Supplementary Information

Cryo-EM structure of the ssDNA-activated SPARTA complex

Minghui Guo[1,4], Yuwei Zhu[1,4], Zhiying Lin[1,4], Dehui Yang[1], Anqi Zhang[1], Changyou Guo[1], and Zhiwei Huang[1,2,3]✉

[1]HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China

[2]Westlake Center for Genome Editing, Westlake Laboratory of Life Sciences and Biomedicine, School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[3]New Cornerstone Science Laboratory, Shenzhen 518054, China

[4]These authors contributed equally: Minghui Guo, Yuwei Zhu, Zhiying Lin

✉*Correspondence:* huangzhiwei@hit.edu.cn

Materials and Methods

Plasmids

The genes that encode *Maribacter polysiphoniae* SPARTA complex were cloned into pACYCDuet-1 vector. The TIR-APAZ, which was linked to an N-terminal-6x-his-TwinStrep-SUMO-tags by PreScission Protease cleavage site-containing linker, was cloned in the MCS I. The Argonaute was cloned in the MCS II of the pACYCDuet-1 vector. Point mutations of SPARTA complex were introduced using Fast Site-Directed Mutagenesis Kit and verified by DNA sequencing.

Protein expression and purification

The SPARTA complex was heterologously expressed in *E. coli* BL21 (DE3) with overnight induction of 0.3 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at 18°C. Cells were harvested by centrifugation at 5895 g for 15 min and were resuspended in binding buffer A (25 mM Tris, PH 8.0, 150 mM NaCl, 1 mM protease-inhibitor PMSF). Cells were lysed by sonication, and the lysate was centrifuged at 23,708 g for 40 min at 4°C. The supernatant was bound to Strep-Tactin resin, and the resin was washed ~5 column volumes (CV) in binding buffer A. Proteins were cleaved by PreScission Protease overnight at 4°C to remove TwinStrep-SUMO tags. The sample was diluted three times with buffer B (25 mM Tris, pH 8.0, 3 mM DTT) and was loaded on a HiTrap Heparin column (Cytiva Life Sciences).

Size-exclusion chromatography

Purified SPARTA complex was mixed in a 1:1.2 molar ratio with 5'-phosphorylated guide RNA/target ssDNA hybrid for 1.5 h at 37°C. The guide RNA and target ssDNA were annealed in 1:1 ratio by incubation at 95°C for 5 min and gradually cooled to room temperature beforehand. The complex was loaded onto Superdex 200 increase 10/300 GL(Cytiva Life Sciences) in buffer (25 mM HEPES, pH 7.5, 125 mM KCl, 2 mM $MgCl_2$). SDS-PAGE gels were stained with Coomassie blue R250 to identify proteins. RNA and ssDNA were monitored using 10% Denaturing TBE-Urea and were visualized by ethidium bromide staining. Purified complexes were concentrated to 1-1.5 mg/ml, flash frozen in liquid nitrogen, and stored at −80°C.

ε-NAD assays

ε-NAD assays were performed in a 60 μL system containing 2 μM SPARTA complex, 1 μM guide RNA, 1 μM ssDNA and 25 μM ε-NAD (MACKLIN). The purified SPARTA complex, guide RNA and ε-NAD were pre-incubated in reaction buffer (10 mM MES, PH 6.5, 125 mM KCl, 2 mM $MgCl_2$) at room temperature for 10 min. After the addition of the target ssDNA, the 96-well plate was transferred to preheated Varioskan LUX (Thermo Fisher Scientific). Reactions were conducted at 37°C for 120 min. Fluorescence intensity was measured in kinetic mode using an excitation wavelength of 310 nm and emission wavelength of 410 nm. All experiments were performed in triplicates, and error bars indicate standard deviations.

Grid preparation and cryo-EM data collection

A 5 μL SPARTA-gRNA-ssDNA complex was applied to glow-discharged holey Au 300 grids (Quantifoil R1.2/1.3). The grids were vitrified in liquid ethane using a FEI Vitrobot Mark IV. The Vitrobot was operated at 4 °C, with a blot force of "0-2", and a blotting time of 2-6 s, and 100% humidity. Cryo-grids were first screened on a FEI 200 kV transmission electron microscope (Talos Arctica equipped with a FEI Falcon 3 camera). Data collection was taken by a Gatan K3 direct electron detector operated at 300 kV (FEI Titan Krios) using the Thermo Fisher Scientific EPU software. The energy filter slit was set to 20 eV. The nominal magnification of 81,000X and the defocus range between -1.0 and -2.5 μm were used for data collection. For each image stack (40 frames), a total dose is about 50.0 electrons/$Å^2$ at a calibrated pixel size 1.0773 Å.

Cryo-EM data processing

We use a similar pipeline for data processing as described before[1]. A total of 4000 micrographs were obtained for data processing (Supplementary information, Fig. S1). Movie stacks were drift-corrected, electron-dose weighted and two-fold binned using MotionCor2[2]. The contrast transfer function (CTF) parameters of drift-corrected micrographs were estimated by the program CTFFIND4[3]. A total of 5,034,199 particles were automatically selected using Blob picker in cryoSPARC[4]. Then, the particles were extracted with 1×binning (1.0773 Å per pixel), and subjected to two rounds of 2D classifications to remove junk particles. About 100,000 particles were selected for ab initio reconstruction to generate initial model. All resulting 2,635,135 particles were then used for two rounds of heterogeneous refinement with four classes. A total of 832,006 particles from the best class were subjected into non-uniform refinement and yielded a reconstruction at 2.84 Å

overall resolution. The resolution was evaluated based on the gold standard Fourier shell correlation (threshold = 0.143), and the local resolution map was estimated by cryoSPARC.

Model building and structure refinement

For the structure of SPARTA-gRNA-ssDNA complex, model building was carried out based on the 2.84 Å reconstruction map (Supplementary information, Fig. S2). The structure of SPARTA complex predicted by AlphaFold[5] was docked into the electron microscopy density map using UCSF Chimera[6] and manual adjusted using COOT[7]. The atomic coordinate of gRNA-ssDNA heteroduplex was built *de novo*. The model was then refined using phenix. real_space_refine[8] application with secondary structure and geometry restraints. The final models were evaluated by MolProbity[9] and Ramachandran plot[10]. Statistics of the map reconstruction and model refinement are presented in Supplementary information, Table S1. Structural figures were prepared with PyMOL and Chimera.

References

1 Ma, X. *et al.* Cryo-EM structures of two human B cell receptor isotypes. *Science* **377**, 880-885, doi:10.1126/science.abo3828 (2022).

2 Zheng, S. Q. *et al.* MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods* **14**, 331-332, doi:10.1038/nmeth.4193 (2017).

3 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216-221, doi:10.1016/j.jsb.2015.08.008 (2015).

4 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, doi:10.1038/nmeth.4169 (2017).

5 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

6 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

7 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

8 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

9 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).

10 Hovmoller, S., Zhou, T. & Ohlson, T. Conformations of amino acids in proteins. *Acta Crystallogr D Biol Crystallogr* **58**, 768-776, doi:10.1107/s0907444902003359 (2002).

