

Structure of Fanzor2 reveals insights into the evolution of the TnpB superfamily

Received: 16 March 2024

Accepted: 27 August 2024

Published online: 1 October 2024

Richard D. Schargel^{1,3}✉, M. Zuhaib Qayyum^{2,3}, Ajay Singh Tanwar^{2,3}, Ravi C. Kalathur² & Elizabeth H. Kellogg²✉

RNA-guided endonucleases, once thought to be exclusive to prokaryotes, have been recently identified in eukaryotes and are called Fanzors. They are classified into two clades, Fanzor1 and Fanzor2. Here we present the cryo-electron microscopy structure of *Acanthamoeba polyphaga* mimivirus Fanzor2, revealing its ω RNA architecture, active site and features involved in transposon-adjacent motif recognition. A comparison to Fanzor1 and TnpB structures highlights divergent evolutionary paths, advancing our understanding of RNA-guided endonucleases.

Among the most widespread genes in all branches of life, transposons are potent agents of genetic change, as they mediate genome rearrangements through a variety of mechanisms. Recently, a group of transposon-encoded accessory proteins, termed obligate mobile element-guided activity (OMEGA) systems, were discovered to possess DNA cleavage activity, guided by a noncoding RNA called ω RNA, and are thought to be ancestors of Cas9 and Cas12 effectors from the class 2 clustered regularly interspaced short palindromic repeats (CRISPR) nuclease family^{1–3}. One class of OMEGA proteins, called TnpB, has evolved into different Cas12 subtypes on multiple occasions, helping to explain the diversity of that family of CRISPR effectors^{4–7}. Indeed, bioinformatics analyses of TnpB homologs indicate a wide range of both structure and function, with diverse protein architectures and catalytic-site geometries^{4,8}. Eukaryotic TnpB homologs, referred to as Fanzors, can be identified in a wide range of organisms, including protists, fungi, arthropods, plants and eukaryotic viruses, and similarly function as RNA-guided nucleases^{6,7,9–11}. Fanzors have attracted considerable interest as genome-editing tools, both for their natural functionality in eukaryotic cells and their substantially smaller size compared to Cas9 and Cas12 proteins. Fanzors are broadly categorized into two distinct clades, Fanzor1 and Fanzor2. Recent structural studies were conducted on Fanzor1 (refs. 9,10) but similar information remained lacking for the more compact Fanzor2. Here, we set out to fill that gap and characterize a representative member of the Fanzor2 clade to understand how these endonucleases recognize target DNA, as well as their relationship to Fanzor1 and the TnpB superfamily.

Results

Structure of ApmFz2– ω RNA–target DNA ternary complex

We reconstituted a ternary complex of *Acanthamoeba polyphaga* mimivirus Fanzor2 (ApmFz2) with the native ω RNA scaffold and target DNA substrate¹¹ and determined its structure. ApmFz2 constitutively associates with its ω RNA (247 nt); therefore, we coexpressed it with an ω RNA scaffold to promote complex stability. The ω RNA construct was designed with a hepatitis delta virus self-cleaving ribozyme at the 3' end to produce a fixed-length 21-nt guide RNA (gRNA)^{12,13}. The resulting ternary complex is biochemically active (Extended Data Fig. 1a) and exhibits cleavage activity consistent with previous results¹¹. The target DNA substrate used in cryo-electron microscopy (cryo-EM) imaging was designed to form an RNA–DNA hybrid and promote ternary complex formation (Fig. 1a).

The resulting 2.99 Å resolution cryo-EM map (Table 1 and Extended Data Fig. 1) enabled nearly complete building of both protein (468 of 520 aa) (Fig. 1b,c) and ω RNA (119 of 247 nt). The nuclear localization signal (NLS; corresponding to residues 1–53) is not observed, consistent with disorder predictions¹¹. ApmFz2 is much more similar to TnpB than to a previously characterized Fanzor1 (ref. 9) (Extended Data Fig. 2), indicating a closer relationship to prokaryotic TnpB, at least at an architectural level, as previously suggested^{9–11}. Like TnpB, ApmFz2 has a recognition (REC) domain, a wedge (WED) domain, a RuvC domain and a zinc finger (ZnF) domain (Fig. 1b). Notably, ApmFz2 has an N-terminal domain (NTD; residues 53–130) that is not observed in any other available TnpB structures (Extended Data Figs. 2 and 3). Part of this NTD (residues 65–91) appears to complete the RuvC fold

¹Department of Microbiology, Cornell University, Ithaca, NY, USA. ²Department of Structural Biology, St. Jude Children's Research Hospital, Memphis, TN, USA. ³These authors contributed equally: Richard D. Schargel, M. Zuhaib Qayyum, Ajay Singh Tanwar. ✉e-mail: rds278@cornell.edu; Elizabeth.Kellogg@stjude.org

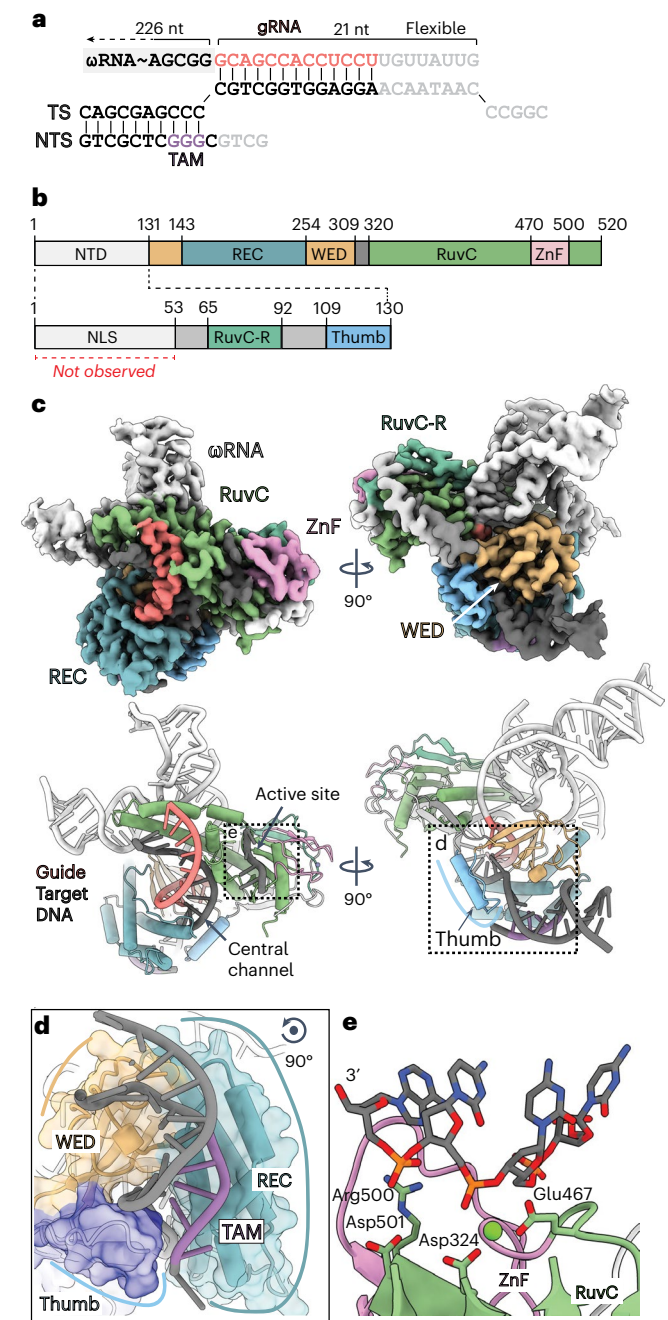


Fig. 1 | Cryo-EM structure of ApmFz2 ternary complex. **a**, Diagram showing target DNA substrate (bottom) annealed to ωRNA (top). Nucleotides not observed in the cryo-EM structure are light gray. The gRNA is pink and the TAM is purple. On the DNA molecule, TS marks the target strand and NTS marks the nontarget strand. **b**, Top: domain organization of ApmFz2, with domain boundaries indicated by residue numbers. NTD, white; REC, aqua; WED, orange-yellow; RuvC, green; ZnF, pink. Bottom: detailed annotation of the NTD. The NLS was not observed in the cryo-EM map. The RuvC-R (blue-green) and thumb (blue) regions structurally reinforce the RuvC and REC domains, respectively. Gray boxes indicate linker regions not specifically assigned to a domain. **c**, Cryo-EM reconstruction (top) and atomic model (bottom) of ApmFz2 ternary complex. Domains and nucleic acid molecules are the same colors as in **a**, **b**, except for ωRNA, which is shown in white. Insets in **d**, **e** are boxed and labeled. **d**, Close-up view of the thumb domain associated with target DNA in conjunction with the WED and REC domains. A transparent surface is overlaid on the atomic model and the thumb domain is colored according to its electrostatic potential, with blue indicating a highly basic surface. **e**, View of the active site bound to the precleaved target DNA substrate. The catalytic triad is shown as green sticks and the magnesium ion is shown as a green sphere.

