Mechanism of R-loop formation and conformational activation of Cas9

Martin Pacesa[1], Martin Jinek[1]

Address:

[1] Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Corresponding author: Martin Jinek (jinek@bioc.uzh.ch)

Abstract

Cas9 is a CRISPR-associated endonuclease capable of RNA-guided, site-specific DNA cleavage[1-3]. The programmable activity of Cas9 has been widely utilized for genome editing applications[4-6]. Despite extensive studies, the precise mechanism of target DNA binding and on-/off-target discrimination remains incompletely understood. Here we report cryo-EM structures of intermediate binding states of *Streptococcus pyogenes* Cas9 that reveal domain rearrangements induced by R-loop propagation and PAM-distal duplex positioning. At early stages of binding, the Cas9 REC2 and REC3 domains form a positively charged cleft that accommodates the PAM-distal duplex of the DNA substrate. Target hybridisation past the seed region positions the guide-target heteroduplex into the central binding channel and results in a conformational rearrangement of the REC lobe. Extension of the R-loop to 16 base pairs triggers the relocation of the HNH domain towards the target DNA strand in a catalytically incompetent conformation. The structures indicate that incomplete target strand pairing fails to induce the conformational displacements necessary for nuclease domain activation. Our results establish a structural basis for target DNA-dependent activation of Cas9 that advances our understanding of its off-target activity and will facilitate the development of novel Cas9 variants and guide RNA designs with enhanced specificity and activity.

Main

Cas9 enzymes rely on a dual guide RNA structure consisting of a CRISPR RNA (crRNA) guide and a trans-activating CRISPR RNA (tracrRNA) coactivator to cleave complementary DNA targets. The archetypical Cas9 ortholog from *Streptococcus pyogenes* (SpCas9) has found widespread use as a programmable DNA targeting tool in genome editing and gene targeting applications[4-6]. Target DNA binding by SpCas9 is predicated on initial recognition of an NGG protospacer-adjacent motif (PAM) downstream of the target site[2,7-9], which triggers local DNA strand separation to initiate its directional hybridization with a 20-nt spacer segment in the guide crRNA to form an R-loop structure[7,10,11]. This process is facilitated by structural pre-ordering of nucleotides 11-20 of the crRNA, termed the seed sequence, in an A form-like conformation[8,12]. Upon formation of a full R-loop, the Cas9 HNH and RuvC nuclease domains become activated to cleave the target (TS) and non-target (NTS) DNA strands, respectively, generating a double-strand DNA break three base pairs (bp) upstream of the PAM[2,8,13]. Although highly specific, SpCas9 can nevertheless cleave off-target genomic sites with imperfect complementarity to the guide RNA[14-18]. The resulting off-target activity is dependent on the number, type, and positioning of base mismatches within the guide-target heteroduplex[15,19-21]. PAM-proximal mismatches within the seed region are discriminated against through substantially increased dissociation rates[11,19,21,22]. In contrast, PAM-distal mismatches are compatible with stable binding, but trap the enzyme in a cleavage-incompetent, dead-end complex[13,23,24]. Structural, biophysical and computational studies of SpCas9 have shed light on the mechanism of guide RNA binding, PAM recognition, and nuclease activation, revealing that enzyme undergoes extensive conformational rearrangements in the process. In particular, high-resolution structures of the fully-bound target DNA complex of SpCas9[25-28] have revealed a target DNA-dependent conformational rearrangement of the Cas9 REC-lobe that is necessary for cleavage activation. However, our structural understanding of the mechanisms that govern conformational activation of SpCas9 and on-/off-target discrimination during R-loop formation remains incomplete.

Cryo-EM analysis of R-loop formation

To investigate the mechanism of SpCas9 R-loop formation, we initially determined the minimal extent of target DNA complementarity necessary for stable binding using fluorescence-coupled size exclusion chromatography, revealing that the presence of six complementary nucleotides in the PAM-proximal region of the target DNA heteroduplex is sufficient for stable association with the SpCas9-guide RNA complex. (**Extended Data Fig. 1**). Subsequently, catalytically-inactive Cas9 (dCas9) was reconstituted with a single-molecule guide RNA (sgRNA) and partially-matched DNA substrates containing 6, 8, 10, 12, 14, and 16 complementary nucleotides (**Fig. 1a, Extended Data Fig. 2**), and the resulting complexes were analysed by cryo-EM, yielding molecular reconstructions at resolutions of 3.0–4.1 Å (**Extended Data Fig. 3, Extended Data Table 1**). The conformational heterogeneity within each partially complementary DNA complex was examined using 3D variability analysis[29]. Most of the detected variability within each complex can be attributed to the PAM-distal duplex and the REC2, REC3, and HNH domains (**Extended Data Fig. 4**), suggestive of conformational equilibrium sampling. The resulting structural models are representative of the most abundant conformational state of each complex (**Extended Data Fig. 5**).

Structural superpositions of each of the partially-bound complexes with the guide RNA-bound binary SpCas9 complex[12] and the catalytically active states of SpCas9[28] provide a reference frame for the reconstruction of the DNA binding mechanism, revealing stepwise domain rearrangements coupled to R-loop formation. All complexes exhibit almost identical conformations of the bridge helix, REC1, RuvC, and PAM interaction domains, as well as the PAM-proximal dsDNA duplex and the sgRNA up to the seed region (**Extended Data Fig. 6a**). Conformational differences are observed in the positioning of the REC2, REC3, and the HNH domain relative to the emerging R-loop, consistent with the 3D variability analysis.

