HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2013 December 06.

Published in final edited form as:

Nature. ; 486(7403): 368–374. doi:10.1038/nature11211.

Structure of yeast Argonaute with guide RNA

Kotaro Nakanishi[1,4], **David E. Weinberg**[2,3,4], **David P. Bartel**[2,3,*], and **Dinshaw J. Patel**[1,*]

[1]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

[2]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA

[3]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Abstract

The RNA-induced silencing complex, comprising Argonaute and guide RNA, mediates RNA interference. Here we report the 3.2 Å crystal structure of *Kluyveromyces* Argonaute (*Kp*AGO) fortuitously complexed with guide RNA originating from small-RNA duplexes autonomously loaded and processed by recombinant *Kp*AGO. Despite their diverse sequences, guide-RNA nucleotides 1–8 are positioned similarly, with sequence-independent contacts to bases, phosphates and 2′-hydroxyl groups pre-organizing the backbone of nucleotides 2–8 in a near–A-form conformation. Compared with prokaryotic Argonautes, *Kp*AGO has numerous surface-exposed insertion segments, with a cluster of conserved insertions repositioning the N domain to enable full propagation of guide–target pairing. Compared with Argonautes in inactive conformations, *Kp*AGO has a hydrogen-bond network that stabilizes an expanded and repositioned loop, which inserts an invariant glutamate into the catalytic pocket. Mutation analyses and analogies to Ribonuclease H indicate that insertion of this glutamate finger completes a universally conserved catalytic tetrad, thereby activating Argonaute for RNA cleavage.

RNA interference (RNAi) is a eukaryote-specific gene-silencing pathway triggered by double-stranded RNA (dsRNA)[1–3]. In this pathway, the RNase III enzyme Dicer first cleaves the dsRNA trigger into small interfering RNAs (siRNAs), which have 5′-monophosphates and pair to each other with 2-nucleotide (nt) 3′ overhangs[4–6]. The siRNA duplex is incorporated into the effector protein Argonaute (AGO), whereupon one of the strands (designated the passenger strand) is cleaved[7–9]. After the cleaved passenger strand is

*Correspondence and requests for materials should be addressed to D.J.P. (pateld@mskcc.org) or D.P.B. (dbartel@wi.mit.edu).

[4]These authors contributed equally to this work

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions All authors designed the study and wrote the manuscript. Structural experiments were performed by K.N. under the supervision of D.J.P. Biochemical experiments were performed by D.E.W. under the supervision of D.P.B.

Author Information The structural coordinates of *Kp*AGO have been deposited in the Protein Data Bank (http://www.rcsb.org/pdb) under accession code 4F1N. RNA-sequencing data have been deposited in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE37725. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests.

discarded, the resulting ribonucleoprotein complex (the RNA-induced silencing complex, or RISC) uses the remaining siRNA strand (designated the guide strand) to specify interactions with target RNAs[10,11]. If sequence complementarity between guide and target is extensive, AGO again catalyzes cleavage, resulting in 'slicing' of the target RNA[12].

The first structures of full-length AGOs were of prokaryotic proteins from *Pyrococcus furiosus* (*Pf*AGO)[12] and *Aquifex aeolicus* (*Aa*AGO)[13]. Early structures revealed that the PIWI domain adopts an RNase H–like fold, thereby implicating AGO as the 'slicer' enzyme that mediates RNAi[12,14,15]. Because these prokaryotic enzymes bind 5′-phosphorylated guide DNAs rather than RNA[16,13], subsequent structures featured the binary complex of *Thermus thermophilus* Ago (*Tt*AGO) with guide DNA[17] and ternary complexes with target RNAs of varying length[18,19]. These studies shed light on the nucleation, propagation and cleavage steps of the AGO catalytic cycle[20,19]. However, the physiological role of prokaryotic AGOs is enigmatic; the origin of the guide DNA is unknown and bacteria lack recognizable components of the RNAi pathway[21]. Therefore, attention has turned to eukaryotic AGOs, which utilize RNA guides and have protein-binding partners absent in bacteria[22]. Eukaryotic AGOs are larger than prokaryotic AGOs, because of additional insertion elements of unknown structure and function. Structures of individual domains and the MID-PIWI lobe within eukaryotic AGO have been determined[23–28], but structural characterization of the entire protein has remained a challenge.

Although *Saccharomyces cerevisiae* lacks RNAi, some closely related budding yeast species have retained RNAi, thereby offering fresh possibilities for the study of the eukaryotic pathway[29]. We previously determined the structure and mechanism of Dicer from the budding yeast *Kluyveromyces polysporus*[30] and thus turned our attention to the AGO of this species.

Cleavage activity of budding-yeast AGO

K. polysporus AGO (Ago1) has the four conserved domains (N, PAZ, MID, PIWI) and two linker regions (L1, L2) found in other AGOs (Fig. 1a). It also has an N-terminal extension, predicted to be disordered, which we removed to facilitate crystallization. The resulting protein, *Kp*AGO, can substitute for the full-length protein when reconstituting RNAi in *S. cerevisiae* (Fig. 1b).

*Kp*AGO and other budding yeast AGOs have acidic side chains at the three positions corresponding to active-site residues in slicing-competent AGOs[31] (Supplementary Fig. 1), which suggested that *Kp*AGO might also cleave target RNAs. Indeed, after incubation with a single-stranded guide RNA, recombinant *Kp*AGO cleaved a matched target RNA at the expected position (Fig. 1c). To examine whether slicing occurs *in vivo*, we performed degradome sequencing from another RNAi-containing yeast, *Saccharomyces castellii*. This procedure identifies polyadenylated RNAs containing 5′-monophosphates, including products of AGO-catalyzed slicing[32,33]. Many *AGO1*-dependent degradome tags mapped to Y′-element transcripts (major targets of *S. castellii* RNAi[29]) and tended to pair to endogenous siRNAs in the register implicating cleavage across from positions 10–11 of the guide RNA, which was diagnostic of slicing[34] (Supplementary Fig. 2). These results

indicate that budding-yeast AGO functions as a slicer during endogenous RNAi, and with the *in vitro* results establish *Kp*AGO as a eukaryotic slicer suitable for structure-function analyses.

Structural architecture of eukaryotic AGO

We crystallized *Kp*AGO purified from *E. coli*. Extensive screening identified several crystals that were free of twinning, one of which diffracted to 3.2 Å resolution. A crystal of selenomethionine-substituted *Kp*AGO yielded reflections suitable for phasing by single-wavelength anomalous dispersion (Supplementary Table 1; representative electron density is shown in Supplementary Fig. 3).

The overall structure of *Kp*AGO resembles the bilobal architecture of its prokaryotic counterparts but with expansions throughout the protein (Fig. 2a and Supplementary Fig. 4). Of the 19 insertion segments not found in prokaryotic AGOs[12,13,17], 11 were conserved segments (cS) found in all eukaryotic AGOs, albeit with some differences in secondary structure and/or length, whereas the remaining eight were variable segments (vS) found in only some eukaryotic AGOs (Supplementary Fig. 1). All insertion segments are external, generating new surfaces for potential interactions with AGO-binding proteins.

Autonomously loaded guide RNA

After modeling the *Kp*AGO protein, the *F*o–*F*c map revealed continuous residual electron density lying along the nucleic acid–binding channel (Fig. 2b). This unanticipated density resembled that of an oligonucleotide and could be fit well with an RNA octamer (Fig. 2c and Supplementary Fig. 5a). Analysis of end-labeled polynucleotides extracted from soluble and crystalline *Kp*AGO confirmed the presence of small RNAs (Fig. 2d and Supplementary Fig. 5b), the high-throughput sequencing of which identified a diverse population with a bimodal length distribution centering at 12 and 17 nt (Supplementary Table 2 and Supplementary Fig. 5c–d).

The location of the co-purifying RNAs suggested that they might represent functional guide RNAs. Supporting this interpretation, they had two features of budding yeast guide RNAs: 5′ uridine enrichment (Fig. 2e) and the presence of 5′ monophosphate, indicated by both electron density (Fig. 2c) and a phosphatase-sensitive block of 5′-end labeling (Supplementary Fig. 5e)[29]. The *Kp*AGO preparation sliced an RNA containing a site complementary to a co-purifying 17-nt RNA comprising ~0.1% of our sequencing reads (Fig. 2f and Supplementary Fig. 5c). Slicing was at the anticipated linkage and sensitive to mismatches to guide nucleotides 10–11. Reactions displayed initial burst kinetics, as observed previously for metazoan AGOs[31, 35–38], although addition of Triton enabled sustained product formation (Supplementary Fig. 5f), perhaps by facilitating a conformational change that promotes product release.

Most co-purifying RNAs mapped to the *Kp*AGO expression plasmid (Fig. 2e and Supplementary Fig. 6), suggesting origins from siRNA-like duplexes loaded into *Kp*AGO with subsequent passenger-strand cleavage (Supplementary Fig. 7). For this to occur, *Kp*AGO must load siRNA duplexes in the absence of RISC-loading factors. Indeed, purified

*Kp*AGO incubated with an siRNA duplex generated products diagnostic of passenger-strand cleavage (Fig. 2g) and formed active RISC able to slice cognate target RNA (Fig. 2h). Loading was more efficient with duplex than with single-stranded guide and occurred asymmetrically in a manner consistent with preference for 5′ uridine on the guide strand (Fig. 2h and Supplementary Fig. 8a–c).