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

ApmFz2 ternary structure (EMD-44046), (PDB 9BOL)	
Data collection and processing	
Magnification	79,000
Voltage (kV)	200
Electron exposure (e ⁻ per Å ²)	51
Defocus range (μm)	-0.5 to -2.5
Pixel size (Å)	1.044
Symmetry imposed	C1
Initial particle images (no.)	1,801,494
Final particle images (no.)	189,912
Map resolution (Å)	2.99
FSC threshold	0.143
Map resolution range (Å)	2.28–5.83
Refinement	
Initial model used (PDB code)	None
Model resolution (Å)	3.2 (unmasked)
FSC threshold	0.5
Model resolution range (Å)	Not applicable
Map sharpening B factor (Å ²)	-86.4
Model composition	5
Nonhydrogen atoms	7,051
Protein residues	466
Ligands	Mg: 1 Zn: 1
B factors (Å ²)	
Protein	108.32
Nucleotide	158.05
Ligand	148.23
Root-mean-square deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.582
Validation	
MolProbity score	1.66
Clashscore	7
Poor rotamers (%)	0.23
Ramachandran plot	
Favored (%)	96.34
Allowed (%)	3.66
Disallowed (%)	0.00

PDB, Protein Data Bank; FSC, Fourier shell correlation.

and interacts with ωRNA (described in detail below); thus, we annotated it as the RuvC-R domain (Fig. 1b,c). Another region of the NTD (residues 109–130) appears to clasp onto the target DNA substrate in coordination with the REC and WED domains (Fig. 1c,d); thus, we call it the ‘thumb’ (Fig. 1b). The thumb is highly basic and seems to sterically invade the target DNA duplex adjacent to the transposon-adjacent motif (TAM) (Fig. 1d). It is possible that the thumb motif may serve to stabilize or further unwind the target DNA and guide the target strand into the central channel.

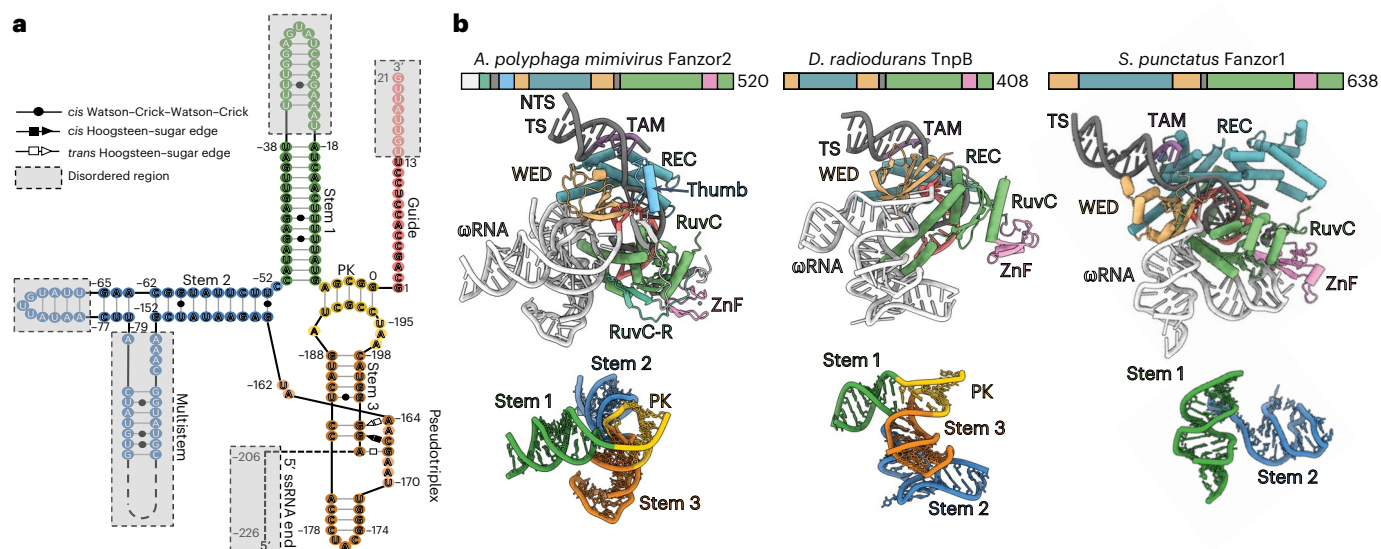


Fig. 2 | ωRNA architecture and structural comparison of existing TnpB structures. **a**, A 2D schematic of ωRNA, showing Watson-Crick base pairing (solid lines) and noncanonical interactions. The ωRNA scaffold and guide region span nucleotides −205 to 0 and 1 to 21, respectively. PK, pseudoknot. Disordered regions are in a dashed gray box. **b**, Comparison of protein and ωRNA structural features across known TnpB ternary complex structures: ApmFz2 (this study),

Deinococcus radiodurans TnpB (PDB 8H1J) and *Sphenodon punctatus* Fanzor1 (PDB 8GKH). Top: protein domain diagram, with domains colored as in Fig. 1a. Middle: RNP complex structure. The ωRNA is colored white. Bottom: the ωRNA structure, with structural features colored as in **a**, shown in the same view as in the ternary structure (middle row).

TAM recognition

We were particularly interested in understanding the basis of TAM recognition, because ApmFz2 has a distinct TAM motif (5'-GGG-3')¹¹ compared to characterized Fanzor1 and TnpB TAM motifs, which tend to be more AT-rich (5'-CATA-3' and 5'-TTGAT-3', respectively)^{2,9}. Consistent with expectations, we observe Arg194 and His215 in the REC domain forming base-specific interactions with the nontarget strand, with dG(0) and dG(−2) (Extended Data Fig. 4a–c). On the target strand, Glu260 in the WED domain forms base-specific interactions with dC(1) (Extended Data Fig. 4d). In contrast to existing Fanzor1 and TnpB structures, we observe only minimal interactions upstream of the TAM (Extended Data Fig. 4b), consistent with the shorter TAM sequence motif in Fanzor2 (3 nt, compared to 4–5 nt in Fanzor1 (ref. 9) and TnpB (ref. 2)).

ApmFz2 ternary complex active-site architecture

Fanzor2, like TnpB, cleaves single-stranded DNA (ssDNA) downstream of the TAM and intriguingly does not exhibit collateral (that is, nonspecific) DNA cleavage activity¹¹. Previous TnpB ternary complexes did not have sufficient density to build a complete RuvC and ZnF active site (Extended Data Fig. 5), which was attributed to the flexibility of the cleavage domain^{14,15}. We observe the RNA–DNA heteroduplex occupying the central channel, formed by the WED, REC and RuvC domains^{9,14,15} (Fig. 1c and Extended Data Fig. 4e–g). In contrast to TnpB, we observe the entire density of the catalytic site within the RuvC and ZnF domains (Fig. 1c), along with four nucleotides of ssDNA occupying the active site (Fig. 1e). The resolution was insufficient for unambiguous assignment of all four bases but it was sufficient for distinguishing a purine at the 3' end (Extended Data Fig. 6a). Therefore, we built a 4-nt 3'-ACCC-5' ssDNA model. In addition, the ssDNA occupying the active site is not cleaved at the expected location according to the active-site geometry (Extended Data Fig. 6a). Within the active site, a catalytic triad is formed by Asp324, Glu467 and Asp501 (Fig. 1e). Notably, Glu467 is shifted 50 residues toward the C terminus relative to its canonical position in the TnpB amino acid sequence^{14,15} (Extended Data Figs. 2 and 7). This rearranged catalytic site is unique to Fanzor and its ancestral TnpB (commonly referred to as TnpB2 or pro-Fanzor), suggesting that they

represent a distinct evolutionary branch, separate from the Cas12 family of CRISPR endonucleases^{10,11}. The catalytic triad coordinates a single Mg²⁺ ion (Fig. 1e and Extended Data Fig. 6a). This contrasts with the postcleavage SpuFz1 structure and cleavage-inhibited Cas12 structures, which both show the catalytic triad coordinating two Mg²⁺ ions^{9,16} (Extended Data Fig. 5). Together with the evidence showing that our purified ApmFz2 purification is biochemically active (Extended Data Fig. 1a), this observation suggests that the ApmFz2 ternary complex captured by cryo-EM is in an inhibited state. This might be because of the differing conditions used to assemble the ternary complex for cryo-EM imaging, with an excess of DNA oligonucleotides (which could cause nonspecific interactions) and 2 mM Mg²⁺ (lower than 10 mM Mg²⁺ used for our biochemical assay). The latter observation suggests that binding of the second Mg²⁺ may be rate limiting.

ωRNA architecture and recognition

Approximately half of the ωRNA scaffold is not visible because of the flexibility of its long stem loops (Fig. 2a). For the remainder of the ωRNA, the high quality of the cryo-EM map allowed us to distinguish purines and pyrimidines (Extended Data Fig. 6b,c) and to assign the nucleotide sequence with confidence. The last 156 nt of the ωRNA scaffold (−160 to −5) are base-paired, as correctly predicted by RNA secondary-structure prediction methods¹¹. However, the first 45 nt (−205 to −161) form long-range interactions, including a pseudotriplex and pseudoknot, with nucleotides at positions −4 to 0 (Fig. 2a). It is worth noting that the RNA pseudoknot is a feature shared between TnpB and ApmFz2 (Fig. 2b); it forms the core of the ωRNA scaffold and is located next to the WED domain, both key characteristics in TnpB and Cas12 families^{14–20}. The WED domain, in tandem with the RuvC and RuvC-R domains, forms a groove that recognizes the pseudoknot (Extended Data Fig. 8a–c). Remarkably, the RuvC-R subdomain along with RuvC appears to make base-specific and backbone interactions, mediated by Asp397 and Lys401 to G(−192):C(−2) and Asp397, Lys401, Ser83 and Asn82 to C(−191):G(−3) (Extended Data Fig. 8c). The remainder of the ωRNA scaffold is recognized solely by the RuvC domain, which appears to make backbone-specific hydrogen-bonding interactions with the stem 3 and stem 2 portions of the ωRNA (Extended

Data Fig. 8d,e). In contrast to the TnpB and Fanzor1 structures, we do not observe any protein interactions with stem I, which is recognized by the RuvC and WED domains in Fanzor1 (ref. 9) and TnpB (ref. 15).