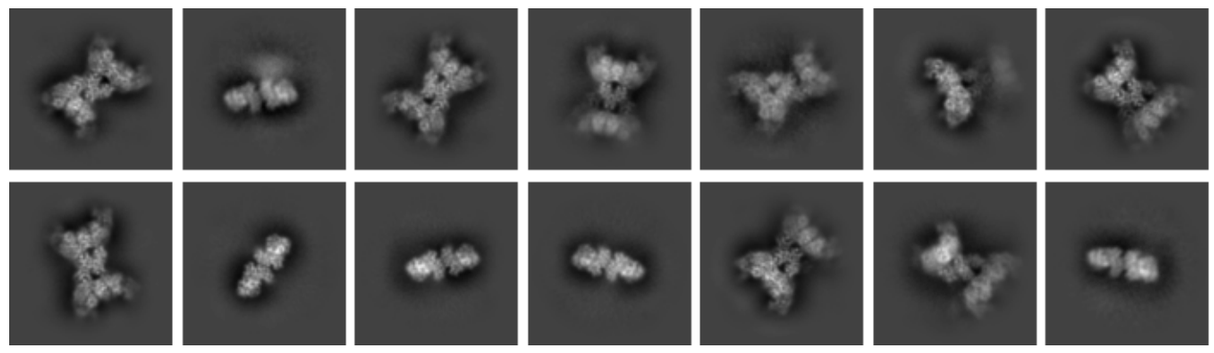
a

4,000 Micrographs

Blob Pick

5,034,199 Particles

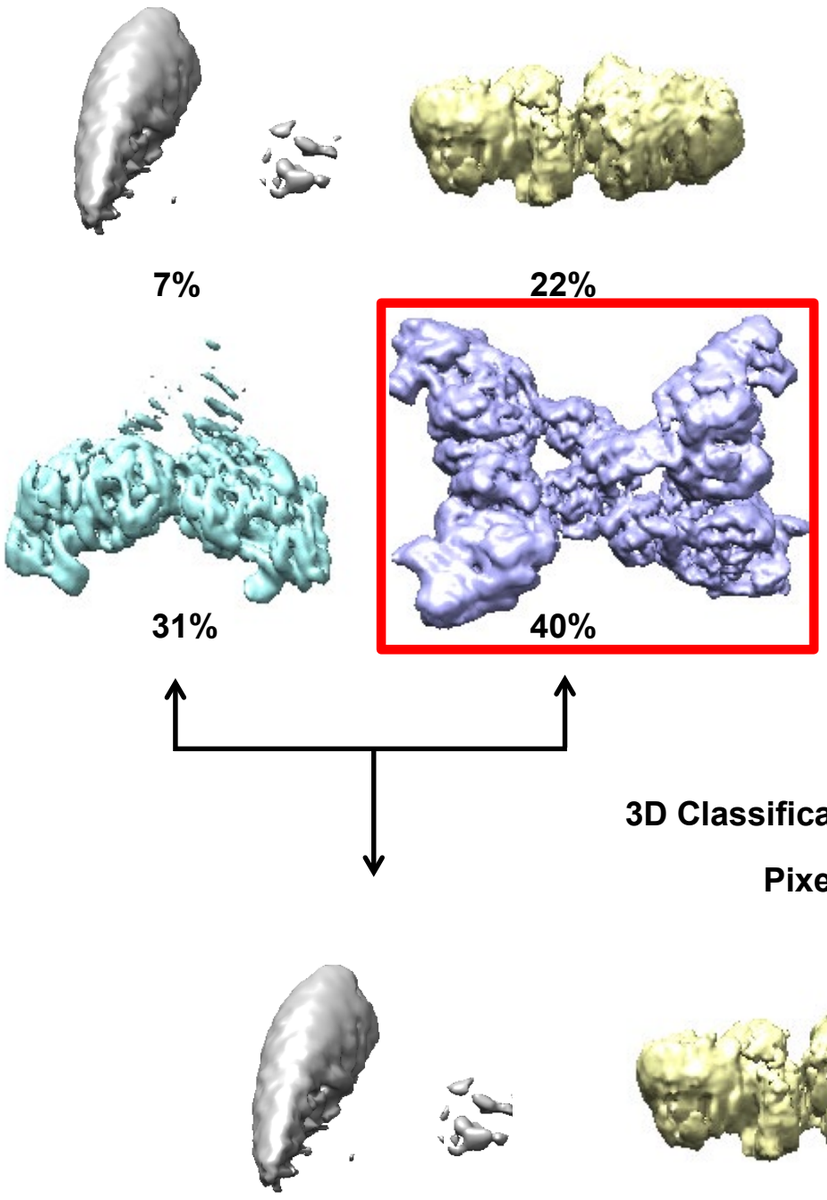
2D Classification



2,635,135 Particles

3D Classification/C1 symmetry

Pixel size: 2.1546 Å



3D Classification/C1 symmetry

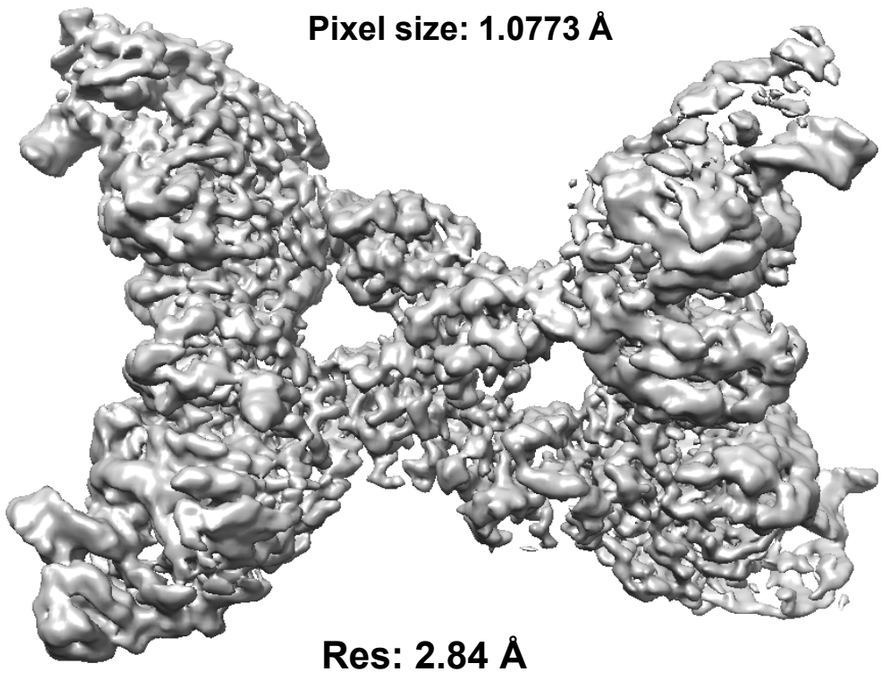
Pixel size: 2.1546 Å

Remove duplicates

non-uniform refinement

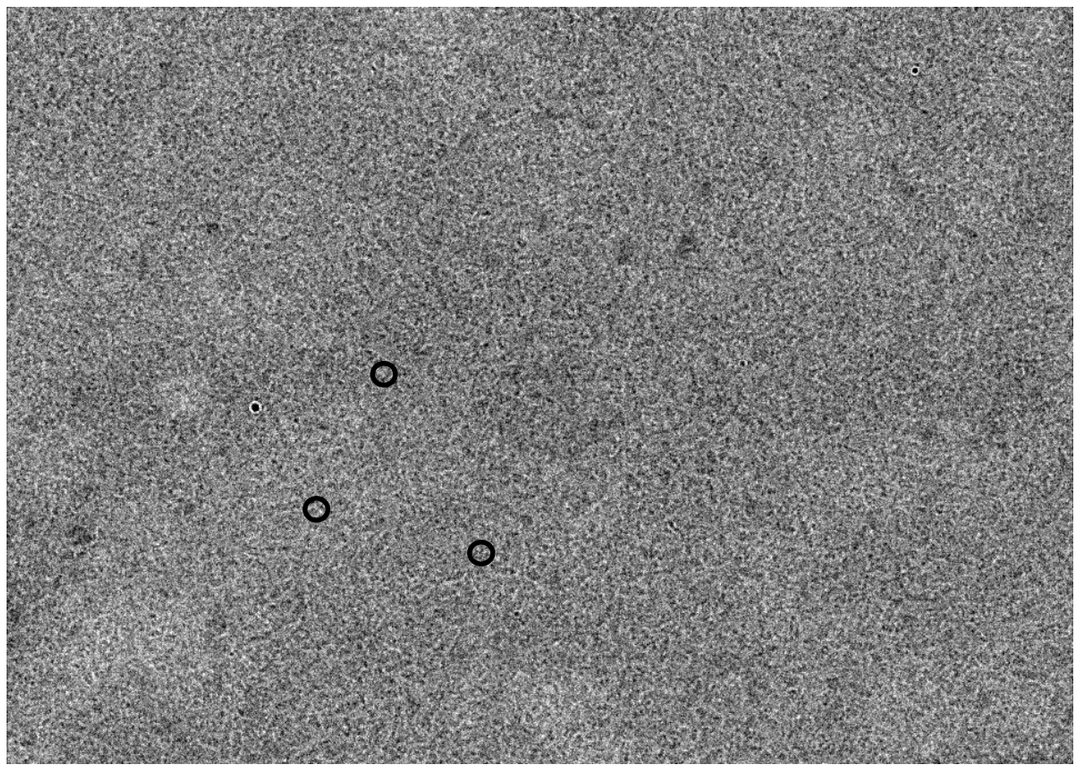
832,006 Particles

Pixel size: 1.0773 Å

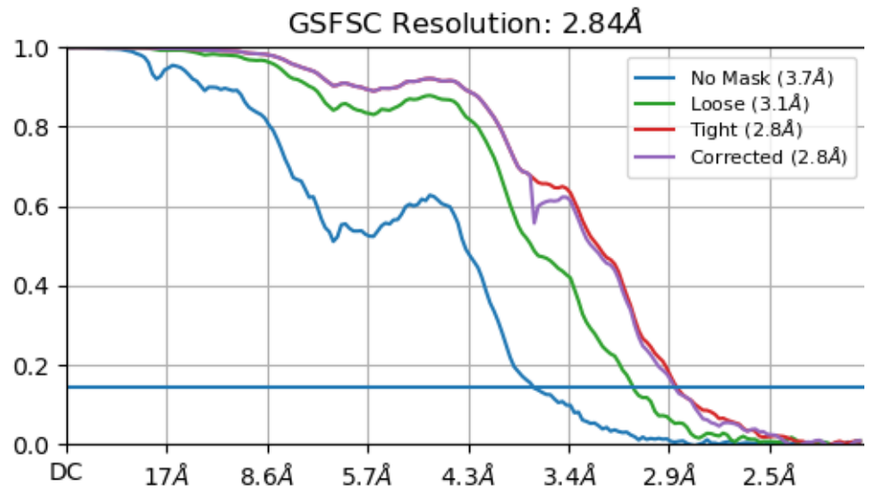


Res: 2.84 Å

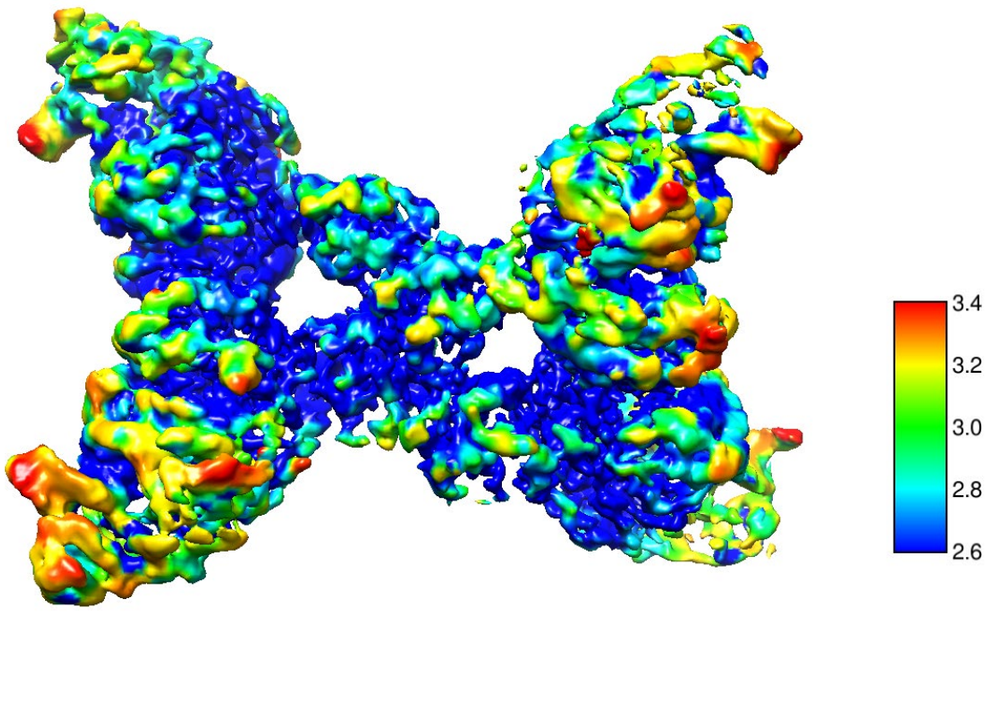
b



c



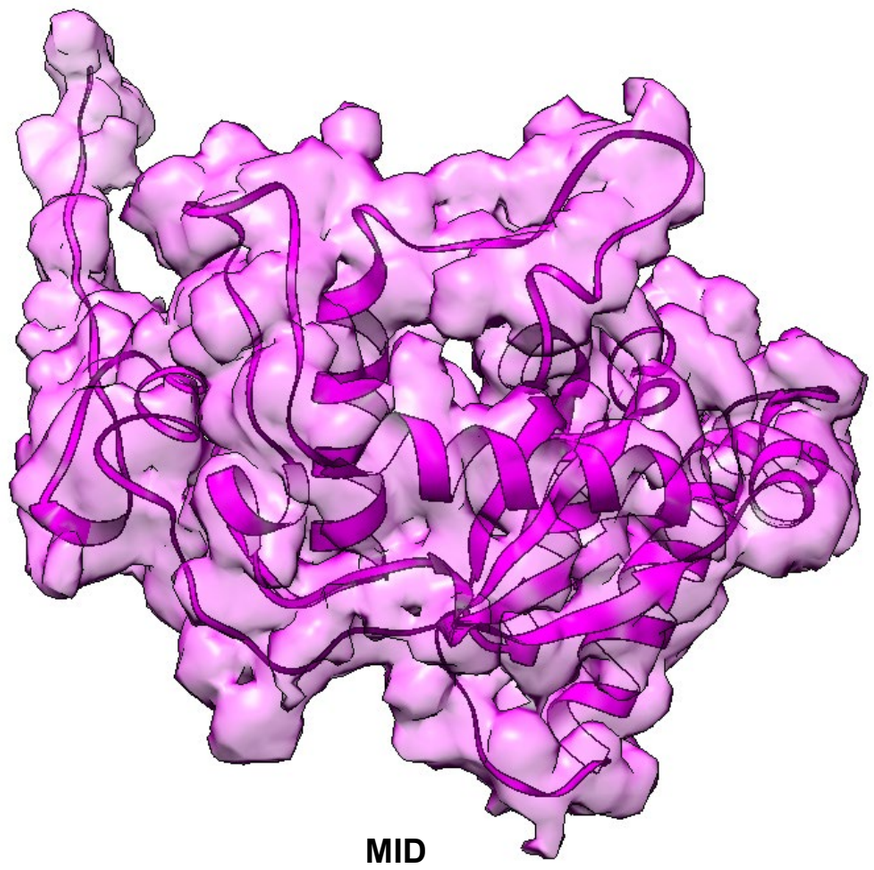
d



Supplementary information, Fig. S1 Cryo-EM image processing procedure of the SPARTA-gRNA-ssDNA complex.

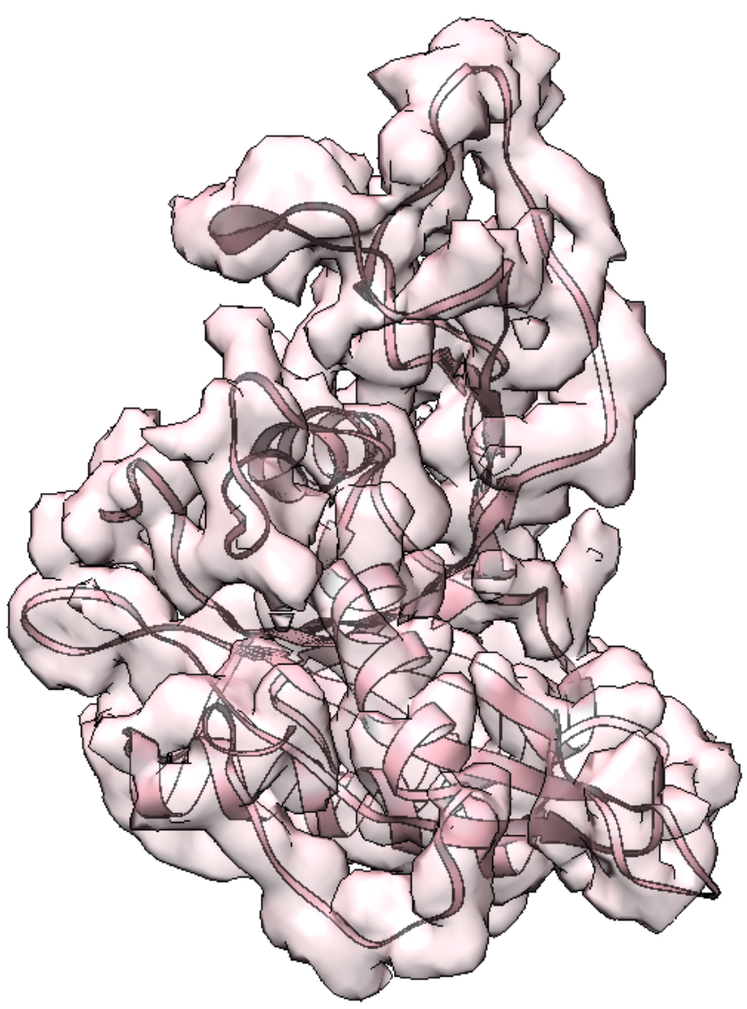
a, Image processing workflow of the SPARTA-gRNA-ssDNA complex. **b**, A representative raw cryo-EM image of the SPARTA-gRNA-ssDNA complex. The particles that were picked for classification are labeled with circles. **c**, Gold standard FSC plot for the final 3D reconstruction of the SPARTA-gRNA-ssDNA complex. **d**, Resolution map for the final 3D reconstruction of the SPARTA-gRNA-ssDNA complex. The color code for resolutions, shown with the unit Å, is calculated using cryoSPARC.

a



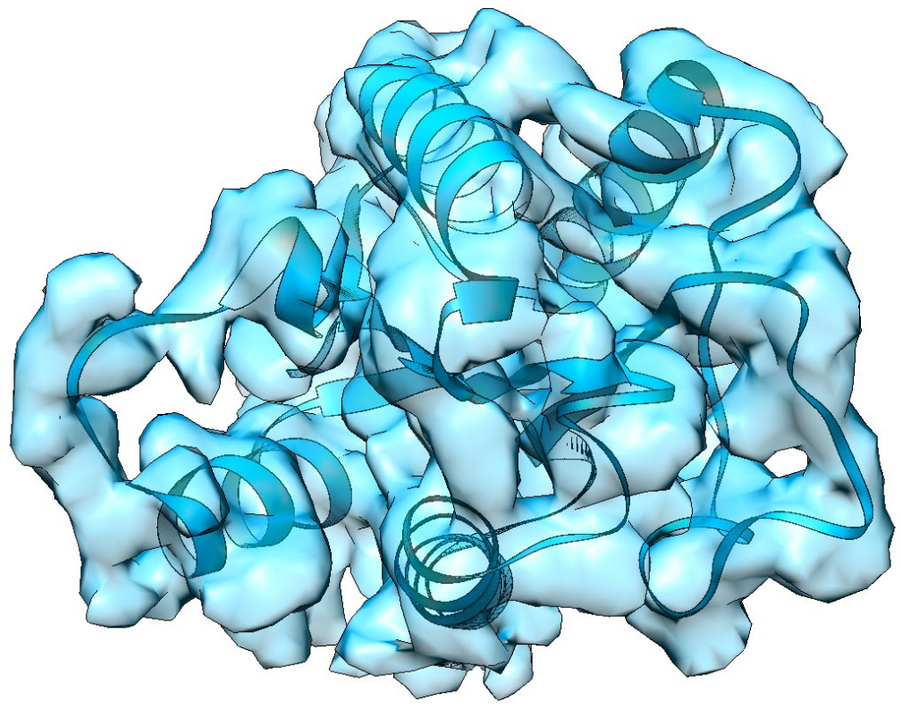
MID

b



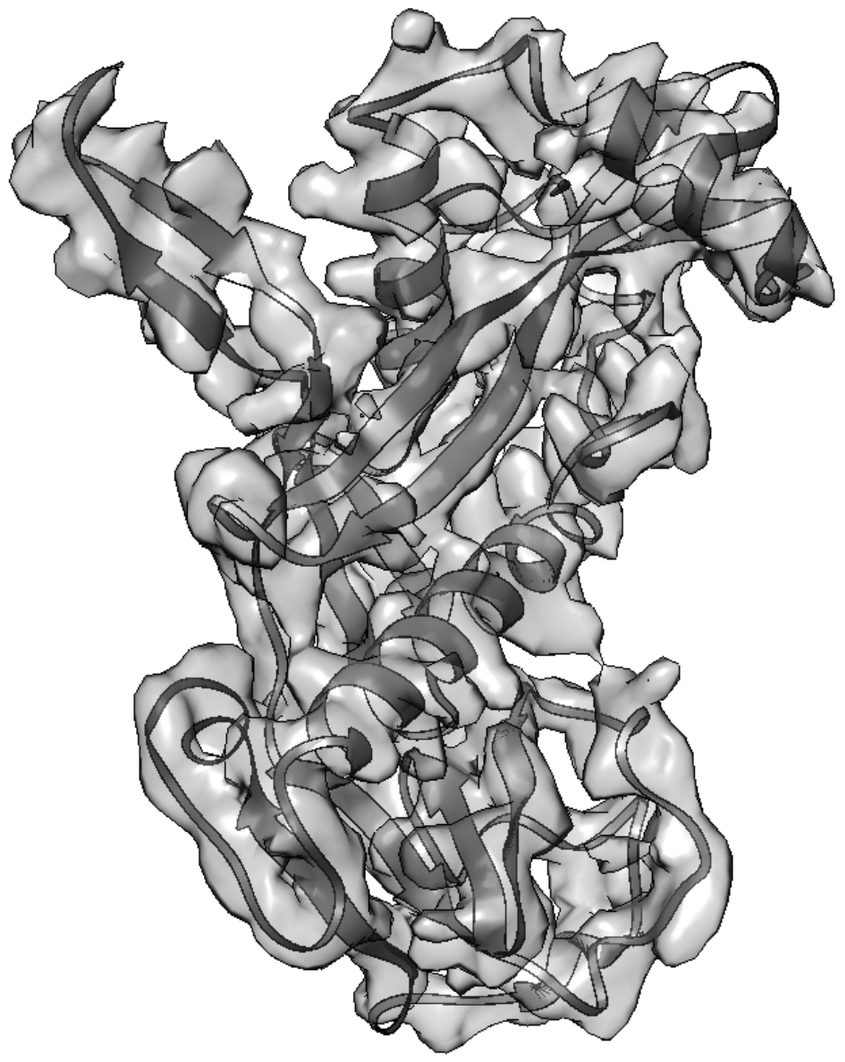
PIWI

c



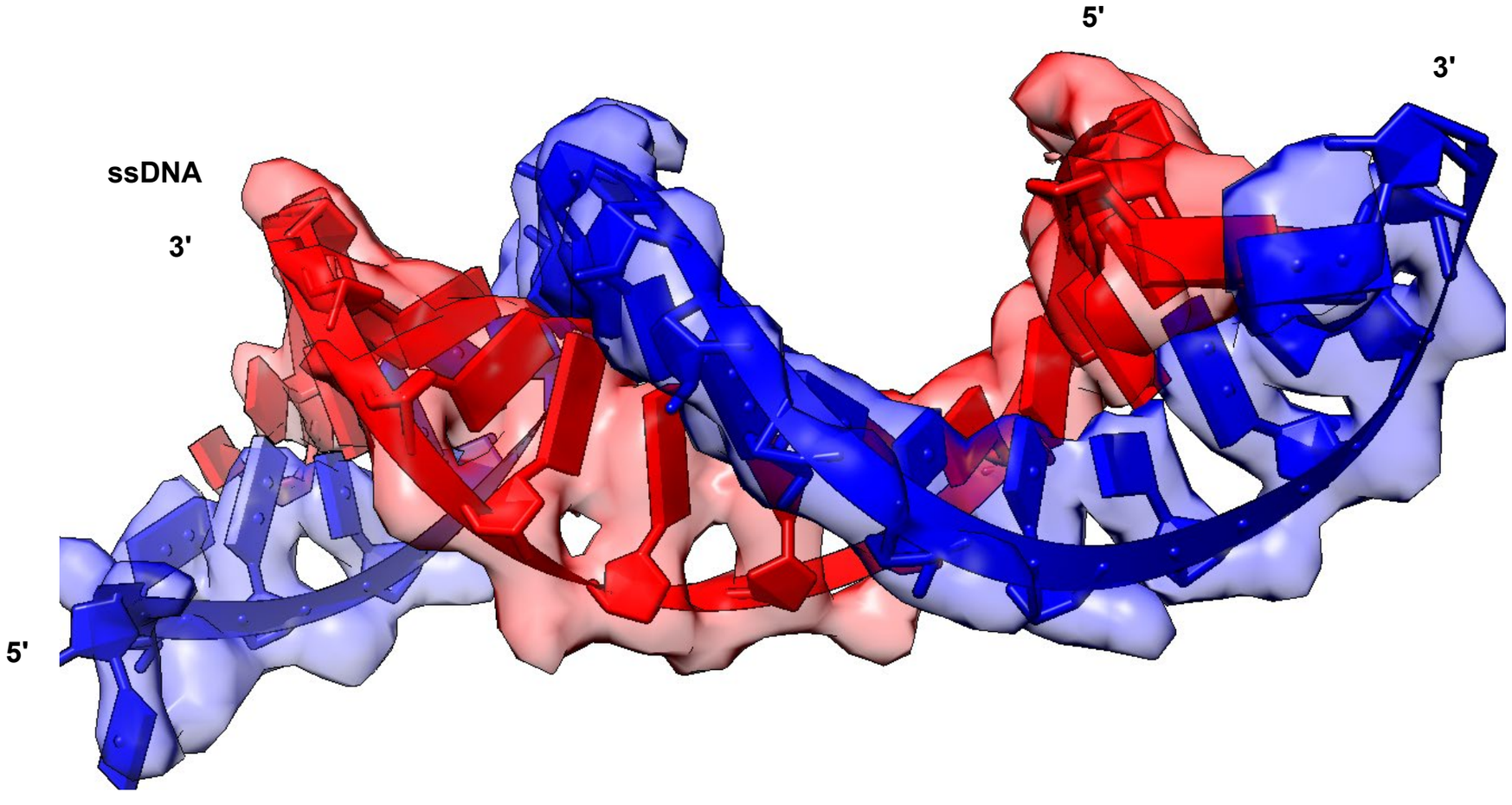
TIR

d



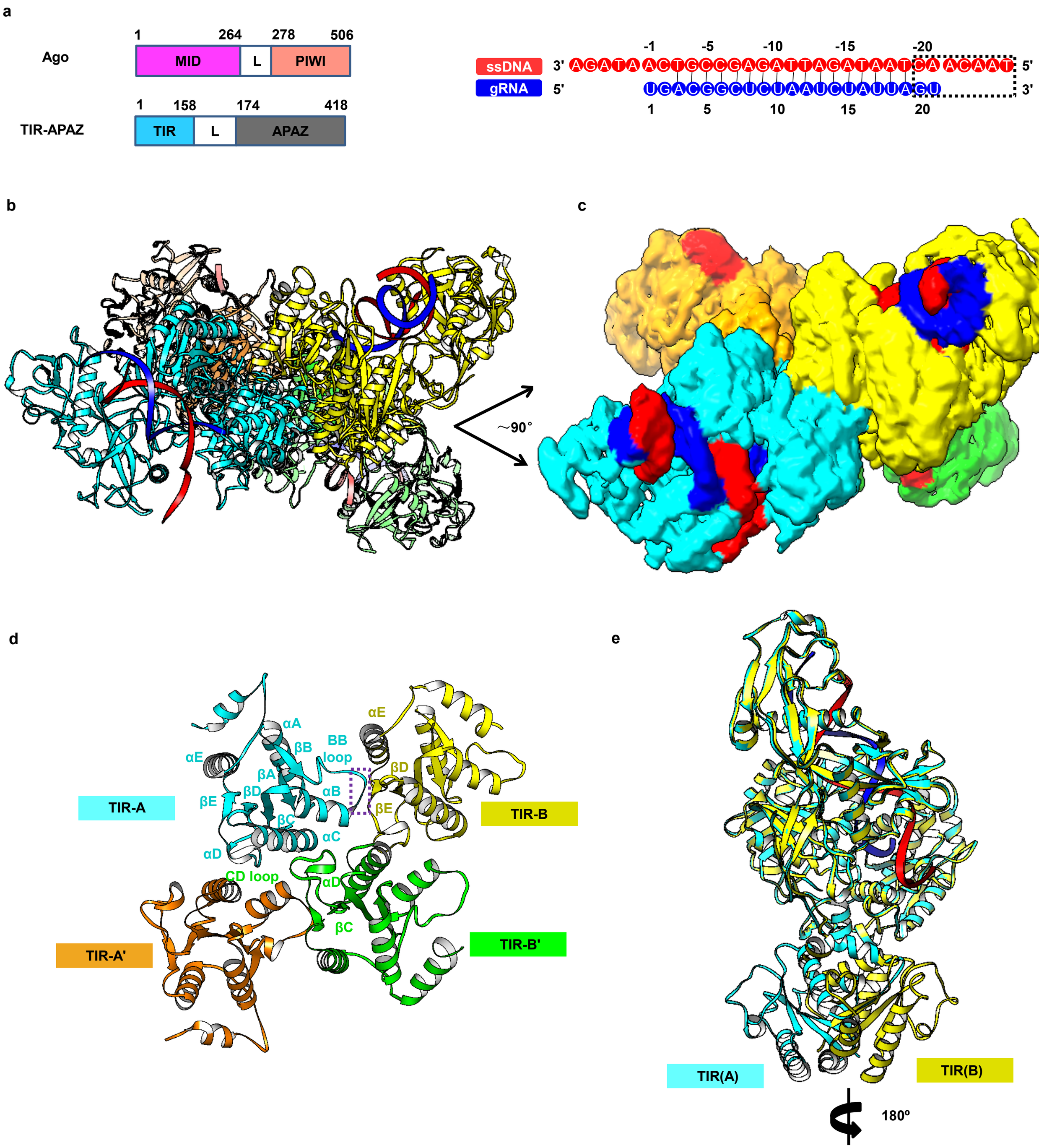
APAZ

e



gRNA

Supplementary information, Fig. S2 Representative local map density for the SPARTA-gRNA-ssDNA complex.
a, Cryo-EM density for the MID domain of Ago protein. **b**, Cryo-EM density for the PIWI domain of Ago protein. **c**, Cryo-EM density for the TIR domain of TIR-APAZ protein. **d**, Cryo-EM density for the APAZ domain of TIR-APAZ protein. **e**, Cryo-EM density for the guide RNA and target ssDNA heteroduplex.



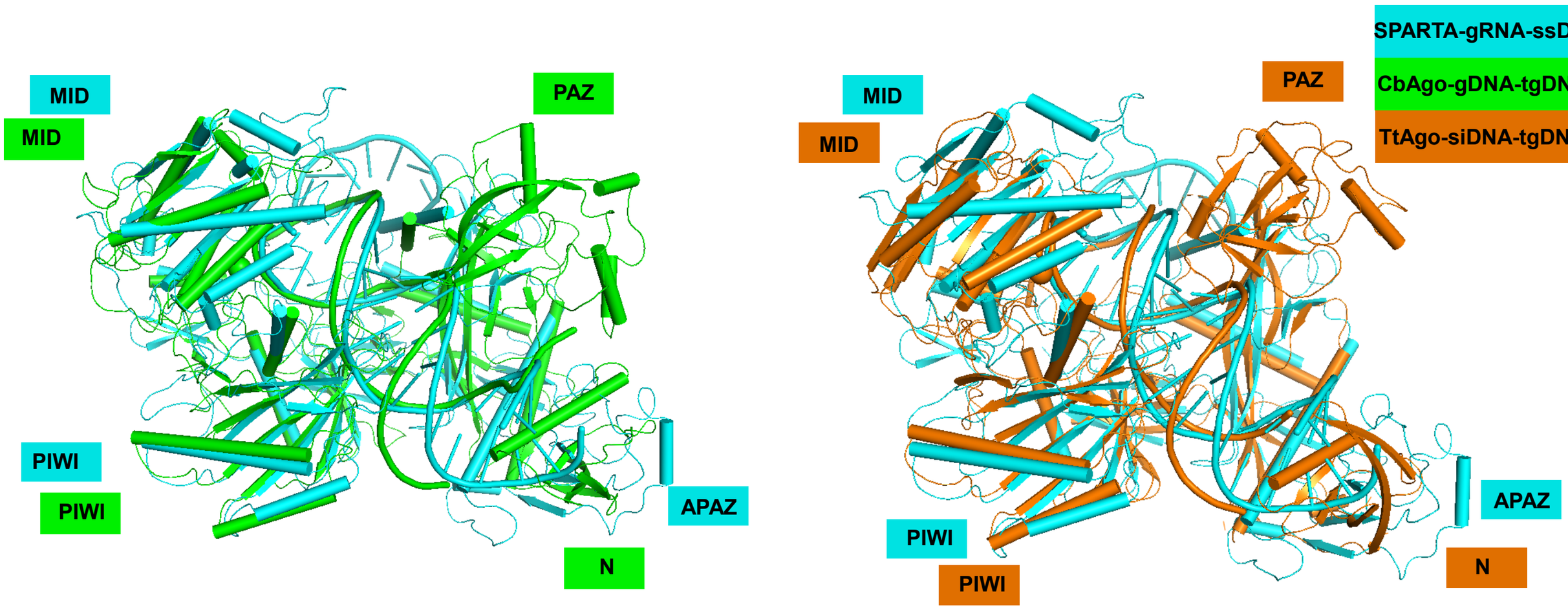
Supplementary information, Fig. S3 Cartoon representation of the SPARTA-gRNA-ssDNA complex.

a, Graphic representation of the domain organization of the Ago and TIR-APAZ proteins. Schematic representation of the gRNA: target ssDNA heteroduplex, while the disordered regions are enclosed in dashed boxes. The MID, PIWI, TIR, APAZ, gRNA and target ssDNA are shown in magenta, pink, sky blue, black, blue, and red, respectively. **b**, Top view of the structure of tetrameric SPARTA-gRNA-ssDNA complex. Four asymmetric molecules are shown in cyan, yellow, orange, and green, respectively. **c**, Cryo-EM density map of the top view of tetrameric SPARTA-gRNA-ssDNA complex. **d**, Structural comparison of monomeric SPARTA-gRNA-ssDNA complex from molecule A and B. **e**, Enlarged view of tetrameric interface of SPARTA-gRNA-ssDNA complex.