We conclude that *Kp*AGO can autonomously load an siRNA duplex, lose the passenger strand and then slice targets. This conclusion counters the prevailing view that loading of siRNA duplexes to form functional RISC requires RISC-loading factors[11]. We suspect that other AGOs may also autonomously load siRNA duplexes and that reports to the contrary resulted from assaying target-RNA slicing under conditions in which AGO retained inhibitory passenger-strand fragments. Autonomous loading explains how *Kp*AGO RISC fortuitously formed in the absence of other RNAi proteins. In contrast to previous preparations of AGO complexes used for structural studies[17], the formation of *Kp*AGO RISC through loading of a duplex resembles the physiological RISC-assembly pathway. From this perspective, the *Kp*AGO structure reflects the natural state of eukaryotic RISC.

Organization of the guide RNA

Electron density corresponding to the base of nucleotide 1 was smaller than that corresponding to most other positions (Supplementary Fig. 9), which agreed with our sequencing results showing that *Kp*AGO-bound RNAs were diverse but enriched for a 5′ uridine (Fig. 2e). Therefore, we modeled the first nucleotide as uridine and the next seven as adenine (the generic nucleotide used to minimize bias during refinement[39]) and refined the final structure as a *Kp*AGO–pUAAAAAAAp binary complex (Supplementary Table 1 and Supplementary Fig. 10).

The guide-strand nucleotides 2–8 run along the nucleic acid–binding channel, from the MID domain to the L2 domain. These nucleotides, including their bases, have electron-density quality resembling that of the *Kp*AGO protein, even though this density represents a composite of thousands of different RNAs. Thus, for this segment of the guide RNA, known as the seed region, diverse RNA sequences are all presented in essentially the same orientation. The electron density disappeared after the ninth nucleotide (Fig. 2b–c), even though most co-purifying RNAs were longer than 9 nt (Supplementary Fig. 5d). This density loss suggests that guide-RNA 3′ halves are either disordered or adopt diverse sequence-specific conformations. In addition, the PAZ domain is not well ordered, as observed in *Tt*AGO complexes in which the PAZ domain has released the 3′ end of the guide[19], consistent with the idea that *Kp*AGO holds the guide RNA without assistance from the PAZ domain.

Like prokaryotic AGOs[15,17–19,40], *Kp*AGO recognizes the 5′ phosphate of the guide, the notable difference being that *Kp*AGO uses the ammonium group of Lys939 rather than a divalent cation (despite Mn^{2+} in the crystallization buffer) to neutralize the negative charge resulting from the close juxtaposition of the C-terminal carboxylate and phosphates 1 and 3 (Fig. 3a and b). The inserted C-terminus is anchored by Lys939 and Lys943 (Fig. 3a), with mutation of either residue impeding guide RNA binding in *Drosophila* AGO1[28]. Another

distinct facet involves Arg1183, which hydrogen bonds with the C-terminal carboxylate and phosphate 4 (Fig. 3c). In the free *Nc*QDE-2 MID-PIWI structure[28] the analogous arginine is in a disordered loop, suggesting that guide RNA recruits Arg1183 to the 5′-phosphate–binding pocket. Notably, conservation of Lys939 and Arg1183 is restricted to eukaryotic AGOs (Supplementary Fig. 1).

The Asn897 main-chain amide interacts with the O2 carbonyl of the uridine at position 1 (Fig. 3a). Because analogous interactions with O2 of cytidine and N3 of purines would be isosteric, this hydrogen bond cannot explain the preference for a 5′ uridine. The preference might instead be attributed to the relatively weak stacking and pairing of uridine, which would facilitate the requisite flipping out of nucleotide 1 during siRNA loading.

*Kp*AGO interacts with phosphates of the seed region primarily using contacts homologous to those observed in prokaryotic AGO complexes[16–19,40] (Fig. 3c). Structures of prokaryotic complexes, however, have not revealed intermolecular contacts to the guide-RNA 2′-OH groups. We find that *Kp*AGO forms hydrogen bonds with most 2′-OH groups of the seed, using main-chain atoms at positions 2, 5 and 6 and hydroxyl groups of Thr1186 and Tyr681 at positions 4 and 7, respectively (Fig. 3d). We also observe an intra-RNA hydrogen bond between the 2′-OH group at position 3 and O4′ at position 4, a type of interaction proposed to facilitate base-pair fluctuations in A-form RNA helices[41]. A second intra-RNA hydrogen bond involves the 2′-OH group at position 1 and a non-bridging oxygen of phosphate 2, as previously observed in *Af*PIWI–siRNA complex structures[16,40].

To examine the contributions of guide-strand 5′ phosphate and 2′-OH groups, we monitored autonomous loading and passenger-strand cleavage of modified siRNA duplexes. Removing the monophosphate or substituting all guide-strand 2′-OH groups with 2′-H (deoxy) greatly impaired activity (Fig. 3e and Supplementary Fig. 11), consistent with observations in transfected human cells[42,43]. To learn more about the 2′-OH groups contributing to this effect, we compared guide RNAs with deoxy substitutions at positions 1, 2–8, 9–14, 15–21 and 22–23. Substitution of the 2′-OH group at position 1 enhanced activity (perhaps by facilitating flipping out of nucleotide 1), whereas substitutions in all other regions impaired activity (Fig. 3e and Supplementary Fig. 11). Deoxy substitution at positions 2–8 impaired activity to a similar degree as at positions 22–23, which are presumably recognized by the PAZ domain[23–25]. Thus, the 2′-OH groups within the seed region contribute to duplex loading or passenger-strand cleavage. Nonetheless, greater effects were observed at positions 9–14 and 15–21, the understanding of which will require structural studies of additional states along the eukaryotic RISC-assembly pathway.

Together, contacts to the phosphate and 2′-OH groups maintain the sugar–phosphate backbone of the single-stranded guide-RNA seed in a near A-form conformation resembling that of the siRNA duplex (Fig. 3f). Maintaining this conformation pre-organizes the seed backbone for pairing to the target, as anticipated from studies of microRNA targeting[1] and supported by structural and biophysical studies[16,40,44]. Also as anticipated, the bases of the seed nucleotides are stacked, with Watson–Crick faces (particularly those of nucleotides 2–4) displayed to solvent and accessible to nucleate pairing to target RNA (Fig. 3g).

The surprising feature of the guide-RNA conformation was the tilting of the bases away from the orientation required for helical pairing (Fig. 3f). *Kp*AGO makes hydrophobic contacts with the bases at positions 2, 5 and 6 while anchoring the sugar–phosphate backbone (Fig. 3c–d). Base 2 packs against Tyr932 (Fig. 3d), which is conserved as Tyr or Thr in eukaryotic AGOs (Supplementary Fig. 1) and thus might represent a conserved hydrophobic interaction that facilitates the flipping out of nucleotide 1 by preventing its stacking on base 2. As observed in structures of prokaryotic AGO complexes[16–19,40], base 2 is recognized at N3 (purines) or O2 (pyrimidines) by the side chain of Asn935 (Fig. 3d), which is conserved throughout all AGOs. Bases 5 and 6 are surrounded by a hydrophobic pocket comprising Ile682, Ala686, Leu1147 and Lys1148 (Fig. 3d). Bases 3 and 4 make no contact with *Kp*AGO but are nonetheless tilted because of continuous stacking of the seed bases (Fig. 3c–d). Untilting of the seed stack, which would accompany nucleation of target pairing at positions 2–4, might disfavor contacts to Ile682 and neighboring residues, thereby facilitating repositioning of α16, a helix that would otherwise block full seed pairing. Such changes in base tilting and α16 might communicate the presence of target RNA.

Potentially unobstructed guide–target pairing

To compare the architectures of eukaryotic and prokaryotic AGOs, we structurally aligned each domain of *Kp*AGO on its *Tt*AGO counterpart. Except for the N domain, each of the domains superimposed well (Supplementary Fig. 12). The structural difference between the N domains is attributed to cS1, cS3 and vS2 (Fig. 4a–b). cS1 and cS3 cluster together with cS7 and cS10 such that they bury a space observed in prokaryotic AGO structures and concomitantly lengthen the nucleic acid–binding channel[12,13,18,17,19] (Fig. 4c–d). These insertion segments interact with the L2 and PIWI domains through a hydrogen-bond network involving residues that are conserved throughout eukaryotic AGOs (Supplementary Fig. 1 and 13a), suggesting that an extended nucleic acid–binding channel is a common feature of eukaryotic AGOs.

In all crystallized conformations of *Tt*AGO, the N domain blocks the channel and prevents propagation of guide–target pairing beyond position 16 (ref 19) (Fig. 4c and Supplementary Fig. 13c–d). In addition to lengthening the nucleic acid–binding channel, the cS1/3/10 cluster positions the *Kp*AGO N domain such that a slight widening of the channel would allow pairing to propagate to the 3′ end of the guide RNA (Fig. 4d and Supplementary Fig. 13b). The potential for unobstructed propagation of guide–target pairing is consistent with the prevalence of pairing throughout the 3′ region of plant small RNAs that guide target cleavage[45] and the contribution of such pairing to the stability of guide–target association *in vitro*[37].

Glu1013 completes a catalytic tetrad

When comparing the structures of *Kp*AGO and the free *Nc*QDE-2 MID-PIWI lobe[28], we observed striking differences in loops L1 and L2 (Supplementary Table 3). In *Kp*AGO, loop L2 expands by partial unfolding of α25 (Fig. 5a) and packs into a cavity, such that the invariant Glu1013 side chain inserts into the catalytic pocket, near the three Asp residues of the active site (Fig. 5b). This conformation is enabled by the movement of loop L1, which

otherwise blocks access to the catalytic pocket (Fig. 5c). Opening of the loop L1 gate in *Kp*AGO is accompanied by a conformational transition of cS11 and hydrophobic packing between aliphatic side chains on loop L1 and cS11. Notably, deletion of cS11 from *Drosophila* AGO1 inhibits guide RNA binding and abolishes silencing activity[28].