Initial DNA binding by REC2/REC3 domains

The structure of the 6-nucleotide complementary substrate (6-nt match) complex shows a 5-bp heteroduplex formed by the sgRNA seed sequence and TS DNA (**Fig. 1b**). Comparisons with the

structure of the Cas9-guide RNA binary complex reveal that target strand hybridization is associated with a displacement of the REC2 domain out of the central binding channel (**Fig. 1b**) in the formation of a positively charged cleft between the REC2 and REC3 domains that accommodates the PAM-distal substrate DNA duplex (**Extended Data Fig. 6b**), which is stabilized by interactions with the REC2 residues Ser219, Thr249 and Lys263 with the NTS backbone (**Extended Data Fig. 6c**), and REC3 residues Arg586 and Thr657 with the TS backbone (**Extended Data Fig. 6d**). Consequently, the NTS is positioned parallel to the guide RNA-TS DNA heteroduplex within the central binding channel (**Fig. 1b**). The 5' end of the sgRNA could not be precisely modelled due to conformational flexibility but residual cryoEM density suggests its placement in a positively charged cleft located between the HNH and PAM-interaction domains (**Extended Data Fig. 6e**).

In the 6-nt match complex, hybridization beyond the fifth seed sequence nucleotide is precluded by base stacking with the side chain of Tyr450, which was previously observed in the structure of the Cas9-sgRNA binary complex[12] (**Fig. 1c**). The structure of the 8-nucleotide complementary substrate (8-nt match) complex reveals that expansion of the R-loop heteroduplex past Tyr450 forces further repositioning of the REC2 and REC3 domains to widen the binding channel as the PAM-distal duplex shifts deeper inside (**Fig. 1d-f; Extended Data Fig. 6f**). Together, these observations suggest that the seed sequence of the Cas9 guide RNA is bipartite and that its hybridization with target DNA proceeds in two steps, consistent with the existence of a short-lived intermediate state observed in FRET studies[11,30]. R-loop propagation and PAM-distal duplex displacement results in the formation of new intermolecular contacts, with Cas9 contacting the PAM-distal duplex backbone through REC2 domain residues Ser217, Lys234 and Lys253, and REC3 residues Arg557 and Arg654 (**Extended Data Fig. 6g,h**). Mutation of these residues, which would further destabilise this intermediate state and thus promote off-target dissociation, presents an opportunity to generate novel high-fidelity SpCas9 variants. As most off-targets are only bound

but not cleaved[19-21,31], these variants could prove to be valuable for biotechnological applications that rely on the fidelity of Cas9 target binding[32-35], such as transcriptional regulation or base editing.

R-loop propagation and remodelling

Further guide RNA-TS hybridisation to form a 10-bp heteroduplex causes a rearrangement of the REC2 and REC3 domains and repositioning of the PAM-distal DNA duplex into the positively charged central binding channel formed by the REC3, RuvC, and the HNH domains (**Fig. 2a**). Although the PAM-distal dsDNA duplex could not be accurately modelled, residual cryo-EM density observed within the channel suggests that it forms a continuous base stack with the sgRNA-TS heteroduplex (**Fig. 2b**). The displaced NTS is positioned underneath the HNH domain and continues to run parallel to the extending guide RNA-TS DNA heteroduplex (**Extended Data Fig. 7c**). X-ray crystallographic analysis of the 10-nt match complex at a resolution of 2.8 Å (**Extended Data Table 2**) confirmed that the TS and NTS remain hybridised at the PAM-distal end of the DNA substrate (**Extended Data Fig. 7a**). The PAM-distal duplex is wedged between the REC3 and RuvC domains, and the L1 HNH linker (**Extended Data Fig. 7a-c**). The relocation of the PAM-distal duplex causes REC2 to shift closer to the binding channel and occlude the cleavage site in TS DNA (**Fig. 2a**). This shift also establishes a new electrostatic interaction between a negatively charged helix in REC2 (Glu260, Asp261, Asp269, Asp272, Asp273, Asp274, Asp276) and a positively charged helix in REC3 (Lys599, Arg629, Lys646, Lys649, Lys652, Arg653, Arg654, Arg655), hereafter referred to as the DDD and RRR helices, respectively (**Fig. 2c**, which are highly conserved across Cas9 orthologs that contain a REC2 domain (**Extended Data Fig. 7d**).

R-loop completion and Cas9 activation

R-loop propagation past the seed region to form a 12-bp heteroduplex does not result in major REC lobe rearrangements (**Fig. 3a**), with the PAM-distal duplex remaining stacked onto the guide RNA-TS DNA heteroduplex. However, the HNH domain, which in the 6-nt, 8-nt and 10-nt match complexes is docked on the RuvC and PI domains with its active site buried at the interface of the three domains (**Fig. 3b**), becomes disordered along with surrounding RuvC and PI loops in the 12-

nt match complex (**Fig. 3c, Extended Data Fig. 4, Extended Data Fig. 5**). The REC lobe conformation is maintained upon extension of the R-loop heteroduplex to 14 bp and the RuvC and PI loops responsible for HNH docking remain structurally disordered (**Fig. 3d, Extended Data Fig. 8a**). Residual density is observed for the HNH domain, due to a contact of the paired heteroduplex with the L2 linker (**Extended Data Fig. 8a**). The PAM-distal region of the substrate DNA becomes disordered, likely due to strand separation, and the NTS can be modelled only a few nucleotides past the PAM region (**Fig. 3d, Extended Data Fig. 8a**). Further extension of the R-loop from 14 to 16 base pairs preserves the orientations of the REC2 and REC3 domains (**Fig. 3a**), but causes a large translation of the HNH domain towards the target heteroduplex within the central binding channel (**Fig. 4a**). Facilitated by the formation of the PAM-distal part of the R-loop, a RuvC domain loop (residues 1030-1040) restructures into a helical conformation, establishing an interaction with the L2 linker of the HNH domain (**Extended Data Fig. 9b**). This interaction alters the positioning of the L2 linker and shifts the HNH domain on top of the heteroduplex to seal off the central binding channel (**Fig. 4a**). This opens up a positively charged cleft between the HNH, RuvC, and PI domains to accommodate the NTS (**Extended Data Fig. 8c**). Although no residual cryo-EM density can be observed for the NTS in the 16-nt match complex, its positioning could be traced to the same positively charged cleft observed in the catalytically active conformation of Cas9[28] (**Extended Data Fig. 8d**).

