# CRISPR–Cas9 bends and twists DNA to read its sequence

Joshua C. Cofsky [1,2,3], Katarzyna M. Soczek [1,2,3], Gavin J. Knott [4], Eva Nogales [1,2,5,6] and Jennifer A. Doudna [1,2,3,5,6,7,8,9] ✉

In bacterial defense and genome editing applications, the CRISPR-associated protein Cas9 searches millions of DNA base pairs to locate a 20-nucleotide, guide RNA-complementary target sequence that abuts a protospacer-adjacent motif (PAM). Target capture requires Cas9 to unwind DNA at candidate sequences using an unknown ATP-independent mechanism. Here we show that Cas9 sharply bends and undertwists DNA on PAM binding, thereby flipping DNA nucleotides out of the duplex and toward the guide RNA for sequence interrogation. Cryogenic-electron microscopy (cryo-EM) structures of Cas9–RNA–DNA complexes trapped at different states of the interrogation pathway, together with solution conformational probing, reveal that global protein rearrangement accompanies formation of an unstacked DNA hinge. Bend-induced base flipping explains how Cas9 'reads' snippets of DNA to locate target sites within a vast excess of nontarget DNA, a process crucial to both bacterial antiviral immunity and genome editing. This mechanism establishes a physical solution to the problem of complementarity-guided DNA search and shows how interrogation speed and local DNA geometry may influence genome editing efficiency.

C RISPR–Cas9 (clustered regularly interspaced short palindromic repeats, CRISPR-associated) nucleases provide bacteria with RNA-guided adaptive immunity against viral infections[1] and serve as powerful tools for genome editing in human, plant and other eukaryotic cells[2]. The basis for Cas9's use is its DNA recognition mechanism, which involves base pairing of one DNA strand with 20 nucleotides of the guide RNA to form an R-loop. The guide RNA's recognition sequence, or 'spacer,' can be chosen to match a desired DNA target, enabling programmable site-specific DNA selection and cutting[3]. The search process that Cas9 uses to comb through the genome and locate rare target sites requires local unwinding to expose DNA nucleotides for RNA hybridization, but it does not rely on an external energy source such as ATP hydrolysis[4,5]. This DNA interrogation process defines the accuracy and speed with which Cas9 induces genome edits, yet the mechanism remains unknown.

Molecular structures of Cas9 (ref. [6]) in pre- and post-DNA bound states revealed that the protein's REC ('recognition') and NUC ('nuclease') lobes can rotate dramatically around each other, assuming an 'open' conformation in the apo Cas9 structure[7] and a 'closed' conformation in the Cas9–guide RNA[8,9] and Cas9–guide RNA–DNA R-loop[9] structures. Furthermore, single-molecule experiments established the importance of protospacer-adjacent motifs (PAMs) (5′-NGG-3′) for pausing at candidate targets[5,10], and R-loop formation was found to occur through directional strand invasion beginning at the PAM[5] (Fig. 1a). However, these findings did not explain the actions that Cas9 performs to interrogate each candidate target sequence. These actions, repeated over and over, comprise the slowest phase of Cas9's bacterial immune function and its induction of site-specific genome editing[11]. Understanding the mechanism of DNA interrogation is critical to determining how

Cas9 searches genomes to find bona fide targets and exclude the vast excess of nontarget sequences.

## Results and discussion

**Covalent cross-linking of Cas9 to DNA stabilizes the interrogation complex.** Evidence that Cas9's target engagement begins with PAM binding[5] indicates that during genome search, there exists a transient 'interrogation state' in which Cas9–guide RNA has engaged with a PAM but not yet formed RNA–DNA base pairs (Fig. 1a). Cas9–guide RNA complexes must repeatedly visit the interrogation state at each surveyed PAM, irrespective of the sequence of the adjacent 20-base-pair (bp) candidate complementarity region (CCR). While this state is the key to Cas9's DNA search mechanism, the interrogation complex has so far evaded structure determination due to its transience, with an estimated lifetime of <30 ms in bacteria[11].