The plugged-in conformation, in which the Glu1013 finger is inserted into the catalytic pocket, is stabilized by an extensive hydrogen-bond network, with Glu1013 bridging His977 and Arg1045, and loop L2 main-chain atoms interacting with Arg1045 and Glu1060 (Fig. 5b). These four residues are conserved throughout eukaryotic AGOs (and even most of the PIWI clade; Supplementary Fig. 1). Glu1013, Arg1045 and Glu1060 are also conserved throughout prokaryotic AGOs, prompting a search for a similar plugged-in conformation in the available structures[12,13,17–19,28]. We found both plugged-in and unplugged conformations of *Tt*AGO, with striking parallels to the eukaryotic hydrogen-bond network and correlated loop movements (Fig. 5a–f). The plugged-in conformation was observed only in complexes in which the PAZ domain released the 3′ end of the guide and *Tt*AGO assumed its catalytically active state (Supplementary Table 3). In contrast, the inactive states —either those of the apo proteins or complexes in which the PAZ domain engages the guide 3′ end—resembled the unplugged conformation. These observations suggest that the plugged-in conformation of loop L2 is correlated with release of the 3′ end of the guide and formation of active RISC. Confirming the functional importance of the plugged-in conformation, mutation of any of the four residues of the *Kp*AGO hydrogen-bond network impaired RNAi (Fig. 5g).

The structures of ternary *Tt*AGO complexes in the plugged-in conformation show the position of loop L2 in the context of guide and target strands. *Tt*AGO loop L2 interacts with the guide DNA at positions 11–15 (Supplementary Table 3)[19]. Moreover, the carboxyl group of the glutamate finger approaches both the 2′-OH of the nucleotide adjacent to the scissile phosphate and one of the two active-site divalent metal ions (Fig. 5h), which suggests that the glutamate finger might act as a catalytic residue. Indeed, simultaneous coordination of the analogous 2′-OH and metal ion is the role of Glu109 in the ‘DEDD’ catalytic tetrad at the active site of *Bacillus halodurans* RNase H1[46]. Although the PIWI domain of AGO has an RNase H fold, only a conserved ‘DDX’ catalytic triad (where ‘X’ is generally Asp or His) had been recognized in AGOs with slicer activity[47,48]. Based on analogy to RNase H, a fourth catalytic residue had been suspected, but previous searches for this missing component had focused on the residues corresponding to Arg1045 and Glu1060[12,47,46,31,13], whose conservation and proximity to the catalytic pocket are now explained instead by their roles in stabilizing the plugged-in conformation (Fig. 5b). In support of the glutamate finger as the missing catalytic residue that helps to coordinate an active-site metal ion (either directly or through outer-sphere contacts), the putative DEDD catalytic tetrads in the plugged-in conformations of both *Tt*AGO and *Kp*AGO are essentially isosteric with the RNase H DEDD tetrad. Moreover, when we assayed RNAi, only mutation of Glu1013 abrograted RNAi to the extent observed for mutation of Asp1046, a previously identified active-site residue (Fig. 5g). Thus, we propose that the glutamate finger constitutes the second residue of a universally conserved RNase H–like DEDX catalytic tetrad at the active site of slicing AGOs.

Our new insights suggest the following model for AGO loading and catalysis. The apo protein in the unplugged conformation binds the siRNA duplex, in part using contacts between the 2-nt overhang of the guide strand and the PAZ domain. As the duplex loads and the 3′ end of the guide stand is released from the PAZ domain, the glutamate finger inserts into the active site, thereby completing the DEDX catalytic tetrad to enable cleavage of the passenger strand. After discarding the passenger-strand fragments, the resulting RISC would remain in a plugged-in conformation resembling that of the current structure and be competent to bind and cleave suitably paired target RNAs.

While our manuscript was in review, a structure of human AGO2 (*Hs*AGO2) with RNA of unknown biochemical origin and function was reported[49]. The authors noted many contacts to the RNA 5′ monophosphate and sugar–phosphate backbone analogous to those of *Kp*AGO. Our inspection of the *Hs*AGO2 structure further revealed that it has an extended nucleic acid–binding channel, an N domain positioned to allow unobstructed guide–target pairing, and a plugged-in glutamate finger that completes a DEDH catalytic tetrad (Supplementary Fig. 14). These similarities indicate that the *Hs*AGO2 structure (for which conserved residues were mutated to improve diffraction) has features of active RISC and that studies of *Kp*AGO will continue to provide insights relevant to metazoan AGOs.

In contrast to RNase H, which forms its active site during initial folding, AGO requires a conformational change to form its active site. What might explain this difference between these two related ribonucleases? The constitutive active site of RNase H is well-suited to its role in nonspecifically cleaving RNA–DNA hybrids, whereas proper AGO function requires high specificity. Coupling siRNA duplex loading (in part through recognition by the PAZ domain) with active-site formation imparts specificity to AGO, thereby preventing it from cleaving any base-paired RNA. After passenger-strand cleavage and removal, activity of the licensed AGO is restricted by its guide RNA. In this way, AGO activity is tightly controlled and spurious endonucleolytic cleavage is prevented. The previous view was that among proteins adopting the RNase H fold, RNase H enzymes were unique in having a catalytic tetrad, whereas the related endonucleases of this protein superfamily (including AGO) were missing the active-site residue corresponding to Glu1013[48]. Our findings revising this view imply that some other proteins for which only a catalytic triad has hitherto been identified (e.g., bacterial UvrC DNA repair protein) might also use the conditional insertion of a “missing” catalytic residue to impart specificity.

METHODS SUMMARY

*Kp*AGO was overexpressed in *Escherichia coli* as a His-Sumo-tagged fusion. Native and SeMet-substituted crystals were obtained by sitting-drop vapor diffusion at 20°C. The phase was determined by the single-wavelength anomalous dispersion method with selenium anomalous signals. Cleavage assays were performed with synthetic or transcribed RNA in 30 mM Tris-HCl pH 7.5, 130 mM KCl, 10 mM NaCl, 1.1 mM $MgCl_2$, 0.1 mM EDTA, 1.3 mM DTT and 5% glycerol, including 0.1% Triton X-100 where indicated. Yeast manipulations, *in vivo* assays and high-throughput sequencing were essentially as described previously[32,29]. Details of all procedures are listed in Methods.

METHODS

Protein purification

DNA encoding *K. polysporus AGO1*(Thr207–Ile1251) was cloned into a modified pRSFDuet vector (Novagen) containing an amino-terminal Ulp1-cleavable His_6-Sumo tag. Protein was overexpressed in *E. coli* BL21(DE3) Rosetta2 (Novagen). Cell extract was prepared using a French Press in Buffer A [10 mM phosphate buffer pH 7.3, 1.5 M NaCl, 25 mM imidazole, 10 mM β-mercaptoethanol (β-M E), 1 mM phenylmethylsulphonyl fluoride] and cleared by centrifugation. The supernatant was loaded onto a nickel column (GE Healthcare) and then washed with Buffer A. The target protein was eluted with a linear gradient of 0.025–1.5 M imidazole. After mixing with Ulp1 protease, the eluted sample was dialyzed against Buffer B (10 mM phosphate buffer pH 7.3, 500 mM NaCl, 20 mM imidazole, 10 mM β-ME) overnight. The digested protein was loaded onto a nickel column to remove the cleaved His_6-Sumo tag. The flow-through sample was dialyzed against Buffer C (5 mM phosphate buffer pH 7.3, 10 mM β-ME) and then loaded onto an SP column (GE Healthcare). The protein was eluted with a linear gradient of 0.0–2.0 M NaCl, mixed with ammonium sulfate (2 M final concentration) and then centrifuged. The supernatant was loaded onto a phenyl-sepharose hydrophobic interaction column (GE Healthcare) in Buffer D (10 mM phosphate buffer pH 7.3, 2 M ammonium sulfate, 10 mM β-ME), and the protein was eluted with a linear gradient of 2.0–0.0 M ammonium sulfate. The eluted protein was dialyzed against Buffer E (300 mM sodium dihydrogen phosphate, 10 mM β-ME) and then loaded onto a MonoQ column (GE Healthcare) in Buffer E. The protein was eluted with a linear gradient of 0.0–2.0 M NaCl. The eluted sample was concentrated by ultrafiltration and loaded onto a HiLoad 200 16/60 column (GE Healthcare) in Buffer F (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT). Purified *Kp*AGO was concentrated to approximately 40 mg ml^{-1} using ultrafiltration and stored at −80°C in Protein Storage Buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl, 5 mM DTT).

