of the guide RNA, suggesting base specific interactions. (*C*) An electrostatic heat map was generated in PyMOL and visualized via a surface representation for TtAgo. The absence of α5 opens up the MID binding pocket, creating space for a 5′ phosphate. The ubiquitous blue surface around the binding pocket denotes the presence of several charged residues available for interaction with the 5′ phosphate of the guide strand. (*D*) A simulated annealing omit map (F_{obs}–F_{calc}) contoured at 3.5σ (green), generated for the first nucleotide of the guide strand. The MpAgo structure was refined to 2.0 Å according to CC*, R_{pim} cut-offs and visual inspection of the resulting map (*Left*; statistics in Table S1) (44), and placement of all major elements are supported using a more conservative cutoff of 2.1 Å (*Right*; with high-resolution shell statistics of completeness = 84.4%, I/σ = 2.2, R_{pim} = 35.2).

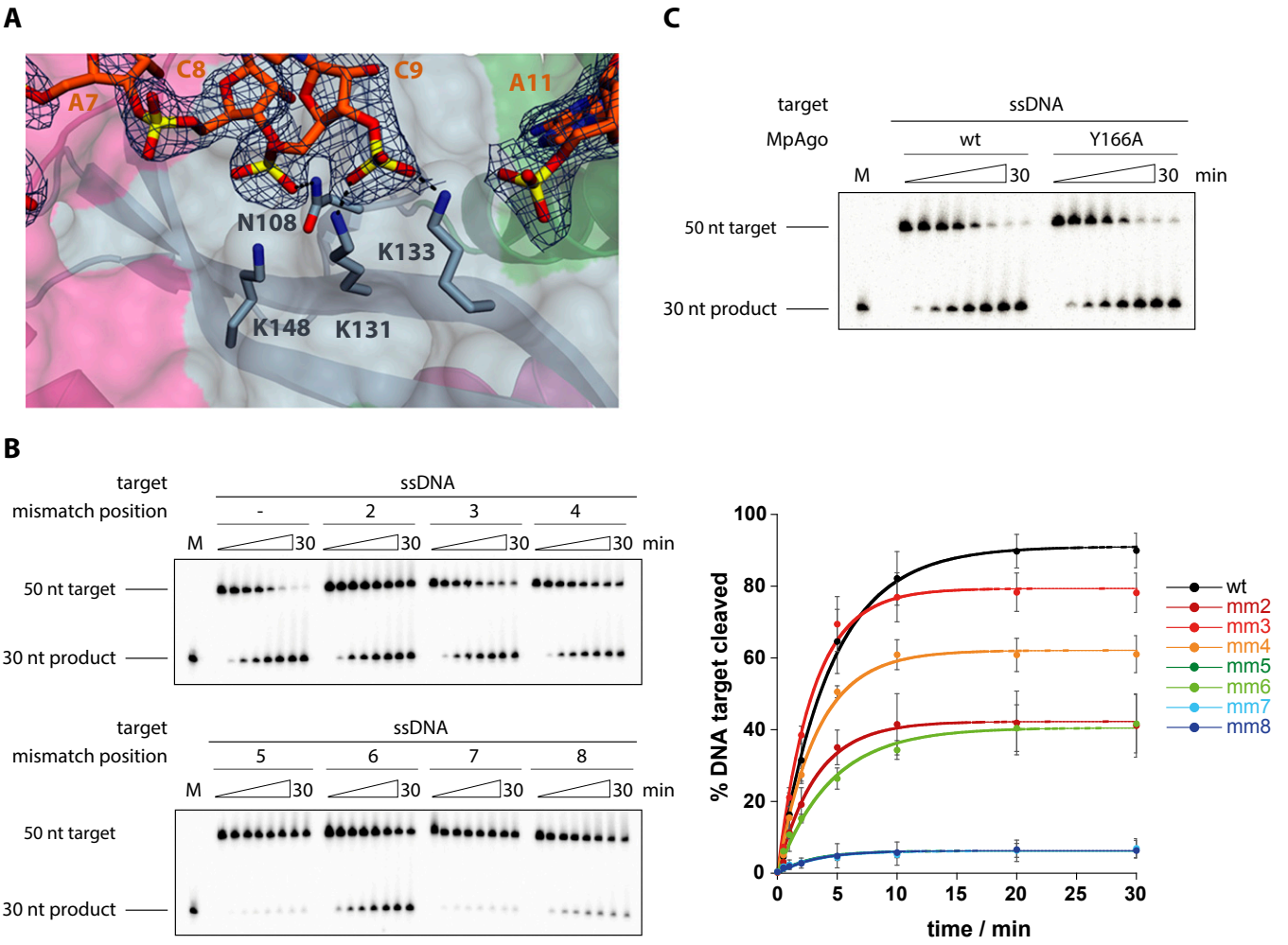


Fig. S4. Structural basis for noncanonical guide-RNA conformation. (*A*) Although the bases of C8 and C9 are unresolved, the ($2F_{obs}–F_{calc}$) electron density map, contoured at 1σ, of the ribose and phosphate backbone shows strong interactions with charged residues from the L1 linker (gray). (*B*) Representative denaturing PAGE (one of three independent experiments) showing MpAgo cleavage kinetics using ssDNA targets bearing single nucleotide mismatches at positions 2–8. Cleavage efficiencies were quantified and observed maximum cleavage is shown for each single nucleotide mismatch (Fig. 4*D*). Error bars represent SD from three independent experiments. (*C*) The protruding Tyrosine 166 was mutated to a smaller residue (alanine) and tested for cleavage activity in a time course experiment.

