Article

# Structural basis for piRNA targeting

https://doi.org/10.1038/s41586-021-03856-x

Received: 20 February 2021

Accepted: 28 July 2021

Published online: 1 September 2021

Check for updates

Todd A. Anzelon[1,3], Saikat Chowdhury[1,2,3], Siobhan M. Hughes[1,3], Yao Xiao[1], Gabriel C. Lander[1] & Ian J. MacRae[1]✉

PIWI proteins use PIWI-interacting RNAs (piRNAs) to identify and silence transposable elements and thereby maintain genome integrity between metazoan generations[1]. The targeting of transposable elements by PIWI has been compared to mRNA target recognition by Argonaute proteins[2,3], which use microRNA (miRNA) guides, but the extent to which piRNAs resemble miRNAs is not known. Here we present cryo-electron microscopy structures of a PIWI–piRNA complex from the sponge *Ephydatia fluviatilis* with and without target RNAs, and a biochemical analysis of target recognition. Mirroring Argonaute, PIWI identifies targets using the piRNA seed region. However, PIWI creates a much weaker seed so that stable target association requires further piRNA–target pairing, making piRNAs less promiscuous than miRNAs. Beyond the seed, the structure of PIWI facilitates piRNA–target pairing in a manner that is tolerant of mismatches, leading to long-lived PIWI–piRNA–target interactions that may accumulate on transposable-element transcripts. PIWI ensures targeting fidelity by physically blocking the propagation of piRNA–target interactions in the absence of faithful seed pairing, and by requiring an extended piRNA–target duplex to reach an endonucleolytically active conformation. PIWI proteins thereby minimize off-targeting cellular mRNAs while defending against evolving genomic threats.

Most animals produce at least two types of small regulatory RNA: miRNAs and piRNAs[4]. miRNAs regulate mRNAs during development, whereas piRNAs protect the germline from transposable elements. At the molecular level, miRNAs and piRNAs function as guides for Argonaute (AGO) and PIWI proteins, respectively, which use sequence information in small RNAs to identify transcripts that are targeted for repression.

Extensive studies have shown that miRNA-class AGO proteins tune the binding properties of miRNAs to recognize short segments of complementarity on target mRNAs[5–8]. AGOs create the miRNA seed region by pre-organizing nucleotides 2–7, lowering the entropic cost of target binding[9,10], and adopt a structure that discourages miRNA–target pairing immediately after the seed[11,12]. miRNA-recognition sites are thus short and precisely defined, enabling individual miRNAs to recognize hundreds of mRNAs and collectively target more than half of the mRNAs in mammals[13].

It has been suggested that piRNAs use a miRNA-like targeting mechanism[2,3]—indeed, seed complementarity is a feature of piRNA targets[2,3,14–17]. However, the relationship between piRNAs and their targets is fundamentally distinct from that of miRNAs. miRNA targets evolve to be recognized to benefit developmental fitness[18]. By contrast, piRNA targets are parasitic genetic elements that are under selective pressure to escape recognition, suggesting a benefit for mechanisms that adapt piRNAs to evolving threats[19]. Additionally, the piRNA repertoire in animals is usually orders of magnitude larger than that of miRNAs[20], indicating that mechanisms must be in place to avoid silencing the entire germline transcriptome.

Crystal structures of insect PIWI proteins, *Drosophila melanogaster* Piwi (*Dm*Piwi) and silkworm Siwi also suggest that piRNAs might be distinct from miRNAs[21,22]. These PIWIs have a 3D domain arrangement that differs from that of AGO proteins, indicating that piRNA–target interactions are shaped in a unique fashion. Previous studies used PIWI samples containing heterogenous mixtures of co-purifying piRNAs, preventing investigation into piRNA–target interactions. As such, the mechanisms that underlie the recognition of transposable elements by piRNAs, as well as the extent to which piRNA targeting resembles that of miRNAs, remain unknown.

## Source of homogenous PIWI–piRNA complexes

PIWI proteins from natural sources co-purify with heterogenous mixtures of endogenous piRNAs[21,22], which obstruct the analysis of targeting. We therefore sought a recombinant system for the preparation of homogenous PIWI–piRNA complexes. A screen of PIWI protein constructs from various animals revealed that Piwi-A from the freshwater sponge *E. fluviatilis*[23] (hereafter referred to as *Ef*Piwi) can be expressed recombinantly, loaded with a chemically defined piRNA and purified as a stable PIWI–piRNA complexes. *Ef*Piwi is normally expressed in *E. fluviatilis* totipotent archeocytes[24]. We also identified a construct to reconstitute the Siwi–piRNA complex, albeit at lower levels than *Ef*Piwi (Extended Data Fig. 1a, b).

## Unique structural features of the extended PIWI family

*Ef*Piwi belongs to an ancient branch of the PIWI family tree that contains *Drosophila* Ago3, Mili, Hili and Zili, and is distinct from the branch that contains Siwi and *Dm*Piwi[25] (Extended Data Fig. 1c). As such, structural features that are common to *Ef*Piwi and Siwi/*Dm*Piwi may be broadly

[1]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA. [2]Present address: CSIR-Centre for Cellular and Molecular Biology, Hyderabad, India. [3]These authors contributed equally: Todd A. Anzelon, Saikat Chowdhury, Siobhan M. Hughes. ✉e-mail: macrae@scripps.edu

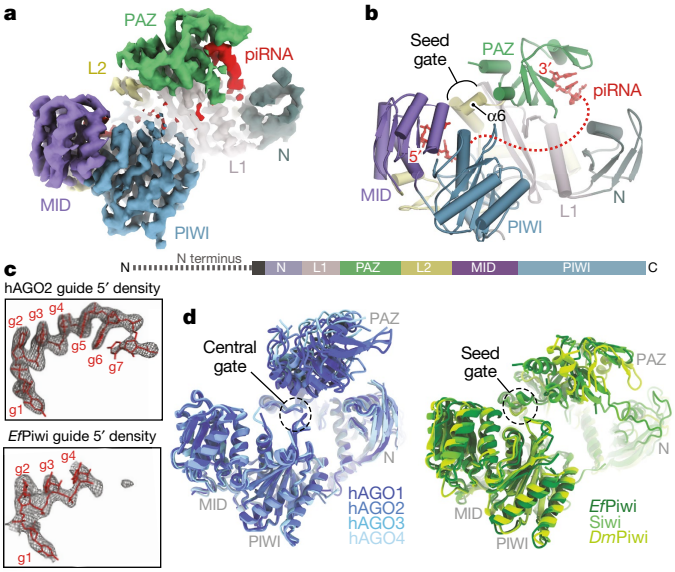



**Fig. 1 | Structural features unique to PIWIs. a**, Segmented *Ef*Piwi–piRNA cryo-EM reconstruction, coloured by domain as indicated. **b**, Cartoon representation of the *Ef*Piwi model. Disordered piRNA nucleotides are indicated by a dashed red line. The linear domain schematic is shown below. **c**, Density maps (grey mesh) in space occupied by seed regions of hAGO2 (PDB: 4OLA; top) and *Ef*Piwi (bottom). Modelled guide RNAs are shown as red sticks. **d**, Superposition of miRNA-class AGO structures (left, PDB: 4KXT, 4OLA, 5VM9, 6OON) and known PIWI structures (right, PDB: 7KX7, 5GUH, 6KR6). AGO central-gate and PIWI seed-gate structures are indicated.

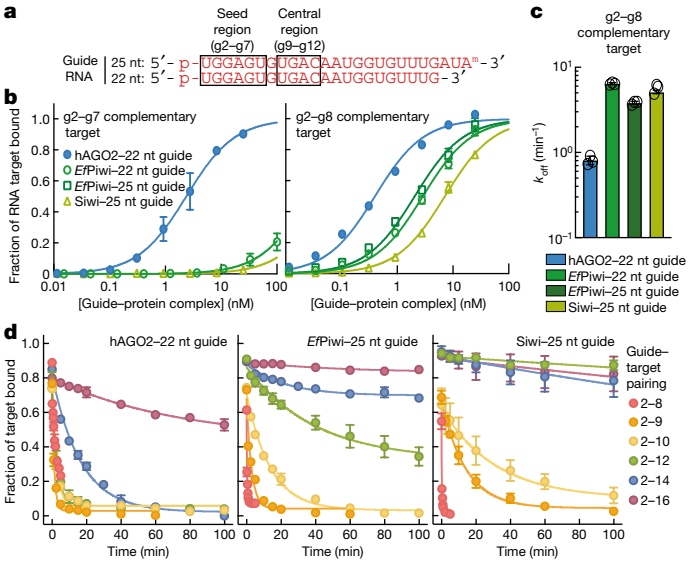



**Fig. 2 | piRNAs are more selective than miRNAs. a**, The guide RNAs used in this experiment. **b**, Fraction of target RNAs bound by hAGO2–guide RNA, *Ef*Piwi–guide RNA and Siwi–guide RNA complexes at equilibrium, plotted against the concentration of the protein–guide RNA complex. **c**, Release rates (dissociation rate, $k_{off}$) of g2–g8 complementary target from protein–guide RNA complexes. **d**, Release of $^{32}$P-labelled target RNAs from the indicated protein–guide complexes in the presence of excess unlabelled target RNA over time. In **b**, **c**, $n = 3$ independent experiments, data are mean ± s.d.

conserved in the PIWI family. We used cryo-electron microscopy (cryo-EM) single-particle analysis to determine the structure of the *Ef*Piwi–piRNA complex to approximately 3.8 Å resolution (Fig. 1a, b, Extended Data Table 1, Extended Data Fig. 2). A comparison of PIWI and metazoan AGO structures revealed two features that were likely to affect targeting by members of the extended PIWI family.

First, although our sample contained a homogenous piRNA, most nucleotides are disordered, including the 3′ half of the seed region (guide (g) nucleotides g5–g7) (Fig. 1c). AGO proteins pre-organize the seed 3′ end by cradling g5–g6 in a kinked loop that is widely conserved in AGOs. The equivalent loop in PIWIs contains bulky residues that prevent the kink that is necessary to cradle g5–g6 (Extended Data Fig. 3a, b). Therefore, unlike AGOs, PIWIs do not pre-organize the full piRNA seed region.

Second, as previously noted[21], the three-way interface of the L1, L2 and PAZ domains in PIWIs differs from that of AGOs. An important consequence of this difference is that residues that correspond to the AGO 'central gate', which restricts guide–target interactions in the miRNA central region[11,12], instead form a small α-helix (helix-6) near the piRNA seed region (Fig. 1b, Extended Data Fig. 3c–g). PIWIs therefore lack a central gate, and instead have a widened central cleft and extended 'seed gate' structure (Fig. 1d).

## piRNAs are more selective than miRNAs

To determine how structural differences affect target recognition, we compared the target-RNA-binding properties of *Ef*Piwi, Siwi and human Argonaute 2 (hAGO2), a miRNA-class AGO protein. *Ef*Piwi and Siwi bound target RNAs with complementarity limited to the seed region (g2–g7) with more than 150-fold lower affinity than hAGO2 loaded with an equivalent guide RNA (Fig. 2a, b). Similarly, both PIWIs bound a target with extended seed (g2–g8) complementarity with more than 7-fold lower affinity than hAGO2. Weak seed-pairing seems to arise from higher target off-rates from the PIWIs, which released the g2–g8 matched target RNA around 7-fold faster than did hAGO2 (Fig. 2c). Extending target complementarity through the guide central region (g9–g12) had little effect on hAGO2 release rates, which is consistent with the model that hAGO2 avoids central pairing due to its central gate[11,12] (Fig. 2d, Extended Data Fig. 4). By contrast, for *Ef*Piwi and Siwi, target release rates decreased to below $1 \times 10^{-3}$ min$^{-1}$ (half-life, $t_{1/2} > 12$ h) as guide-target complementarity extended through the central region (Fig. 2d, Extended Data Fig. 4).

We propose that differences in the structures of PIWI and AGO lead to distinct targeting properties for piRNAs and miRNAs. PIWIs pre-organize a minimal seed, rendering the piRNA seed far weaker than the miRNA seed. However, unlike AGOs, PIWIs lack a central gate and therefore can compensate for weak seed-pairing by extending piRNA–target interactions into the central region. piRNAs are thereby more selective than miRNAs when identifying targets, and can remain associated with recognized targets for much longer periods of time.