a

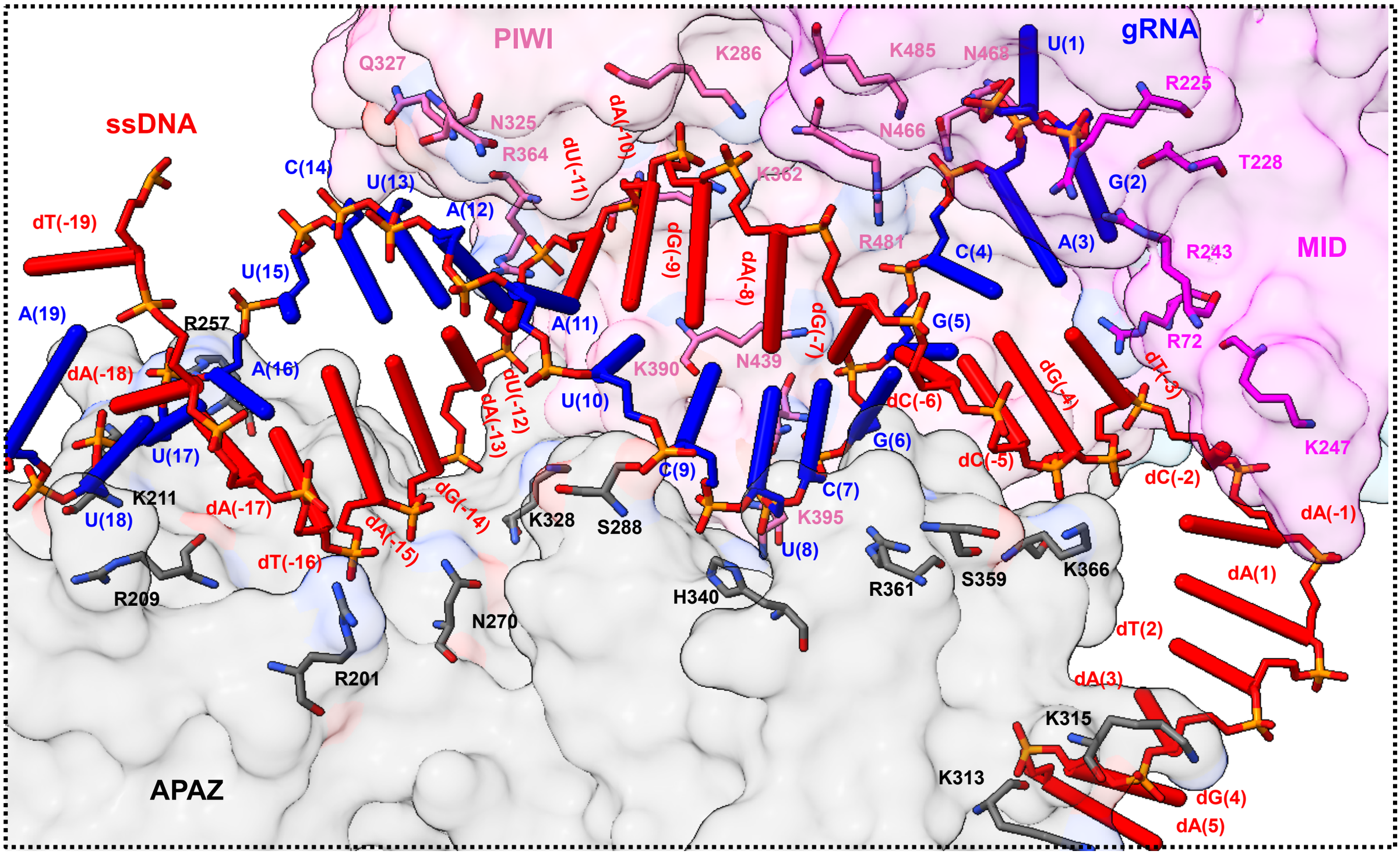
```
MapAgo      MKELIYIEEPKILFAHGQKCTDARDGLALFGPLNNLYGIKSGVIGTKQGLKIFRDYLDHIQKPIYNSNSITRPMFPGFEA   80
CbAgo       ..........................................................TNFDFMGFESGVLEEPVLMGANG   206
TtAgo       ...........................................................................
consensus                                                                * *  *  **  *  *

MapAgo      .VFDCKWESTGITFKEVTNEDIGKFLYNSSTHKRTYDL.VSLFIDKIISANKNEDENVDVWFVIVPDEIYKYCRPNSVLP   158
CbAgo       KIKDKKQIFINGFFKNPKENVKFGVLYPEGCMENAQSI.ARSILDFATAGKYNKQENK....YISKNLMNIGFKPSECIF   429
TtAgo       ...........AVRAQAYRLSIPKLMGRRAVSKPADALRVGFYRAQETALALLRLDGAQGWPEFLRRALLRAFGASGASL   376
consensus    * * *     * ***     * **** * * * * * *** *** ** **   *  ***  * ******  *

MapAgo      KEMVQTKALMSKSKAKSFRYEPSLFPDINIELKEQEKEAETYNYDAQFHDQFKARLLKHTIPTQIFRESTLAWRDFKNAF   238
CbAgo       ESYKLGDITEYKATARKLKEHEKVGFVIAVIPDMNELEVENP......YNPFKKVWAKLNIPSQMITLKTT..EKFKN..   499
TtAgo       RLHTLHAHPSQGLAFREALRKAKEEGVQAVLVLTPPMAWED.......RNRLKALLLREGLPSQILNV.............   437
consensus   *   *       *  **  *   **  ******  **** !          * *!*****  *!*!** * *     ***

MapAgo      GLPIRDFSKIEGHLAWTISTAAFYKAGGKPWKLSDVRNGV.CYLGLVYKKVEKSKNPRNACCAAQMFLDNGDGTVFKGEV   317
CbAgo       ...IVDKSGL..YYLHNIALNILGKIGGIPWIIKDMPGNIDCFIGLAVGTREKGIH.FPACSV..LFDKYGKLINYYKPT   571
TtAgo       ..PLREEER...HRWENALLGLLAKAGLQVVALSGAYP.AELAVGFDAGGRESFRFGGAACAVG...GDGGHLLWTLPEA   508
consensus     **** ***   *    ** *****!*!* ** ****    *****!*  * *!* *     !! ** ** * ! ** *   *

MapAgo      GPWYNPKNGQQYHLEPKEAKALLSQSLQSYKEQIGEYPKEVFIHAKTRFNHQEWDAFLEVTPKETNLVGVT.ISKTKPLKL   396
CbAgo       IPQSGEKIAETILQE.....IFDNVLISYKEENGEYPKNIVIH.RAGFSRENIDWYKEYFDKKGIKFNIIEVKKNIPVKI   645
TtAgo       ..QAGERIPQEVVWD.....LLEETLWAFRRKAGRLPSRVLLLRDGRVPQDEFALALEALAREGIAYDLVSVRKSGGGRV   581
consensus    ** ****** ** *      *****! **** !**!* **** **** **** **!   **** * ** **!* ****

MapAgo      YKTEGDYTILR.GNAYVVNERSAFLWTVGYVPKIQTALSMEVPNPLFIEINKGEADIKQVLKDILSLTKLNYNACIFADG   475
CbAgo       AKVVGSNICNPIKGSYVLKNDKAFIVTTDIKDGVAS......PNPLKIEKTYGDVEMKSILEQIYSLSQIHVGSTKSLR.   718
TtAgo       YPVQGRLA....DGLYVPLEDKTFLLLTVHRDFRGT......PRPLKLVHEAGDTPLEALAHQIFHLTRLYPASGFAFPR   651
consensus   *** !          * !!* ****!****   ** ***       !*!!***   !* *** ** *!**!***   **

MapAgo      EPVTLRFADKIGEILTA..STDIKTPPLAFKYYI   507
CbAgo       LPITTGYADKICKAIEY..IPQGVVDNRLFFL..   748
TtAgo       LPAPLHLADRLVKEVGRLGIRHLKEVDREKLFFV   685
consensus   *!*** *!!** * *     *   **    * * ***
```

b



Supplementary information, Fig. S4 Sequence alignment and structural comparison of *Map*Ago, *Cb*Ago and *Tt*Ago.

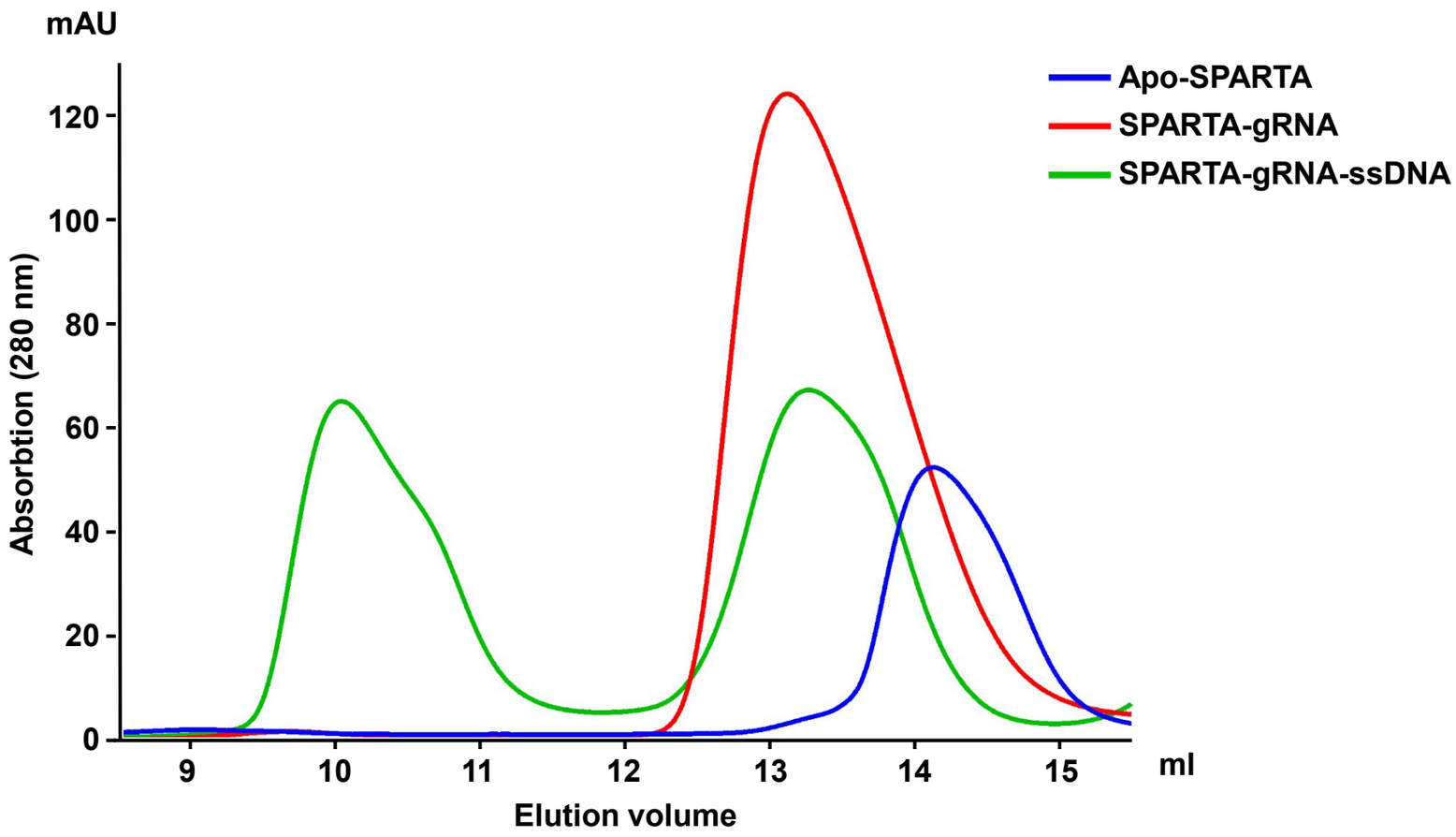
a, The MID and PIWI domains of *Map*Ago, *Cb*Ago and *Tt*Ago were aligned and analyzed. **b**, Structural comparison of SPARTA-gRNA-ssDNA with *Cb*Ago-gDNA-tgDNA and *Tt*Ago-siDNA-tgDNA, respectively.



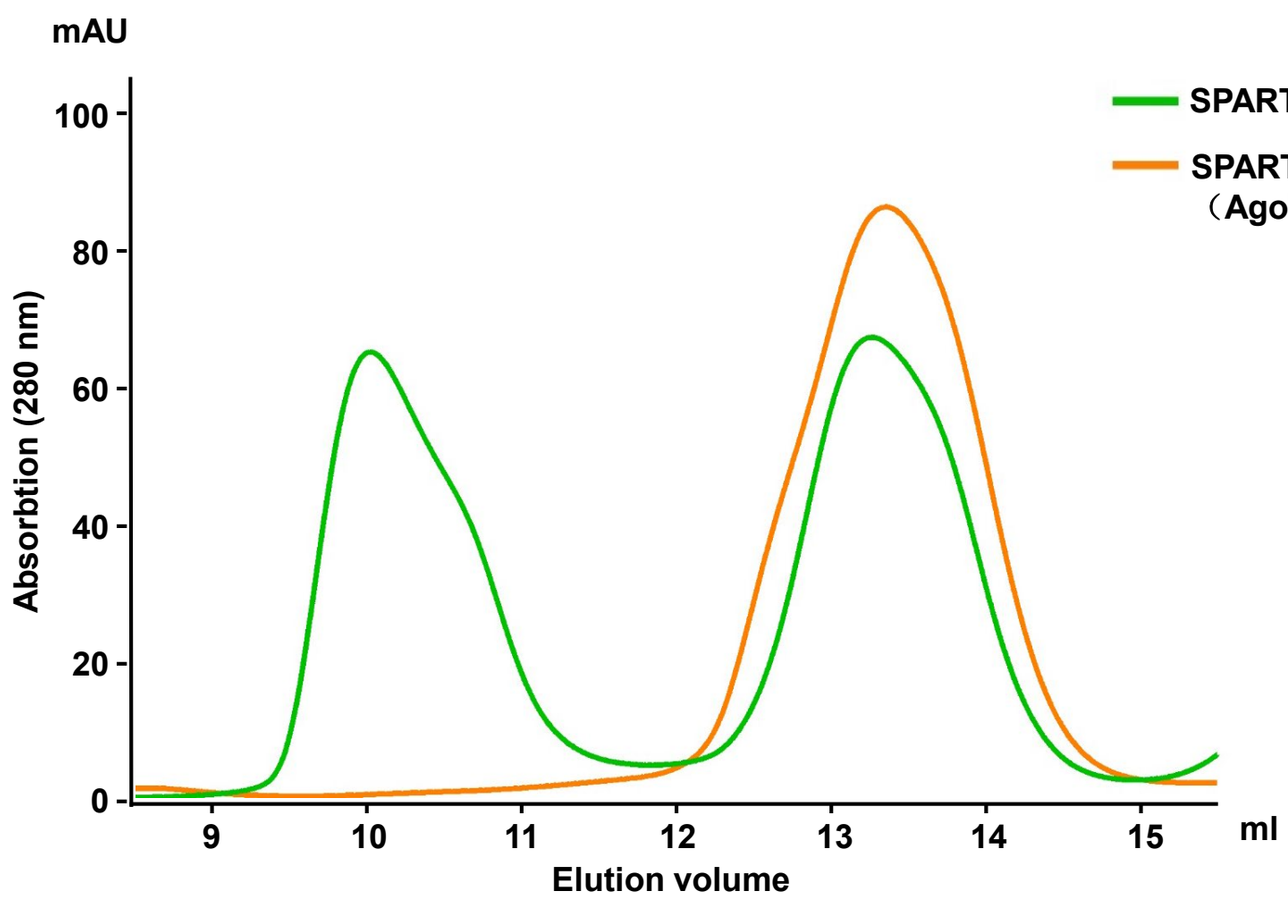
Supplementary information, Fig. S5 Interaction of SPARTA complex with the backbone phosphates of the gRNA-ssDNA heteroduplex.

Details show of the interaction of SPARTA complex with the backbone phosphates of the gRNA-ssDNA heteroduplex. The residues involved in the interaction are labeled and shown as sticks. The monomeric SPARTA-gRNA-ssDNA complex is colored as Fig. 1c.