Notably, the orientation of REC2/3 domains in the 12-,14- and 16-nt match complexes is consistent with a catalytically inactive state of Cas9 proposed in biophysical studies[23,24,36]. The catalytic site of the HNH domain remains positioned away from the TS cleavage site in the 16-bp heteroduplex complex (**Fig. 3d**). To reach the catalytically active state[28], the Cas9 REC2 domain must dissociate from the TS cleavage site, which is facilitated by concerted movements of the HNH and REC3 domains upon full R-loop formation (**Fig. 4a**), and consistent with prior biophysical, structural, and computational studies[23,24,37,38]. REC2 dissociation enables a ~140° rotation of the HNH domain towards the scissile phosphate in the TS (**Fig. 4b**), which is accompanied by the

extension of the L2 linker helix, and leads to activation of the RuvC domain for NTS cleavage[13]. HNH domain repositioning is likely induced by the proximity of the extended R-loop heteroduplex, as supported by observations that PAM-distal end positioning allosterically modulates HNH domain conformation[36].

Conclusions

In sum, our structural analysis of SpCas9 along its DNA binding pathway reveals a mechanism whereby R-loop formation is allosterically and energetically coupled to domain rearrangements necessary for nuclease activation (**Extended Data Fig. 9**). Directional formation of the guide RNA-target DNA heteroduplex causes a rearrangement of the REC2 and REC3 domains and repositions the HNH nuclease domain towards the TS DNA substrate, facilitated by a network of electrostatic and hydrogen bonding interactions between the DNA substrate and Cas9, and between individual protein domains. Incomplete target pairing fails to conformationally activate the HNH domain for cleavage, explaining why off-target complexes remains trapped in an inactive state. This model further highlights the importance of maintaining guide-target complementarity and proper heteroduplex conformation, as domain mispositioning occludes interaction surfaces necessary for downstream rearrangements and nuclease activation, consistent with biophysical and computational studies showing that the conformation of the R-loop heteroduplex strongly affects off-target binding[11,39]. These findings thus have important implications for ongoing experimental and computational studies aiming to uncover the effects of off-target mismatches on Cas9 activity, and could inform the development of new high-fidelity SpCas9 variants and guide RNA designs.

Methods

Expression and purification of Cas9 proteins

Catalytically inactive *Streptococcus pyogenes* Cas9 (D10A/H840A mutant) was expressed in *E.coli* Rosetta 2 (DE3) (Novagen) for 16 hours at 18 °C as fusion proteins with an N-terminal His_6-MBP-TEV tag. Bacterial pellets were resuspended and lysed in 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 5 mM imidazole, and protease inhibitors. Cell lysates were clarified using ultracentrifugation and loaded on a 15 ml Ni-NTA Superflow column (QIAGEN) and washed with 7 column volumes of 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 5 mM imidazole. Tagged Cas9 was eluted with 10 column volumes of 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 200 mM imidazole. Salt concentration was adjusted to 250 mM KCl and the protein was loaded on a 10 ml HiTrap Heparin HP column (GE Healthcare) equilibrated in 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT. The column was washed with 5 column volumes of 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT, and dCas9 was eluted with 15 column volumes of 20 mM HEPES-KOH pH 7.5, 1.5 M KCl, 1 mM DTT, in a 0-50% gradient (peak elution around 500 mM KCl). His6-MBP tag was removed by TEV protease cleavage overnight at 4 °C with gentle shaking. The untagged protein was concentrated and further purified on a Superdex 200 16/600 gel filtration column (GE Healthcare) in 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 1 mM DTT. Pure fractions were concentrated to 10 mg/ml, flash frozen in liquid nitrogen and stored at 80 °C.

sgRNA *in vitro* transcription

The single guide RNA (sgRNA) was transcribed from a dsDNA template in a 5 ml transcription reaction (30 mM Tris-HCl pH 8.1, 25 mM MgCl2, 2 mM spermidine, 0.01% Triton X-100, 5 mM CTP, 5 mM ATP, 5 mM GTP, 5 mM UTP, 10 mM DTT, 1 µM DNA transcription template, 0.5 units inorganic pyrophosphatase (Thermo Fisher), 250 µg T7 RNA polymerase). The transcription reaction was incubated at 37 °C for 5 hours, after which the dsDNA template was degraded for 30 minutes with 15 units of RQ1 DNAse (Promega). The transcribed sgRNA was PAGE purified on an

8% denaturing polyacrylamide gel containing 7 M urea, ethanol precipitated and dissolved in DEPC-treated water.

Gel filtration binding assay

The dCas9-gRNA complex was assembled by incubating 371 picomoles of dCas9 with 400 picomoles of the sgRNA in 20 mM HEPES-KOH pH 7.5, 200 mM KCl, 2 mM $MgCl_2$ for 10 minutes at room temperature. Then 250 picomoles of Cy5-labeled dsDNA substrate was added and incubated another 15 minutes. The volume was adjusted up to 100 µl with reaction buffer and the mixture was centrifuged to remove possible precipitates. Individual reactions were transferred to a 96-well plate and analysed using a Superdex 200 Increase 5/150 GL gel filtration column (GE Healthcare) attached to an Agilent 1200 Series Gradient HPLC system. The 260 nm, 280 nm, and Cy5 signals were exported and plotted as a function of the retention volume in GraphPad Prism 9.