Structure determination and refinement

Initial crystals of recombinant *Kp*AGO diffracted poorly but could be improved by addition of 1,4-dioxane to the crystallization buffer. Native crystals of *Kp*AGO were obtained at 20°C by sitting-drop vapour diffusion in 100 mM MIB buffer pH 5.0 (2 Na-malonate : 3 imidazole: 3 borid acid), 3% 1,4-dioxane, 19% PEG3350, 12 mM $MnCl_2$ and 3% ethanol. SeMet-substituted crystals were grown at 20°C by sitting-drop vapour diffusion in 100 mM MIB pH 5.0, 3% 1,4-dioxane, 19% PEG3350, 12 mM $MnCl_2$, 3% ethanol and 9 mM sarcosine. The native and SeMet-substituted crystals of *Kp*AGO were soaked in collection buffer (1.2-fold concentrated reservoir solution) and cryoprotected with 20% glycerol. Both derivative data sets were collected at the Advanced Photon Source NE-CAT beamlines. Data were processed with HKL2000[50]. Data collection and refinement statistics are listed (Supplementary Table 1). A total of 33 selenium sites were found using peak data with HKL2MAP[51] and were used for phase calculation at 4.2 Å resolution with Phaser-EP[52]. The initial phases were improved by solvent flattening, electron density histogram and non-crystallographic symmetry averaging with Parrot and DM[53]. The initial model was built manually with Coot[54] and was improved by iterative cycles of refinement with Phenix[55]. Molecular replacement was performed with MOLREP[56] using the SeMet structure as a

search model. The final model was improved using the native data processed at 3.2 Å. The Ramachandran plot analysis by PROCHECK[53] showed 82.0%, 17.3% and 0.8% of the protein residues in the most favorable, additionally allowed and generously allowed regions, respectively, with no residues in disallowed regions. The simulated-annealing omit map was calculated by CNS[57]. All figures of structures were generated with PYMOL[58].

RNAs

A list of RNA oligonucleotide sequences is provided (Supplementary Table 4). To generate cap-labeled target RNAs, RNA was transcribed *in vitro* with T7 RNA polymerase using DNA oligonucleotide templates. DNase-treated transcripts were purified on a denaturing gel and capped using the ScriptCap m^{7G} Capping System (CellScript) according to the manufacturer's directions, except that high-specific-activity RNA was prepared by omitting GTP and including 5 μl [α-^{32}P]GTP (6000 Ci/mmol), and low-specific-activity RNA was prepared by using a 1500:1 molar ratio of GTP:[α-^{32}P]GTP (6000 Ci/mmol). Cap-labeled RNA was gel purified and quantified by scintillation counting, and 10X stocks were prepared in water supplemented with 1 μM DNA carrier oligonucleotide.

5′-phosphorylated guide RNA and its 2′-deoxy-substituted variants were chemically synthesized (IDT) and gel purified. To prepare 5′ end-labeled RNAs, 5′-OH RNAs were chemically synthesized (Dharmacon), deprotected, purified on a denaturing gel, phosphorylated with [γ-^{32}P]ATP (6000 Ci/mmol) using T4 Polynucleotide Kinase (PNK, NEB) and again gel-purified. To prepare 3′ end-labeled RNAs, 5′-phosphorylated RNAs lacking the terminal nucleotide (i.e., 22-nucleotide variants) were chemically synthesized (IDT), gel-purified, extended using cordycepin 5′-[α-^{32}P]triphosphate (5000 Ci/mmol) and Yeast Poly(A) Polymerase (USB) and again gel-purified.

siRNA duplexes were prepared by annealing synthetic ssRNAs. Complementary RNAs designed to hybridize to generate 21-bp duplexes with 2-nucleotide 3′ overhangs, were combined (using at least 3-fold excess unlabeled RNA) in dsRNA Annealing Buffer (30 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA) and slow-cooled from 90°C to room temperature over >2 hr. Annealed RNAs were separated from ssRNAs on native 20% polyacrylamide gels, and duplexes were eluted from gel slices in 0.3 M NaCl overnight at 4°C, ethanol precipitated and stored in dsRNA Storage Buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.1 mM EDTA). RNA was quantified by scintillation counting, and 10X stocks were prepared in dsRNA Storage Buffer supplemented with 1 μM DNA carrier oligonucleotide.

AGO activity assays

For all biochemical assays, *Kp*AGO was diluted and stored at −20°C in Protein Dilution Buffer (5 mM Tris-HCl pH 7.5, 100 mM NaCl, 2.5 mM DTT, 50% glycerol). The concentration of *Kp*AGO was determined by absorbance at 280 nm. For the slicing assay in Fig. 1c, 1.1 μM *Kp*AGO was pre-incubated with 110 nM guide RNA in 1.1X Reaction Buffer (1X Reaction Buffer: 30 mM Tris-HCl pH 7.5, 130 mM KCl, 1.1 mM $MgCl_2$, 1 mM DTT, 0.1 mM EDTA) for 1 hr at 25°C. To initiate the slicing reaction, 1 μl cap-labeled target RNA (final concentration, 200 nM) was added to 9 μl of the pre-incubated mixture.

Reactions were incubated at 30°C, and 3 μl aliquots were removed at the indicated time and quenched by addition to 12 μl Formamide Loading Buffer (95% formamide, 18 mM EDTA, 0.025% sodium dodecyl sulfate, 0.025% xylene cyanol, 0.025% bromophenol blue). The slicing assay in Fig. 2h was conducted similarly except pre-incubation was performed with 110 nM *Kp*AGO and 50 pM guide RNA, and target was subsequently added to a final concentration of 100 pM. The slicing assay in Fig. 2f was guided by co-purifying RNA and thus did not involve pre-incubation. These reactions contained 1X Reaction Buffer supplemented with 0.1% Triton X-100, 100 nM *Kp*AGO (or an equal volume of Protein Dilution Buffer) and 100 pM target RNA. Reactions were incubated at 30°C, and 5 μl aliquots were removed at the indicated time and quenched by addition to 15 μl Formamide Loading Buffer. The passenger-strand cleavage reactions in Fig. 3e contained 1X Reaction Buffer, 100 μg/ml Ultrapure BSA (Ambion), 10 nM *Kp*AGO and 50 pM substrate. All other passenger-strand cleavage reactions contained 1X Reaction Buffer supplemented with 0.1% Triton X-100, 100 nM *Kp*AGO (or an equal volume of Protein Dilution Buffer) and 50 pM substrate. Reactions were incubated at 30°C, and 5 μl aliquots were removed at the indicated time and quenched by addition to 10–15 μl Formamide Loading Buffer.

To monitor cleavage, RNAs were resolved on denaturing (7.5 M urea) polyacrylamide gels (15% gel for target cleavage using synthetic guide RNA, 20% for target cleavage using co-purifying guide RNA or 22.5% for passenger-strand cleavage), and radiolabeled products were visualized by phosphorimaging (Fujifilm BAS-2500) and quantified using Multi Gauge (Fujifilm). For kinetic analyses, at each time point (t) the fraction product was measured as F_P = product/(product + substrate). Data in Fig. 3e were fit with a smoothed curve using the cubic spline method implemented in KaleidaGraph.

Analysis of co-purifying RNA

To avoid loss of especially small fragments, polynucleotides were extracted without subsequent precipitation. For analysis of pooled crystals, ~100 crystals were collected, stored in Harvest Buffer [100 mM MIB buffer pH 5.0, 3% 1,4-dioxane, 20% PEG 3350, 12 mM $MnCl_2$, 3% ethanol, 6 mM sarcosine, 25% glycerol] and immediately frozen. After thawing, the mixture was diluted with an equal volume of water and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) followed by extraction with chloroform. The aqueous phase was retained and diluted ~1:20 in water for use in labeling reactions or used undiluted to prepare sequencing libraries. For analysis of soluble protein, polynucleotides were similarly extracted from 1.5 nmol *Kp*AGO. For analysis of individual crystals, each single crystal was collected, stored in 1 mM EDTA and immediately frozen. After thawing, the mixture was heated at 90°C for 3 min, chilled on ice for 5 min and extracted with an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, Sigma) followed by two extractions with chloroform. The aqueous phase was retained and used undiluted in labeling reactions.

Prior to 5′ labeling, polynucleotides were dephosphorylated in a 20 μl reaction containing 2 μl diluted polynucleotides (or water) and 1X PNK Buffer (NEB) in the presence or absence of 2 units Thermosensitive Alkaline Phosphatase (TSAP, Promega) for 30 min at 37°C. To inactivate TSAP, the reaction was quenched with 1 μl 220 mM EDTA and incubated at 74°C

for 15 min. 5′ phosphorylation was performed in a 30 μl reaction containing 21 μl heat-inactivated TSAP reaction, 3 units T4 PNK (NEB), 0.04 μl [γ-^{32}P]ATP (8000 Ci/mmol), 0.8 μl 10X PNK Buffer and 1 μl 240 mM $MgCl_2$ for 1 hr at 37°C. Reactions were quenched with an equal volume of 2X urea loading buffer and products were resolved on a denaturing 22.5% polyacrylamide gel. For analysis of nuclease sensitivity, 15 μl aliquots were removed from PNK reactions and incubated with 2 μl RNase I (Ambion) or RQ1 RNase-free DNase (Promega) for 30 min at 37°C prior to gel analysis.

To monitor preparation of sequencing libraries, trace amounts of synthetic 3′-pCp[5′-^{32}P]-labeled 7- and 23-nucleotide RNA internal standards were added to 2 μl undiluted polynucleotides isolated from soluble or crystalline *Kp*AGO (or a water-only mock control). Dephosphorylation was performed in a 30 μl reaction containing 3 units TSAP (Promega) and 1X PNK Buffer (NEB) supplemented with 2 μl manganese-chelating mix (10 mM $MgCl_2$, 10 mM EDTA) for 30 min at 37°C. To inactivate TSAP, the reaction was quenched with 1.5 μl 240 mM EDTA and incubated at 74°C for 15 min. RNA was ligated to pre-adenylated adaptor DNA in a 50 μl reaction containing 32 μl heat-inactivated TSAP reaction, 100 pmol adaptor DNA[59], 45 units T4 RNA Ligase 1 (NEB), 10% PEG8000 (NEB), 2 μl 10X PNK Buffer and 1 μl 390 mM $MgCl_2$ for 2.5 hr at room temperature. After phenol extraction and precipitation, 28–50-nt ligation products were gel purified and 5′ phosphorylated in a 50 μl reaction containing 20 units T4 PNK (NEB) and 1X PNK Buffer supplemented with 1 μl [γ-^{32}P]ATP (6000 Ci/mmol) for 30 min at 37°C, followed by a chase with 10 μl cold reaction mixture [1X PNK Buffer, 28 units T4 PNK, 6 mM ATP] and incubation for an additional 30 min at 37°C. After desalting, phenol extraction and precipitation, RNA was ligated to a 5′-adaptor RNA, gel-purified, converted to cDNA, amplified 10 cycles (soluble) or 12 cycles (crystalline) and sequenced using the Illumina SBS platform. The library prepared without input polynucleotides did not yield an observable PCR product, indicating minimal contamination from polynucleotides that might co-purify with the enzymes used for library construction.

