BIOLOGICAL SCIENCES – Biochemistry

Programmable RNA recognition using a CRISPR-associated Argonaute

Audrone Lapinaite[a], Jennifer A. Doudna[a,b,c,d,e] and Jamie H. D. Cate[a,b,e,1]

[a]Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720; [b]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720; [c]Howard Hughes Medical Institute, University of California, Berkeley, CA 94720; [d]Innovative Genomics Institute, University of California, Berkeley, CA 94720; and [e]Department of Chemistry, University of California, Berkeley, CA 94720

[1]Corresponding author. Email: jcate@lbl.gov

KEYWORDS. Argonaute; CRISPR; programmable RNA recognition; small noncoding RNA; RNA editing; inosine

ABSTRACT

Argonaute proteins (Agos) are present in all domains of life. While the physiological function of eukaryotic Agos in regulating gene expression is well documented, the biological roles of many of their prokaryotic counterparts remain enigmatic. In some bacteria, Agos are associated with CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) loci and use non-canonical 5'-hydroxyled guide RNAs (gRNAs) for nucleic acid targeting. Here we show that *in vitro* reconstituted CRISPR-associated *Marinitoga piezophila* Argonaute-gRNA complexes (MpAgo RNPs) are destabilized upon binding to highly complementary RNA substrates. Quantitative and site-specific covalent crosslinking of the gRNA to MpAgo stabilizes the MpAgo RNP and significantly improves its specificity and affinity for RNA targets. Using crosslinked MpAgo RNPs, we mapped the seed region of the gRNA that contributes to specific RNA substrate binding, and identified the nucleotides of the gRNA that play the most significant role in targeting specificity. We also show that crosslinked MpAgo RNPs can be programmed to distinguish between substrates that differ by a single nucleotide, using permutations at the 6th and 7th positions in the gRNA. Using these specificity features, we employed MpAgo RNPs to detect specific Adenosine to Inosine edited RNAs in a complex mixture. These findings broaden our mechanistic understanding of the interactions of Argonautes with guide and substrate RNAs, and demonstrate that crosslinked MpAgo RNPs can be used as a highly-specific RNA-targeting platform to probe RNA biology.

SIGNIFICANCE

Argonaute proteins are present in bacteria, archaea and eukaryotes. They play an important role in a wide range of biological processes, from transcriptional and translational gene expression regulation to defense against viruses and silencing of mobile genetic elements. Here we present mechanistic insights into the interactions of the CRISPR-associated *Marinitoga piezophila* Argonaute (MpAgo) with its guide RNA (gRNA) and RNA substrates. By engineering site-specific covalent crosslinking of the gRNA to MpAgo, we demonstrate that the crosslinked MpAgo RNP is easily programmable, has high affinity to fully complementary RNA substrates, and can discriminate between substrates that differ by only a single nucleotide. These crosslinked MpAgo RNPs should be useful for probing endogenous RNAs in living cells.

INTRODUCTION

Argonautes are nucleic acid-guided proteins present in organisms from all three domains of life: bacteria, archaea and eukaryotes (1). In eukaryotes, Argonautes (eAgos) play central roles in RNA interference (RNAi) and micro-RNA (miRNA) pathways that coordinate a wide range of cellular processes including transcriptional and translational gene regulation (2, 3), silencing of mobile genetic elements (e.g. transposons) (4, 5), and host defense (6). eAgos bind single stranded RNAs, either small interfering RNAs (siRNAs) or miRNAs, which act as templates for the recognition of complementary single-stranded RNA (ssRNA) target sites, leading either to cleavage of the targeted RNA by catalytically active Ago (7) or to the recruitment of additional components of the RNA degradation machinery (9–11).

Although prokaryotes lack RNAi pathways (12), prokaryotic Argonautes (pAgos) are thought to contribute to host defense against foreign DNA (8). Despite their structural similarity to eAgos–including a bi-lobed architecture that binds the 5'- and 3'-ends of guides (**Fig. 1A**), pAgos are more diverse in their biochemical behavior than their eukaryotic counterparts. For example, some pAgos can use either DNA or RNA guides, and can target either DNA or RNA substrates (8, 13–15). Differing from eAgos and other pAgo family members, the CRISPR locus-associated pAgo subfamily found in *Marinitoga piezophila*, *Thermotoga profunda* and *Marinitoga sp. 1155)* (16) binds 5'-hydroxylated (5'-OH) guide RNAs (gRNAs) rather than 5'-phosphorylated guides (**Fig. 1A**). Furthermore, the *Marinitoga piezophila* Ago (MpAgo) cleaves both single stranded DNA and RNA *in vitro* in a gRNA dependent manner (16). These findings emphasize the diversity of pAgos but leave the origin of endogenous MpAgo guides and the physiological function of MpAgo enigmatic.

Here we probe guide-dependent RNA targeting by MpAgo. We show that MpAgo is easily programmed with gRNAs to form RNA-protein complexes (RNPs). However, reconstituted MpAgo RNPs are destabilized upon interaction with highly complementary RNA substrates, resulting in the release of MpAgo from the guide-substrate duplex. We therefore site-specifically crosslinked the gRNA to MpAgo to stabilize the MpAgo-gRNA complex (MpAgo xRNP), to better explore MpAgo's gRNA-dependent targeting specificity. Using MpAgo xRNPs, we identified a novel extended seed region of the gRNA for binding to RNA targets. Moreover, we show that MpAgo xRNPs can selectively bind substrates that differ by only one nucleotide. We also show that MpAgo xRNPs enable isolation of inosine-edited RNA substrates from a complex mixture.

RESULTS

***In vitro* reconstituted MpAgo-gRNA complexes are destabilized upon binding highly complementary RNA substrates.** The CRISPR locus associated MpAgo can be programmed with a short 5'-hydroxylated gRNA to cleave single stranded DNA and RNA substrates *in vitro* (16). We used catalytically active MpAgo for cleavage assays and catalytically inactive MpAgo for binding assays (16). Since the endogenous guides and targets for *M. piezophila* are not known, we investigated MpAgo interactions with a previously published 5'-hydroxylated 21 nucleotide (nt) gRNA (**Table S1**) (16). We determined the equilibrium dissociation constant of MpAgo for its gRNA at 55 °C using a filter binding (FB) assay, since *M. piezophila* grows at 45 – 70 °C (17). We observed that MpAgo binds the 5'-hydroxylated gRNA with ~50 nM affinity (**Fig. 1B**). We also tested if we could reconstitute the active MpAgo RNP at 37 °C, which would

enable expression and assembly of MpAgo RNPs in mesophilic cells. Although MpAgo binds the 5'-hydroxylated gRNA at 37 °C with a similar affinity as at 55 °C, the overall binding of gRNA is diminished by ~24% at the lower temperature (**Fig. 1B**).

To assess the stability of the reconstituted MpAgo RNP *in vitro,* we performed competition-binding assays, in which preassembled MpAgo RNP complexes were titrated with either 5'-hydroxylated or 5'-phosphorylated competitor gRNAs. The observed half-inhibitory concentrations (IC_{50}) for 5'-hydroxylated and 5'-phosphorylated gRNAs reveal that MpAgo has a very low affinity for 5'-phosphorylated gRNAs (**Fig. 1C**), in agreement with previously published structural and biochemical data on MpAgo gRNA interactions (16). The IC_{50} values also demonstrate that the 5'-hydroxylated gRNA is not easily displaced from MpAgo RNP by either 5'-hydroxylated or by 5'-phosphorylated RNAs.

Since different temperatures did not affect the affinity of the MpAgo for the gRNA but only impacted the efficiency of gRNA loading (**Fig. 1B**), we tested the stability of MpAgo RNP complexes reconstituted at 37 °C and 55 °C in the presence of a fully complementary RNA target (**Fig. S1A**). MpAgo RNPs immobilized on beads through a biotin tag on the gRNA were incubated with buffer (negative control), a 10 fold excess of mRNA containing a fully complementary target site (ON-target sample), or an excess (~2.8 μg) of total RNA purified from HEK293T cells (OFF-target sample). Notably, in the presence of a fully complementary RNA target, we observed more than 50% displacement of MpAgo from MpAgo RNPs assembled at 37 °C. Even with MpAgo RNPs reconstituted at 55 °C, nearly 30% of the MpAgo RNP complexes dissociated upon substrate binding (**Fig. S1B**). Collectively, these results indicate that reconstitution at higher temperatures increases proportion of stable MpAgo RNPs. However,

binding to fully complementary target RNAs destabilizes these RNPs regardless of reconstitution temperature.

Previously, it was shown that MpAgo does not have a preference for the 5'-terminal nucleotide of the gRNA in targeting ssDNA substrates for cleavage (16). Since the cellular target of MpAgo is not known, and some Argonautes including MpAgo target ssRNA *in vitro* (16), we next tested whether the 5'-end nucleotide of the gRNA affects the ability of MpAgo RNPs to bind and cleave RNA. We performed *in vitro* ssRNA cleavage assays using MpAgo RNPs loaded with one of four guides that differed only in the 5'-terminal nucleotide (**Fig. S2A**). Regardless of the 5' nucleotide identity, MpAgo cleaved ssRNA substrates with similar rates and efficiencies (**Fig. S2B**). We then examined if the 5'-terminal nucleotide of the gRNA influences ssRNA substrate binding efficiency and specificity *in vitro*. We performed binding assays using 0.1 nM substrate that was either fully complementary (ON target) or non-complementary (OFF target) to the gRNA, and 200 nM of MpAgo RNP programmed with the gRNAs differing at the 5'-terminal position (**Fig. 1D**). Whereas ON target binding efficiencies of all four RNPs are comparable, the specificity is affected by the identity of the 5'-terminal position of the gRNA. RNPs programmed with 5'-G and 5'-U gRNAs are more specific than the RNPs programmed with 5'-A and 5'-C gRNAs, although they only yielded a ~2-fold difference between ON and OFF target binding. To determine if this is due to the higher affinity of MpAgo for 5'-G and 5'-U gRNAs, we measured the equilibrium dissociation constants of MpAgo for all four gRNAs (**Fig. S2C**). The 5'-terminal nucleotide of the gRNA has no appreciable effect on MpAgo gRNA binding as we found the affinities to be similar in all four cases. The fact that MpAgo RNP cleaves only complementary substrates (16) indicates that non-specific binding of the ssRNA substrates obtained here is likely gRNA independent and could be caused by incomplete

assembly of the MpAgo RNP, allowing non-specific RNA binding to the gRNA binding cleft of free MpAgo (16).

Crosslinking the 5'-terminal nucleotide of the gRNA to MpAgo improves the ssRNA substrate binding specificity and affinity. Despite low nanomolar affinity of the MpAgo RNP for ssRNA substrates (**Fig. S3**), the low specificity (**Fig. 1D**) and destabilization of the MpAgo complex upon binding fully complementary RNA substrate (**Fig. S1B**) prompted us to engineer a more stable MpAgo RNP. The previously published structure of MpAgo bound to a gRNA revealed that the 5'-terminal nucleotide of the gRNA is flipped into a binding pocket formed by MpAgo's MID and PIWI domains (**Fig. 1A**), where it stacks on the aromatic ring of tyrosine 379 (Y379) (**Fig. 2A**) (16). We took advantage of the close proximity of the 5'-terminal nucleotide and the Y379 aromatic ring to crosslink the 5'-terminal nucleotide of the gRNA to MpAgo. We used a gRNA containing 5-bromo-deoxyuridine at the 5'-end position (**Table S1**), and found that crosslinking with 305 nm UV light was essentially quantitative (**Fig. 2B**). We then performed ssRNA cleavage assays using crosslinked MpAgo RNP (MpAgo xRNP) to test if crosslinking the gRNA impacts the performance of MpAgo (**Fig. 2C, S4**). The obtained k_{obs} values of MpAgo RNP and MpAgo xRNP are similar, indicating that crosslinking the gRNA to MpAgo does not significantly affect MpAgo cleavage activity (**Fig. 2C**). Moreover, in optimized conditions (1:1.5 MpAgo:gRNA ratio for crosslinking, 10 μg/mL of heparin, and 10 nM MpAgo xRNP), ON target RNA substrate binding efficiency was ~45%, whereas OFF target binding was diminished to a practically undetectable level of < 1% (**Fig. 2D, Fig. S4**).

To obtain MpAgo xRNP affinities for ssRNA substrates, we determined equilibrium dissociation constants using FB assays and found that the MpAgo xRNP binds the fully complementary ssRNA substrate (21-nt complementarity) very tightly with ~500 pM affinity (**Fig. 3**), whereas the non-crosslinked MpAgo RNP has ~5 times lower affinity (**Fig. S3**). We then investigated if mismatches of 3 or 6 nts at the 3'-end of the gRNA affect MpAgo xRNP binding to ssRNA substrate (**Fig. 3A**). Surprisingly, we found that reducing complementarity between the ssRNA and gRNA to 18 nts or 15 nts enhanced the affinity of MpAgo xRNP substantially, with K_d values of ~100 pM (**Fig. 3B**).