## The seed-gate enforces piRNA target-binding fidelity

Although the piRNA seed is weak, perfect seed complementarity is required for the recognition of piRNA targets in vivo[2,3,14–17]. To understand this requirement, we determined the structure of the *Ef*Piwi–piRNA complex bound to a target RNA with complementarity to piRNA nucleotides g2–g16, to approximately 3.5 Å resolution (Fig. 3a, b, Extended Data Table 1, Extended Data Fig. 5). The reconstruction contains a conspicuous 15-bp piRNA–target RNA duplex (Fig. 3a), which seems to drive *Ef*Piwi into a more opened conformation (Fig. 3c). *Ef*Piwi opening involves a shift in the seed gate, which moves approximately 12 Å to avoid clashing with base pairs 5–8 of the piRNA–target duplex (Fig. 3d) and docks with the duplex at the 3′ end of the seed (Fig. 3e). These contacts, as well as contacts from the MID and PIWI domains, probe the piRNA–target duplex backbone and minor groove, probably enforcing ideal pairing to the piRNA seed.

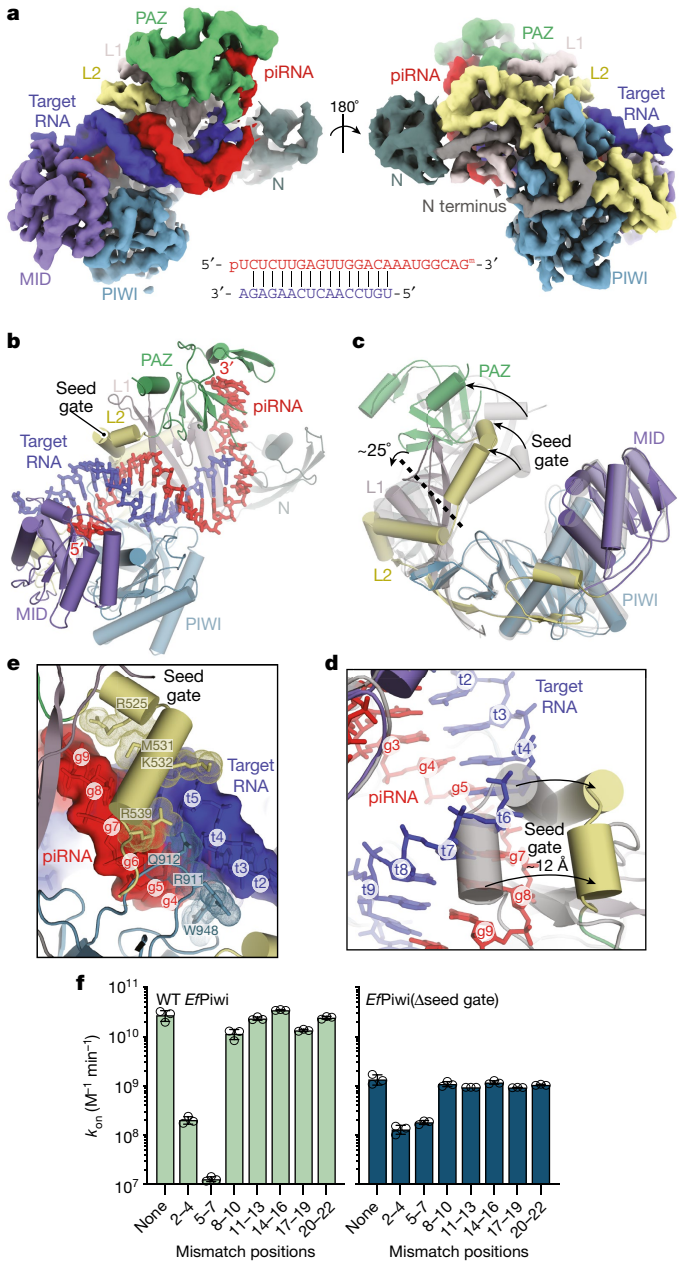



**Fig. 3 | Structural basis for piRNA target binding. a**, Segmented cryo-EM reconstruction of an *Ef*Piwi–piRNA–target RNA complex, coloured as in Fig 1. The piRNA–target base-pairing schematic is shown below. **b**, Cartoon representation of the *Ef*Piwi–piRNA–target RNA model. **c**, Superposition of *Ef*Piwi–piRNA (grey, semi-transparent) and *Ef*Piwi–piRNA–target (coloured and solid) structures. The arrows indicate movements to the target-bound structure, and the dashed line indicates the hinge in the L1 stalk. **d**, Close-up view showing shifts in the seed gate of approximately 12 Å upon target binding. The arrows indicate the direction of movement from the guide-only to the target-bound conformation. **e**, Interactions between *Ef*Piwi and the piRNA–target duplex around the seed region. **f**, Association rates of target RNAs with three consecutive mismatches binding wild-type *Ef*Piwi(left) and *Ef*Piwi(Δseed gate) (right). $n = 3$ independent experiments, data are mean ± s.d.

Notably, piRNA–target mismatches in the 3' half of the seed (g5–g7) nearly abolish target binding, reducing the association rate $k_{on}$ by more than 2,000-fold compared to a target with no mismatches (Fig. 3f, Extended Data Fig. 6). Binding to the g5–g7 mismatched target was restored by removing the seed gate (*Ef*Piwi(Δseed gate), L520–H537 replaced with a $Gly_6$ linker), which increased $k_{on}$ 14-fold compared to wild-type *Ef*Piwi (Fig. 3f). *Ef*Piwi(Δseed gate) bound a target with mismatches at the seed 5' end (g2–g4) at a similar rate to wild-type *Ef*Piwi. By contrast, *Ef*Piwi(Δseed gate) bound all targets with intact seed complementarity about 10-fold slower than the wild-type. Siwi(Δseed gate) behaved similarly (Extended Data Fig. 7a, b). The seed gate therefore enforces targeting fidelity by facilitating propagation of piRNA–target pairing initiated at the seed 5' end, and by inhibiting duplex propagation in the absence of faithful pairing to the seed 3' end.

## PIWI accommodates mismatches beyond the seed

Beyond the seed-paired region, *Ef*Piwi contacts the piRNA–target duplex backbone but does not probe the shape of the minor groove (Extended Data Fig. 7c). *Ef*Piwi might thereby recognize the overall helical structure of the duplex in a manner that is tolerant of deviations from A-form geometry, and therefore insensitive to intermittent mismatches beyond the seed. Indeed, 3-nt mismatched segments after the seed have only minor effects (2.5-fold or less) on target-binding rates (Fig. 3f) and most bound targets dissociate over the course of hours (Extended Data Figs. 6b, 7d, e). *Ef*Piwi can therefore engage targets containing 3 or fewer consecutive non-seed mismatches with excellent affinity (dissociation constant, $K_D$ values in the pM–fM range) (Extended Data Fig. 7f). Tolerance of mismatches beyond the seed might enable piRNAs to recognize related or evolving transposable elements.

## Extensive piRNA 3′ pairing activates target cleavage

PIWIs often endonucleolytically cleave target RNAs[21,22,26]. Cleavage activity of both *Ef*Piwi and Siwi is far weaker than that of hAGO2 but can be stimulated by $Mn^{2+}$ and increasing temperature (Extended Data Fig. 8a–e), enabling us to characterize target substrate preferences. *Ef*Piwi cleavage activity increased as complementarity extended towards the piRNA 3′ end, reaching a plateau when extended to g18 (Fig. 4a, Extended Data Fig. 9a, b). Siwi cleavage also increased with 3' complementarity, reaching a plateau at g16 pairing (Extended Data Fig. 9e, f).

We examined how mismatches towards the piRNA 3′ end influence cleavage (Extended Data Fig. 9c, d). A pool of 256 target RNA sequences with combinations of mismatches to g11–g18 was treated with excess *Ef*Piwi–piRNA. RNA sequencing of cleavage products revealed that the top 10% (26 out of 256 sequences) accounted for 80% of all cleavage events and contained no segments with 3 or more consecutive mismatches (Fig. 4b, Extended Data Fig. 9f). *Ef*Piwi nuclease activity is therefore driven by extended 3' pairing, optimally with 2 or fewer mismatches towards the piRNA 3' end.

## Extended pairing drives *Ef*Piwi into an open conformation

We also examined the structure of *Ef*Piwi–piRNA bound to a target RNA with complementarity to g2–g25 by cryo-EM. Although the piRNA–target duplex was obvious in 2D averages, *Ef*Piwi appeared smaller than expected (Fig. 4b). An approximately 7.0 Å resolution 3D reconstruction was consistent with the MID and PIWI domains bound to the end of the piRNA–target duplex (Fig. 4c). Docking the full *Ef*Piwi–piRNA–target model (with pairing to g2–g16) shows that, after position 15, the extended piRNA–target duplex sterically clashes with segments of the N, L1 and PAZ domains (Fig. 4d). Extended piRNA–target pairing therefore drives central cleft opening to such an extent that the entire N–L1–PAZ–L2 lobe becomes conformationally uncoupled from the MID–PIWI lobe.

We suggest that *Ef*Piwi nuclease activity might be activated by widening of the central cleft, driven by the formation of an extended piRNA–target duplex. Mismatched segments of at least 3 nt in length could impart flexibility to the guide RNA–target duplex, reducing

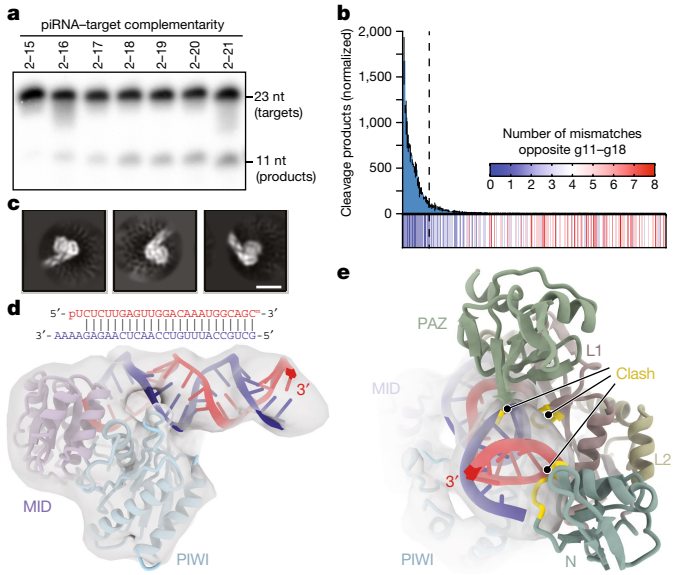



**Fig. 4 | Extensive pairing activates piRNA–target cleavage. a**, Representative denaturing gel showing intact and cleaved target RNAs, with increasing amounts of complementarity to a piRNA, after incubation with excess *Ef*Piwi–piRNA for 1 h. Results are representative of 3 replicates. **b**, Relative number of RNA sequencing reads of cleavage products after treating a pool of 256 target RNAs with excess *Ef*Piwi–piRNA. The heat map indicates the number of mismatches opposite piRNA nucleotides g11–g18 in each sequence. The dashed line denotes the 26 most abundant products. $n = 3$ independent experiments, data are mean ± s.d. **c**, Three examples of 2D class averages of *Ef*Piwi–piRNA complex bound to a target with complementarity from g2–g25. Scale bar, 5 nm. **d**, 3D cryo-EM reconstruction fit with the MID/PIWI lobe and extended piRNA–target duplex. The piRNA 3′ end is indicated. **e**, Docking the *Ef*Piwi–piRNA–target model with pairing limited to g2–g16 (Fig. 3) reveals that the N, L1 and PAZ domains sterically clash (gold highlights) with the extended piRNA–target duplex.

the propensity to drive open the cleft and stimulate cleavage. piRNA cleavage products identified in mouse testes are generally free of mismatched segments 3 nt or more in length[15,17,27], and target cleavage by immunopurified Miwi also requires extended piRNA–target complementarity[26].

## Discussion

piRNAs and miRNAs are ancient genetic regulators that might have helped to usher in the era of multicellular animal life[4]. Many animals also produce small interfering RNAs (siRNAs), which typically associate with a separate siRNA class of AGO proteins[25] that remains uncharacterized at the structural level. Our results show that PIWI and miRNA-class AGO proteins possess distinct structural features, enabling piRNAs and miRNAs to carry out discrete roles in animal evolution.

The primary function of piRNAs is to seek out and silence transposable elements. Considering the immense diversity of the piRNA repertoire, targeting must be stringent enough to avoid inadvertently silencing cellular RNAs. *Ef*Piwi accomplishes this by creating a weak seed and closely monitoring pairing via the seed gate. Seed complementarity is therefore necessary but insufficient for target recognition. Stringency is further established for target cleavage by the requirement of a piRNA–target duplex that is strong enough to activate the endonuclease mechanism. Stringency might also be regulated by a recently discovered cleavage stimulating factor[28]. Selectivity at the target-cleavage step could help to guard against cellular RNAs entering piRNA biogenesis pathways via the ping-pong cycle[1].