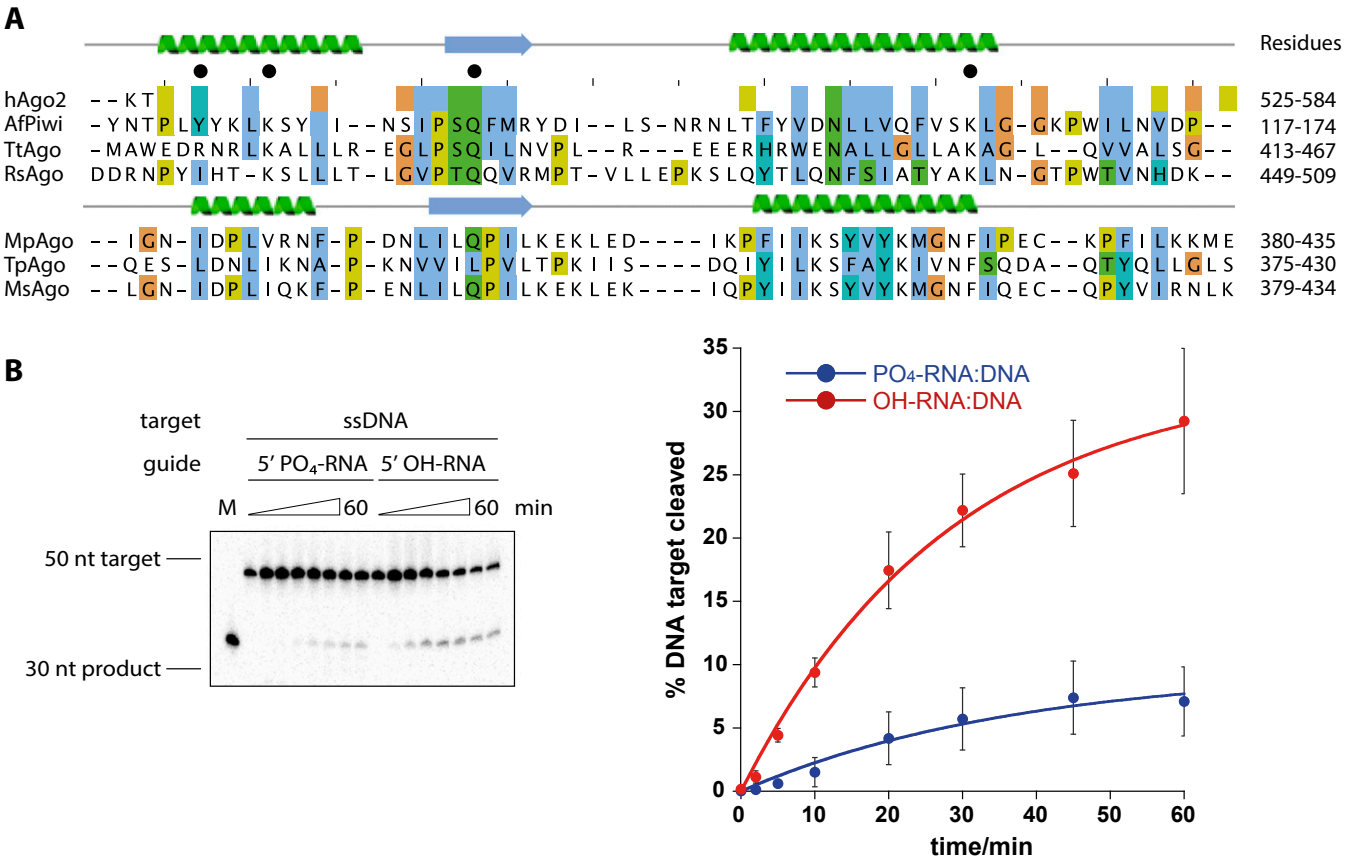


Fig. S5. Structure based sequence alignment of the MID and PIWI domain. (*A*) Structure-based sequence alignment of the 5′ end binding pocket. Residues coordinating the 5′ phosphate in the TtAgo structure are highlighted with a black dot. Experimentally determined secondary structures for human Ago2 are shown on top and are conserved for AfPiwi, TtAgo, and RsAgo; the experimentally determined secondary structures for MpAgo, which are also predicted to be present in TpAgo and MsAgo, are shown in the middle. Residues are color-coded using Clustal X color scheme. (*B*) In vitro ssDNA cleavage of TpAgo programmed with either a 5′-hydroxylated or 5′-phosphorylated 21 nt RNA guide.

Table S1. List of *M. piezophila* subtype III-B *cas* genes

Accession number	Gene product
AEX84842.1	Cas6
AEX84843.1–AEX84849.1	Cmr1-7
AEX84846.1	Predicted primase
AEX84847.1	Cas1
AEX84848.1	Cas2
AEX84849.1	MpAgo