To trap the Cas9 interrogation complex, we replaced residue Thr1337 with cysteine in *Streptococcus pyogenes* Cas9, the most widely used genome editing enzyme, and combined this protein with a single-guide RNA (sgRNA)[3] and a 30-bp DNA molecule functionalized with an $N^4$-cystamine cytosine modification[12] (Fig. 1b). The DNA included a PAM but lacked any complementarity to the sgRNA spacer (Fig. 1c). Reaction of the cysteine thiol with the cystamine creates a protein–DNA disulfide cross-link on the side of the PAM distal to the site of R-loop initiation (Fig. 1b,d). The position of the cross-link was chosen based on previous high-resolution structures of the Cas9–PAM interface[9,13] (Extended Data Fig. 1a). Incubation of Cas9 T1337C with sgRNA and the modified DNA duplex resulted in a decrease in electrophoretic mobility for roughly 70% of the total protein mass under denaturing but nonreducing conditions (Extended Data Fig. 1b), consistent with

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [2]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, CA, USA. [3]Innovative Genomics Institute, University of California, Berkeley, CA, USA. [4]Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Melbourne, Victoria, Australia. [5]Howard Hughes Medical Institute, University of California, Berkeley, CA, USA. [6]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]Department of Chemistry, University of California, Berkeley, CA, USA. [8]Gladstone Institutes, University of California, San Francisco, CA, USA. [9]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA. ✉e-mail: doudna@berkeley.edu
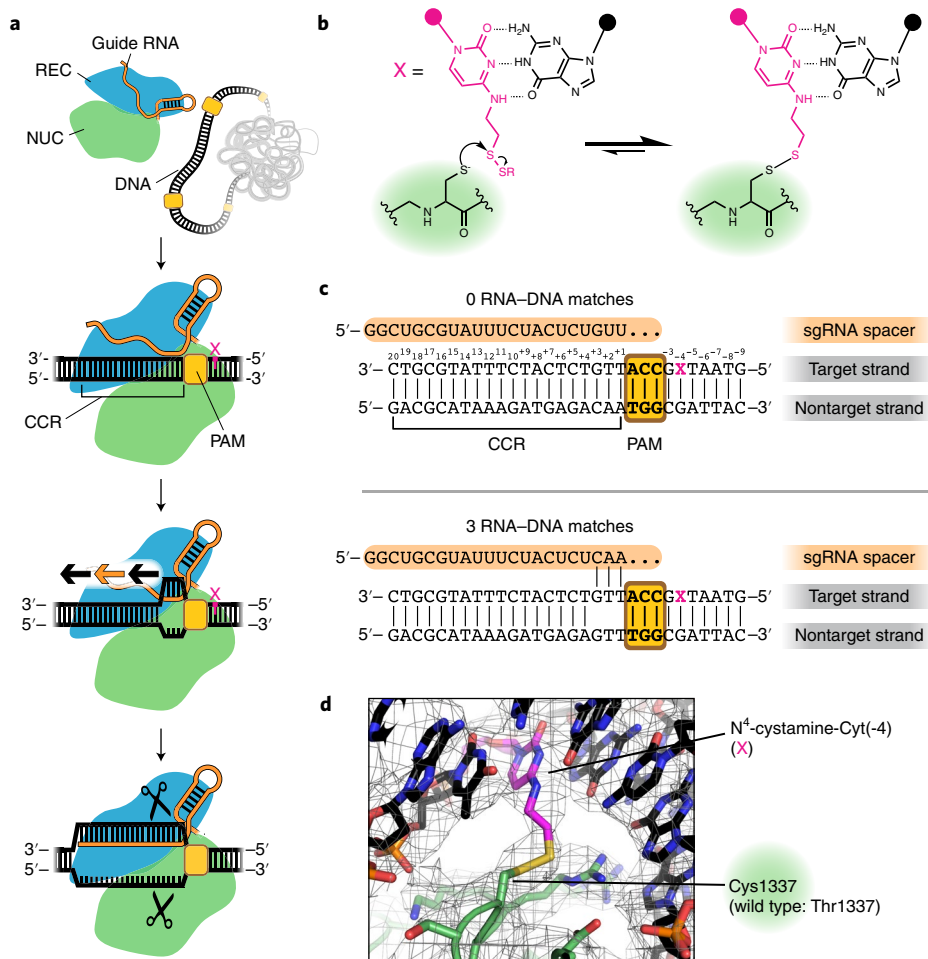
**Fig. 1 | Trapping the Cas9 interrogation complex. a**, Known steps leading to Cas9-catalyzed DNA cleavage. Orange and black arrows indicate the direction of guide RNA strand invasion into the DNA helix. The magenta X indicates the location of the cystamine modification. **b**, Chemistry of the protein–DNA cross-link. **c**, RNA and DNA sequences used in structural studies. **d**, Sharpened cryo-EM map (threshold 7σ) and model of the cross-linked Cas9–sgRNA–DNA complex (0 RNA–DNA matches, bent DNA), centered on density contributed by the nonnative thioalkane cross-link.