On the other hand, flexibility in target recognition would enable piRNAs to respond to evolving transposable element sequences[19]. Targeting flexibility, in terms of binding, is provided beyond the seed, where the PIWI central cleft is tolerant of helical imperfections that arise from piRNA–target mismatches such that high-affinity binding requires less complementarity than cleavage in vitro. A notable feature of the PIWI–piRNA complex is that the release of bound target molecules is exceptionally slow, similar to the lifetimes of the most stable mRNAs in embryonic stem cells[29]. As such, after PIWI engages a target, it could potentially remain bound for the remainder of the transcript's existence. We propose that the recognition of transposable elements might involve the accumulation of multiple PIWI–piRNA complexes on each target transcript, leading to multivalent assemblies that recruit histone and/or DNA methylation factors to transposable-element loci in the nucleus[30–33], or traffic cytoplasmic transcripts to phase-separated compartments that are associated with silencing and piRNA production[34]. We suggest that by shaping piRNA–target interactions, PIWIs can leverage the vast piRNA pool to restrict the escape of transposable elements from surveillance while minimizing off-targeting, and have thereby maintained metazoan germlines over the past 800 million years.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03856-x.


1. Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D. & Zamore, P. D. PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* **20**, 89–108 (2019).
2. Shen, E. Z. et al. Identification of piRNA binding sites reveals the Argonaute regulatory landscape of the *C. elegans* germline. *Cell* **172**, 937–951.e18 (2018).
3. Gou, L. T. et al. Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res.* **24**, 680–700 (2014).
4. Grimson, A. et al. Early origins and evolution of microRNAs and Piwi-interacting RNAs in animals. *Nature* **455**, 1193–1197 (2008).
5. Lewis, B. P., Shih, I. H., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. Prediction of mammalian microRNA targets. *Cell* **115**, 787–798 (2003).
6. Wee, L. M., Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* **151**, 1055–1067 (2012).
7. Salomon, W. E., Jolly, S. M., Moore, M. J., Zamore, P. D. & Serebrov, V. Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell* **162**, 84–95 (2015).
8. Chandradoss, S. D., Schirle, N. T., Szczepaniak, M., MacRae, I. J. & Joo, C. A dynamic search process underlies microRNA targeting. *Cell* **162**, 96–107 (2015).
9. Parker, J. S., Parizotto, E. A., Wang, M., Roe, S. M. & Barford, D. Enhancement of the seed-target recognition step in RNA silencing by a PIWI/MID domain protein. *Mol. cell* **33**, 204–214 (2009).
10. Schirle, N. T. & MacRae, I. J. The crystal structure of human Argonaute2. *Science* **336**, 1037–1040 (2012).
11. Schirle, N. T., Sheu-Gruttadauria, J. & MacRae, I. J. Structural basis for microRNA targeting. *Science* **346**, 608–613 (2014).
12. Sheu-Gruttadauria, J., Xiao, Y., Gebert, L. F. & MacRae, I. J. Beyond the seed: structural basis for supplementary microRNA targeting by human Argonaute2. *EMBO J.* **38**, e101153 (2019).
13. Friedman, R. C., Farh, K. K., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* **19**, 92–105 (2009).
14. Zhang, D. et al. The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science* **359**, 587–592 (2018).
15. Goh, W. S. et al. piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.* **29**, 1032–1044 (2015).
16. Halbach, R. et al. A satellite repeat-derived piRNA controls embryonic development of *Aedes*. *Nature* **580**, 274–277 (2020).
17. Zhang, P. et al. MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.* **25**, 193–207 (2015).
18. Nozawa, M. et al. Evolutionary transitions of microRNA–target pairs. *Genome Biol. Evol.* **8**, 1621–1633 (2016).
19. Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**, 1089–1103 (2007).
20. Wang, J. et al. piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.* **47**, D175–D180 (2019).
21. Matsumoto, N. et al. Crystal structure of silkworm PIWI-clade Argonaute Siwi bound to piRNA. *Cell* **167**, 484–497.e9 (2016).
22. Yamaguchi, S. et al. Crystal structure of *Drosophila* Piwi. *Nat. Commun.* **11**, 858 (2020).

23. Funayama, N., Nakatsukasa, M., Mohri, K., Masuda, Y. & Agata, K. Piwi expression in archeocytes and choanocytes in demosponges: insights into the stem cell system in demosponges. *Evol. Dev.* **12**, 275–287 (2010).
24. Alie, A. et al. The ancestral gene repertoire of animal stem cells. *Proc. Natl Acad. Sci. USA* **112**, E7093–E7100 (2015).
25. Wynant, N., Santos, D. & Vanden Broeck, J. The evolution of animal Argonautes: evidence for the absence of antiviral AGO Argonautes in vertebrates. *Sci. Rep.* **7**, 9230 (2017).
26. Reuter, M. et al. Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature* **480**, 264–267 (2011).
27. Wu, P. H. et al. The evolutionarily conserved piRNA-producing locus *pi6* is required for male mouse fertility. *Nat. Genet.* **52**, 728–739 (2020).
28. Arif, A. et al. The tiny, conserved zinc-finger protein GTSF1 helps PIWI proteins achieve their full catalytic potential. Preprint at https://doi.org/10.1101/2021.05.04.442675 (2021).
29. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
30. Sienski, G., Donertas, D. & Brennecke, J. Transcriptional silencing of transposons by Piwi and Maelstrom and its impact on chromatin state and gene expression. *Cell* **151**, 964–980 (2012).
31. Le Thomas, A. et al. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. *Genes Dev.* **27**, 390–399 (2013).
32. Aravin, A. A. et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31**, 785–799 (2008).
33. Kuramochi-Miyagawa, S. et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* **22**, 908–917 (2008).
34. Nott, T. J. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).

# Methods

### Recombinant PIWI cloning, mutagenesis and expression

DNA fragments encoding truncations of PIWI proteins from various animals (including human, silk moth, fruit fly, worm, zebrafish, flour beetle and sponge) were cloned into a modified form of pFastBac HTA (Thermo Fisher) to generate expression plasmids for the Bac-to-Bac baculovirus expression system (Thermo Fisher). Resulting baculoviruses were used to infect Sf9 cells and recombinant protein expression levels were assessed 3 days post infection by Ni-NTA purification (Qiagen) and SDS–PAGE. The most highly expressed constructs for the top two PIWIs were N-terminal truncations of *E. fluviatilis* Piwi A (NCIB: AB533505, residues 219–987) and *Bombyx mori* Siwi (NCIB: AB332313, residues 88–899).

### Preparation of protein–guide RNA complexes

*Ef*Piwi–guide RNA, Siwi–guide RNA and hAGO2–guide RNA samples were purified as described previously for hAGO2[35], following a variation on the Arpon method[36], with the exception that the final size-exclusion purification step was omitted from PIWI preparations to maintain sufficient yields. Purified PIWI–guide complexes were stored at −80 °C at a protein concentration range of 5 to 50 μM in 150 mM NaCl, 0.5 mM TCEP, 5–10% glycerol, 20 mM Tris pH 8.

### Grid preparation for cryo-EM

*Ef*Piwi–piRNA-target ternary complexes were formed by adding 1.2 molar equivalents of target RNA to purified *Ef*Piwi–guide complex and incubating on ice for 10 min in the following buffer: 150 mM NaCl, 0.5 mM TCEP, 50 mM Tris, pH 8. Divalent ions were excluded to prevent target cleavage. Nonetheless, final models include two $Mg^{2+}$ ions (one bound to the 5′ phosphate and one in the endonuclease active site), which seem to have co-purified with *Ef*Piwi. Additionally, for the extended (2–25 paired) target, the procedure used to form the ternary complex seems to have been sub-optimal as many particles were not bound to the target RNA (possibly connected to reduced *Ef*Piwi activity at low temperatures). This turned out to be fortunate as it also enabled determination of the *Ef*Piwi–piRNA binary structure (shown in Fig. 1a) from the same dataset. To prepare samples for cryo-EM, 3.5 μl *Ef*Piwi-guide–target complex at 2.5 mg $ml^{-1}$ was added onto freshly plasma-cleaned (75% nitrogen, 25% oxygen atmosphere at 15 W for 7 s in Solarus plasma cleaner, Gatan) 300 mesh holey gold grids (UltrAuFoil R1.2/1.3, Quantifoil). Excess sample solution was removed from grids by blotting with Whatman No.1 filter paper for 5–7 s. Samples were immediately vitrified by plunge-freezing in liquid ethane at −179 °C using a manual plunge-freezing device. Grid vitrification was performed in a cold room maintained at 4 °C with relative humidity between 95–98% to minimize sample evaporation.

### Cryo-EM data acquisition

Cryo-EM data were acquired on a 200 kV Talos Arctica (Thermo Fisher Scientific) transmission electron microscope. Micrographs were acquired using a K2 Summit (Gatan) direct electron detector, operated in electron-counting mode, using the automated data collection software Leginon[37] by image shift-based movements from the centre of four adjacent holes to target the centre of each hole for exposures. Each micrograph for the *Ef*Piwi–piRNA–target (2–25) complex (which fortuitously also contained many *Ef*Piwi–piRNA binary complexes) was collected as 48 dose-fractionated movie frames over 12 s and with a cumulative electron exposure of 47.33 $e^{-}$ $Å^{-2}$. For the *Ef*Piwi–piRNA–target (2–16) complex, each micrograph was acquired as 64 dose-fractionated movie frames over 16 s with a cumulative electron exposure of 47.33 $e^{-}$ $Å^{-2}$. Both datasets were collected at a nominal magnification of 36 kx, corresponding to 1.15 Å per pixel on the detector, with random nominal defocus values varying between 1 μm and 1.6 μm. A total of 1,765 micrographs were collected for the *Ef*Piwi–piRNA–extended target (pairing to 2–25) complex (these micrographs also contained the *Ef*Piwi–piRNA binary complex). A total of 1,881 micrographs were collected for the *Ef*Piwi–PiwipiRNA–target (pairing to 2–16) complex.

### Image processing and 3D reconstruction

Beam-induced motion correction and radiation damage compensation over spatial frequencies (dose-weighting) of the raw movies was performed using UCSF MotionCor2[38] implemented in the Appion[39] image-processing workflow. Motion-corrected, summed micrographs were imported into the RELION 2.0[40] data processing pipeline. Contrast transfer function (CTF) parameters for these micrographs were estimated using CTFFind4[41]. Laplacian of Gaussian-based automated particle-picking program in RELION was used for picking 3,280,351 and 2,551,046 particles from the *Ef*Piwi–piRNA and *Ef*Piwi–piRNA-target micrographs, respectively. Picked particles were extracted from the micrographs with a 160-pixel box and subjected to 2D classification in RELION. After discarding particles that belonged to classes containing non-particle features, aggregates and low-resolution features, new stacks of particles from 2D classes containing different orientations of the complexes and high-resolution features were created. A subset of 2D classes from the *Ef*Piwi–piRNA–target (g2–g25 paired) dataset resolved features corresponding to a smaller complex with an extended RNA duplex. The 608,488 particles belonging to these classes were isolated into a new particle stack for further processing. To serve as an initial model for 3D analyses of the *Ef*Piwi–piRNA and *Ef*Piwi–piRNA–target complexes, a 40 Å low pass filtered map was generated from the Siwi crystal structure (Protein Data Bank (PDB) ID: 5GUH) using the molmap function in UCSF Chimera[42,43]. An initial model for 3D analyses of the smaller subcomplex particles was generated using the Cryosparc v1[44] ab initio reconstruction program. The selected particle stacks corresponding to the three distinct complexes were subjected to multiple iterations of 3D classification in RELION and particles belonging to the most well-resolved 3D class for each complex were selected for downstream 3D processing. After 3D classification, 125,041, 118,493 and 116,655 particles from the best-resolved class for *Ef*Piwi–piRNA, *Ef*Piwi–piRNA–target (g2–g16 paired) and the smaller *Ef*Piwi–piRNA–target (g2–g25 paired), respectively, were re-extracted with a 160-pixel box from the respective micrographs with re-centred coordinates. These particles were then subjected to 3D refinement in RELION. 3D binary masks for refinement were generated using 15 Å low-pass filtered selected class volume for each of the complexes with a 5-pixel expansion and 8-pixel Gaussian fall-off in RELION. The final reconstructed maps for *Ef*Piwi–piRNA, *Ef*Piwi–piRNA–target (g2–g16 paired) and *Ef*Piwi–piRNA–target (g2–g26 paired) were at 3.8 Å, 3.5 Å and 8.6 Å (at a Fourier shell correlation (FSC) value of 0.143), respectively. Local resolution for these complexes was determined using the local resolution estimation program in RELION and the local-resolution-based filtered maps were used for atomic model building. Maps for *Ef*Piwi–piRNA and *Ef*Piwi–piRNA–target complexes were trimmed to a box size of 90 pixels for deposition into the Electron Microscopy Data Bank (EMDB). Directional FSCs were estimated using the 3DFSC server[45] (https://3dfsc.salk.edu).