MpAgo utilizes an extended seed region on the gRNA with highest sensitivity at the 3rd, 6th, 7th and 9th nts. The seed sequence in the gRNA is one of the major factors that contributes to the specificity of Ago RNP interactions with target substrates. The canonical seed region of all eAgos and some pAgos is composed of the 2nd to 8th nucleotides of the gRNA and has an A-form helical conformation in the Ago RNP complex (18–23). Furthermore, the 2nd to 4th bases of the seed sequence are exposed to the solvent for interaction with the complementary substrate (22). Notably, in contrast to eAgos and pAgos, the previously determined structure of an MpAgo-gRNA complex revealed that the gRNA and the predicted seed region have a unique conformation: the guide RNA contains a sharp kink between the 6th and 7th nucleotide that interferes with the A-like helical conformation of the seed by disrupting continuous 2nd – 8th base stacking (16). Interestingly, the measured tolerance for dinucleotide mismatches across the MpAgo gRNA when targeting ssDNA revealed that the mismatches at the 5th to 15th nucleotides of the gRNA abolish the cleavage of ssDNA substrates the most (24). These structural and

biochemical studies suggest that the seed region of the MpAgo may differ from the canonical seed used by previously characterized eAgos and pAgos.

Due to the high affinity of 15-nucleotide target RNAs (Fig. 3B), we mapped the seed region of MpAgo when targeting ssRNA substrates using a baseline of 15 nucleotides of complementarity to target ssRNAs (**Fig. 3A**). We introduced single nucleotide mismatches at positions 1 to 15 in a 21 nucleotide long gRNA keeping the sequence of the gRNA constant (**Table S1, Fig. S5A**) and determined the equilibrium dissociation constants using filter binding assays (**Fig. S5B**). The average K_d values for each substrate were plotted against the mismatched position in the gRNA (**Fig. 4A**). As expected, based on the location of the 1st gRNA nucleotide in a separate binding pocket (**Fig. 2A**) and our use of a 5'-nt crosslink, there is no effect of a mismatch at the first position. We noticed that the K_d values for the substrates that have mismatches spanning the 2nd to 12th positions, are higher relative to the K_d values for the fully complementary ssRNA substrate, or for mismatches in the 13th-15th positions (**Fig. 4A**).

We also tested whether the type of mismatch affects the observed K_d value for target binding. In the context of scanning mismatches in one gRNA, purine-purine mismatches (A-G mismatch in gRNA–substrate at the 6th position, and G-A mismatches at the 2nd and 8th positions) in general weakened target binding substantially more than pyrimidine-pyrimidine mismatches (**Fig. 4A, Fig. S5**). To check whether the negative impact of the mismatch at the 6th position is due to the nature of an A-G mismatch, we exchanged the gRNA sequence to obtain a G-A mismatch at the 6th position, which resulted in a K_d very similar to those for substrates that form G-A mismatches at the 2nd and 8th positions (**Fig. 4A, Fig. S7A**). We performed additional filter binding assays using diverse gRNAs and substrates (**Table S1**) to probe purine-purine mismatches at the 3rd, 7th, 9th, 10th, 12th and 14th positions to more rigorously define the seed

region for MpAgo RNPs (**Fig. S6**). The determined K_d values indicate that the MpAgo RNP seed region spans nucleotides 2 – 9, and is highly sensitive to mismatches at the 3rd, 6th, 7th and 9th positions (**Fig. 4B**). Taken together, our data suggests that MpAgo xRNP has an extended seed compared to the canonical seed in eAgos and some pAgos (18–23), when targeting ssRNAs (**Fig. 4C**).

MpAgo xRNPs can be programmed to distinguish between single nucleotide variants. Using the 6th and 7th nucleotides located at the kink of the gRNA (16) to probe the specificity of the MpAgo xRNP's interaction with its RNA substrate, we investigated whether MpAgo xRNPs could distinguish between substrates that differ by only one nucleotide at the corresponding complementary position. We performed filter binding assays using MpAgo xRNPs programmed with gRNAs that have A, G, U or C either at the 6th or at the 7th position and RNA substrates with all permutations at the corresponding positions (**Fig. S7, Fig. S8**). The obtained average dissociation constants for the 6th and 7th position are presented in the form of heat maps (**Fig. 5A, B**). Notably, at the 6th position MpAgo xRNP has the highest affinity for the fully complementary substrates (gRNA:substrate base pairs G:C, U:A, A:U and C:G), the substrates which form pyrimidine-pyrimidine mismatches (U–U and C–U), and the substrate that forms a G•U wobble. It is notable that substrates forming U–U and C–U mismatches bind with high affinity, suggesting that these mismatches are well accommodated in the gRNA-substrate RNA duplex at this position. Interestingly, only the G:C base pair forming substrate binds more strongly than the substrate that creates a G•U wobble, whereas the binding affinities of the substrates that form U:A, A:U and C:G Watson-Crick base pairs are ≥ 2× weaker. Purine-purine mismatches destabilize substrate RNA binding most, with A–A, A–G, G–A and G–G

mismatches increasing the K_d values 20- to 56-fold relative to Watson-Crick pairs. Moreover, MpAgo xRNP binding affinity to ssRNA also depends on which nucleotide of the mismatch is on the gRNA strand. The binding affinity of the substrate forming a G•U wobble is 21× tighter than that of the substrate forming a U•G wobble. Similar trends are observed for other types of mismatches, e.g. G–A vs. A–G, C–A vs. A–C, and C–U vs. U–C (**Fig. 5A**). Even in the case of perfectly matching gRNA-substrate duplexes, a C:G base pair binds 13x more tightly than a G:C base pair.

In contrast to the 6th position, at the 7th position MpAgo xRNP has the highest affinity only for the fully complementary substrates (gRNA:substrate base pairs G:C, U:A, A:U and C:G). Even though pyrimidine-pyrimidine mismatches and the G•U and U•G wobble affect the binding affinities the least, they increase the K_d values ~10- to 20-fold relative to Watson-Crick pairs, suggesting that even these mismatches are not well accommodated in the gRNA-substrate RNA duplex at this position. Purine-purine mismatches (A–A, A–G, G–A and G–G) and purine-pyrimidine mismatch A–C destabilize substrate RNA binding even more, increasing the K_d values 10- to 100-fold relative to Watson-Crick pairs. Similarly to the 6th position, MpAgo xRNP binding affinity to ssRNA also depends on which nucleotide of the mismatch is on the gRNA strand. The binding affinity of the substrate forming a C–A mismatch is 47× tighter than that of the substrate forming an A–C mismatch. Similar trends are observed for other types of mismatches, e.g. G–A vs. A–G, G•U vs. U•G, and C–U vs. U–C (**Fig. 5B**). Collectively, the results of the permutations at the 6th and 7th positions of the gRNA and at the corresponding positions of the ssRNA substrate indicate that MpAgo monitors not only the type of mismatch, but also the identity of the bases forming the mismatch.

Next we tested whether MpAgo has high enough specificity to recognize modified nucleotides in the RNA substrates. For instance, adenosine deaminases acting on RNA (ADARs) catalyze adenosine to inosine (A-to-I) editing of certain mRNAs (25, 26). Compared to adenosine, inosine has very different base-paring properties (27). We wondered if MpAgo xRNPs could be programmed to distinguish between A-to-I edited and non-edited RNA substrates. Thus we performed binding assays using MpAgo xRNPs programmed with gRNAs that have A, G, U or C at either the 6th or at the 7th positions and substrate with inosine at the corresponding complementary positions (**Fig. 5A, B, Fig. S7, Fig. S8**). Although MpAgo xRNPs have low picomolar affinity for the inosine containing substrate irrespective of which nucleotide is used at the 6th position of the gRNA, the presence of an A at the 6th position of the gRNA leads to ~10-fold weaker binding to substrates with an A instead of I (**Fig. 5A, Fig. S7**). Intriguingly at the 7th position, MpAgo xRNP binds the I-containing substrate weaker than at the 6th position with the exception of the MpAgo xRNP that has C at it 7th position (**Fig. 5B, Fig. S8**). Moreover, the presence of a C at the 7th position of the gRNA leads to ~33-fold weaker binding to substrates with an A instead of I (**Fig. 5B, Fig. S8**).

We then investigated whether MpAgo xRNP programmed with the gRNA containing either A at its 6th or C at its 7th position could be used to isolate A-to-I edited RNAs from a complex mixture. We used filter binding assays, in which we spiked in either a radiolabeled, inosine-containing RNA substrate or a radiolabeled, adenosine-containing RNA substrate into HEK293T total RNA, and incubated the mixtures with MpAgo xRNP programmed with a gRNA containing either A at its 6th or C at its 7th position. We also substituted bulk tRNA for heparin in the binding reactions to make the conditions of binding reactions more physiological. At the 6th position, MpAgo xRNP bound an I-containing substrate with ~19% efficiency, while the A-

harboring RNA substrate was bound only with 5% efficiency. At the 7th position, MpAgo xRNP bound an I-containing substrate with ~42% efficiency, while the A-harboring RNA substrate was bound only with 4% efficiency (**Fig. 5C**). These results are consistent with the specificity seen with pure substrates in FB assays (Fig. S7, Fig. S8) at the concentration of MpAgo xRNP used with the complex mixture (500 pM MpAgo xRNP). These results demonstrate that the MpAgo xRNP can distinguish between the RNA substrates that differ by only one nucleotide with high specificity, and that it can be programmed to isolate A-to-I edited substrates from complex mixtures.

DISCUSSION

As opposed to all other known Agos, the CRISPR-associated Agos in the *M. piezophila* family bind chemically distinct 5'-hydroxylated gRNAs rather than phosphorylated guides. This implies that the origin and biological assembly pathway of these Ago RNPs are also likely to be unique. It is likely that MpAgo gRNAs are generated by nucleases associated with the CRISPR operon, namely Cas6, Csx1, or Csm6 (28–30). All three nucleases produce 5'-hydroxylated short RNAs (31–33) that could serve as MpAgo guides. We speculate that either Csx1 or Csm6 but not Cas6 is responsible for generating gRNAs for MpAgo, as Cas6 cleaves CRISPR array transcripts within repeat sequences producing short RNAs that contain a conserved (8-nt) repeat sequence at their 5′ end (32). The conserved sequence spans the most important region of the seed sequence (**Fig. 4**), which would defeat the functional role of this region of the gRNA. The fact that the source of guides is likely unique for these CRISPR-associated Agos suggests that their biological

role could differ from other pAgos (8, 13–15), a question that will require future experiments to answer.

In addition to the unknown origin of MpAgo gRNAs, the physiological targets for MpAgo are currently obscure. The *in vitro* reconstituted MpAgo-gRNA complexes are destabilized upon binding to highly complementary RNA substrates. This behavior is similar to that of human Argonaute 2, which releases its gRNA-substrate duplex upon binding a perfectly complementary substrate, at least *in vitro* (34). This suggests that the MpAgo RNP may bind ssRNA substrates using a functionally similar mechanism as the human Argonaute 2 RNP. This MpAgo release mechanism could serve as means to re-load MpAgo with new gRNAs *in vivo*.

The destabilization of the MpAgo RNP upon its interaction with a target RNA substrate has implications for engineering a stable and specific MpAgo RNP. We site-specifically crosslinked the 5'-end nucleotide of the gRNA to MpAgo with an exceptionally high (~100%) efficiency compared to previously reported crosslinking efficiencies of 20% to 40% (35, 36). Importantly, crosslinking the gRNA to MpAgo significantly improved MpAgo xRNP's stability, affinity, and specificity for RNA substrates. MpAgo xRNP can bind complementary RNA substrates up to 300 times more efficiently than non-complementary substrates with a mismatch at a single nucleotide (**Fig. 4B**), while in the case of non-crosslinked MpAgo RNPs the difference is barely two-fold.

The specificity of the Ago RNP's interaction with its substrates is determined by the seed sequence in the guide RNA. We have shown that MpAgo RNP has an extended seed region spanning the 2nd to 9th nucleotides of the gRNA. Although the position of the seed region is similar to that previously reported for other Agos, the 3rd, 6th, 7th and 9th positions of the seed

region impact the MpAgo RNP's specificity in binding target RNAs the most, rather than the 4th and 5th nt of the gRNA that contribute the most to the specificity of eAgos (37). In MpAgo, the 6th-7th positions, may serve as conformational checkpoints, since the guide RNA is kinked between the 6th and 7th nt (**Fig. 1A**). We hypothesize that base paring of the 6th and 7th nucleotides releases the distortion in the gRNA and allows further 5' to 3' propagation of the guide-substrate base pairing. The structure of MpAgo RNP bound to a DNA substrate suggests that two protein helices of the L2 domain and a loop of the PIWI domain could enable MpAgo to sense the shape of the minor groove at the 6th and 7th position of the gRNA and ensure that the gRNA-RNA substrate duplex is complementary (**Fig. S9**) (24), although MpAgo RNP interactions with DNA and RNA substrates are likely to be distinct (**Fig. 4**). The permutations at the 6th and at the 7th nt of the gRNA and at the corresponding complementary position of the substrate RNA demonstrate that MpAgo not only monitors the position and type of the mismatch but also the identity and position of the bases forming the mismatch. This discovery reveals a coding feature of the MpAgo xRNP that enables it to discriminate between substrate RNAs differing by only a single nucleotide.