protein–DNA cross-link formation. The cross-link did not inhibit Cas9's ability to cleave sgRNA-complementary DNA (Extended Data Fig. 1c), suggesting that the enzyme is not grossly perturbed by the introduced disulfide. More importantly, mechanistic hypotheses revealed by cross-linked complexes can be tested in noncross-linked complexes.

We subjected the cross-linked interrogation complex to cryo-EM imaging and analysis (Table 1, Extended Data Figs. 2 and 3a,b). Ab initio volume reconstruction, refinement and modeling revealed two structural states of the complex. In one, the DNA lies as a linear duplex across the surface of the open form of the Cas9 ribonucleoprotein (Fig. 2a). In the other state, Cas9's two lobes pinch the DNA into a V shape whose helical arms meet at the site of R-loop initiation, using a bending mode that underwinds the DNA duplex (Fig. 2b).

**The linear-DNA conformation reveals a DNA scanning state of Cas9.** In the 'linear-DNA' conformation (Fig. 2a), the interface of the DNA with the PAM-interacting domain is similar to that seen in the crystal structure of Cas9–R-loop[9] (Extended Data Fig. 3c). However, the REC lobe of the protein is in a position radically different from that observed in all previous structures of nucleic acid-bound Cas9 (refs. [8,9,13,14]), having rotated away from the NUC lobe into an open-protein conformation that resembles the apo Cas9 crystal structure[7] (Fig. 2a).

Notably, a linear piece of DNA docked into the PAM-binding cleft would result in a severe structural clash[4] in either the Cas9–R-loop[9] or the Cas9–sgRNA (ref. [8]) crystal structure but only a minor one in the apo Cas9 crystal structure[7], which can be relieved by slightly tilting and bending the DNA (Extended Data Fig. 3c). We propose that the open-protein conformation, originally thought to be unique to nucleic acid-free Cas9, can also be adopted by the sgRNA-bound protein to enable its interaction with linear DNA. Indeed, cryo-EM analysis of the Cas9–sgRNA complex revealed only particles in the open-protein state (Table 1, Extended Data Fig. 4), indicating that the original crystal structure of Cas9–sgRNA (ref. [8]), which was in a closed-protein state, represented only one possible conformation of the complex that happened to be captured in that crystal form. Single-molecule Förster resonance energy transfer experiments also support the ability of Cas9–sgRNA to access both closed and open conformations[15]. The linear-DNA/open-protein conformation captured in our cryo-EM structure, then, may represent the conformation of Cas9 during any process for which it must accommodate a piece of linear DNA, such as during sliding[10] or initially engaging with a PAM.