### Model building and refinement

An initial model for *Ef*Piwi was obtained by threading the *Ef*Piwi primary sequence onto the Siwi crystal structure (PDB: 5GUH) using SWISS-MODE[46]. Discrete domains were then docked into the *Ef*Piwi–piRNA reconstruction using UCSF Chimera, followed by manual model building using Coot[47]. The *Ef*Piwi–guide–target model was built in a similar fashion, using the *Ef*Piwi–guide structure as an initial model. Models were refined through iterative rounds of manual building and fixing of geometric and rotameric outliers in Coot and real-space refinement optimizing global minimization, atomic displacement parameters and local grid search using PHENIX[48]. Base pairing between guide and target RNAs was maintained by including hydrogen atoms. Most of the

residues within the N domain of both models were truncated to alanine at the end of refinement to reflect a lack of supporting cryo-EM density. Model validation was performed using MolProbity (http://molprobity.biochem.duke.edu)[49] and PDB validation servers (https://www.rcsb.org). Structural figures were made using PyMOL (Schrödinger, LLC) and UCSF ChimeraX[50]. Prediction of seed-gate structure in extended Piwi family members was performed by PSIPRED 4.0[51,52]. The structure of *Ef*Piwi was compared to previous human AGO structures[53–55] in Fig. 1.

**Equilibrium target-RNA-binding assays**

Equilibrium dissociation constants for seed-matched target RNAs were determined as described previously[11]. Dissociation constants for targets with extensive complementarity were too small to be accurately measured and were therefore calculated as the quotient of measured target dissociation ($k_{off}$) and association ($k_{on}$) rates.

**Target-RNA dissociation assays**

A dot-blot apparatus (GE Healthcare) was used to separate protein–RNA complexes from free (unbound) RNA. Protein–RNA complexes were immobilized on Protran nitrocellulose membrane (0.45 µm pore size, Whatman, GE Healthcare) and unbound RNA was immobilized on Hybond Nylon membrane (Amersham, GE Healthcare). Membranes were stacked so that the sample is first pulled through the protein-binding membrane, and any unbound RNA passes through and binds to the RNA-binding membrane.

Target dissociation rates were determined by incubating guide-loaded *Ef*Piwi, Siwi or hAGO2 samples with 0.1 nM $^{32}$P 5′-radiolabelled target RNA in binding reaction buffer (30 mM Tris pH 8.0, 100 mM potassium acetate, 0.5 mM TCEP, 0.005% (v/v) NP-40, 0.01 mg ml$^{-1}$ baker's yeast tRNA) in a single reaction with a volume of 100 µl per time point planned for the experiment (for example, 1,000 µl for 10 time points) at room temperature for 60 min. The concentration of hAGO2, *Ef*Piwi or Siwi–guide RNA complex was 5 nM.

After sample equilibration, a zero-time point was taken by applying 100 µl of the reaction to the dot-blot apparatus under vacuum, followed by 100 µl of ice-cold wash buffer (30 mM Tris pH 8.0, 0.1 M potassium acetate, 0.5 mM TCEP). The dissociation time course was started by the addition of 300 nM (final concentration) unlabelled target RNA. Aliquots of 100 µl were taken at various times and immediately applied to a dot-blot apparatus under vacuum, followed by 100 µl of ice-cold wash buffer. Time points ranged from 0.25 to 100 min. Membranes were air-dried and visualized by phosphorimaging. Quantification of the $^{32}$P signal was performed using ImageQuant TL (GE Healthcare). The fraction of target RNA bound was calculated as the ratio of bound to total (bound + free) target RNA for various concentrations of AGO2–guide, *Ef*Piwi–guide or Siwi–guide RNA complexes. Dissociation rates were calculated by plotting data as fraction bound versus time and fitting to a one-phase exponential curve using Prism v.8.0 (GraphPad).

**Target RNA association assays**

Target association rates were determined by incubating 5 pM $^{32}$P 5′-radiolabelled target RNA in binding reaction buffer (30 mM Tris pH 8.0, 100 mM potassium acetate, 0.5 mM TCEP, 0.005% (v/v) NP-40, 0.01 mg ml$^{-1}$ baker's yeast tRNA) in a single large reaction with a volume of 100 µl per time point planned for the experiment (for example, 1,000 ul for 10 time points) at room temperature for 15 min.

A zero-time point was taken by applying 100 µl of the reaction to a dot-blot apparatus under vacuum, followed by 100 µl of ice-cold wash buffer (30 mM Tris pH 8.0, 0.1 M potassium acetate, 0.5 mM TCEP). The association reaction was started by the addition of guide-loaded *Ef*Piwi or Siwi at concentrations ranging from 0.05 to 15 nM. Aliquots of 100 µl were taken at various time points and applied to a dot-blot apparatus under vacuum, followed by 100 µl of ice-cold wash buffer as before. Time points ranged from 0.25 to 15 min. Membranes were air-dried and visualized, the signal was quantified, and the bound fraction was calculated as described above. Because target dissociation was very slow compared with association, and the target RNA concentrations were always at least 10-fold less than the total *Ef*Piwi–piRNA or Siwi–piRNA concentration, association was treated as an irreversible pseudo first-order system. Data were fit to a one-phase exponential curve using Prism v.8.0 (GraphPad) to determine an observed binding rate constant ($k_{obs}$). $k_{obs}$ values were proportional to protein concentration over the range used in our experiments. The association rate ($k_{on}$) was calculated by dividing $k_{obs}$ by the concentration of the *Ef*Piwi–piRNA or Siwi–piRNA complex used in the experiment.

**Target RNA cleavage assays**

Purified *Ef*Piwi–guide, Siwi–guide, or hAGO2–guide RNA complexes (100 nM, final concentration) were incubated at 37 °C with complementary 5′-$^{32}$P-labelled target RNAs (10 nM, final concentrations) in reaction buffer composed of 30 mM Tris pH 8.0, 0.1 M NaCl, 2 mM $MgCl_2$, 0.5 mM TCEP, and 0.01 mg ml$^{-1}$ baker's yeast tRNA. Target cleavage was stopped at various times by mixing aliquots of each reaction with an equal volume of denaturing gel loading buffer (98% w/v formamide, 0.025% xylene cyanol, 0.025% w/v bromophenol blue, 10 mM EDTA pH 8.0). Intact and cleaved target RNAs were resolved by denaturing PAGE (15%) and visualized by phosphorimaging. Quantification of signal was performed using ImageQuant TL (GE Healthcare).

Adjustments were made to this protocol to determine which divalent cations were catalytic with *Ef*Piwi or Siwi, by replacing 2 mM $MgCl_2$ with 2 mM $MnCl_2$, $CoCl_2$, $CaCl_2$ or $NiCl_2$. Similar adjustments were made to titrate in $MgCl_2$ or $MnCl_2$.

To determine the piRNA-target pairing length requirement for cleavage, 100 nM protein–guide RNA complex and 1 nM radiolabelled target RNA of complementarity g2–g15, g2–g16, g2–g17, g2–g18, g2–g19, g2–g20 or g2–g21 were combined in buffer with a mixture of 2 mM $MgCl_2$ + 2 mM $MnCl_2$. Reactions proceeded for 1 h at 37 °C before an equal volume of denaturing gel loading buffer was added to stop the reaction. Reactions were run on denaturing PAGE (15%) and analysed as described above.

To determine the effects of the mismatch triplets on slicing, 100 nM protein–guide RNA complex and 1 nM radiolabelled target RNA with mismatches from g2–g4, g5–g7, g8–g10, g11–g13, g14–g16, g17–g19 or g20–g22 were combined in buffer with 2 mM $MgCl_2$ + 2 mM $MnCl_2$. Reactions proceeded for 1 h at 37 °C before an equal volume of denaturing gel loading buffer was added to stop the reaction. Reactions were run on denaturing PAGE (15%) and analysed as described above. Uncropped and unprocessed scans of gels and quantification of replicate data can be found in Supplementary Fig. 1.

**Pooled target cleavage assay**

A pool of 256 target RNAs containing molecules with combinations of mismatched pairing to g11–g18 of our standard piRNA sequence was synthesized by Integrated DNA Technologies. Synthesis included piRNA-matched and mismatched phosphoramidites mixed in a 1:1 ratio at steps generating t11–t18, resulting in 256 unique target sequences. All mismatches are expected to reside on the 5' cleavage product because *Ef*Piwi cleaves targets between t10–t11.

The target RNA pool (27 nM) was incubated with *Ef*Piwi–miR122 25mer complex (133 nM) in cleavage buffer (100 mM NaCl, 20 mM Tris pH 8, 2 mM $MgCl_2$, 2 mM $MnCl_2$, 0.05 mg ml$^{-1}$ tRNA, 0.5 mM TCEP) for 60 min at 37 °C. Cleavage products from the 5' end of the target (complementary to the piRNA 3' end) were isolated by denaturing gel electrophoresis, cut out of the gel, solubilized and purified by ethanol precipitation followed by passage through an Oligo Clean and Concentrator column (Zymo Research). Following this step, cleavage products had 5' and 3' adapters ligated with the T4 RNA Ligase and T4 RNA Ligase 2 truncated enzymes (New England Biolabs), respectively, and were subjected to a reverse transcriptase reaction using the SuperScript II reverse transcriptase (Lifetech). Resulting cDNA was sequenced on a

NextSeq 500 (Illumina). To control for variations in 3' adapter ligation efficiency to the cleaved sequences, a 'pre-cleaved' control was used, which was a pool of 256 RNAs synthesized to mimic all possible 5' cleavage products, but with a three-base barcode distinct from the intact target sequences. Experimental and pre-cleaved target pools were combined before adapter ligation, reverse transcription and sequencing.

After sequencing, unique reads were trimmed of the adapter sequences and those of the correct read length were retained using cutadapt[56]. Remaining reads were sorted into two subsets, experimental or ligation ('pre-cleaved') control, using the three-base barcodes. A read count analysis was performed on each subset, and counts were normalized by dividing the experimental counts by the pre-cleaved counts and multiplying by 1,000 so that the total counts for each observed sequence was ≥1. Targets were then sorted by abundance, resulting in the histogram in Fig. 4c.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

## Data availability

Maps for the *Ef*Piwi–piRNA and *Ef*Piwi–piRNA-target complexes have been deposited in the EMDB under accession numbers EMD-23061 and EMD-23063, respectively. Corresponding atomic models have been deposited in the PDB under accession numbers 7KX7 and 7KX9. The *Ef*Piwi(MID/Piwi)–piRNA–long-target complex map has been deposited in the EMDB under accession number EMD-23062. Source data are provided with this paper.


35. Sheu-Gruttadauria, J. & MacRae, I. J. Phase transitions in the assembly and function of human miRISC. *Cell* **173**, 946–957.e16 (2018).
36. Flores-Jasso, C. F., Salomon, W. E. & Zamore, P. D. Rapid and specific purification of Argonaute–small RNA complexes from crude cell lysates. *RNA* **19**, 271–279 (2013).
37. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
38. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
39. Lander, G. C. et al. Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J. Struct. Biol.* **166**, 95–102 (2009).
40. Kimanius, D., Forsberg, B. O., Scheres, S. H. & Lindahl, E. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016).
41. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
42. Goddard, T. D., Huang, C. C. & Ferrin, T. E. Visualizing density maps with UCSF Chimera. *J. Struct. Biol.* **157**, 281–287 (2007).
43. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
44. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
45. Tan, Y. Z. et al. Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat. Methods* **14**, 793–796 (2017).
46. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
47. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
48. Liebschner, D. et al. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D* **75**, 861–877 (2019).
49. Williams, C. J. et al. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
50. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
51. Buchan, D. W. A. & Jones, D. T. The PSIPRED protein analysis workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
52. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999).
53. Nakanishi, K. et al. Eukaryote-specific insertion elements control human ARGONAUTE slicer activity. *Cell Rep.* **3**, 1893–1900 (2013).
54. Park, M. S. et al. Human Argonaute3 has slicer activity. *Nucleic Acids Res.* **45**, 11867–11877 (2017).
55. Park, M. S. et al. Multidomain convergence of Argonaute during RISC assembly correlates with the formation of internal water clusters. *Mol. Cell* **75**, 725–740.e6 (2019).
56. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).