Sequencing reads were filtered by requiring that they contain a perfect match to the first 12 nucleotides of the 3′ adaptor and that every nucleotide up to the beginning of the 3′ adaptor have a Phred+64 quality score of at least '^'. After removing the internal-standard reads and trimming away the adaptor sequences, reads representing the small RNAs were collapsed to a non-redundant set of 8–24-nucleotide sequences. To examine the origins of the co-purifying RNAs, 15–24-nucleotide sequences were mapped sequentially to the *Kp*AGO expression plasmid, the chloramphenicol-resistance gene found on pRARE2, and the BL21(DE3) genome[60], allowing no mismatches and recovering all hits. (The 15-nucleotide lower bound for mapping was chosen because this was the minimum read length that achieved a <1% genome-mapping rate for random or shuffled small-RNA sequences). Because there were fewer fortuitous matches to the *Kp*AGO expression plasmid, analysis of 12-nucleotide sequences was performed on reads that mapped to the plasmid. For mapping-independent analyses, sequences with <10 reads were not considered.

For analysis of nucleotide composition, information content was calculated by determining the relative frequency of each nucleotide at position *X* compared to the relative frequency at

all other positions combined. The selectivity for a given nucleotide *n* at position 1 was calculated using the following equation:

$$S_n=[\sum_{i=A,U,G,C}([f(i,X)/f(n,X)]/[f(i,\sim X)/f(n,\sim X)])]^{-1}$$

where *f*(*i*, *X*) is the frequency of nucleotide *i* at position *X* and *f*(*i*, ~*X*) is the frequency of nucleotide *i* at all other positions

Information content scores were then calculated using the following equation:

$$I_n=S_n*[\log_2(S_n)+2]$$

For phasing analysis, the frequency of distances separating 5′-end pairs (*i*, *j*) mapping to opposite DNA strands was calculated using the following equation:

$$Frequency_D=\sum_{i,j}\min(Reads_i, Reads_j)_D$$

where *D* = (distance between small-RNA 5′ ends) + 1

Yeast manipulations

S. castellii and *S. cerevisiae* were grown at 25°C and 30°C, respectively, on standard *S. cerevisiae* plate and liquid media (e.g., YPD and SC). Transformations of *S. castellii* were performed as described previously[29]. Transformations of *S. cerevisiae* were performed as described[61]. For FACS analyses, strains were inoculated in SC, in either non-inducing (2% glucose) or inducing (1% galactose and 1% raffinose) conditions, and grown overnight. Fresh cultures were then seeded from the overnight cultures, and cells were grown to log phase. Cells were analyzed using FACSCalibur (BD Biosciences); data were processed with CellQuest Pro (BD Biosciences) and FlowJo (Tree Star).

Plasmids and strains used and generated in this study are listed (Supplementary Tables 5 and 6). Vectors pRS404CYC1-*Kp*Ago1 and pRS405TEF-*Kp*Dcr1 were constructed by insertion of the coding sequencing of the respective *K. polysporus* genes between the *CYC1* or *TEF* promoter and *CYC1* terminator (cloned from p416CYC or p416TEF[62]) of the appropriate vector[63] using SpeI and XhoI sites (*Kp*Ago1) or BamHI and XhoI sites (*Kp*Dcr1). Vector pRS404CYC1-*Kp*Ago1(207–1251) was constructed similarly, with the insertion of an ‘ATG’ codon upstream of amino acid 207. Vector pRS404CYC1-$FLAG_3$-*Kp*Ago1 was generated by PCR-based insertion of the sequencing encoding the $FLAG_3$ epitope downstream of the ‘ATG’ codon of pRS404CYC1-*Kp*Ago1. Point mutations were introduced by PCR-based mutagenesis to generate vectors encoding mutant FLAG-tagged Ago1. pRS402GPD-GFP(S65T) was constructed by insertion of the coding sequence of GFP(S65T) [amplified from pFA6a-GFP(S65T)-kanMX6[64]] between the *GPD* promoter and *CYC1* terminator (cloned from p416GPD[62]) of pRS402[63] using SpeI and XhoI sites. To reconstitute RNAi in *S. cerevisiae*, GFP(S65T), *Kp*Ago1 and *Kp*Dcr1 expression vectors were integrated into W303-1B variants already containing other components of the GFP-

silencing system[29], using standard protocols[61]. To generate *S. castellii* strains DPB267 and DPB268 for degradome sequencing, *XRN1* was deleted in DPB005 and DPB007, respectively, using the kanMX6 cassette[64].

Degradome sequencing and analysis

Total RNA was isolated from mid-log phase (OD600 ~0.6) cultures of strains DPB267 and DPB268 using the hot-phenol method. Degradome libraries were constructed from 5 μg poly(A)+ RNA essentially as described[32] and sequenced on the Illumina SBS platform. After removing adaptor sequences and generating each reverse complement, reads representing degradome-cleavage tags were collapsed to a non-redundant set. To analyze tags deriving from Y′-element loci, 20–21-nucleotide sequences were mapped to a consensus *S. castellii* Y′ element as described previously[29], and 49% of reads in the *AGO1* library were randomly sampled (to normalize for higher sequencing yield, Supplementary Fig. 2b) and used for subsequent analyses. Mapping data was then used to generate a single-nucleotide-resolution plot of the consensus Y′ element. For phasing analysis, the frequency of distances separating opposite-strand pairs of 5′ ends of 20–21-nucleotide degradome tags (*i*), and 5′ ends of 22–23-nt small RNAs (*j*) was calculated using the following equation:

$$Frequency_D = \sum_{i,j} Reads_i * Reads_j / Norm_i$$

where D = position of 5′ end of degradome tag with respect to 5′ end of small RNA, and $Norm_i$ = number of reads for small RNAs in which the 5′ end of degradome tag i falls. Fractional frequencies were calculated for each D by dividing $Frequency_D$ by the total number of reads corresponding to degradome-tag 5′ ends that map opposite 22–23-nucleotide small RNAs.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank K. Rajashankar for assistance with data processing, V. Auyeung and D. Shechner for discussions, the Whitehead Genome Technology Core for high-throughput sequencing, and the NE-CAT beamline at the Advanced Photon Source. This work was supported by National Institutes of Health grants AI068776 (D.J.P.) and GM61835 (D.P.B.), a Human Frontier Science Program Long-term Fellowship (K.N.), a fellowship from the Japan Society for the Promotion of Science for Research Abroad (K.N.), and a National Science Foundation graduate research fellowship (D.E.W.). D.P.B. is an Investigator of the Howard Hughes Medical Institute.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell. 2004; 116:281–97. [PubMed: 14744438]
2. Meister G, Tuschl T. Mechanisms of gene silencing by double-stranded RNA. Nature. 2004; 431:343–9. [PubMed: 15372041]
3. Malone CD, Hannon GJ. Small RNAs as guardians of the genome. Cell. 2009; 136:656–68. [PubMed: 19239887]

4. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001; 409:363–6. [PubMed: 11201747]
5. Elbashir SM, Lendeckel W, Tuschl T. RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes Dev. 2001; 15:188–200. [PubMed: 11157775]
6. Hutvagner G, et al. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science. 2001; 293:834–8. [PubMed: 11452083]
7. Matranga C, Tomari Y, Shin C, Bartel DP, Zamore PD. Passenger-strand cleavage facilitates assembly of siRNA into Ago2-containing RNAi enzyme complexes. Cell. 2005; 123:607–20. [PubMed: 16271386]
8. Miyoshi K, Tsukumo H, Nagami T, Siomi H, Siomi MC. Slicer function of Drosophila Argonautes and its involvement in RISC formation. Genes Dev. 2005; 19:2837–48. [PubMed: 16287716]
9. Rand TA, Petersen S, Du F, Wang X. Argonaute2 cleaves the anti-guide strand of siRNA during RISC activation. Cell. 2005; 123:621–9. [PubMed: 16271385]
10. Tomari Y, Zamore PD. Perspective: machines for RNAi. Genes Dev. 2005; 19:517–29. [PubMed: 15741316]
11. Carthew RW, Sontheimer EJ. Origins and Mechanisms of miRNAs and siRNAs. Cell. 2009; 136:642–55. [PubMed: 19239886]
12. Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science. 2004; 305:1434–7. [PubMed: 15284453]
13. Yuan YR, et al. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Mol Cell. 2005; 19:405–19. [PubMed: 16061186]
14. Liu J, et al. Argonaute2 is the catalytic engine of mammalian RNAi. Science. 2004; 305:1437–41. [PubMed: 15284456]
15. Parker JS, Roe SM, Barford D. Crystal structure of a PIWI protein suggests mechanisms for siRNA recognition and slicer activity. EMBO J. 2004; 23:4727–37. [PubMed: 15565169]
16. Ma JB, et al. Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature. 2005; 434:666–70. [PubMed: 15800629]
17. Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. Nature. 2008; 456:209–13. [PubMed: 18754009]
18. Wang Y, et al. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature. 2008; 456:921–6. [PubMed: 19092929]
19. Wang Y, et al. Nucleation, propagation and cleavage of target RNAs in Ago silencing complexes. Nature. 2009; 461:754–61. [PubMed: 19812667]
20. Parker JS. How to slice: snapshots of Argonaute in action. Silence. 1:3. [PubMed: 20226069]
21. Makarova KS, Wolf YI, van der Oost J, Koonin EV. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol Direct. 2009; 4:29. [PubMed: 19706170]
22. Meister G, et al. Identification of novel argonaute-associated proteins. Curr Biol. 2005; 15:2149–55. [PubMed: 16289642]
23. Lingel A, Simon B, Izaurralde E, Sattler M. Structure and nucleic-acid binding of the Drosophila Argonaute 2 PAZ domain. Nature. 2003; 426:465–9. [PubMed: 14615801]
24. Yan KS, et al. Structure and conserved RNA binding of the PAZ domain. Nature. 2003; 426:468–74. [PubMed: 14615802]
25. Ma JB, Ye K, Patel DJ. Structural basis for overhang-specific small interfering RNA recognition by the PAZ domain. Nature. 2004; 429:318–22. [PubMed: 15152257]
26. Boland A, Tritschler F, Heimstadt S, Izaurralde E, Weichenrieder O. Crystal structure and ligand binding of the MID domain of a eukaryotic Argonaute protein. EMBO Rep. 2010; 11:522–7. [PubMed: 20539312]
27. Frank F, Sonenberg N, Nagar B. Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. Nature. 2010; 465:818–22. [PubMed: 20505670]

28. Boland A, Huntzinger E, Schmidt S, Izaurralde E, Weichenrieder O. Crystal structure of the MID-PIWI lobe of a eukaryotic Argonaute protein. Proc Natl Acad Sci U S A. 2011; 108:10466–71. [PubMed: 21646546]
29. Drinnenberg IA, et al. RNAi in budding yeast. Science. 2009; 326:544–50. [PubMed: 19745116]
30. Weinberg DE, Nakanishi K, Patel DJ, Bartel DP. The inside-out mechanism of Dicers from budding yeasts. Cell. 2011; 146:262–76. [PubMed: 21784247]
31. Rivas FV, et al. Purified Argonaute2 and an siRNA form recombinant human RISC. Nat Struct Mol Biol. 2005; 12:340–9. [PubMed: 15800637]
32. Addo-Quaye C, Eshoo TW, Bartel DP, Axtell MJ. Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol. 2008; 18:758–62. [PubMed: 18472421]
33. German MA, et al. Global identification of microRNA-target RNA pairs by parallel analysis of RNA ends. Nat Biotechnol. 2008; 26:941–6. [PubMed: 18542052]
34. Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J. 2001; 20:6877–88. [PubMed: 11726523]
35. Haley B, Zamore PD. Kinetic analysis of the RNAi enzyme complex. Nat Struct Mol Biol. 2004; 11:599–606. [PubMed: 15170178]
36. Martinez J, Tuschl T. RISC is a 5′ phosphomonoester-producing RNA endonuclease. Genes Dev. 2004; 18:975–80. [PubMed: 15105377]
37. Ameres SL, Martinez J, Schroeder R. Molecular basis for target RNA recognition and cleavage by human RISC. Cell. 2007; 130:101–12. [PubMed: 17632058]
38. Forstemann K, Horwich MD, Wee L, Tomari Y, Zamore PD. Drosophila microRNAs are sorted into functionally distinct argonaute complexes after production by dicer-1. Cell. 2007; 130:287–97. [PubMed: 17662943]
39. Frazao C, et al. Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. Nature. 2006; 443:110–4. [PubMed: 16957732]
40. Parker JS, Roe SM, Barford D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature. 2005; 434:663–6. [PubMed: 15800628]
41. Pan Y, MacKerell AD Jr. Altered structural fluctuations in duplex RNA versus DNA: a conformational switch involving base pair opening. Nucleic Acids Res. 2003; 31:7131–40. [PubMed: 14654688]
42. Schwarz DS, Hutvagner G, Haley B, Zamore PD. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol Cell. 2002; 10:537–48. [PubMed: 12408822]
43. Chiu YL, Rana TM. siRNA function in RNAi: a chemical modification analysis. RNA. 2003; 9:1034–48. [PubMed: 12923253]
44. Parker JS, Parizotto EA, Wang M, Roe SM, Barford D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. Mol Cell. 2009; 33:204–14. [PubMed: 19187762]
45. Mallory AC, et al. MicroRNA control of PHABULOSA in leaf development: importance of pairing to the microRNA 5′ region. EMBO J. 2004; 23:3356–64. [PubMed: 15282547]
46. Nowotny M, Gaidamakov SA, Crouch RJ, Yang W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. Cell. 2005; 121:1005–16. [PubMed: 15989951]
47. Hall TM. Structure and function of argonaute proteins. Structure. 2005; 13:1403–8. [PubMed: 16216572]
48. Nowotny M. Retroviral integrase superfamily: the structural perspective. EMBO Rep. 2009; 10:144–51. [PubMed: 19165139]
49. Schirle NT, Macrae IJ. The Crystal Structure of Human Argonaute2. Science. 2012
50. Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter, CW.; Sweet, RM., editors. Methods in Enzymology. Vol. 276. Academic Press; San Diego: 1997. p. 307-326.Macromolecular Crystallography, Part A

51. Pape T, Schneider TR. HKL2MAP: a graphical user interface for phasing with SHELX programs. J Appl Cryst. 2004; 37:843–844.
52. McCoy AJ, et al. Phaser crystallographic software. J Appl Crystallogr. 2007; 40:658–674. [PubMed: 19461840]
53. Collaborative Computational Project. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr. 1994; 50:760–3. [PubMed: 15299374]
54. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr A. 1991; 47 (Pt 2):110–9. [PubMed: 2025413]
55. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr. 2002; 58:1948–54. [PubMed: 12393927]
56. Vagin A, Teplyakov A. An approach to multi-copy search in molecular replacement. Acta Crystallogr D Biol Crystallogr. 2000; 56:1622–4. [PubMed: 11092928]
57. Brunger AT, et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallographica Section D-Biological Crystallography. 1998; 54:905–921.
58. DeLano WL, Lam JW. PyMOL: A communications tool for computational models. Abstracts of Papers of the American Chemical Society. 2005; 230
59. Grimson A, et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. Nature. 2008; 455:1193–7. [PubMed: 18830242]
60. Jeong H, et al. Genome sequences of Escherichia coli B strains REL606 and BL21(DE3). J Mol Biol. 2009; 394:644–52. [PubMed: 19786035]
61. Gietz RD, Schiestl RH. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat Protoc. 2007; 2:31–4. [PubMed: 17401334]
62. Mumberg D, Muller R, Funk M. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene. 1995; 156:119–22. [PubMed: 7737504]
63. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. Genetics. 1989; 122:19–27. [PubMed: 2659436]
64. Longtine MS, et al. Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast. 1998; 14:953–61. [PubMed: 9717241]

