

Insights into the compact CRISPR–Cas9d system

Jie Yang^{1,2#}, Tongyao Wang^{1,2#}, Ying Huang^{3,4,5#}, Zhaoyi Long^{1#}, Xuzichao Li¹, Shuqin Zhang¹, Lingling Zhang¹, Zhikun Liu¹, Qian Zhang¹, Huabing Sun⁶, Minjie Zhang⁷, Hang Yin⁸, Zhongmin Liu^{3,4,5*}, Heng Zhang^{1,2*}

1. State Key Laboratory of Experimental Hematology, Tianjin Institute of Immunology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Key Laboratory of Immune Microenvironment and Disease (Ministry of Education), School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China.

2. Department of Biochemistry and Molecular Biology, Tianjin Key Laboratory of Cellular Homeostasis and Disease, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China.

3. Shenzhen Key Laboratory of Biomolecular Assembling and Regulation, School of Life Sciences, Southern University of Science and Technology Shenzhen 518055, Guangdong, China

4. Institute for Biological Electron Microscopy, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

5. Department of Immunology and Microbiology, School of Life Sciences, Southern University of Science and Technology, Shenzhen 518055, Guangdong, China

6. Tianjin Key Laboratory on Technologies Enabling Development of Clinical Therapeutics and Diagnostics, Department of Chemical Biology, School of Pharmacy, Tianjin Medical University, Tianjin, China

7. Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

8. Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China

These authors contributed equally

*Correspondence: liuzm@sustech.edu.cn (Z.L.), zhangheng134@gmail.com (H.Z.)

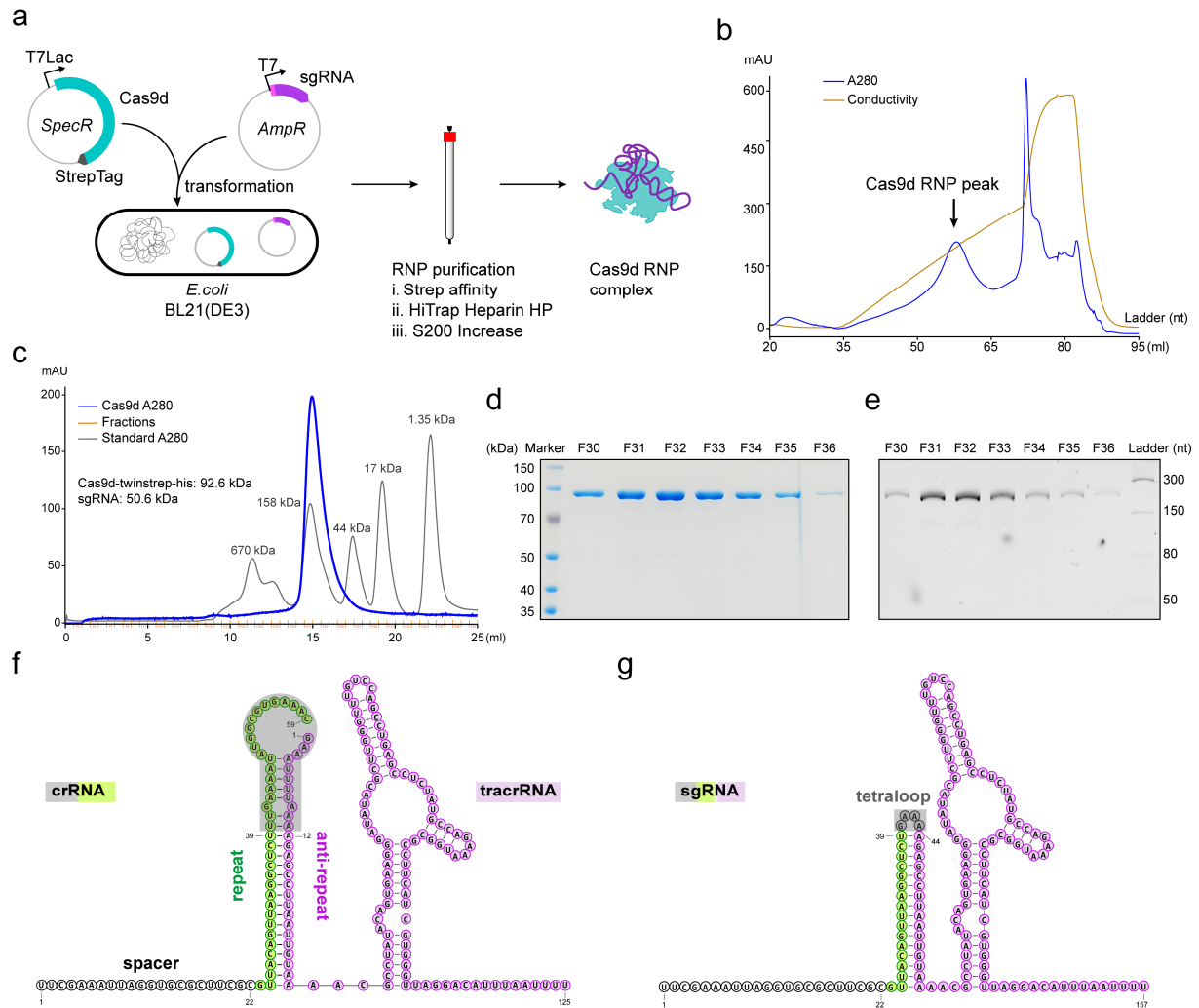


Figure S1. The purification of Cas9d RNP.

(a) A schematic illustrating the co-expression of Cas9d protein and sgRNA in *E. coli* and the subsequent purification process.

(b) Heparin affinity chromatography using a HiTrap Heparin HP column. The linear NaCl gradient was from 150 mM to 750 mM.

(c) Size-exclusion chromatography using an S200 Increase column. The protein standard is shown in gray.

(d–e) The SDS-PAGE (d) and Urea-PAGE (e) of gel-filtration fractions F30–F36 as indicated in (c).

(f–g) The crRNA and tracrRNA forms a hybrid dual-RNA guide (f) and can be trimmed at the repeat:anti-repeat region (shaded) to form a single-guide RNA (sgRNA) (g) for effective cleavage.

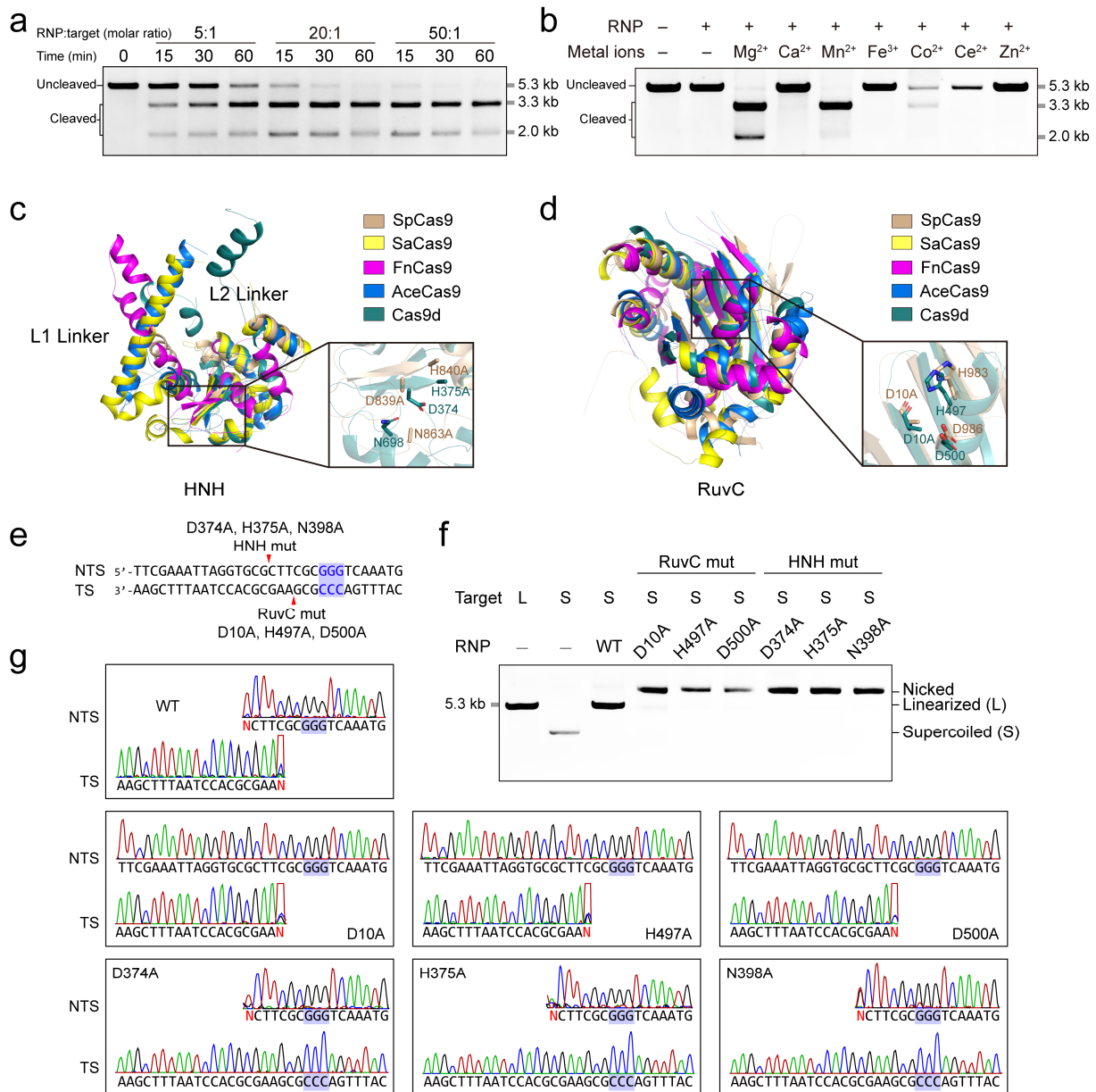


Figure S2. Nuclease activity of Cas9d.

(a) In vitro cleavage assays showed that WT Cas9d cleaves dsDNA target in time- and dose-dependent manners.

(b) In vitro cleavage assay was performed using a variety of divalent metal ions as indicated.

(c–d) Structural alignments of the conserved nuclease domains HNH (c) and RuvC (d) between Cas9d (green) and orthologs including Spy- (wheat, PDB ID: 6O0Z), Sau- (yellow, PDB ID: 5CZZ), Fno- (magenta, PDB ID: 5B2O), and Ace- (marine, PDB ID: 6WBR) Cas9s.

(e) Schematics of the cleavage sites in the targeting (TS) and non-targeting strands (NTS). The PAM was highlighted in blue. The mutated catalytic residues were listed.

(f–g) In vitro cleavage assays were performed using supercoiled plasmids as targets. The cleavage products were resolved on an agarose gel (f), and the major single bands were gel-purified and subjected to Sanger sequencing (g).

The gel images in (a), (b) and (f) are from three biological replicates.

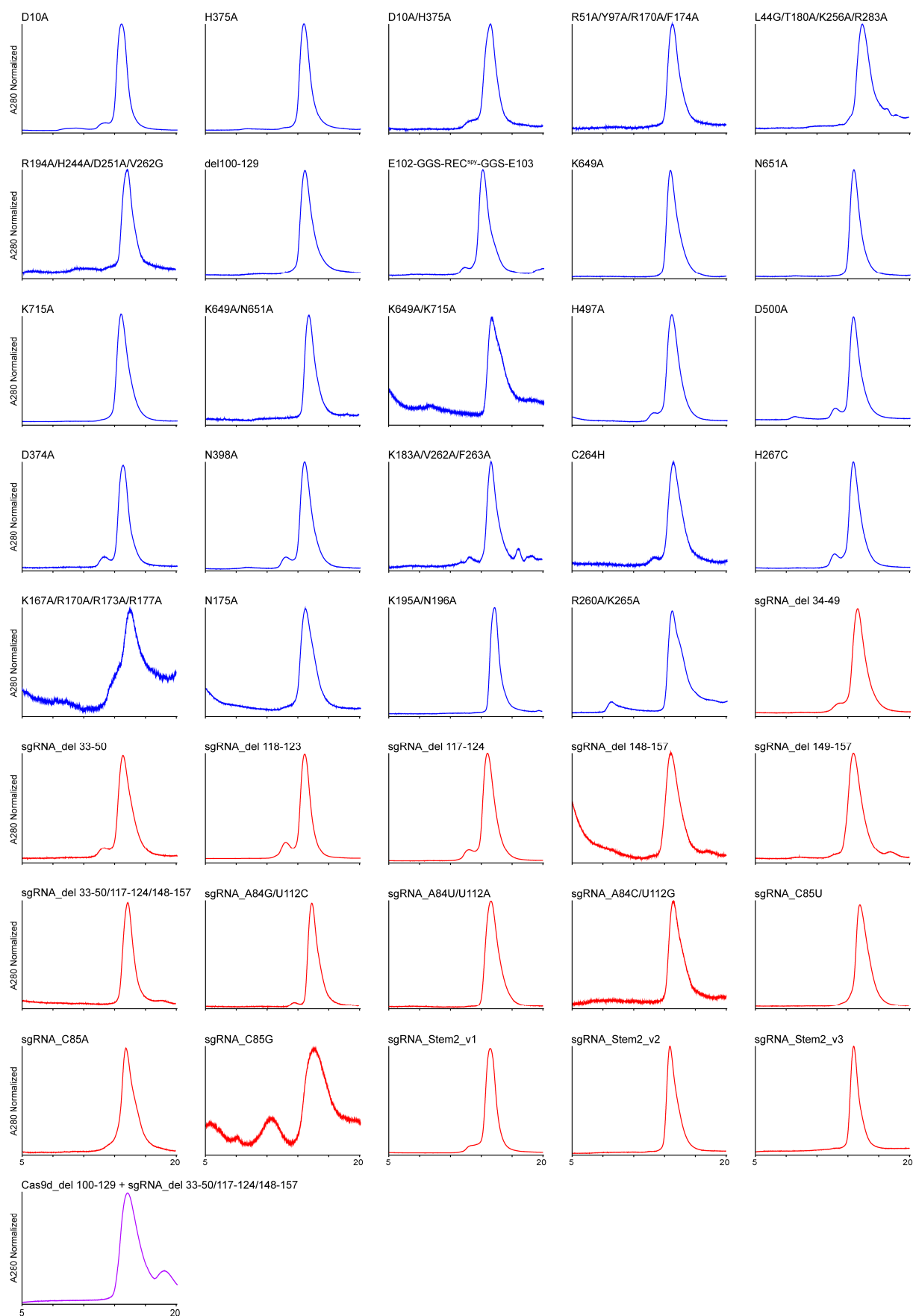


Figure S3. Size-exclusion chromatographies of all tested Cas9d RNP mutants.

The A280 profiles of Cas9d variants are shown as blue lines, sgRNA variants as red lines, and the combined mutant is shown as a purple line.

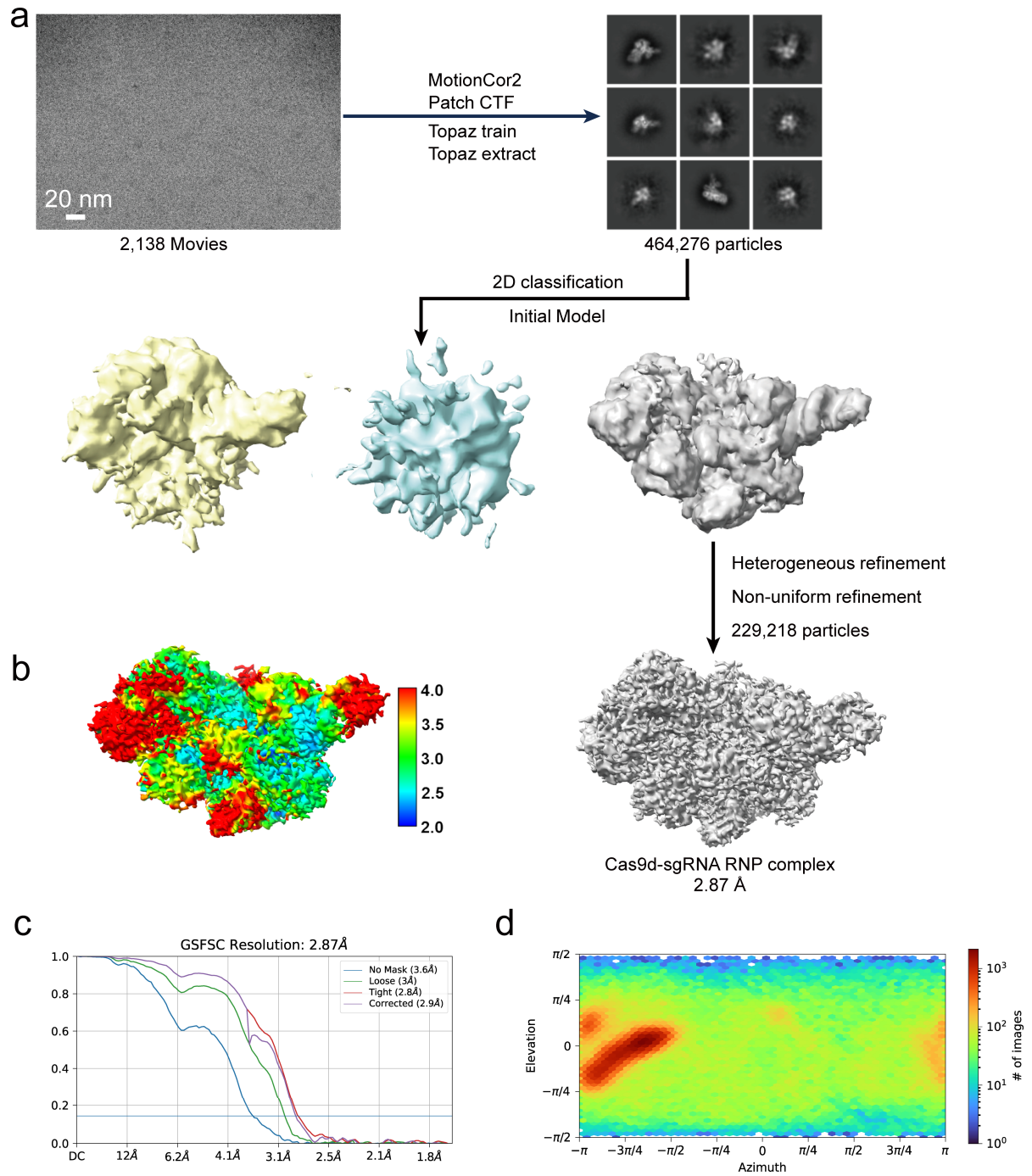


Figure S4. Cryo-EM analysis of Cas9d-sgRNA binary complex (target-free state).

- (a) Single-particle cryo-EM data processing pipeline with representative images of the motion corrected micrograph (scale bar, 20 nm), selected 2D class averages, and 3D classification.
- (b) Cryo-EM density map colored according to the local resolution.
- (c) Fourier shell correlation (FSC) curve at 0.143 cut-off determines an overall resolution at 2.87 Å.
- (d) Angular distribution of selected particles for 3D reconstruction.

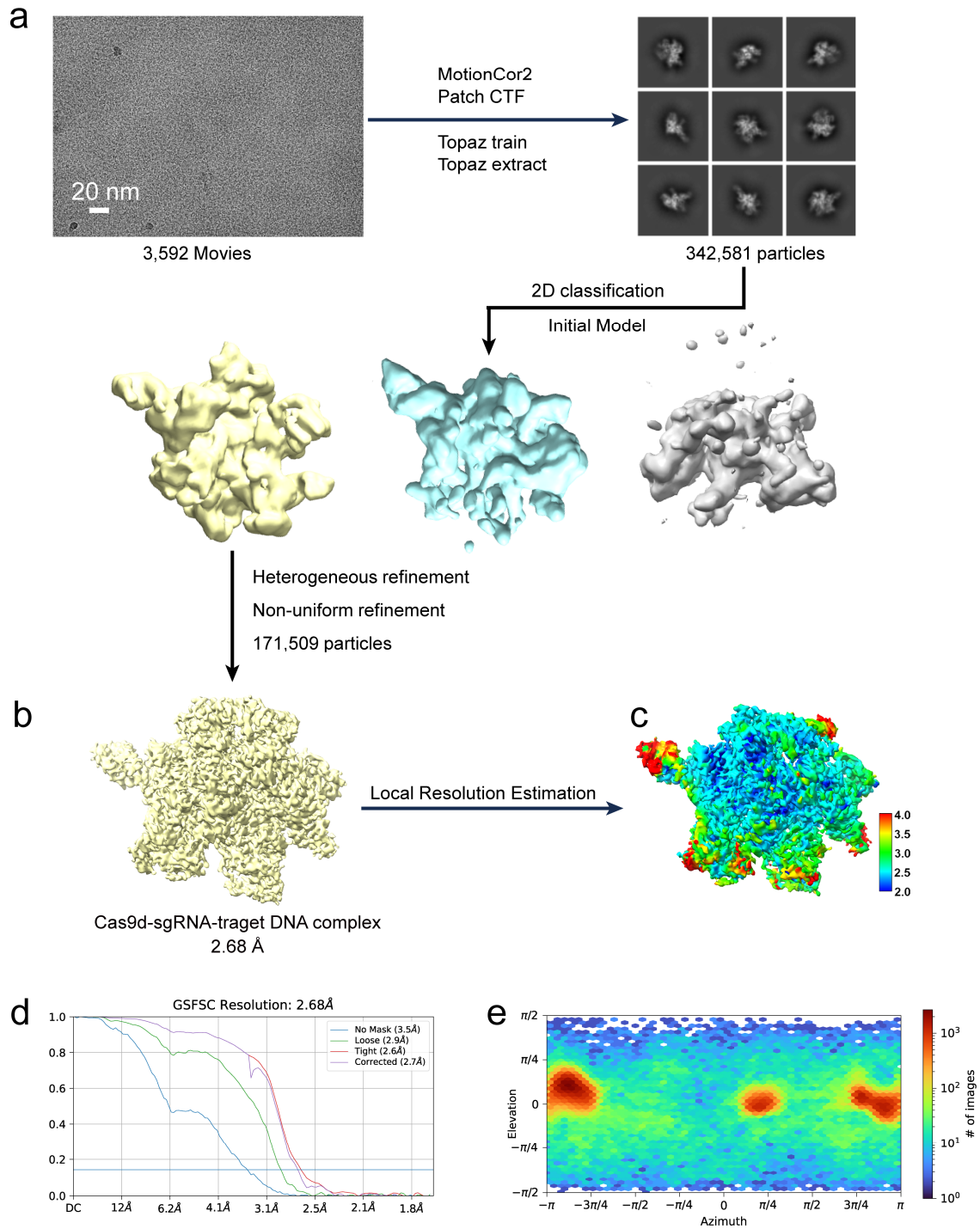


Figure S5. Cryo-EM analysis of Cas9d–sgRNA–target ternary complex (target-bound state).

- (a) Single-particle cryo-EM data processing pipeline with representative images of the motion corrected micrograph (scale bar, 20 nm), selected 2D class averages, and 3D classification.
- (b) Cryo-EM density map.
- (c) Cryo-EM density map colored according to the local resolution.
- (d) FSC plot determines an overall resolution at 2.68 Å.
- (e) Angular plot of particles for final reconstruction.

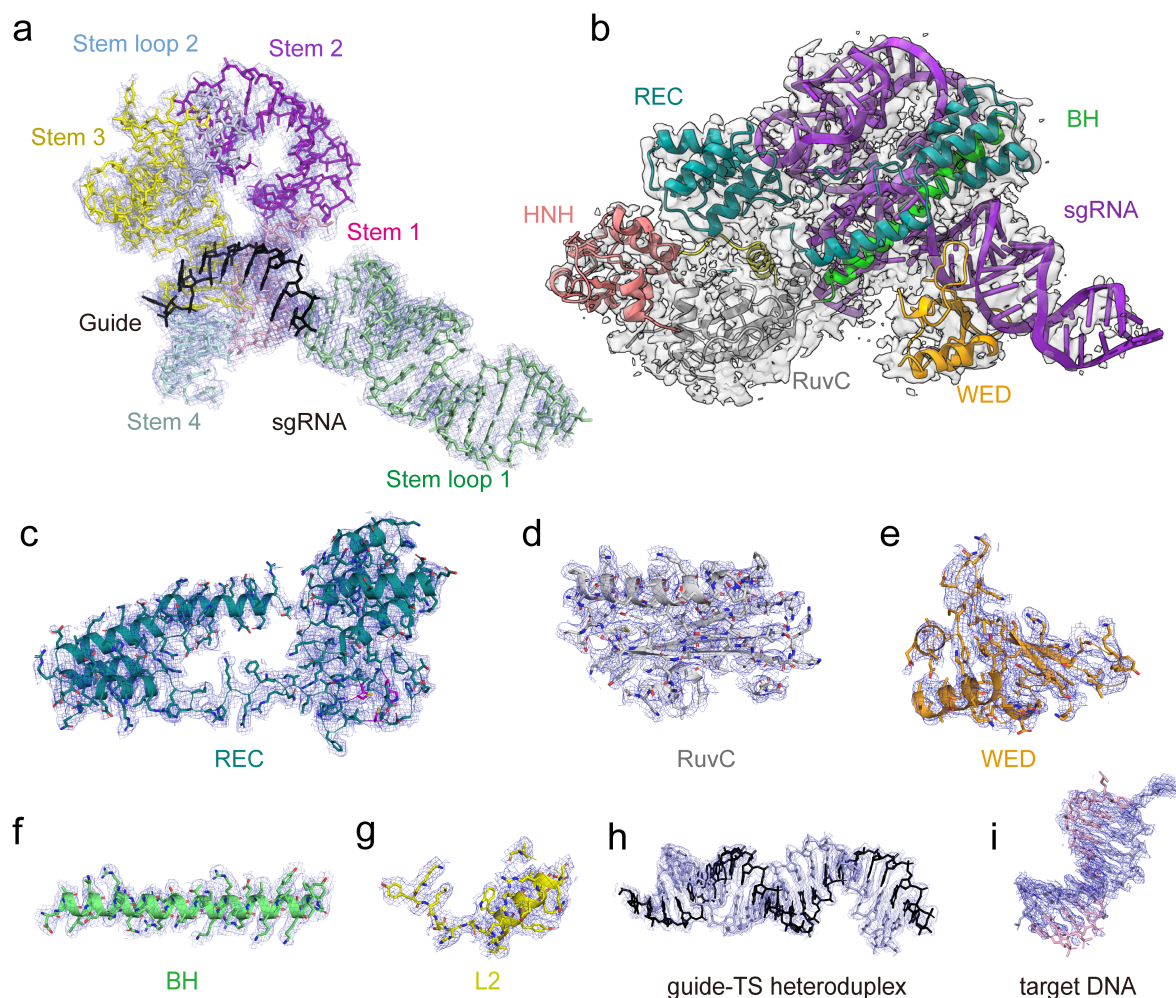


Figure S6. Cryo-EM density maps of Cas9d-sgRNA and Cas9d-sgRNA-target complexes.

(a) Cryo-EM density map and model for the sgRNA in Cas9d-sgRNA complex. The color scheme matches that in **Fig. 2**.

(b) Predicted model of HNH domain (salmon) docked into the cryo-EM density of Cas9d-sgRNA complex at a low threshold.

(c-g) Cryo-EM density maps and models for the REC (c), RuvC (d), WED (e), BH (f), and L2 linker (g) in Cas9d-sgRNA complex. The color scheme matches that in **Fig. 1**.

(h-i), Cryo-EM density maps and models for the guide-TS heteroduplex (h) and dsDNA substrate (i) in Cas9d-sgRNA-target ternary complex.

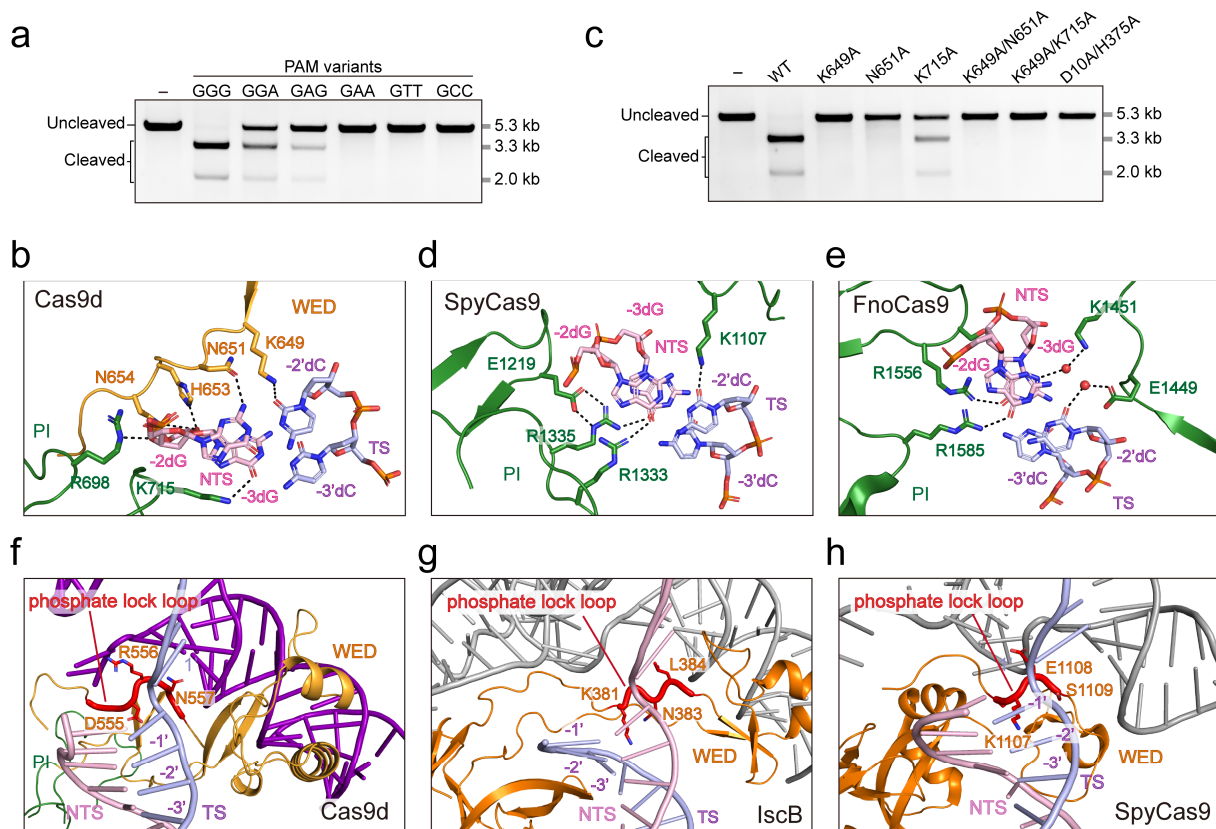


Figure S7. PAM recognition by Cas9d

(a) Cas9d recognizes NGG PAM. In vitro cleavage assays were performed using target DNA with various PAM sequences as indicated.

(b) Detailed interactions between PAM sequence and Cas9d.

(c) In vitro cleavage assays were performed using Cas9d mutants with alanine substitutions at the PAM-interacting residues. The gel images in (a) and (c) are from three biological replicates.

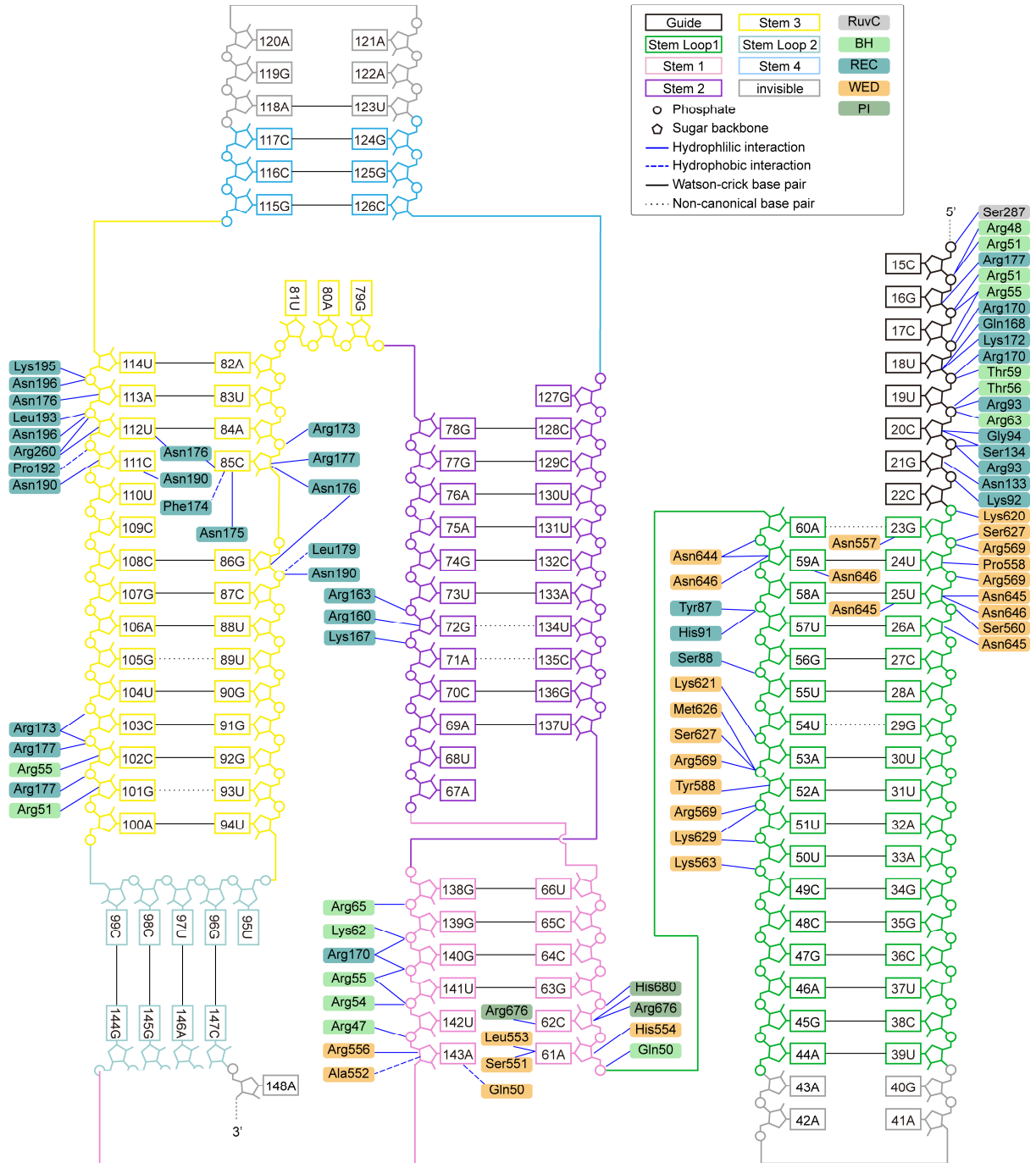
(d–e), Structure basis of NGG PAM recognition by SpyCas9 (d, PDB ID: 5F9R) and FnoCas9 (e, PDB ID: 5B2O).

For (b), (d) and (e), key interacting residues are shown as sticks. Unlike Cas9d which uses both the WED and PI domains for PAM recognition, SpyCas9 and FnoCas9 solely rely on their PI domains. Specifically, in SpyCas9 and the Cas9d, the -2dG^{NTS} and the -2'dC^{TS} are individually recognized through hydrogen bonding. In SpyCas9, Lys1335 forms a hydrogen bond with -2dG^{NTS}, while Lys1107 hydrogen-bonds with -2'dC^{TS}. In contrast, FnoCas9 only specifies -2dG^{NTS} through a direct hydrogen bond with Arg1585, while recognition of -2'dC^{TS} involves a hydrogen bond mediated by a water molecule, facilitated by Glu1449. In all three Cas9 systems, the -3G^{NTS} is specified by basic residues, while no interactions are observed for the recognition of the -3'C^{TS}. In SpyCas9, the Arg1335 residue, which recognizes -3dG^{NTS}, is stabilized by the acidic residue Glu1219. However, such stabilizing interactions are not observed in Cas9d or FnoCas9.

(f) A phosphate lock loop contacts the phosphate group of 1'dG^{TS} and unwinds the dsDNA target.

(g–h) Similar phosphate lock loops for PAM-proximal dsDNA unwinding in SpyCas9 (g, PDB ID: 6O0Z) and IscB systems (h, PDB ID: 7XHT).

a



b

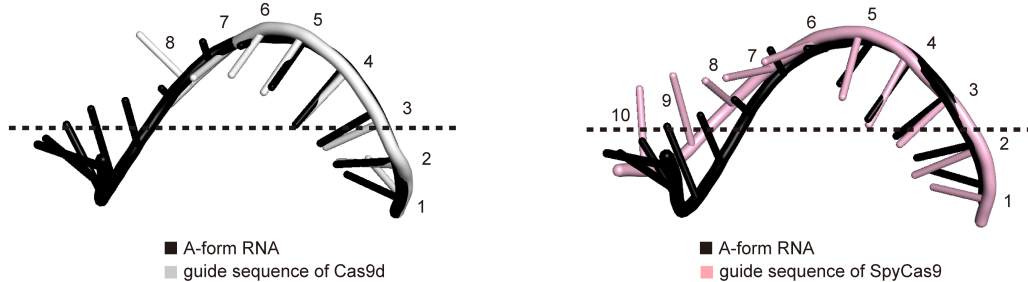
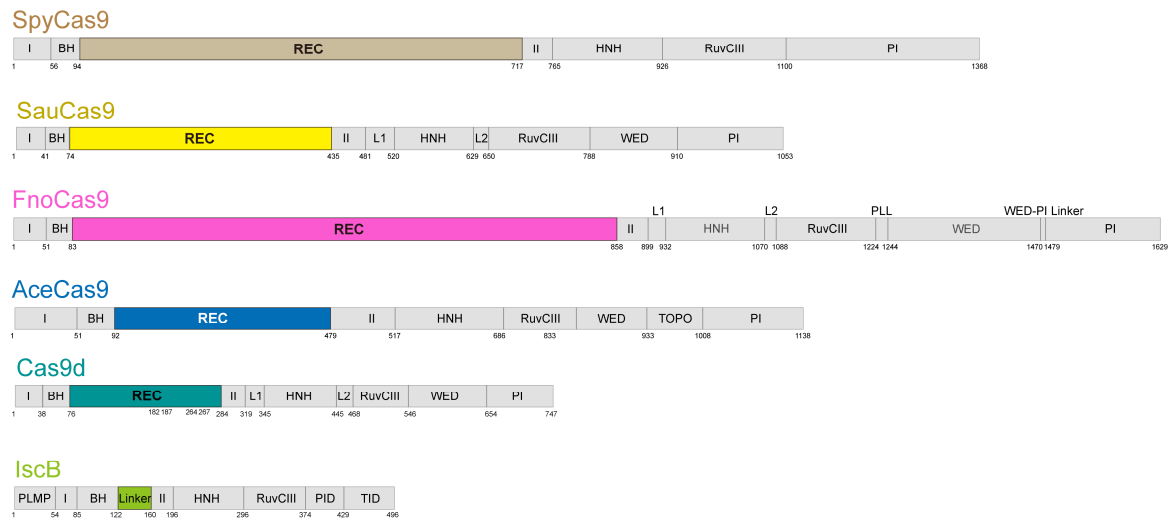


Figure S8. Detailed Cas9d-sgRNA interactions

(a) Interactions between Cas9d and the sgRNA scaffold and seed region. The color codes match that of **Fig. 1 and 2**.

(b) Superimposed A-form RNA molecule (black, PDB ID: 413D) and sgRNA seed regions of Cas9d (gray, current study) and SpyCas9 (light pink, PDB ID: 4ZT0). Helical axis is represented by black dotted line.

a



b

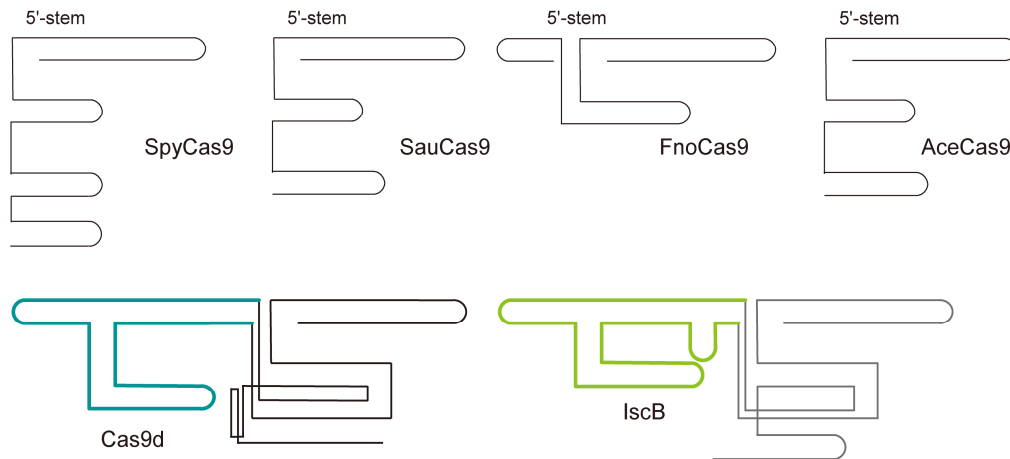


Figure S9. Comparison of REC and sgRNA scaffold in Cas9 variants and ancestral IscB.
 (a) Domain organization of type II-A, -B, -C and -D effector proteins, i.e., SpyCas9, FnoCas9, AceCas9, Cas9d and OgeuIscB. REC domains are highlighted in different colors.
 (b) Schematic diagram of sgRNA scaffold in Cas9 and ω RNA scaffold in IscB, RNA scaffolds for target recognition in Cas9d and IscB are highlighted with bolded and colored lines.

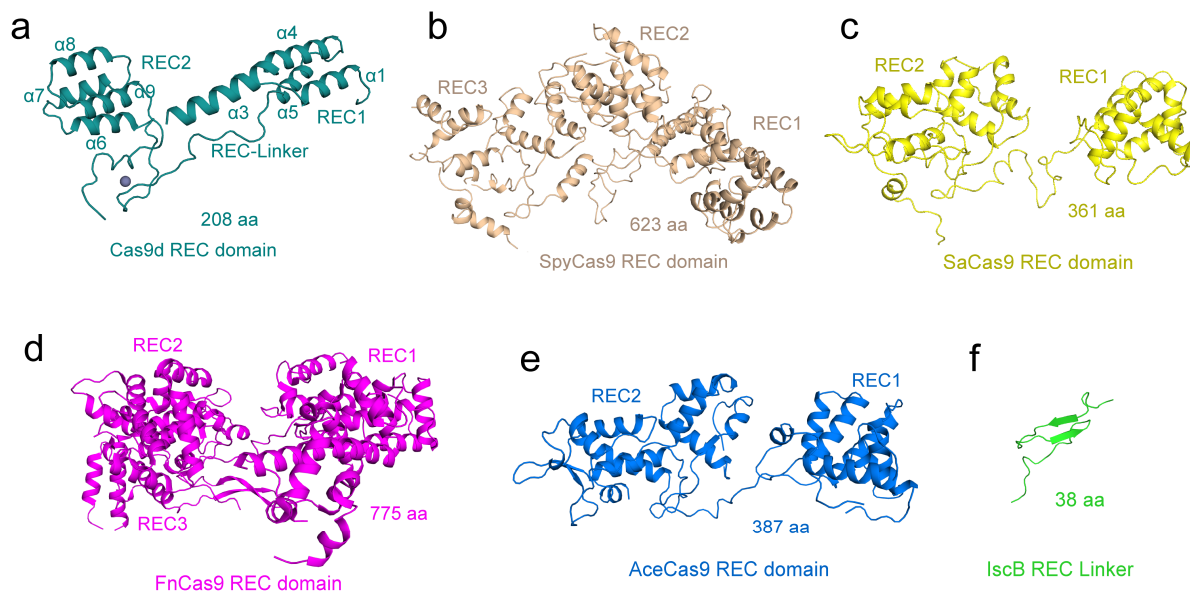


Figure S10. Comparison of the REC domains in the Cas9 orthologs and IscB.

(a–f) Structural comparison of the REC domains from Cas9d (deep teal), SpyCas9 (wheat, PDB ID:6O0Z), SaCas9 (yellow, PDB ID:5CZZ), FnCas9 (magenta, PDB ID:5B2O), AceCas9 (marine, PDB ID:6WBR), and OgeuIscB (lime, PDB ID:7XHT).

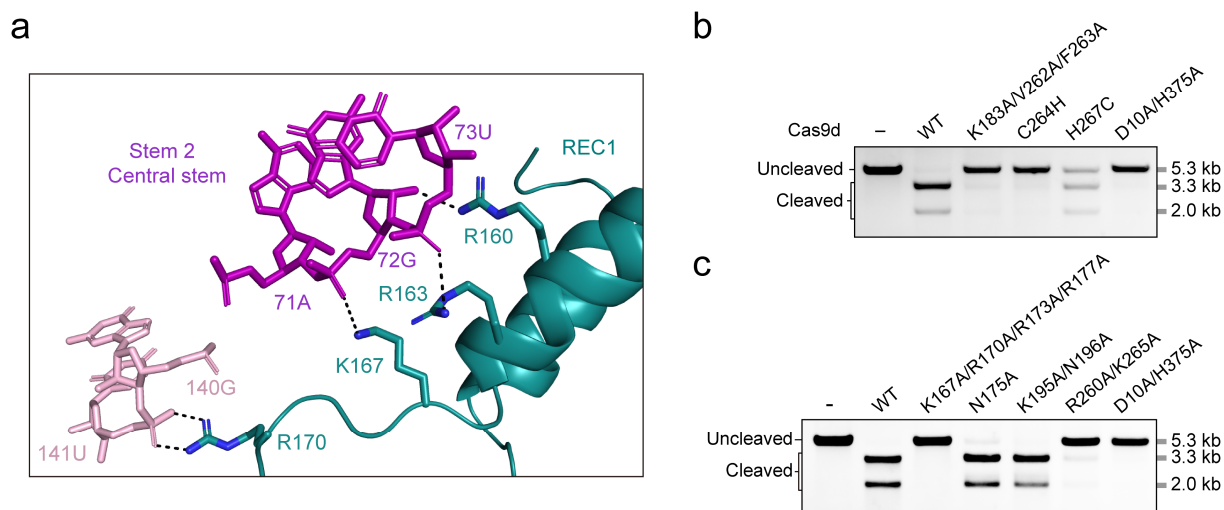


Figure S11. Zinc finger motif and REC-sgRNA interaction is critical for target DNA cleavage by Cas9d.

(a) Detailed insights into the interaction interface between the REC1 domain and the sgRNA scaffold. Key interacting residues and nucleotides are represented as sticks.

(b) In vitro cleavage assay of target DNA by WT and mutant Cas9d. Mutations were introduced into the cysteine and histidine residues within the zinc-containing motif.

(c) In vitro DNA cleavage assay with WT and mutant Cas9d. Residues on the REC-sgRNA interaction interface are mutated.

The gels in (b) and (c) are from three biological replicates. The cleavage percentage was calculated as described in Fig. 2.

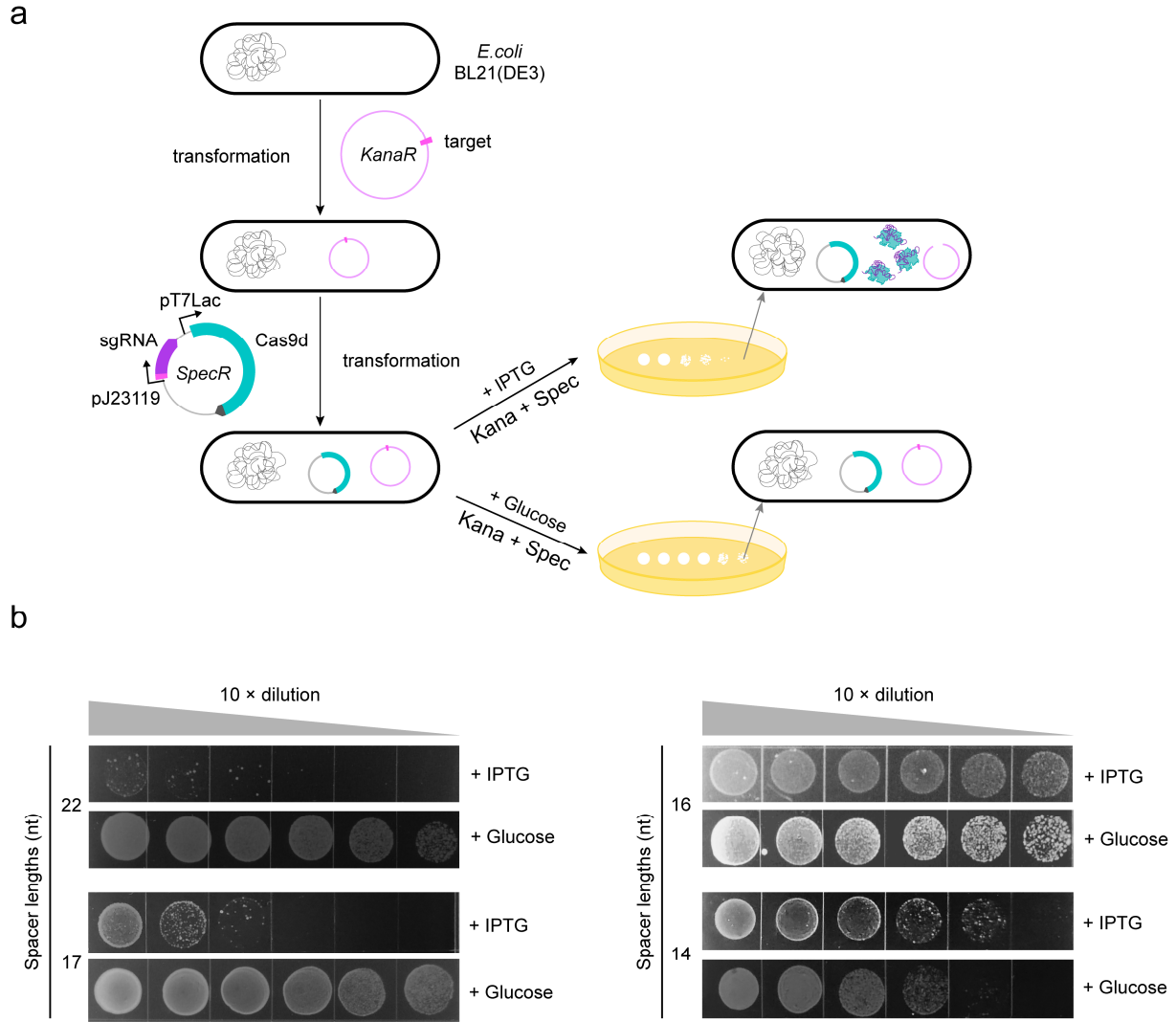


Figure S12. Cas9d-mediated plasmid clearance in *E. coli*.

(a) A schematic illustrating the experimental setup. *E. coli* BL21(DE3) cells were transformed with Kanamycin-resistant target plasmids and made competent. CDFDuet plasmids (Spectinomycin-resistant) encoding the expression cassette for Cas9d and sgRNA with varying spacer lengths were then transformed into the cells. The *E. coli* cells were split into two equal parts: one was induced with IPTG, and the other was grown in LB medium supplemented with glucose as a control. These cells were spotted onto IPTG- and glucose-containing agar plates with serial tenfold dilutions and selected with both antibiotics, respectively.

(b) Plasmid clearance assays for Cas9d with sgRNA of varying spacer lengths as indicated. The images are representative of three biological replicates.

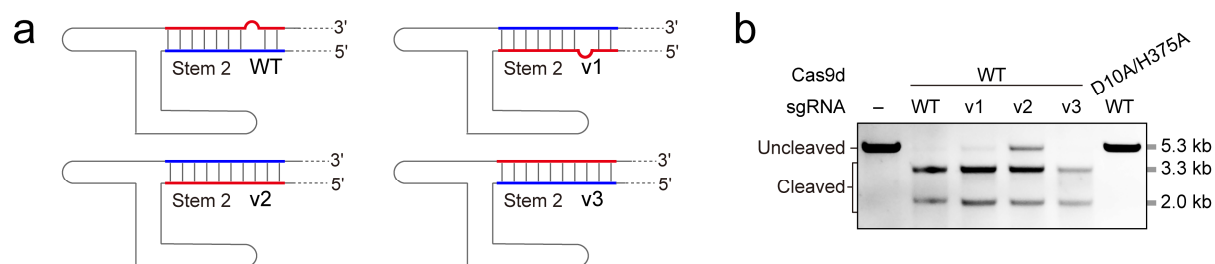


Figure S13. Stem 2 serves as the structural support of Cas9d.