**Acknowledgements** We thank N. Funayama for the *E. fluviatilis piwi-a* cDNA clone, Y. Tomari for the Siwi cDNA clone, and I. H. Segel for advice about measuring binding reactions with very slow off-rates. The research of G.C.L. is supported by NIH grant R21AG067594 and an Amgen Young Investigator Award. The research of I.J.M. is supported by NIH grant R35GM127090.


**Author contributions** T.A.A. prepared *Ef*Piwi–piRNA, Siwi–piRNA, and hAGO2–miRNA samples, performed biochemical experiments, built *Ef*Piwi models and co-wrote the manuscript. S.C. prepared cryo-EM samples, collected data, produced high-resolution reconstructions and assisted with model building. S.M.H. identified and developed *Ef*Piwi as a source of active Piwi protein. Y.X. helped to develop *Ef*Piwi and established purification protocols. G.C.L. provided structural insights and guidance in cryo-EM data collection and analysis. I.J.M. provided structural and mechanistic insights and co-wrote the manuscript.

**Competing interests** The authors declare no competing interests.

**Additional information**

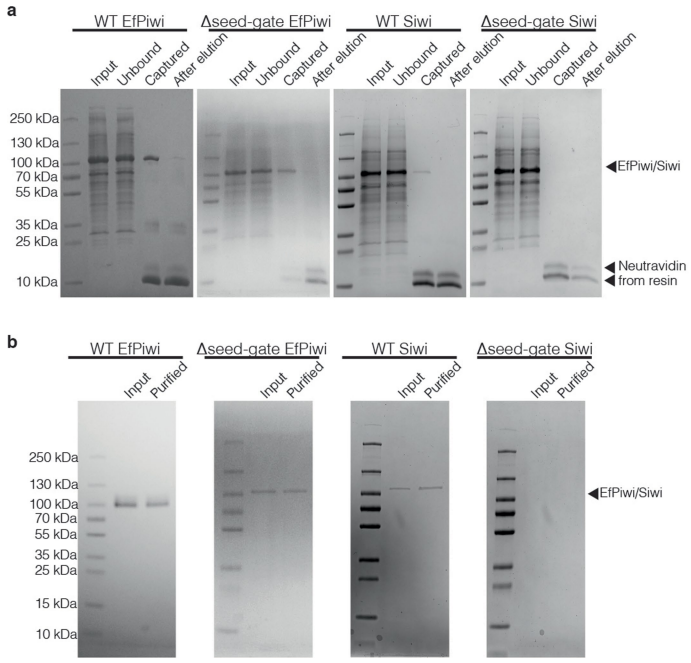

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03856-x.

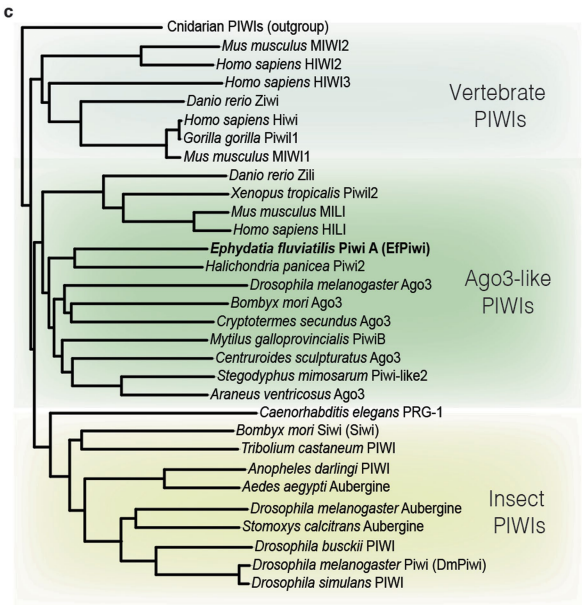


**Correspondence and requests for materials** should be addressed to I.J.M.

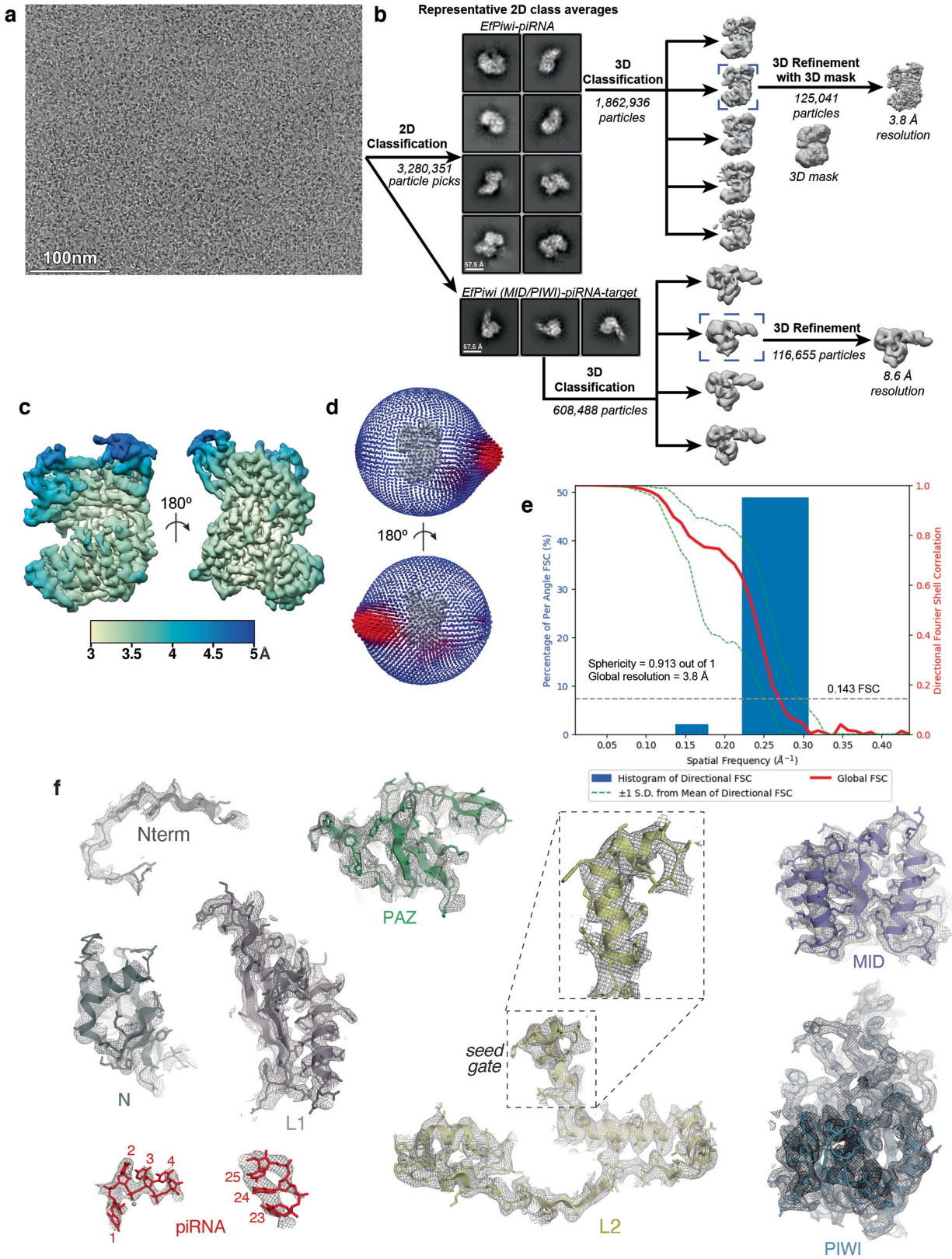


**Peer review information** *Nature* thanks Hong-Wei Wang and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



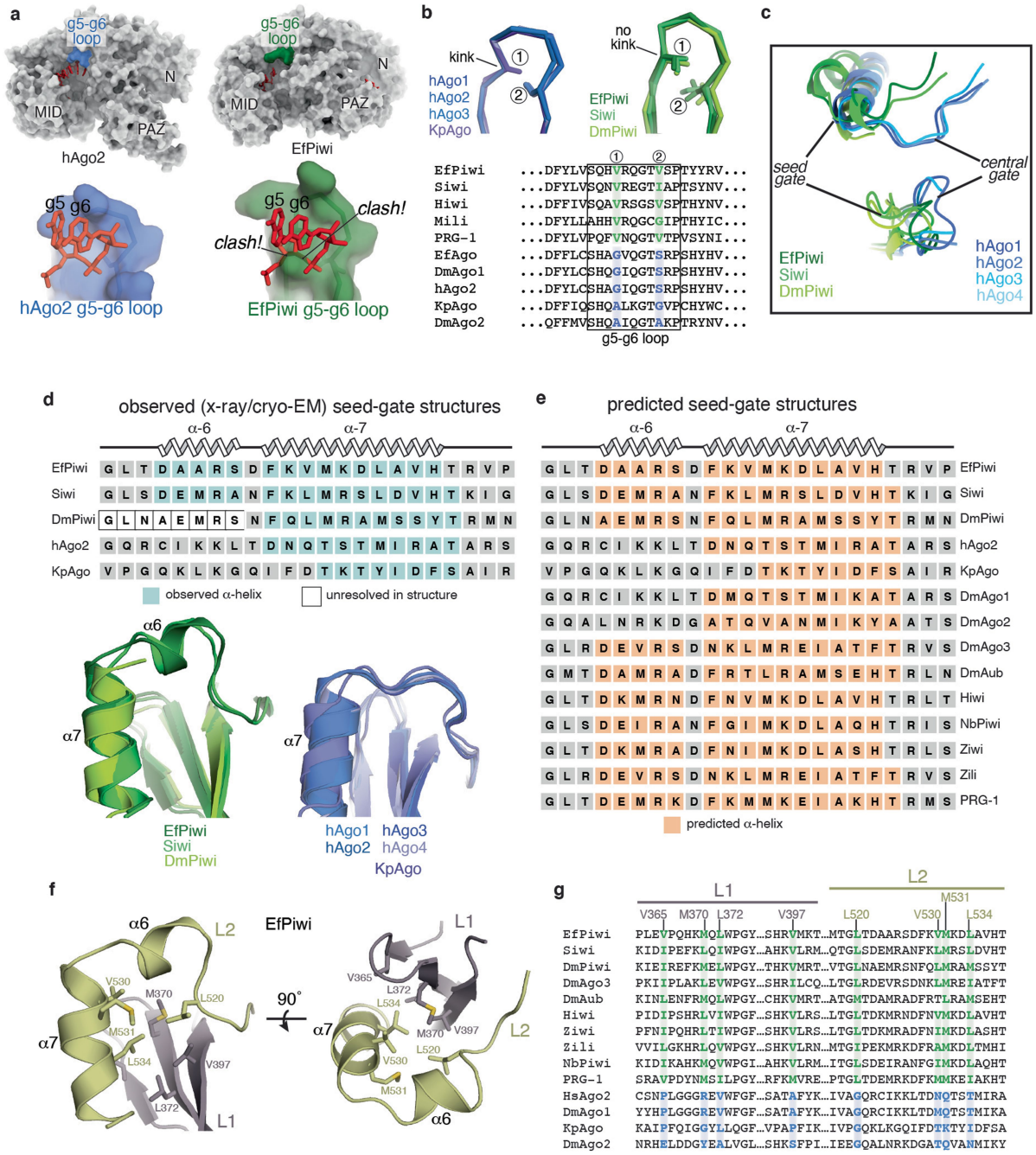




**Extended Data Fig. 1 | Piwi protein purification and extended Piwi family tree. a**, Coomassie-stained SDS PAGE of piRNA-loaded PIWI proteins captured using an immobilized complementary oligonucleotide. Input shows partially purified protein samples that were incubated with capture resin. Unbound shows protein that did not bind the resin. Captured shows protein retained on the resin after washing (eluted by boiling in SDS). After elution shows protein retained on the resin after incubation with the competitor oligonucleotide (eluted by boiling in SDS). **b**, Capture-purified PIWI proteins before and after anion exchange purification. Input fraction shows samples after elution by competitor oligonucleotide in capture-purification step. Purified indicates the final purification products. Note: Δseed-gate Siwi was captured at such low levels that it was unclear whether any active Siwi was obtained until observing the sample's ability to specifically bind $^{32}$P-labeled target RNAs. **c**, Phylogenetic tree of PIWI proteins shows EfPiwi belongs to the ancient *Drosophila* Ago3-like branch.