The MpAgo xRNP can also be programmed to distinguish between target RNAs that have been post-transcriptionally modified, i.e. that contain common base modifications such as inosine. Despite inosine's ability to base pair with all four canonical bases, it forms the most stable base pair with cytosine. Therefore, it is recognized as guanine by the translational machinery when it is present in coding mRNA sequences and can cause amino acid substitutions (38). Additionally, perturbations of A-to-I editing were found to correlate with several neurological diseases (39). Despite the prevalence of A-to-I editing, the functions of most editing sites remain unknown. The method primarily used to identify editing sites is based on the

comparison of cDNA and corresponding genomic DNA sequences (38). However, there is no reliable method to investigate the function, cellular localization, and interaction networks of specific, edited RNA transcripts *in vivo*. The ability of MpAgo xRNPs to isolate A-to-I edited RNAs from a complex mixture should enable deeper exploration of the biological function of inosine modifications in RNAs. In principle, MpAgo RNPs could also be used to isolate mRNAs with C-to-U deaminations (40) using gRNAs with an A at the 7th position (Fig. 5B).

MpAgo has moderate affinity for its 5'-hydroxylated gRNA (K_d ~50 - 100 nM), when compared to other Agos (K_d ~1 – 7 nM for *Rhodobacter sphaeroides* Ago (41) and human Argonaute 2, respectively (42)). Thus, MpAgo is relatively easily programmed with defined gRNAs to target diverse RNA substrates of interest. With subsequent covalent crosslinking of gRNAs to MpAgo, MpAgo xRNPs bind fully complementary ssRNA substrates with much higher affinities than most RNA binding proteins currently employed for RNA targeting. Whereas the MS2 and PP7 coat proteins, the λN protein, and repurposed RCas9 bind their RNA substrates with 1.6 nM – 5 nM affinities, only the CRISPR nuclease Csy4 binds its RNA hairpin substrate with similar affinity to MpAgo, ~50 pM (43–47). Of these systems, only RCas9 is easily programmable, and requires a PAMmer DNA oligonucleotide for efficient RNA binding (46, 48, 49). Notably, the cellular concentration of mRNAs in mammalian cells is generally in the 10 – 100 pM range, except for the most abundant transcripts (50, 51), matching the affinity range for MpAgo xRNPs. Finally, MpAgo xRNPs can be used to discriminate single nucleotide variants and to identify A-to-I edited target RNAs. In the future, it should be possible to use MpAgo xRNPs to analyze, image and manipulate untagged RNAs in live cells with single nucleotide resolution.

MATERIALS AND METHODS

A full description of the materials and methods used in this study, including *in vitro* RNA synthesis and radiolabeling, *in vitro* reconstitution of MpAgo RNP, MpAgo unloading assays, *in vitro* binding and cleavage assays, MpAgo-gRNA crosslinking, A-to-I edited substrate isolation assays, and a list of oligonucleotides, is provided in SI Materials and Methods and Table S1.

ACKNOWLEDGMENTS

We thank the members of the Cate and the Doudna laboratories, especially Steven C. Strutt, Emine Kaya, and Kevin W. Doxzen for their technical assistance and helpful discussions. A.L. acknowledges support from the Human Frontiers in Science Program (HFSP). This work was supported by the US National Institutes of Health (NIH) grant R01-GM065050 to J.H.D.C. and by a Frontiers Science award from the Paul Allen Institute to J.A.D. J.A.D. is an Investigator of the Howard Hughes Medical Institute (HHMI). The authors have submitted a patent related to this work.

REFERENCES

1. Swarts DC, et al. (2014) The evolutionary journey of Argonaute proteins. *Nat Struct Mol Biol* 21(9):743–753.
2. Volpe TA, et al. (2002) Regulation of Heterochromatic Silencing and Histone H3 Lysine-9 Methylation by RNAi. *Science (80-)* 297(5588):1833–1837.
3. Sigova A, Rhind N, Zamore PD (2004) A single Argonaute protein mediates both transcriptional and posttranscriptional silencing in Schizosaccharomyces pombe. *Genes Dev* 18(19):2359–2367.
4. Tabara H, et al. (1999) The rde-1 gene, RNA interference, and transposon silencing in C. elegans. *Cell* 99(2):123–32.

5. Houwing S, et al. (2007) A Role for Piwi and piRNAs in Germ Cell Maintenance and Transposon Silencing in Zebrafish. *Cell* 129(1):69–82.

6. Wilkins C, et al. (2005) RNA interference is an antiviral defence mechanism in Caenorhabditis elegans. *Nature* 436(7053):1044–1047.

7. Liu J, et al. (2004) Argonaute2 Is the Catalytic Engine of Mammalian RNAi. *Science (80-)* 305(5689):1437–1441.

8. Swarts DC, et al. (2014) DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* 507(7491):258–261.

9. Wu L, Fan J, Belasco JG (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proc Natl Acad Sci U S A* 103(11):4034–9.

10. Giraldez AJ, et al. (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312(5770):75–9.

11. Behm-Ansmant I, et al. (2006) mRNA degradation by miRNAs and GW182 requires both CCR4:NOT deadenylase and DCP1:DCP2 decapping complexes. *Genes Dev* 20(14):1885–98.

12. Shabalina SA, Koonin E V (2008) Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol* 23(10):578–87.

13. 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.

14. Yuan Y-R, et al. (2005) Crystal Structure of A. aeolicus Argonaute, a Site-Specific DNA-Guided Endoribonuclease, Provides Insights into RISC-Mediated mRNA Cleavage. *Mol Cell* 19(3):405–419.

15. 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.

16. Kaya E, et al. (2016) A bacterial Argonaute with noncanonical guide RNA specificity. *Proc Natl Acad Sci U S A* 113(15):4057–62.

17. Prieur D, et al. (2002) Marinitoga piezophila sp. nov., a rod-shaped, thermo-piezophilic bacterium isolated under high hydrostatic pressure from a deep-sea hydrothermal vent. *Int J Syst Evol Microbiol* 52(4):1331–1339.

18. Elkayam E, et al. (2012) The Structure of Human Argonaute-2 in Complex with miR-20a. *Cell* 150(1):100–110.

19. Nakanishi K, Weinberg DE, Bartel DP, Patel DJ (2012) Structure of yeast Argonaute with guide RNA. *Nature* 486(7403):368–374.

20. Schirle NT, MacRae IJ (2012) The Crystal Structure of Human Argonaute2. *Science (80-)* 336(6084):1037–1040.

21. Schirle NT, Sheu-Gruttadauria J, MacRae IJ (2014) Structural basis for microRNA targeting. *Science (80-)* 346(6209):608–613.

22. Wang Y, et al. (2008) Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. *Nature* 456(7224):921–6.

23. 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.

24. Doxzen KW, et al. (2017) DNA recognition by an RNA-guided bacterial Argonaute. *PLoS One* 12(5):e0177097.

25. Wagner RW, Smith JE, Cooperman BS, Nishikura K (1989) A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proc Natl Acad Sci U S A* 86(8):2647–51.

26. Bass BL, Weintraub H (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55(6):1089–98.

27. Kawase Y, Iwai S, Inoue H, Miura K, Ohtsuka E (1986) Studies on nucleic acid interactions. I. Stabilities of mini-duplexes (dG2A4XA4G2-dC2T4YT4C2) and self-complementary d(GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. *Nucleic Acids Res* 14(19):7727–36.

28. Huynen M, Snel B, Lathe W, Bork P (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res* 10(8):1204–10.

29. Galperin MY, Koonin E V. (2000) Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol* 18(6):609–613.

30. Aravind L (2000) Guilt by association: contextual information in genome analysis. *Genome Res* 10(8):1074–7.

31. Niewoehner O, Jinek M (2016) Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA* 22(3):318–329.

32. 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.

33. Sheppard NF, Glover CVC, Terns RM, Terns MP (2016) The CRISPR-associated Csx1 protein of *Pyrococcus furiosus* is an adenosine-specific endoribonuclease. *RNA* 22(2):216–224.

34. De N, et al. (2013) Highly Complementary Target RNAs Promote Release of Guide RNAs from Human Argonaute2. *Mol Cell* 50(3):344–355.

35. Willis MC, LeCuyer KA, Meisenheimer KM, Uhlenbeck OC, Koch TH (1994) An RNA-protein contact determined by 5-bromouridine substitution, photocrosslinking and sequencing. *Nucleic Acids Res* 22(23):4947–52.

36. Gott JM, Willis MC, Koch TH, Uhlenbeck OC (1991) A specific, UV-induced RNA-protein cross-link using 5-bromouridine-substituted RNA. *Biochemistry* 30(25):6290–5.

37. Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD (2012) Argonaute Divides Its RNA Guide into Domains with Distinct Functions and RNA-Binding Properties. *Cell* 151(5):1055–1067.

38. Burns CM, et al. (1997) Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 387(6630):303–308.

39. Maas S, Kawahara Y, Tamburro KM, Nishikura K (2006) A-to-I RNA editing and human disease. *RNA Biol* 3(1):1–9.

40. Teng B, Burant CF, Davidson NO (1993) Molecular cloning of an apolipoprotein B messenger RNA editing protein. *Science* 260(5115):1816–9.

41. Miyoshi T, Ito K, Murakami R, Uchiumi T (2016) Structural basis for the recognition of guide RNA and target DNA heteroduplex by Argonaute. *Nat Commun* 7:11846.

42. Deerberg A, Willkomm S, Restle T (2013) Minimal mechanistic model of siRNA-dependent target RNA slicing by recombinant human Argonaute 2 protein. *Proc Natl Acad Sci U S A* 110(44):17850–5.

43. Sternberg SH, Haurwitz RE, Doudna JA (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* 18(4):661–672.

44. LeCuyer KA, Behlen LS, Uhlenbeck OC (1995) Mutants of the bacteriophage MS2 coat protein that alter its cooperative binding to RNA. *Biochemistry* 34(33):10600–6.

45. Cilley CD, Williamson JR (1997) Analysis of bacteriophage N protein and peptide binding to boxB RNA using polyacrylamide gel coelectrophoresis (PACE). *RNA* 3(1):57–67.

46. O'Connell MR, et al. (2014) Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* 516(7530):263–266.

47. Chao JA, Patskovsky Y, Almo SC, Singer RH (2008) Structural basis for the coevolution of a viral RNA–protein complex. *Nat Struct Mol Biol* 15(1):103–105.

48. Nelles DA, et al. (2016) Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* 165(2):488–496.

49. Batra R, et al. (2017) Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. *Cell* 170(5):899–912.e10.

50. Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 14(9):618–30.

51. Albayrak C, et al. (2016) Digital Quantification of Proteins and mRNA in Single Mammalian Cells. *Mol Cell* 61(6):914–24.

52. Milligan JF, Groebe DR, Witherell GW, Uhlenbeck OC (1987) Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* 15(21):8783–98.

53. Rio DC (2012) Filter-Binding Assay for Analysis of RNA-Protein Interactions. *Cold Spring Harb Protoc* 2012(10):pdb.prot071449-prot071449.

FIGURES

Figure 1. Stability and specificity of *in vitro* reconstituted MpAgo RNPs. (A) (Top) Linear depiction of the organization of MpAgo's domains and (bottom) the structure (PDB ID: 5I4A) of MpAgo (grey, in surface representation) bound to its guide RNA (orange, in cartoon representation). The bi-lobed structure includes a PAZ (PIWI-Argonaute-Zwille) domain that anchors the 3'-end of the guide RNA, while the second lobe includes a MID (middle) and a PIWI (P-element Induced WImpy testis) domain that binds the 5'-end of the guide. Catalytic residues in the PIWI domain are marked. The structure of MpAgo reveals the unique conformation of its gRNA, which is bent at the 6th and 7th nts. **(B)** MpAgo and gRNA binding curves obtained at two temperatures: 37 °C (orange) and 55 °C (black). The fraction of the bound gRNA is plotted as a function of the MpAgo concentration. The data were fit with a standard binding isotherm (solid lines) to extract the binding affinities. MpAgo binds the gRNA at 37 °C and 55 °C with K_d values of 52 ± 4 nM and 49 ± 5 nM, respectively. Data are represented as the mean ± standard deviation (SD) from three independent experiments. **(C)** Competition-binding assay at 37 °C to examine displacement of 3'-radiolabeled gRNA from the RNP complex reconstituted at 55 °C, with 5'-OH RNA competitor (black) and 5'-PO_4 RNA competitor (orange). The half-inhibitory concentrations (IC_{50}) are 0.55 ± 0.08 μM and > 4.8 μM for the 5'-OH RNA and 5'-PO_4 RNA competitors, respectively. Data are represented as the mean ± SD from three independent experiments. **(D)** Substrate ssRNA filter-binding assays with catalytically inactive MpAgo RNP formed with a 5'-G, 5'-A, 5'-U, or 5'-C gRNA, and either the fully complementary substrate (ON target, black) or non-complementary ssRNA substrate (OFF target, red). The fraction of the ssRNA substrate bound to the MpAgo RNP was quantified and plotted as the mean ± SD from

three independent experiments.