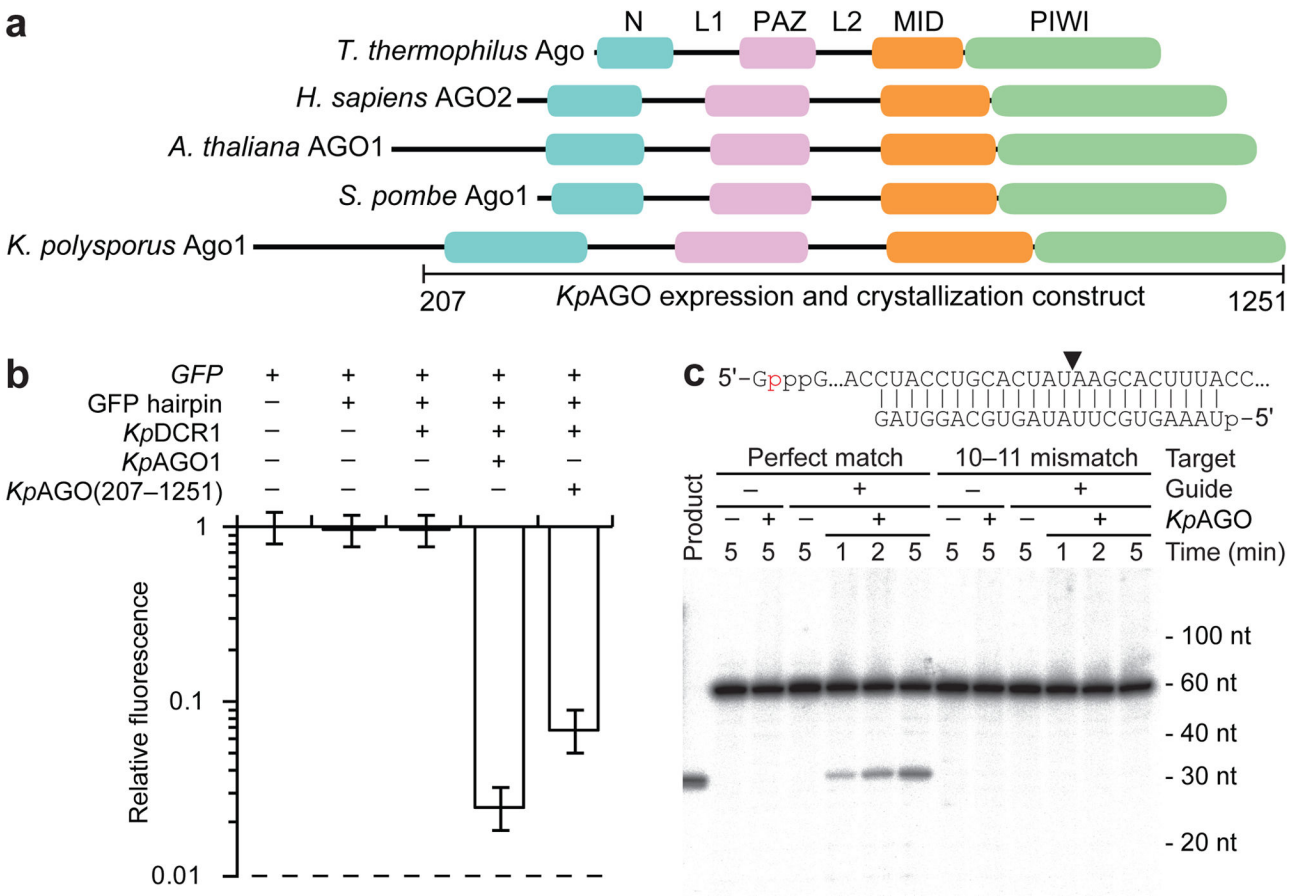

Figure 1. Cleavage activity of budding-yeast AGO

a, Domain architectures of AGO proteins. **b**, RNAi reconstituted in *S. cerevisiae* using *K. polysporus* genes. Median GFP intensity is plotted as a fraction of GFP-only control. Error bars, quartiles; dashed line, background fluorescence. **c**, Cleavage activity of *Kp*AGO. RNAs labeled at a cap phosphate (red) and matching the guide (either perfectly or with mismatches to positions 10–11) were incubated with/without (+/−) *Kp*AGO that had been pre-incubated in the presence/absence (+/−) of synthetic guide RNA. Product was resolved on a denaturing polyacrylamide gel alongside cap-labeled synthetic product (left) and RNA standards (migration shown on right).

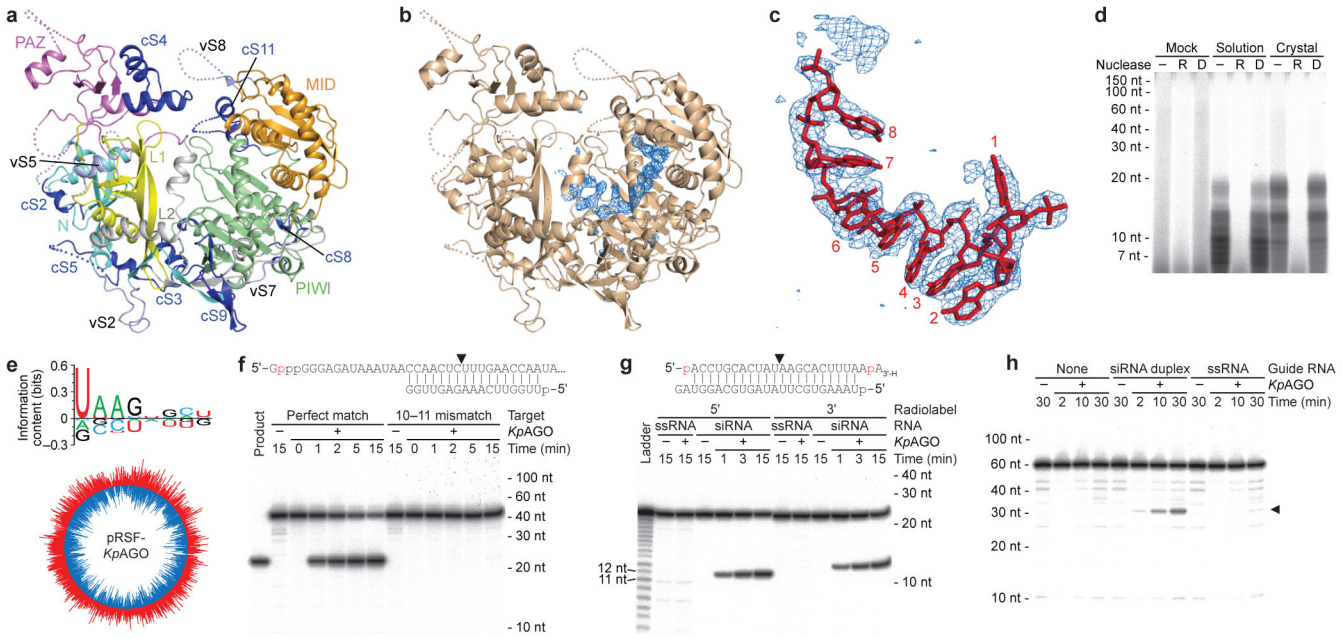

Figure 2. *Kp*AGO architecture and co-purifying RNA

a, *Kp*AGO protein structure, with N (cyan), linker L1 (yellow), PAZ (violet), linker L2 (grey), MID (orange), and PIWI (green) domains in ribbon representation. Constant (cS) and variable (vS) insertion segments, blue and slate, respectively; disordered regions, dotted lines. **b**, *Fo*–*Fc* map (blue) contoured at 2.8 σ before modeling RNA. **c**, Simulated-annealing omit map (blue) contoured at 3.5 σ around final RNA model (red). **d**, Nuclease sensitivity of co-purifying nucleic acid. End-labeled polynucleotides extracted from the indicated *Kp*AGO samples were either untreated (−) or incubated with RNase (R) or DNase (D) before analysis on a denaturing gel. **e**, Nucleotide composition and origin of co-purifying RNA. Sequences were analyzed for enriched or depleted nucleotides (positive or negative bits, respectively) at each of the first eight positions (top). Numbers of sequencing reads mapping along each strand of the *Kp*AGO expression plasmid are indicated (bottom, log scale). **f**, Cleavage activity guided by co-purifying RNA. As in Fig. 1c, except labeled RNAs were designed to match the indicated co-purifying RNA. **g**, Autonomous duplex loading and passenger-strand cleavage. Labeled ssRNA or siRNA duplex was incubated +/− *Kp*AGO. **h**, Autonomous duplex loading and target cleavage. As in Fig. 1c, except *Kp*AGO and RNA concentrations were reduced by 90% and 99.95%, respectively.

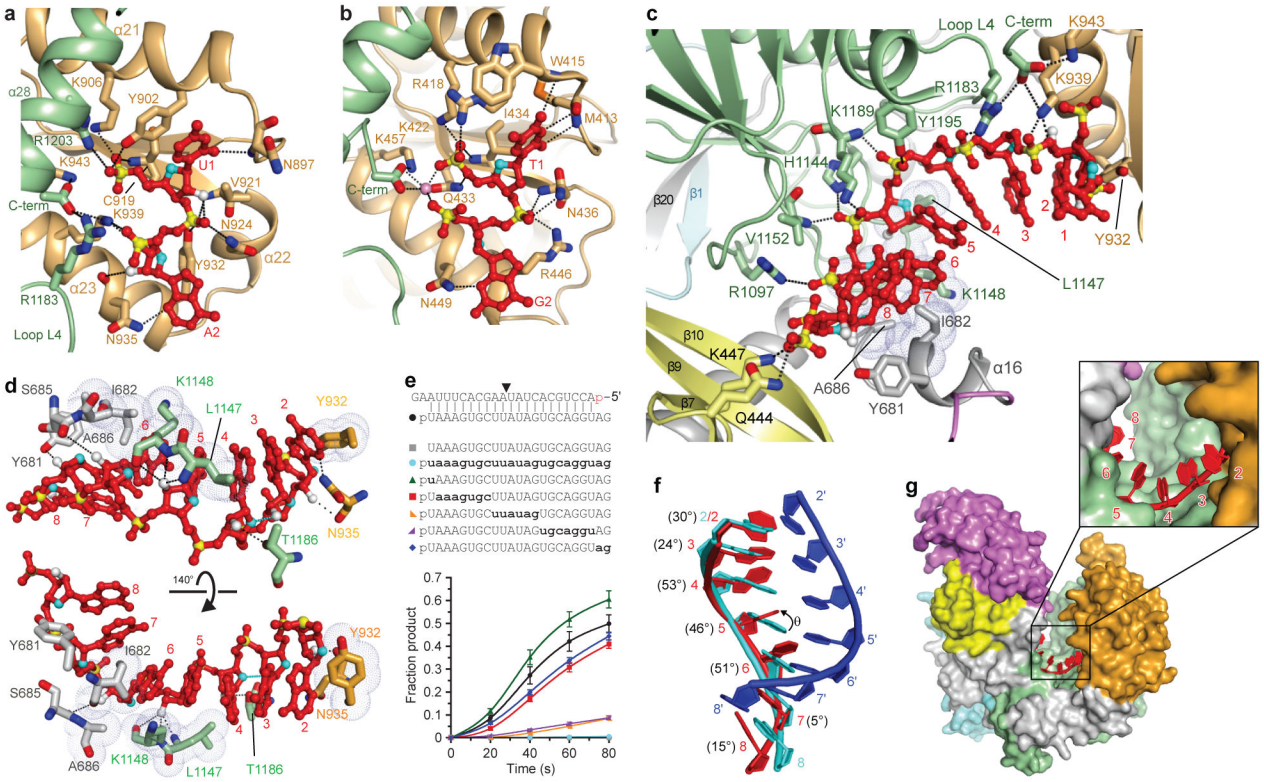

Figure 3. Organization of the guide RNA

a–b, The 5′-nucleotide–binding pockets of *Kp*AGO (a) and *Tt*AGO[17] (b). Colors, as in Fig. 2a; protein, ribbon representation; highlighted residues and RNA, stick representation; O2′, O4′ and phosphate, white, cyan and yellow, respectively; hydrogen bonds, dotted lines. **c–d,** Interactions involving bases and either phosphates (c) or 2′-OH groups (d) of the seed region. Intermolecular (black) and intramolecular (blue) hydrogen bonds, dotted lines; hydrophobic interactions, van der Waals radii. **e,** Effects of guide-strand modifications on duplex loading and passenger-strand cleavage. *Kp*AGO was incubated with siRNA duplexes with the indicated guide strands; p, 5′ monophosphate; upper case, 2′-OH; lower case, 2′-deoxy. The fraction of labeled passenger strand cleaved is plotted (average of three independent replicates; error bars, standard deviations; points connected by smooth curves). **f**, Superposition of guide-RNA nucleotides 2–8 (red) on A-form RNA (cyan and blue). Dihedral angles (θ) between guide-RNA bases and those of A-form RNA are in parenthesis. **g**, Solvent-exposed seed nucleotides (red). *Kp*AGO surface is rendered, domains colored as in Fig. 2a.

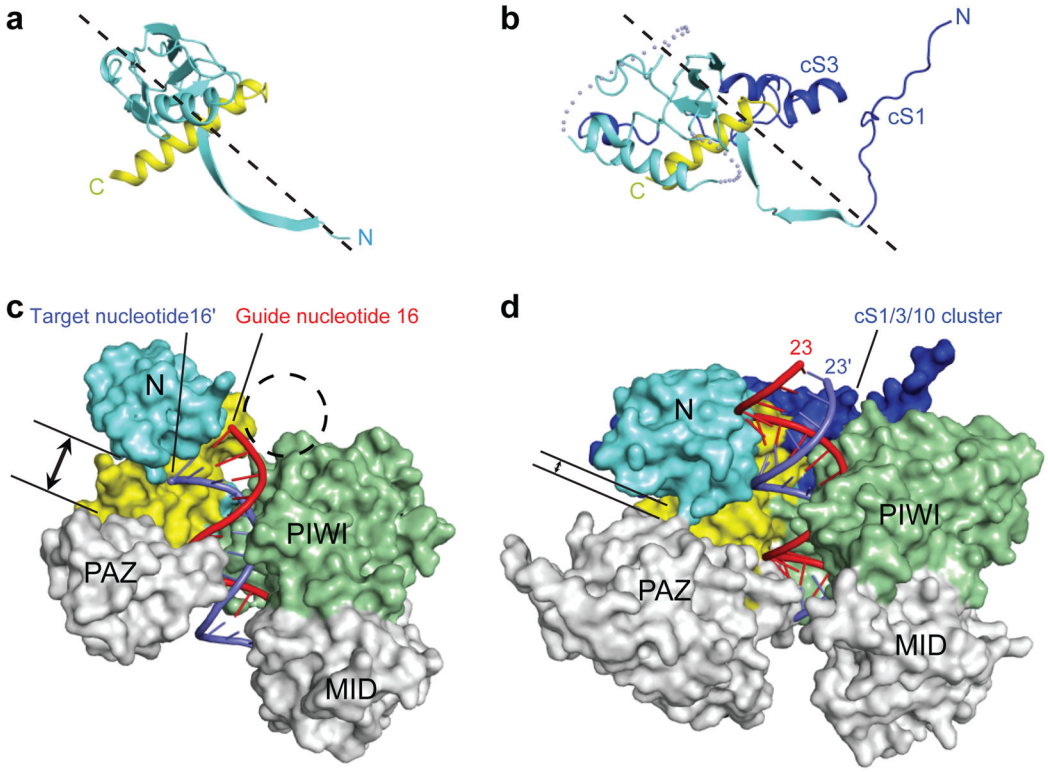

Figure 4. An extended, potentially unobstructed nucleic acid–binding channel in *Kp*AGO

a–b, Position of N domain relative to L1-linker domain in *Tt*AGO[19] (a) and *Kp*AGO (b). Domains are oriented based on their N-terminal beta strands (dashed line connects strand termini). Colors as in Fig. 2a. **c–d,** Channels of *Tt*AGO ternary complex[19] (c) and *Kp*AGO with modeled A-form duplex (d). Protein surfaces are rendered, highlighting distances between the N and PAZ domains (parallel lines) and the cS1/3/10 cluster (blue), which fills a cavity (dashed circle).

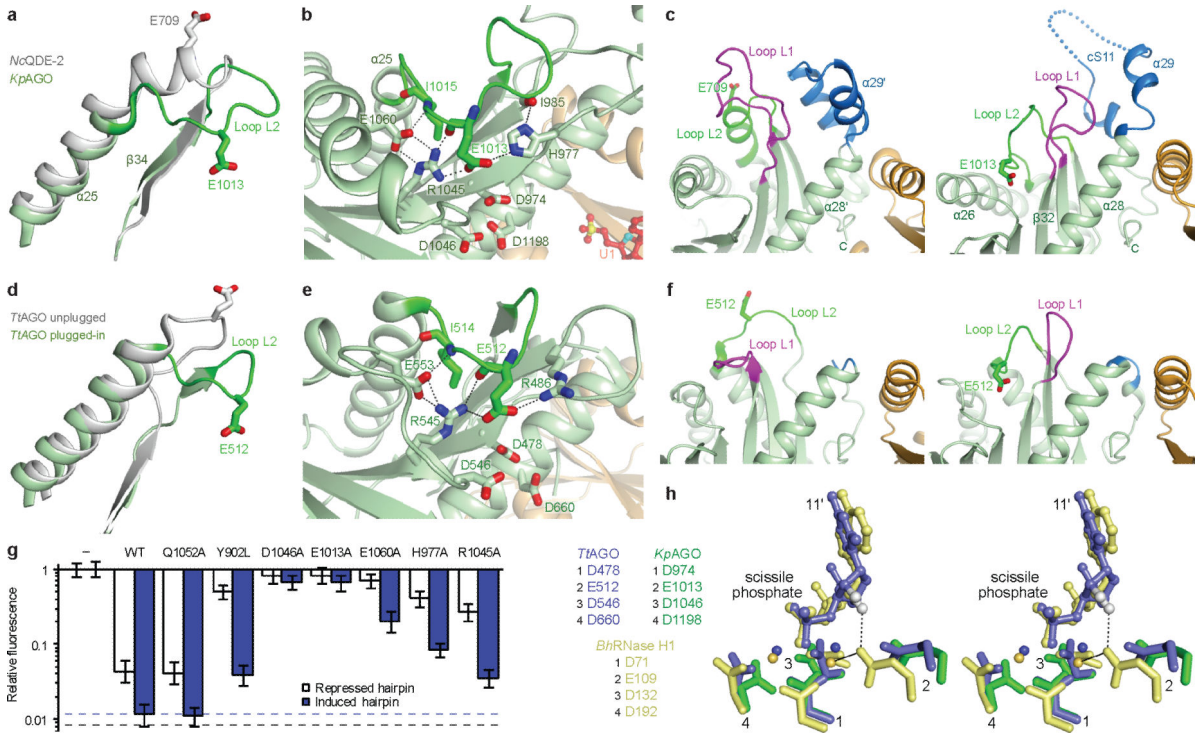

Figure 5. A plugged-in glutamate finger at the active site

a, Superposition of α25 and β34 of *Kp*AGO (green) on counterparts of *Nc*QDE-2 MID-PIWI lobe (grey), highlighting the extended loop L2 (dark green). **b**, Hydrogen-bond network stabilizing the plugged-in loop L2. Loop L2, dark green; otherwise, as in Fig. 3a. **c**, Closed (left) and open (right) configurations of the loop L1 gate (purple) in *Nc*QDE-2 MID-PIWI lobe and *Kp*AGO, respectively. cS11, blue; otherwise, as in panel b. **d**, Superposition of the region flanking loop L2 in the unplugged (grey) and plugged-in (green) conformations of *Tt*AGO, depicted as in panel a. **e**, Hydrogen-bond network stabilizing the plugged-in loop L2 in *Tt*AGO, depicted as in panel b. **f**, Closed and open configurations of the loop L1 gate in the unplugged and plugged-in conformations of *Tt*AGO, respectively, depicted as in panel c. **g**, RNAi reconstituted in *S. cerevisiae* using wild-type *K. polysporus AGO1* WT or genes with the indicated substitutions. Silencing was monitored under either permissive (induced hairpin, blue bars) or stringent (repressed hairpin, open bars) conditions. Q1052 and Y902, conserved residues insensitive and sensitive to substitution, respectively[28], were included as controls. Dashed lines (blue and black), background fluorescence in permissive and stringent conditions, respectively; otherwise, as in Fig. 1b. **h**, Stereoview of *Kp*AGO catalytic residues (green) superpositioned with catalytic residues, divalent cations, scissile phosphate and adjacent nucleoside in *Tt*AGO (blue) and *Bh*RNase H1 (yellow) ternary complexes.

Table 1

X-ray refinement statistics

Resolution (Å)	38.12–3.19
R_{work}/R_{free}	16.83/21.62
Number of atoms	
Protein	14,034
RNA	358
Water	34
B-factors	
Protein	52.47
RNA	42.62
Water	28.60
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (°)	1.347