Crystallisation and X-ray structure determination

The 10-nt complementary ternary complex of dCas9 was assembled by first incubating dCas9 with the sgRNA in a 1:1.5 molar ratio, and pre-purifying the binary complex on a Superdex 200 16/600 gel filtration column (GE Healthcare) in 20 mM HEPES-KOH pH 7.5, 500 mM KCl, 1 mM DTT. The binary complex was diluted in 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT to 2.5 mg/ml and the partially complementary dsDNA substrate was added in 1:1.5 molar excess. For crystallisation, 1 µl of the ternary complex (1.5-2.5 mg/ml) was mixed with 1 µl of the reservoir solution (0.1 M sodium cacodylate pH 6.5, 0.8-1.2 M ammonium formate, 12-14% PEG4000) and crystals were grown at 20 °C using the hanging drop vapour diffusion setup. Crystals were harvested after 3-4 weeks, cryoprotected in 0.1 M Na cacodylate pH 6.5, 1.0 M ammonium formate, 13% PEG4000, 20% glycerol, 2 mM $MgCl_2$, and flash-cooled in liquid nitrogen. Diffraction data was measured at the beamline PXIII of the Swiss Light Source at a temperature of 100 K (Paul Scherrer Institute, Villigen, Switzerland) and processed using the autoPROC and STARANISO package with anisotropic cut-off[43]. Phases were obtained by molecular replacement using the Phaser module of the Phenix package[44] using the NUC lobe of the PDB ID: 5FQ5 as initial search

model. The crystals belonged to the P1 space group and contained two copies of the complex in the asymmetric unit.

Cryo-EM sample preparation and data acquisition

To assemble the 6-, 8-, 10-, 12-, 14-, and 16-nt match complexes, dCas9 protein was mixed with the sgRNA in a 1:1.5 molar ratio, and incubated at room temperature for 10 minutes in buffer 20 mM HEPES-KOH pH 7.5, 150 mM KCl, 1 mM DTT. The respective partially complementary dsDNA substrate was then added in a 1:3 Cas9:DNA molar ratio and incubated another 20 minutes at room temperature. The complexes were then purified using a Superdex 200 Increase 10/300 GL gel filtration column (GE Healthcare) and eluted in 20 mM HEPES-KOH pH 7.5, 250 mM KCl, 1 mM DTT. Concentration of the monomeric peak was determined using the Qubit 4 Fluorometer Protein Assay, and then diluted to 0.275 mg/ml in 20 mM HEPES-KOH pH 7.5, 250 mM KCl cold buffer. 3 µl of diluted complex was applied to a glow discharged 200-mesh holey carbon grid (Au 1.2/1.3 Quantifoil Micro Tools), blotted for 1.5-2.5 s at 90% humidity, 20 °C, plunge frozen in liquid propane/ethane mix (Vitrobot, FEI) and stored in liquid nitrogen. Data collection was performed on a 300 kV FEI Titan Krios G3i microscope equipped with a Gatan Quantum Energy Filter and a K3 direct detection camera in super-resolution mode. Micrographs were recorded at a calibrated magnification of 130,000 x with a pixel size of 0.325 Å. Data acquisition was performed automatically using SerialEM with three shots per hole at -0.8 µm to -2.2 µm defocus.

Cryo-EM data processing

Acquired cryoEM data was processed using cryoSPARC[45]. Gain-corrected micrographs were imported and binned to a pixel size of 0.65 Å during patch motion correction. After patch CTF estimation, micrographs with a resolution estimation worse than 5 Å and full-frame motion distance larger than 100 Å were discarded. Initial particles were picked on denoised micrographs using Topaz 0.2.4 with the pre-trained ResNet16 (64 units) model[46]. Particles were extracted with a box size of 384 x 384 pixels, down-sampled to 192 x 192 pixels. After 2D classification, templates were generated using good classes and particle picking was repeated using the template picker. Particle

picks were inspected and particles with NCC scores above 0.4 were extracted as before. After 2D classification, duplicate particles were removed and used for *ab initio* 3D reconstruction. All partially bound complexes displayed several conformational states. After several rounds of 3D classification, classes with most detailed features were reextracted using full 384 x 384 pixel box size and subjected to non-uniform refinement to generate high-resolution reconstructions[47]. Each map was sharpened using the appropriate B-factor value to enhance structural features, and local resolution was calculated and visualised using ChimeraX[48].

Structural model building, refinement, and analysis

Map sharpening and density modification to facilitate X-ray and cryoEM model building was performed using Phenix.auto_sharpen[49]. Manual Cas9 domain placement, model adjustment and nucleic acid building was completed using COOT [50]. Atomic model refinement was performed using Phenix.refine for X-ray data[51] and Phenix.real_space_refine for cryoEM[52]. The quality of refined models was assessed using MolProbity[53]. Protein-nucleic acid interactions were analysed using the PISA web server[54]. Characterisation of the guide-protospacer duplex was performed using the 3DNA 2.0 web server[55]. Structural figures were generated using ChimeraX[48].

Protein sequence alignment

Protein sequences of Cas9 orthologues harbouring the REC2 domain were obtained from UniProt[40]. Sequence alignment was performed using MUSCLE with default parameters[41]. Alignment was visualised using Jalview with highlighting only the conservation of charged residues[42].

Author contributions

M.P. and M.J. conceived the study and designed experiments. M.P. purified Cas9, performed *in vitro* cleavage assays, crystallized 10-bp heteroduplex complex, prepared cryo-EM samples and solved the structures. M.P. and M.J. performed structural analysis and wrote the manuscript.

Acknowledgements

This work was supported by the Swiss National Science Foundation Grant 31003A_182567 (to M.J.). M.J. is an International Research Scholar of the Howard Hughes Medical Institute and Vallee Scholar of the Bert L & N Kuggie Vallee Foundation. We thank Simona Sorrentino, Marta Sawicka, Luuk Loeff, Irma Querques, and Lena Muckenfuss for assistance with cryo-EM data collection, and Franziska Boneberg and Christelle Chanez for their help with preparing reagents. We thank members of the Jinek laboratory for discussion and critical reading of the manuscript. We are grateful to Josh Cosfsky, Katarzyna Soczek and Jennifer Doudna for sharing unpublished data and helpful comments.

References

1 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71, doi:10.1038/nature09523 (2010).

2 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

3 Sapranauskas, R. *et al.* The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* **39**, 9275-9282, doi:10.1093/nar/gkr606 (2011).

4 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).

5 Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2**, e00471, doi:10.7554/eLife.00471 (2013).

6 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).

7 Mekler, V., Minakhin, L. & Severinov, K. Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc Natl Acad Sci U S A* **114**, 5443-5448, doi:10.1073/pnas.1619926114 (2017).

8 Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67, doi:10.1038/nature13011 (2014).

9 Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* **111**, 9798-9803, doi:10.1073/pnas.1402597111 (2014).

10 Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573, doi:10.1038/nature13579 (2014).

11 Ivanov, I. E. *et al.* Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc Natl Acad Sci U S A* **117**, 5853-5860, doi:10.1073/pnas.1913445117 (2020).

12 Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481, doi:10.1126/science.aab1452 (2015).

13 Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113, doi:10.1038/nature15544 (2015).

14 Cameron, P. *et al.* Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600-606, doi:10.1038/nmeth.4284 (2017).

15 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).

16 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).

17 Lazzarotto, C. R. *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR-Cas9 genome-wide activity. *Nat Biotechnol*, doi:10.1038/s41587-020-0555-7 (2020).

18 Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197, doi:10.1038/nbt.3117 (2015).

19 Boyle, E. A. *et al.* Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. *Sci Adv* **7**, doi:10.1126/sciadv.abe5496 (2021).

20 Jones, S. K., Jr. *et al.* Massively parallel kinetic profiling of natural and engineered CRISPR nucleases. *Nat Biotechnol*, doi:10.1038/s41587-020-0646-5 (2020).

21 Zhang, L. *et al.* Systematic in vitro profiling of off-target affinity, cleavage and efficiency for CRISPR enzymes. *Nucleic Acids Research*, doi:10.1093/nar/gkaa231 (2020).

22 Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat Commun* **7**, 12778, doi:10.1038/ncomms12778 (2016).

23 Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410, doi:10.1038/nature24268 (2017).

24 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, doi:10.1126/sciadv.aao0027 (2017).

25 Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell* **61**, 895-902, doi:10.1016/j.molcel.2016.02.020 (2016).

26 Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871, doi:10.1126/science.aad8282 (2016).

27 Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949, doi:10.1016/j.cell.2014.02.001 (2014).

28 Zhu, X. *et al.* Cryo-EM structures reveal coordinated domain motions that govern DNA cleavage by Cas9. *Nat Struct Mol Biol* **26**, 679-685, doi:10.1038/s41594-019-0258-2 (2019).

29 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol* **213**, 107702, doi:10.1016/j.jsb.2021.107702 (2021).

30 Sung, K., Park, J., Kim, Y., Lee, N. K. & Kim, S. K. Target Specificity of Cas9 Nuclease via DNA Rearrangement Regulated by the REC2 Domain. *J Am Chem Soc* **140**, 7778-7781, doi:10.1021/jacs.8b03102 (2018).

31 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, doi:10.1038/nbt.2916 (2014).

32 Gaudelli, N. M. *et al.* Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471, doi:10.1038/nature24644 (2017).

33 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424, doi:10.1038/nature17946 (2016).

34 Maeder, M. L. *et al.* CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977-979, doi:10.1038/nmeth.2598 (2013).

35 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).

36 Yang, M. *et al.* The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep* **22**, 372-382, doi:10.1016/j.celrep.2017.12.048 (2018).

37 Palermo, G. *et al.* Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q Rev Biophys* **51**, doi:10.1017/S0033583518000070 (2018).

38 Palermo, G., Miao, Y., Walker, R. C., Jinek, M. & McCammon, J. A. Striking Plasticity of CRISPR-Cas9 and Key Role of Non-target DNA, as Revealed by Molecular Simulations. *ACS Cent Sci* **2**, 756-763, doi:10.1021/acscentsci.6b00218 (2016).

39 Newton, M. D. *et al.* DNA stretching induces Cas9 off-target activity. *Nat Struct Mol Biol* **26**, 185-192, doi:10.1038/s41594-019-0188-z (2019).

40 UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* **49**, D480-D489, doi:10.1093/nar/gkaa1100 (2021).

41 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

42 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191, doi:10.1093/bioinformatics/btp033 (2009).

43 Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**, 293-302, doi:10.1107/S0907444911007773 (2011).

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

45 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).

46 Bepler, T. *et al.* Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nat Methods* **16**, 1153-1160, doi:10.1038/s41592-019-0575-8 (2019).

47 Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221, doi:10.1038/s41592-020-00990-8 (2020).

48 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).

49 Terwilliger, T. C., Sobolev, O. V., Afonine, P. V. & Adams, P. D. Automated map sharpening by maximization of detail and connectivity. *Acta Crystallogr D Struct Biol* **74**, 545-559, doi:10.1107/S2059798318004655 (2018).

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

51 Liebschner, D. *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877, doi:10.1107/S2059798319011471 (2019).

52 Afonine, P. V. *et al.* Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol* **74**, 531-544, doi:10.1107/S2059798318006551 (2018).

53 Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci* **27**, 293-315, doi:10.1002/pro.3330 (2018).

54 Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774-797, doi:10.1016/j.jmb.2007.05.022 (2007).

55 Li, S., Olson, W. K. & Lu, X. J. Web 3DNA 2.0 for the analysis, visualization, and modeling of 3D nucleic acid structures. *Nucleic Acids Res* **47**, W26-W34, doi:10.1093/nar/gkz394 (2019).

Figure 1

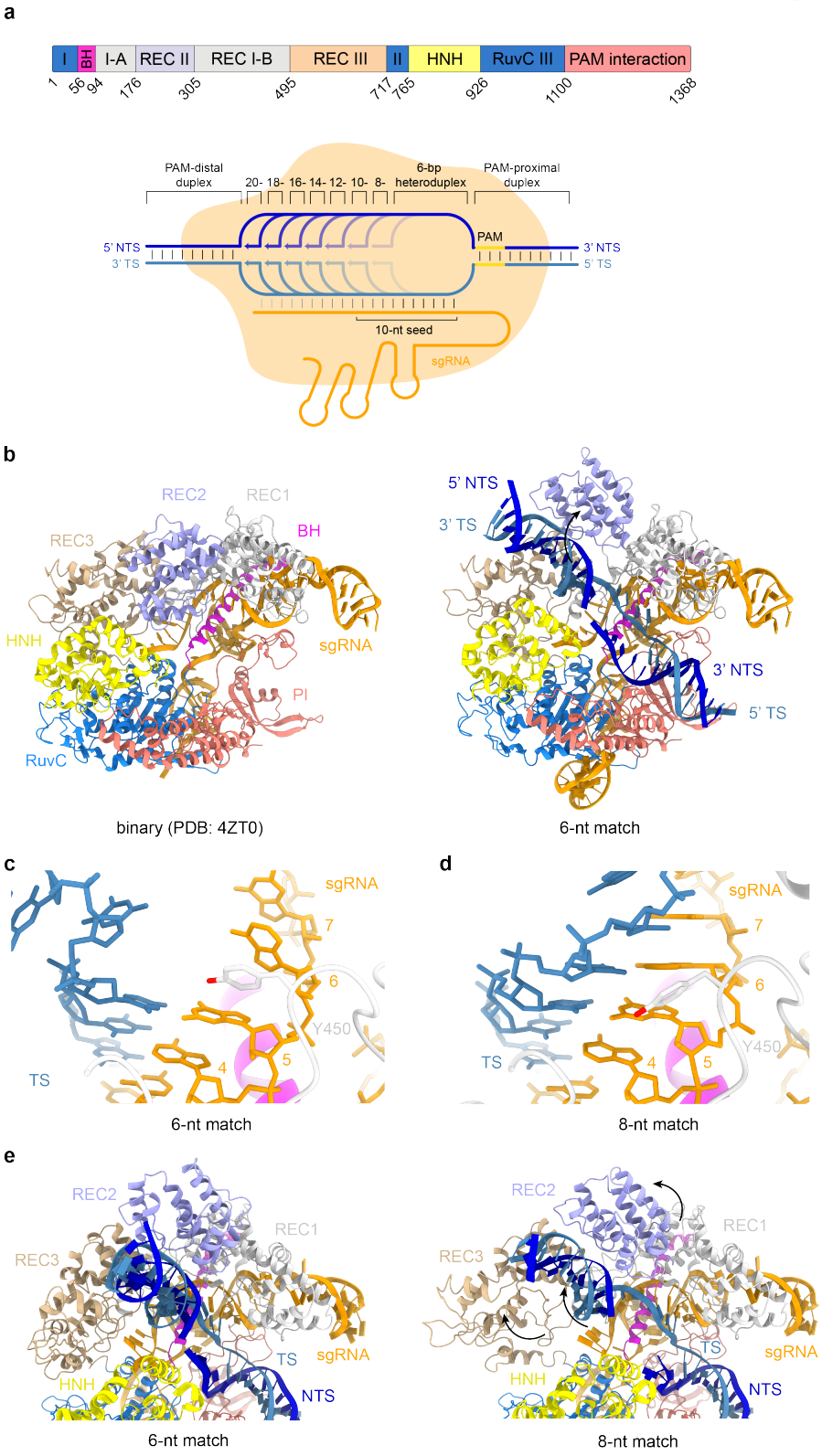

Figure 1 | Target DNA binding induces restructuring of the Cas9 REC lobe.
a, Top, Schematic representation of the domain composition of the *Streptococcus pyogenes* Cas9 nuclease, BH: bridge helix; Bottom, Schematic depicting DNA-bound complexes with increasing complementarity to guide RNA. **b,** Structural comparison of the SpCas9 binary complex (left) and the 6-nt match complex (right). **c,** Zoom-in view of the seed region og the sgRNA-TS DNA heteroduplex in the 6-nt match complex. Tyr450 stacks between the 5th and 6th nucleotide, counting from the PAM-proximal end of the heteroduplex. **d,** Zoom-in view of the seed region og the sgRNA-TS DNA heteroduplex in the 8-nt match complex. **e,** Structural comparison of the 6- and 8-nt match complexes. Arrows indicate inferred domain repositioning during the 6-nt to 8-nt transition.

Figure 2

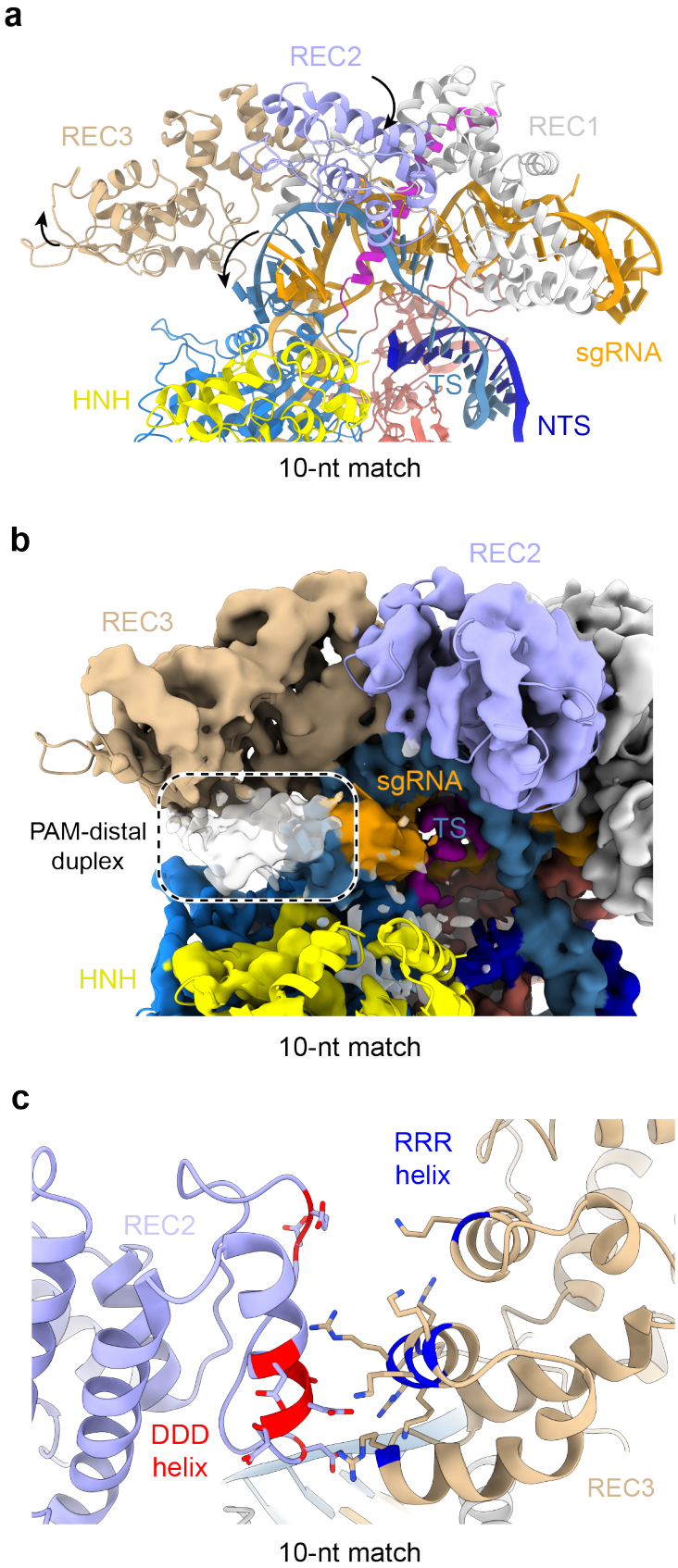

Figure 2 | 10-bp duplex formation drives DNA repositioning within Cas9.
a, Zoom-in view of the R-loop bound in the 10-nt match complex. **b,** Zoom-in view of the PAM-distal duplex (white density) in the 10-nt match complex. The cryo-EM map is coloured according the schematic in **Fig. 1a**. **c,** Zoom-in view of the interaction between the REC2 domain DDD helix and the REC3 RRR helix.

Figure 3

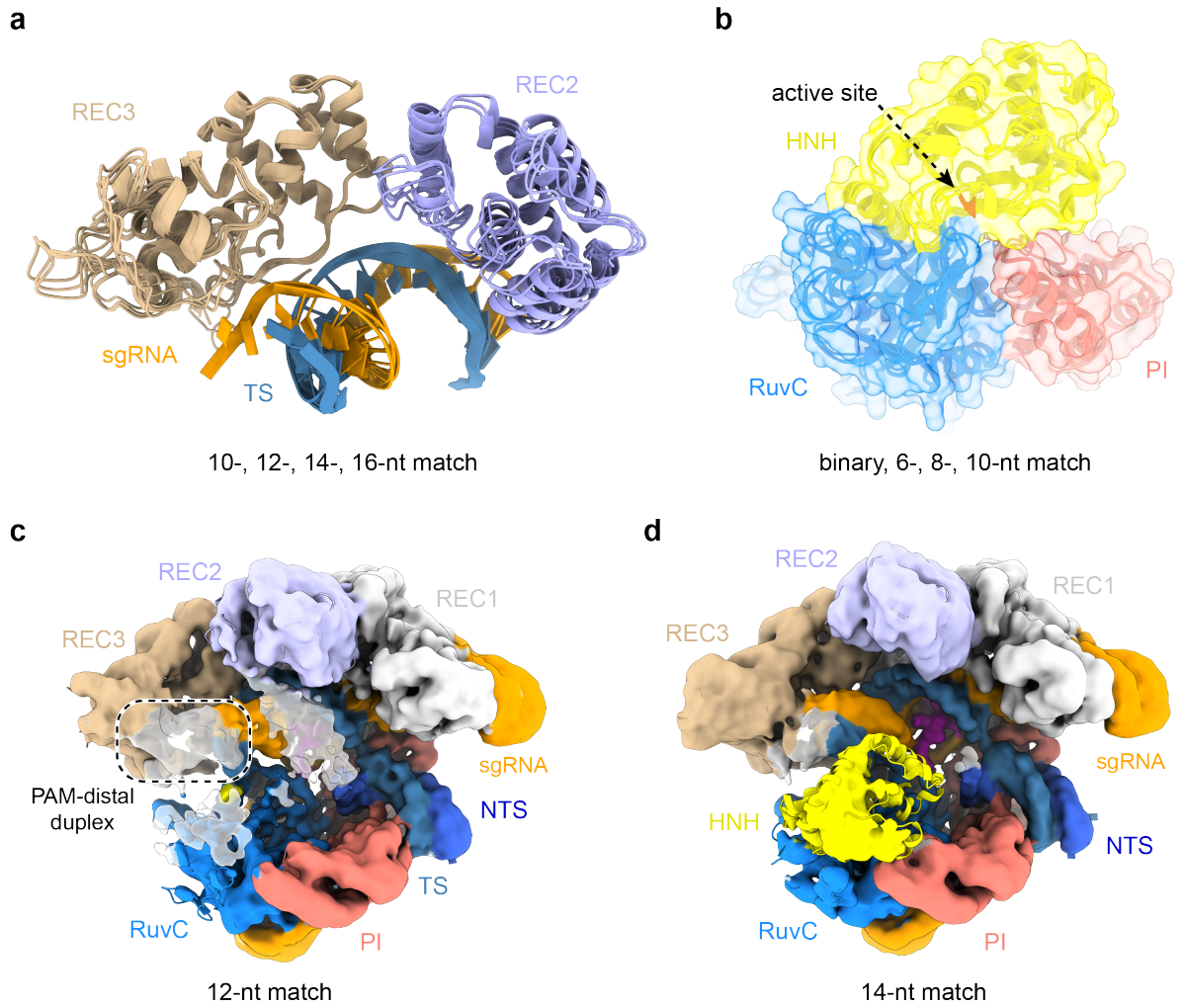

Figure 3 | Target pairing past the seed region undocks the HNH nuclease domain.
a, Structural overlay of the REC2 and REC3 domains in the 10-, 12-, 14-, and 16-nt match complexes. **b.** Position of the HNH catalytic site in the 6-, 8-, and 10-nt match complexes. **c,** Overview of the 12-nt match complex model overlaid with unsharpened cryo-EM map. Residual density corresponding to the PAM-distal duplex (white) is highlighted, the adjacent density is presumed to correspond to the NTS. HNH density is disordered. **d,** Overview of the 14-nt match complex model overlaid with unsharpened cryo-EM map, showing residual density corresponding to the HNH domain. No density is visible for NTS. Cryo-EM maps are coloured according the schematic in **Fig. 1a**.

Figure 4

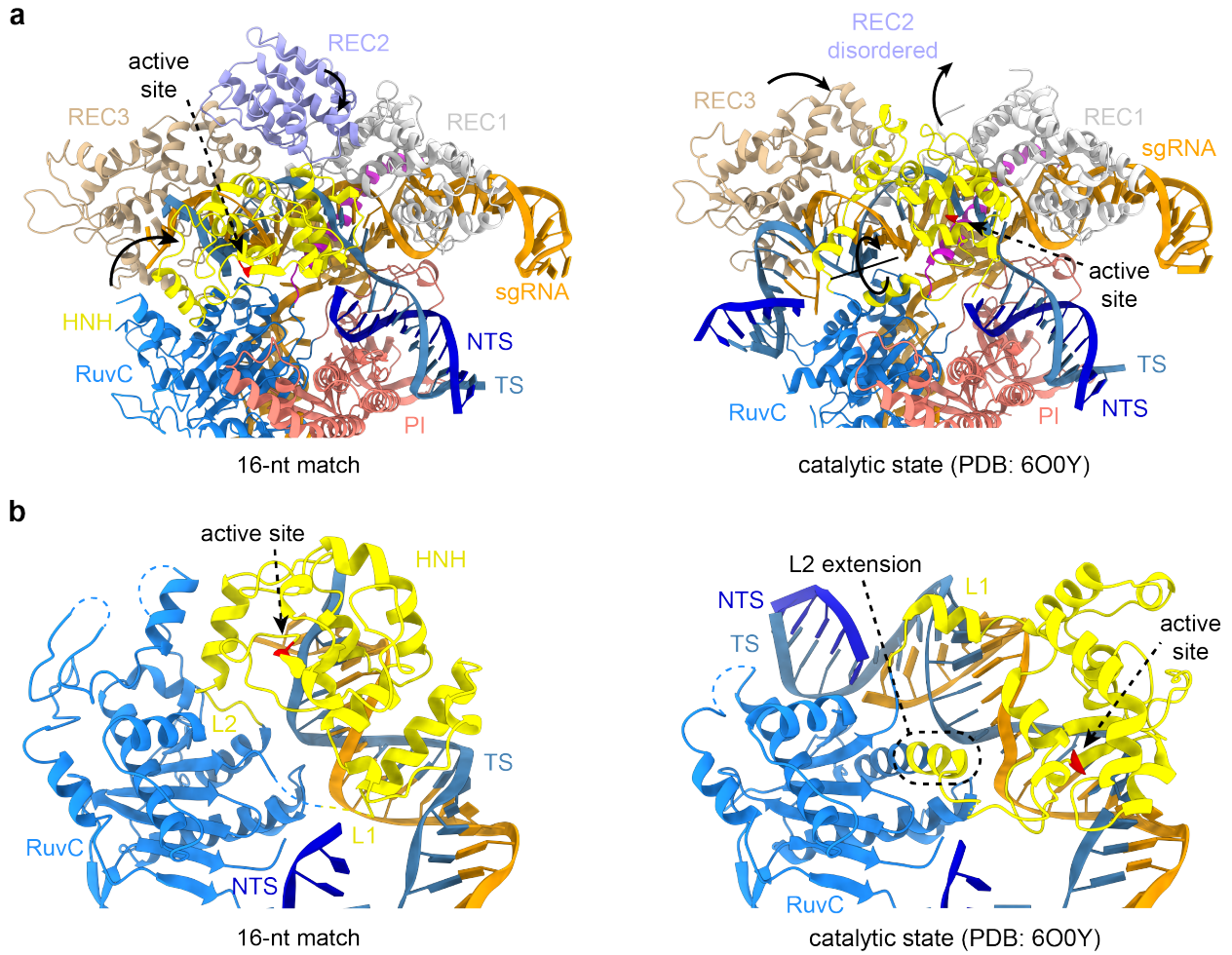

Figure 4 | HNH rotation is required TS cleavage and RuvC nuclease activation.
a, Structural comparison of the 16-nt match complex (left) and the catalytic state of SpCas9 (right). **b.** Detailed views of the HNH and RuvC nuclease domains in the 16-nt match complex (left) and the catalytic state (right).