a



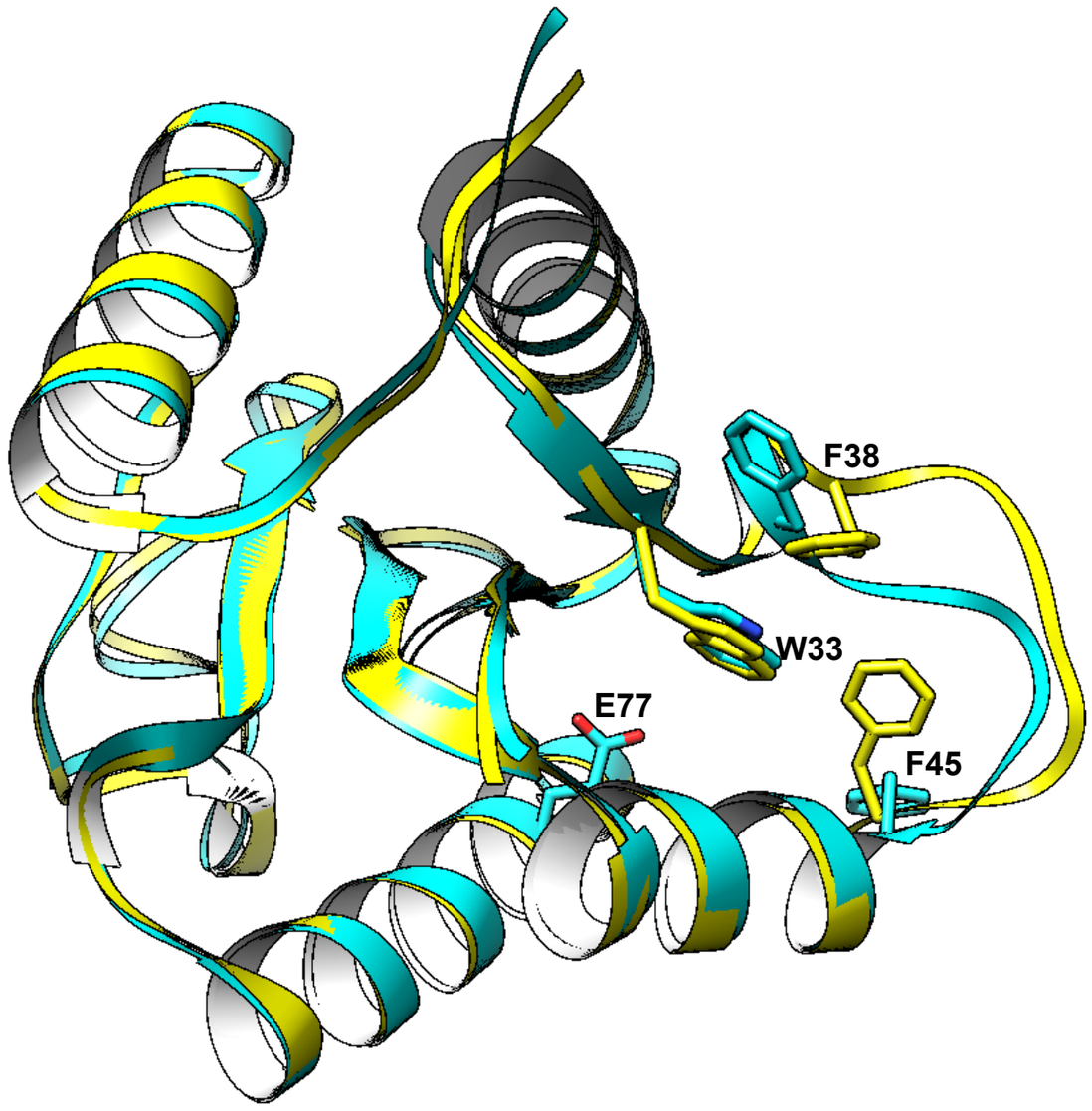
b



Supplementary information, Fig. S6 SPARTA oligomerizes upon guide RNA-mediated target ssDNA binding.

a, SPARTA was incubated with guide RNA and target ssDNA in a 1:1.2:1.2 molar ratio and subjected to size-exclusion chromatography analysis. Colour codes are indicated. **b**, The size exclusion chromatography results of SPARTA complex and SPARTA complex ($Ago^{N131A/E132A-K314A}$). Colour codes are indicated.

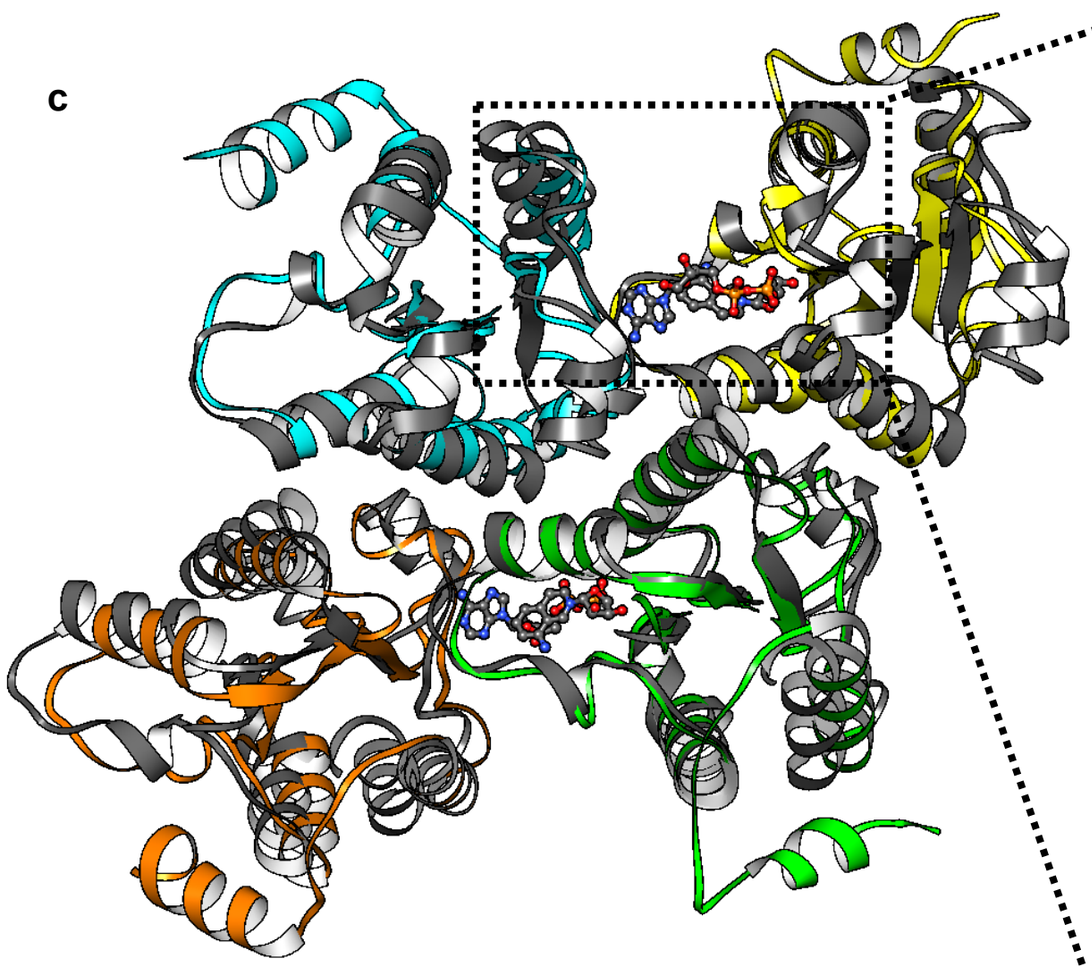
a



b

```
Map-TIR:1-140     MRNK..IFISHATPDDNDFTRWLALKLIGLGYEVWCDILFLDKGVDFWSNIEKV.I.REDTCKF..LLVSSSYSN
Ab-TIR:131-267    SNAEYDLFISHASEDKEDFVRPLAETLQQLGVNVWYDEFTLKVG.D..SLRQKIDSGLRNS.KYGTVVLSTDFI.
consensus               *!!!!!* !  !! ! !!  !  !!  !! ! !  !  ! !  !    !*      * !*  ***!* *

Map-TIR:1-140     QREGVLKELAVAAKVKKQLKDDKFIIPLAIDEQLSYDDINIDIV.RLN.AIDFKMSWARGLKDILEAFEKQKV
Ab-TIR:131-267    KKDWTNYELD.GL.VAREMNGHKMILPIW..HKITKNDV.LDYSPNLADKVALNTS.VNSIEEIAHQLADVIL
consensus         **    !!  *  !  * *   !  !*!*     **  !*  *!    !    *    !     * *!       *
```

c



Supplementary information, Fig. S7 The putative active center of the tetrameric TIR domains.

a, Structural superimposition of the TIR domains between Molecule A (A') and Molecule B (B'). The TIR domains in TIR-APAZs are colored as Fig. 1b. **b**, Sequence alignment of the TIR domains from TIR-APAZ and *Ab*Tir proteins. **c**, Structural comparison of the TIR domains in TIR-APAZ and *Ab*Tir (PDB ID: 7UXU). The TIR domains in TIR-APAZs are colored as Fig. 1b, and the TIR domains in *Ab*Tir are colored black. Enlarged view of the active sites of the TIR domains in TIR-APAZ and AbTir is indicated by dashed lines.

Supplementary information, Table S1

Table S1. Cryo-EM data collection, refinement and validation statistics

	SPARTA-gRNA-ssDNA
Data collection and processing	
Magnification	81,000
Voltage (kV)	300
Electron exposure (e^{-}/Å^2)	50.0
Defocus range (μm)	−1.0 to −2.5
Pixel size (Å)	1.0773
Symmetry imposed	C1
Initial particle images (no.)	5,034,199
Final particle images (no.)	832,006
Map resolution (Å)	2.84
FSC threshold	0.143
Map resolution range (Å)	4.80−2.84
Refinement	
Initial model used (PDB code)	N/A
Model resolution (Å)	2.84
FSC threshold	(0.143)
Model resolution range (Å)	4.80-2.84
Map sharpening *B* factor (Å^2)	−96.3
Model composition	
Nonhydrogen atoms	31820
Protein residues	3488
Nucleotide	163
B factors (Å^2)	
Protein	59.13
Nucleotide	57.11
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.657
Validation	
MolProbity score	2.57
Clashscore	12.75
Poor rotamers (%)	4.68
Ramachandran plot	
Favored (%)	93.07
Allowed (%)	6.90
Disallowed (%)	0.03