(a) Schematic representation of Stem 2 variants, where the two strands of the stem were either swapped (v1), bulge-removed (v3), or both (v2).

(b) In vitro cleavage assays were conducted using the WT and mutant sgRNAs (v1, v2, and v3) in complex with WT Cas9d. The gel represents three biological replicates. The cleavage percentage was calculated as described in Fig. 2.

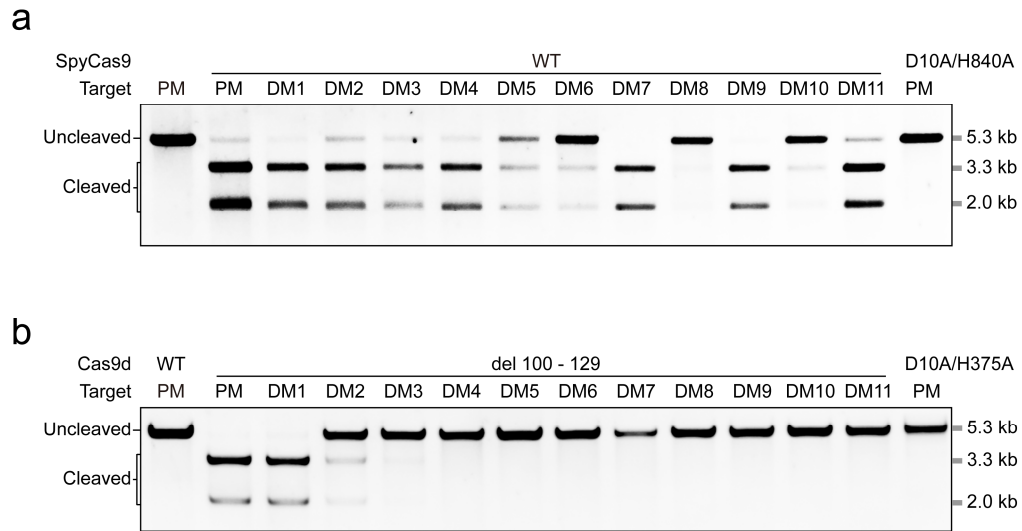


Figure S14. Double-mismatch tolerance profiles of SpyCas9 and the Cas9d deletion mutant.

(a–b) In vitro cleavage assays were conducted using WT SpyCas9 (a) and deletion mutant of Cas9d (b) to cleave a series of double mismatch-containing targets (DM1–DM11) as illustrated in **Fig. 5d**. Gels are representative from three independent experiments each. The cleavage percentage was determined as described in Fig. 2.

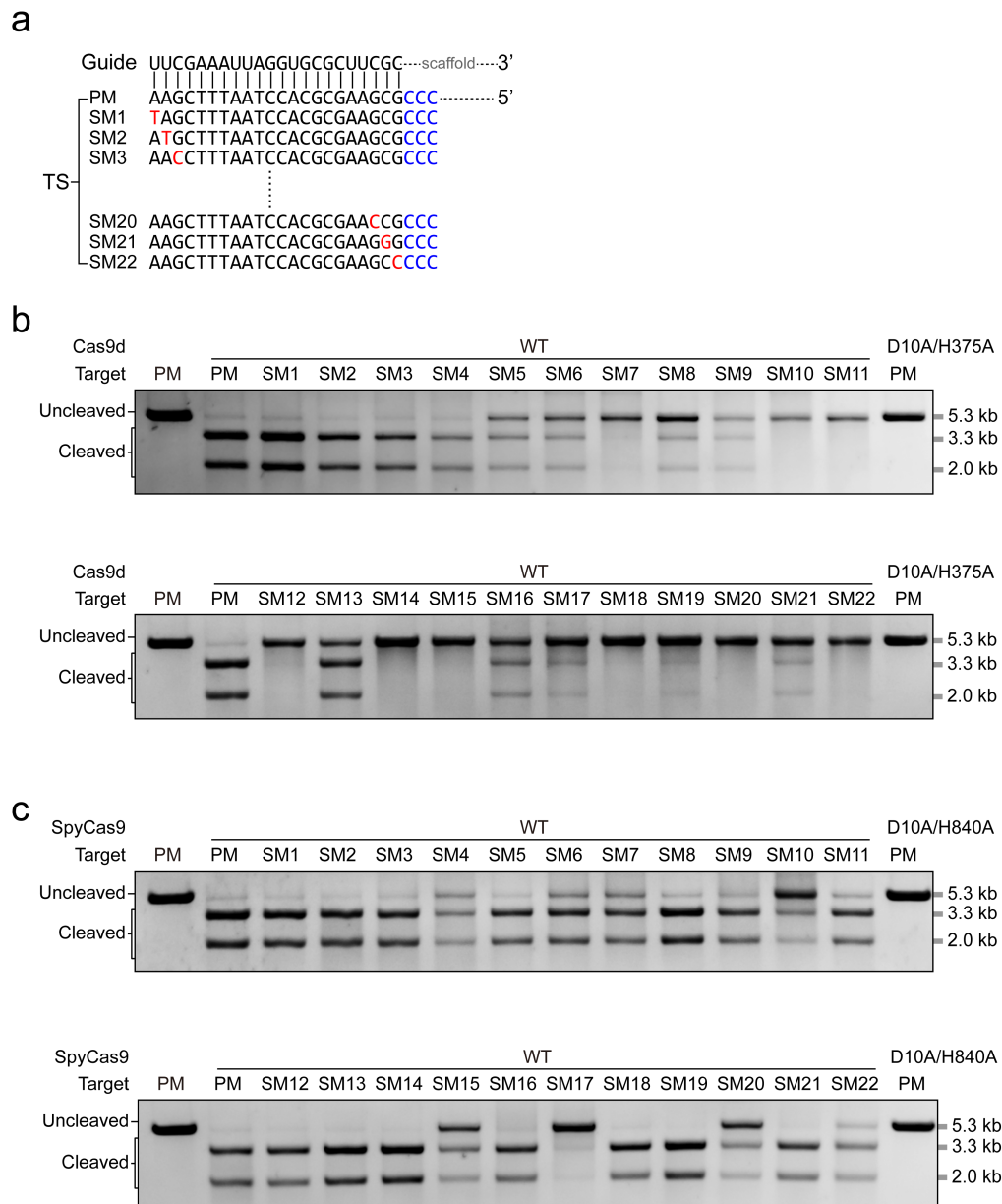


Figure S15. Single mismatch tolerance profiles of Cas9d and SpyCas9.

(a) Schematic representation of single mismatches on the TS (highlighted in red), with the complementary NTS omitted for clarity.

(b–c) In vitro cleavage assays were conducted using WT Cas9d (b) and SpyCas9 (c) to cleave a series of single mismatch-containing targets (SM1–SM22) as illustrated in (a). Gels are representative from three independent experiments each. The cleavage percentage was calculated as described in Fig. 2.

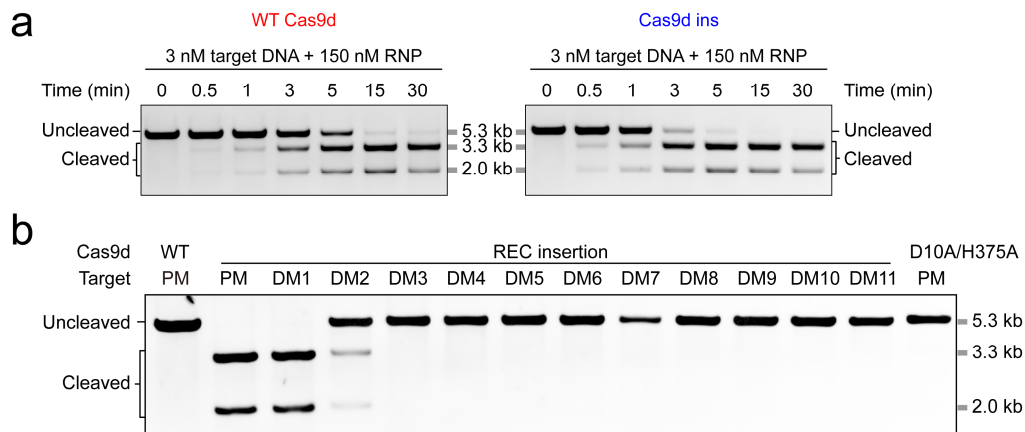


Figure S16. Nuclease activity and mismatch tolerance profile of the Cas9d insertional mutant.

(a) In vitro cleavage assays were conducted at various time points, comparing full-length WT Cas9d with the insertional mutant.

(b) In vitro cleavage assays were performed using Cas9d insertional mutant to cleave a series of mismatch-containing targets as described in **Fig. 5d**.

For (a) and (b), gels are representative from three independent experiments each. The cleavage percentage was measured as described in Fig. 2.

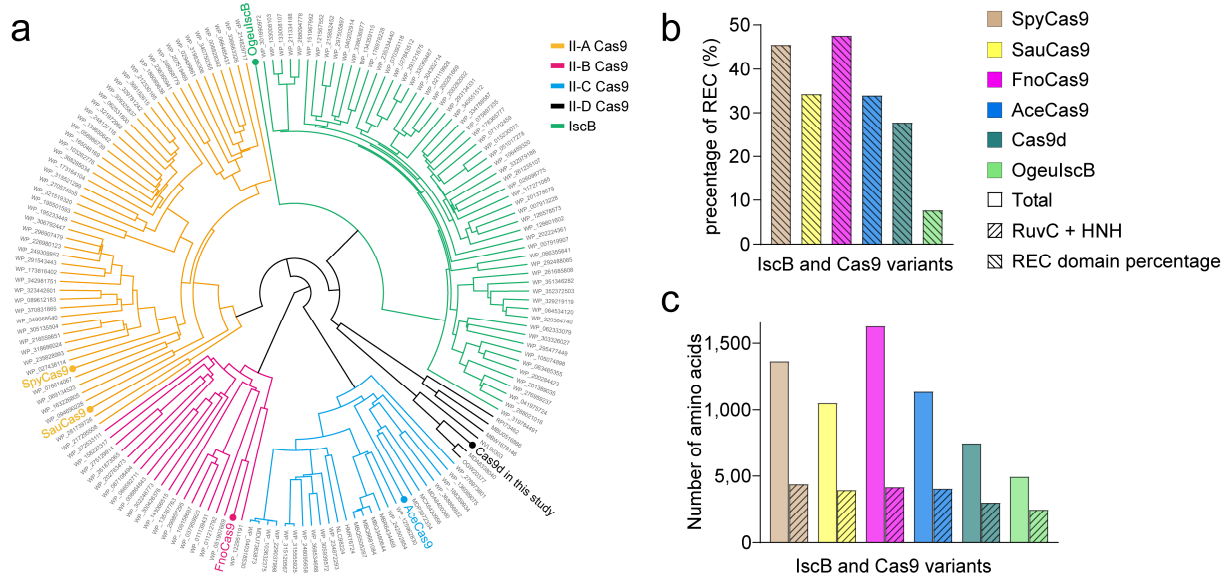


Figure S17. An evolutionary trend in Cas9 systems.

(a) Phylogenetic analysis of type II Cas9s and ancestral IscB. Representative genomes were curated from NCBI database. Obtained protein sequences of Cas9s and IscBs were aligned with MAFFT G-INS-i under default parameters, employing the BLOSUM62 scoring matrix. A phylogenetic tree was subsequently generated using the UPGMA method.

(b) Comparative analysis of REC domain prevalence in Cas9 proteins and IscB.

(c) Comparison of amino acid numbers in Cas9 proteins, IscB, and their nuclease domains.

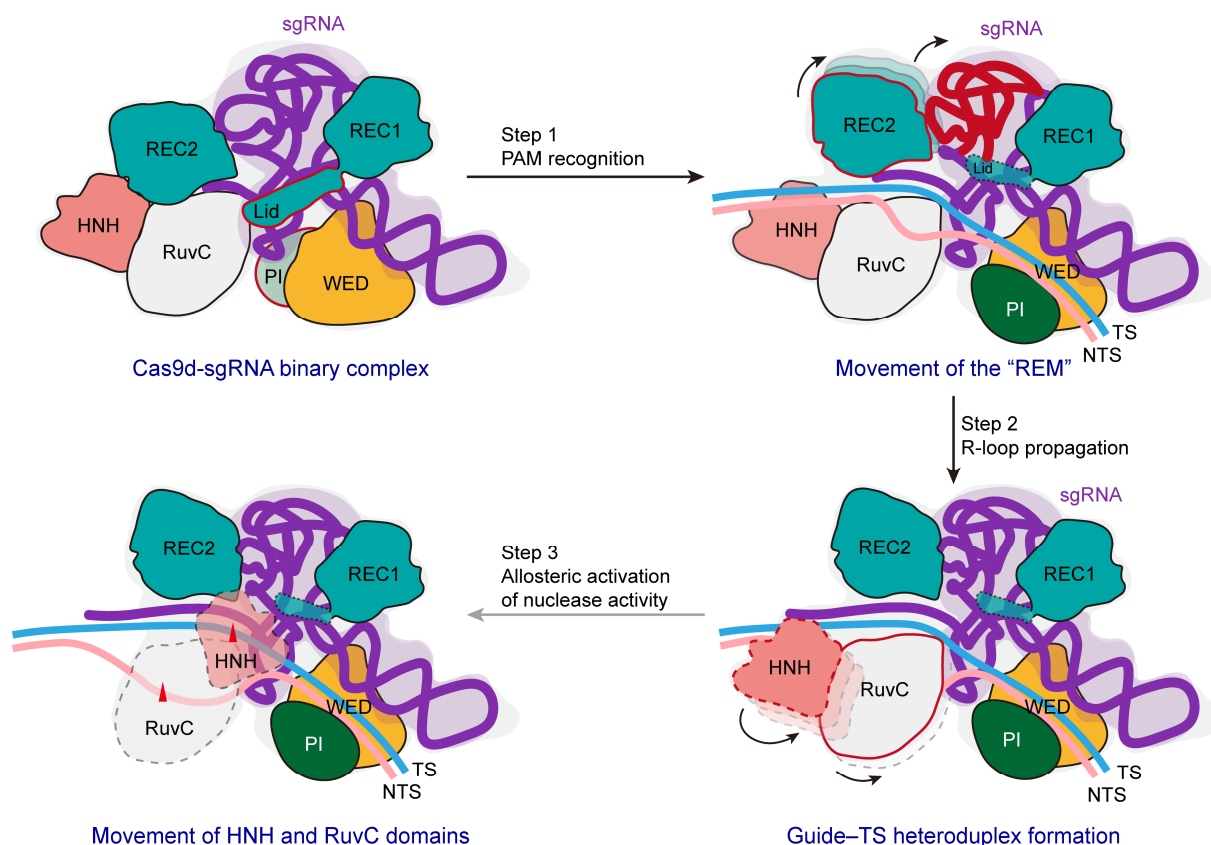


Figure S18. Schematic representation of target DNA recognition and cleavage by Cas9d. The color scheme matches that of **Fig. 1**. Steps 1, 2, and 3 illustrate (1) PAM recognition facilitated by PI and WED domains, (2) TS-guide heteroduplex formation coordinated by the hybrid REC domain and sgRNA stems, and (3) target triggered conformational changes of nuclease domains enabling DNA cleavage. The edges of domains or essential motifs undergoing significant movement in each step are highlighted in red, with proposed trajectories indicated by black arrows. The final panel depicts the predicted catalytic state of Cas9d.

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

	Cryo-EM structure of the Cas9d-sgRNA complex (EMDB-60379) (PDB 8ZQ9)	Cryo-EM structure of the Cas9d-sgRNA- target DNA complex (EMDB-60006) (PDB 8ZDR)
Data collection and processing		
Magnification	105,000×	105,000×
Voltage (kV)	300	300
Electron exposure (e-/Å ²)	50	50
Defocus range (μm)	-0.8 to -1.6	-0.8 to -1.6
Pixel size (Å)	0.83	0.83
Symmetry imposed	C1	C1
Initial particle images (no.)	464,276	342,581
Final particle images (no.)	229,218	171,509
Map resolution (Å)	2.87	2.68
FSC threshold	0.143	0.143
Map resolution range (Å)	2.0 to 4.0	2.0 to 4.0
Refinement		
Initial model used (PDB code)	AlphaFold	AlphaFold
Model resolution (Å)	3.0	2.9
FSC threshold	0.5	0.5
Model resolution range (Å)	2.7 to 3.6	2.6 to 2.9
Map sharpening <i>B</i> factor (Å ²)	69.4	82.3
Model composition		
Non-hydrogen atoms	6211	7440
Protein residues	433	446
Nucleotides	125	187
<i>B</i> factors (Å ²)		
Protein	77.21	38.01
Nucleotides	93.88	44.37
R.m.s. deviations		
Bond lengths (Å)	0.008	0.005
Bond angles (°)	1.226	0.667
Validation		
MolProbity score	2.28	1.70
Clashscore	12.58	7.72
Poor rotamers (%)	2.11	0.55
Ramachandran plot		
Favored (%)	93.53	95.95
Allowed (%)	6.47	4.05
Disallowed (%)	0	0