Figure 2. Optimized conditions for crosslinked MpAgo RNP assembly and its specificity for binding complementary ssRNA substrates. (A) MpAgo (blue) (PDB ID: 5I4A) forms a binding pocket for the 5'-end nucleotide (U1) of the gRNA (grey) that stacks on the aromatic ring of tyrosine residue Y379. The close proximity of the 5'-end nucleotide and the Y379 residue allows site-specific crosslinking of 5-bromo-deoxyU modified gRNA and MpAgo. **(B)** The *in vitro* reconstituted and UV crosslinked MpAgo xRNP was resolved on a 4-20% gradient SDS-PAGE. The gel was first stained with SYBR Gold to detect the gRNA band and then stained with Comassie to image the MpAgo protein band. **(C)** Substrate ssRNA cleavage assay using either crosslinked (orange) or non-crosslinked (black) MpAgo RNP. The amount of cleaved 5'-radiolabed ssRNA substrate was plotted as a function of time. The data were fit with a single exponential to extract the apparent cleavage rate constant (k_{obs}). Average k_{obs} are: 0.31 ± 0.03 min^{-1} and 0.59 ± 0.07 min^{-1} for xRNP and RNP respectively. Data are represented as the mean ± SD from three independent experiments. **(D)** Substrate ssRNA filter-binding assays with initial and final MpAgo:gRNA concentrations and binding conditions. Assays were performed at 37 °C, incubating 200 nM of MpAgo RNP (reconstituted at 1:1.1 ratio of MpAgo:gRNA) and 10 nM of crosslinked MpAgo xRNP (reconstituted at 1:1.5 ratio of MpAgo:gRNA) with 0.1 nM of either the fully complementary (ON target, black) or the non-complementary ssRNA substrate (OFF target, red) in the presence of either 2 μg/mL (for non-crosslinked RNP) or 10 μg/mL of heparin (for xRNP). The fraction of ssRNA substrate bound to the RNP was quantified and plotted as the mean ± SD from three independent experiments.

Figure 3. Crosslinked MpAgo RNP binds ssRNA substrates with high affinity. (A) Cartoon of three duplexes of the gRNA and different ssRNA substrates with varying degree of complementarity to the gRNA used in binding assays: full 21-nt complementarity (top, red), 18-nt complementarity with 3 mismatches at the 3'-end of the gRNA (middle, orange), and 15-nt complementarity with 6 mismatches at the 3'-end of the gRNA (bottom, black). The first nt at the 5'-end of the gRNA does not base pair even if it is complementary (**Fig. 2A**). **(B)** Filter-binding assays with MpAgo xRNP and ssRNA substrates with 21 nt complementarity (red), 18 nt complementarity (orange), or 15 nt complementarity (black). The amount of bound ssRNA substrate is plotted as a function of MpAgo xRNP concentration. The obtained average K_d's for MpAgo xRNP and the substrates with 21-nt, 18-nt, and 15-nt complementarity are 500 ± 115 pM, 107 ± 6 pM, and 83 ± 8 pM, respectively. Data are represented as the mean ± SD from three independent experiments.

Figure 4. Determination of the gRNA seed region of MpAgo xRNPs. (A) Substrate ssRNA filter-binding assays were performed with crosslinked MpAgo xRNP and ssRNA substrates that contain a single nucleotide mismatch along the gRNA from the 1st to 15th position. The sequence of the gRNA was kept constant, and the obtained average dissociation constants were plotted against the position of the mismatch. Purine-purine and pyrimidine-pyrimidine mismatches are represented as white and black bars respectively. The red dash line indicates the value equal to ~5 times the K_d for the fully complementary ssRNA substrate. Mismatches at the 2nd through 12th guide nucleotides increase the dissociation constant ≥5 times the K_d for the fully complementary

ssRNA substrate. Asterisk (*), the mismatch at the 6th position is A–G (guide RNA–substrate RNA). The K_d for the substrate that has G–A mismatch (guide RNA–substrate RNA) at the 6th position is 779 ± 71 pM, using a different gRNA (**Table S1**), and is similar to the K_d's for the substrates with G–A mismatch at the 2nd and 8th positions. **(B)** The ratios of the average K_d's of the substrates that create purine-purine mismatch (K_d^{mm}) to the fully complementary substrate (K_d^{Compl}) (Fold change = K_d^{mm}/ K_d^{Compl}) were plotted against the position of the mismatch. The red dash line indicates a value equal to 5. The purine–purine mismatches at the 2nd through 12th positions decrease the affinity of MpAgo xRNP more than 5 fold. The 3rd, 6th, 7th and 9th positions reduce the affinity of MpAgo xRNP the most, 32-, 22-, 301- and 327-fold respectively. **(C)** The gRNA-substrate RNA duplex with the extended seed region (orange box).

Figure 5. Crosslinked MpAgo RNP ssRNA substrate binding specificity at the 6th and 7th positions. (A) Substrate ssRNA filter-binding assays were performed using MpAgo xRNPs programmed with A, G, U or C gRNA at the 6th position and ssRNA substrates containing U, C, A, G or I at the corresponding position. The average K_d's were extracted and are represented in a heat map. **(B)** Substrate ssRNA filter-binding assays were performed using MpAgo xRNPs programmed with A, G, U or C gRNA at the 7th position and ssRNA substrates containing U, C, A, G or I at the corresponding position. The average K_d's were extracted and are represented in a heat map. **(C)** A filter-binding assay to isolate the A-to-I edited substrate (black) but not the non-edited substrate (red) from the 500 ng of total RNA purified from HEK239T cells in the presence of 200 ng/μL of yeast tRNA. For this assay MpAgo xRNP programmed with a gRNA containing either A at its 6th or C at its 7th position was used. The fraction of the ssRNA substrate bound to

MpAgo xRNP was quantified and plotted as the mean ± SD from three independent experiments.

Figure 1

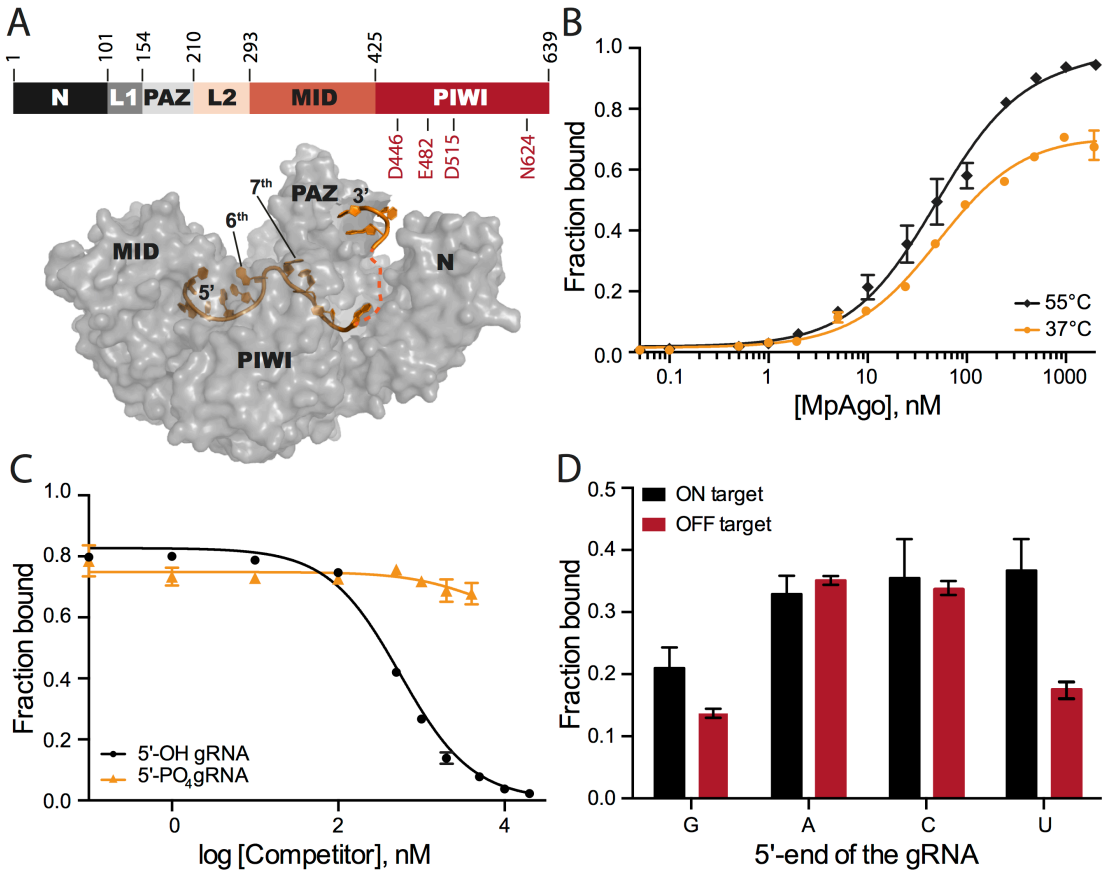

A
1
101
154
210
293
425
639
N
L1
PAZ
L2
MID
PIWI
D446
E482
D515
N624
7th
PAZ
3'
6th
MID
5'
N
PIWI
B
Fraction bound
[MpAgo], nM
55°C
37°C
C
Fraction bound
log [Competitor], nM
5'-OH gRNA
5'-PO4gRNA
D
Fraction bound
5'-end of the gRNA
ON target
OFF target
G
A
C
U

Figure 2

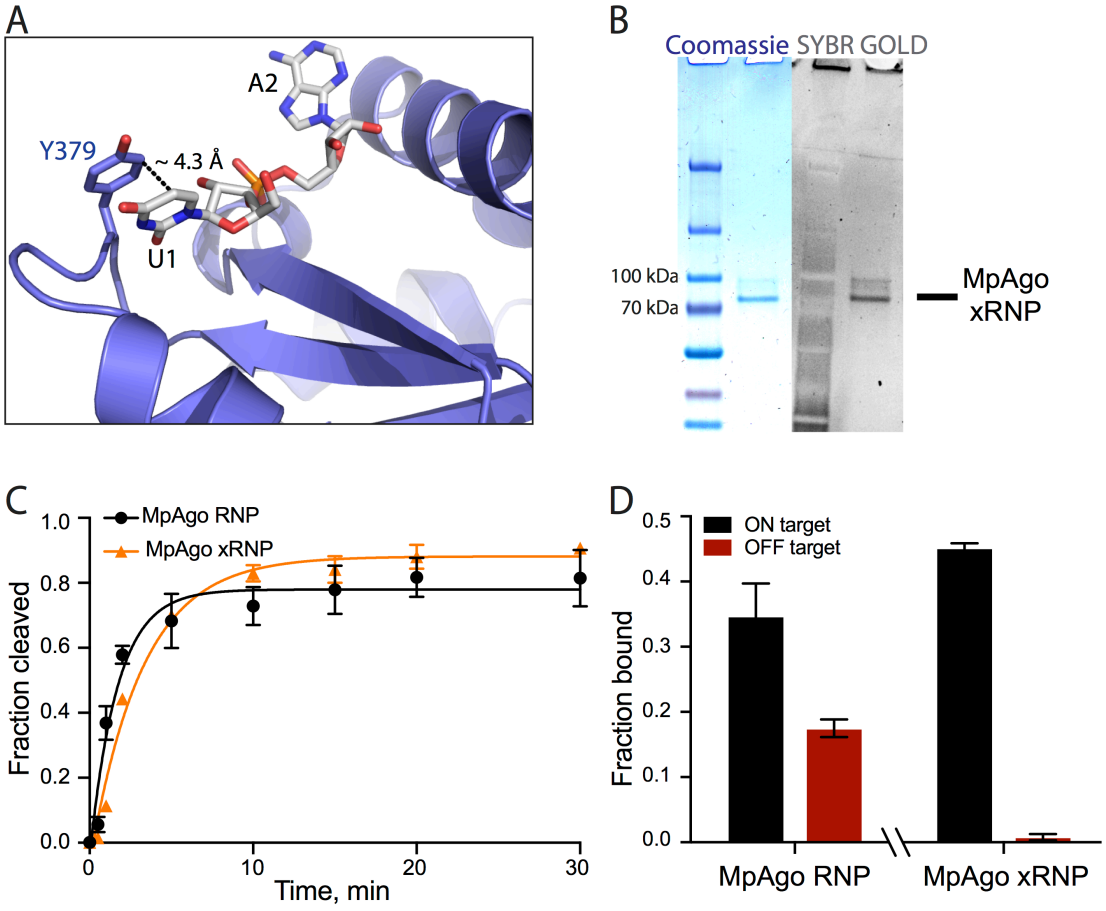

A
A2
Y379
~ 4.3 Å
U1
B
Coomassie
SYBR GOLD
100 kDa
70 kDa
MpAgo
xRNP
C
MpAgo RNP
MpAgo xRNP
Fraction cleaved
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
Time, min
D
ON target
OFF target
Fraction bound
0.5
0.4
0.3
0.2
0.1
0.0
MpAgo RNP
MpAgo xRNP

Figure 3

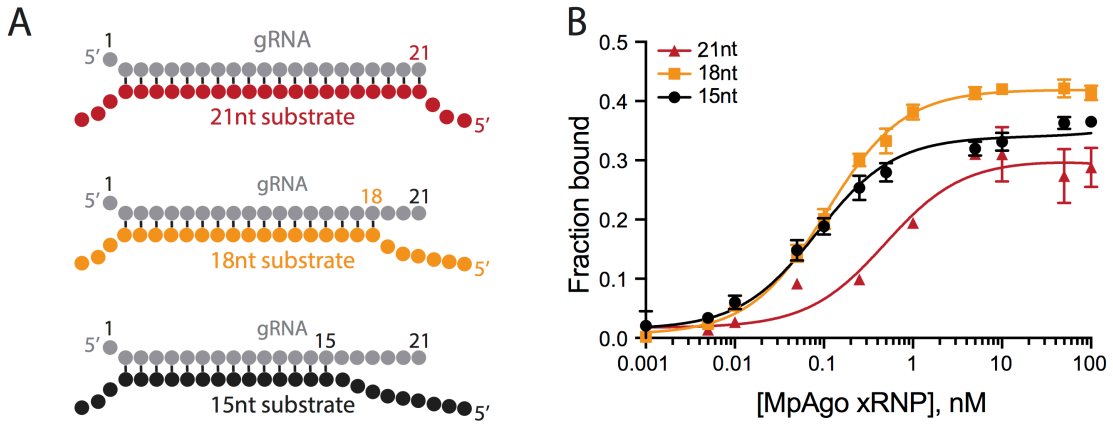

A
5′
1
gRNA
21
21nt substrate
5′
5′
1
gRNA
18
21
18nt substrate
5′
5′
1
gRNA
15
21
15nt substrate
5′
B
21nt
18nt
15nt
Fraction bound
0.0
0.1
0.2
0.3
0.4
0.5
0.001
0.01
0.1
1
10
100
[MpAgo xRNP], nM

Figure 4

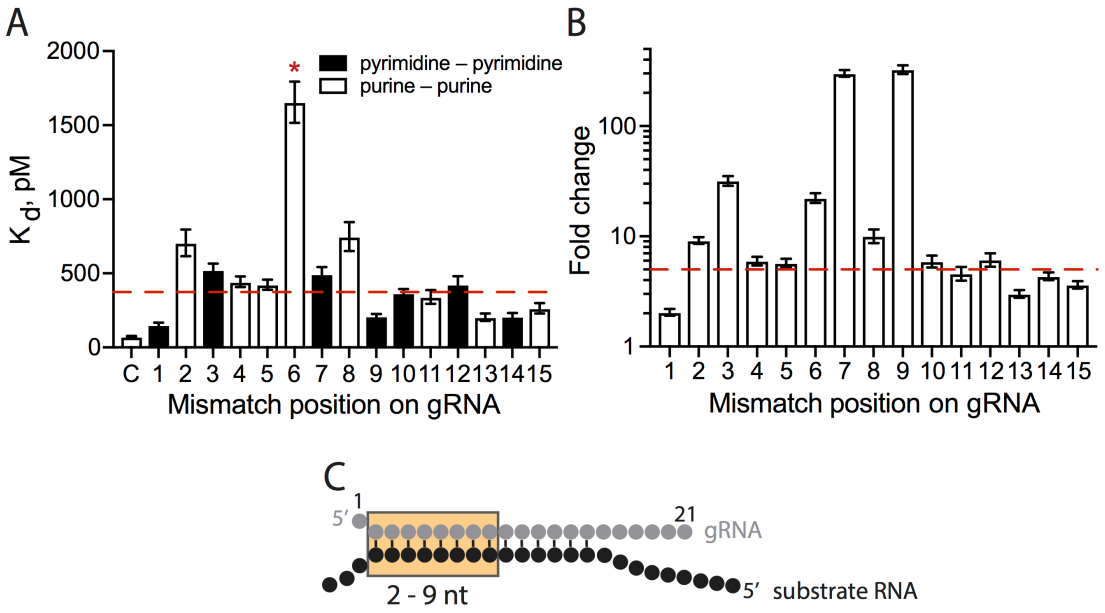
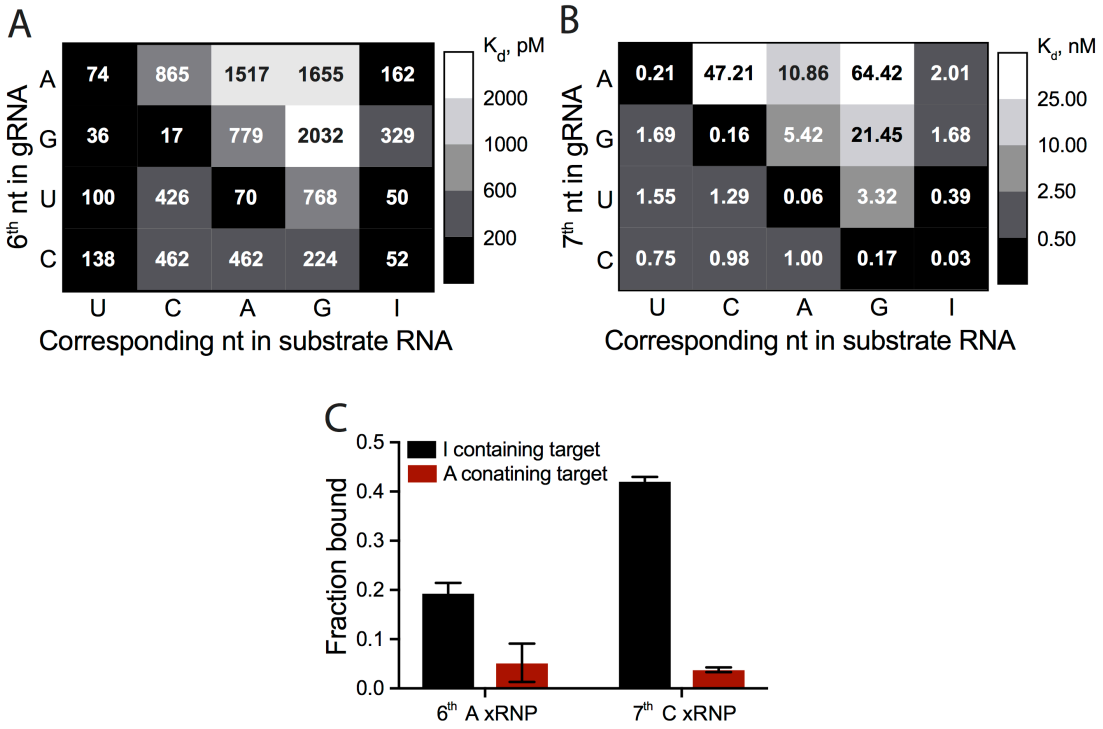

A
2000
1500
1000
500
0
K d, pM
*
pyrimidine – pyrimidine
purine – purine
C 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Mismatch position on gRNA
B
100
10
1
Fold change
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Mismatch position on gRNA
C
5′
1
21
gRNA
2 - 9 nt
5′
substrate RNA

Figure 5


A
6th nt in gRNA
A
G
U
C
74
865
1517
1655
162
36
17
779
2032
329
100
426
70
768
50
138
462
462
224
52
U
C
A
G
I
Corresponding nt in substrate RNA
Kd, pM
2000
1000
600
200
B
7th nt in gRNA
A
G
U
C
0.21
47.21
10.86
64.42
2.01
1.69
0.16
5.42
21.45
1.68
1.55
1.29
0.06
3.32
0.39
0.75
0.98
1.00
0.17
0.03
U
C
A
G
I
Corresponding nt in substrate RNA
Kd, nM
25.00
10.00
2.50
0.50
C
Fraction bound
0.5
0.4
0.3
0.2
0.1
0.0
I containing target
A conatining target
6th A xRNP
7th C xRNP

SUPPORTING INFORMATION

SI MATERIALS AND METHODS

Protein Expression and Purification. Recombinant wild-type (WT) MpAgo (UniProtKB accession code H2J4R4) and catalytically inactive mutant D446A MpAgo (dMpAgo) were expressed in *E. coli* strain BL21 (DE3) (New England Biolabs) and purified as described previously (16). The purified proteins were concentrated to ~8 mg/mL, flash-frozen in liquid nitrogen, and stored at −80 °C. Catalytically active MpAgo was used for cleavage assays, whereas the catalytically inactive mutant was used in all binding assays. The abbreviation "MpAgo" is used for both versions of the protein.

RNA Synthesis and Labeling. RNAs longer than 29 nt were transcribed *in vitro* using synthetic DNA templates carrying a T7 promoter sequence (52). After transcription, the RNA was purified by denaturing 12% polyacrylamide gel electrophoresis (PAGE) and extracted from the gel using the crush-and-soak method followed by ethanol precipitation. The resulting transcribed RNAs were dephosphorylated using alkaline phosphatase CIP (NEB). After the reaction, the CIP was removed by phenol/chloroform extraction and the RNA was purified by ethanol precipitation. The RNAs were resuspended in RNase-free water and stored at −80 °C. Short and modified RNA and all DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). All nucleic acids used in this study are listed in **Table S1**.

For *in vitro* cleavage and binding assays, 10 pmol of RNA oligonucleotide was 5'-radiolabeled using [γ-^{32}P] ATP (Perkin-Elmer) and 5 units of T4 polynucleotide kinase (NEB) in 1× T4 PNK buffer (NEB) at 37 °C for 30 min. The kinase was inactivated by

incubation at 65 °C for 20 min and the labeling reaction was spun through an illustra™ MicroSpin G-25 column (GE Life Sciences) to remove free nucleotides.

For the *in vitro* gRNA binding assay and the competition assays, 10 pmol of gRNA was 3'-radiolabeled using [^{32}P] pCp (Perkin-Elmer) and 10 units of T4 RNA Ligase I (NEB). The labeling was performed overnight at 16 °C. The ligase was inactivated by incubation at 95 °C for 2 min and the reaction was purified from free nucleotides by spinning through an illustra™ MicroSpin G-25 column (GE Life Sciences).

Assembly of the MpAgo RNP Complex. The MpAgo-gRNA complex was prepared by mixing MpAgo and gRNA with a ratio of 1:1.1 in RNP assembly buffer (20 mM Tris-HCl, 150 mM KCl, 1 mM $MnCl_2$, 2 μg/mL heparin, 10 μg/mL BSA, 5% (v/v) glycerol, 2 mM DTT, pH_{RT} 7.5). After mixing the components, the samples were incubated either at 37 °C for 45 min or at 55 °C for 15 min and cooled down slowly to room temperature.

MpAgo-gRNA Filter-Binding Assays. The filter-binding assays were performed as described previously to obtain equilibrium binding isotherms (53). The gRNA binding experiments were carried out in RNP assembly buffer described above. 0.1 nM of 3'-radiolabeled gRNA was incubated with increasing concentrations of WT MpAgo in a 50 μL reaction at either 37 °C for 45 min or at 55 °C for 15 min. 45 μL of each sample were applied to a dot-blot apparatus with low vacuum and three stacked membranes (Tuffryn (Pall Corporation), Protran (GE Healthcare), Hybond-N^+ (GE Healthcare)) to separate MpAgo-gRNA complexes from free gRNA. Membranes were washed with washing buffer (20 mM Tris-HCl, 100 mM KCl, 5 mM $MgCl_2$, pH_{RT} 7.5), air-dried and visualized by phosphor-imaging. The intensities of the free and MpAgo-bound gRNAs were analyzed using GelEval 1.35 (FrogDance Software). The obtained data of three independent experiments was plotted

as function of MpAgo concentration and fit with a standard binding isotherm using the Prism 6 program (GraphPad Software, Inc.).

MpAgo Unloading Assay. The MpAgo RNP was prepared by mixing catalytically inactive MpAgo and biotinylated gRNA at a ratio of 1:2 in RNP assembly buffer. The samples were incubated either at 37 °C for 45 min or at 55 °C for 15 min and cooled down slowly to room temperature. 30 pmol of reconstituted RNP were combined with 60 μL of magnetic beads (Dynabeads MyOne Streptavidin M280; Life Technologies) in bead-binding buffer (10 mM Tris-HCl, 150 mM KCl, 2 mM $MgCl_2$, 5% (v/v) glycerol, 0.05% (v/v) Tween 20, pH_{RT} 7.5). The binding reaction was incubated at 25 °C for 30 min on a rotating mixer. The beads were placed on a magnetic rack to remove the supernatant and were washed 5 times by adding 500 μL of bead-binding buffer followed by 5 min incubation on a rotating mixer. The beads with immobilized MpAgo RNP were resuspended in 60 μL of bead-binding buffer. Then the sample was split three ways. Bead-binding buffer, fully complementary target site containing mRNA, or total RNA was added to the three samples, to a final volume of 50 μL for each reaction. After incubation at 37 °C for 1 h, the samples were placed on a magnetic rack. The flow-through fractions were saved for SDS-PAGE analysis. The beads were washed 3 times with 500 μL each wash of bead-binding buffer. To elute MpAgo RNPs, the beads were resuspended in 20 μL of 2× SDS-PAGE loading buffer (100 mM Tris-HCl, 4% (w/v) SDS, 20% (v/v) glycerol, 200 mM DTT, 0.2% (w/v) Bromophenol blue, pH_{RT} 6.8) and boiled at 95 °C for 5 min. The supernatant (elution sample) was separated from the beads using a magnetic rack. The flow-through and elution samples were analyzed by SDS-PAGE.

Competition-Binding Assay. The MpAgo-gRNA complex was reconstituted by incubating WT MpAgo and 3'-end radiolabeled gRNA at 55 °C for 15 min in RNP assembly buffer. After adding increasing concentrations of unlabeled 5'-OH or 5'-PO_4 gRNA, the

competition-binding reactions were overlaid with 30 μL of mineral oil (Sigma) to avoid evaporation and were incubated overnight at 37 °C. 45 μL of each sample was applied to a dot-blot apparatus with low vacuum and three stacked membranes (Tuffryn (Pall Corporation), Protran (GE Healthcare), Hybond-N^+ (GE Healthcare)) to separate the MpAgo-gRNA complexes and the free gRNA. The membranes were washed with washing buffer, air-dried, and visualized by phosphor-imaging. The intensities of free and MpAgo-bound gRNA were analyzed using GelEval 1.35 (FrogDance Software). The obtained data of three independent experiments was plotted as a function of logarithm of competitor (either 5'-OH or 5'-PO_4 gRNA) concentration and fit with a one-site competitive binding equation using the Prism 6 program (GraphPad Software, Inc.), and the IC_{50} values (i.e. the concentration of competitor needed to displace 50% of the gRNA from the RNP) were determined.

Single Turnover *In Vitro* Cleavage Assays. For the single turnover experiments, the MpAgo–gRNA complexes were reconstituted by mixing 5 nM MpAgo with 5.5 nM gRNA in cleavage buffer (20 mM Tris-HCl, 150 mM KCl, 2 mM $MnCl_2$, 5% (v/v) glycerol, 2 mM DTT, pH_{RT} 7.5) and incubating at 55 °C for 15 min. The cleavage reactions were initiated by adding 5'-radiolabeled ssRNA substrates to a final concentration of 0.5 nM and then incubated at 58 °C. 15 μL aliquots were removed at the 0, 0.5, 1, 2, 5, 10, 20, and 30 min time points and quenched by mixing them with an equal volume of formamide gel loading buffer (95% formamide, 100 mM EDTA, 0.025% SDS, and 0.025% (w/v) Bromophenol Blue). Samples were heated to 90 °C for 2 min, resolved on 13% denaturing polyacrylamide gel, and visualized by phosphor-imaging. Assays were performed in three independent replicates, and the intensities of the uncleaved and cleaved RNA were analyzed using GelEval 1.35 (FrogDance Software). The mean ± standard deviation (SD) at each time point of the three independent experiments was plotted as a function of time and the data were fit with an exponential one-phase decay curve using the Prism 6 software (GraphPad Software,

Inc.). The first-order cleavage rate constants were determined and reported as apparent k_{obs}, because the nature of the rate-limiting step is not known.

ssRNA Substrate Filter-Binding Assay with MpAgo RNPs. Uncrosslinked MpAgo-gRNA complexes were reconstituted by incubating catalytically inactive MpAgo and gRNA at 55 °C for 15 min in RNP assembly buffer with a ratio of 1:1.1, if not specified otherwise. The binding reaction was performed in substrate binding buffer (20 mM Tris-HCl, 150 mM KCl, 2 mM $MgCl_2$, 5 % (v/v) glycerol, 2 mM DTT, 10 µg/mL BSA, 2 µg/mL heparin, pH_{RT} 7.5) by mixing 200 nM (if not specified otherwise) of MpAgo RNP and either 0.1 nM of 5'-radiolabeled complementary ssRNA substrate or non-complementary ssRNA substrate to assess MpAgo RNP's binding specificity. To obtain MpAgo RNP – substrate RNA equilibrium binding isotherm, 0.1 nM of 5'-radiolabeled RNA substrate was incubated with increasing concentrations of preassembled MpAgo RNP in a final volume of 50 µL.

The binding reactions were incubated for 1 h at 37 °C. 45 µL of each sample was applied to a dot-blot apparatus with low vacuum and three stacked membranes (Tuffryn (Pall Corporation), Protran (GE Healthcare), Hybond-N^+ (GE Healthcare)) to separate the free ssRNA substrate and the ssRNA substrate bound to the MpAgo RNP. Membranes were washed with washing buffer, air-dried, and visualized by phosphor-imaging. The intensities of the free ssRNA and the ssRNA bound to the MpAgo RNP were analyzed using GelEval 1.35 (FrogDance Software).

Preparation of the Crosslinked MpAgo RNP (MpAgo xRNP). The MpAgo-gRNA complex was prepared by mixing MpAgo and guide RNA that contains 5-bromo-deoxyU at the 5'-terminal 1st position with a ratio of 1:1.5 (if not specified otherwise) in RNP assembly buffer described above. After mixing the components, the sample was incubated at 55 °C for 15 min and cooled down slowly to room temperature. Next, crosslinking was performed by

irradiating reconstituted MpAgo RNP in an uncapped 0.5 mL tube at ~ 305 nm for 1 h on ice (using Hg-UVB lamp bulbs and Spectrolinker XL-1500 UV Crosslinker (Spectronics Corporation Spectroline)).

MpAgo xRNP binding assays with ssRNA substrates. Under conditions of high xRNP concentrations, like those used for non-crosslinked RNPs (0.1 nM of the substrate and 200 nM of the xRNP), crosslinking improved ON target substrate binding efficiency by ~ 20% while the OFF target binding remained unchanged (**Fig. 2D, S4B**). The binding of the OFF target was reduced from ~ 15% to ~ 7% by increasing the MpAgo:gRNA ratio from 1:1.1 to 1:1.5 during the reconstitution of the MpAgo xRNP, with further increases in the protein:gRNA ratio failing to improve OFF target binding (**Fig. S4B**). We therefore used the 1:1.5 ratio of MpAgo:gRNA for reconstitution and crosslinking in all further experiments.

The ON target binding efficiency was further improved up to ~ 51% while non-specific OFF target binding was reduced to ~ 3% by supplementing the reaction with 10 μg/mL of heparin. Further increases in heparin concentration diminished not only OFF target binding but also ON target binding (**Fig. S4C**). Thus, we used 10 μg/mL of heparin in subsequent experiments to assess ssRNA binding.

The optimal crosslinked MpAgo xRNP concentration for binding specificity and efficiency was determined by ssRNA substrate filter-binding assays performed at 37 °C, with 0.1 nM of either fully complementary substrate or non-complementary substrate in the presence of 10 μg/mL heparin. The lowest concentration of the xRNP tested (10 nM) provided the most specific binding (**Fig. S4D**).

To obtain MpAgo xRNP – substrate RNA equilibrium binding isotherms, 10 pM of 5'-radiolabeled RNA substrate was incubated with increasing concentrations of preassembled

MpAgo xRNP in a final volume of 100 μL. The binding reactions were incubated for 1 h at 37 °C. 95 μL of each sample was applied to a dot-blot apparatus with low vacuum and three stacked membranes (Tuffryn (Pall Corporation), Protran (GE Healthcare), Hybond-N^+ (GE Healthcare)) to separate the free ssRNA substrate and the ssRNA substrate bound to the MpAgo xRNP. Membranes were washed with washing buffer, air-dried, and visualized by phosphor-imaging. The intensities of the free ssRNA and the ssRNA bound to the MpAgo xRNP were analyzed using GelEval 1.35 (FrogDance Software).

Isolation of A-to-I Edited ssRNA Substrates from a Complex Mixture. The MpAgo-gRNA complex was reconstituted by incubating catalytically inactive MpAgo and gRNA at 55 °C for 15 min in the RNP assembly buffer with a ratio of 1:1.5. Next, crosslinking was performed by incubating reconstituted MpAgo RNP under UVB light (λ_{max} ~ 305 nm) for 1 h on ice. The MpAgo xRNP binding reaction was performed in substrate binding buffer (20 mM Tris-HCl, 150 mM KCl, 2 mM $MgCl_2$, 5 % (v/v) glycerol, 2 mM DTT, 10 μg/mL BSA, pH_{RT} 7.5) by mixing 0.5 nM of the MpAgo xRNP and 0.1 nM of 5'-radiolabeled ssRNA target containing either I or A, in the presence of 500 ng of total RNA purified from HEK 293T cells and 200 ng/μL of yeast tRNA. The binding reactions were incubated for 1 h at 37 °C. 45 μL of each sample was applied to a dot-blot apparatus with low vacuum and three stacked membranes (Tuffryn (Pall Corporation), Protran (GE Healthcare), Hybond-N^+ (GE Healthcare)) to separate the free ssRNA substrate and the ssRNA substrate bound to the MpAgo xRNP. Membranes were washed with washing buffer, air-dried, and visualized by phosphor-imaging. The intensities of the free ssRNA and the ssRNA bound to the MpAgo xRNP were analyzed using GelEval 1.35 (FrogDance Software).

SI FIGURES

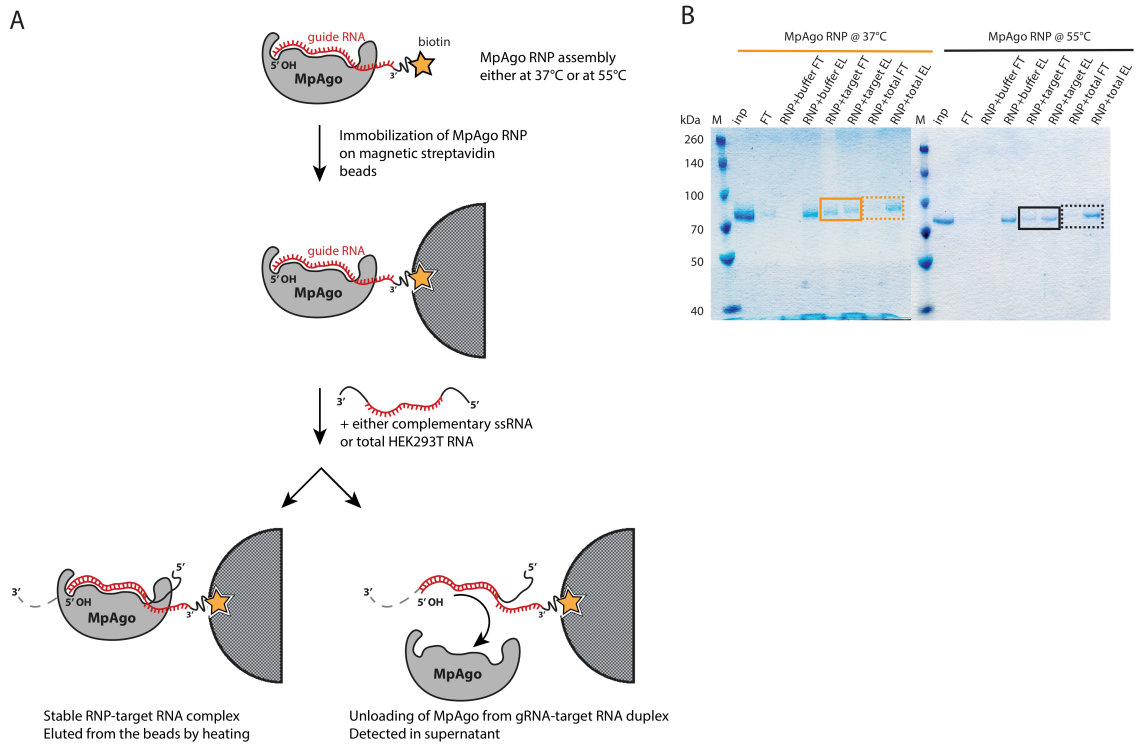


Figure S1. *In vitro* reconstituted MpAgo RNP destabilized upon complementary substrate RNA binding. (A) A scheme depicting the MpAgo unloading assay. MpAgo RNP was assembled either at 37 °C or at 55 °C, using catalytically inactive MpAgo and biotinylated guide RNA, and then immobilized on streptavidin-coated magnetic beads. The beads were washed 5× with bead-washing buffer to remove unbound RNPs. Next, the immobilized RNPs were incubated with only the buffer, the fully complementary substrate RNA (ON), or a pool of non-complementary RNAs (OFF). The flow-through fractions were saved for SDS-PAGE analysis. The beads were washed 3× with bead-washing buffer. MpAgo RNPs bound to the beads were eluted by boiling. The flow-through and elution samples were analyzed by 12% SDS-PAGE. **(B)** SDS gel of MpAgo bound to and released from the gRNA upon adding the complementary ssRNA substrate. The MpAgo unloading

assay was performed to test the temperature at which the assembled RNP is more stable in the presence of the ssRNA substrate. The collected flow-through (FT) and elution (E) fractions were resolved on 12% SDS-PAGE. MpAgo RNPs assembled at 37 °C (solid, orange box); MpAgo RNPs assembled at 55 °C (solid, black box). No MpAgo protein was detected in the FT fraction after adding non-specific RNAs (dashed, orange and black boxes).

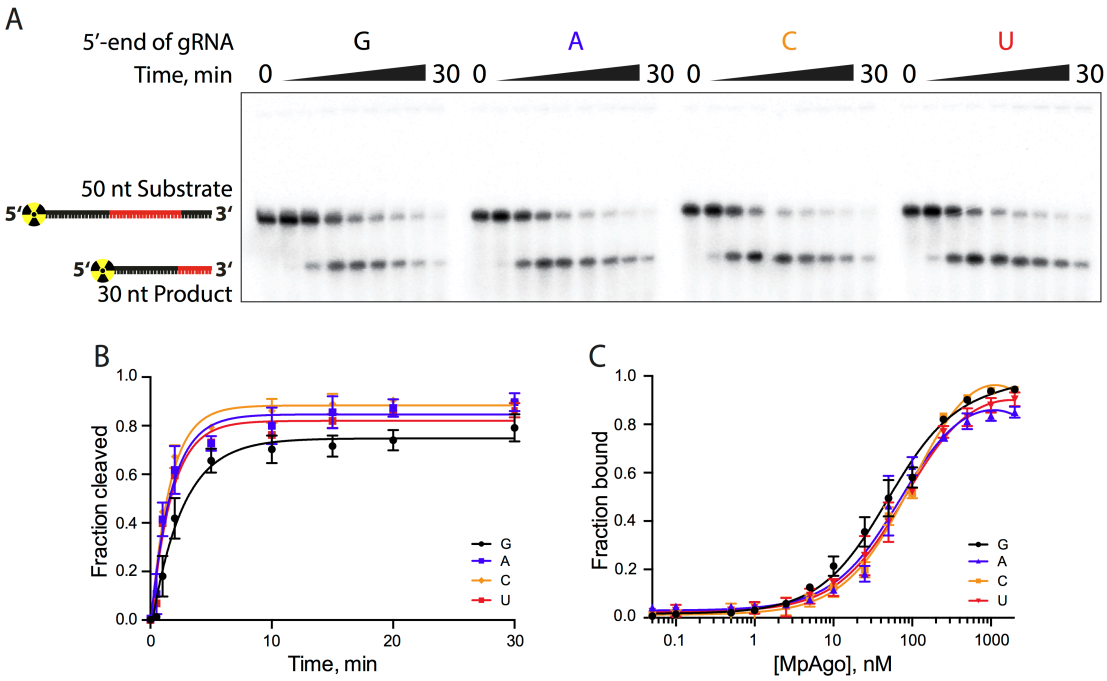


Figure S2. The role of the 5'-terminal nucleotide of the gRNA in substrate ssRNA cleavage by MpAgo. (A) A representative polyacrylamide denaturing gel of the 5'-radiolabeled ssRNA substrate cleavage reactions performed with MpAgo RNP programmed with a gRNA containing 5'-treminal guanine (G), adenine (A), cytosine (C) or uracil (U). **(B)** The fraction of cleaved ssRNA substrate plotted as a function of time. The data were fit with a single exponential to extract the apparent cleavage rate constants (k_{obs}): 5'-G k_{obs} = 0.39 ± 0.05 min^{-1}; 5'-A k_{obs} = 0.59 ± 0.09 min^{-1}; 5'-U k_{obs} = 0.59 ± 0.07 min^{-1}; and 5'-C k_{obs} = 0.64 ± 0.09 min^{-1}. Data are represented as the mean ± SD from three independent experiments. **(C)** Binding curves of the MpAgo and the 5'-G, 5'-A, 5'-U, or 5'-C guide RNA obtained from filter binding assays at 55 °C. The fraction of bound gRNA was plotted as a function of the MpAgo concentration. The obtained average K_d's for MpAgo and the gRNAs with 5'-guanine, adenine, uracil or cytosine are: 5'-G K_d = 49 ± 5 nM; 5'-A K_d = 73 ± 10 nM; 5'-U

K_d = 81 ± 9 nM; and 5'-C K_d = 104 ± 9 nM. Data are represented as the mean ± standard deviation (SD) from three independent experiments.

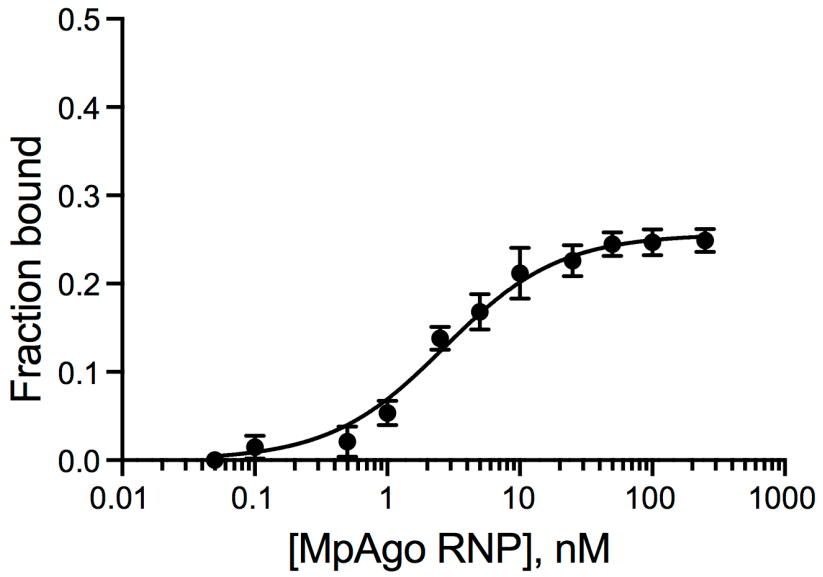


Figure S3. The binding affinity of non-crosslinked MpAgo RNP for ssRNA substrate. The ssRNA filter-binding assays were performed with non-crosslinked MpAgo RNP and the ssRNA substrate with 21-nt complementarity. The fraction of bound ssRNA substrate was plotted as a function of the MpAgo RNP concentration. The obtained K_d is 2.6 ± 0.4 nM. Data are represented as the mean ± SD from three independent experiments.

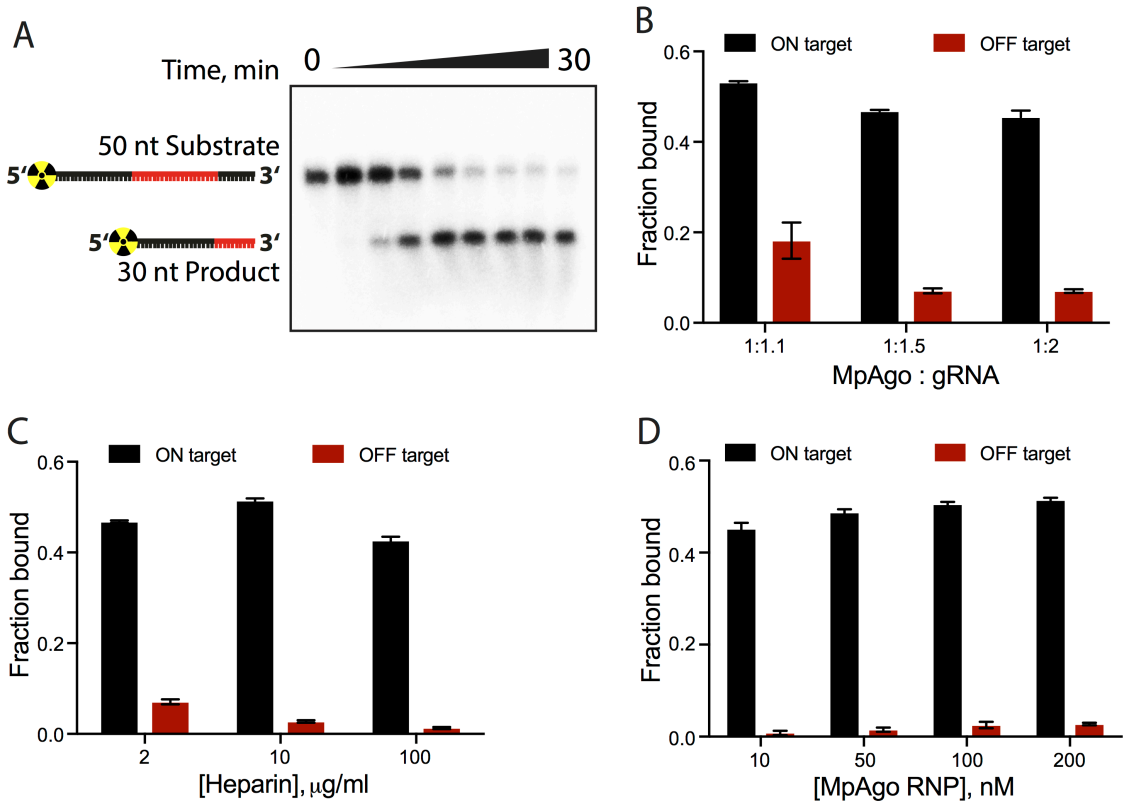


Figure S4. ssRNA cleavage and binding by crosslinked MpAgo xRNP. (A) A representative denaturing gel of the 5'-radiolabeled ssRNA cleavage reaction performed with cosslinked MpAgo xRNP. **(B)** Optimal MpAgo:gRNA for reconstitution of crosslinked MpAgo xRNP, determined using ssRNA substrate filter-binding assays performed at 37 °C, where the preassembled MpAgo xRNPs were incubated with either fully complementary substrate (ON target, black) or non-complementary substrate (OFF target, red) at 0.1 nM concentration. The fraction of the ssRNA substrate bound to the RNP was quantified and plotted as the mean ± SD from three independent experiments. **(C)** Optimal heparin concentration in the ssRNA binding reaction determined by ssRNA substrate filter-binding assays performed at 37 °C, with 0.1 nM of either fully complementary substrate (ON target, black) or non-complementary substrate (OFF target, red) at three heparin concentrations: 2 μg/mL, 10 μg/mL and 100 μg/mL. The fraction of the ssRNA substrate bound to the RNP

was quantified and plotted as the mean ± SD from three independent experiments. **(D)** The optimal crosslinked MpAgo xRNP concentration determined by ssRNA substrate filter-binding assays performed at 37 °C, with 0.1 nM of either fully complementary substrate (ON target, black) or non-complementary substrate (OFF target, red) in the presence of 10 μg/mL heparin. The fraction of the ssRNA substrate bound to the RNP was quantified and plotted as the mean ± SD from three independent experiments.

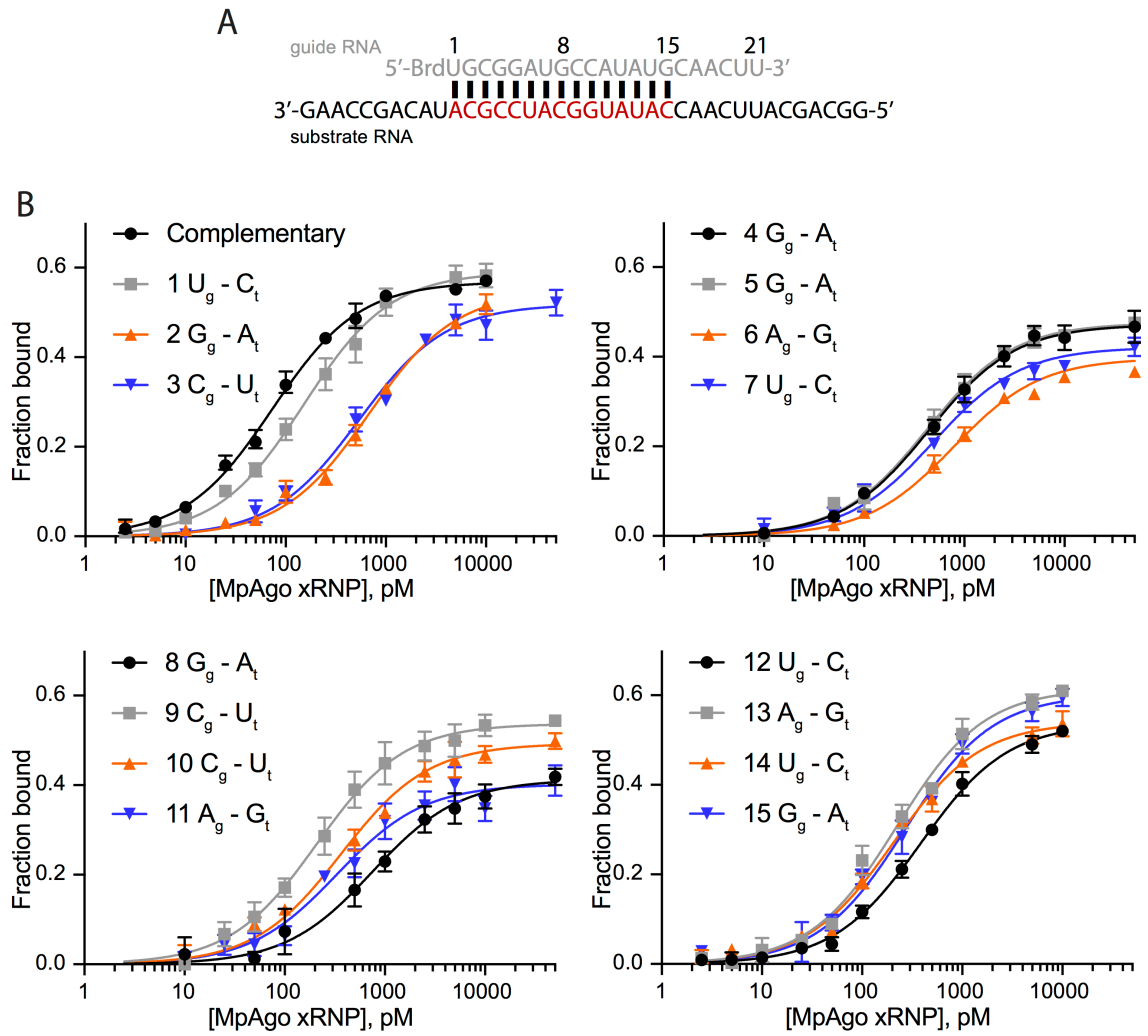


Figure S5. Binding of crosslinked MpAgo xRNP and the ssRNA substrates containing single mismatches across the gRNA. (A) A schematic depiction of the sequences of the 5-bromo-U modified gRNA (grey) and the complementary ssRNA substrate (black and complementary region in red) used for binding assays. Single nucleotide mismatches were introduced across the red region keeping the sequence of the gRNA constant. **(B)** The binding curves of crosslinked MpAgo xRNP and the ssRNA substrates containing mismatches at various positions of the gRNA from the 1st to the 15th position. The position and nature of the mismatches are indicated in the legends of the plots, where a nucleotide with a subscript 'g'

is located in the gRNA and a nucleotide with a subscript 't' is located in the ssRNA target strand. The data were fit with a standard binding isotherm (solid lines) to extract the binding affinities. The ratios of obtained average dissociation constants for the purine-purine mismatches (K_d^{mm}) to the fully complementary substrate (K_d^{Compl}) were calculated and plotted against the position (1st, 2nd, 4th, 5th, 6th, 8th, 11th, 13th, and 15th) of the mismatch in **Fig. 4B**.

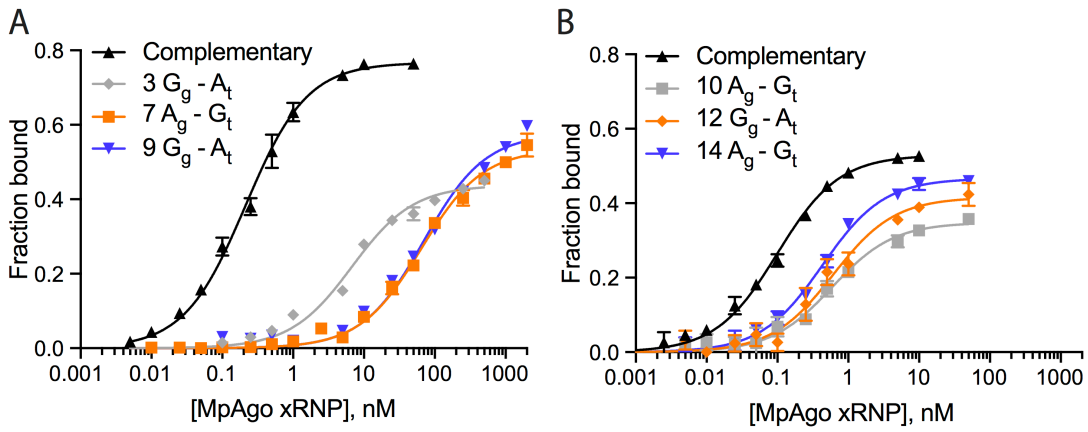


Figure S6. Binding of crosslinked MpAgo xRNPs and ssRNA substrates containing purine-purine mismatches across the guide RNA. To obtain purine-purine mismatches at all 15 positions, we used two additional gRNAs (**Table S1**) to program MpAgo. **(A)** Binding curves of crosslinked MpAgo xRNP and ssRNA substrates that create purine-purine mismatches at the 3rd, 7th and 9th positions of the gRNA. **(B)** Binding curves of crosslinked MpAgo xRNP and ssRNA substrates that create purine-purine mismatches at the 10th, 12th and 14th positions of the gRNA. The position and nature of the mismatches are indicated in the legends of the plots, where a nucleotide with a subscript 'g' is located in the gRNA and a nucleotide with a subscript 't' is located in the ssRNA target strand. The data were fit with a standard binding isotherm (solid lines) to extract the binding affinities. The ratios of the obtained average dissociation constants for the purine-purine mismatches (K_d^{mm}) to the fully complementary substrate (K_d^{Compl}) were calculated and plotted against the position of the mismatch in **Fig. 4B**.

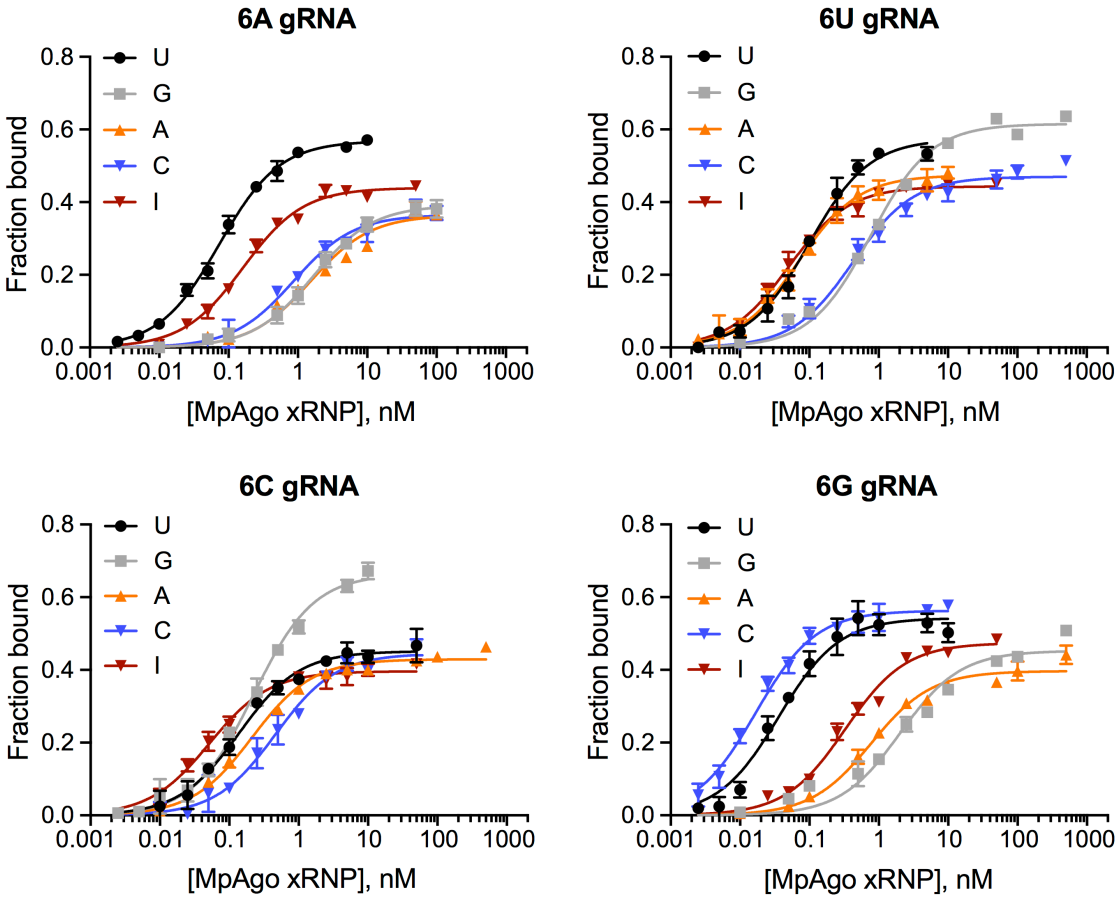


Figure S7. Binding of crosslinked MpAgo xRNP programmed with gRNAs containing permutations at the 6th position and ssRNA substrates containing permutations at the corresponding position. The ssRNA filter-binding assays were performed using MpAgo xRNPs programmed with A, G, U or C gRNA at the 6th position and ssRNA targets containing U, C, A, G or I at the corresponding position. The average K_d's were extracted and are represented in a heat map **(Fig. 5A)**.

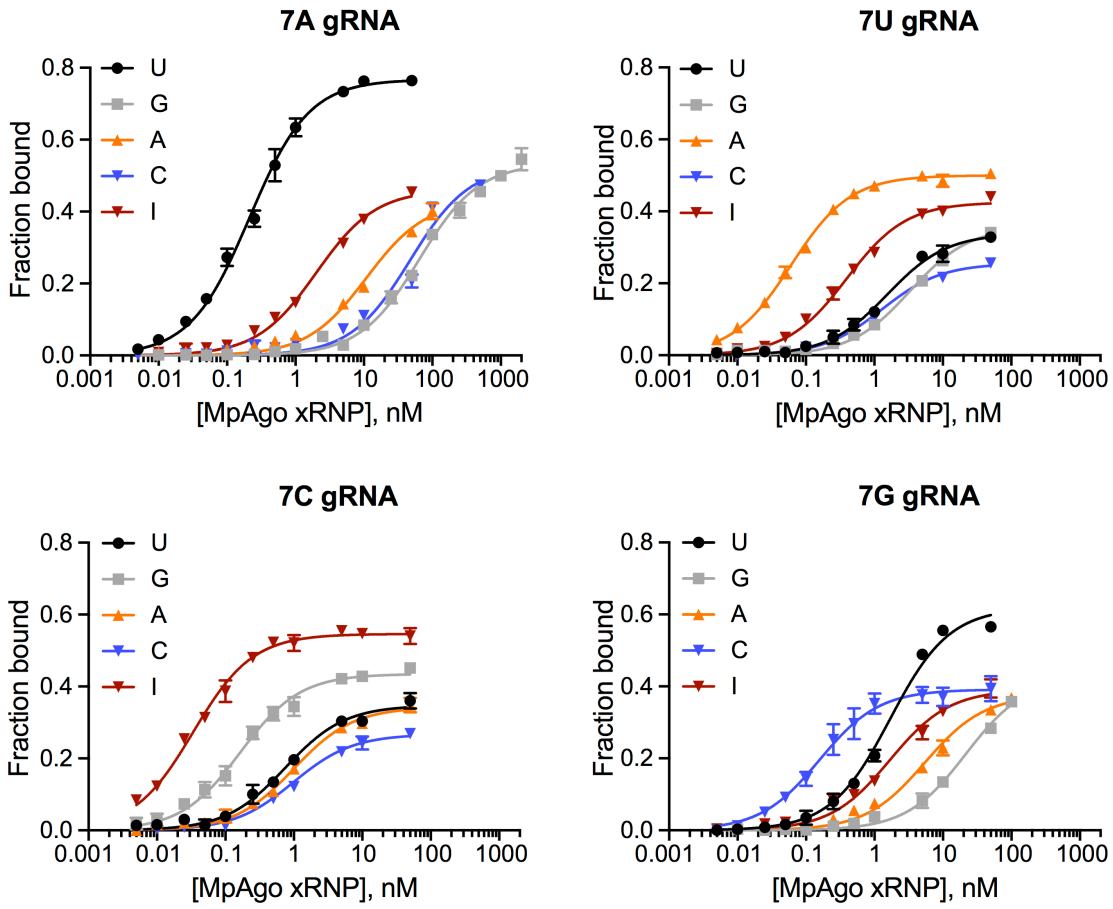


Figure S8. Binding of crosslinked MpAgo xRNP programmed with gRNAs containing permutations at the 7th position and ssRNA substrates containing permutations at the corresponding position. The ssRNA filter-binding assays were performed using MpAgo xRNPs programmed with A, G, U or C gRNA at the 7th position and ssRNA targets containing U, C, A, G or I at the corresponding position. The average K_d's were extracted and represented in a heat map **(Fig. 5B)**.

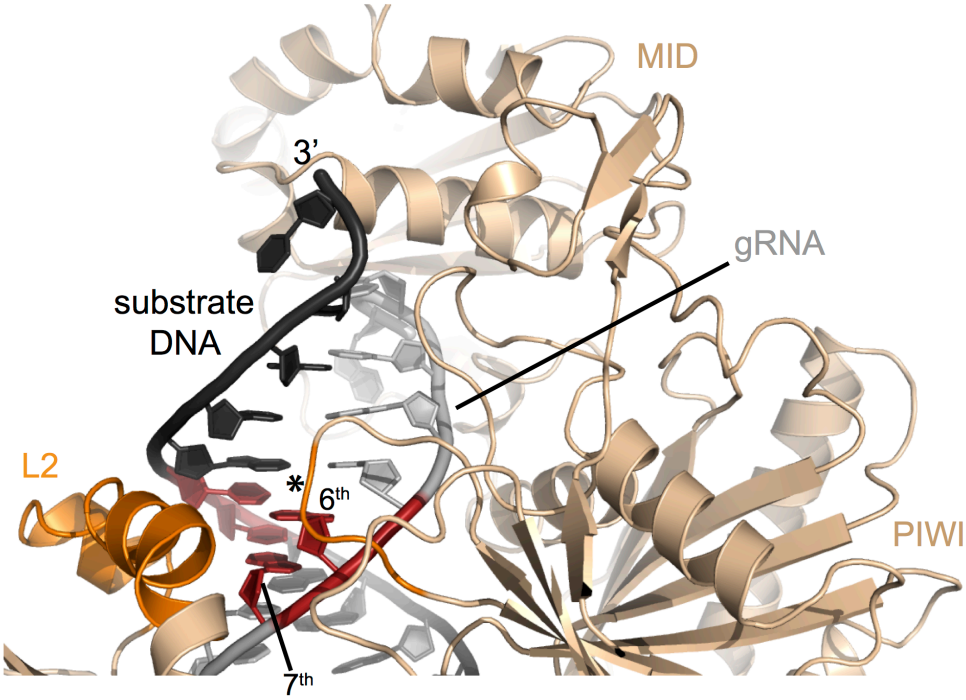


Figure S9. Structure of MpAgo RNP bound to a ssDNA substrate. MpAgo (light orange) bounds to its gRNA (grey) and a fully complementary ssDNA substrate (black). Two helices of the L2 domain and a loop (asterisk) of the PIWI domain (dark orange) probe the shape of the minor groove of the gRNA-ssDNA duplex at the 6th and 7th position (red) of the gRNA. Structural model derived from PDB coordinates 5UX0 (24).