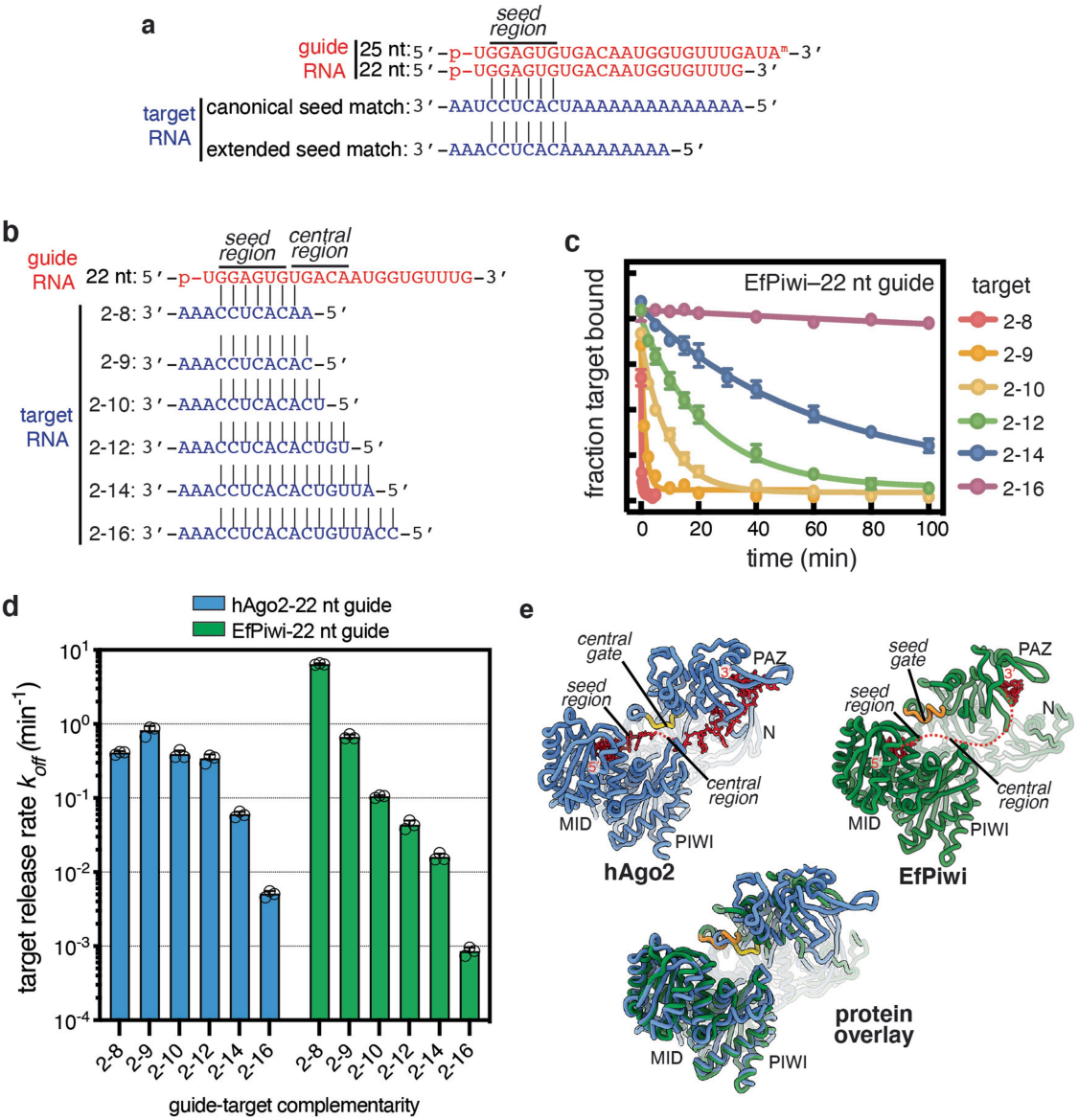



**Extended Data Fig. 2** | See next page for caption.

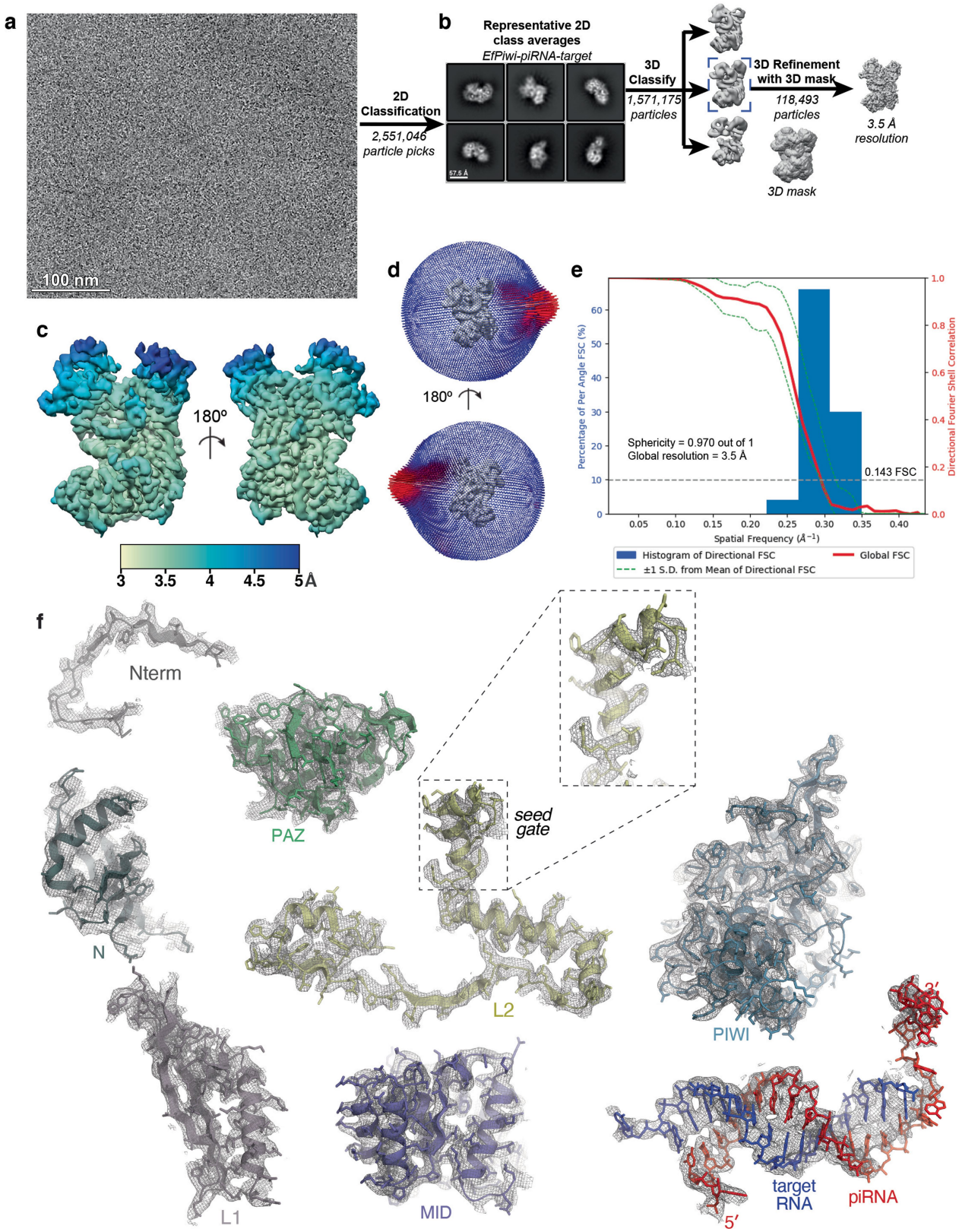
**Extended Data Fig. 2 | Imaging and processing of the EfPiwi-piRNA complex (and EfPiwi-piRNA-long target complex). a**, Representative cryo-EM micrograph (1,765 micrographs collected in total). Input sample contained EfPiwi-piRNA and a long target RNA (complementary to piRNA nucleotides g2–g25). **b**, Cryo-EM data processing workflow. The data set contained two populations of well resolved particles, one for the binary EfPiwi-piRNA complex and another for the ternary EfPiwi-piRNA-long target complex. Particles isolated from micrographs were sorted by reference-free 2D classification. Only particles containing high-resolution features for the intact complex were selected for downstream processing. 3D classification was used to further remove low-resolution or damaged particles, and the remaining particles were refined to obtain a 3.8 Å reconstruction for the EfPiwi-guide complex, and 8.6 Å for the ternary EfPiwi-piRNA-long target complex. **c**, The final 3D map for the EfPiwi-piRNA complex coloured by local resolution values, where the majority of the map was resolved between 3.5 Å and 4 Å with the flexible PAZ and N domains having lower resolution. **d**, Angular distribution plot showing the Euler angle distribution of the EfPiwi-piRNA particles in the final reconstruction. The position of each cylinder corresponds to the 3D angular assignments and their height and colour (blue to red) corresponds to the number of particles in that angular orientation. **e**, Directional Fourier Shell Correlation (FSC) plot representing 3D resolution anisotropy in the reconstructed map, with the red line showing the global FSC, green dashed lines correspond to ±1 standard deviation from mean of directional resolutions, and the blue histograms correspond to percentage of directional resolution over the 3D FSC. **f**, EM density quality of EfPiwi-piRNA complex. Individual domains of EfPiwi fit into the EM density, EM density shown in mesh; molecular models (coloured as in Fig. 2) shown in cartoon representation with side chains shown as sticks; piRNA shown in stick representation.




**Extended Data Fig. 3 | Conserved structural features in extended Piwi family. a**, Surface of hAGO2 (left) and EfPiwi (right), highlighting g5-g6 nucleotide-binding loops. Superimposing g5-g6 nucleotides (red sticks) from hAGO2 onto EfPiwi results in steric clashes. **b**, g5-g6 loop in AGO structures (left) is kinked, enabling pre-organization of seed 3' end. Equivalent loop in Piwi structures (right) cannot kink due to bulky residues (labeled positions 1 and 2), conserved in Piwi family. **c**, Close up superposition of central-gate and seed-gate structures in AGO and PIWI proteins, respectively. **d**, Superposition of seed-gate regions from all known Piwi (left) and AGO (right) structures, with secondary structure schematics shown above. **e**, Secondary structure predictions indicate the α6 extension is a defining feature of the Piwi family. Predictions were by PSIPRED 4.0. **f**, L1-L2 interface near seed-gate in EfPiwi. Hydrophobic residues buried at the L1-L2 interface are shown. **g**, Sequence alignment shows L1-L2 interface residues in EfPiwi are broadly conserved in Piwis (green) and distinct from the equivalent residues in AGOs (blue).