Table S2. List of nucleic acid sequences used in this study

No.	Purpose	Sequence 5′-3′
1	Oligo for preparing double-stranded T7 promoters for in vitro transcription	TAATACGACTCACTATA
2	ssDNA template for transcribing RNA target	GACCTCCTGATATACAACCTACTACCTCATTAATATCTAGATCATAGT**TATAGTGAGTCGTATTA**
3	ssRNA target	GGACUAUGAUCUAGAUAUUAAUGAGGUAGUAGGUUGUAUAUCAGGAGGUC
4	50 nt ssDNA target	GGACTATGATCTAGATATTAATGAGGTAGTAGGTTGTAT**N**TCAGGAGGTC
5	50 nt ssDNA nontarget (reverse complement of 4)	GACCTCCTGATATACAACCTACTACCTCATTAATATCTAGATCATAGTCC
6	30 nt ssDNA marker	GGACTATGATCTAGATATTAATGAGGTAGT
7	37 nt ssDNA marker	GGATCCTTTTATGTTGTAAGTTATGATATTACAAATG
8	ssDNA template for transcribing RNA marker	ACTACCTCATTAATATCTAGATCATAGT**TATAGTGAGTCGTATTA**
9	30 nt RNA marker	GGACUAUGAUCUAGAUAUUAAUGAGGUAGU
10	40 nt guide RNA	UAUACAACCUACUACCUCAUUAAUAUCUAGAUCAUAGUCC
11	30 nt guide RNA	UAUACAACCUACUACCUCAUUAAUAUCUAG
12	25 nt guide RNA	UAUACAACCUACUACCUCAUUAAUA
13	21 nt guide RNA	**N**AUACAACCUACUACCUCAUU
14	21 nt guide DNA	TATACAACCTACTACCTCATT
15	20 nt guide RNA	UAUACAACCUACUACCUCAU
16	19 nt guide RNA	UAUACAACCUACUACCUCA
17	18 nt guide RNA	UAUACAACCUACUACCUC
18	17 nt guide RNA	UAUACAACCUACUACCU
19	16 nt guide RNA	UAUACAACCUACUACC
20	15 nt guide RNA	UAUACAACCUACUAC
21	14 nt guide RNA	UAUACAACCUACUA
22	13 nt guide RNA	UAUACAACCUACU
23	12 nt guide RNA	UAUACAACCUAC GTTGTATC
24	50 nt ssDNA target, mismatch at position 2	GGACTATGATCTAGATATTAATGAGGTAGTAGGTTGTAACTCAGGAGGTC
25	50 nt ssDNA target, mismatch at position 3	GGACTATGATCTAGATATTAATGAGGTAGTAGGTTGTTTCTCAGGAGGTC
26	50 nt ssDNA target, mismatch at position 4	GGACTATGATCTAGATATTAATGAGGTAGTAGGTTGAATCTCAGGAGGTC
27	50 nt ssDNA target, mismatch at position 5	GGACTATGATCTAGATATTAATGAGGTAGTAGGTTCTATCTCAGGAGGTC
28	50 nt ssDNA target, mismatch at position 6	GGACTATGATCTAGATATTAATGAGGTAGTAGGTAGTATCTCAGGAGGTC
29	50 nt ssDNA target, mismatch at position 7	GGACTATGATCTAGATATTAATGAGGTAGTAGGATGTATCTCAGGAGGTC
30	50 nt ssDNA target, mismatch at position 8	GGACTATGATCTAGATATTAATGAGGTAGTAGCTTGTATCTCAGGAGGTC

The reverse complement of the T7 promoter is indicated in bold. **N**, permutations of all four nucleotides at this position.

Table S3. Data collection and refinement statistics

		MpAgo (SeMet)	
Data collection	MpAgo	Inflection/peak	Remote
Space group	*P* 1	*P* 1	
Cell dimensions			
a, b, c (Å)	50.88, 75.42, 102.21	50.75, 75.44, 101.71	
α, β, γ (°)	110.87, 90.55, 90.09	110.72, 89.94, 90.09	
Resolution (Å)	48.33–1.95 (1.98–1.95)*	48.98–2.99 (3.18–2.99)*	
R_{merge} (%)	8.2 (113.6)	11.5 (88.9)	11.5 (90.0)
R_{pim} (%)	4.8 (69.6)	4.9 (38.1)	4.9 (38.6)
I/σ	10.6 (1.2)	15.0 (2.6)	14.9 (2.5)
$CC_{1/2}$	99.8 (36.9)	99.7 (66.2)	99.7 (67.3)
Completeness (%)	96.7 (82.1)	98.5 (95.1)	98.5 (95.0)
Redundancy	3.9 (3.5)	6.5 (6.3)	6.4 (6.3)
Refinement			
Resolution (Å)	48.33–1.95		
No. reflections	100,295 (4,224)		
R_{work}/R_{free}	21.6/23.9		
No. atoms			
Protein	10,664		
RNA	437		
Water	625		
Average B-factors (Å^2)			
Protein	42.9		
RNA	58.8		
Water	43.6		
RMS deviations			
Bond lengths (Å)	0.003		
Bond angles (°)	0.578		
Ramachandran			
Favored (%)	96.4		
Allowed (%)	3.4		
Outliers (%)	0.2		

Single crystals were used to collect native and anomalous datasets.
*Highest resolution shell is shown in parentheses.