

# Insights into the compact CRISPR–Cas9d system

Received: 20 September 2024

Accepted: 22 February 2025

Published online: 12 March 2025



Jie Yang<sup>1,2,9</sup>, Tongyao Wang<sup>1,2,9</sup>, Ying Huang<sup>3,4,5,9</sup>, Zhaoyi Long<sup>1,9</sup>, Xuzichao Li<sup>1</sup>, Shuqin Zhang<sup>1</sup>, Lingling Zhang<sup>1</sup>, Zhikun Liu<sup>1</sup>, Qian Zhang<sup>1</sup>, Huabing Sun<sup>6</sup>, Minjie Zhang<sup>7</sup>, Hang Yin<sup>8</sup>, Zhongmin Liu<sup>3,4,5</sup>✉ & Heng Zhang<sup>1,2</sup>✉

Cas9d, the smallest known member of the Cas9 family, employs a compact domain architecture for effective target cleavage. However, the underlying mechanism remains unclear. Here, we present the cryo-EM structures of the Cas9d–sgRNA complex in both target-free and target-bound states. Biochemical assays elucidated the PAM recognition and DNA cleavage mechanisms of Cas9d. Structural comparisons revealed that at least 17 base pairs in the guide–target heteroduplex is required for nuclease activity. Beyond its typical role as an adaptor between Cas9 enzymes and targets, the sgRNA also provides structural support and functional regulation for Cas9d. A segment of the sgRNA scaffold interacts with the REC domain to form a functional target recognition module. Upon target binding, this module undergoes a coordinated conformational rearrangement, enabling heteroduplex propagation and facilitating nuclease activity. This hybrid functional module precisely monitors heteroduplex complementarity, resulting in a lower mismatch tolerance compared to SpyCas9. Moreover, structure-guided engineering in both the sgRNA and Cas9d protein led to a more compact Cas9 system with well-maintained nuclease activity. Altogether, our findings provide insights into the target recognition and cleavage mechanisms of Cas9d and shed light on the development of high-fidelity mini-CRISPR tools.

The Class 2 type II CRISPR–Cas is a dual-RNA guided defense system composed of a single effector Cas9 nuclease, CRISPR RNA (crRNA), and trans-activating crRNA (tracrRNA). These components form a ribonucleoprotein (RNP) complex to recognize and cleave invasive nucleic acids, such as bacteriophages and plasmids<sup>1,2</sup>. Targets containing a protospacer sequence complementary to the crRNA spacer, along with a downstream protospacer-adjacent motif (PAM), can be recognized and cleaved by the CRISPR system<sup>1,3,4</sup>. To streamline the application of CRISPR–Cas9 in genome editing, the dual-RNA guide was simplified by linking the crRNA and tracrRNA with a tetraloop, creating a chimeric single-guide RNA (sgRNA)<sup>3,4</sup>.

The Class 2 type II CRISPR–Cas9 system is further categorized into four subtypes: II-A, II-B, II-C, and II-D, based on the genomic

architecture of the CRISPR loci<sup>5–7</sup>. Among these, the type II-A Cas9 from *Streptococcus pyogenes* (SpyCas9) is the most widely used due to its high editing efficiency. However, its broader application is restricted by the payload of adeno-associated virus (AAV) delivery vectors, which are typically limited to 4.7 kb<sup>8</sup>, as well as by off-target editing effects<sup>9,10</sup>. To overcome these limitations, numerous Cas9 homologs with smaller molecular sizes have been identified through bioinformatic pipelines<sup>6,7,11–15</sup>. Emerging alternatives such as the Cas9 ancestor IscBs offer more compact genome editing tools. The effector proteins of these systems pair with extended guide RNAs or omega RNAs to execute robust nuclease activity<sup>7,16–19</sup>. Comprehensive biochemical and structural characterization, along with engineering efforts of Cas9 and IscB variants, have led to the

A full list of affiliations appears at the end of the paper. ✉ e-mail: [liuzm@sustech.edu.cn](mailto:liuzm@sustech.edu.cn); [zhangheng134@gmail.com](mailto:zhangheng134@gmail.com)

development of versatile genome-editing tools, advancing biological and biomedical research<sup>9,19–26</sup>.

A type II-D Cas9 ortholog, the newly identified MG34-1 Cas9 (hereafter referred to as Cas9d) from *Deltaproteobacteria* is 747 amino acids in size<sup>6</sup>, smaller than SpyCas9 (1368 amino acids)<sup>3</sup> but larger than IscBs (typically ~500 amino acids)<sup>7</sup>. The sgRNA scaffold used by Cas9d (135 nt) is longer than that of SpyCas9 (76 nt) but shorter than the omega RNAs used by IscBs (~200 nt). Importantly, Cas9d's expression cassette, including the promoters and terminators for both the protein and the sgRNA, is less than 3.4 kb, making it suitable for single AAV delivery. Cas9d has been demonstrated to exhibit double-stranded (ds)DNA cleavage activity in vitro, plasmid clearance in *E. coli*, and base editing in human cells with an efficiency up to 22%<sup>6,27</sup>. However, despite its potential in genome editing, the mechanisms by which Cas9d sustains its nuclease activity remain elusive. Further comparison of Cas9d with other Cas9 orthologs and IscBs, focusing on their structural and biochemical properties, could offer valuable insights for the development of more compact and precise genome editing tools.

In this study, we characterized Cas9d as a high-fidelity nuclease that cleaves dsDNA with a downstream NGG PAM, resolved its cryo-EM structures in target-free and target-bound states, and identified a novel functional module for target engagement, comprising its REC domain and a segment of the sgRNA scaffold. We term this distinct

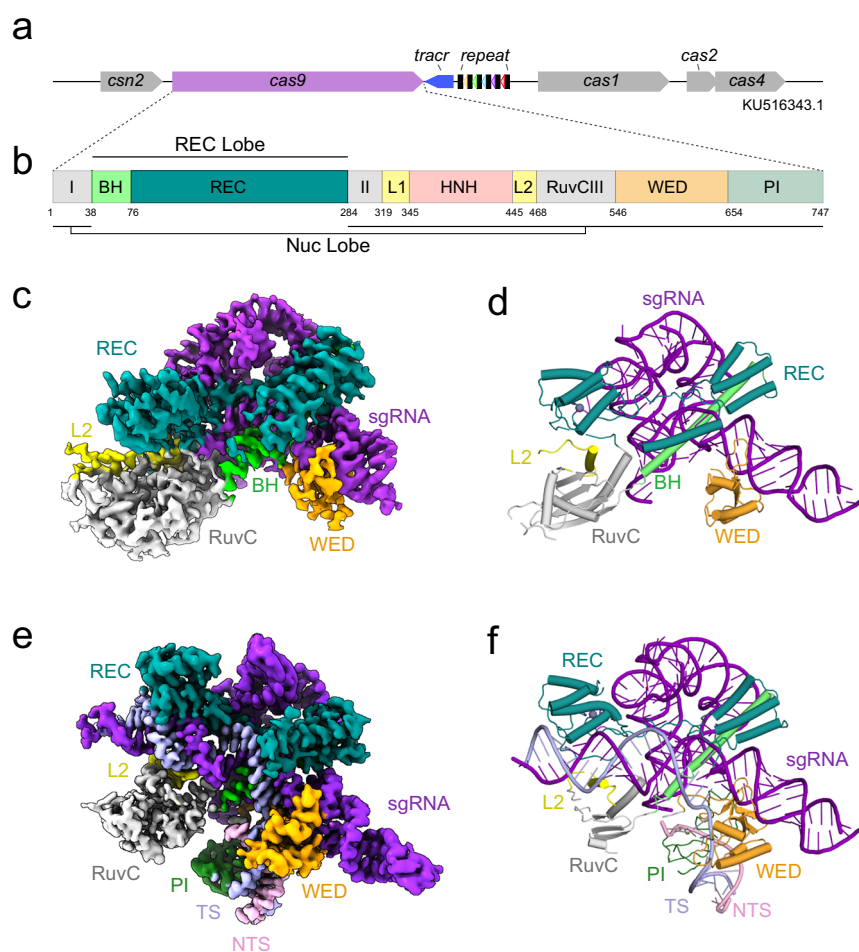
protein–RNA hybrid module the “REM” (RNA-coordinated target Engagement Module), which forms the structural basis for Cas9d's nuclease activity and is essential for target recognition and mismatch discrimination. Leveraging on these insights, we engineered a more compact version of Cas9d system by optimizing both its protein and sgRNA components, laying the groundwork for future engineering of Cas9d and the advancement of genome-editing tools.

## Results

### Biochemical characterization of Cas9d system

We co-expressed the 747 amino-acid Cas9d protein from *Deltaproteobacteria* with a 157-nt single-guide (sg)RNA, a fusion of the natural crRNA and tracrRNA linked by a GAAA tetraloop (Fig. S1). The co-localization of Cas9 with the CRISPR array and the adaptation machinery Cas1–Cas2 (Fig. 1a) suggests that the Cas9d system may be an active defense system<sup>6</sup>. Structural predictions using AlphaFold2<sup>28</sup>, along with structure and sequence alignments with other Cas9 orthologs<sup>25,29–31</sup>, indicate that Cas9d contains a bridge helix (BH), a target recognition domain (REC), a wedge domain (WED), a PAM interacting domain (PI), and two nuclease domains: the split RuvC domain (I, II, III) and the HNH domain (Fig. 1b).

