Mechanism of type II-B CRISPR-Cas9 enhanced specificity

Grace N. Hibshman[1,2], David W. Taylor[1-4]

[1]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX 78712, USA.
[2]Interdisciplinary Life Sciences Graduate Programs, Austin, TX 78712, USA.
[3]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX 78712, USA.
[4]LIVESTRONG Cancer Institutes, Dell Medical School, Austin, TX 78712, USA.

Abstract

The Cas9 endonuclease from *Streptococcus pyogenes* (SpCas9) has revolutionized genome editing, as it can be programmed by an RNA sequence to target and cleave DNA. However, SpCas9 is prone to deleterious off-target effects, prompting efforts to develop and discover high-fidelity variants. In the search for a more precise genome editor, many intrinsically specific Cas9 orthologs have been found in nature. One such ortholog is from the human pathogen *Francisella novicida* (FnCas9), yet the molecular basis of its enhanced specificity remains unknown. Here, we uncover the mechanism of FnCas9 specificity through kinetic characterization in concert with structural determination. We identify a novel region within the REC3 domain, the REC3 clamp, as critical for mismatch discrimination through interaction with the PAM-distal heteroduplex. Mutational analysis confirmed the role of the REC3 clamp in reducing cleavage of off-target sequences. Further, a small CRISPR-associated RNA (scaRNA) within the *F. novicida* CRISPR locus has been shown to guide transcriptional repression of an endogeneous gene without the consequence of cleavage. This secondary function of FnCas9 enables host subversion and heightened virulence, yet the mechanism distinguishing target repression versus cleavage is unclear. Through experimental and simulated structural analysis, we identify the REC3 clamp as the key to differentiating DNA sequences targeted for repression versus cleavage, ultimately bolstering the virulence of *F. novicida*. The conservation of the REC3 clamp across type II-B CRISPR systems highlights its evolutionary significance, and potential as a target for developing specific genome editors and novel antibacterial effectors.

Introduction

The ongoing arms race between bacteria and invading genetic elements has resulted in a plethora of bacterial defense modules. One defense mechanism harnessed by bacteria and archaea is the CRISPR-Cas system[1]. The discovery of the endonuclease Cas9 quickly prompted optimization of the system for genome engineering. Cas9 from *S. pyogenes* is the most widely used CRISPR-Cas enzyme for genome editing, which utilizes a guide RNA (gRNA) to direct the enzyme to the desired DNA editing site[2]. Only 15 years after its initial discovery, CRISPR-Cas9 therapeutics are now approved for clinical use[3,4].

Despite the widespread utility of Cas9, the enzyme has its pitfalls, most notably, targeting of unintended sequences, also known as off-target effects[5,6]. This deleterious phenomenon has been combated through years of fine-tuning Cas9 through directed evolution and rational engineering to minimize off-target effects[7–11]. The resulting high-fidelity Cas9 variants have been rigorously characterized and are typically less efficient than wildtype Cas9[12–14]. This tradeoff reminds us that nature has often already optimized enzymes for their intended use. Therefore, we chose to explore the diversity of type II CRISPR-Cas systems for a Cas9 ortholog that may exhibit higher fidelity.

The Cas9 ortholog from *Francisella novicida* (FnCas9) has been less studied than SpCas9, despite initial findings indicating that FnCas9 is intrinsically more specific than SpCas9[15]. SpCas9 is known to be particularly susceptible to mismatches that occur away from the protospacer adjacent motif (PAM) sequence, corresponding to the final base pairs of the R-loop[16,17]. Intriguingly, unlike SpCas9, FnCas9 does not tolerate mismatches in this PAM-distal region. Therefore, we wanted to interrogate the molecular basis of this enhanced specificity by FnCas9.

Here, we employ structural determination informed by kinetic analyses to capture FnCas9 in multiple states during nuclease activation. This technique enabled visualization of distinct checkpoints during FnCas9 activation. One novel checkpoint emerged in which a region of the REC3 domain becomes ordered only in productive conformations. This region of REC3, which we call the REC3 clamp, docks onto the PAM-distal heteroduplex, indicating the REC3 clamp confers FnCas9 specificity. Mutation of REC3 clamp residues diminishes cleavage of substrates with PAM-distal mismatches, providing further support for the role of the REC3 clamp in FnCas9 mismatch discrimination.

F. novicida is a human pathogen that evades recognition by its host, enabling replication in the cytosol[18]. The native *F. novicida* CRISPR locus contains Cas9, Cas1, Cas2, Cas4, CRISPR RNA (crRNA), trans-activating CRISPR RNA (tracrRNA), and a small CRISPR-associated RNA (scaRNA)[19,20]. The unique scaRNA was found to target an endogenous bacterial lipoprotein (BLP) sequence stifling its expression, ultimately preventing activation of host Toll-like receptor 2 (TLR2), thus subverting host detection. The mechanism by which FnCas9 differentiates target repression versus cleavage is posited

to be deciphered by the length of complementarity between the RNA and the DNA target sequence. Truncated spacer complementarity is a well-characterized mode of cleavage inhibition across CRISPR-Cas systems as in the case of tracr-L, a long-form natural single gRNA that guides Cas9 to transcriptionally repress its promoter[21]. Through our structural analyses, we uncovered that the REC3 clamp senses R-loop complementarity, enabling repression of DNA sequences without concomitant cleavage. Furthermore, we demonstrate that the role of the REC3 clamp in enhanced specificity and virulence extends across type II-B CRISPR-Cas systems, which contain similarly enlarged REC3 domains. Our findings reveal the molecular and evolutionary basis of heightened specificity and virulence exhibited by type II-B CRISPR-Cas systems. These results provide insight into how type II-B CRISPR-Cas systems may be improved for genome editing, and how they can be targeted to dampen antibiotic resistance in pathogenic bacteria.

Results

FnCas9 undergoes distinct structural changes during nuclease activation

Previous structural studies of FnCas9 were performed using inactive enzyme, where the catalytic residue N995 in the HNH domain was mutated to alanine[22]. Inactivating HNH may hinder structural rearrangements required for enzyme activation[23,24]. To elucidate the conformational changes associated with FnCas9 activation, we prepared cryo-EM grids of fully active FnCas9 in complex with gRNA and 55-bp target DNA after a one-hour incubation. The target DNA substrate contained the preferred PAM for FnCas9, NGG[22]. This single cryo-EM dataset yielded four unique, high-resolution reconstructions of FnCas9 throughout the process of enzyme activation. (**Fig. 1, Extended Data Fig. 1, Extended Data Fig. 2 and Extended Data Table 1**). The reconstructions were distinguished by the presence or absence of the two catalytic domains, HNH and RuvC.

All four reconstructions demonstrate that FnCas9 recognizes the NGG PAM primarily through hydrogen bonding between G2 and R1585, and G3 and R1556 (**Extended Data Fig. 3**), consistent with previous reports[22]. The two non-productive structures of FnCas9 do not contain resolvable HNH density, indicating that the HNH domain is flexibly tethered and inactive in these states. We also noticed that a large portion of REC3 is missing from the non-productive reconstructions. In the non-productive reconstruction where RuvC is observed, only 13 bp of the R-loop are resolved (**Fig. 1c**). At lower thresholds, the PAM-distal R-loop can be seen sticking out away from the enzyme, hence why it was averaged out in the high-resolution reconstruction (**Extended Data Fig. 4**). In the other non-productive structure without RuvC density, we observe the R-loop in its entirety, including contacts formed between REC3 and the PAM-distal R-loop (**Fig. 1b**). REC3 residue Q801 contacts the gRNA phosphate backbone at position 15 of the R-loop. N791 similarly contacts the TS phosphate backbone, but between R-loop positions 17 and 18. Together, the non-productive structures of FnCas9 point to a checkpoint during enzyme activation

where approximately 100 amino acids in REC3 are disordered prior to docking onto the PAM-distal heteroduplex. We call this portion of REC3 the REC3 clamp.

The productive FnCas9 reconstructions from this dataset contain well-resolved HNH domains in the active conformation (**Fig. 1d-e**). Furthermore, the phosphodiester backbone of the TS is cleaved at the anticipated HNH endonuclease site three nucleotides upstream of the PAM. In addition to HNH, the REC3 clamp is well-resolved, and forms stabilizing contacts with the completed R-loop in both productive conformations. In the productive structure lacking RuvC density, we observe flexibility in the portion of REC3 outside of the clamp that typically interfaces with RuvC, although the REC3 clamp remains in contact with the PAM-distal heteroduplex. These productive structures support an activation mechanism where the REC3 clamp must dock onto the PAM-distal heteroduplex before HNH repositions into the active conformation.

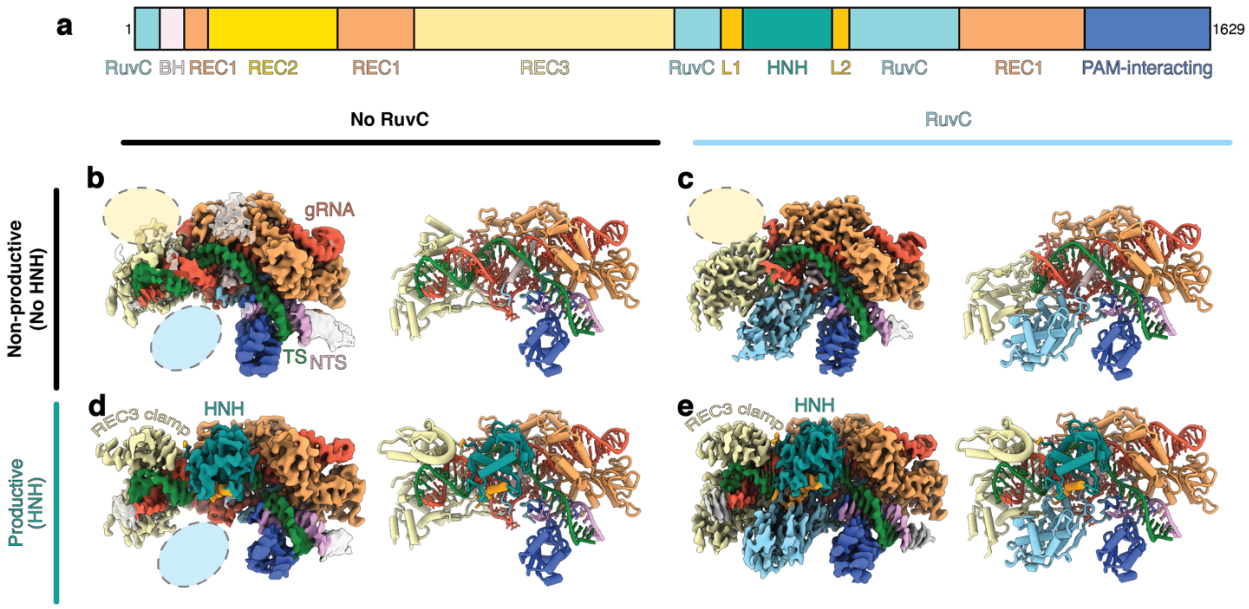


Fig. 1. Visualization of distinct stages of FnCas9 nuclease activation. a, FnCas9 domain organization. b, 2.9 Å cryo-EM reconstruction of FnCas9 in a non-productive state with no RuvC density resolved. Yellow oval highlights missing REC3 density, and blue oval highlights missing RuvC density. The corresponding models are depicted to the right of the cryo-EM reconstructions. c, 2.9 Å cryo-EM reconstruction of FnCas9 in a non-productive state with RuvC density. d, 3.0 Å cryo-EM reconstruction of FnCas9 in a productive state with HNH in the active conformation and no RuvC density. e, 2.6 Å cryo-EM reconstruction of FnCas9 in a productive state with RuvC density. The structures are colored as in (a) and as follows: TS, green; NTS, pink; gRNA, red.

The REC3 clamp confers FnCas9 specificity

SpCas9 is known to cleave off-target sequences, particularly when mismatches are located in the PAM-distal region[10,25,26]. In contrast, intial studies have established FnCas9 to be very sensitive to PAM-distal mismatches[15]. Even a single mismatch can abolish FnCas9 cleavage activity[15,27]. The underlying molecular mechanism of the intrinsic specificity of FnCas9 is not well characterized. To establish the mechanism of enhanced specificity displayed by FnCas9, we measured the cleavage kinetics of FnCas9 with gRNA and target DNA sequences containing a single mismatch at each position of the protospacer (**Fig. 2b**).

In comparison to perfectly matched DNA, substrates with PAM-distal mismatches had apparent cleavage rates up to 155-fold slower (**Extended Data Table 2**). Mismatches at TS positions 6, 7, and 11 displayed increased apparent cleavage rates in comparison to perfectly matched DNA. Structural analysis uncovered few protein contacts formed between FnCas9 and the TS at these positions enabling blind spots to mismatches. Although, close inspection of our productive FnCas9 structure with RuvC density revealed many contacts being formed between the REC3 clamp and the PAM-distal heteroduplex, especially where FnCas9 is sensitive to mismatches (**Fig. 2a**). REC3 clamp residues S792 and K788 coordinate one well-resolved water molecule in conjuction with the phosphodiester backbone of the TS. K789 forms electrostatic interactions with the TS backbone between positions 18 and 19. Of particular note, Y794 stacks upon the ribose at position 16 of the TS (**Fig. 2a & Extended Data Fig. 5**). To probe the relevance of this interaction, we prepared cryo-EM grids of FnCas9 in complex with gRNA and DNA containing a single mismatch at position 16 after a two hour incubation to ensure that the cleavage reaction had gone to completion. This cryo-EM dataset yielded two high-resolution structures of FnCas9 in both the productive and non-productive states (**Fig. 2c-d, Extended Data Fig. 6, and Extended Data Table 1**). We analyzed the distribution of particles between the two conformations and found only 30% of the particles were in the productive state where HNH had rearranged into its active conformation. In this state, Y794 is able to stack upon ribose at position 16 despite the mismatch within the R-loop. Although, the majority (70%) of the particles were in a non-productive state resembling our structure of FnCas9 with a perfectly matched target that was in the non-productive state with RuvC present. This conformation highlights the importance of the REC3 clamp in distinguishing on- from off-target sequences at the single mismatch scale. The REC3 clamp senses PAM-distal target complementarity, establishing an additional step during the stepwise process of nuclease activation.

To uncover the role of Y794 in FnCas9 mismatch discrimination, we mutated Y794 to alanine, and measured the rate of TS cleavage for on- and off-target substrates (**Fig. 2e**). While the observed cleavage rate for an on-target substrate was comprable to WT FnCas9, the observed cleavage rate for an off-target substrate with a mismatch at position 16 was 27-fold slower for the Y794A mutant than WT FnCas9. For all cleavage assays with the Y794A mutant, we observed less product formed, even after long time points. These findings indicate that the ribose stacking interaction between Y794 and position 16 of the TS is important for REC3 docking. Without Y794 able to stack upon the

ribose at position 16, about half of the Y794A molecules get trapped in a non-productive state where the REC3 clamp cannot dock and cleavage cannot occur. The other half of Y794A molecules are able to clamp onto the PAM-distal duplex favoring HNH rearrangement and DNA cleavage. We further deleted REC3 (Δ 450-858) and measured the rate of TS cleavage and observed no cleavage after two hours (**Fig. 2e**). Interestingly, we still observed binding of our REC3 deletion mutant to on-target DNA, although it is diminished in comparison to WT FnCas9 binding (**Fig. 2f**). These results establish REC3 docking as a prerequisite to nuclease activation. Together, our biochemical analyses highlight the significance of REC3 docking in FnCas9 target identification and activation.

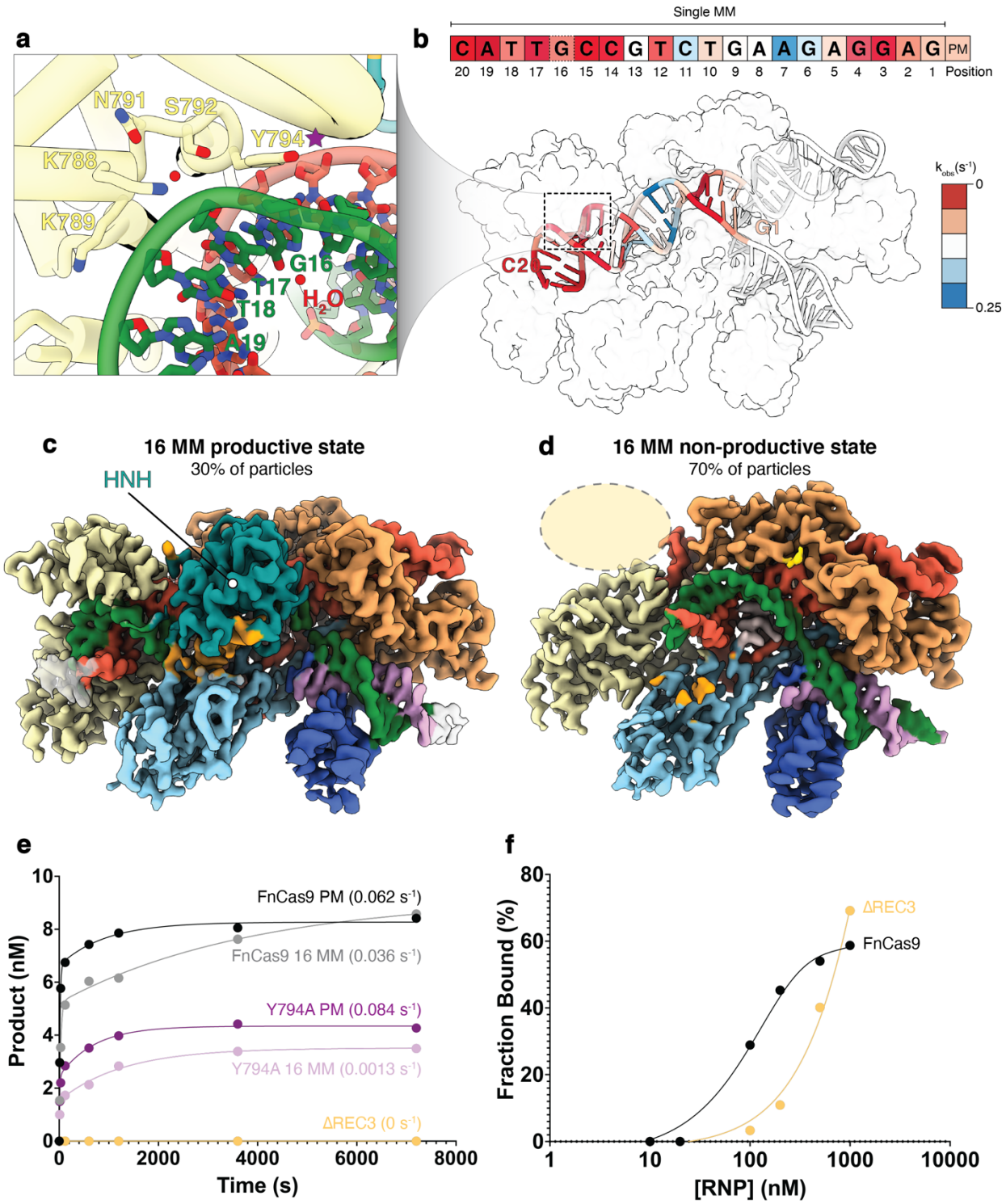


Fig. 2. The REC3 clamp enhances FnCas9 specificity. a, Detailed view of REC3 clamp residues contacting the PAM-distal heteroduplex. Purple star indicates Y794 which is subsequently mutated to alanine in e-f. b, Surface representation of FnCas9 in the product state with the nucleotides colored by observed cleavage rate for a single mismatch at that TS position. Observed cleavage rates were measured for each

mismatch via capillary electrophoresis. Red indicates a slower observed cleavage rate, and blue indicates a faster observed cleavage rate. c-d, FnCas9 was incubated with DNA containing a mismatch at position 16 of the TS for 1 hour. The reaction was quenched via vitrification and analyzed via cryo-EM. c, 30% of the particles (79,858 particles) from this dataset yielded a 2.9 Å cryo-EM reconstruction of FnCas9 in a productive state bound to DNA with a mismatch at position 16. d, 70% of the particle (128,469 particles) from this dataset comprised a 3.0 Å cryo-EM reconstruction of FnCas9 in a non-productive state bound to DNA with a mismatch at position 16. e, 100 nM of each FnCas9 variant was mixed with 25 nM of each substrate and the cleaved DNA product was monitored via capillary electrophoresis. Observed cleavage rates were calculated using KinTek Explorer. f, Various concentrations of preformed FnCas9:gRNA (RNP) were mixed with 10 nM of on-target DNA and allowed to equilibrate for two hours. Binding was measured by EMSA.

Structural basis of heightened virulence in *F. novicida*

The *F. novicida* CRISPR locus contains a non-canonical small RNA (scaRNA) that complexes with its tracrRNA (similarly to a crRNA) (**Extended Data Fig. 7**), but instead of targeting foreign genetic elements, the FnCas9:scaRNA:tracrRNA complex (scaRNA complex) targets its own genome[19,28] (**Fig. 3a**). More specifically, in the *F. novicida U112* strain, the scaRNA targets an endogenous bacterial lipoprotein (BLP) that is typically detected by host Toll-like receptor 2, stimulating inflammatory signaling. Interestingly, instead of targeting the BLP gene for cleavage, the scaRNA complex represses BLP expression by binding to the 5'UTR of genes encoding for the BLP, thereby subverting host detection. Limited spacer-protospacer complementarity prevents cleavage by the scaRNA complex, yet the molecular mechanisms dictating whether a target is repressed or cleaved by FnCas9 is not known (**Fig. 3b-c**).

To elucidate the mechanism by which FnCas9 differentiates targets to be repressed versus cleaved, we employed AlphaFold3 (AF3) to predict the structure of FnCas9 in complex with a gRNA mimicking the scaRNA sequence and the 5'UTR of *FTN_1101* (1101), one of the genes encoding the BLP[28,29]. To our surprise, AF3 outputted predicted structures with varying R-loop lengths. Three intermediate states dominated the five predicted AF3 models: one with a 5 bp R-loop, one with a 14 bp R-loop, and one with a 16 bp R-loop (**Fig. 3d**). The structures all depict targeting of the same region on the 1101 TS beginning with 5'-TTACA-3'. These structures illustrate that limited target complementarity prevents R-loop completion, consequently preventing the REC3 clamp from docking. Without the REC3 clamp properly docked onto the PAM-distal heteroduplex, the HNH domain cannot rearrange into the catalytically competent conformation. This is evidenced by the HNH linker 1 (L1) conformation, where residues 900-932 are folded in an inactive conformation due to the distorted scaRNA:1101 R-loop[10,30](**Fig. 3e,f,g**). Given a perfectly matched target, the R-loop propagates to completion, and L1 extends across the heteroduplex properly positioning HNH in the active conformation (**Fig. 3c**). Our analyses of experimental and simulated structures

suggest that the additional step during nuclease activation imposed by the REC3 clamp determines whether a target is repressed or cleaved by FnCas9.

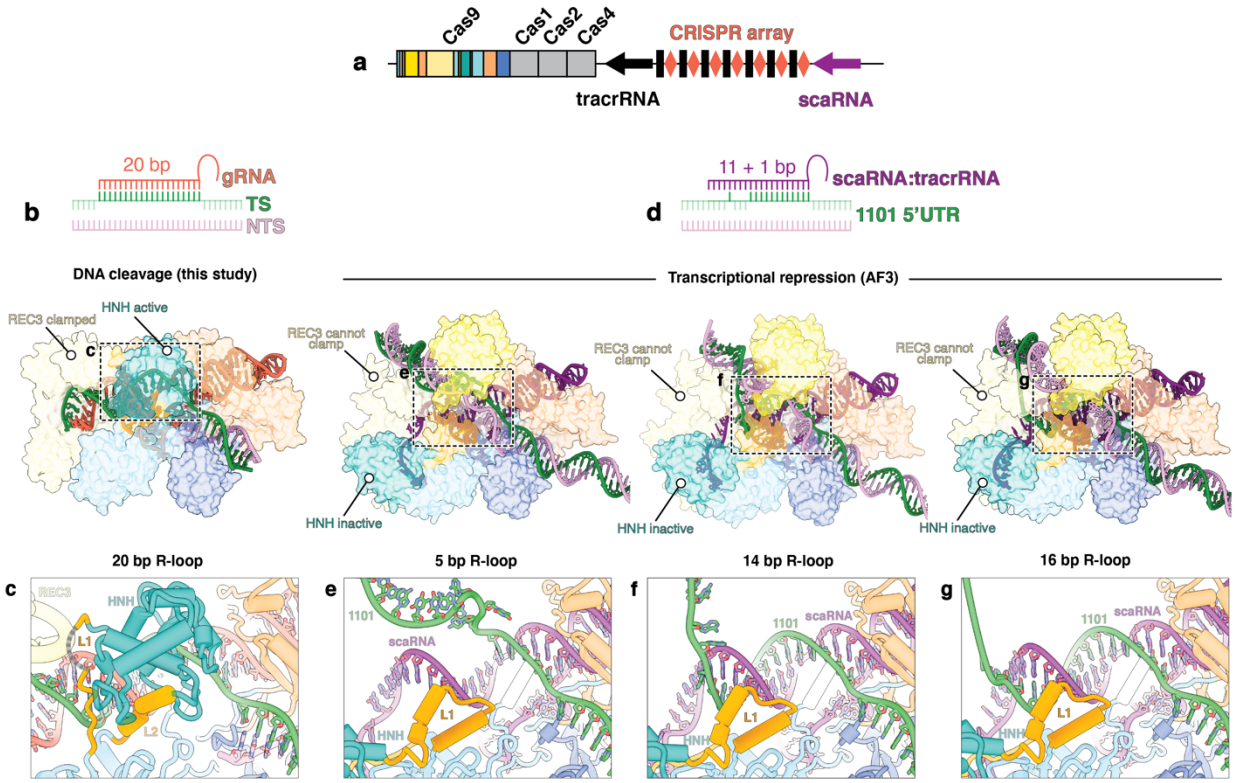


Fig. 3. The REC3 clamp senses target complementarity distinguishing DNA cleavage from transcriptional repression. a, Schematic of the *Francisella novicida* chromosomal locus consisting of Cas genes, tracrRNA, crRNA, and scaRNA. b, Schematic of canonical FnCas9 targeting where 20 bp of R-loop form between the gRNA and TS. c, Detailed view of L1 when the R-loop has propagated to completion and HNH is in the active conformation. d, Schematic of scaRNA-mediated targeting where 11 + 1 bp of complementarity exists to the target. AF3 predicted three distinct models of scaRNA-mediated targeting by FnCas9 with 5, 14, and 16 bp of R-loop formed. e, Detailed view of L1 in 5 bp R-loop intermediate. f, Detailed view of L1 in 14 bp R-loop intermediate. g, Detailed view of L1 in 16 bp R-loop intermediate. The *FTN_1101* NTS and REC3 are hidden in e, f, and g for clarity.

The REC3 clamp is conserved across type II-B CRISPR systems

Type II-B CRISPR-Cas systems typically contain larger Cas9 effectors than other type II systems[31]. This observation led us to question whether other type II-B Cas9 effectors have extended REC3 domains, and potentially similar activation mechanisms as FnCas9.

To test this theory, we performed a conservation analysis using ConSurf[32], where 51 type II-B Cas9 homologues were compared. Our conservation analysis revealed that REC3 clamp residues are highly conserved across type II-B CRISPR systems (**Fig. 4a-c**). More specifically, Q798 that intercalates between the gRNA and TS duplex is one of the most highly conserved residues. Additionally, S792 that coordinates a water molecule with the TS backbone, and Y794 that stacks upon the TS at position 16 are also highly conserved. Notably, the REC3 clamp residues are as highly conserved as the catalytic residues found in both the HNH and RuvC domains (**Fig. 4c**). The high conservation of REC3 residues bolsters their importance during the process of nuclease activation across type II-B CRISPR systems.

The conformational changes associated with Cas9 nuclease activation ensure that foreign genetic elements are targeted for cleavage, as opposed to the host chromosome. Here we find that type II-B CRISPR systems contain extended REC3 domains that sense PAM-distal mismatches encompassing an additional step during nuclease activation, ultimately hindering off-target cleavage (**Fig. 4d**). In addition to enhancing specificity, we also identify the REC3 clamp as responsible for the ability of FnCas9 to transcriptionally repress its own genome without detrimental cleavage. This unique feature enables subversion of host detection during infection by *F. novicida*. Taken together, the REC3 clamp found in type II-B Cas9 confers increased specificity and virulence. This attribute distinguishes type II-B systems from the more commonly studied type II CRISPR-Cas systems and provides a blueprint for more specific genome editors.

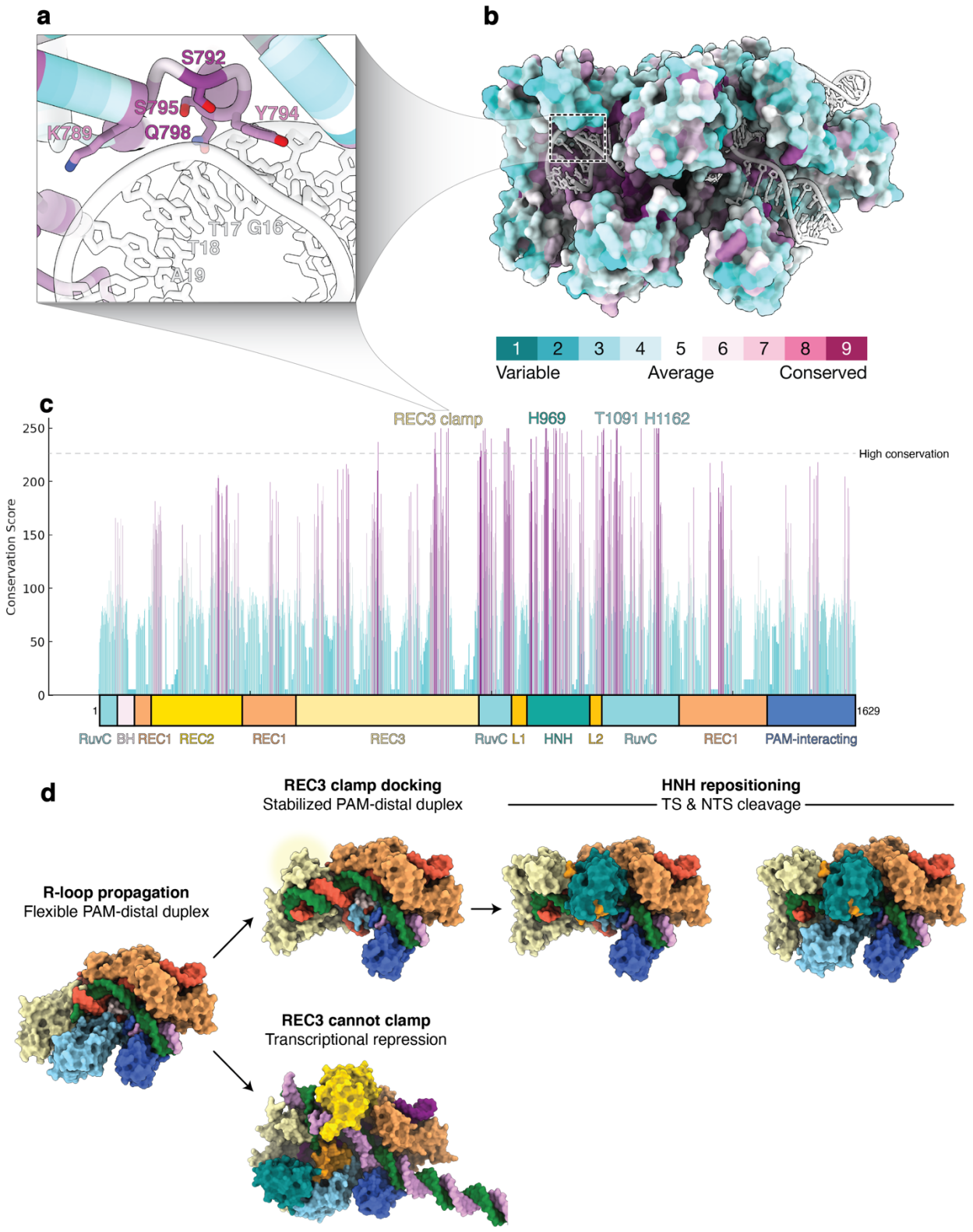

a
S792
S795
Y794
K789
Q798
T17
G16
T18
A19
b
1
2
3
4
5
6
7
8
9
Variable
Average
Conserved
c
REC3 clamp
H969
T1091
H1162
High conservation
Conservation Score
250
200
150
100
50
0
1
1629
RuvC
BH
REC1
REC2
REC1
REC3
RuvC
L1
HNH
L2
RuvC
REC1
PAM-interacting
d
REC3 clamp docking
Stabilized PAM-distal duplex
HNH repositioning
TS & NTS cleavage
R-loop propagation
Flexible PAM-distal duplex
REC3 cannot clamp
Transcriptional repression

Fig. 4. Type II-B CRISPR-Cas systems contain extended REC3 domains that add an additional checkpoint during enzyme activation. a, Detailed view of REC3 clamp residues contacting the PAM-distal heteroduplex colored by conservation score. The location of these residues is highlighted in b-c. b, Surface representation of FnCas9 in the product state where each residue is colored by conservation, where purple represents more conserved residues, and cyan denotes more variable residues. c, Domain schematic of FnCas9 with the corresponding conservation score for each residue. Residues with a conservation score of 225 or higher are considered highly conserved. d, Model of type II-B CRISPR-Cas system enzyme activation. If the R-loop propagates to completion, the REC3 clamp docks onto the PAM-distal heteroduplex prompting HNH repositioning and DNA cleavage. Alternatively, if REC3 cannot clamp due to incomplete R-loop formation, type II-B CRISPR-Cas systems repress transcription without the consequence of cleavage.

Discussion

Through kinetic analysis and structural determination, we resolved distinct stages of FnCas9 nuclease activation, revealing a novel step toward activation called REC3 clamping. This unique region of REC3 responsible for clamping onto the PAM-distal heteroduplex senses R-loop complementarity, inhibiting cleavage of off-target DNA. The location of the REC3 clamp makes FnCas9, and other type II-B CRISPR systems, particularly sensitive to PAM-distal mismatches. This differs from the more commonly used SpCas9, which has a “blind spot” to mismatches at positions 12-14 and stabilizes mismatches at positions 18-20[10](**Extended Data Fig. 8**). Additionally, SpCas9 tolerates single mismatches throughout the R-loop via diverse mechanisms[26]. Pinpointing the residues responsible for FnCas9 enhanced specificity provides a blueprint for merging domains from different Cas9 enzymes into an ideal genome editor. Since the REC3 clamp expands the REC3 domain, it would be judicious to add the FnCas9 REC3 domain onto a miniature Cas effector due to delivery size limitations. Given the recent advancements in *in silico* atomic modeling [29], one could test the architecture of said “Frankenstein” Cas effectors with ease. Beyond the newfound simplicity of enzyme engineering, evolution has already provided countless iterations of Cas enzymes to investigate in pursuit of the ideal genome editor.

Beyond its role in mismatch discrimination, the REC3 clamp may also differentiate targets destined for cleavage versus repression. This distinctive feature enables CRISPR-Cas systems with scaRNA located in their CRISPR loci to hinder expression of endogenous sequences that trigger a host immune response without the consequence of cleavage. In this way, scaRNA-mediated gene repression increases bacterial virulence[20,33,34]. Furthermore, most bacteria containing scaRNA are human pathogens, and exhibit resistance to common antibiotics[35]. Knocking out type II-B CRISPR genes in *Legionella pneumophila* increased susceptibility to antibiotics, illuminating a potential antibacterial approach, where an essential CRISPR component is targeted in combination with an

antibiotic. In our current age of rising antibiotic resistance and diminishing antibiotic discovery, it is crucial to pursue novel antibiotic approaches.

Our findings reveal the molecular mechanism enabling type II-B CRISPR-Cas systems to differentiate transcriptional repression of its own genome from DNA cleavage of foreign genetic material. Inhibiting the repressive function of type II-B CRISPR-Cas systems found in pathogenic bacteria may decrease virulence and increase sensitivity to antibiotics. By leveraging the unique properties of the REC3 clamp in FnCas9, we can potentially enhance genome editing specificity and develop novel antibacterial strategies that mitigate the threat of antibiotic resistance.

Methods

Protein expression and purification

FnCas9 was expressed from pET-His6-FnCas9GFP purchased from Addgene (Addgene plasmid #130966; http://n2t.net/addgene:130966; RRID:Addgene_130966)[15]. FnCas9 was then expressed and purified as previously described[15], except NiCo21(DE3) (New England Biolabs) were used instead of Rosetta2 (DE3). Additionally, after affinity purification with Ni-NTA beads, the lysate was exposed to TEV protease to liberate GFP. The remainder of the purification was executed as previously described.

FnCas9 Y794A mutant was cloned using Q5 Hot Start Hi-Fidelity polymerase and KLD kit (New England Biolabs). The mutated sequences were verified by Plasmidsaurus and purified in the same manner as WT FnCas9.

Nucleic acid preparation

DNA duplexes 55 nt long were prepared from PAGE-purified oligonucleotides synthesized by Integrated DNA Technologies as previously described[36]. The sequences of the synthesized oligonucleotides, including the positions of mismatches, are listed in **Extended Data Table 1**.

Buffer composition and kinetic reactions

Cleavage reactions were performed in 1X cleavage buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 10mM $MgCl_2$, 5% glycerol, 1 mM DTT) at 37 °C.

DNA cleavage kinetics

The reaction of FnCas9 with on- and off-target DNA was performed by preincubating FnCas9.gRNA (500 nM active-site concentration of FnCas9, 750 nM gRNA) in 5X cleavage buffer supplemented with 0.2 mg/mL molecular biology grade BSA for 10 minutes at room temperature. The reaction was initiated by the addition of 5'-FAM-labeled DNA duplex at the final concentrations of 10 nM DNA, and 100 nM preincubated enzyme. The reactions were carried out at 37° C and stopped at various times by mixing with 0.5 M EDTA. Reaction products were resolved and quantified using an Applied Biosystems DNA sequencer (ABI 3130xl) equipped with a 36 cm capillary array and nanoPOP7 polymer (MCLab)[37]. Data fit to equations were fit using either a single or double-exponential equations shown below:

Single exponential equation:

$$Y = A_1 e^{-\lambda_1 t} + C$$

where Y represents concentration of cleavage product, A_1 represents the amplitude, λ_1 represents the observed decay rate (eigenvalue) and C is the endpoint. The half-life was calculated as $t_{1/2} = \ln(2)/\lambda_1$.

Double exponential equation:

$$Y = A_1 e^{-\lambda_1 t} A_2 e^{-\lambda_2 t} + C$$

where Y represents concentration of cleavage product, A_1 represents the amplitude and λ_1 represents the observed rate for the first phase. A_2 represents the amplitude and λ_2 represents the observed rate for the second phase, and C is the endpoint.

Cryo-EM sample preparation, data collection, and processing

FnCas9 bound to different DNA substrates were assembled by mixing FnCas9 with gRNA in a 1:1.5 molar ratio and incubated at room temperature for 10 minutes in reaction buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 10 mM MgCl2, 5% glycerol, and 5 mM DTT). Each DNA substrate was then added in a 1:1 molar ratio with either 8 μM FnCas9 gRNP for the on-target DNA substrate, or 10μM FnCas9 gRNP for the off-target DNA substrate. The on-target complex was incubated for 1 hour at room temperature, and the off-target complex was incubated for 2 hours at room temperature. The reactions were quenched by vitrification. 2.5 μl of sample was applied to glow discharged holey carbon grids (Quantifoil 1.2/1.3), blotted for 6 s with a blot force of 0, and rapidly plunged into liquid nitrogen-cooled ethane using an FEI Vitrobot MarkIV.

All datasets were collected on an FEI Titan Krios cryo-electron microscope equipped with a K3 Summit direct electron detector (Gatan, Pleasanton, CA). Images were recorded with SerialEM v4.1[38] with a pixel size of 0.83 Å. Movies were recorded at 13.3 electrons/pixel/second for 6 s (80 frames) to give a total dose of 80 electrons/pixel. All datasets were collected with a defocus range of –1.5 to –2.5 μm. Motion correction, CTF estimation and particle picking was performed on-the-fly using cryoSPARC Live v4.0.0-privatebeta.2[39]. Further data processing was performed with cryoSPARC v.3.2. A total of 5,378 movies were collected for the on-target dataset, and 7,158 movies were collected for the off-target dataset.

The initial stages of processing were similar for both the on- and off-target datasets, where blob picker was used with a minimum particle diameter of 120 Å and a maximum particle diameter of 220 Å. The off-target dataset particles were then subjected to a single round of 2D classification. The particles selected from 2D classification were processed via ab initio reconstruction, followed by heterogeneous refinement. The on-target dataset particles were not subjected to 2D classification, but rather, went straight into ab initio reconstruction, followed by heterogeneous refinement. The best class from heterogeneous refinement was then fed to non-uniform refinement. The mask from the non-uniform refinement was used for 3D variability, which was displayed using the cluster

output mode, with 10 clusters, and a 5 Å filter resolution. The clusters were grouped by the presence of the HNH, RuvC, and REC3 domains. The particles corresponding to a single state were extracted with a 400-pixel box size for the on-target particles, and a 384-pixel box size for the off-target particles. The extracted particles were then globally and locally CTF refined, and fed to non-uniform refinement which yielded the final reconstructions.

Structural model building and refinement

Inactive FnCas9 (Protein Data Bank (PDB) code: 5B2O) was used as a starting model for the on-target FnCas9 structure with RuvC and HNH present. The HNH domain from the starting model was removed and replaced through rigid body fitting with a model of the HNH domain folded by AlphaFold2[40]. HNH L1 and L2 were built, and nucleic acid alterations were made in Coot v1.1.07[41]. Further modeling was performed using Isolde v1.6[42]. The model was ultimately subjected to real-space refinement implemented in Phenix v1.21[43]. The remaining on-target structures used the FnCas9 structure with RuvC and HNH present as a starting model. The productive off-target structure also used the FnCas9 structure with RuvC and HNH present as a starting model, but the non-productive off-target structure used the non-productive structure of FnCas9 with RuvC present as its starting model. These structures were then subjected to a similar refinement workflow, where nucleic acid alterations were made in Coot v1.1.07[41], modeling was done in Isolde v1.6[42], and ultimately subjected to real-space refinement implemented in Phenix v1.21[43]. All structural figures and videos were generated using ChimeraX v1.2[44]. The Cas9 sequence from *F. novicida U112* was supplied to the AlphaFold3 web server[29], along with a gRNA mimicking scaRNA and a 72 bp DNA comprising the 5'UTR of *FTN_1101*. Five models were outputted with one 5 bp R-loop, one 14 bp R-loop, and three 16 bp R-loop intermediates. The majority of all five structures were predicted with very high confidence (plDDT > 90), the 5 bp R-loop was predicted with confidence (90 > plDDT > 70), and the final bp of the 14 and 16 bp R-loops were predicted with lower confidence (plDDT < 70).

Conservation analysis

The on-target FnCas9 structure with RuvC and HNH present was given to ConSurf[32] for conservation analysis. Homologues were collected from the UniProt[45] database via a HMMR search[46]. The E-value cutoff was set to 0.001 with three iterations. The CD-HIT cutoff was set to 99%, maximum number of homologues to 150, and minimum sequence identity to 10%. The minimum query sequence coverage was set to 60%. A multiple sequence alignment was then built using MAFFT, and conservation scores calculated by the Bayesian method. The conservation scores were then mapped onto the on-target FnCas9 structure with RuvC and HNH present. The bar graph depicting amino acid position versus conservation score was made using Jalview[47].

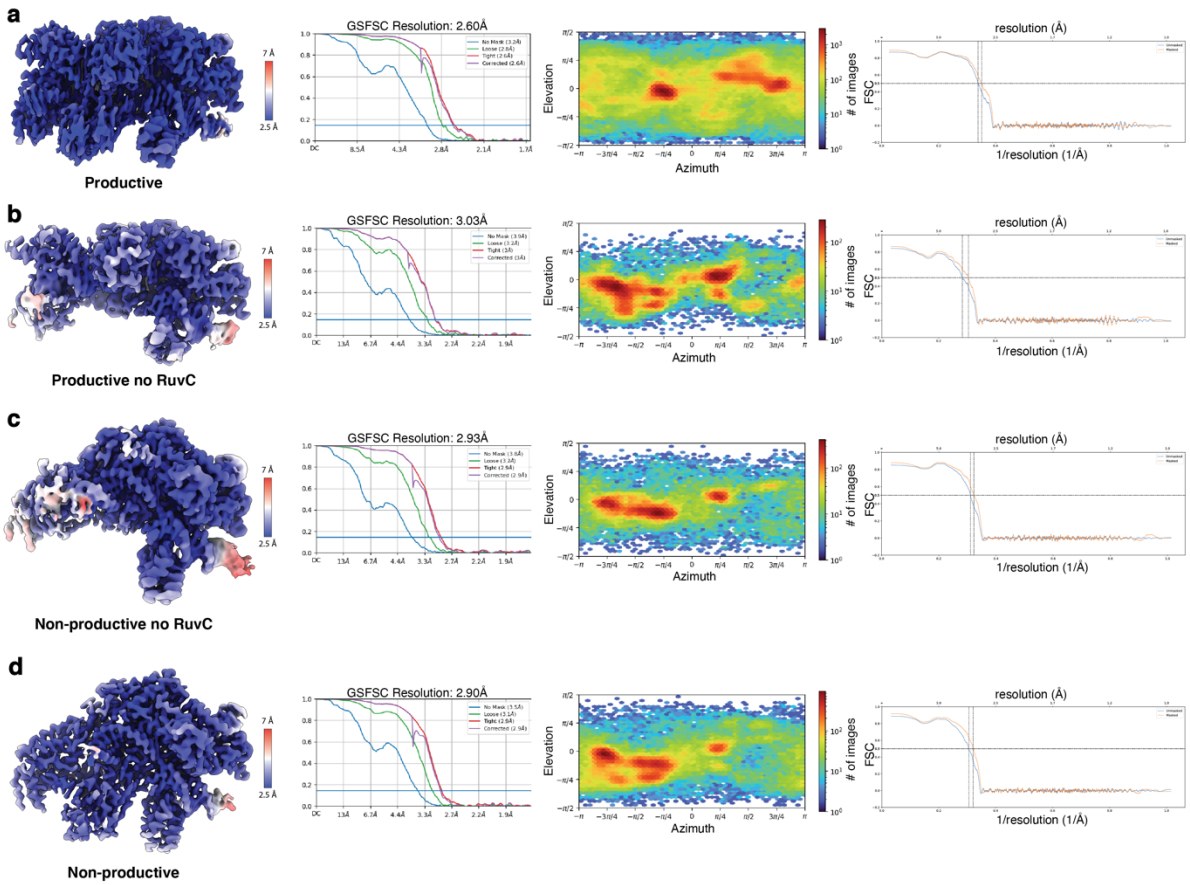
References

1. Sorek, R., Lawrence, C. M. & Wiedenheft, B. CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu Rev Biochem* **82**, 237–266 (2013).
2. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (1979)* **337**, 816–821 (2012).
3. Li, T. *et al.* CRISPR/Cas9 therapeutics: progress and prospects. *Signal Transduction and Targeted Therapy 2023 8:1* **8**, 1–23 (2023).
4. Philippidis, A. CASGEVY Makes History as FDA Approves First CRISPR/Cas9 Genome Edited Therapy. *Hum Gene Ther* **35**, 1–4 (2024).
5. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology 2013 31:9* **31**, 822–826 (2013).
6. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature 2020 578:7794* **578**, 229–236 (2020).
7. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature 2015 529:7587* **529**, 490–495 (2016).
8. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature 2017 550:7676* **550**, 407–410 (2017).
9. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science (1979)* **351**, 84–88 (2016).
10. Bravo, J. P. K. *et al.* Structural basis for mismatch surveillance by CRISPR–Cas9. *Nature 2022 603:7900* **603**, 343–347 (2022).
11. Slaymaker, I. M. & Gaudelli, N. M. Engineering Cas9 for human genome editing. *Curr Opin Struct Biol* **69**, 86–98 (2021).
12. Liu, M.-S. *et al.* Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Communications 2020 11:1* **11**, 1–13 (2020).
13. Kim, N. *et al.* Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nature Biotechnology 2020 38:11* **38**, 1328–1336 (2020).
14. Kim, Y. hoon *et al.* Sniper2L is a high-fidelity Cas9 variant with high activity. *Nature Chemical Biology 2023 19:8* **19**, 972–980 (2023).
15. Acharya, S. *et al.* Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing. *Proc Natl Acad Sci U S A* **116**, 20959–20968 (2019).
16. Sternberg, S. H., Lafrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature 2015 527:7576* **527**, 110–113 (2015).
17. Singh, D. *et al.* Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nature Structural & Molecular Biology 2018 25:4* **25**, 347–354 (2018).
18. Jones, C. L. & Weiss, D. S. TLR2 Signaling Contributes to Rapid Inflammasome Activation during F. novicida Infection. *PLoS One* **6**, e20609 (2011).

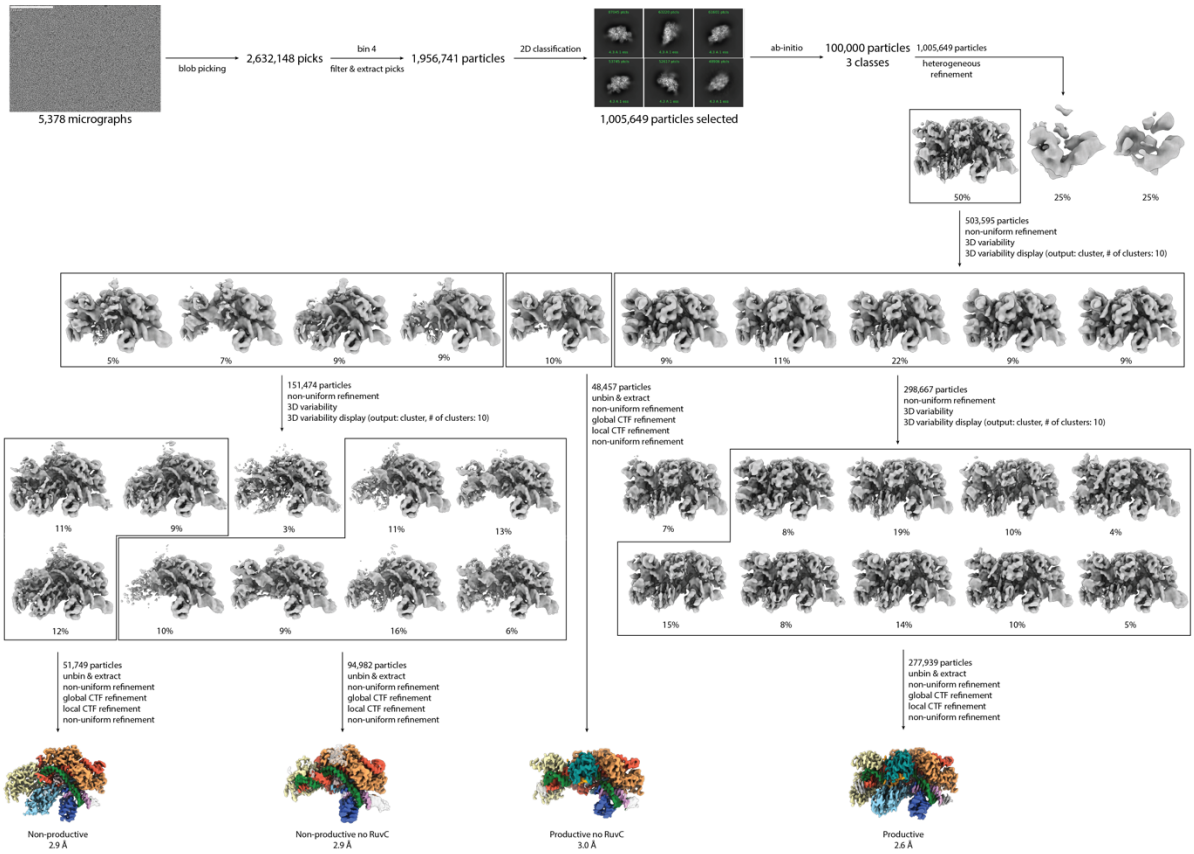
19. Sampson, T. R., Saroj, S. D., Llewellyn, A. C., Tzeng, Y. L. & Weiss, D. S. A CRISPR-CAS System Mediates Bacterial Innate Immune Evasion and Virulence. *Nature* **497**, 254 (2013).
20. Ratner, H. K. *et al.* Catalytically Active Cas9 Mediates Transcriptional Interference to Facilitate Bacterial Virulence. *Mol Cell* **75**, 498-510.e5 (2019).
21. Workman, R. E. *et al.* A natural single-guide RNA repurposes Cas9 to autoregulate CRISPR-Cas expression. *Cell* **184**, 675-688.e19 (2021).
22. Hirano, H. *et al.* Structure and Engineering of Francisella novicida Cas9. *Cell* **164**, 950–961 (2016).
23. Sun, W. *et al.* Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. *Mol Cell* **76**, 938 (2019).
24. Pacesa, M. *et al.* R-loop formation and conformational activation mechanisms of Cas9. *Nature 2022 609:7925* **609**, 191–196 (2022).
25. Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S. & Yang, S. H. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
26. Pacesa, M. *et al.* Structural basis for Cas9 off-target activity. *Cell* **185**, 4067-4081.e21 (2022).
27. Azhar, M. *et al.* Rapid and accurate nucleobase detection using FnCas9 and its application in COVID-19 diagnosis. *Biosens Bioelectron* **183**, 113207 (2021).
28. Ratner, H. K. *et al.* Catalytically Active Cas9 Mediates Transcriptional Interference to Facilitate Bacterial Virulence. *Mol Cell* **75**, 498-510.e5 (2019).
29. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature 2024 630:8016* **630**, 493–500 (2024).
30. Zhang, Y. *et al.* Catalytic-state structure and engineering of Streptococcus thermophilus Cas9. *Nature Catalysis 2020 3:10* **3**, 813–823 (2020).
31. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 6091 (2014).
32. Yariv, B. *et al.* Using evolutionary data to make sense of macromolecules with a “face-lifted” ConSurf. *Protein Science* **32**, e4582 (2023).
33. 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 2013 497:7448* **497**, 254–257 (2013).
34. Djordjevic, M. *et al.* In silico analysis suggests common appearance of scaRNAs in type II systems and their association with bacterial virulence. *Front Genet* **9**, 17 (2018).
35. Xu, P. X. *et al.* Distribution characteristics of the Legionella CRISPR-Cas system and its regulatory mechanism underpinning phenotypic function. *Infect Immun* **92**, (2024).
36. Hibshman, G. N. *et al.* Unraveling the mechanisms of PAMless DNA interrogation by SpRY-Cas9. *Nature Communications 2024 15:1* **15**, 1–15 (2024).
37. Dangerfield, T. L., Huang, N. Z. & Johnson, K. A. High throughput quantification of short nucleic acid samples by capillary electrophoresis with automated data processing. *Anal Biochem* **629**, 114239 (2021).

38. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36–51 (2005).
39. Punjani, A. Real-time cryo-EM structure determination. *Microscopy and Microanalysis* **27**, 1156–1157 (2021).
40. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature 2021 596:7873* **596**, 583–589 (2021).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486 (2010).
42. Croll, T. I. ISOLDE: A physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol* **74**, 519–530 (2018).
43. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213 (2010).
44. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science* **30**, 70–82 (2021).
45. Bateman, A. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158–D169 (2017).
46. Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**, W29 (2011).
47. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

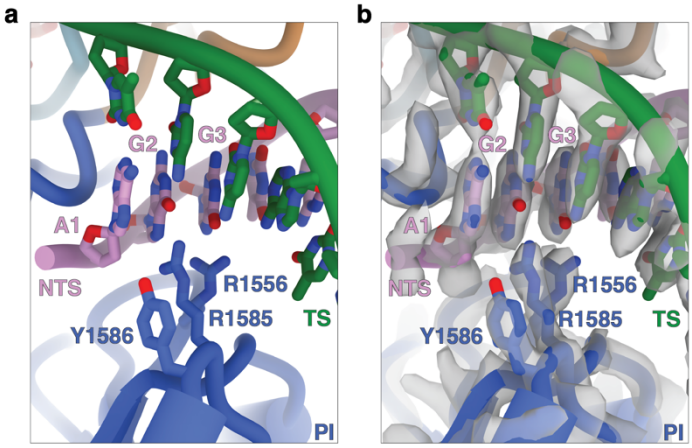
Extended Data Figures



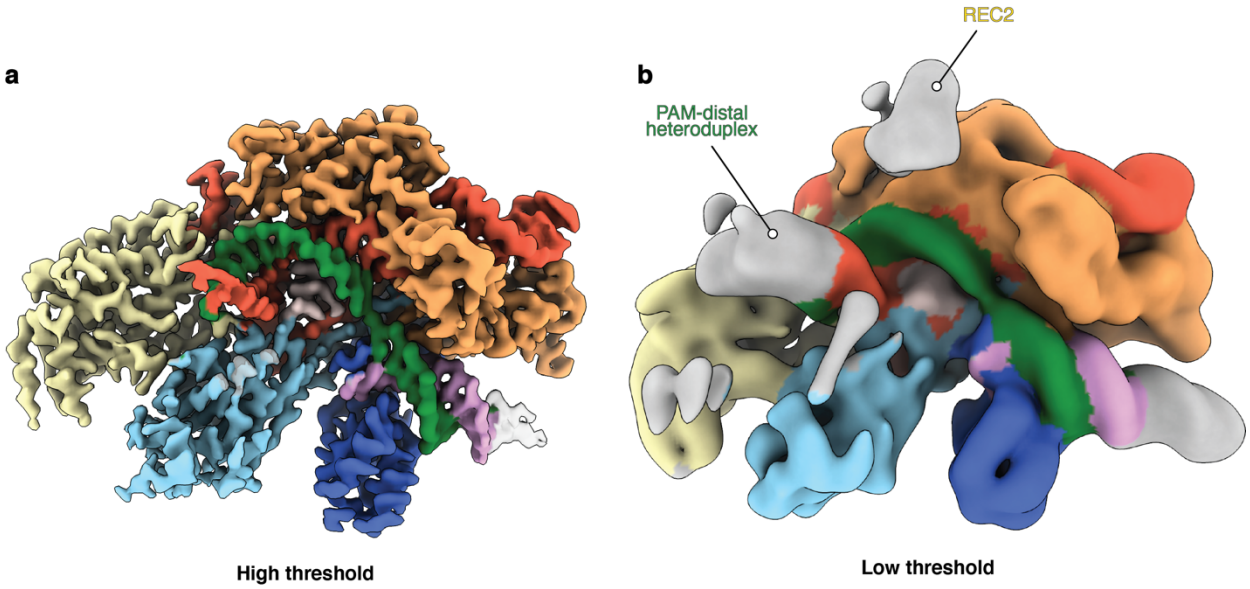
Extended Data Figure 1: Cryo-EM data analysis of FnCas9 bound to PM DNA in the a, productive b, productive no RuvC c, non-productive no RuvC and d, non-productive states.



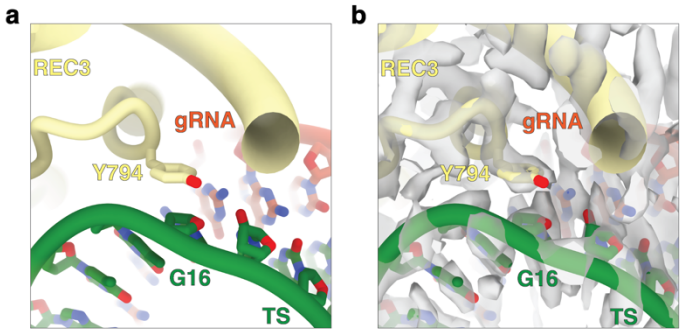
Extended Data Figure 2: Cryo-EM data processing workflow for FnCas9 bound to PM DNA after a 1-hour incubation at room temperature. Final reconstructions are colored by domain.



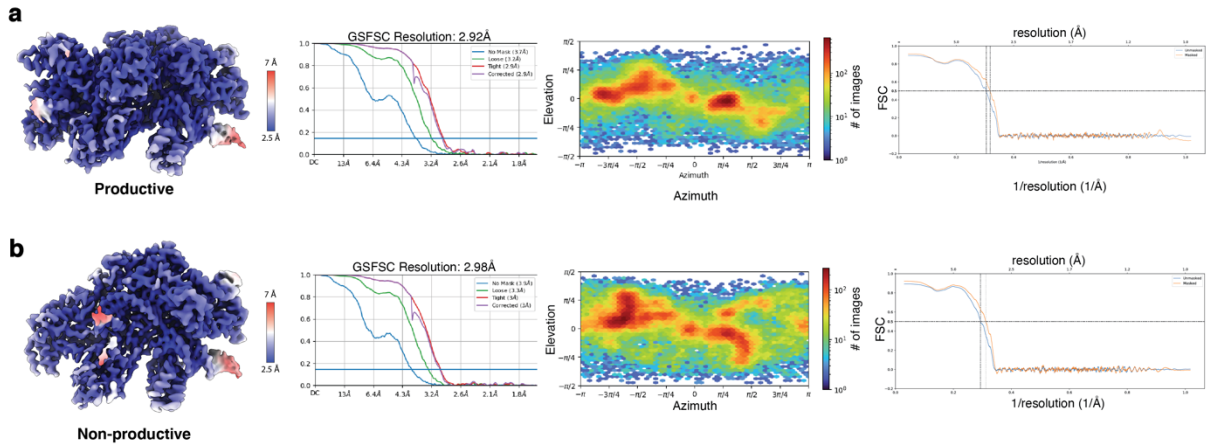
Extended Data Figure 3: FnCas9 residues R1556 and R1585 recognize the NGG PAM. a, Detailed view of the PAM site interaction when FnCas9 is in the product state. b, Cryo-EM density of the FnCas9 PAM site interaction.



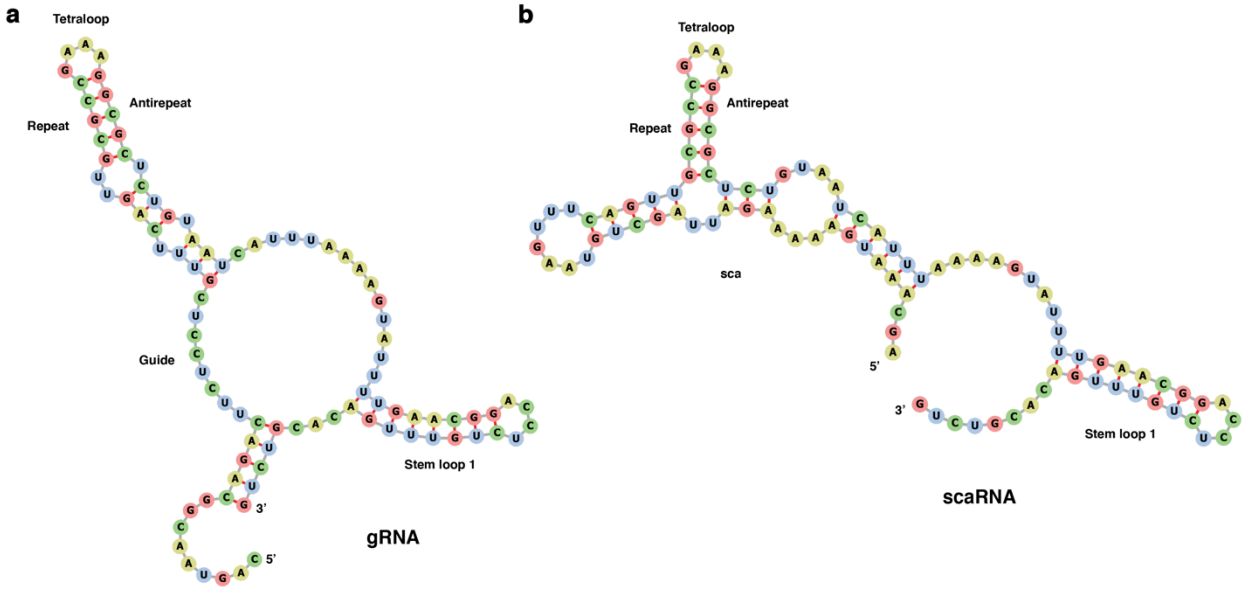
Extended Data Figure 4: The PAM-distal heteroduplex is flexible prior to REC3 docking. a, 2.9 Å cryo-EM reconstruction of FnCas9 in a non-productive conformation at a high threshold where the PAM-distal heteroduplex is averaged out, and b, at a low threshold where the PAM-distal heteroduplex is resolved.



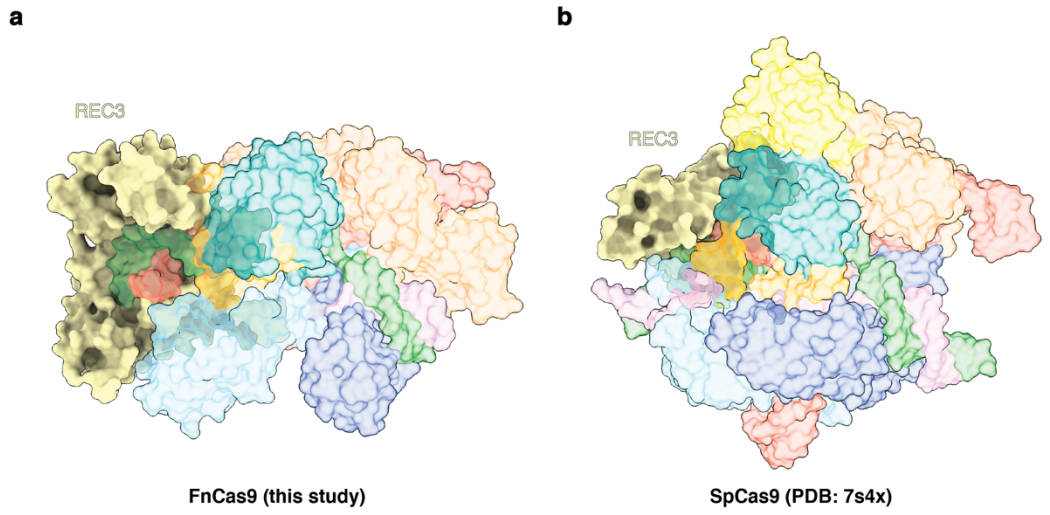
Extended Data Figure 5: REC3 clamp residue Y794 stacks upon the ribose at position 16 of the TS. a, Detailed view of the REC3 ribose stacking interaction when FnCas9 is in the product state. b, Cryo-EM density for the REC3 ribose stacking interaction at 2.6 Å resolution.



Extended Data Figure 6: Cryo-EM data analysis of FnCas9 bound to 16 MM DNA in the a, productive and b, non-productive states.



Extended Data Figure 7: Comparison between FnCas9 a, gRNA and b, scaRNA single guide mimic secondary structures.



Extended Data Figure 8: FnCas9 and SpCas9 REC3 domain comparison. a, FnCas9 residues 450-858 comprise the REC3 domain. b, SpCas9 residues 496-718 comprise the REC3 domain. REC3 domains are opaque.

	FnCas9 PM Productive RuvC (EMDB-xxxx) (PDB xxxx)	FnCas9 PM Productive No RuvC (EMDB-xxxx) (PDB xxxx)	FnCas9 PM Non-Productive RuvC (EMDB-xxxx) (PDB xxxx)	FnCas9 PM Non-Productive No RuvC (EMDB-xxxx) (PDB xxxx)	FnCas9 16MM Productive (EMDB-xxxx) (PDB xxxx)	FnCas9 16MM Non-Productive (EMDB-xxxx) (PDB xxxx)
Data collection and processing						
Voltage (kV)	300 kV	300 kV	300 kV	300 kV	300 kV	300 kV
Electron exposure (e–/Å^2)	80	80	80	80	80	80
Defocus range (μm)	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5	-1.5 to -2.5
Pixel size (Å)	0.8332	0.8332	0.8332	0.8332	0.8332	0.8332
Symmetry imposed	C1	C1	C1	C1	C1	C1
Initial particle images (no.)	1,956,741	1,956,741	1,956,741	1,956,741	2,495,949	2,495,949
Final particle images (no.)	277,939	48,457	51,749	94,982	79,858	63,288
Map resolution (Å)	2.6	3.0	2.9	2.9	2.9	3.0
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5
Map resolution range (Å)	2.5-7	2.5-7	2.5-7	2.5-7	2.5-7	2.5-7
Refinement						
Initial model used (PDB code)	5B2O	FnCas9 PM Productive RuvC	FnCas9 PM Productive RuvC	FnCas9 PM Productive No RuvC	FnCas9 PM Productive RuvC	FnCas9 PM Non-Productive RuvC
Model resolution (Å)	2.8	3.0	3.1	3.1	3.1	3.2
FSC threshold	0.5	0.5	0.5	0.5	0.5	0.5
Model resolution range (Å)						
Map sharpening *B* factor (Å^2)	70.2	67.6	72.5	80.1	79.3	78.2
Model composition						
Non-hydrogen atoms	13,904	12,096	11,138	10,192	13,901	11,136
Protein residues	1,356	1,125	1,057	891	1,356	1,057
Nucleotides	132	132	116	132	132	116
Ligands	2 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}	2 Mg^{2+}
R.m.s. deviations						
Bond lengths (Å)	0.005	0.005	0.005	0.005	0.005	0.005
Bond angles (°)	1.019	0.994	1.033	1.058	1.122	1.126
Validation						
MolProbity score	1.11	1.03	0.98	1.11	0.98	0.96
Clashscore	2.23	2.45	2.03	3.17	2.11	1.98
Poor rotamers (%)	0.66	0.68	0.21	0.98	0.66	0.84
Ramachandran plot						
Favored (%)	97.47	98.56	97.99	98.63	98.07	98.95
Allowed (%)	2.53	1.44	2.01	1.37	1.93	1.05
Disallowed (%)	0.00	0.00	0.00	0.00	0.00	0.00

Extended Data Table 1: Cryo-EM data collection, refinement and validation statistics.

DNA substrate	k_{obs} (s^{-1})
PM	0.066
1 MM	0.074
2MM	0.045
3 MM	0.004
4 MM	0.007
5 MM	0.078
6 MM	0.149
7 MM	0.222
8 MM	0.102
9 MM	0.099
10 MM	0.086
11 MM	0.148
12 MM	0.027
13 MM	0.101
14 MM	0.012
15 MM	0.001
16 MM	0.036
17 MM	0.001
18 MM	0.034
19 MM	0.001
20 MM	0.001

Extended Data Table 2: FnCas9 observed cleavage rates for perfect match, and single mismatch (MM) DNA substrates.