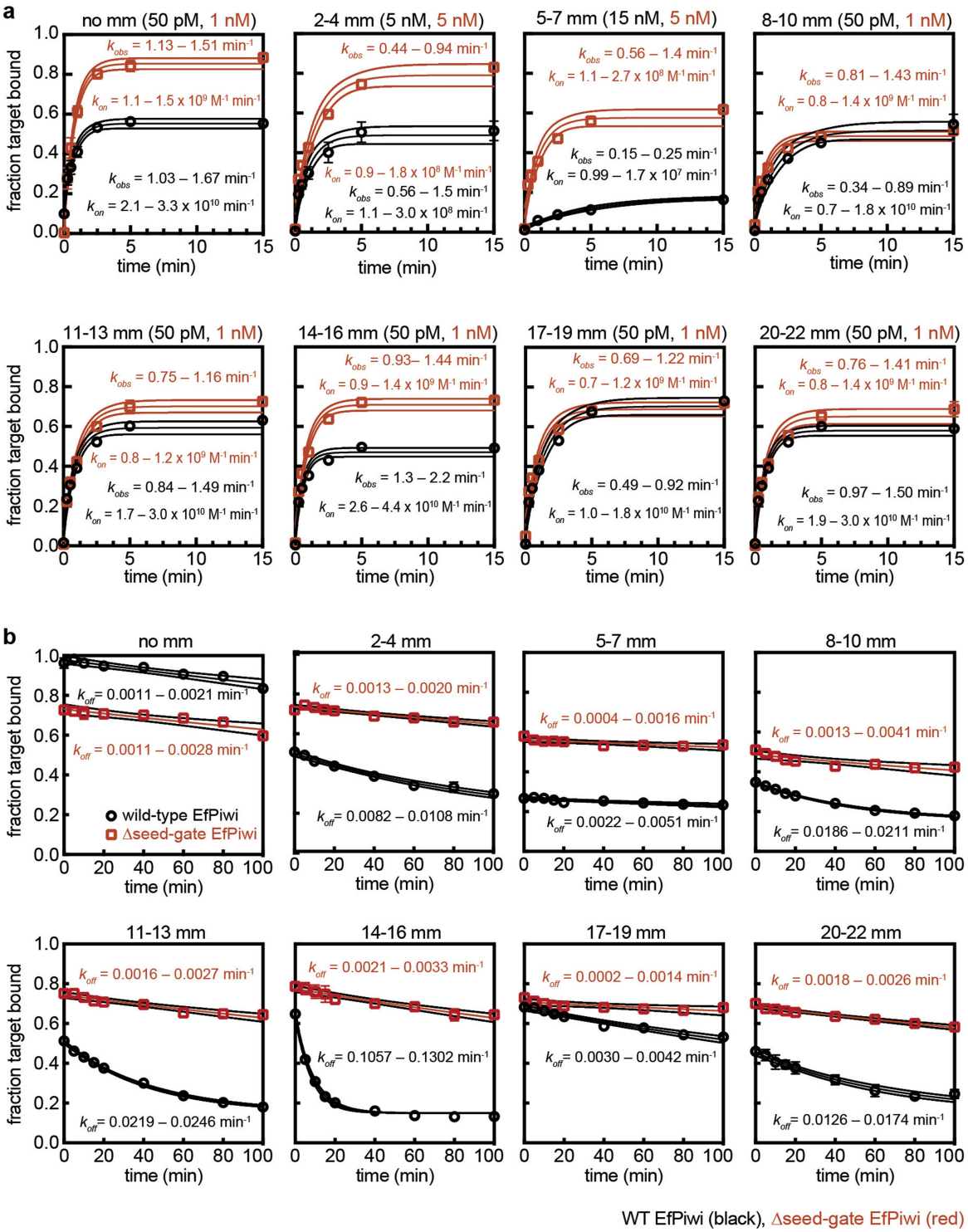



**Extended Data Fig. 4 | Target release from hAGO2 and EfPiwi loaded with identical guides. a**, Schematic of pairing between guide RNAs and seed-matched target RNAs used in main text Fig. 2b, c. **b**, Schematic of pairing between 22 nt guide RNA and target RNAs spanning the seed and central regions. **c**. Release of $^{32}$P-labeled target RNAs from EfPiwi-22 nt guide in the presence of excess unlabeled target RNA over time. **d**, Release rates of target RNAs from hAGO2-22 nt guide (data from Fig. 2d, left) and EfPiwi-22nt guide (**c**). Results show hAGO2 and EfPiwi create distinct binding properties for the same guide RNA. All plotted data are the mean values of triplicate measurements. Error bars indicate SD. **e**, Ribbon representation of hAGO2, EfPiwi, and an overlay illustrating relative positions of the central-gate and seed-gate. In **c**, **d**, $n = 3$ independent experiments, data are mean ± s.d.

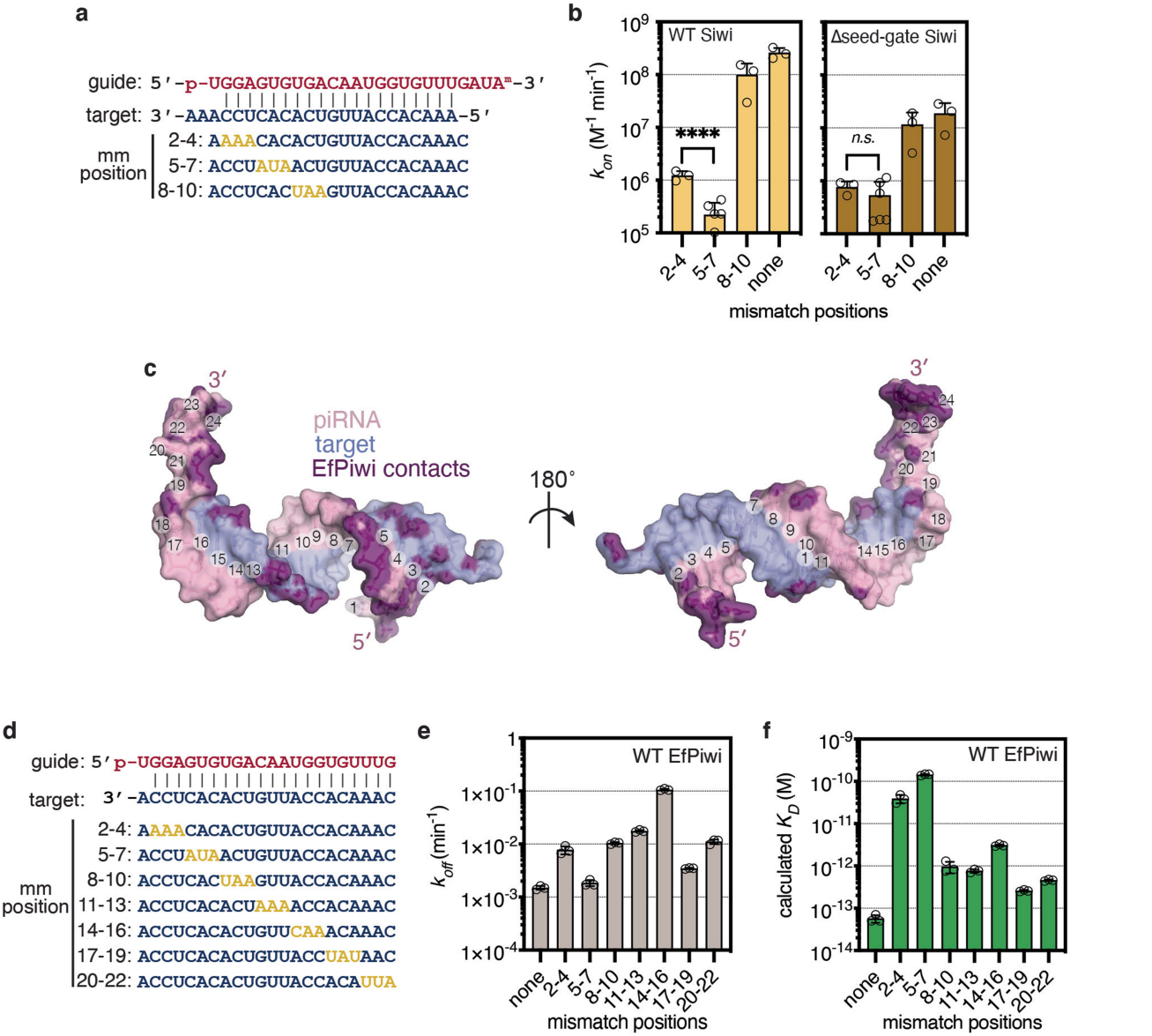



**Extended Data Fig. 5** | See next page for caption.

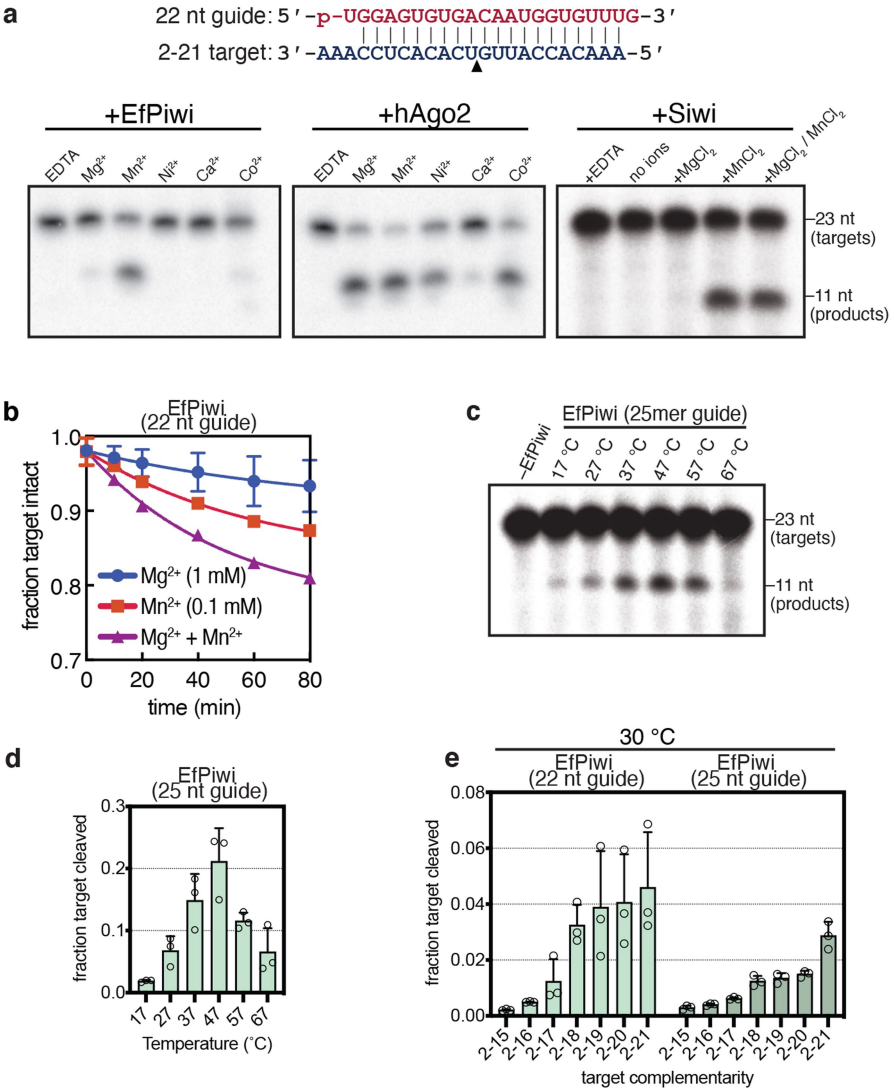
**Extended Data Fig. 5 | Imaging and processing of the EfPiwi-piRNA-target complex. a**, Representative cryo-EM micrograph of EfPiwi-piRNA-target complex (1,881 micrographs collected in total). **b**, Workflow for processing EfPiwi-piRNA-target complex dataset. Particles isolated from micrographs were sorted by reference-free 2D classification. Only particles containing high-resolution features for the intact complex were selected for downstream processing. 3D classification was used to further remove low-resolution or damaged particles, and the remaining particles were refined to obtain a 3.5 Å map. **c**, The EfPiwi-piRNA-target complex map coloured by local resolution. **d**, Euler angle distribution plot for the EfPiwi-piRNA-target complex particles. **e**, Directional Fourier Shell Correlation (FSC) plot representing 3D resolution anisotropy in the reconstructed map. Red line shows global FSC; green dashed lines ±1 standard deviation from mean of directional resolutions; blue histograms indicate percentage of directional resolution over the 3D FSC. **f**, EM density quality of EfPiwi-piRNA-target complex. Individual domains of EfPiwi and RNAs fit into the EM density; EM density shown in mesh; protein models shown in cartoon representation (coloured as in Fig. 1) with side chains shown as sticks; RNAs shown in stick representation.