Biochemical analyses showed that Cas9d cleaves DNA targets in dose- and time-dependent manners, and requires magnesium ions for



**Fig. 1 | Overall structure of the Cas9d–sgRNA and Cas9d–sgRNA–target complex. a** Genomic architecture of the Cas9d-containing CRISPR locus (accession: KU516343.1). **b** Domain organization of Cas9d. BH, bridge helix; L1 and L2, Linker 1 and 2; REC, target recognition domain; PI, PAM interacting domain; WED, wedge domain. RuvC is omitted for I and II due to the space limitation. **c, d** Cryo-EM density map (**c**) and atomic model (**d**) of the Cas9d–sgRNA complex. The BH and

REC domain in the REC lobe are illustrated in lime and deep teal, respectively. The WED, PI, HNH and RuvC domains in NUC lobe are colored orange, forest green, salmon and grey, respectively. L2 linker is colored in yellow. **e, f** Cryo-EM density map (**e**) and atomic model (**f**) of the Cas9d–sgRNA–target complex, using the same color scheme as in (**c, d**). The target strand (TS) and non-target strand (NTS) are shown in light blue and pink, respectively.

nuclease activity (Fig. S2a–b). Structural alignments to other type II Cas9 orthologs revealed highly conserved folding patterns and catalytic pockets in both nuclease domains (Fig. S2c–d). Alanine substitutions at the conserved catalytic residues<sup>32</sup> in either nuclease domain produced nicks in the supercoiled target plasmids (Figs. S2e–g and S3). Sequencing of the nicked products revealed that HNH cleaves the target strand (TS) three nucleotides upstream of the PAM, while RuvC cleaves the non-target strand (NTS) six nucleotides upstream, generating sticky ends with three-nucleotide 5' overhangs (Fig. S2e, g).

### Cryo-EM structures of Cas9d system

Using cryo-electron microscopy (cryo-EM), we determined the structure of the Cas9d–sgRNA ribonucleoprotein (RNP) complex at a resolution of 2.87 Å (Figs. 1c–d and S4). Additionally, by assembling the catalytically inactive (D10A/H375A) complex with a 45-bp target DNA containing a 5'-GGG-3' PAM, we resolved the cryo-EM structure of the Cas9d–sgRNA–target DNA ternary complex at a resolution of 2.68 Å (Figs. 1e–f and S5).

In the RNP complex, nearly all nucleotides of the sgRNA were reliably resolved (Figs. 1d and S6a). The Cas9d protein adopts a typical bilobed architecture with a domain organization similar to those of the other reported Cas9 orthologs (Figs. 1b and S6). The BH domain, most of the REC domain and the RuvC domain could be clearly traced (Figs. 1c–d and S6b–d). The HNH domain could only be docked into the EM density at a low threshold (Fig. S6b). The entire PI domain and a part of the WED domain (aa 557–633) were missing in the cryo-EM model, likely due to their flexibility (Figs. 1c–d and S6b, e).

In the target DNA-bound state, a 22-bp heteroduplex comprising the guide and the TS was reliably modeled (Figs. 1e–f and S6h). The NTS strand was absent, except for the PAM and the downstream dsDNA region, which were clearly defined in the map (Fig. S6i). The BH domain and majority of the REC domain were well resolved. The WED domain of Cas9d, stabilized by the target, was well-traced, and part of the PI domain (aa 670–686, 695–701, 713–721, 734–743) surrounding the PAM binding site was also built in the EM model. However, only a portion of the RuvC domain could be modeled and the entire HNH domain was missing at this state (Fig. 1e–f).

### Structural basis of PAM recognition

Cas9d cleaves dsDNA targets with an NGG PAM located 3' downstream of the protospacer<sup>6</sup>. Structural and biochemical analyses revealed the basis of PAM recognition (Fig. S7a). In the cryo-EM structure of the target-bound state, the PAM region is accommodated by both the WED and PI domains (Fig. S7b), distinguishing Cas9d from SpyCas9 and *Francisella novicida* (Fno) Cas9, which rely solely on the PI domain for NGG PAM recognition<sup>25,33</sup>. Investigation of the PAM-interacting interface revealed that Asn651 and Lys649 in the WED domain are hydrogen-bonded with -2dG<sup>NTS</sup> and -2'dC<sup>TS</sup>, respectively. Lys715 in the PI domain also forms a hydrogen bond with -3dG<sup>NTS</sup> (Fig. S7b). These interactions typically specify a dC:dG base pair at positions (-2) and (-3), indicating that target recognition by the Cas9d system is initiated through the precise recognition of the downstream NGG PAM. Moreover, Arg698 in the PI domain, along with Asn654 and His653 in the WED domain, interact with the phosphate backbone of the PAM region on the NTS through electrostatic interactions, thereby stabilizing the dsDNA target (Fig. S7b). Consistent with these structural observations, alanine substitution of these base-interacting residues completely or partially abolished target cleavage, demonstrating that Asn651, Lys649 and Lys715 are key residues in PAM recognition (Fig. S7c). Interestingly, the wild-type (WT) Cas9d exhibits activity toward the GAG and GGA PAMs (Fig. S7a), and the K715A mutant retains some cleavage activity, suggesting the potential for engineering PAM-relaxed variants. Notably, in SpyCas9, the presence of extra stabilizing residues supporting the PAM-interacting residues imposes the most stringent

PAM requirement, whereas no such stabilizing residues are present in Cas9d or FnoCas9 (Fig. S7).

Immediately adjacent to the PAM, the phosphate group of 1'dG<sup>TS</sup> interacts with a phosphate lock loop (residues D555–N557, Fig. S7f). This loop is present in most of the reported Cas9 structures, and also the ancestor IscB (Fig. S7g–h), emphasizing its importance in DNA duplex unwinding and the guide–TS heteroduplex formation<sup>16,17,33</sup>.

### Structural features of the sgRNA

Distinct from the other type II members, an extended sgRNA is found to associate with a relatively small Cas9d protein. The sgRNA of Cas9d system consists of a 22-nt guide segment (1U–22 C) and a 135-nt RNA scaffold (Fig. 2a–b). Only eight nucleotides (15C–22 C) of the guide sequence were built in the RNP complex (Fig. 2a–b), whose backbone phosphate groups are anchored by a cluster of arginine residues in the BH domain (Fig. S8a). The anchored guide region adopts an A-form-like geometry, similar to the 10-nt seed region of the SpyCas9 guide<sup>34</sup>, suggesting a 7-nt seed region (16G–22 C) for target recognition (Fig. S8b).

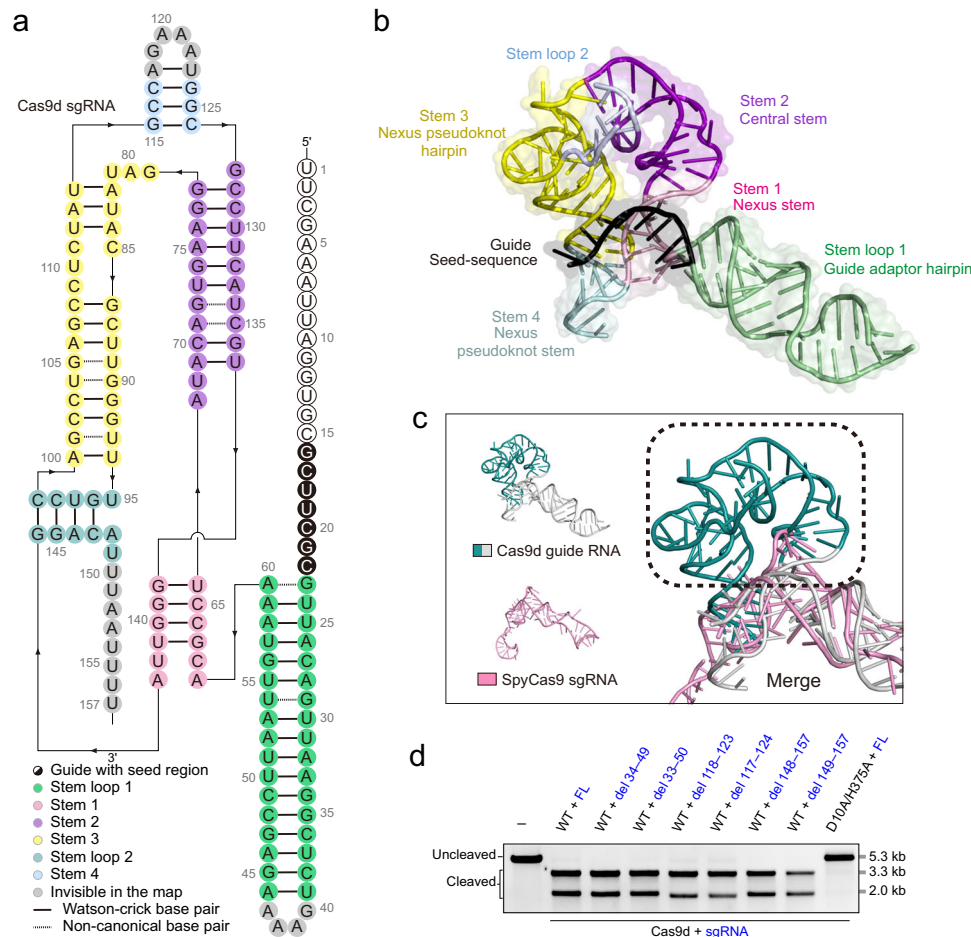
The sgRNA scaffold folds into a complex tertiary RNA structure that includes two stem loops (Stem Loops 1 and 2) and four stems (Stems 1–4). Downstream of the guide segment, the repeat:anti-repeat region forms a guide adaptor hairpin, termed as Stem Loop 1, comprising 17 base pairs (23G:60A–39U:44A). Following this region, the nucleotides fold upward to form a nexus stem, the Stem 1, harboring a 4-bp duplex (63A:138G–66U:141A). Extending from this nexus stem is a 17-bp central stem (Stem 2), which bifurcates into Stem Loop 2 and the nexus pseudoknot hairpin (Stem 3). Stem Loop 2, containing a 4-bp duplex (115G:126C–118A:123U), projects outward from the sgRNA scaffold. Conversely, the nexus pseudoknot hairpin turns downward, stacking with the central stem and nexus stem in a back-to-back arrangement. The nucleotides extending from the nexus pseudoknot hairpin flip out to form a GC-rich nexus pseudoknot stem (Stem 4).

Compared to SpyCas9, the Cas9d sgRNA retains conserved Stem Loop 1 and Stem 1 regions (Fig. 2c). The Stem Loop1 interacts with the REC and WED domains (Fig. S8a). However, the distal paired region beyond 33A:50U showed minimal interactions with the protein, and truncation in this region still maintained cleavage activity of Cas9d (Fig. 2d). Downstream of Stem Loop 1, Stem 1 forms extensive contacts with BH domain, making it crucial for the assembly with Cas9d (Fig. S8a).

A key difference between Cas9d and SpyCas9 is the presence of Stems 2 and 3 in sgRNA (Fig. 2c). The two stems are enveloped by the compact REC domain, suggesting that the sgRNA in Cas9d plays a role beyond simply serving as an adaptor between the enzyme and DNA targets. Stems 2 and 3 likely provide additional structural support and functional coordination with Cas9d (Figs. 2d and S8a). In contrast, Stem 4 and Stem Loop 2, which are also unique to Cas9d, show limited interactions with the REC domains. Stem 4 interacts with the RuvC domain, stabilizing the folding of the sgRNA scaffold. On the other hand, Stem Loop 2 exhibits minimal interactions with the protein, and part of it was not visible in the EM density. Partial truncation of Stem Loop 2 also resulted in maintained cleavage activity (Fig. 2d). The lack of resolved density for nucleotides after 150U indicates flexibility at the 3'-end of the sgRNA. Truncations beyond 148 A indeed had no impact on DNA cleavage (Fig. 2d).

### The RNA-coordinated target engagement module

The structural observations illustrated the extensive interactions between the extended sgRNA scaffold, particularly Stems 2 and 3, and the REC domain (Fig. S8a), indicating an intricate structural and functional coordination between them. The compact REC domain (208 aa) in Cas9d consists of REC1 and REC2 parts that flank the sgRNA scaffold on opposite sides and are interconnected by a positively charged REC-linker that passes through the Stems 2 and 3 (Fig. 3a). In



**Fig. 2 | Structure of the Cas9d sgRNA. a, b** Schematic representation (**a**) and atomic model (**b**) of the Cas9d sgRNA. The guide sequence is depicted in black and white, with black-filled circles indicating the seed region. Stem loop 1, Stem 1, Stem 2, Stem 3, Stem loop 2, and Stem 4 are highlighted in green, pink, purple, yellow, teal, and light blue, respectively. Canonical Watson–Crick base pairs are illustrated with solid lines, while non-canonical base interactions are shown with dashed lines. Nucleic acids that are not visible in the cryo-EM map have been omitted from the atomic model. **c** Comparison of sgRNA scaffolds between Cas9d (teal and grey) and SpyCas9 (pink, PDB: 6O0Z). The sgRNA of Cas9d is shown in two parts: the grey

region resembles the sgRNA of SpyCas9, while the teal regions are absent in SpyCas9. **d** In vitro cleavage assays were performed using WT Cas9d in combination with various sgRNA truncations. The catalytically inactive mutant D10A/H375A served as a control. The sgRNAs were truncated at the nucleotide positions indicated, as shown in **a**. The representative gel from three biological replicates is shown. The molecular weight markers are labeled on the right side of the gel. The uncropped gel image and cleavage percentages are included in the source data. Source data are provided as a Source Data file.

contrast to the REC domains of SpyCas9 and other reported Cas9s<sup>25,29,31</sup>, the central region of the REC domain is occupied by Stems 2 and 3 of the sgRNA scaffold in Cas9d (Figs. 3b and S9–10). This indicates that, in Cas9d, the combination of REC domain with these two stems of the sgRNA may form a hybrid functional module involved in target engagement process, analogous to the REC domains in SpyCas9 or other Cas9 members (Fig. 3a, b).

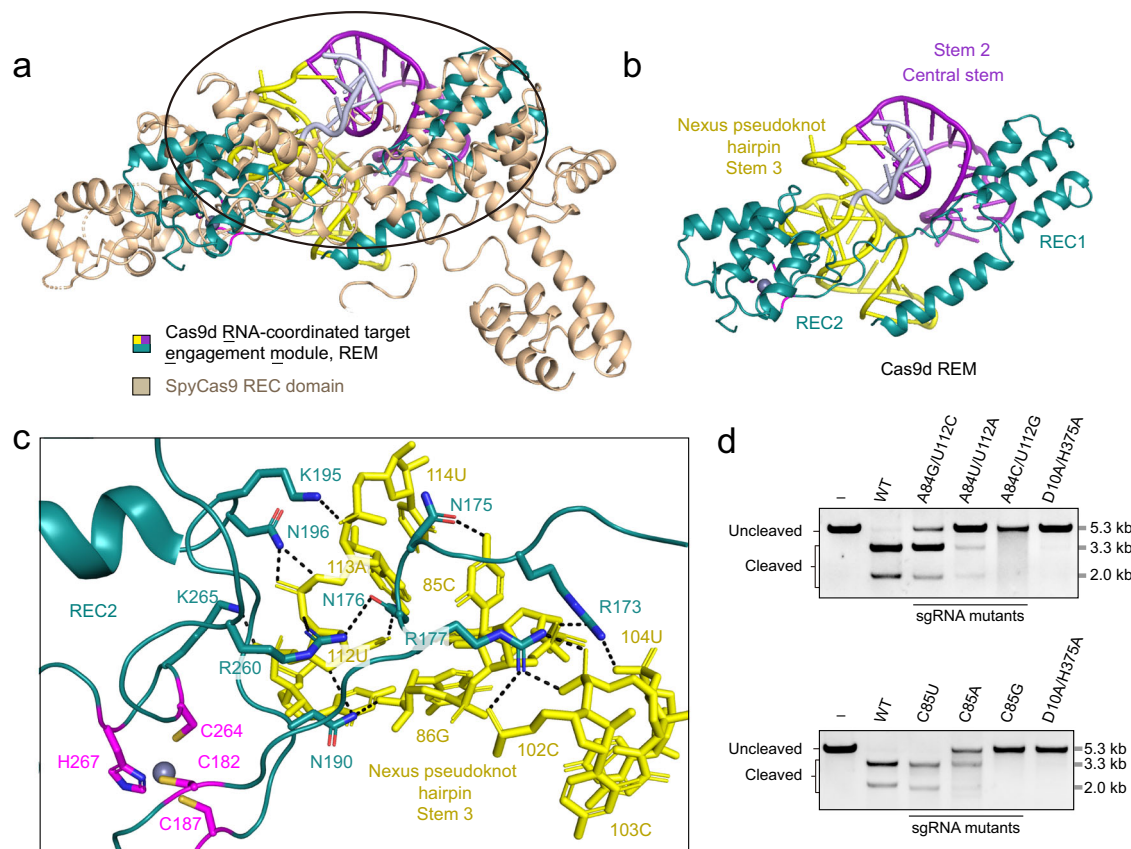
Structural analyses revealed that the REC1 region is stabilized by Stem 2 through non-specific hydrogen bonding between the sugar-phosphate backbone and the amino acid side chains (Figs. S8a and S11a). Specifically, two basic residues, Arg160 and Lys167, interact with the 2'-hydroxyl group and the non-bridging oxygen of the phosphate backbone at 72G on Stem 2 of the sgRNA, while Arg163 contacts the phosphate backbone at 73U (Fig. S11a).

A search against DALI server<sup>35</sup> using Cas9d REC2 yielded no significant hits, implying that the sgRNA binding mode in Cas9d is distinct from those in other Cas9 orthologs or the ancestral IscB members. Intriguingly, the REC2 part features a zinc-containing motif comprising three cysteine residues (Cys182, Cys187, and Cys264) and one histidine residue (His267) (Fig. 3c), positioned face-to-face with Stem 3, involving in the sgRNA binding. Members of the

CCCH family, commonly found in plants and animals<sup>36</sup>, typically harbor the consensus sequence C-X<sub>6-14</sub>-C-X<sub>4-5</sub>-C-X<sub>3</sub>-H<sup>37</sup>, functioning as RNA-binding motifs<sup>36,37</sup>. The atypical CCCH zinc-containing motif in Cas9d, although markedly different in sequence and structure from its eukaryotic counterparts, retains its RNA-binding function. This highlights the functional conservation of zinc-containing RNA binding motifs across both eukaryotic and prokaryotic systems, suggesting a possible ancient evolutionary origin for these motifs. Mutations in the cysteine or histidine of the zinc finger (C264H and H267C), as well as mutations in the residues stabilizing the zinc-finger motif (K183A/V262A/F263A), resulted in compromised or completely abolished cleavage activity of Cas9d (Fig. S11b). Additionally, two polar residues (Asn190, Asn196) and three positively charged residues (Lys195, Arg260, Lys265) within the zinc-finger motif form hydrogen bonds with the sugar-phosphate backbone of Stem 3 (Fig. 3c). The R260A/K265A mutants exhibited a completely abolished nuclease activity (Fig. S11c), indicating the critical role of intramolecular contacts between the REC2 and Stem 3.

More importantly, the REC linker intercalates into the minor groove of the A-form Stem 3, stabilizing the RNA structure through both non-specific and base-specific interactions (Fig. 3c). Asn176





**Fig. 3 | The hybrid REM of Cas9d comprising the REC domains and Stem 2 and stem 3 of the sgRNA. a** Structural alignment of the Cas9d (RNA-coordinated target Engagement Module, REM) with the SpyCas9 REC domain. **b** Close-up view of the Cas9d REM, as highlighted by the ellipse in **a**. **c** Interaction interface of REC domain

with sgRNA scaffold, with key residues and nucleotides involved in the intra-modular interactions shown as sticks. **d** In vitro cleavage assays were performed using WT Cas9d with specific mutations in the sgRNA. Each gel is representative of three independent experiments. Source data are provided as a Source Data file.

recognizes 84A:112U base pairing within Stem 3. Mutating 84A:112U to G:C, U:A, or C:G resulted in impaired target DNA cleavage (Fig. 3d, top). Additionally, Asn175 forms two hydrogen bonds with the flipped nucleotide 85C. A C-to-G transversion at position 85 completely abolished the target cleavage, while a C-to-U transition maintained the activity of Cas9d (Fig. 3d, bottom). Particularly, Arg173 simultaneously interacts with the backbone of 85C, 103C, and 104U, while Arg177 hydrogen bonds with the backbone of 102C and 103C. A combined mutation of K167A/R170A/R173A/R177A also led to a complete loss of activity (Fig. S11c). Collectively, the interactions between REC1 and Stem 2, REC2 and Stem 3, and the REC linker with Stem 3 are crucial for cleavage activity. We, therefore, proposed that these components form an RNA-coordinated target Engagement Module (REM), where the extensive intramolecular interactions between the REC domain and the two stems provide the structural foundation for the stability and nuclease activity of Cas9d.

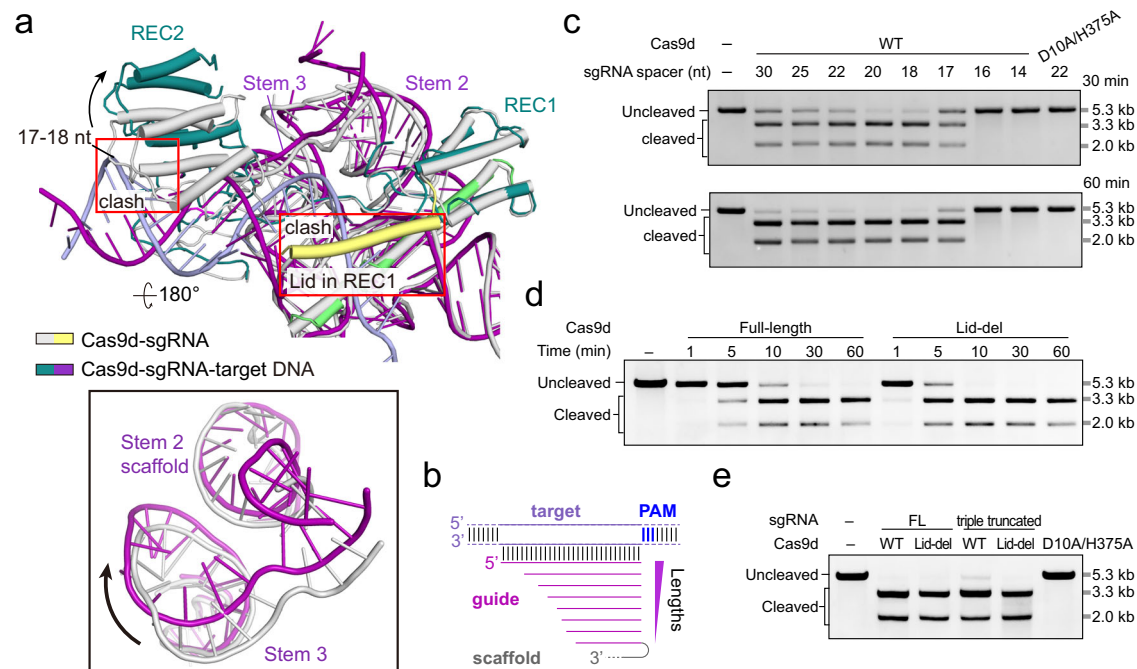
### Concerted movement of REC domain and sgRNA upon target binding

To further elucidate the target DNA engagement mechanism of REM, we compared the cryo-EM structures of the Cas9d system in both target-free and target-bound states. Upon target DNA binding, the REC2 domain shifts outward to accommodate the guide-TS heteroduplex, similar to the observations in canonical Cas9s. Meanwhile, Stem 3 of the sgRNA scaffold moves outward by ~6 Å, highlighting the coordinated rearrangement of both Stem 3 and REC2 in target engagement (Fig. 4a). Disrupting interactions between REC2 and Stem 3 would impede their synergistic movement, impair guide-TS engagement, and lead to Cas9d inactivation (Fig. 3d).

Furthermore, structural comparison between the target-free and target-bound complexes indicated clashes between the  $\alpha 8$ - $\alpha 9$  loop of the REC2 domain and the 17–18 nucleotides of the TS. This observation suggested that at least 17 base pairs between the guide and TS may be required to trigger the conformational changes of REC2–Stem 3. Consistent with these structural observations, in vitro cleavage assays and in vivo plasmid clearance assays showed that Cas9d did not cleave target DNA with a spacer sequence shorter than 17 nucleotides (Figs. 4b, c and S12), aligning with previous findings for canonical Cas9s that also require a minimum of 17 base pairs to induce effective REC domain movement<sup>38</sup>. Thus, a heteroduplex of at least 17 base pairs may interact sterically with the  $\alpha 8$ - $\alpha 9$  loop of the REC2, acting as a checkpoint to initiate the conformational change of the REM.

In contrast to the large-scale movement of the REC2–Stem 3, REC1–Stem 2 remains structurally rigid with minimal conformational changes upon target binding (Fig. 4a). Only non-specific backbone interactions were observed between REC1 and Stem 2 (Figs. S8a and S11a), indicating that Stem 2 functions primarily as structural support, holding the REC2–Stem 3 for the hinge-like motion to facilitate the guide-TS heteroduplex propagation (Fig. 4a). To test this hypothesis, we performed large-scale nucleotide substitutions in the base-paired regions in Stem 2 (69A:137U–78G:128C) while preserving the complementarity. The target DNA cleavage activity of Cas9d with these mutant sgRNAs remained comparable to that with the WT sgRNA, corroborating the structurally supportive role of Stem 2 (Fig. S13).

Additionally, structural comparison between the target-free and target-bound states showed clashes between the  $\alpha 3$  helix of the REC1 domain and the TS at the cleavage site (4'dA–3'dG, Fig. 4a). This clash could potentially hinder early R-loop formation at PAM-proximal sites,



**Fig. 4 | Cas9d requires a minimum 17-bp heteroduplex for effective cleavage.** **a** Structural alignment of Cas9d at target-free and target-bound states. The lid region present in the RNP complex is highlighted in yellow. Two potential clash regions upon target binding are highlighted with red rectangles. The zoomed-in view shows the movement of the two sgRNA stems upon target engagement. **b** A schematic representation of base pairing between the target DNA and the sgRNA guide, showing variations in spacer length. **c** In vitro cleavage assays were

performed at 30 and 60 min using 2.5 mM  $Mg^{2+}$  and various spacer lengths ranging from 14 nt to 30 nt, as indicated. **d** In vitro cleavage assays at various time points using WT and the lid-deleted Cas9d, with the deletion spanning residues 100–129. **e** In vitro cleavage assays using WT and the lid-deleted Cas9d in complex with the full-length (FL, 135-nt scaffold) and the triple truncated (107-nt scaffold) sgRNAs. **c–e** Each gel represents three biological replicates. Source data are provided as a Source Data file.

indicating that the “lid-off” conformation of this region is essential for guide–TS propagation. Reversible changes to the “lid-on” conformation could lead to unsuccessful heteroduplex formation. The deletion of this region slightly enhanced target DNA cleavage compared to WT Cas9d (Fig. 4d). Interestingly, the lid-deleted Cas9d in complex with the triple-truncated sgRNA exhibited nuclease activity comparable to the WT Cas9d system (Fig. 4e), again demonstrating the engineerability of the Cas9d.

#### Guide–target heteroduplex recognition and cleavage fidelity

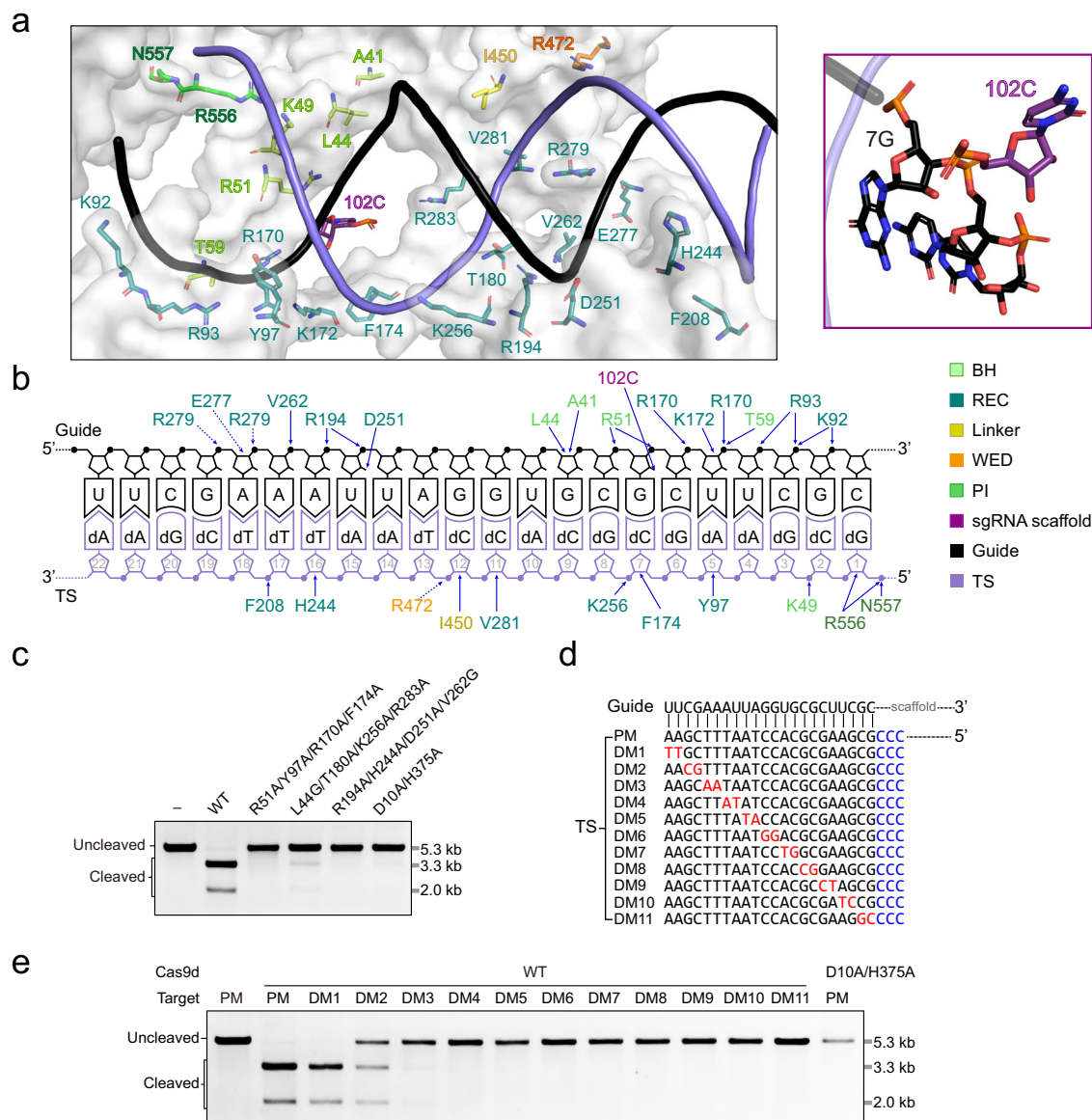
The guide–TS heteroduplex adopts a canonical A-form geometry at the four PAM-distal base pairs, while the remaining 18 base pairs exhibit a distorted A-form due to extensive contacts with Cas9d (Fig. 5a). Charged and hydrophobic residues primarily from the REC lobe, along with residues from the L2 linker and WED domain, engage with the sugar-phosphate backbone of the 1–18 bp region in the guide–TS duplex through non-specific contacts (Fig. 5a–b). These interactions monitor the complementarity of each base pair in this region. Mutations in heteroduplex-interacting residues, such as R51A/Y97A/R170A/F174A, L44G/T180A/K256A/R283A, and R194A/H244A/D251A/V262G, completely abolished target cleavage, underscoring the critical role of stringent target monitoring of Cas9d (Fig. 5c). Notably, the 2'-OH of 7G within the seed region is stabilized by 102C within Stem 3, further demonstrating the versatile functions of sgRNA scaffold in heteroduplex recognition (Fig. 5a–b).

Previous structural evidence from Cas9 with mismatched DNA showed that a lack of interaction between the guide–TS heteroduplex often leads to mismatch tolerance at certain sites, the so-called “blind spots”<sup>9,10</sup>. In contrast, the precise monitoring of the guide–TS duplex by Cas9d may result in a low tolerance to mismatches. To verify this, we introduced consecutive double mismatches tiling along the target (Fig. 5d) and assessed cleavage activity using both WT Cas9d and

SpyCas9 under identical reaction conditions. Mismatches in the PAM-distal region (21–22) were well tolerated by Cas9d, while mismatches in the 19–20 region compromised activity. Mismatches in the 18-bp PAM-proximal region were not tolerated, leading to a lack of cleavage. In contrast, SpyCas9 showed higher tolerance to the same mismatches (Fig. S14a), consistent with previous observations<sup>9,10</sup>, indicating that Cas9d has higher fidelity compared to SpyCas9 in in vitro reactions (Figs. 5e and S14a). The lid-deleted Cas9d was also tested for mismatch tolerance, showing no significant impact from the deletion (Fig. S14b). Furthermore, at corresponding positions, single mismatches were generally better tolerated than double mismatches by both Cas9d and SpyCas9 (Fig. S15), likely due to the lower degree of mismatch-induced heteroduplex distortions. Notably, Cas9d exhibited higher overall fidelity, as evidenced by its lower tolerance for single mismatches at most positions compared to SpyCas9. Taken together, the in vitro DNA cleavage assays and structural analyses suggested that Cas9d provides stringent monitoring of the guide–TS heteroduplex through the REC lobe and the sgRNA scaffold, exhibiting low tolerance for mismatches in vitro, particularly in the PAM-proximal region.

#### REC-insertion mutation accelerates target cleavage

Structural comparisons among type II Cas9 orthologs (including II-A to II-D) and the ancestral IscB revealed conserved nuclease domains (Fig. S2c–d). However, their varying cleavage activities may, at least in part, be attributed to their distinct structural modules responsible for target engagement (Figs. 6a and S9–10). Previous phylogenetic studies suggested that the acquisition of the REC domain during Cas9 evolution likely enhanced the host's immune defense to foreign nucleic acids<sup>7,16,17</sup>. To test this hypothesis, we inserted the REC2 domain from SpyCas9 into the REC domain of Cas9d and assessed its cleavage efficiency (Fig. 6b–c). The Cas9d mutant with the REC insertion exhibited increased cleavage activity compared to WT Cas9d, without



**Fig. 5 | Interactions between guide-TS heteroduplex and Cas9d. a** Cas9d accommodates the guide-TS duplex within a central channel. The channel region of Cas9d is shown as surface representation (left panel), with interacting residues and nucleotide depicted as sticks and the duplex illustrated in cartoon form. The right panel represents the polar interaction between 102C in the sgRNA scaffold and 7G in the seed region, with nucleotides shown in sticks representation. The color scheme corresponds to Fig. 1e-f. **b** Schematic representation of guide-TS heteroduplex recognition by the Cas9d system. Side-chain interactions with the

heteroduplex are depicted as solid blue arrows, while backbone-mediated interactions are shown as blue dashed arrows. **c** In vitro cleavage assays were performed using WT Cas9d and its mutants. **d** Schematic representation of double-mismatches (DM) on the TS (highlighted in red), with the complementary NTS omitted for clarity. **e** In vitro cleavage assays were performed using WT Cas9d RNP with targets bearing various mismatches as depicted in d. c, e The gel images are both representative of three independent experiments each. Source data are provided as a Source Data file.

affecting mismatch tolerance (Figs. 6d and S16). These findings support the notion that REC domain acquisition positively impacts Cas9 functionality.

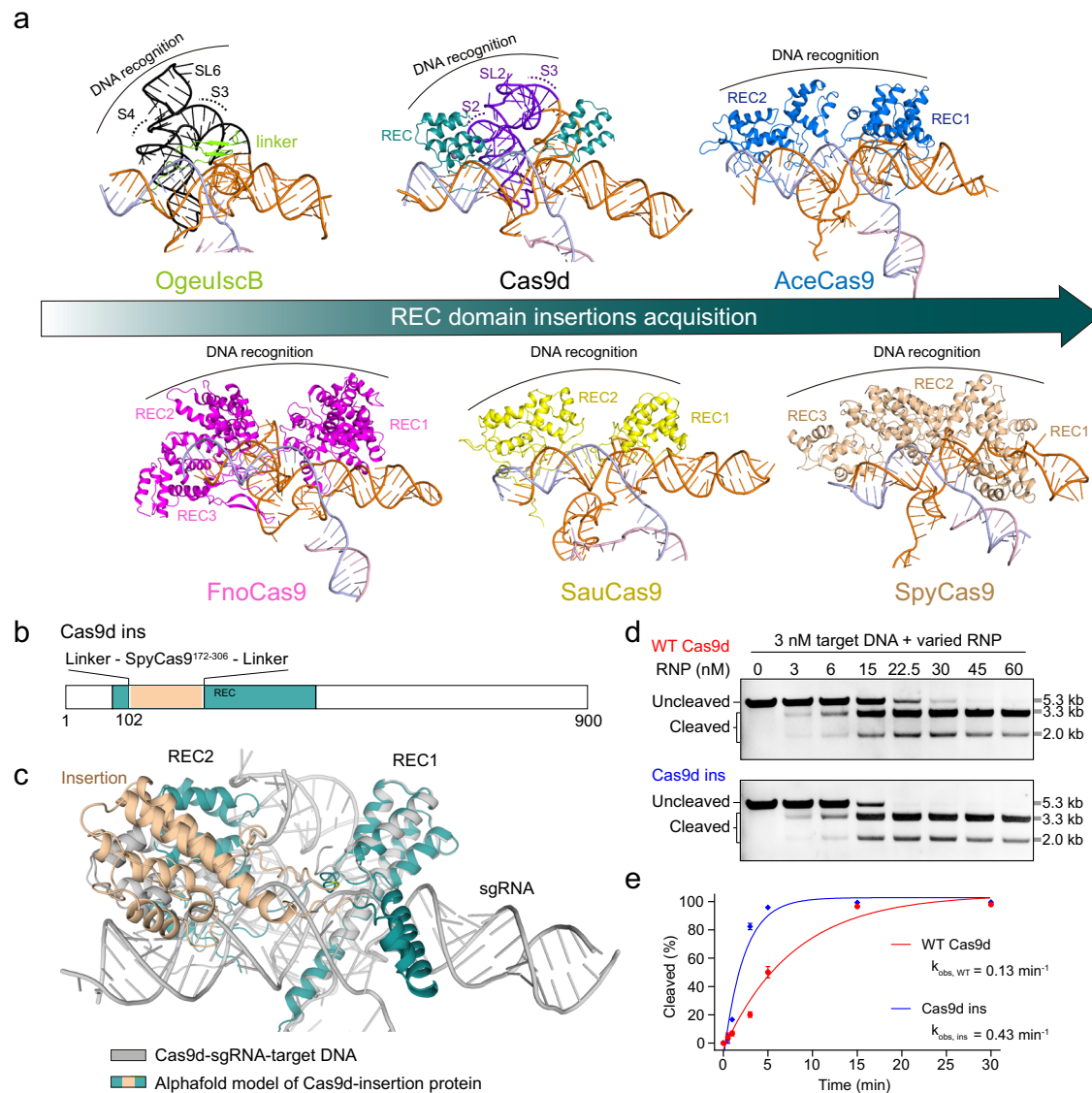
## Discussion

In this work, we uncovered the unique features of the Cas9d system, which employs a hybrid architecture of protein and sgRNA for target engagement. Cas9d protein harbors a relatively small REC domain (208 aa), situated between the minimal REC domain of IscB (38 aa) and those of other Cas9 proteins, such as SpyCas9 (623 aa), FnoCas9 (775 aa), and *Acidothermus cellulolyticus* (Ace)Cas9 (387 aa) (Fig. S9–10, S17). The variation in REC domain sizes not only reflects their evolutionary positions (Fig. S17), but also implies the functional diversity among Cas9 proteins and their ancestors. Cas9d also features a

relatively long and complex sgRNA scaffold, which acts not only as an adaptor between the protein and the target, as reported for other Cas9 orthologs<sup>25,30,31,33</sup>, but also provides structural support and plays an important role in guide-TS engagement, as observed in IscB proteins<sup>16,17</sup>. Specifically, Stem 2 and Stem 3 of the sgRNA work in concert with the REC domains, forming a hybrid functional module, the REM.

Based on our structural and biochemical analyses of the Cas9d system, we propose an allosteric activation model for Cas9d's nuclease activity (Fig. S18, Supplementary Movie 1). Following PAM positioning by the WED and PI domain in Cas9d, the dsDNA is unwound by the phosphate lock loop, enabling the TS base pairing with the sgRNA guide sequence immediately adjacent to the PAM. Similar to canonical REC domains in other Cas9 orthologs, the REM in Cas9d undergoes





**Fig. 6 | Cas9d REC domain insertion accelerates target cleavage.** **a** Structural comparisons among Cas9 orthologs and the Cas9 ancestor IscB illustrate the diversification of the REC domain and the trend of REC domain acquisition. **b** Schematic design of REC-insertion Cas9d fusion protein. The SpyCas9<sup>172-306</sup> with flanking (GGG)<sub>3</sub> linkers was inserted into Cas9d between Glu102 and Glu103. **c** Structural alignment of the Cas9d ternary complex (gray) with the AlphaFold2-predicted model of the insertional mutant. The REC insertion is colored wheat. **d** In vitro cleavage assays were performed at various target-to-RNP ratios, comparing

WT Cas9d with the insertional mutant. Gels are representative of three independent experiments. Cleavage (%) was determined as described in Fig. 2. **e** In vitro cleavage assays using 3 nM of DNA substrate and 150 nM of RNP were performed at different time points. The scatter plot shows the mean  $\pm$  standard deviation (SD) from three biological replicates. Cleavage (%) was quantified, and the observed rate constant ( $k_{obs}$ ) was estimated using one-phase exponential regression. Representative gels are shown in Supplementary Fig. S16a. Source data are provided as a Source Data file.

critical conformational changes, which are essential for full R-loop formation and for accommodating the guide-TS heteroduplex. Formation of the guide-TS heteroduplex exceeding 17 base pairs facilitates the movement of the nuclease domains to the cleavage site, enabling target DNA cleavage. Moreover, Cas9d shows high fidelity in discriminating mismatched targets, owing to the stringent monitoring of the guide-TS heteroduplex complementarity by the REM. Notably, during the preparation of our manuscript, a similar work also addressed that the sgRNA stems are critical for the nuclease activity of Cas9 through concerted movements with the REC domains<sup>27</sup>. Their elegant kinetic assays suggested that Cas9d exhibits a higher level of discrimination against mismatches, consistent with our mismatched target cleavage assay<sup>27</sup>.

Our study offers structural and mechanistic insights into the further engineering of Cas9d. The compact Cas9d itself can be further

truncated. A lid-like region in the REC domain prevents the central channel's accessibility to the target. Removing this region accelerates target cleavage by Cas9d without compromising mismatch discrimination. Moreover, the sgRNA can be truncated by up to 20% in its scaffold region, as demonstrated in a recent, insightful study of Cas9d using *E. coli* in vivo assays<sup>27</sup>. Combined truncations in both the Cas9d protein and sgRNA maintained Cas9d's nuclease activity, demonstrating the versatility of this engineered system. Furthermore, the PAM sequence showed slight tolerance for adenine at positions (-2) and (-3). Mutating the PAM-interacting residue (K715A) retains target cleavage, indicating the potential for developing PAM-relaxed variants. Interestingly, a domain insertional engineering enhances the cleavage efficiency of Cas9d, suggesting the adaptability of recognition modules among Cas9 orthologs.

Our study underscores the value of combining structural data with evolutionary insights to guide Cas9d engineering and improve



genome-editing tools. The engineering of Cas9d in our study relies primarily on experimentally determined structural and biochemical data, which provide crucial insights into the mechanisms of target engagement and nuclease activity activation. The structure-guided approach has been widely applied in other studies, such as the engineering of high fidelity SuperFi-Cas9<sup>9</sup> and activity-enhanced Cas13<sup>39</sup>. While evolutionary data helps guide the identification of both conserved and adaptively evolving features, as well as functional hypotheses, it alone would not have been sufficient for precise engineering. More recently, AI-driven methods have also been employed to develop genome-editing tools<sup>40,41</sup>. However, these approaches still depend on prior structural and functional knowledge for initial screening and require experimental validation. Additionally, experimental techniques like phage-assisted evolution offer alternative strategies for engineering genome editors without the need for prior structural knowledge<sup>42,43</sup>. Combining these complementary approaches will significantly enhance protein engineering efforts, paving the way for developing more efficient and versatile genome-editing tools.

## Methods

### Plasmid construction

The gene encoding the full-length Cas9d (Sangon) was codon optimized for *E. coli* expression and assembled into a modified pET vector (2Bc-T, Addgene #37236) with a C-terminal thrombin-TwinStrepII-His-tag using Gibson assembly (New England Biolabs, E2611L). Mutations in Cas9d and SpyCas9 (WT expression plasmid, Addgene #101199) were introduced using the QuickChange Mutagenesis kit (Takara, Cat# 638949) following the manufacturer's instructions. The WT and variant sgRNAs for Cas9d and SpyCas9 (oligos from Genewiz) were individually cloned into a co-transformation vector (13S-A, Addgene #48323), each under the control of a T7 promoter and terminator. Cloning was performed using Golden Gate assembly with BsaI-HFv2, T4 Ligase and T4PNK (New England Biolabs, R3733L, M0202L, and M0201L) following the manufacturer's protocols. The target plasmids were Golden Gate cloned into a modified pCMV vector (Addgene #205277) with Kanamycin resistance. The sequences of nucleic acids used in this study were provided in Source Data.

### Co-expression and purification of the Cas9d-sgRNA complex

Two plasmids, one encoding either the WT or mutant Cas9d and the other the appropriate sgRNA, were co-transformed into *E. coli* Rosetta (DE3) cells (Novagen, Cat# 70954). The cells were grown in TB culture media and induced with 0.25 mM IPTG upon reaching an OD<sub>600</sub> of ~0.8. After 12 h of induction at 18 °C, the cells were harvested and resuspended in a lysis buffer containing 50 mM HEPES (pH 7.2), 300 mM NaCl, and 5% glycerol, freshly supplemented with 1 mM DTT, 0.2 mg/mL lysozyme, and 10 µg/mL DNase. The suspension was left on ice for 30 min before sonication. The lysates were clarified by centrifugation at 20,000 × *g* for 40 min at 4 °C. The supernatant was then filtered through a 0.22 µm membrane and subjected to Strep-Tactin (IBA, Cat# 2-5010-002) affinity purification. The eluate was further purified by affinity chromatography (HiTrap Heparin HP, Cytiva), with elution performed using a linear gradient of 0.15–1 M NaCl in a buffer containing 50 mM HEPES (pH 7.2), 1 mM DTT, and 5% glycerol. The concentrated peak samples were then purified by size-exclusion chromatography (Superdex 200 Increase 10/300 GL, Cytiva) in gel-filtration buffer (25 mM HEPES pH 7.2, 2 mM DTT, 350 mM NaCl). Peak fractions were confirmed by SDS-PAGE and Urea-PAGE, pooled, and concentrated to a final concentration of ~5 mg/mL in gel-filtration buffer containing 30% glycerol. The aliquots were snap-frozen in liquid nitrogen and stored at –80 °C until use. The SpyCas9 and its sgRNA were co-expressed and purified as described above, with the minor modification that HEPES buffer was adjusted to pH 7.9.

### In vitro target DNA cleavage assay

The purified RNP complex was pre-incubated with supercoiled or FspI (New England Biolabs, R0135L) digested linearized target plasmid DNA for 15 min at 37 °C. The reactions were initiated by adding MgCl<sub>2</sub> and further incubated at 37 °C for the specified time. The final reaction mixture contained 200 ng of DNA and a defined amount of RNP in a buffer consisting of 25 mM HEPES (pH 7.5), 100 mM KCl, 5 mM Mg<sup>2+</sup>, 0.1 mg/mL BSA, 5% glycerol, and 1 mM DTT. For reactions with supercoiled plasmids, the RNP to target DNA molar ratio was 20:1 with a 30-min incubation, while for linearized plasmids, the ratio was 50:1 with a 60-min incubation, unless otherwise specified. The reactions using SpyCas9 for cleavage were performed in identical conditions with those of using Cas9d. Reactions were terminated by adding proteinase K (ThermoFisher, EO0491) to a final concentration of 1 mg/mL and incubating at 37 °C for 15 min. Samples were then mixed with the 6× DNA loading dye and resolved by 1% agarose gel electrophoresis in 1× TAE buffer. Gel images were visualized and obtained by using an Azure 400 imaging system. The percentage of cleavage was calculated by normalizing the products' band intensity to the total intensity of the substrates and products (ImageJ). The uncropped raw gel images and the cleavage percentages are provided in Source Data.

### Plasmid clearance assay in *E. coli*

Plasmids (pCDFDuet-1, Novagen) expressing sgRNA with varying spacers lengths and the WT Cas9d (spectinomycin-resistant) were constructed by Gibson assembly, and transformed into BL21(DE3) competent cells carrying target plasmids (Kanamycin-resistant). The transformed cells were plated on LB agar supplemented with spectinomycin, kanamycin, and 0.4% glucose, and cultured overnight at 37 °C. Single colonies were then selected and inoculated into LB medium at 37 °C, grown to an OD<sub>600</sub> of 0.6–0.8. The culture was split into two: one was induced with 0.1 mM IPTG, while the other continued growing in LB with glucose to prevent leaky expression. Each culture was subjected to serial tenfold dilutions, and 10 µL of each dilution was spotted onto selective LB agar plates with either glucose or IPTG for further analysis.

### Cryo-EM sample preparation and data collection

Purified Cas9d RNP complex were concentrated to approximately 1 mg/mL. Prior to use, holey carbon grids (Quantifoil Au R1.2/1.3, 300 mesh) were glow discharge for 80 seconds at a current of 15 mA and a pressure of 0.39 mBar in air. Using an FEI Vitrobot Mark IV system (Thermo Fisher Scientific), 4 µL aliquots of the RNP complex were applied to the grids at 4 °C and 100% humidity, with a blotting duration of 3.5 seconds at a blot force of -2 and a draining time of 11 seconds. Subsequently, the grids were promptly plunge-frozen in liquid ethane and carefully transferred to liquid nitrogen for storage until data collection. In the preparation of the ternary complex, the RNP was combined with pre-annealed DNA at a molar ratio of 1:1.4 and incubated at 37 °C for 30 min prior to vitrification.

The grids were subsequently transferred to a Titan Krios G3i microscope from Thermo Fisher Scientific, operating at an accelerating voltage of 300 kV. The microscope was coupled with a Gatan Quantum-LS Energy Filter with a 20 eV slit width, as well as a Gatan K3 Summit direct electron detector operated in super-resolution mode. During the imaging process, the images were captured at a nominal magnification of 105,000× (0.83 Å pixel size). The acquisition of movie stacks was performed using EPU software. The defocus range was set between -0.8 to -1.6 µm, and an image-beam-shift multiple recording strategy was employed to enhance data quality. Throughout the imaging procedure, a total accumulated dose of approximately 50 electrons per square angstrom (e<sup>-</sup>/Å<sup>2</sup>) was directed toward each specimen. Notably, each micrograph stack was composed of 32 individual frames, which contributed to the overall data set.

## Cryo-EM data processing

For all datasets, motion correction was applied using either MotionCor2<sup>44</sup> or the patch motion correction method in cryoSPARC<sup>45</sup>. Contrast transfer function (CTF) parameters were calculated using patch CTF estimation method in cryoSPARC. Micrographs deemed unsuitable for analysis were excluded after visual inspection. All subsequent image processing steps were conducted exclusively within cryoSPARC. A Topaz model was initially trained on a curated set of 100 micrographs using the “manually curate exposures” task. Particle picking was performed using the Blob picker, extracting particles within a 300-pixel box and subsequently Fourier-cropped to 100 pixels for computational efficiency. These particles underwent iterative rounds of 2D classification. Selected particles were used to refine the Topaz model through the “Topaz train” process. All micrographs were then processed using the trained Topaz model to extract particles, which were further subjected to 2D classification. High-quality particles were selected for ab initio reconstruction and heterogeneous refinement. Particles in good classes underwent non-uniform and local refinement to produce the final map. The overall resolution of this final map was determined using the 0.143 criterion of the gold-standard Fourier shell correlation<sup>46</sup>. Local resolution estimation was also performed to assess resolution across different regions of each map.

For the Cas9d-sgRNA RNP complex, a total of 464,276 particles were initially picked using Topaz. These particles were subjected to 2D classification, and the high-quality particles were selected for ab-initio reconstruction followed by heterogeneous refinement. A refined class consisting of 229,218 particles was then subjected to non-uniform (NU) refinement, yielding a final reconstruction with an overall resolution of 2.87 Å. Similarly, for the Cas9d-sgRNA-target DNA complex, 342,581 particles were picked using Topaz. After 2D classification and the selection of good particles, ab-initio reconstruction and heterogeneous refinement were performed. A class comprising 171,509 particles was s NU-refined, resulting in a final reconstruction with an overall resolution of 2.99 Å. Detailed EM image parameters for each sample are provided in Supplementary Table 1.

## Model building and refinement

To model the Cas9d-sgRNA binary complex, an AlphaFold2-predicted model<sup>28</sup> of Cas9d was segmented according to its domain organization. Individual domain models were fitted into the cryo-EM density map in ChimeraX<sup>47</sup> at a low threshold. These docked models were then manually refined in Coot<sup>48</sup> based on the EM density, with residues missing in the density omitted. The initial Cas9d protein model was used to segment the sgRNA density in ChimeraX<sup>47</sup>. An initial sgRNA model was manually built in Coot<sup>48</sup>. The manually built Cas9d-sgRNA complex model was iteratively refined using Coot and the phenix.real\_space\_refine<sup>49</sup> program in phenix. For the Cas9d-sgRNA-target DNA ternary complex, the pre-refined RNP complex model was fitted into the cryo-EM density map and manually refined in Coot<sup>48</sup>. The guide-TS duplex and dsDNA were also manually traced based on the EM density. The manually built Cas9d-sgRNA-target DNA ternary complex model was then subjected to iterative refinement using Coot and phenix.real\_space\_refine<sup>49</sup>. The quality of the final models was validated with MolProbity in Phenix<sup>49</sup>. Data collection and model refinement statistics are concluded in Supplementary Table 1.

## Reporting summary

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

## Data availability

The atomic coordinates have been deposited in the Protein Data Bank under accession codes [8ZQ9](#) (Cas9d-sgRNA RNP complex) and [8ZDR](#) (Cas9d-sgRNA-target DNA complex). Cryo-EM maps have been

deposited in the Electron Microscopy Data Bank under corresponding accession codes [EMDB-60379](#) and [EMDB-60006](#). Structural data supporting the findings of this study are accessible in the PDB (accession: [4I3D](#), [4ZT0](#), [5B20](#), [5CZZ](#), [5F9R](#), [600Z](#), [6WBR](#), [7XHT](#)). The uncropped gel images and the nucleic acid sequences used in this study are provided in the Source Data file. Source data are provided with this paper.

## References

- Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl Acad. Sci.* **109**, E2579–E2586 (2012).
- Sapranas, R. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
- Aliaga Goltsman, D. S. et al. Compact Cas9d and HEARO enzymes for genome editing discovered from uncultivated microbes. *Nat. Commun.* **13**, 7602 (2022).
- Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
- Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
- Bravo, J. P. K. et al. Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature* **603**, 343–347 (2022).
- Pacesa, M. et al. Structural basis for Cas9 off-target activity. *Cell* **185**, 4067–4081.e4021 (2022).
- Ran, F. A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).
- Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y.-L. & Weiss, D. S. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**, 254–257 (2013).
- Tsui, T. K. M., Hand, T. H., Duboy, E. C. & Li, H. The impact of DNA topology and guide length on target selection by a cytosine-specific Cas9. *ACS Synth. Biol.* **6**, 1103–1113 (2017).
- Chylinski, K., Le Rhun, A. & Charpentier, E. The tracrRNA and Cas9 families of type II CRISPR–Cas immunity systems. *RNA Biol.* **10**, 726–737 (2013).
- Gasiunas, G. et al. A catalogue of biochemically diverse CRISPR–Cas9 orthologs. *Nat. Commun.* **11**, 5512 (2020).
- Schuler, G., Hu, C. & Ke, A. Structural basis for RNA-guided DNA cleavage by LscB-ωRNA and mechanistic comparison with Cas9. *Science* **376**, 1476–1481 (2022).
- Kato, K. et al. Structure of the LscB-ωRNA ribonucleoprotein complex, the likely ancestor of CRISPR–Cas9. *Nat. Commun.* **13**, 6719 (2022).
- Han, D. et al. Development of miniature base editors using engineered LscB nickase. *Nat. Methods* **20**, 1029–1036 (2023).
- Yan H. et al. Assessing and engineering the LscB-ωRNA system for programmed genome editing. *Nat. Chem. Biol.* **20**, 1617–1628 (2024).
- Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
- Huang, T. P. et al. High-throughput continuous evolution of compact Cas9 variants targeting single-nucleotide-pyrimidine PAMs. *Nat. Biotechnol.* **41**, 96–107 (2023).
- Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

23. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
24. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
25. Hirano, H. et al. Structure and engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).
26. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
27. Ocampo, R. F. et al. DNA targeting by compact Cas9d and its resurrected ancestor. *Nat. Commun.* **16**, 457 (2025).
28. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
29. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
30. Nishimasu, H. et al. Crystal Structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).
31. A, Das et al. Coupled catalytic states and the role of metal coordination in Cas9. *Nat. Catal.* **6**, 969–977 (2023).
32. Li, X. et al. Structural basis for the type I-F Cas8-HNH system. *EMBO J.* **43**, 4667–4667 (2024). 4656.
33. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
34. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9–guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
35. Holm, L. Dali server: structural unification of protein families. *Nucleic Acids Res.* **50**, W210–W215 (2022).
36. Fu, M. & Blackshear, P. J. RNA-binding proteins in immune regulation: a focus on CCCH zinc finger proteins. *Nat. Rev. Immunol.* **17**, 130–143 (2017).
37. Wang, D. et al. Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. *BMC Genomics* **9**, 44 (2008).
38. Pacesa, M. et al. R-loop formation and conformational activation mechanisms of Cas9. *Nature* **609**, 191–196 (2022).
39. Yang, J. et al. Engineered LwaCas13a with enhanced collateral activity for nucleic acid detection. *Nat. Chem. Biol.* **19**, 45–54 (2023).
40. Jiang K. et al. Rapid in silico directed evolution by a protein language model with EVOLVEpro. *Science* **387**, 6732 (2024).
41. Nguyen E. et al. Sequence modeling and design from molecular to genome scale with Evo. *Science* **386**, 6723 (2024).
42. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
43. Doman, J. L. et al. Phage-assisted evolution and protein engineering yield compact, efficient prime editors. *Cell* **186**, 3983–4002.e3926 (2023).
44. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
45. 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).
46. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* **9**, 853–854 (2012).
47. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
48. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132 (2004).
49. Williams, C. J. et al. MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).

## Acknowledgements

Zhongmin Liu is an investigator at the SUSTech Institute for Biological Electron Microscopy. We thank the cryo-EM Centre, Southern University of Science and Technology for supporting cryo-EM data acquisition. This work was supported by the National Natural Science Foundation of China (22377091 to J.Y., 32371275 to Z.L. and 32322040 to H.Z.), the Natural Science Foundation of Tianjin Municipal Science and Technology Commission (23JCZDJC00410 to H.Z.), Tianjin Science and Technology Plan Project (23ZYCGSY00750 to H.Z.), Scientific Research Program of Tianjin Municipal Education Commission (2023ZD011 to H.Z. and 2024ZD037 to J.Y.), Shenzhen Municipal Basic Research projects (No. JCYJ20210324105007020 to Z.L.), Guangdong Innovative and Entrepreneurial Research Team Program (No. 2021ZT09Y104 to Z.L.), Guangdong basic and applied basic research foundation (2024A1515011538 to Z.L.), Shenzhen Municipal Basic Research projects (No. ZDSYS2022040211000001 to Z.L.) and Southern University of Science and Technology.

## Author contributions

J.Y., Zhongmin Liu, and H.Z. conceived the project. T.W., Y.H., Z.Long., S.Z., L.Z., Zhikun Liu, and Q.Z. carried out the experimental work. Data analysis was performed by X.L., H.S., M.Z., H.Y., Zhongmin Liu, and H.Z. J.Y. and X.L. wrote the manuscript with contributions from all authors.

## 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/s41467-025-57455-9>.

**Correspondence** and requests for materials should be addressed to Zhongmin Liu or Heng Zhang.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**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) 2025



<sup>1</sup>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, China. <sup>2</sup>Department of Biochemistry and Molecular Biology, Tianjin Key Laboratory of Cellular Homeostasis and Disease, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. <sup>3</sup>Shenzhen Key Laboratory of Biomolecular Assembling and Regulation, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>4</sup>Institute for Biological Electron Microscopy, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>5</sup>Department of Immunology and Microbiology, School of Life Sciences, Southern University of Science and Technology, Shenzhen, Guangdong, China. <sup>6</sup>Tianjin Key Laboratory on Technologies Enabling Development of Clinical Therapeutics and Diagnostics, Department of Chemical Biology, School of Pharmacy, Tianjin Medical University, Tianjin, China. <sup>7</sup>Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. <sup>8</sup>Department of Pharmacology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. <sup>9</sup>These authors contributed equally: Jie Yang, Tongyao Wang, Ying Huang, Zhaoyi Long. ✉ e-mail: [liuzm@sustech.edu.cn](mailto:liuzm@sustech.edu.cn); [zhangheng134@gmail.com](mailto:zhangheng134@gmail.com)