Our findings show that some essential structural features within the ω RNA scaffold, such as the pseudoknot, may be obscured by secondary-structure predictions. This highlights the importance of experimental structure determination to reveal functional features that are conserved across large evolutionary timescales in ribonucleoprotein (RNP) complexes. Fanzor1 lacks the aforementioned pseudoknot RNA motif (Fig. 2b), establishing that the pseudoknot is not conserved across Fanzors⁹. We modeled the long stem loops that lack density (Fig. 2a), predicting that they project away from the core of the structure (Extended Data Fig. 9), which could indicate that they may not all be essential for activity. Indeed, previous studies showed that truncation of the RNA long stem loops of TnpB and Fanzor1 did not affect activity^{9,14,15}. Furthermore, ω RNAs across TnpB homologs typically contain predicted long stem loops, suggesting that the ω RNA scaffold has a wide range of structural variability, consistent with the idea that not all ω RNA structural features are essential. Lastly, our Fanzor2 structure reveals a distinct evolutionary trajectory in which every RNA domain has increased in size, in contrast with the available Fanzor1 structure⁹, which shows the RNA domains have become truncated or have disappeared entirely (Fig. 2 and Extended Data Fig. 10).

Discussion

Collectively, our structure shows how the compact eukaryotic Fanzor2 carries out ω RNA recognition, TAM recognition and target DNA loading. We also reveal the architecture of its ω RNA, highlighting key features that distinguish Fanzor2 and Fanzor1. The unique structured N-terminal extension of Fanzor2 has subdomains that reinforce the core of the protein and interacts with the DNA duplex and RNA pseudoknot. The rearranged catalytic site highlights the plasticity of TnpB family effectors. Our findings provide a framework for future protein engineering directions and advance our understanding of the evolution from prokaryotic TnpB proteins to eukaryotic Fanzor proteins.

Online content

Any methods, additional references, Nature Portfolio 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/s41594-024-01394-4>.

References

- Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
- Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
- Kapitonov, V. V., Makarova, K. S. & Koonin, E. V. ISC, a novel group of bacterial and archaeal DNA transposons that encode Cas9 homologs. *J. Bacteriol.* **198**, 797–807 (2015).
- Altae-Tran, H. et al. Diversity, evolution, and classification of the RNA-guided nucleases TnpB and Cas12. *Proc. Natl Acad. Sci. USA* **120**, e2308224120 (2023).
- Siguier, P., Gourbeyre, E. & Chandler, M. Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
- Shmakov, S. et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
- Bao, W. & Jurka, J. Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mob. DNA* **4**, 12 (2013).
- Wiegand, T. et al. TnpB homologues exapted from transposons are RNA-guided transcription factors. *Nature* **631**, 439–448 (2024).
- Saito, M. et al. Fanzor is a eukaryotic programmable RNA-guided endonuclease. *Nature* **620**, 660–668 (2023).
- Yoon, P. H. et al. Eukaryotic RNA-guided endonucleases evolved from a unique clade of bacterial enzymes. *Nucleic Acids Res.* **51**, 12414–12427 (2023).
- Jiang, K. et al. Programmable RNA-guided DNA endonucleases are widespread in eukaryotes and their viruses. *Sci. Adv.* **9**, eadk0171 (2023).
- Berkhout, B., Gao, Z. & Herrera-Carrillo, E. Design and evaluation of guide RNA transcripts with a 3'-terminal HDV ribozyme to enhance CRISPR-based gene inactivation. *Methods Mol. Biol.* **2167**, 205–224 (2021).
- Gao, Z., Herrera-Carrillo, E. & Berkhout, B. Improvement of the CRISPR–Cpf1 system with ribozyme-processed crRNA. *RNA Biol.* **15**, 1458–1467 (2018).
- Sasnauskas, G. et al. TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* **616**, 384–389 (2023).
- Nakagawa, R. et al. Cryo-EM structure of the transposon-associated TnpB enzyme. *Nature* **616**, 390–397 (2023).
- Pausch, P. et al. DNA interference states of the hypercompact CRISPR–CasΦ effector. *Nat. Struct. Mol. Biol.* **28**, 652–661 (2021).
- Dong, D. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).
- Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).
- Kurihara, N. et al. Structure of the type V-C CRISPR–Cas effector enzyme. *Mol. Cell* **82**, 1865–1877 (2022).
- Yang, H., Gao, P., Rajashankar, K. R. & Patel, D. J. PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR–Cas endonuclease. *Cell* **167**, 1814–1828 (2016).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Methods

Protein production and purification

ApmFz2 was overexpressed in *Escherichia coli* BL21 star (DE3) cells. The cells were cotransformed with pCDF-ApmFz2- ω RNA and pET15b-ApmFz2 expression plasmids (Supplementary Table 1). A single colony was used to grow a starter culture overnight. Then, 10 ml of starter culture was used to inoculate 1 L of 2xYT medium containing 100 $\mu\text{g ml}^{-1}$ ampicillin and 50 $\mu\text{g ml}^{-1}$ spectinomycin at 37 °C with shaking until the cell density reached an optical density at 600 nm of ~0.7. The overexpression of protein was induced by 0.6 mM IPTG and grown at 18 °C for 18 h. The cells were harvested by centrifugation at 6,240g for 15 min (4 °C) and the cell pellet was frozen at –80 °C until needed. The cell pellet was resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol, EDTA-free cOmplete protease inhibitor cocktail (Roche) and 1 mM DTT). Cells were disrupted using a cell disruptor (Constant Systems) at 20,000 psi. The lysate was cleared by centrifugation at 48,380g for 30 min at 4 °C. The lysate was applied to a gravity column after incubation with pre-equilibrated Strep-Tactin Sepharose resin (IBA Life Sciences) for 30 min at 4 °C. The column was washed with 15 column volumes of 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol and 1 mM DTT. The protein was eluted with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5% glycerol, 1 mM DTT and 5 mM d-desthiobiotin. The eluted fractions were verified by running 4–20% SDS–PAGE gels. The fractions were pooled together and diluted with 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 5% glycerol and 1 mM DTT buffer until the NaCl concentration reached 200 mM. Next, the resulting sample was loaded onto a 5-ml HiTrap Heparin column (Cytiva) pre-equilibrated with buffer A (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% glycerol and 1 mM DTT). The RNA-bound samples were eluted with 40–60% buffer B (20 mM Tris-HCl pH 8.0, 1 M NaCl, 5% glycerol and 1 mM DTT). The eluted fractions were verified by SDS–PAGE and the desired protein fractions were pooled together and concentrated using a 50 kDa-cutoff membrane filter unit (Millipore). The concentrated sample was injected onto a Superose 6 Increase 10/300 GL column (Cytiva) equilibrated with buffer C (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 2 mM MgCl_2 , 10 $\mu\text{M ZnCl}_2$, 5% glycerol and 1 mM DTT). The purified protein was concentrated to 2.4 mg ml^{-1} .

In vitro cleavage assay

The DNA substrate was produced using PCR from gBlocks (Integrated DNA Technologies) as templates (Supplementary Table 1). The amplified DNA substrate was purified using a QIAquick PCR purification kit (Qiagen).

The cleavage reaction (25 μl) was set up by mixing 1.6 μM purified DNA substrate with 5 μM freshly purified Fz2 in reaction buffer (100 mM NaCl, 50 mM Tris-HCl pH 7.9, 10 mM MgCl_2 and 1 mM DTT) at 25 °C and incubated at 37 °C for 2 h. The tube was then incubated on ice for 10 min to quench the reaction. It was then heated at 95 °C for 15 min followed by cooling to 50 °C after the addition of 10 μg of RNase A (Qiagen) for 10 min. Next, 50 μl of buffer-saturated phenol was added and the tube was vortexed and spun immediately. Then, 20 μl of the aqueous phase was aspirated out and mixed with 3 μl of 6 \times New England Biolabs DNA loading dye. The sample was then run on a 5% Mini-PROTEAN TBE gel (Bio-Rad) as per the protocol recommended by the manufacturer. Gels were stained with 1 \times SYBR gold (Thermo Fisher Scientific) and imaged on a ChemiDoc gel imager (Bio-Rad). Each in vitro cleavage assay was performed in triplicate.

Target DNA preparation

The reaction mixture contained equimolar oligo concentrations of Fz2_CEM_sub_top and Fz2_CEM_sub_bot (Supplementary Table 1) in nuclease-free water. The mixture was annealed in a thermocycler (95 °C to 10 °C, $\Delta 1.5$ °C ramp per min).

Sample preparation for cryo-EM

The ApmFz2–RNA–DNA complex was reconstituted in vitro by mixing 20 μM ApmFz2–RNA with 30 μM assembled target DNA in buffer C

(20 mM Tris-HCl pH 8.0, 200 mM NaCl, 2 mM MgCl_2 , 10 $\mu\text{M ZnCl}_2$, 5% glycerol and 1 mM DTT) at 37 °C for 10 min.

Cryo-EM grid preparation

UltrAuFoil grids (R1.2/1.3 300 mesh, MiteGen) were cleaned for 40 s in a Solarus II plasma cleaner (Gatan) before the application of 3.5 μl of the sample (~ 1.5 mg ml^{-1}) and plunge-freezing in liquid ethane using a Vitrobot mark IV (FEI) with 95% chamber humidity at 10 °C.

Cryo-EM data acquisition and processing

ApmFz2–RNA–DNA complex. Data were collected on a Talos Arctica 200 keV microscope (Thermo Fisher Scientific) equipped with a K3 direct electron detector and a BioQuantum energy filter. Sample grids were imaged at 200 kV, with an intended defocus range of –2.25 to –0.5 μm and a magnification of $\times 79,000$ in electron counting mode (1.044 Å per pixel). Videos were collected with a total dose of 51 $\text{e}^- \text{per Å}^2$. A total of 3,205 videos were recorded with EPU software (Thermo Fisher Scientific). Downstream processing was performed in cryoSPARC 4.4.0 (ref. 21). Movies were motion-corrected and summed using Patch motion correction in cryoSPARC. The contrast transfer function (CTF) was estimated using Patch CTF in cryoSPARC. Initially, 1,000 micrographs were processed, particles were picked using blob picker and extracted and two-dimensional (2D) classes were generated. The good classes were used as templates to perform template-based particle picking and 2D classification. Subsequently, good 2D classes were used for training the Topaz model for the entire dataset²². Particles were extracted with a 360-pixel box and subjected to multiple rounds of 2D classification to remove junk particles followed by multiple rounds of heterogeneous refinements. The resulting best class was used for nonuniform refinement in cryoSPARC. CTF parameters were refined on a per-micrograph and per-particle basis using cryoSPARC global CTF refinement and local CTF refinement, respectively. Particles were then subjected to local motion correction and then to homogenous refinement followed by nonuniform refinement²³. The resolution was estimated using the gold-standard method. Local resolution was estimated using cryoSPARC.

Model building, refinement and analysis. The ApmFz2– ω RNA–DNA ternary structure was generated using an initial model of the protein predicted from AlphaFold2 and an initial model of the ω RNA predicted from RNAcomposer^{24,25}. The resulting models were first docked into the cryo-EM density and manually rebuilt using Coot. Certain parts of the model were manually remodeled or rebuilt using Coot version 0.9.8.2 (ref. 26). For the DNA substrate, real-space refinement was carried out in PHENIX 1.21, with both base-pair and secondary-structure restraints being enforced²⁷. The final structure was validated using PHENIX²⁷ and MolProbity²⁸. Structural representations for figures were created using UCSF ChimeraX²⁹ and Adobe Illustrator (<https://adobe.com/products/illustrator>).

Reporting summary

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

Data availability

The cryo-EM map and model for the ApmFz2– ω RNA–DNA ternary structure were deposited to the EM Data Bank and PDB, respectively, under accession codes EMD-44046 and 9BOL. Other atomic coordinates used in this study for comparison purposes are available from the PDB under the following accession codes: 8GKH for SpuFz1– ω RNA–DNA, 8HIJ for ISDra2TnpB– ω RNA–RNA, 8EXA for ISDra2TnpB–reRNA–RNA and 7LYT for Cas12j–crRNA–(phosphorothioate-DNA). Source data are provided with this paper.

References

21. 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).
22. Bepler, T. et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat. Methods* **16**, 1153–1160 (2019).
23. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
24. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
25. Antczak, M. et al. New functionality of RNAComposer: an application to shape the axis of miR160 precursor structure. *Acta Biochim. Pol.* **63**, 737–744 (2016).
26. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
27. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
28. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
29. Goddard, T. D. et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

Acknowledgements

This work was performed at the St. Jude Cryo-EM Center, which is generously supported by the St. Jude Children's Research Hospital. We thank I. Chen for advice and editorial feedback on the manuscript and J. Peters for his supporting role (to R.D.S.) throughout the project. This research is supported by the National Institutes of Health (R01GM144566 to E.H.K.), Pew Biomedical Foundation (to E.H.K.)

and the National Science Foundation Graduate Research Fellowship Program (DGE - 2139899 to R.D.S.).

Author contributions

R.D.S. and E.H.K. conceptualized the project and designed the experiments. A.S.T. purified protein. M.Z.Q. prepared and screened cryo-EM samples and then collected and refined images to produce the final cryo-EM map. R.D.S. built and interpreted the cryo-EM map. M.Z.Q. and A.S.T. performed the in vitro cleavage assays. R.C.K. and E.H.K. supervised the project. R.D.S. and E.H.K. wrote the initial draft and all authors contributed to the writing and revision of the paper and figures.

Competing interests

The authors declare no competing interests.

Additional information

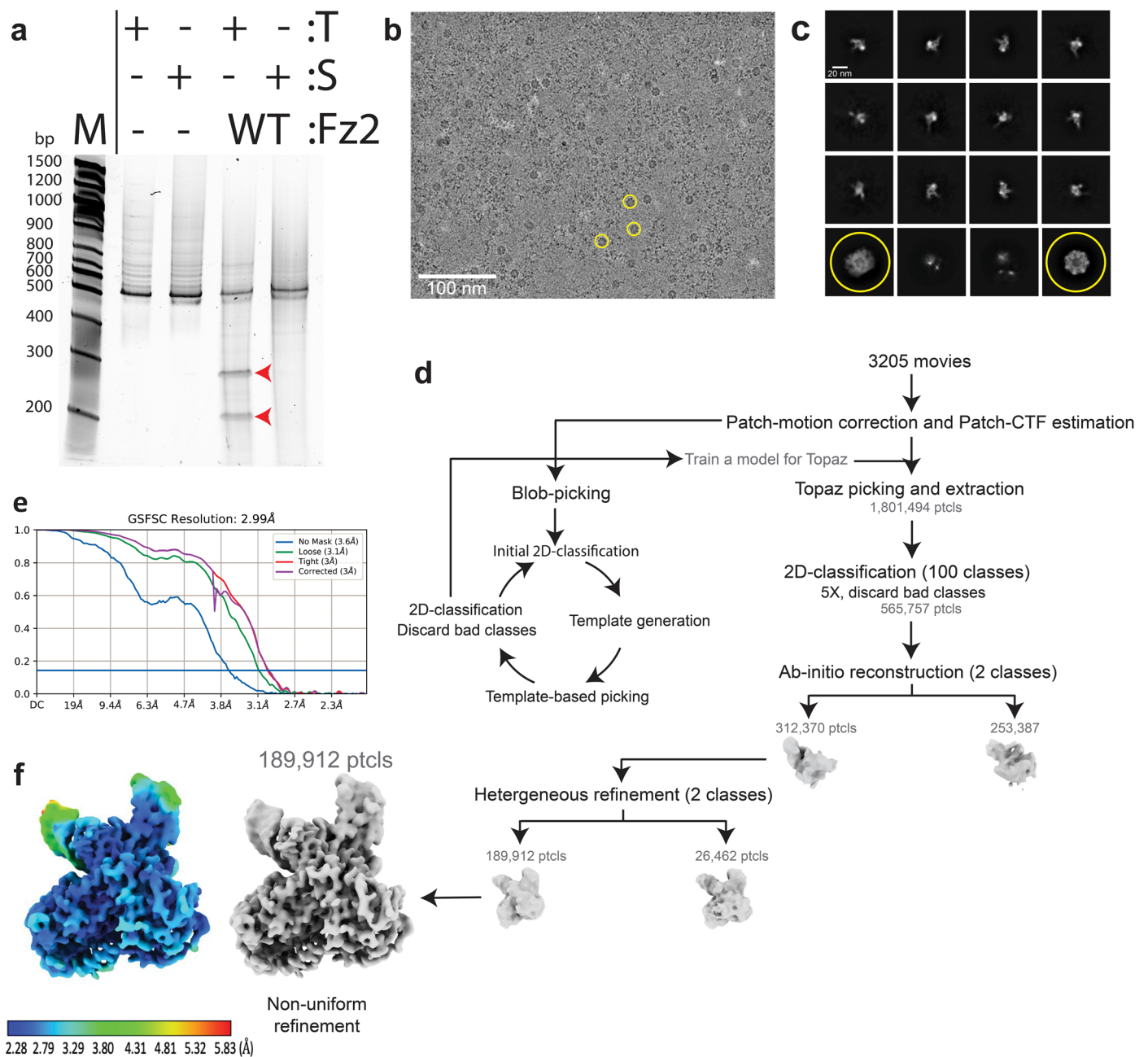
Extended data is available for this paper at <https://doi.org/10.1038/s41594-024-01394-4>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-024-01394-4>.

Correspondence and requests for materials should be addressed to Richard D. Schargel or Elizabeth H. Kellogg.

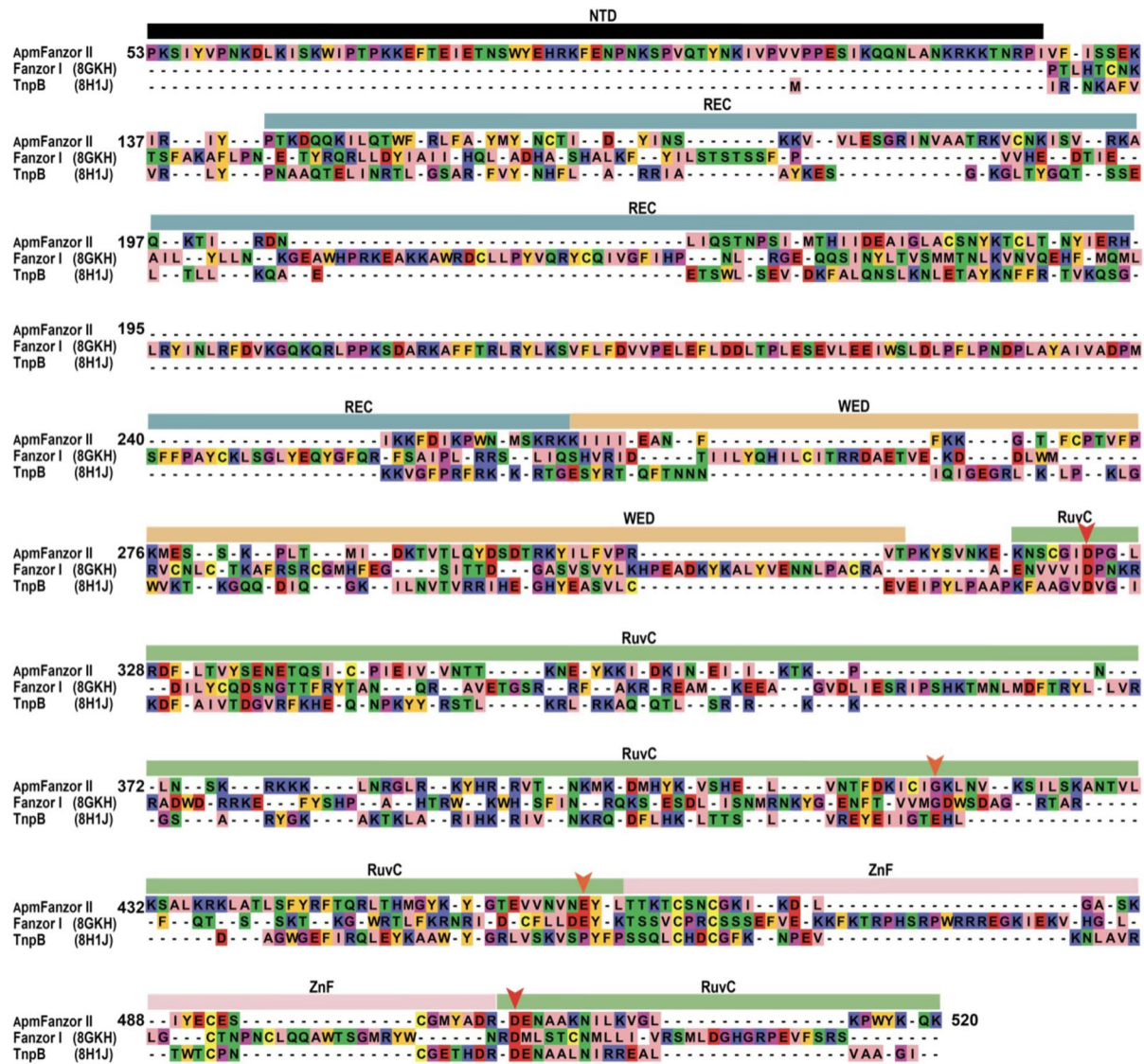
Peer review information *Nature Structural & Molecular Biology* thanks Hiroshi Nishimasu and David Taylor for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Dimitris Typas, in collaboration with the *Nature Structural & Molecular Biology* team.

Reprints and permissions information is available at www.nature.com/reprints.



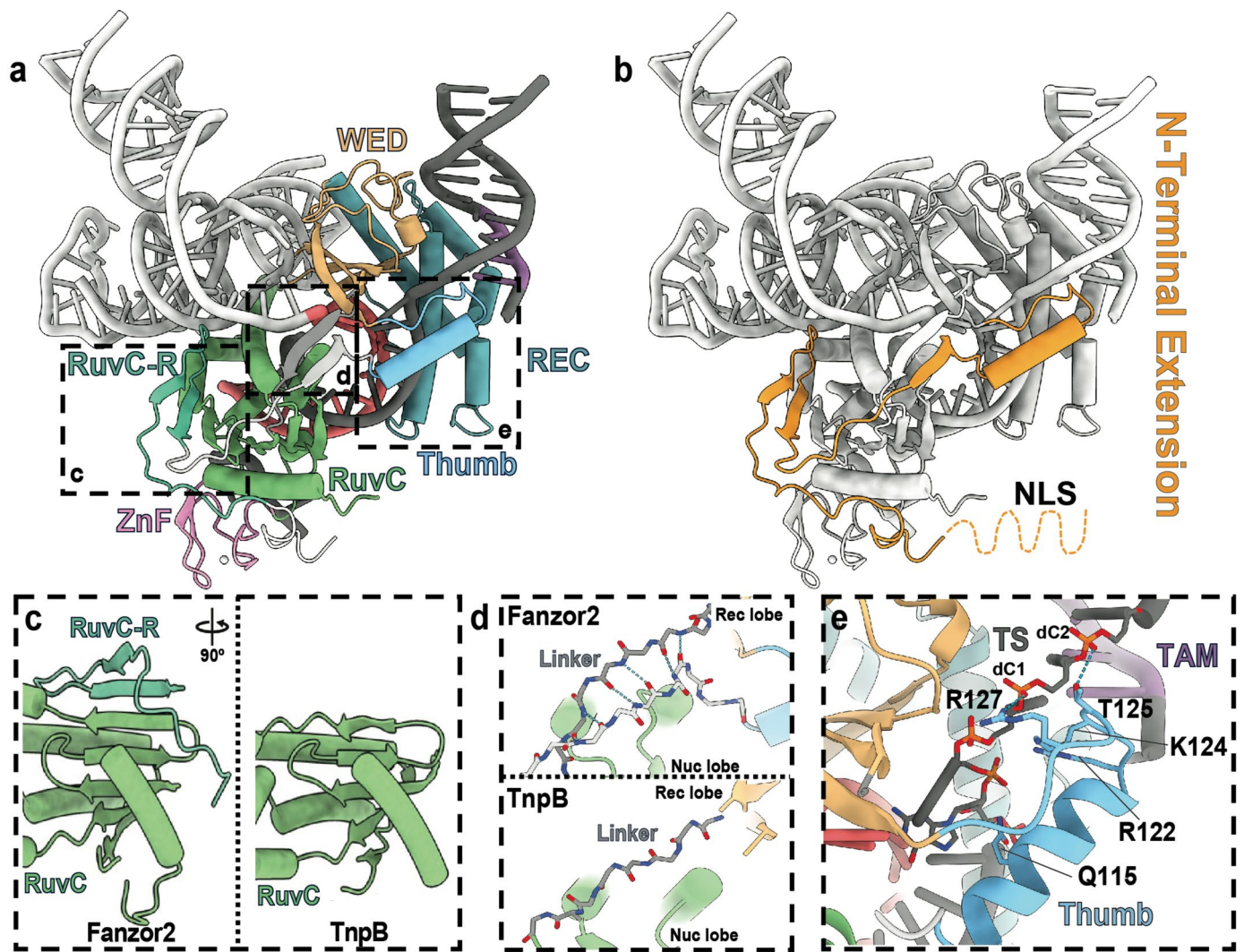
Extended Data Fig. 1 | Validation of biochemical activity, cryo-EM imaging, and image processing pipeline of the ApmFz2 ternary complex.
a. Biochemical validation of Fanzor2 (Fz2) cleavage activity tested on a linear DNA substrate containing a 5'-GGG TAM with either a correct target spacer (T) or a scrambled spacer (S) ($n = 3$). Red arrowheads indicate cleavage product.
b. Representative cryo-EM micrograph from reconstituted ApmFz2 ternary complex ($n = 3,205$). Yellow circles indicate contaminants. Scale bar, 100 nm.

c. 2D classification in cryoSPARC 4.4.0 on template-picked particles. Scale bar, 20 nm **d.** Image processing workflow used to analyze the cryo-EM data. **e.** Fourier shell correlation (FSC) curve. Resolution is reported at the 0.143 cut-off (solid blue line). **f.** Final ApmFz2 ternary complex refined reconstruction (right) and local resolution filtered reconstruction (left). Colors indicate the local resolution range, from 2.3 Å to 5.8 Å.



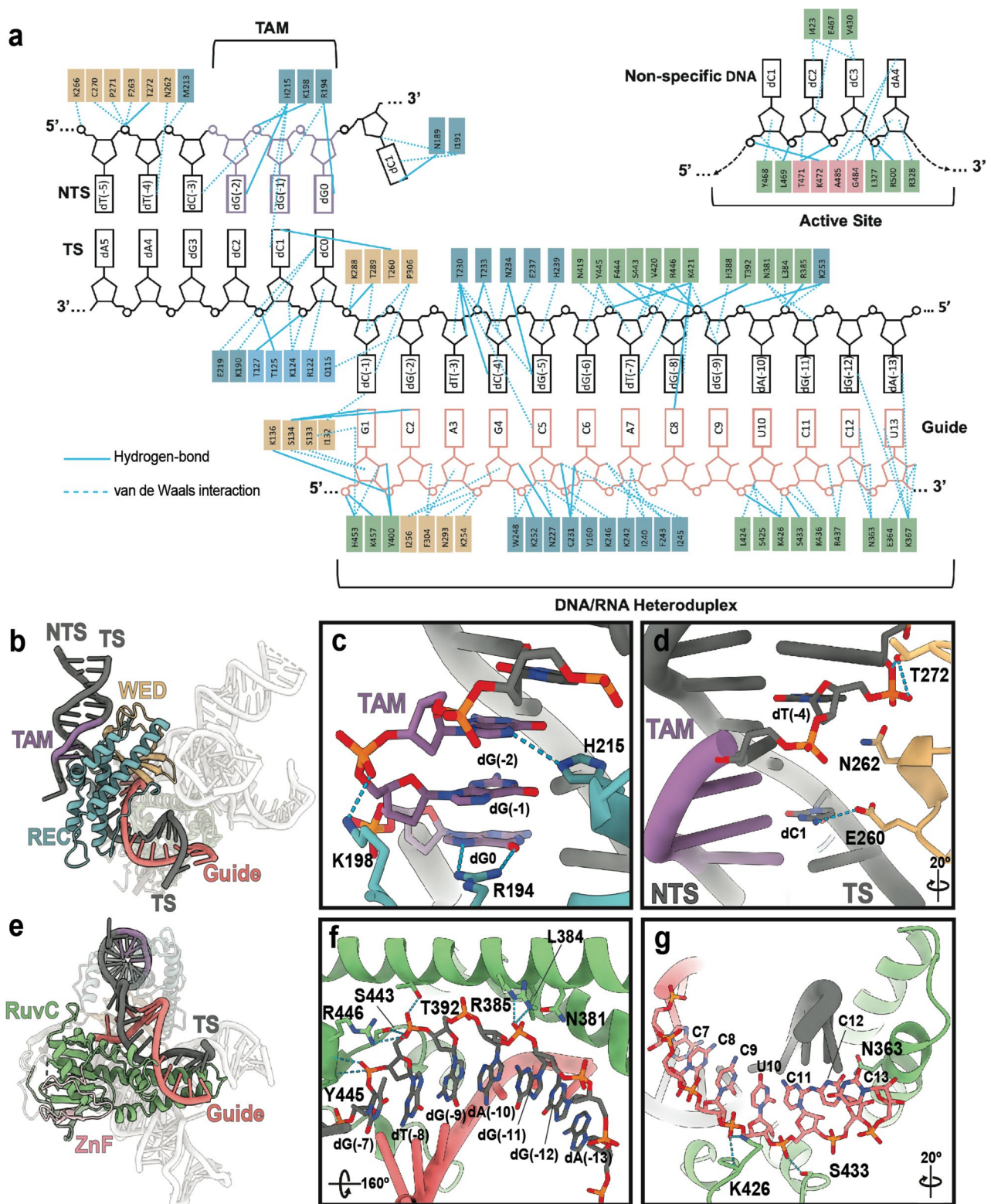
Extended Data Fig. 2 | Structure-based sequence alignment of ApmFanzor2 and its homologs. Sequence alignment includes comparison to Fanzor I (PDB 8GKH) and TnpB (PDB 8H1J). Domains are annotated according to ApmFz2 and labeled according to the convention used in Fig. 1. Red arrowheads indicate

conserved catalytic triad positions, orange arrowheads indicate catalytic glutamate for TnpB and for Fanzor2, as this position is not conserved among homologs.



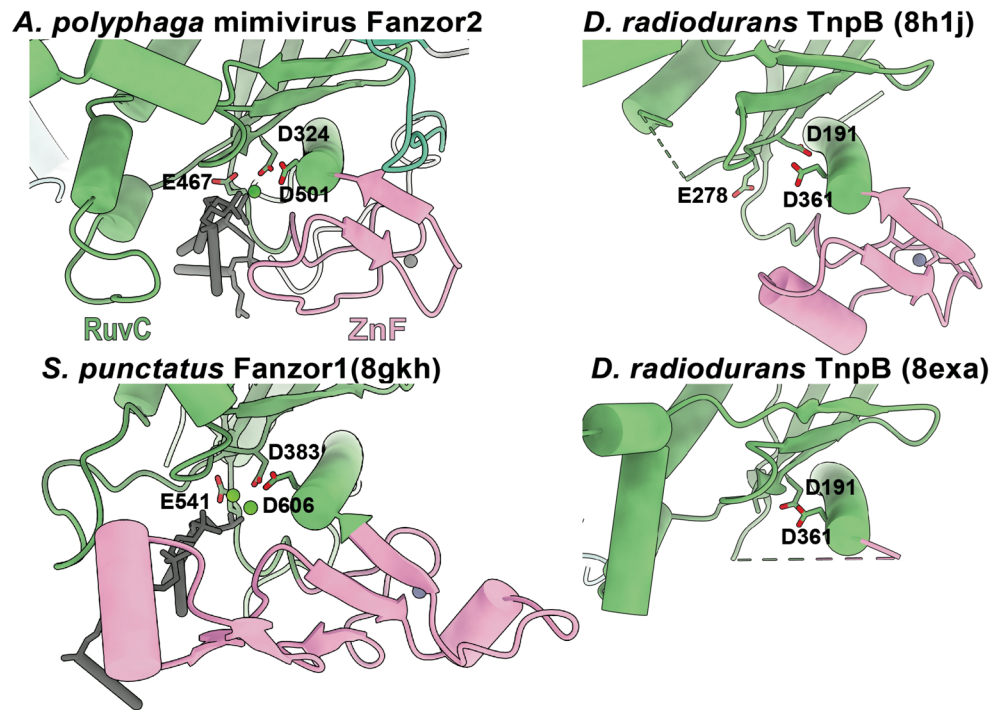
Extended Data Fig. 3 | N-Terminal extension in ApmFz2 forms structural contacts with conserved and important functional domains. a–b. Unique N-terminal extension in ApmFz2 contacting protein functional domains. ApmFz2 domains colored individually (**a**) or with only N-terminal extension colored in orange (**b**). Unstructured nuclear-localization signal (NLS) is shown as an orange dotted line. **c.** Comparison of ApmFz2 and TnpB RuvC domain. RuvC-R

domain binds to the canonical RuvC domain. Rotation relative to **a** is indicated. **d.** Comparison of ApmFz2 and TnpB linker domain that connects the Rec and Nuc lobes. The N-terminal extension in ApmFz2 forms a short β -sheet with the linker domain. **e.** Thumb domain forms hydrogen bonds with target strand within and downstream of TAM duplex. Electrostatic interactions are represented as dashed cyan lines.



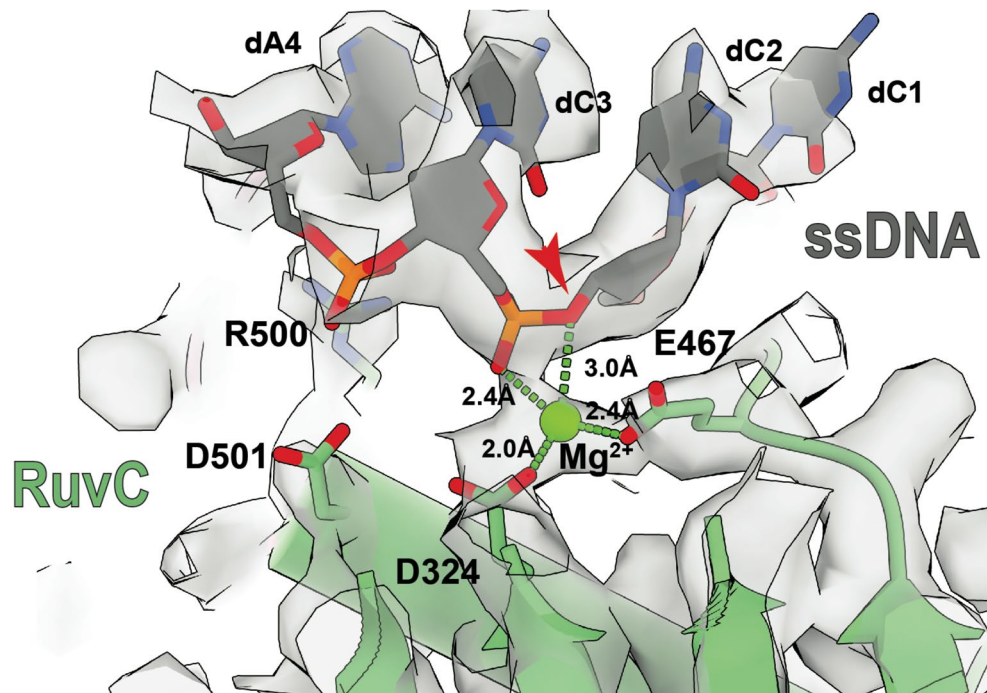
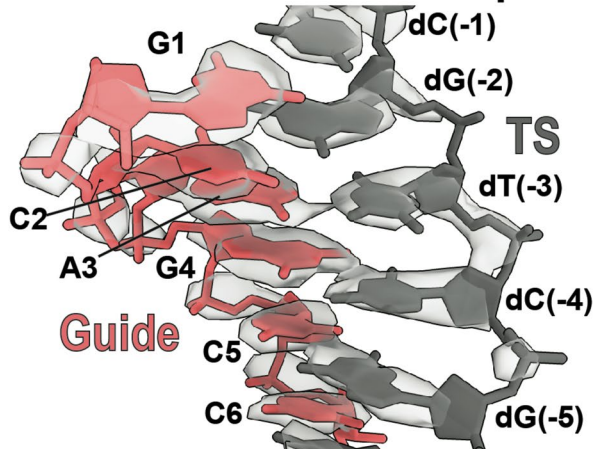
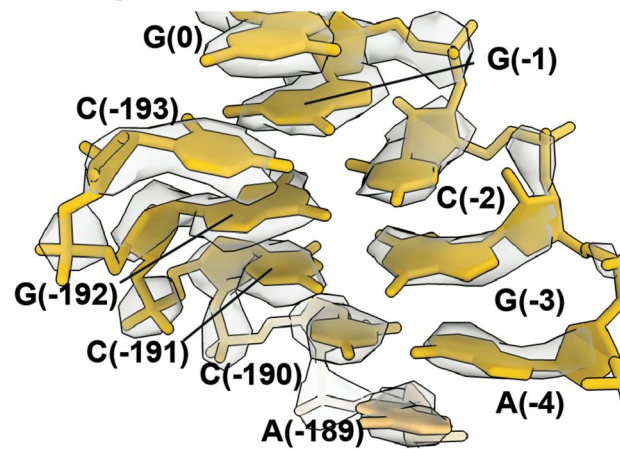
Extended Data Fig. 4 | Structural analysis of TAM and heteroduplex substrate recognition. **a.** Schematic representation of DNA substrate recognition by AmpFz2 including TAM duplex, non-target strand, target strand, and a four-nucleotide single-stranded DNA within the active site. **b.** Atomic model of AmpFz2 recognition of TAM duplex and upstream DNA substrate. TAM duplex is bound in the groove formed by WED and REC. **c.** Recognition of 5'-GGG-3' TAM

by the REC domain. **d.** Recognition of TAM duplex and upstream nucleotides by WED domain. In **c,d**, rotation relative to **b** is indicated in the bottom right corner. **e.** Atomic model of DNA-RNA heteroduplex recognition by RuvC. **f.** Recognition of heteroduplex target-strand by RuvC. **g.** Recognition of heteroduplex guide-RNA by RuvC. In **f-g**, rotation relative to **e** is indicated in the bottom right corner. In **b-g**, hydrogen bond interactions are represented by cyan dashed lines.



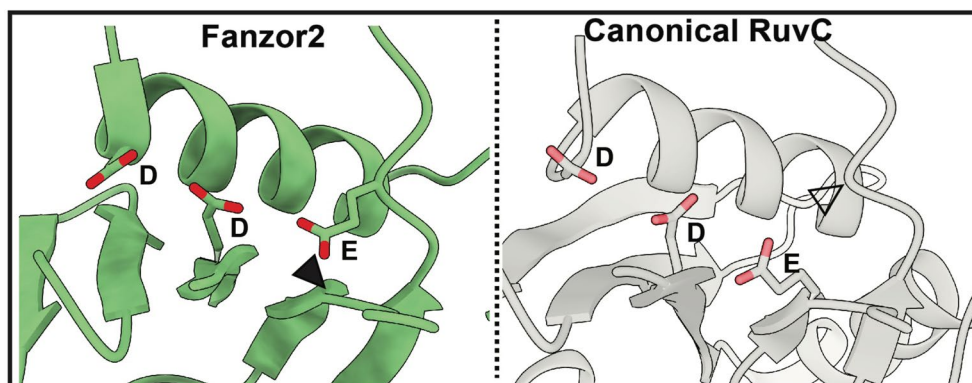
Extended Data Fig. 5 | Comparison between active sites of Fanzor2, Fanzor1 and TnpB structures. Structural comparison of active sites of ApmFz2 (top left), SpuFz1 (bottom left) and TnpB structures (right top and bottom). PDB IDs are

indicated. Dashed lines represent missing sections in structure. DNA is shown as gray sticks, catalytic residues in stick representation, Mg^{2+} as green spheres. The zinc finger (ZnF) domain is pink, with the zinc ion shown as gray sphere.

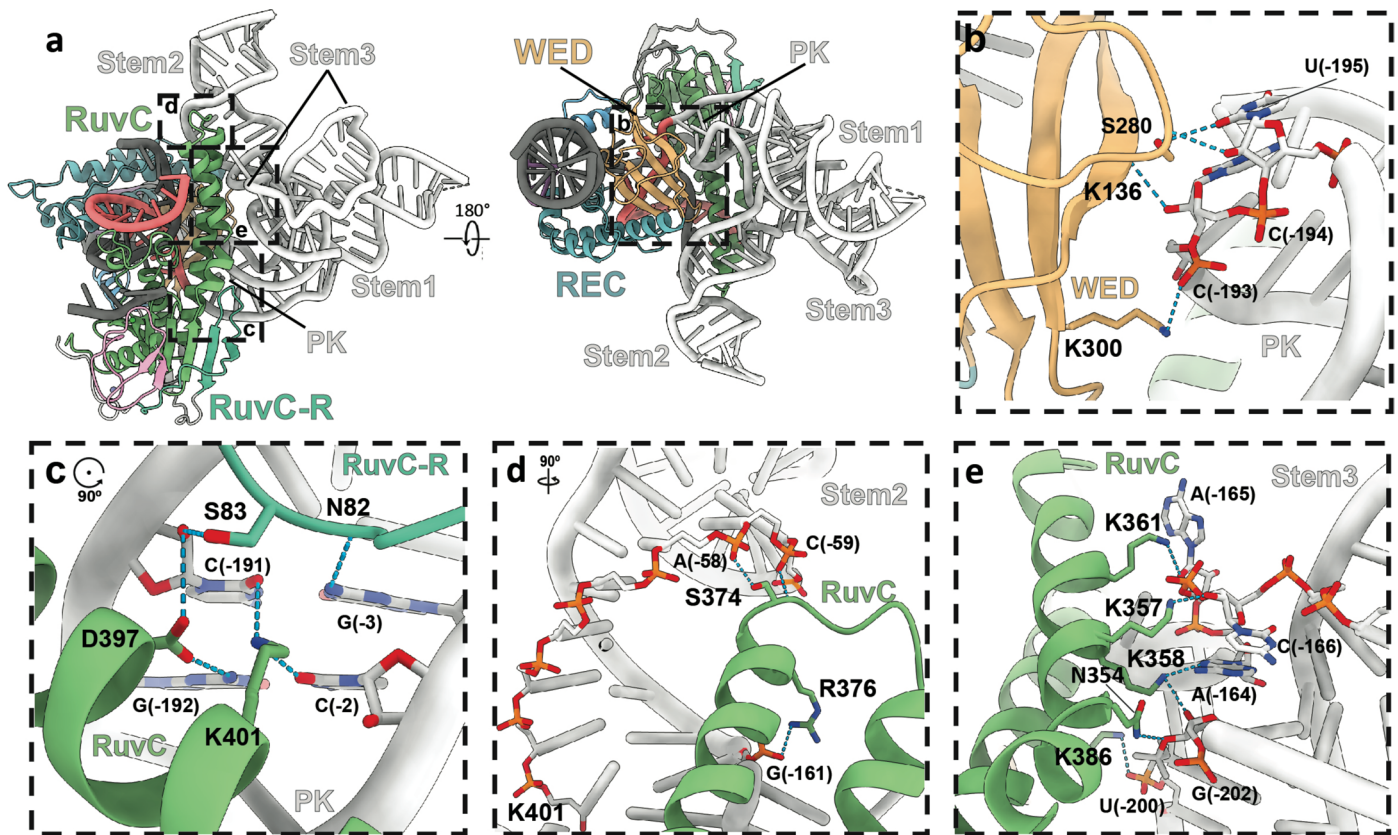
a *A. polyphaga* mimivirus Fanzor2 Active Site**b** DNA-RNA Heteroduplex**c** Pseudoknot

Extended Data Fig. 6 | Quality of the Cryo-EM map is sufficient to distinguish purines and pyrimidines. a. Cryo-EM density (transparent surface) is shown for select regions of the map. The atomic model is docked and colored as defined in Fig. 1. Catalytic residues are shown in stick representation, with distances to the

Mg²⁺ within active site indicated (green dashed lines). Red arrowhead indicates expected cleavage location. Cryo-EM density for DNA-RNA heteroduplex (b) and pseudoknot (c).

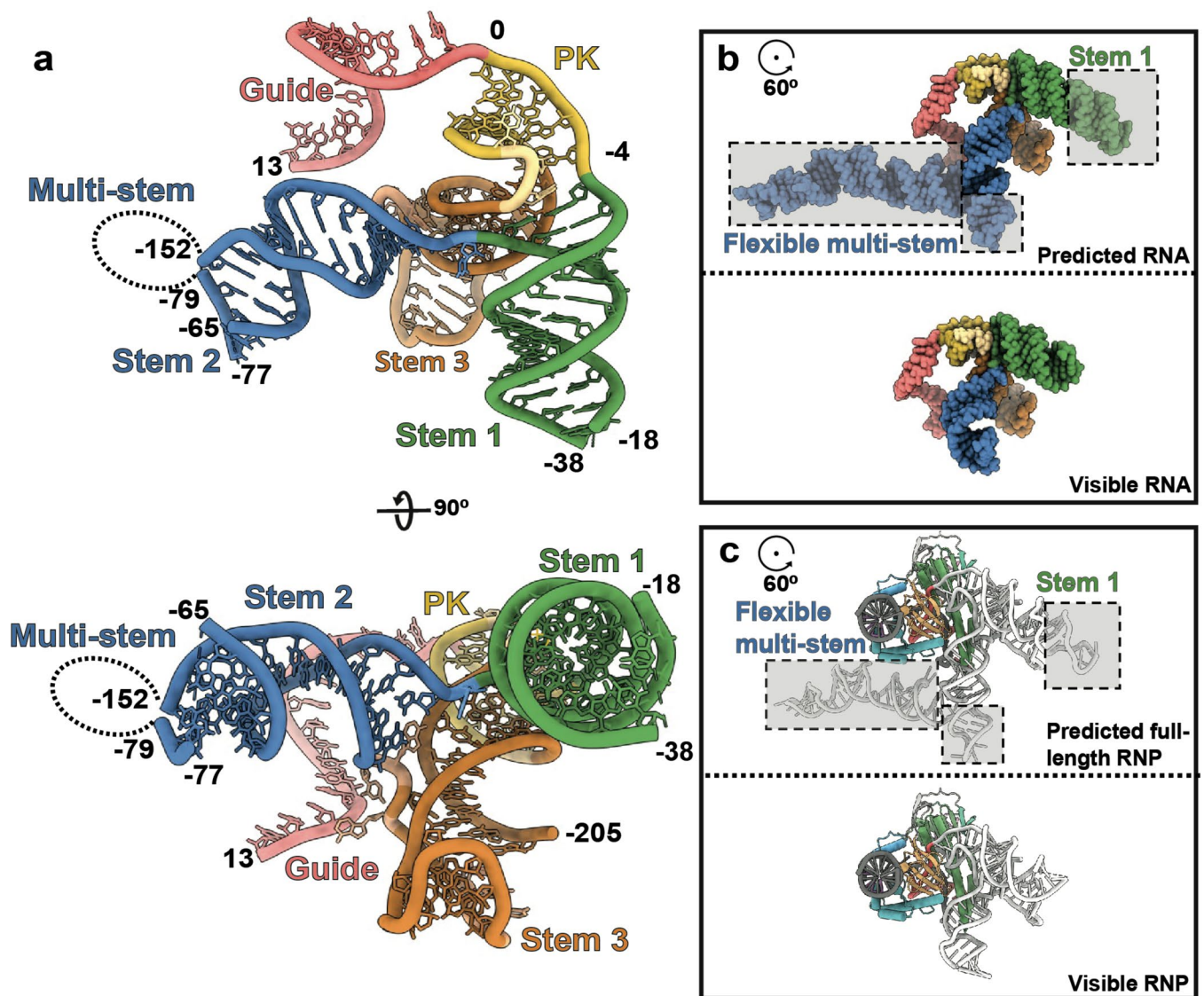


Extended Data Fig. 7 | Comparison between Fanzor2 and Canonical RuvC catalytic site. Structural comparison between Fanzor2 catalytic site and canonical catalytic site (Cas12j, PDB [7LYT](#)). Filled and open triangles indicate canonical and non-canonical catalytic residue position, respectively.



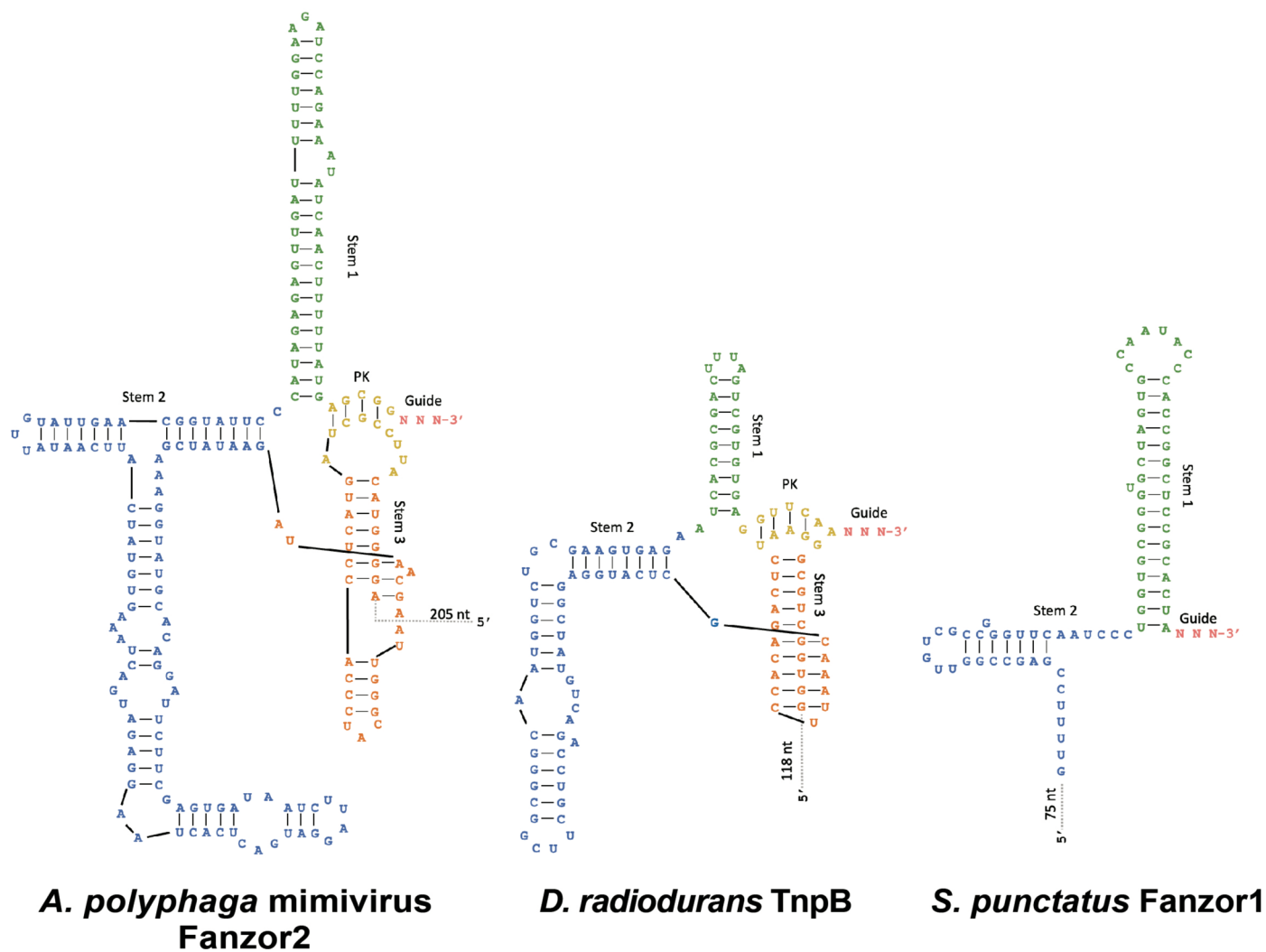
Extended Data Fig. 8 | ω RNA Recognition is mediated by multiple protein domains in ApmFz2. **a.** Recognition of ω RNA by the ApmFz2 protein using the WED, RuvC and RuvC-R domains. Panels **b–e** display hydrogen bonding (dotted

cyan lines) to ω RNA backbone or bases. The pseudoknot (PK) is recognized by the WED domain (**b**) and by RuvC and RuvC-R (**c**). Recognition of Stem2 (**d**) and Stem3 (**e**) by RuvC. Rotations relative to **a** are indicated in **c,d**.



Extended Data Fig. 9 | Comparison between observed and full-length predicted ω RNA architecture. **a.** Atomic model of ω RNA scaffold (including nucleotides -205 to -152, -79 to -77, -65 to -38, -18 to 13). Black dashed curve represents unresolved multi-stem branching from Stem2. **b-c.** Structure of predicted full-length ω RNA (top) and observed ω RNA (bottom). Only ω RNA

is shown in **b**, whereas protein and ω RNA are shown in **c**. The predicted ω RNA structure was generated by docking RNAcomposer predictions of the individual RNA stem loops onto the final cryo-EM model. Disordered regions are displayed in a dashed grey box. Rotations relative to **a** (top) are indicated. Top and bottom panels in **b** and **c** depict the same viewing direction.



Extended Data Fig. 10 | Comparison between full-length ω RNA architecture of TnpB and Fanzor variants. Comparison between TnpB and Fanzor variant predicted and experimentally determined ω RNA shown in a two-dimensional representation. RNA elements colored relative to Fig. 2.

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

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

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | 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) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | 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: ThermoFisher Talos Arctica microscope at 200kV (equipped with a K3 direct electron detector and a BioQuantum energy filter). Data collected using EPU software v3.6.
Data analysis	Cryo-EM data: cryoSPARC (v4.4.0); Structure visualization: UCSF ChimeraX (v1.6.1); Model building: Coot (v0.9.8.2), PHENIX (v1.21); Validation: Molprobity (v4.5.2), Phenix (v1.21).

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 Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Cryo-EM map and model for ApmFz2-ωRNA-DNA ternary structure were deposited in the EMDB and PDB, respectively, with accession codes EMD-44046 and PDB

9BOL. Other atomic coordinates used in this study for the comparison purposes are available from the PDB under the following accession codes: 8GKH for SpuFz1-ωRNA-DNA, 8H1J for ISDra2TnpB-ωRNA-RNA, 8EXA for ISDra2TnpB-reRNA-RNA, and 7LYT for Cas12j-crRNA-(Phosphorothioate-DNA).

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender N/A

Reporting on race, ethnicity, or other socially relevant groupings N/A

Population characteristics N/A

Recruitment N/A

Ethics oversight N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

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](https://www.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 189,912 particles originating from 3,205 movies were used for our final cryo-EM reconstruction.

Data exclusions Particles in 2D classes showing contaminants and 3D classes showing unsatisfactory structural features were excluded from the final reconstructions in the datasets analyzed.

Replication Biochemical assays were performed three independent times with similar results.

Randomization No randomization was involved as this study did not involve use of sample collection from different experimental groups involving participants.

Blinding Blinding is not applicable as this study did not involve sampling different experimental groups involving participants.

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

☒ ☐ Clinical data

☒ ☐ Dual use research of concern

☒ ☐ Plants

Methods

n/a Involved in the study

☒ ☐ ChIP-seq

☒ ☐ Flow cytometry

☒ ☐ MRI-based neuroimaging

Seed stocks	Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.
Novel plant genotypes	Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.
Authentication	Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.