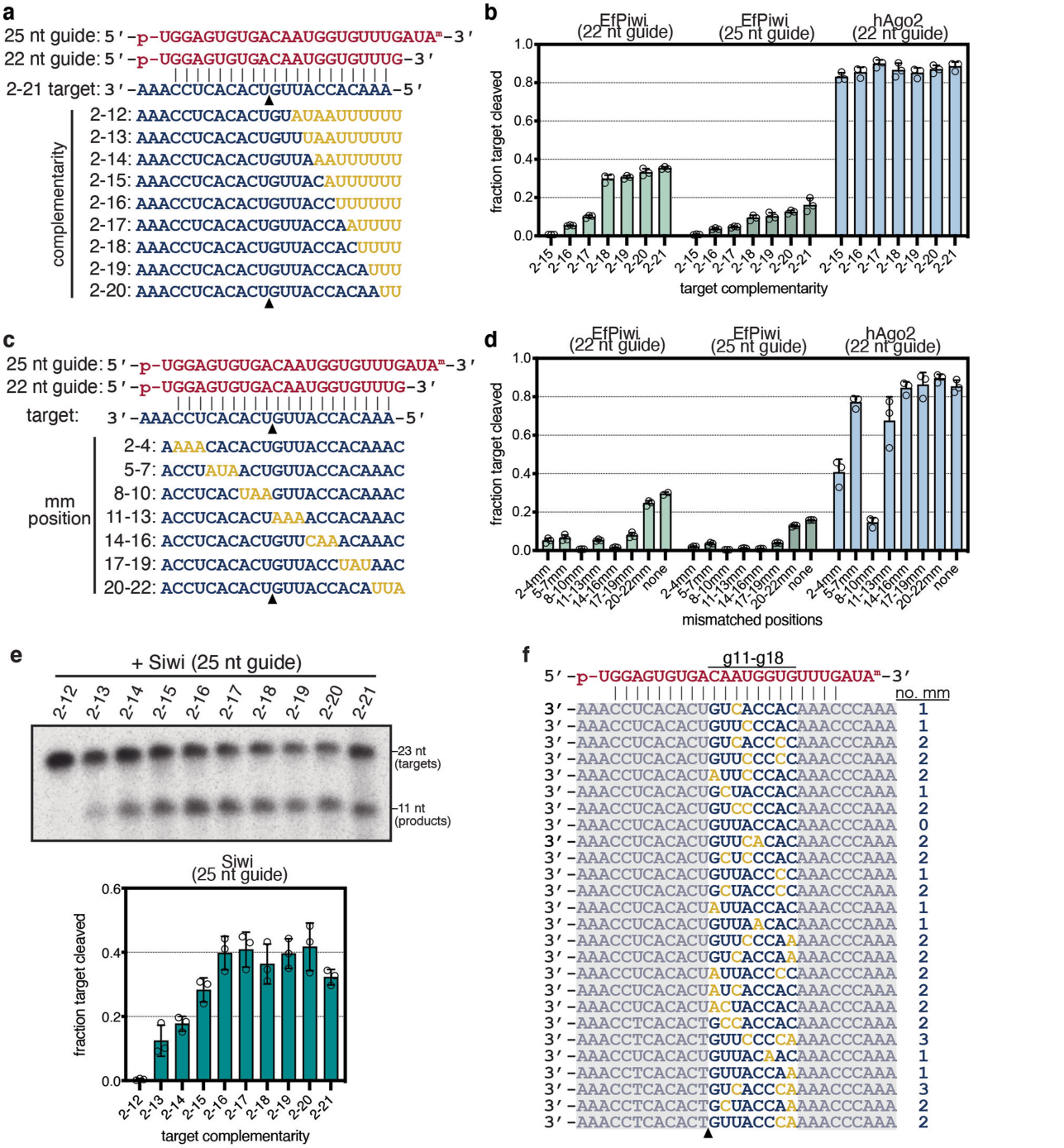

**Extended Data Fig. 6 | EfPiwi target binding data. a**, Raw data for $k_{on}$ values shown in Fig. 3f. Plots of target RNAs with mismatches (sequences shown in Extended Data Fig. 6c) binding to EfPiwi-piRNA complexes over time. Protein concentrations used in each experiment are indicated at top of each graph. 95% confidence limits of observed association rates ($k_{obs}$) and $k_{on}$ values indicated. WT EfPiwi (black), Δseed-gate EfPiwi (red). **b**, Raw data for $k_{off}$ values shown in Extended Data Fig. 7e. Plots of target RNAs with mismatches (mm) dissociating from EfPiwi-piRNA complexes over time. All data were fit to a plateau value of 0.15. 95% confidence limits of $k_{on}$ values indicated. All data points were measured three times. Error bars indicate SEM. Center line indicates best fit to data. Surrounding lines indicate 95% confidence limits. In all panels, $n = 3$ independent experiments, data are mean ± s.d.




**Extended Data Fig. 7 | Target binding with mismatches. a**, Guide-target pairing schematic for select mismatched targets binding Siwi-piRNA complexes. Mismatches coloured gold. **b**, Association rates of target RNAs (shown in **a**) with wild-type Siwi (left) and Δseed-gate Siwi. Indicated p-values from two-sided t-test are $6.11 \times 10^{-5}$ and 0.205 for wild-type and Δseed-gate Siwi, respectively. **c**, Guide-target pairing schematic for mismatched targets used in main text Fig. 3f, and panels d and e here. Mismatches coloured gold. **d**, Dissociation rates of $^{32}$P-labeled target RNAs with three consecutive mismatches from wild-type EfPiwi. Most mismatched segments had moderate (~10-fold) effects on $k_{off}$, except 14–16 mismatches, which increased $k_{off}$ ~70 fold. **e**, Dissociation constants ($K_D$) calculated from $k_{on}$ and $k_{off}$ values for target RNAs binding wild-type EfPiwi-guide complex. **f**, Surface representation of the modeled piRNA-target duplex. piRNA nucleotides numbered at the Watson-Crick face. Non-hydrogen RNA atoms positioned ≤ 4 Å from an EfPiwi atom coloured purple. In **b**, **d**, and **e**, $n = 3$ independent experiments, data are mean ± s.d.




**Extended Data Fig. 8 | Target cleavage by EfPiwi and Siwi (part 1).** **a**, Denaturing gels showing cleavage of g2-g21 matched $^{32}$P-labeled target RNA by EfPiwi, hAGO2 or Siwi in the presence of various divalent cations (2 mM each). Schematic of piRNA-target pairing shown (top). Gels are representative results for $n = 3$ independent experiments for EfPiwi and hAGO2, and $n = 2$ independent experiments for Siwi. **b**, Time course showing cleavage of g2-g21 paired $^{32}$P-labeled target RNA by EfPiwi in the presence of $Mg^{2+}$, $Mn^{2+}$, or both at approximate physiological divalent cation concentrations. **c**, Cleavage of g2-g21 matched $^{32}$P-labeled target RNA by EfPiwi at various temperatures shows activity over the full physiological range (17–30 °C). Gel is representative of $n = 3$ independent experiments. **d**, Quantification of results (and replicates) in **c**. **e**, Cleavage of target RNAs with varying degrees of 3' complementarity by EfPiwi at 30 °C. In **b**, **d** and **e**, $n = 3$ independent experiments, data are mean ± s.d.




**Extended Data Fig. 9 | Fig. S8. Target cleavage by EfPiwi and Siwi (part 2).** **a**, Guide-target pairing schematic for targets with varying degrees of complementarity to piRNA 3' end used in Fig. 4a and **b**, **e**. **b**, Quantification target RNAs (1 nM) cleaved after treating with excess (100 nM) EfPiwi loaded with a 22 or 25 nt guide or hAGO2 loaded with a 22 nt guide for 1 h. $n = 3$ independent experiments, data are mean ± s.d. **c**, Guide-target pairing schematic for targets with 3 nt mismatched regions. **d**, Quantification mismatched target RNAs (1 nM) cleaved after treating with excess (100 nM) EfPiwi loaded with a 22 or 25 nt guide or hAGO2 loaded with a 22 nt guide for 1 h. $n = 3$ independent experiments, data are mean ± s.d. **e**, Cleavage of targets with varying degrees of complementarity to piRNA 3' end (shown panel a) by Siwi (2 mM $MnCl_2$, 37 °C). Gel is representative of $n = 3$ independent experiments, with data plotted as mean ± s.d. shown below. **f**, From Fig. 4c: sequences of the 26 target RNAs, with 0–8 mismatches opposite piRNA nucleotides g11–g18, that were most readily cleaved by EfPiwi (listed in order of cleavage product abundance). Mismatched nucleotides coloured yellow. Greyed out sequences indicate constant regions shared by all target RNAs. Triangle indicates cleavage site.

**Extended Data Table 1 | Cryo-EM data collection, refinement, and validation statistics**

| Sample name | EfPiwi-piRNA-target (2-16) | EfPiwi-piRNA | EfPiwi (MID/PIWI)-piRNA-long-target (2-25) |
|---|---|---|---|
| EMDB ID | EMD-23063 | EMD-23061 | EMD-23062 |
| PDB ID | 7KX9 | 7KX7 | |
| | | | |
| Microscope | Talos Arctica | Talos Arctica | |
| Detector (Mode) | Gatan K2 Summit, counting mode | Gatan K2 Summit, counting mode | |
| Voltage (kV) | 200 | 200 | |
| Magnification (nominal / at detector) | 36,000x / 43,478x | 36,000x / 43,478x | |
| Total electron fluence ($e^-/Å^2$) | 47.33 | 47.33 | |
| Electron flux ($e^-$/pixel/sec) | 3.91 | 5.22 | |
| Defocus range (μm) | -1.0 to -1.6 | -1.0 to -1.6 | |
| Pixel size (Å) | 1.15 | 1.15 | |
| Total exposure time (sec) | 16 | 12 | |
| Total frames/micrograph | 64 | 48 | |
| Exposure per frame ($e^-/Å^2$/frame) | 0.74 | 0.99 | |
| Micrographs collected (no.) | 1,881 | 1,765 | |
| Total extracted particles (no.) | 2,551,046 | 3,280,351 | |
| Particles used for 3D analyses (no.) | 1,571,175 | 1,862,936 | 608,488 |
| Final refined particles (no.) | 118,493 | 125,041 | 116,655 |
| Accuracy of rotations (deg.) / translations (pixels) | 1.972 / 0.581 | 2.748 / 0.785 | 8.374 / 1.739 |
| Symmetry imposed | C1 | C1 | C1 |
| Global Resolution (Å) | | | |
| FSC 0.5 (masked/unmasked)<br>FSC 0.143 (masked/unmasked) | 3.8/4.0<br>3.4/3.6 | 4.3/4.5<br>3.8/3.9 | 8.6 / 10.8<br>8.6 / 10.8 |
| FSC Sphericity | 0.97 | 0.91 | |
| Local resolution range (Å) | 3.5 – 6.0 | 3.5 – 6.0 | |
| Map Sharpening *B* factors ($Å^2$) | -80 | -80 | |
| **Model composition** | | | |
| Non-hydrogen atoms | 6578 | 5828 | |
| Protein residues | 734 | 729 | |
| RNA residues | 40 | 7 | |
| Mg2+ ions | 2 | 2 | |
| **Refinement** | | | |
| Refinement package (s) | Phenix | Phenix | |
| Map Correlation Coefficient | | | |
| Global | 0.79 | 0.82 | |
| Local | 0.78 | 0.83 | |
| R.m.s. deviations | | | |
| Bond lengths (Å) | 0.008 | 0.008 | |
| Bond angles (°) | 1.183 | 1.092 | |
| **Validation** | | | |
| EMRinger score | 2.84 | 2.98 | |
| MolProbity score | 2.22 | 2.18 | |
| Clashscore | 13.11 | 13.87 | |
| Rotamer outliers (%) | 0.17 | 0.34 | |
| Cβ deviations (%) | 0 | 0 | |
| Ramachandran plot | | | |
| Favored (%) | 88.46 | 90.57 | |
| Allowed (%) | 11.54 | 9.43 | |
| Disallowed (%) | 0 | 0 | |
| CaBLAM outliers (%) | 3.74 | 3.8 | |

nature research

Corresponding author(s): Ian J. MacRae

Last updated by author(s): Jul 9, 2021

# Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☒ | ☐ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☐ | ☒ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| | |
|---|---|
| Data collection | Cryo-EM data were collected using the automated data collection software Leginon. |
| Data analysis | Cryo-EM data were processed and analyzed using UCSF MotionCor2 (doi: 10.1038/nmeth.4193), CTFFind4 (doi: 10.1016/j.jsb.2015.08.008), RELION 2 (doi: 10.7554/eLife.18722), Cryosparc (doi: 10.1371/journal.pbio.3000122), and the 3DFSC server (www.3dfsc.salk.edu). Atomic models were assessed using PHENIX (doi: 10.1107/S2059798319011471), Molprobity (molprobity.biochem.duke.edu), and PDB validation servers (www.wwpdb.org). Phosphorimages were analyzed using ImageQuant TL (GE Healthcare), and resulting data analyzed using Prism version 8.0 (GraphPad Software, Inc.). |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Maps for the EfPiwi-piRNA and EfPiwi-piRNA-target complexes were deposited in the Electron Microscopy Data Bank under accession IDs EMD-23061 and EMD-23063, respectively. Corresponding atomic models were deposited in the Protein Data Bank under accession IDs 7KX7 and 7KX9. The EfPiwi(MID/PIWI)-piRNA-long-target complex map was deposited in the Electron Microscopy Data Bank under accession ID EMD-23062.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Biochemical experiments were performed using at least two different preparations of each RNA-protein complex. Cryo-EM data were collected on several occasions using several different samples and multiple freezing conditions. Only samples that gave interpretable data were used. |
| Data exclusions | Uninterpretable cryo-EM data were excluded. |
| Replication | All biochemical experiments were replicated at least three times. |
| Randomization | Randomization was not performed. |
| Blinding | Investigators were not blinded. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ Antibodies |
| ☒ | ☐ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☒ | ☐ Animals and other organisms |
| ☒ | ☐ Human research participants |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|---|---|
| ☒ | ☐ ChIP-seq |
| ☒ | ☐ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |