Article

Structural basis for mismatch surveillance by CRISPR–Cas9

Jack P. K. Bravo[1,5], Mu-Sen Liu[1,5], Grace N. Hibshman[1,2], Tyler L. Dangerfield[1,2], Kyungseok Jung[1], Ryan S. McCool[1,2], Kenneth A. Johnson[1,2 ✉] & David W. Taylor[1,2,3,4 ✉]

CRISPR–Cas9 as a programmable genome editing tool is hindered by off-target DNA cleavage[1-4], and the underlying mechanisms by which Cas9 recognizes mismatches are poorly understood[5-7]. Although Cas9 variants with greater discrimination against mismatches have been designed[8-10], these suffer from substantially reduced rates of on-target DNA cleavage[5,11]. Here we used kinetics-guided cryo-electron microscopy to determine the structure of Cas9 at different stages of mismatch cleavage. We observed a distinct, linear conformation of the guide RNA–DNA duplex formed in the presence of mismatches, which prevents Cas9 activation. Although the canonical kinked guide RNA–DNA duplex conformation facilitates DNA cleavage, we observe that substrates that contain mismatches distal to the protospacer adjacent motif are stabilized by reorganization of a loop in the RuvC domain. Mutagenesis of mismatch-stabilizing residues reduces off-target DNA cleavage but maintains rapid on-target DNA cleavage. By targeting regions that are exclusively involved in mismatch tolerance, we provide a proof of concept for the design of next-generation high-fidelity Cas9 variants.

For therapeutic applications of CRISPR–Cas9, off-target DNA cleavage must be minimized[1-3]. Although a variety of high-fidelity Cas9 variants with improved mismatch discrimination have been developed[7,9], their enhanced specificity comes at the cost of severely reduced rates of on-target DNA cleavage[5,11]. Mismatches induce alternative Cas9 conformations[12,13]; however, the structures used to guide rational redesign of such variants were bound to on-target DNA and in inactive conformations[14,15]. To understand the molecular mechanisms that govern off-target recognition, here we used kinetic analysis to guide sample preparation for cryo-electron microscopy (cryo-EM) and obtained structural snapshots of Cas9 pre-cleavage activation intermediates in the presence of various guide RNA–DNA target strand (gRNA–TS) mismatches.

Kinetics of Cas9 on mismatched DNA

We measured the rates of target strand cleavage by Cas9 in the presence of contiguous triple nucleotide mismatches at different positions along the gRNA–TS duplex (Extended Data Fig. 1a, Extended Data Table 1). Compared to rapid on-target cleavage (around 1.0 s^{-1}) the well-characterized protospacer adjacent motif (PAM)-distal 18–20 MM[5,9,12,13] (three mismatches 18–20 bp distal from the PAM) caused a reduction in rate of around 40-fold. Other mismatches (6–8 MM, 9–11 MM and 15–17 MM) resulted in a greater-than-2,000-fold reduction in cleavage rates, with only 20% of the DNA cleaved after 2 h of incubation (Extended Data Fig. 1b).

Notably, the 12–14 MM allowed Cas9 activation but with rates around 10-fold slower than those of the 18–20 MM. Although Cas9 cleavage is markedly slower for both 12–14 MM-and 18–20 MM-containing DNA than for on-target DNA, more than 80% of either substrate was cleaved within an hour of incubation with Cas9. This time frame for off-target cleavage poses problems for genome-editing applications, which typically occur on the time scale of days to weeks[16].

Structures of Cas9 with mismatched DNA

To understand the structural basis for Cas9 activation of mismatched DNA, we vitrified Cas9 with 12–14 MM DNA after a 5-min reaction, in which only around 10% of DNA was cleaved (Extended Data Table 2). We determined a cryo-EM structure at a global resolution of 3.6 Å (Fig. 1a, Extended Data Fig. 2, Extended Data Table 3). The target-strand-cleaving HNH endonuclease domain was not observed, indicating conformational heterogeneity before activation[17,18]. Of note, the distal end of the gRNA–TS duplex was in a linear conformation relative to the PAM-proximal DNA–DNA duplex—a state that differs from previously determined on-target DNA-bound Cas9 structures that depict a kinked duplex (around 70°)[14,18], although this state is reminiscent of early R-loop formation intermediates[19].

We then vitrified samples of Cas9 with 12–14 MM DNA after a 1-h incubation in which around 80% of the DNA was cleaved (Fig. 1b). Two distinct conformations were observed: a linear duplex conformation consistent with the 5-min structure of 12–14 MM and the kinked duplex conformation described above (Fig. 1a, c). The Cas9 conformations in the two 12–14 MM structures are identical (Fig. 2), but the PAM-distal gRNA–TS duplex end was shifted by around 30 Å and stably docked with REC3 (Fig. 2c). We propose that the linear duplex

[1]Department of Molecular Biosciences, University of Texas at Austin, Austin, TX, USA. [2]Interdisciplinary Life Sciences Graduate Programs, University of Texas at Austin, Austin, TX, USA. [3]Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, TX, USA. [4]Livestrong Cancer Institutes, Dell Medical School, University of Texas at Austin, Austin, TX, USA. [5]These authors contributed equally: Jack P. K. Bravo, Mu-Sen Liu. ✉e-mail: kajohnson@utexas.edu; dtaylor@utexas.edu
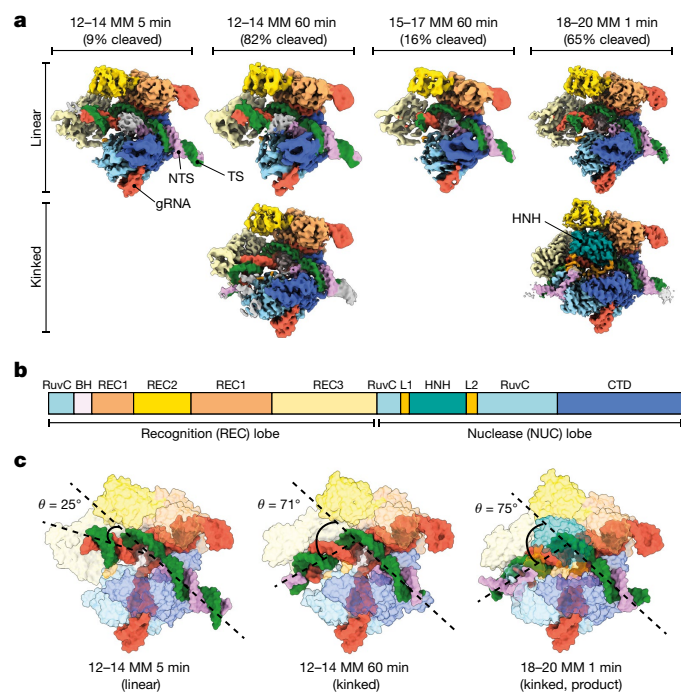
Fig. 1 | Mismatch-induced Cas9 conformational intermediates. a, Cryo-EM reconstructions of Cas9 in complex with various partially mismatched DNA substrates, determined at nominal resolutions ranging from 2.8 to 3.6 Å. Cryo-EM structures are coloured according to the domain map for Cas9. Nucleotides are coloured: target strand (TS), green; NTS, pink; and gRNA, red. The fraction of target strand DNA cleaved by Cas9 containing contiguous triple mismatches at the position and time point used for structural determination is shown above each structure. **b**, Domain organization of SpCas9. CTD, C-terminal domain. **c**, Models of Cas9 in complex with mismatched DNA substrates shown as isosurface representations. The angle between PAM-proximal and PAM-distal duplexes (θ) is shown. θ is equivalent to around 25° for all linear conformations observed.

Fig. 2 | Positions 12–14 of the gRNA–TS duplex occupy a blind spot for REC3 mismatch detection. a, **b**, Structures of 12–14 MM at 5 min (**a**) and 1 h (**b**) in linear and kinked conformations, respectively. The position of the 12–14 MM is shown as light green and light pink for the gRNA and the target strand, respectively. Models are shown as isosurface representations. **c**, Conformational change of the PAM-distal gRNA–TS duplex. The Cas9 protein structure is largely unchanged (root-mean-square deviation (RMSD) of less than 2 Å for equivalent C-alpha atoms), but the PAM-distal gRNA–TS duplex end undergoes a 30 Å conformational change, docking with REC3. **d**, Close-up view of positions 12–14, showing that because of the phase of the gRNA–TS duplex, REC3 makes no contacts with these base pairs. **e**, Schematic of interactions between REC3 and positions 9–17 of the gRNA–TS duplex. No interactions occur between Cas9 REC3 and positions 12–14 MM. Position 1 of the duplex is the first base of the target strand that hybridizes with the gRNA spacer.

conformation corresponds to an early intermediate of Cas9, before HNH rearrangement and docking to cleave the DNA[9,18]. This is supported by recent structural analyses of catalytically dead Cas9 in complex with various R-loop formation intermediates, several of which exhibit linear gRNA–TS duplex conformations that are similar to our linear duplex structures[20].

Notably, positions 12–14 of the gRNA–TS make no direct contacts with the REC3 domain of Cas9 (Fig. 2). Although positions 9–11 and 15–17 make considerable contacts with REC3, the alignment of the gRNA–TS duplex leaves positions 12–14 without any engagement with this domain (Fig. 2d, e). Because REC3 has a critical role in sensing PAM-distal mismatches[9], the 12–14 MM is likely to be able to evade mismatch discrimination by REC3 as it is positioned in a blind spot.

We reasoned that mismatches that prevent the PAM-distal gRNA–TS duplex from docking on REC3 would be unable to assume the kinked conformation, leading to considerably reduced DNA cleavage. To test this hypothesis, we determined a structure of Cas9 with 15–17 MM double-stranded DNA (dsDNA) substrate after 1 h of incubation with the enzyme (Fig. 1b). This mismatch inhibits cleavage by Cas9, but still permits DNA binding as measured by high-throughput profiling[21]. We observed only the linear duplex conformation (Fig. 1a, c). These structures support a model in which a linear duplex conformation precedes the canonical kinked duplex conformation that is required for activation, and mismatches that block formation of the kinked conformation escape DNA cleavage by Cas9.
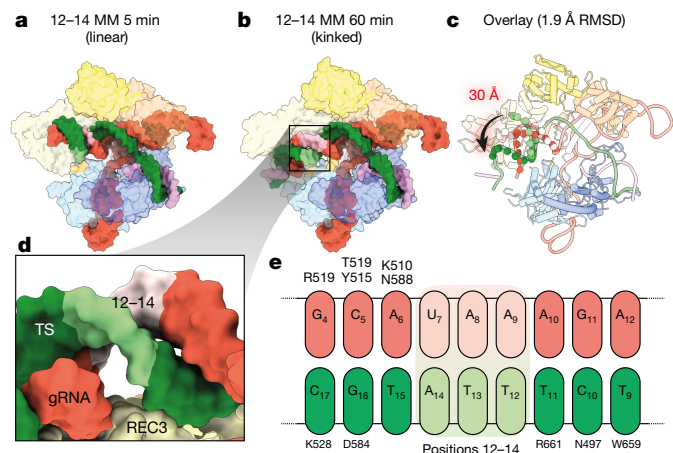
The 18–20 mismatch supports Cas9 activation

We next sought to understand how certain mismatches can evade Cas9 discrimination to allow more efficient Cas9 activation and DNA cleavage relative to other mismatches. We examined Cas9 after incubation with 18–20 MM DNA at the 1-min time point at which around 65% of the DNA was cleaved (Extended Data Fig. 1b), to determine whether this more tolerated mismatch undergoes the same structural transition as that of 12–14 MM DNA. Consistent with the fraction of product formation, we observed a mixed population of particles including the linear (Fig. 1a, c) and the kinked duplex conformation. In the kinked duplex structure, we observed HNH docked at the target site scissile phosphate, indicating the fully active conformation. This arrangement of HNH is entirely consistent with the previously observed active Cas9 conformation[12,18]. These results suggest that the population of particles showing a linear conformation represents an early intermediate in the pathway, and that the kinking of the gRNA–TS duplex is linked to HNH docking.

We observed target strand cleavage between nucleotides 3 and 4 (Fig. 3, Extended Data Fig. 3) and non-target strand (NTS) cleavage at the canonical site three bases upstream from the PAM. We report a direct observation of an RuvC active site with the non-target strand bound in the product state (Fig. 3, Extended Data Fig. 3). R986 is in the 'down' conformation, stabilizing the two magnesium ions as predicted by molecular dynamics simulations[22] (Fig. 3), whereas F916 wedges between the −2 and −3 bases through stacking interactions and positions the −3 position within the RuvC active site. These observations are in agreement with previous structural and mutagenesis studies[23,24]. Our structure suggests a histidine-mediated catalytic mechanism, consistent with two-metal-ion-dependent catalysis[25] and supported by quantum-classical simulations[26]. Furthermore, our product state reveals that the two Mg^{2+} ions are around 4.2 Å from each other, in agreement with the product state of the histidine-mediated mechanism (Extended Data Fig. 3).

The fully active configuration requires marked conformational rearrangements, including an approximately 140° rotation of the HNH domain from the inactive state. Furthermore, our structures reveal the molecular mechanisms that underlie this rearrangement. The L1 and L2 linker domains tether HNH to the rest of Cas9 and are often missing from crystal structures, presumably owing to their intrinsic flexibility. However, in our active structure, we observe high-quality density for both L1 and L2. Notably, the L1 helix docks against the minor groove of the PAM-distal gRNA–TS duplex and forms an extended network of interactions, including multiple water-mediated hydrogen bonds with both strands (Fig. 3). As L1 docks on the minor groove, these interactions are gRNA–TS structure-specific rather than sequence-specific and can only occur when the PAM-distal duplex end is in the kinked conformation. This provides a structural basis for our observation that the kinked duplex conformation is an intermediate that precedes Cas9 activation and DNA cleavage. Comparisons of our model with Cas9 structures in inactive (Electron Microscopy Data Bank (EMDB) code EMD-3276) and active (EMD-0584) conformations confirmed that L1 docking against the gRNA–TS duplex is correlated with HNH rearrangement and Cas9 activation (Extended Data Fig. 4). Furthermore, our observation of L1 and L2 'locking' HNH in an active conformation is supported by the slow rate of dissociation of Cas9 from target DNA after cleavage[27].

Residue F916 stabilizes the NTS and is within the L2 linker domain; however, within the inactive Cas9 conformation, L2 is positioned more than 20 Å away from the RuvC active site. L1-facilitated positioning of HNH on the target strand enables relocation of L2, which in turn enables positioning of the NTS within the RuvC active site (Extended Data Fig. 4). This mechanism provides a structural explanation for the observed coupling of target strand and NTS cleavage, in which HNH docking precedes alignment of the NTS at the RuvC site for cleavage[5,28]. The HNH and RuvC cleavage reactions appear to occur simultaneously because the alignment is rate-limiting.

Although previous studies have noted the importance of L1 docking onto the gRNA–TS duplex for HNH repositioning[23,29], our observation that a linear gRNA–TS duplex conformation induced by PAM-distal mismatches precludes L1 docking provides a structural explanation for why certain PAM-distal mismatched substrates are able to bind Cas9, while not triggering rapid DNA cleavage[21].

The 18–20 mismatch reorders an RuvC loop

The 18–20 MM contains an unusual duplex conformation at the site of the mismatch. The C:C mismatch at position 18 on the target strand, TS(18), is stabilized by stacking interactions with adjacent Watson–Crick base pairs. However, the gRNA is otherwise distorted with gRNA position 2 (gRNA(2)) flipped out by around 180° so that gRNA(1) then intercalates between TS(19) and TS(20). TS(19) participates in water-mediated hydrogen bonds to Q1027, and TS(20) resumes base-pairing with NTS (Fig. 4, Extended Data Fig. 5).

This unusual nucleic acid conformation is stabilized by RuvC and appears to facilitate the binding of this mismatch. The residues within RuvC that contact and stabilize this distorted configuration are absent in previous on-target structures[14,15,18,30] (Extended Data Fig. 6), despite the overall similarity between our model and a previously determined active on-target Cas9 (Extended Data Fig. 7). This indicates that these resolved RuvC residues are involved only in mismatch binding and not in on-target activation (Fig. 4). Although this mechanism to accommodate certain mismatches may provide an essential mechanism for bacteria to restrict phage variants, it is counterproductive for the use of Cas9 in gene editing.

Previous rationally engineered variants 'hyper-accurate Cas9' (HypaCas9; N692A, M694A, Q695A and H698A mutations) and 'high-fidelity Cas9' (Cas9-HF1; N467A, R661A, Q695A and Q926A mutations) achieve somewhat higher fidelity at the expense of up to 100-fold reduced efficiency of on-target DNA cleavage[5,8,9]. The mutated residues are mainly located within the REC3 domain and make numerous interactions only with the kinked duplex end. Therefore, by abolishing interactions between REC3 and the PAM-distal duplex, these high-fidelity variants reduce the capacity of Cas9 to stabilize the kinked duplex configuration that is required for the docking of L1, and thereby reduce
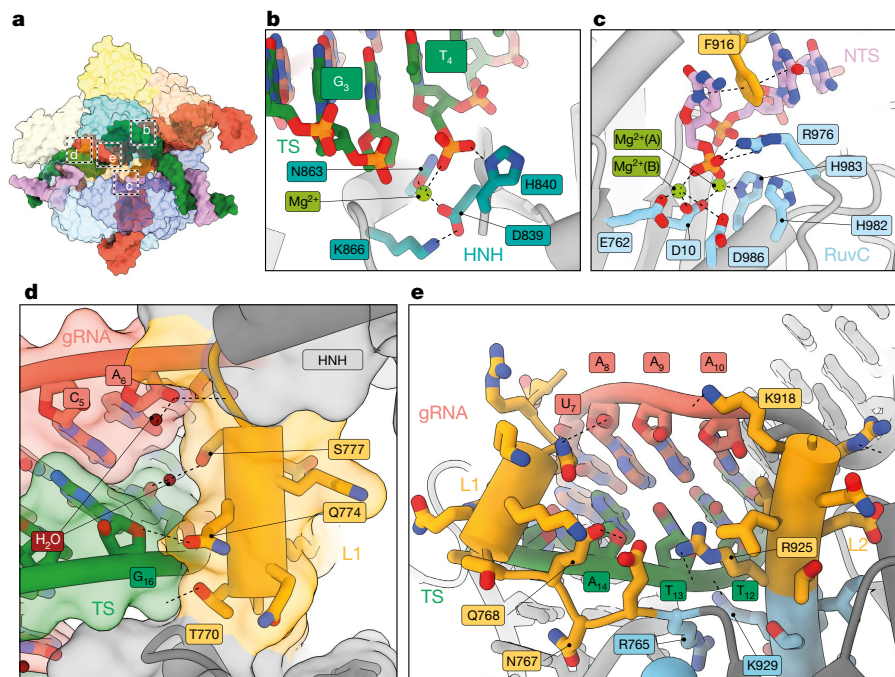

Fig. 3 | Linkers L1 and L2 mediate the structural transition to the active state. a, Overview of the 18–20 MM active conformation. **b**, **c**, Detailed view of HNH (**b**) and RuvC (**c**) active sites. **d**, Docking of the L1 linker helix against the PAM-distal gRNA–TS duplex, shown as an isosurface representation. **e**, Interactions of L1 and L2 regions with the minor groove of the gRNA–TS duplex. HNH extending from L1 and L2 linkers has been removed for clarity and does not interact with this region of the gRNA–TS duplex.
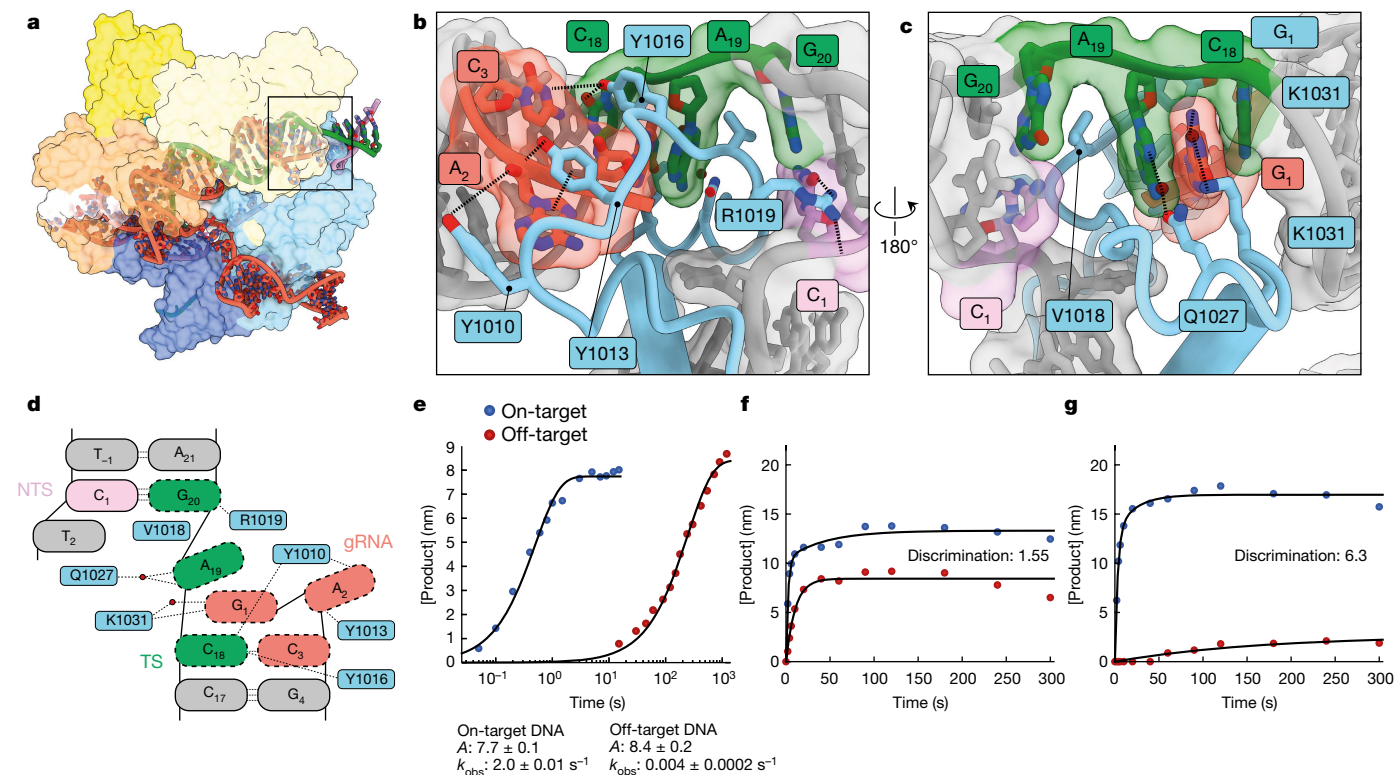
Fig. 4 | Stabilization of distorted 18–20 MM by the RuvC domain and improved fidelity of SuperFi-Cas9. a, Overall structure of the 18–20 MM active conformation viewed from the back. **b**, **c**, Magnified views of Cas9 interacting with the distal end of the duplex. Flipped gRNA base position 2 is accommodated by stacking interactions and hydrogen bonding with RuvC tyrosine side-chains, whereas a network of interactions (including a water-mediated hydrogen bond) stabilizes the stretched target strand configuration, which allows TS(20) to resume base-pairing with the NTS.

d, Schematic of distorted PAM-distal gRNA–TS duplex. Red circles correspond to water molecules. **e**, Kinetics of on-target and off-target (18–20 MM) Mg^{2+}-initiated cleavage by the 7-D Cas9 mutant (SuperFi-Cas9). **f**, **g**, Cleavage competition assay for wild-type Cas9 (**f**) and SuperFi-Cas9 (**g**). 25 nM of either Cas9 variant was mixed with 50 nM of each substrate and the cleaved DNA product was monitored. Discrimination in favour of the on-target DNA is defined by the ratio of amplitudes for on-target and off-target product formed.

HNH repositioning and cleavage activity. Our data provide a structural explanation for why these high-fidelity Cas9 variants reduce the activation of Cas9[9] by off-target substrates, but also reduce on-target Cas9 activity.

To test the role of this loop for mismatch stabilization, we designed a 7-D mutant (in which all seven of the stabilizing residues in Fig. 4b are mutated to aspartic acid) and tested the effects of this mutant on DNA cleavage. Although this 7-D mutant cleaved on-target DNA

Fig. 5 | Model for Cas9 activation. During R-loop propagation (step 1), the gRNA–TS duplex adopts a linear conformation. After R-loop completion, the PAM-distal end of the linear duplex is captured by REC3 (steps 2 and 3). Mismatches in the PAM-distal region appear to prevent REC3 docking and thereby block subsequent steps of Cas9 activation. Once the kinked R-loop conformation has been formed, L1 and L2 linkers use the gRNA–TS duplex as a scaffold to position the HNH domain at the scissile phosphate of the target strand and to position the NTS in the RuvC site (step 4), which enables Cas9 to make a double-strand break (step 5). According to this model, mutations in the RuvC loop (corresponding to SuperFi-Cas9) inhibit formation of the kinked conformation and subsequent cleavage of the gRNA–TS duplex with mismatches at the PAM-distal end.

at a similar rate to wild-type *Streptococcus pyogenes* Cas9 (SpCas9) (2 s[−1]), we observed that cleavage of 18–20 MM DNA was 500-fold slower (0.004 s[−1]) (Fig. 4e). This indicates that this loop is critical for stabilizing the distorted mismatch-induced PAM-distal duplex conformation, thereby allowing the duplex to adopt the kinked conformation that is prerequisite for Cas9 activation. We refer to our designed high-fidelity variant that retains wild-type on-target cleavage rates as 'SuperFi-Cas9'.
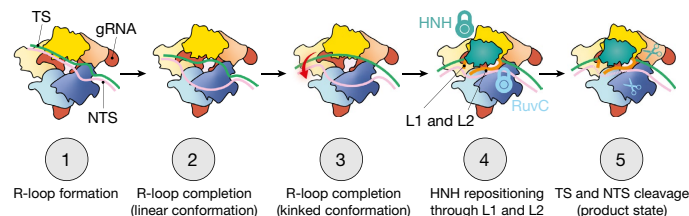
Because enzyme specificity is a kinetic phenomenon that is not determined solely by the rates of the chemical reaction, we performed a direct competition assay, in which on-target and off-target (18–20 MM) dsDNA substrates were mixed simultaneously with enzyme and cleavage was monitored over time. Although wild-type Cas9 showed some preference for on-target substrates (a 1.55-fold specificity ratio favouring the on-target over 18–20 MM off-target DNA), SuperFi-Cas9 showed rapid cleavage of on-target DNA and minimal cleavage of 18–20 MM DNA (6.3-fold preference for on-target DNA) (Fig. 4f, g). The ability to discriminate between on- and off-target DNA substrates without compromising DNA cleavage efficiency appears to be unique to SuperFi-Cas9[11]. Although further studies are needed to fully define the kinetic basis for the change in discrimination, our current data constitute a proof of concept and provide a rationale for engineering improved variants of Cas9 using our structure.

Discussion

Through kinetics-guided structural determination, we have described a gRNA–TS duplex conformational intermediate that precedes Cas9

activation (Fig. 5). Notably, we observe that the well-characterized and widespread off-target cleavage of DNA containing mismatches at the extreme PAM-distal end (positions 18–20 (refs. [5,9,12,31,32])) is attributed to a unique mechanism that stabilizes a highly distorted duplex conformation, involving a domain loop in RuvC that penetrates the duplex. This region is missing in previously determined structures of Cas9, which suggests that it has a role solely in mismatch tolerance at these positions. Our results provide molecular insights into the underlying structural mechanisms that govern off-target effects of Cas9, and provide a molecular blueprint for the design of next-generation high-fidelity Cas9 variants that reduce off-target DNA cleavage while retaining efficient cleavage of on-target DNA.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-04470-1.

1. Jinek, M. et al. RNA-programmed genome editing in human cells. *eLife* **2**, e00471 (2013).
2. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).
4. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).
5. Liu, M. et al. Engineered CRISPR/Cas enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nat. Commun.* **11**, 3576 (2020).
6. Kim, D., Luk, K., Wolfe, S. A. & Kim, J. S. Evaluating and enhancing target specificity of gene-editing nucleases and deaminases. *Annu. Rev. Biochem.* **88**, 191–220 (2019).
7. Slaymaker, I. M. & Gaudelli, N. M. Engineering Cas9 for human genome editing. *Curr. Opin. Struct. Biol.* **69**, 86–98 (2021).
8. Kleinstiver, B. P. et al. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
9. Chen, J. S. et al. Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).
10. Slaymaker, I. M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).
11. Kim, N. et al. Prediction of the sequence-specific cleavage activity of Cas9 variants. *Nat. Biotechnol.* **38**, 1328–1336 (2020).
12. Sternberg, S. H., Lafrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR–Cas9. *Nature* **527**, 110–113 (2015).
13. Singh, D. et al. Mechanisms of improved specificity of engineered Cas9s revealed by single-molecule FRET analysis. *Nat. Struct. Mol. Biol.* **25**, 347–354 (2018).
14. Jiang, F. et al. Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
15. 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).
16. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
17. Dagdas, Y. S., Chen, J. S., Sternberg, S. H., Doudna, J. A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR–Cas9. *Sci. Adv.* **3**, eaao0027 (2017).
18. Zhu, X. et al. Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat. Struct. Mol. Biol.* **26**, 679–685 (2019).
19. Cofsky, J. C., Soczek, K. M., Knott, G. J., Nogales, E. & Doudna, J. A. CRISPR–Cas9 bends and twists DNA to read its sequence. Preprint at https://doi.org/10.1101/2021.09.06.459219 (2021).
20. Pacesa, M. & Jinek, M. Mechanism of R-loop formation and conformational activation of Cas9. Preprint at https://doi.org/10.1101/2021.09.16.460614 (2021).
21. Jones, S. K. et al. Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat. Biotechnol.* **39**, 84–93 (2021).
22. Palermo, G. Structure and dynamics of the CRISPR–Cas9 catalytic complex. *J. Chem. Inf. Model.* **59**, 2394–2406 (2019).
23. Zhang, Y. et al. Catalytic-state structure and engineering of *Streptococcus thermophilus* Cas9. *Nat. Catal.* **3**, 813–823 (2020).
24. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
25. Steitz, T. A. & Steitz, J. A. A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl Acad. Sci. USA.* **90**, 6498–6502 (1993).
26. Casalino, L., Nierzwicki, Ł., Jinek, M. & Palermo, G. Catalytic mechanism of non-target DNA cleavage in CRISPR–Cas9 revealed by ab initio molecular dynamics. *ACS Catal.* **10**, 13596–13605 (2020).
27. Aldag, P. et al. Probing the stability of the SpCas9–DNA complex after cleavage. *Nucleic Acids Res.* **49**, 12411–12421 (2021).
28. Gong, S., Yu, H. H., Johnson, K. A. & Taylor, D. W. DNA unwinding is the primary determinant of CRISPR–Cas9 activity. *Cell Rep.* **22**, 359–371 (2018).
29. Sun, W. et al. Structures of *Neisseria meningitidis* Cas9 complexes in catalytically poised and anti-CRISPR-inhibited states. *Mol. Cell* **76**, 938–952 (2019).
30. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
31. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–198 (2015).
32. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).

Article

Methods

Protein expression and purification

SpCas9 was expressed and purified as described previously[5].

Nucleic acid preparation

DNA duplexes (55 nt) were prepared from PAGE-purified oligonucleotides synthesized by Integrated DNA Technologies. DNA duplexes used in cleavage assays were prepared by mixing 6-FAM- or Cy3-labelled target strands with unlabelled non-target strands at a 1:1.15 molar ratio in annealing buffer (10 mM Tris-HCl pH 8, 50 mM NaCl and 1 mM EDTA), heating to 95 °C for 5 min, then cooling to room temperature over the course of 1 h. The sgRNA was purchased from Synthego and annealed in annealing buffer using the same protocol as for the duplex DNA substrates. The sequences of the synthesized oligonucleotides, including the positions of mismatches, are listed in Extended Data Table 1.

Kinetics

Buffer composition for kinetic reactions. Cleavage reactions were performed in 1× cleavage buffer (20 mM Tris-Cl, pH 7.5, 100 mM KCl, 5% glycerol and 1 mM DTT) at 37 °C.

DNA cleavage kinetics. The reaction of Cas9 with on- and off-target DNA was performed by preincubating Cas9.gRNA (28 nM active-site concentration of Cas9, 100 nM gRNA) with 10 nM DNA with a 6-FAM label on the target strand in the absence of Mg^{2+}. The reaction was initiated by adding Mg^{2+} to 10 mM, then stopped at various times by mixing with 0.3 M EDTA (Extended Data Fig. 1). Products of the reaction were resolved and quantified using an Applied Biosystems DNA sequencer (ABI 3130xl)[33]. Data were fit using either a single or a double-exponential equation, as shown below.

Single exponential equation:

$$Y = A_1 e^{-\lambda 1t} + C \qquad (1)$$

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

Double exponential equation:

$$Y = A_1 e^{-\lambda 1t} A_2 e^{-\lambda 2t} + C \qquad (2)$$

in which Y represents the concentration of the 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.

Kinetic competition assay

Enzyme specificity is a kinetic phenomenon that is a function of all steps leading up to and including the first largely irreversible step in the pathway and it is common for mutants to introduce a change in specificity determining steps[34]. Therefore, we designed an assay to monitor relative rates of cleavage for on- and off-target DNA when the enzyme was presented with both substrates simultaneously. The competition assay was performed by mixing a solution of 25 nM (active site concentration) Cas9 and 100 nM sgRNA, in the presence of 10 mM Mg^{2+}, with 50 nM on-target DNA and 50 nM off-target DNA, in which the DNA contained a 5'-6-FAM label or a 5'-Cy3 label on the target or off-target DNA, respectively. Time points were collected by mixing with 0.3 M EDTA and reaction products were resolved and quantified by capillary electrophoresis, as described above. On-target cleavage data were fit to a single exponential function and off-target cleavage data were fit to a double exponential function. Discrimination was calculated as the ratio of the total amplitude of on-target cleavage divided by the amplitude for off-target cleavage to derive the relative specificity constants for the on-target DNA compared to the off-target DNA.

Cryo-EM sample preparation, data collection and processing

Cas9 in complex with various mismatched DNA substrates was frozen at different time points, on the basis of kinetic analysis (Extended Data Fig. 1). A non-productive mismatch complex (15–17 MM, 1 h); a slow productive mismatch (12–14) at early (5 min) and late (1 h) time points; and a fast productive mismatch (18–20, 1 min) were chosen. MDCC-Cas9 was used for structure determination to couple structural analysis with ongoing kinetic studies monitoring changes in fluorescence. It has previously been shown that the kinetics of MDCC-Cas9 were indistinguishable from those of wild-type enzyme[5]. The cleavage reaction was triggered by mixing 10 μM DNA duplex preincubated with 10 mM $MgCl_2$ and 8 μM MDCC-labelled Cas9: 8 μM gRNA was preincubated with 10 mM $MgCl_2$, in reaction buffer (19 mM Tris-Cl, pH 7.5, 95 mM KCl, 4.75% glycerol and 5 mM DTT) at a 1:1 ratio. Four microlitres of sample was applied to glow-discharged holey carbon grids (C-flat 2/2, Protochips), blotted for 1 s with a blot force of 4 and rapidly plunged into liquid nitrogen-cooled ethane using an FEI Vitrobot MarkIV. Reactions were quenched through vitrification.

Data were collected on an FEI Titan Krios cryo-electron microscope equipped with a K3 Summit direct electron detector (Gatan). Images were recorded with SerialEM[35] with a pixel size of 1.1 Å for 12–14 MM datasets, and 0.81 Å for 18–20 MM and 15–17 MM datasets, over a defocus range of −1.5 to −2.5 μm. During collection of the 12–14 MM 5-min time-point dataset, a preferred orientation was observed. To ameliorate this, a second dataset was collected at 30° tilt. Movies were recorded at 13.3 electrons per pixel per s for 6 s (80 frames) to give a total dose of 80 electrons per pixel. CTF correction, motion correction and particle picking were performed in real-time using cryoSPARC Live. Further data processing was performed with cryoSPARC v.3.2[36].

Multiple rounds of 3D classification within cryoSPARC yielded reconstructions of six distinct Cas9 complexes at resolutions ranging from 2.7 to 3.6 Å (Extended Data Table 3). To aid the separation of multiple Cas9 conformational states from within the same dataset, 3D variability analysis was performed within CryoSPARC. First and last frames from suitable eigenvector trajectory were then used as references for heterogeneous refinement (that is, reference-based 3D classification), and particles from resulting classes were refined using non-uniform refinement and used for final reconstructions[37]. Active Cas9 (Protein Data Bank (PDB) code: 6OOX) was rigid-body fitted into each map using ChimeraX[38]. Regions of the model not present in a given map were truncated, and flexible fitting was performed using Namdinator[39]. Further modelling was performed using Isolde[40], and the models were ultimately subjected to real-space refinement as implemented in PHENIX.

Reporting summary

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

Data availability

The structures of 12–14 MM 5 min, 12–14 MM 60-min linear and 18–20 MM 1-min kinked active, and their associated atomic coordinates, have been deposited into the EMDB and the PDB with EMDB accession codes EMD-24833, EMD-24835 and EMD-24838 and PDB accession codes 7S4U, 7S4V and 7S4X, respectively. Maps of 12–14 MM 60-min linear, 15–17 MM 60-min linear and 18–20 1-min linear have been deposited into the EMDB with accession codes EMD-23834, EMD-24836 and EMD-24837, respectively.

33. 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).

34. Johnson, K. A. *Kinetic Analysis for the New Enzymology* (KinTek, 2019).

35. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).

36. 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).
37. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat. Methods* **17**, 1214–1221 (2020).
38. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
39. Kidmose, R. T. et al. Namdinator—automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ* **6**, 526–531 (2019).
40. Croll, T. I. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519–530 (2018).
41. Pausch, P. et al. DNA interference states of the hypercompact CRISPR–CasΦ effector. *Nat. Struct. Mol. Biol.* **28**, 652–661 (2021).
42. Huang, X. et al. Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.* **11**, 5241 (2020).
43. Nishimasu, H. et al. Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-022-04470-1.
Correspondence and requests for materials should be addressed to Kenneth A. Johnson or David W. Taylor.
Peer review information *Nature* thanks Daan Swarts, John van der Oost and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.
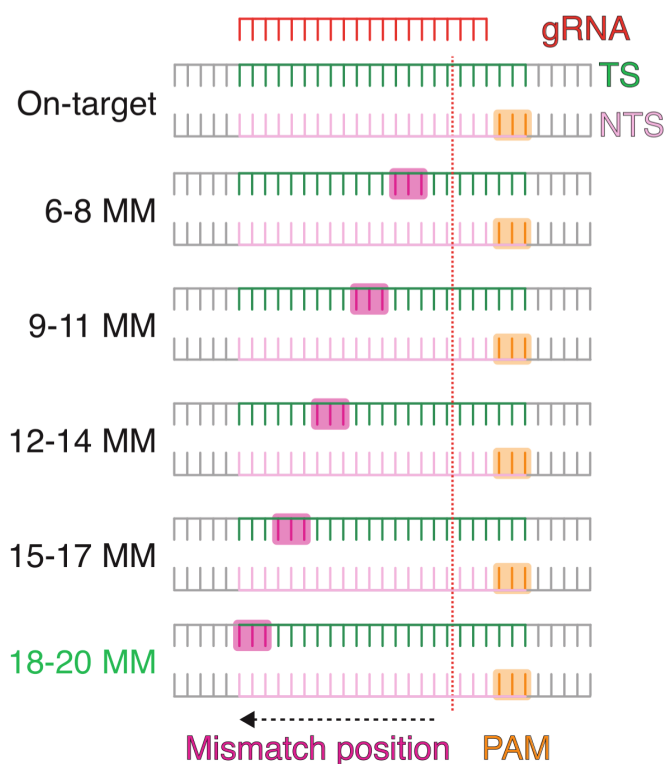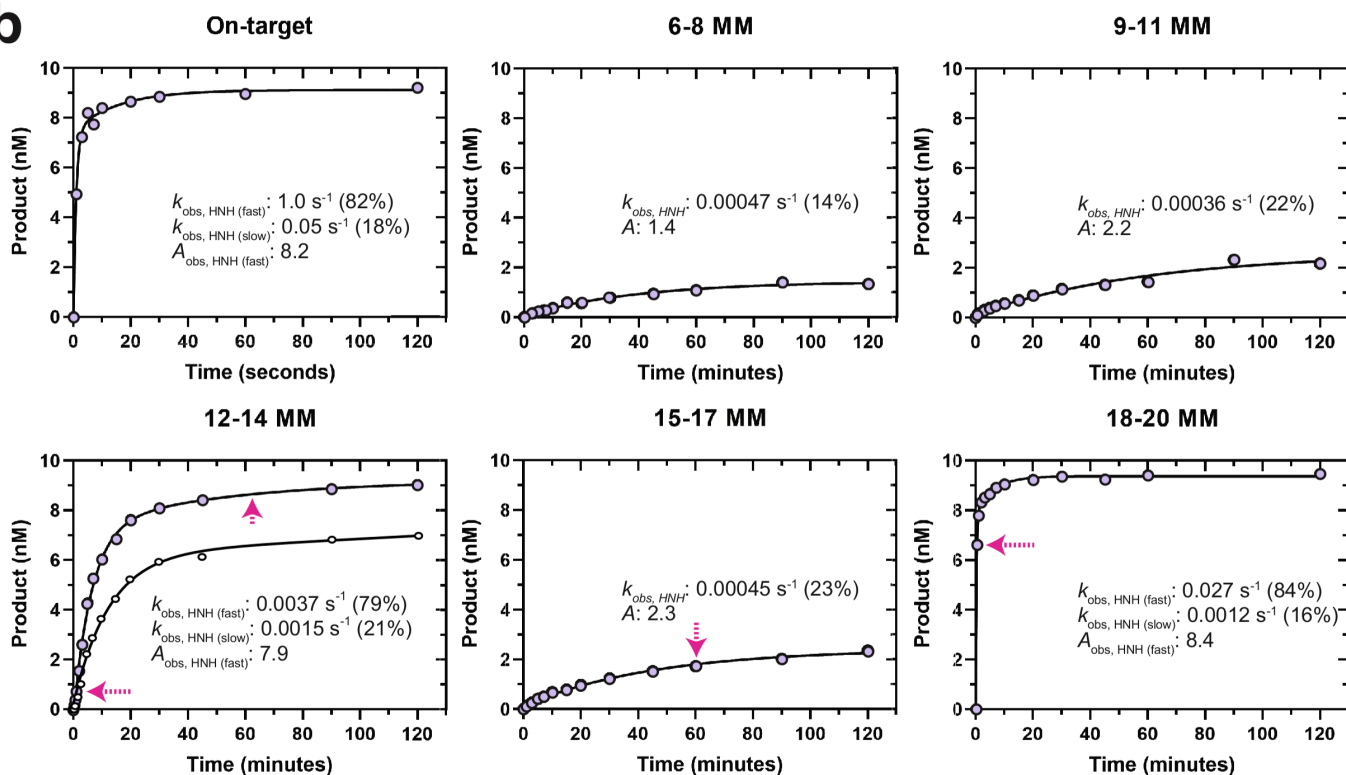
Extended Data Fig. 1 | Kinetic basis for mismatch discrimination by Cas9.
a, Schematic representation of mismatch constructs used for kinetic analysis.
b, Time course of cleavage of on-target and mismatched DNA (10 nM) by Cas9.
Magenta arrows correspond to time-points used to prepare cryo-EM samples.
A_{obs} corresponds to amplitude of product formed (i.e. total cleavage).

For 12–14 MM, target strand cleavage is shown with larger filled circles, while
NTS cleavage is given with smaller open circles. For other mismatches we only
show target strand cleavage. We previously reported NTS cleavage data for
on-target[28] and 18–20 MM substrates[5].

a

> 7Å
3Å

> 7Å
3Å

> 7Å
3Å

> 7Å
3Å

> 6Å
2Å

> 6Å
2Å

b

12-14 MM 5 min (linear)
GSFSC Resolution: 3.57Å

No Mask (4.1Å)
Spherical (3.9Å)
Loose (3.6Å)
Tight (3.5Å)
Corrected (3.6Å)

12-14 MM 60 min (linear)
GSFSC Resolution: 3.47Å

No Mask (4.4Å)
Spherical (4.1Å)
Loose (3.6Å)
Tight (3.4Å)
Corrected (3.5Å)

12-14 MM 60 min (kinked)
GSFSC Resolution: 3.28Å

No Mask (3.8Å)
Spherical (3.7Å)
Loose (3.4Å)
Tight (3.2Å)
Corrected (3.3Å)

15-17 MM 60 min (linear)
GSFSC Resolution: 3.33Å

No Mask (3.9Å)
Spherical (3.7Å)
Loose (3.5Å)
Tight (3.3Å)
Corrected (3.3Å)

18-20 MM 1 min (linear)
GSFSC Resolution: 2.87Å

No Mask (3.4Å)
Spherical (3.2Å)
Loose (3Å)
Tight (2.8Å)
Corrected (2.9Å)

18-20 MM 1 min (kinked)
GSFSC Resolution: 2.76Å

No Mask (3.4Å)
Spherical (3.2Å)
Loose (3Å)
Tight (2.7Å)
Corrected (2.8Å)

c

Extended Data Fig. 2 | Resolution estimates and orientation distributions of cryo-EM maps. a, Unsharpened maps coloured according to local resolution. **b**, Gold-standard FSC curves for cryo-EM reconstructions. Resolutions were estimated at FSC=0.143. **c**, Euler diagrams showing orientation distributions of cryo-EM reconstructions.
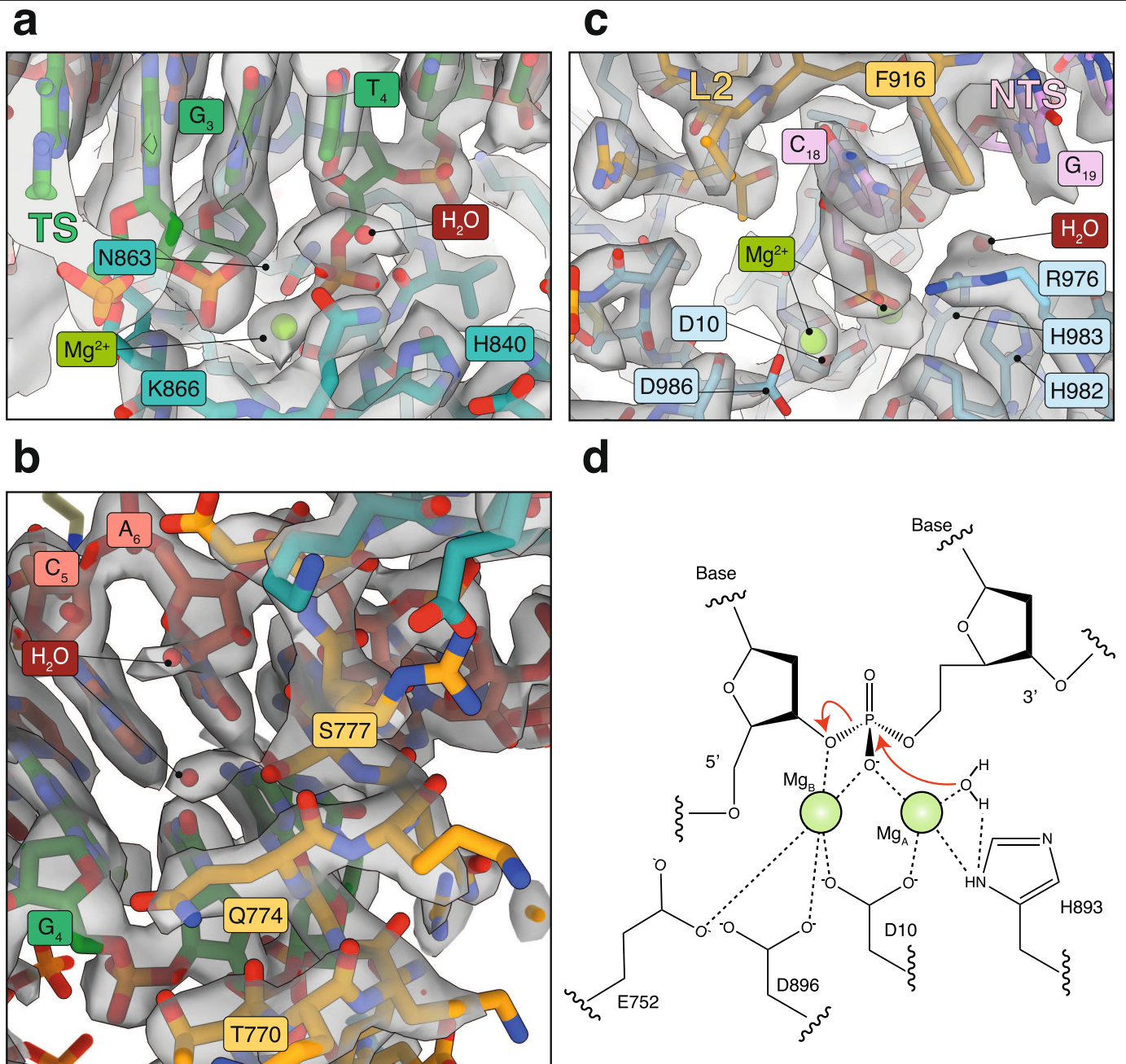
Extended Data Fig. 3 | Representative cryo-EM densities for 18–20 MM 1-min kinked (product) structure. a, HNH active site, showing cleaved target strand. **b**, L1 linker docked on PAM-distal kinked gRNA–TS duplex. Two water molecules are involved within the network of interactions that stabilize the L1 helix conformation. **c**, RuvC active site, showing cleaved NTS, and positioning of two Mg^{2+} ions. **d**, RuvC DNA cleavage mechanism. This is a typical two-metal-ion mechanism as described in[25] and agrees with QM/MM simulations for histidine-mediated activation[26], and the proposed mechanisms of Cas12j and Cas12i[41,42].

a
12-14 MM 5 min linear
vs
12-14 MM 1h kinked

18-20 MM 1min kinked
vs
12-14 MM 1h kinked

RMSD 1.904 Å

RMSD 4.647 Å

b
18-20 MM 1min kinked
vs
12-14 MM 1h kinked

60°

c

RuvC loop

REC3
T519
Y515
K526 K510 R586

L1 & L2
R661 S777
R780
N803 R783

Bridge Helix
K918
I852 D850 R919

R71
R66 R78
R63 R70 R74

REC1
R165
R447 G166

N407
Y72

gRNA

V1009 Y1016*
G1011 Y1013*
Y1010*
A2
C3

M694
Q695 G1

R1019* A19

G20

C18

V1018*

K1031*
Q1027*

N692

F1038
K1035 K1024
C1

T2 G3

K1003

G4 C5

Y1001 R1058

G4 C5 A6 U7 A8 A9 A10 G11 A12 U13 G14 A15 G16 A17 C18 G19 C20

C18 C17 G16 T15 A14 T13 T12 T11 C10 T9 A8 C7 T6 C5 T4 G3 C2 G1

Q768 E779 R925
Q774 R765
T770 K929 Q926

R661 N497 W659 R655
T657

S834 K810
V838 N863
K866 H840

D839
S1109

A-1 C-2 C-3

TS

HNH

F916

A6 U7 A8

F1037

N1041 T770

RuvC

T924
C18 G19 C20

D762
D10
K974 H983 R976

K1096

T21 G22 G23

R1333 R1335

PAM site

NTS

d

HNH
L1
L2

Binary

Inactive

e
Inactive
(EMD-3276)

Active
(EMD-0584)

No L1
density

L1
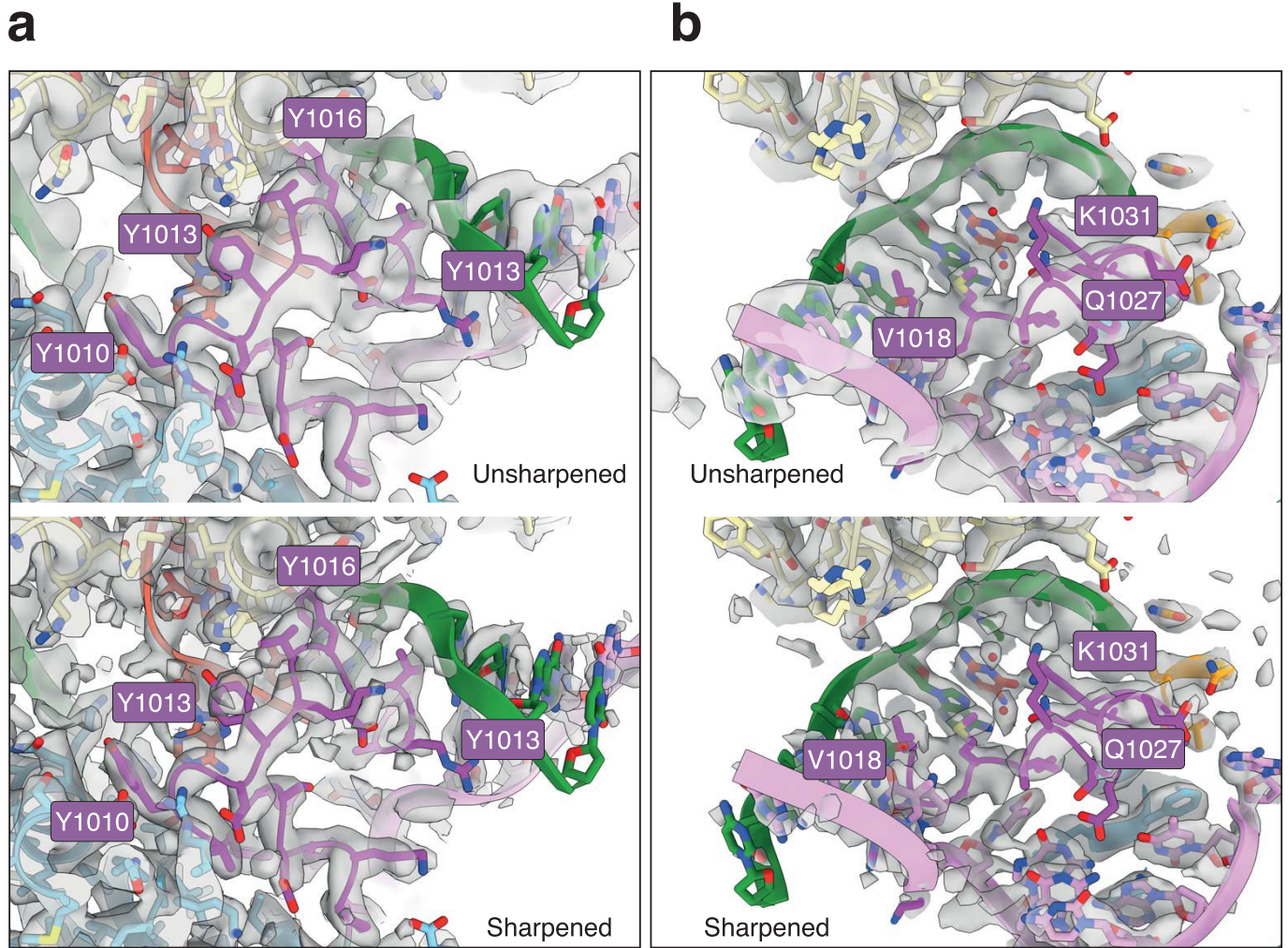density

f

evoCas9

Cas9-HF1

HypaCas9

**Extended Data Fig. 4** | See next page for caption.

Article

Extended Data Fig. 4 | Structural analysis of Cas9. a, Left, comparison of Cas9 protein only between 12–14 MM 60 min linear (colour) and 12–14 MM 1-h kinked (grey) models. Right, comparison of Cas9 protein only active conformation (18–20 MM 1 min linear, colour) and kinked pre-active (12–14 MM 60 min kinked, grey) models. While there is no significant conformational change associated between transition from linear to kinked pre-active (root-mean standard deviation (RMSD) between equivalent Cα atoms of 1.904 Å), the change from kinked pre-active to active conformations is associated with a larger conformational change (4.647 Å, most of which occurs within the REC3 domain). **b**, Close-up view of REC3 conformational changes that occur upon activation, as viewed from one angle. REC3 moves forwards towards the kinked duplex by ~15 Å upon activation and HNH repositioning. **c**, Schematic representation of Cas9–nucleic acid contacts in the context of 18–20 MM. Residues mutated in SuperFi-Cas9 are denoted by an asterisk. **d**, Conformations of HNH domain (green) and L1 (gold) and L2 (purple) linkers in the context of Cas9 binary complex (i.e. with gRNA, PDB 4ZT0), Cas9–gRNA complex bound to dsDNA in an inactive conformation (PDB 5F9R), and in the active Cas9 18–20 MM structure presented in this work. Upon activation, HNH is repositioned at the target strand cleavage site, driven by large conformational changes in the L1 and L2 linkers. **e**, Comparison with the active Cas9 18–20 MM structure presented in this work and previously determined cryo-EM maps (transparent grey) of inactive (left, EMD-3276[14]) and active (right, EMD-0584[18]) Cas9 bound to on-target dsDNA. The inactive Cas9 has no density for L1 helix at the kinked distal-docked gRNA–TS site, whereas there is clear density for L1 at this site in the active Cas9 cryo-EM map. **f**, Mapping of residues mutated to alanine in selected high-fidelity Cas9 variants. EvoCas9 (yellow) – M495, Y515, R661, K526. Cas9-HF1 (red) – N497, R661, Q695, Q926. HypaCas9 (blue) – N692, M694, Q695, H698. Residues shared between Cas9-HF-1 and either EvoCas9 or HypaCas9 are shown as orange and purple, respectively.
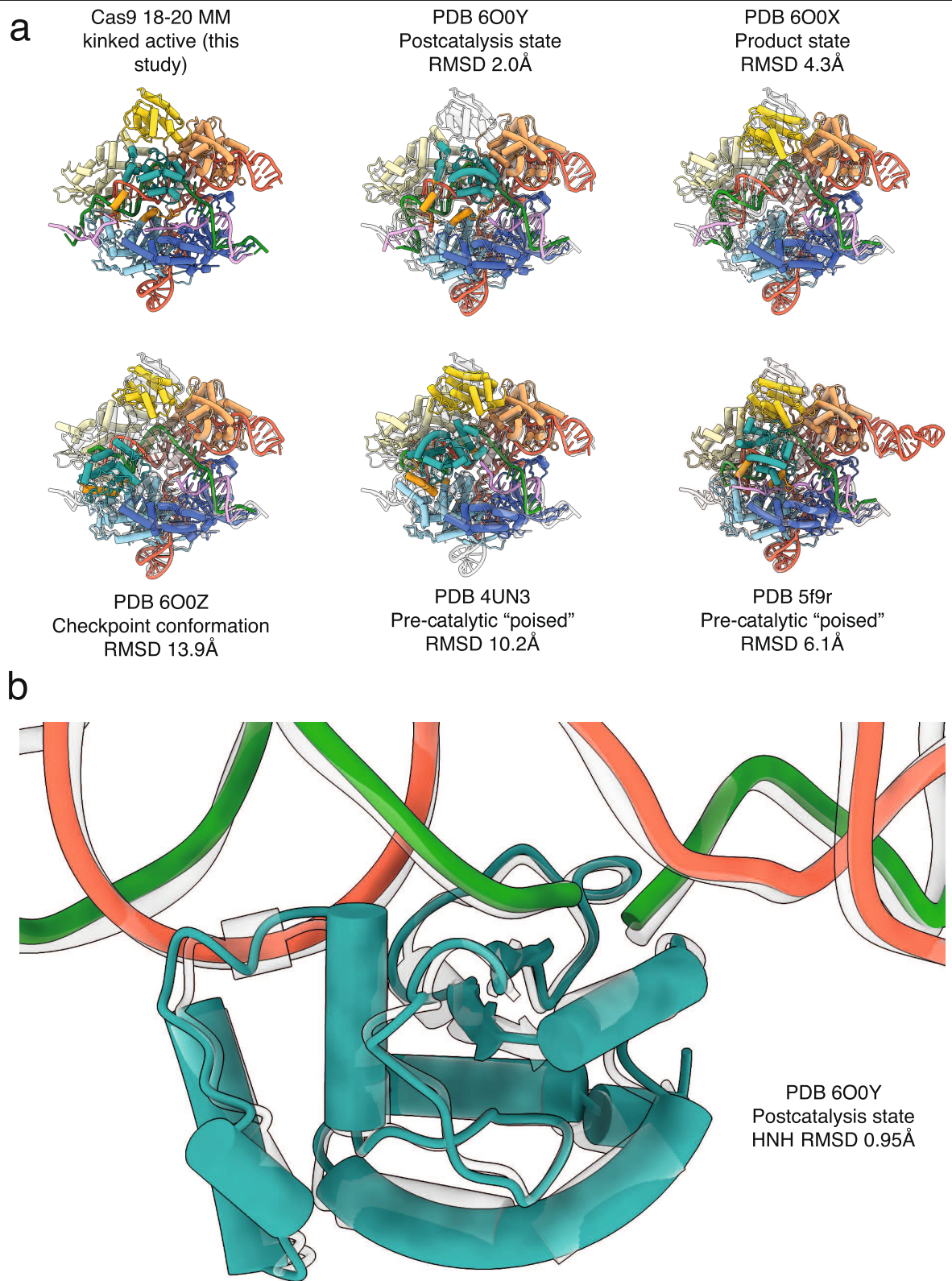
Extended Data Fig. 5 | Representative cryo-EM density for the RuvC loop. Two different views are shown (**a**, **b**). Unsharpened and B-factor sharpened maps are shown for each view with the RuvC loop shown as dark magenta. Key residues involved in stabilizing this distorted conformation are labelled.
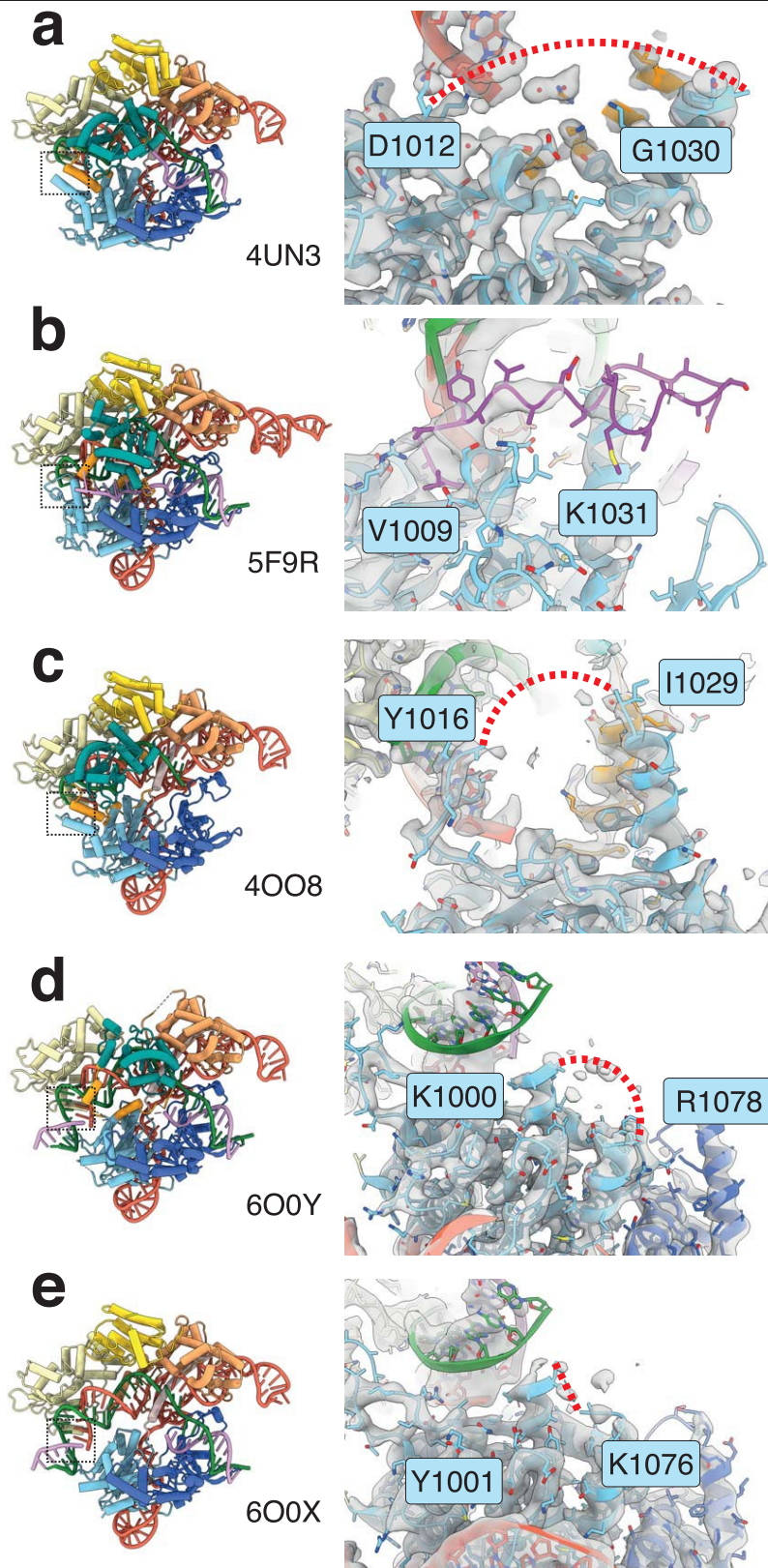
Extended Data Fig. 6 | RuvC loop in on-target SpCas9 structures.
a, On-target inactive Cas9 bound to dsDNA (PDB 4UN3)[15]. RuvC loop is missing between 1013–1029. **b**, On-target inactive (primed – HNH rearranged and adjacent to target strand scissile phosphate) Cas9 bound to dsDNA (PDB 5F9R)[14]. RuvC loop has been built primarily as alanine 'stub' residues, but electron density is very poor and diffuse for this region. **c**, On-target inactive Cas9 bound to dsDNA (PDB 4OO8)[43]. RuvC loop is missing between 1017–1028.

d, On-target active Cas9 bound to dsDNA in postcatalysis state[18]. RuvC loop is missing between 1001–1077. **e**, On-target active Cas9 bound to dsDNA in product state[18]. RuvC loop is missing between 1000–1075. In **a**–**c**, electron density is displayed as a grey surface, and in **d**, **e** cryo-EM density is shown as a grey surface. In all structures, missing residues are depicted as a red dashed line with the RuvC loop in **b** shown as magenta. Position of RuvC loop is denoted by a black dashed box in the left panel for each model.

a

Cas9 18-20 MM
kinked active (this
study)

PDB 6O0Y
Postcatalysis state
RMSD 2.0Å

PDB 6O0X
Product state
RMSD 4.3Å

PDB 6O0Z
Checkpoint conformation
RMSD 13.9Å

PDB 4UN3
Pre-catalytic "poised"
RMSD 10.2Å

PDB 5f9r
Pre-catalytic "poised"
RMSD 6.1Å

b

PDB 6O0Y
Postcatalysis state
HNH RMSD 0.95Å

Extended Data Fig. 7 | Comparison of Cas9 with previous structures.
a, Comparison of 18–20 MM kinked product state Cas9 with a selection of
previously determined structures. RMSD between equivalent C-alpha atoms is
shown. **b**, Alignment of HNH from the 18–20 MM kinked product state

presented here (transparent grey) and the previously determined
'post-catalysis' state (PDB 6O0Y). The catalytically competent HNH
conformation between these two structures is highly similar.

Extended Data Table 1 | List of nucleotide sequences used in the study

Name	Sequence (5'-3')	Source
On-target TS	/6-FAM/agc tga cgt ttg tac tcc agc gtc tca tct tta tgc gtc agc aga gat ttc tgc t	IDT
On-target NTS	agc aga aat ctc tgc tga cgc ata aag atg aga cgc tgg agt aca aac gtc agc t	IDT
6-8 MM TS	/6-FAM/agc tga cgt ttg tac tcc agc gtc agt tct tta tgc gtc agc aga gat ttc tgc t	IDT
6-8 MM NTS	agc aga aat ctc tgc tga cgc ata aag aac tga cgc tgg agt aca aac gtc agc tct cg	IDT
9-11 MM TS	/6-FAM/agc tga cgt ttg tac tcc agc gtc tca aga tta tgc gtc agc aga gat ttc tgc t	IDT
9-11 MM NTS	agc aga aat ctc tgc tga cgc ata atc ttg aga cgc tgg agt aca aac gtc agc tct cg	IDT
12-14 MM TS	/6-FAM/agc tga cgt ttg tac tcc agc gtc tca tct aat tgc gtc agc aga gat ttc tgc t	IDT
12-14 MM NTS	agc aga aat ctc tgc tga cgc aat tag atg aga cgc tgg agt aca aac gtc agc tct cg	IDT
15-17 MM TS	/6-FAM/agc tga cgt ttg tac tcc agc gtc tca tct tta acg gtc agc aga gat ttc tgc t	IDT
15-17 MM NTS	agc aga aat ctc tgc tga ccg tta aag atg aga cgc tgg agt aca aac gtc agc tct cg	IDT
18-20 MM TS	/6-FAM or Cy3/agc tga cgt ttg tac tcc agc gtc tca tct tta tgc cag agc aga gat ttc tgc t	IDT
18-20 MM NTS	agc aga aat ctc tgc tct ggc ata aag atg aga cgc tgg agt aca aac gtc agc t	IDT
sgRNA	GACGCAUAAAGAUGAGACGC + 80-mer SpCas9 scaffold	Synthego

Extended Data Table 2 | Correlation between fraction of DNA cleaved and fraction of cryo-EM particles in linear or kinked duplex conformations

Mismatch	Timepoint	% DNA cleaved	% of particles in linear/kinked conformation
12-14	5 min	9	100/0
12-14	60 min	82	42/58
15-17	60 min	19	100/0
18-20	1 min	63	61/39

Extended Data Table 3 | Cryo-EM data collection, refinement and validation statistics

	12-14 MM 5 min (linear) (EMD-24833) (PDB 7S4U)	12-14 MM 60 min (linear) (EMD-24834)	12-14 MM 60 min (kinked pre-active) (EMD-24835) (PDB 7S4V)	15-17 MM 60 min (linear) (EMD-24836)	18-20 MM 1 min (linear) (EMD-24837)	18-20 MM 1 min (kinked) (EMD-24838) (PDB 7S4X)
Data collection and processing						
Magnification	22,500	22,500	22,500	29,000	29,000	29,000
Voltage (kV)			300 kV			
Electron exposure (e–/Å²)	80	80	80	70	70	70
Defocus range (μm)			-1.5 to -2.5			
Pixel size (Å)	1.1	1.1	1.1	0.81	0.81	0.81
Symmetry imposed			C1			
Initial particle images (no.)	1,546,987	1,185,683	1,185,683	1,200,112	997,043	997,043
Final particle images (no.)	376,601	139,113	198,005	222,693	163,647	104,658
Map resolution (Å) FSC threshold	3.57	3.47	3.28	3.33	2.87	2.76
Map resolution range (Å)		3 - 7			2 - 6	
Refinement						
Initial model used (PDB code)	6O0X	N/A	6O0X	N/A	N/A	6O0X
Model resolution (Å) FSC threshold	3.7	N/A	3.4	N/A	N/A	3.0
	0.5	0.5	0.5			0.5
Map sharpening B factor (Å²)	202.1	92.0	96.8	111.7	59.3	68.8
Model composition						
Non-hydrogen atoms	11934	N/A	12221	N/A	N/A	14495
Protein residues	1092		1122			1354
Nucleotides	142		145			161
Ligands	0		0			5 Mg^{2+}, 7 H_2O
Mean B factors (Å²)						
Protein	43.8	N/A	30.98	N/A	N/A	54.68
Nucleotides	88.68		53.58			80.66
R.m.s. deviations						
Bond lengths (Å)	0.004	N/A	0.004	N/A	N/A	0.006
Bond angles (°)	0.75		0.555			0.588
Validation						
MolProbity score	1.55	N/A	1.58	N/A	N/A	1.66
Clashscore	6.11		4.66			6.97
Poor rotamers (%)	0		0			0
Ramachandran plot						
Favored (%)	96.65	N/A	95.05	N/A	N/A	95.99
Allowed (%)	3.35		4.95			4.01
Disallowed (%)	0		0			0

nature portfolio

Corresponding author(s): Kenneth A. Johnson and David W. Taylor

Last updated by author(s): Jan 10, 2022

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected on a Titan Krios equipped with a K3 direct electron detector. SerialEM 3.8 (ref 38) was used for automated cryo-EM data collection.
Data analysis	Cryo-EM data acquisition was monitored by on-the-fly pre-processing in cryoSPARC Live. Data were further processed using cryoSPARC v3.2 (ref 39). Model building and refinement of the structures were performed using Map to Model in PHENIX v1.18rc5 as well as Namdinator (ref 42). Models were also manually built in Isolde (ref 43). Model validation was performed using MolProbity as implemented within PHENIX. Visualization was performed using ChimeraX v1.0 (ref 41).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The structures of 12-14MM 5 min, 12-14MM 60 min linear and 18-20MM 1 min kinked active have been and their associated atomic coordinates have been deposited into the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) with accession codes EMD-24833, EMD-24835, EMD-24838 and PDB codes 7S4U, 7S4V and 7S4X, respectively. Maps of 12-14MM 60 min linear, 15-17MM 60 min linear and 18-20 1 min linear have been deposited into the Electron

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of between 1,848 - 2,172 micrographs were collected for each dataset. Each dataset contained at least 997,043 particles, and at least 104,658 were used for the final reconstruction. These are typical image numbers for cryo-EM datasets to obtain high resolution reconstructions.
Data exclusions	2D and 3D classification procedures were used to exclude damaged and 'bad' particles. This is standard practice in cryo-EM and is necessary in order to obtain homogeneous high resolution cryoEM structures.
Replication	Cryo-EM datasets were collected with multiple samples in separate imaging sessions. Linear gRNA:TS conformations present within all 4 datasets were nearly identical.
Randomization	No randomization was performed. Randomization is not relevant to this study because the work did not involve human subjects or live animals.
Blinding	No blinding was performed. Blinding is not relevant to this study because the work did not involve human subjects or live animals.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	Antibodies
☒	Eukaryotic cell lines
☒	Palaeontology and archaeology
☒	Animals and other organisms
☒	Human research participants
☒	Clinical data
☒	Dual use research of concern

Methods

n/a	Involved in the study
☒	ChIP-seq
☒	Flow cytometry
☒	MRI-based neuroimaging