**The bent-DNA conformation reveals PAM-adjacent DNA unwinding by Cas9.** In the 'bent-DNA' Cas9 interrogation complex, the protein grips the PAM as in the linear-DNA complex. The CCR, on the other hand, is tilted at a 50° angle to the PAM-containing

**Table 1 | Cryo-EM data collection, refinement and validation statistics**

| | Cas9–sgRNA–DNA (*S. pyogenes*) with 0 RNA–DNA base pairs, open-protein/linear-DNA conformation (EMD-24823) (PDB 7S3H) | Cas9–sgRNA–DNA (*S. pyogenes*) with 0 RNA–DNA base pairs, closed-protein/bent-DNA conformation (EMD-24817) (PDB 7S36) | Cas9–sgRNA (*S. pyogenes*) in the open-protein conformation (EMD-24818) (PDB 7S37) | Cas9–sgRNA–DNA (*S. pyogenes*) forming a 3-base-pair R-loop (EMD-24819) (PDB 7S38) |
|---|---|---|---|---|
| **Data collection and processing** | | | | |
| Magnification | ×81,000 | ×81,000 | ×36,000 | ×81,000 |
| Voltage (kV) | 300 | 300 | 200 | 300 |
| Electron exposure (e−/Å²) | 50 | 50 | 50 | 50 |
| Defocus range (μm) | −0.8 to −2 | −0.8 to −2 | −0.8 to −2 | −0.8 to −2 |
| Pixel size (Å) | 1.05 (counting)/0.525 (super-resolution) | 1.05 (counting)/0.525 (super-resolution) | 1.115 (counting)/0.5575 (super-resolution) | 1.05 (counting)/0.525 (super-resolution) |
| Symmetry imposed | C1 | C1 | C1 | C1 |
| Initial particle images (no.) | 8,386,564 | 8,386,564 | 1,594,370 | 5,006,233 |
| Final particle images (no.) | 545,450 | 18,121 | 87,130 | 17,424 |
| Map resolution (Å) | 2.5 | 3.2 | 3.2 | 3.3 |
| FSC threshold | 0.143 | 0.143 | 0.143 | 0.143 |
| Map resolution range (Å) | | | | |
| **Refinement** | | | | |
| Initial model used (PDB code) | 4ZT0, 5FQ5, 5F9R, 4CMP | 4ZT0, 5FQ5, 5F9R | 4ZT0, 5F9R, 4CMP | 4ZT0, 5FQ5, 5F9R |
| Model resolution (Å) | 3.0 | 3.2 | 5.3 | 3.3 |
| FSC threshold | 0.5 | 0.5 | 0.5 | 0.5 |
| Model resolution range (Å) | | | | |
| Map sharpening *B* factor (Å²) | | | | |
| Model composition | | | | |
| Nonhydrogen atoms | 9,966 | 13,821 | 9,392 | 13,987 |
| Protein residues | 1,030 | 1,353 | 1,033 | 1,353 |
| Nucleotides | 76 | 131 | 46 | 139 |
| Ligands | 1 | 1 | 0 | 1 |
| *B* factors (Å²) | | | | |
| Protein | 56.34 | 80.16 | 109.19 | 85.61 |
| Nucleotide | 68.92 | 104.85 | 225.80 | 119.07 |
| Ligand | 58.68 | 89.24 | – | 70.37 |
| R.m.s. deviations | | | | |
| Bond lengths (Å) | 0.007 | 0.007 | 0.006 | 0.007 |
| Bond angles (°) | 1.070 | 0.980 | 1.037 | 0.980 |
| Validation | | | | |
| MolProbity score | 1.42 | 1.36 | 1.49 | 1.33 |
| Clashscore | 3.77 | 2.98 | 4.66 | 2.62 |
| Poor rotamers (%) | 0.11 | 0.16 | 0.11 | 0.08 |
| Ramachandran plot | | | | |
| Favored (%) | 96.28 | 96.07 | 96.29 | 95.85 |
| Allowed (%) | 3.72 | 3.93 | 3.71 | 4.15 |
| Disallowed (%) | 0 | 0 | 0 | 0 |

helix and leans against the REC lobe, which has risen into the same 'closed' position as in the Cas9–sgRNA crystal structure (Figs. 2b and 3a). Compared to the high-resolution cryo-EM density contributed by the PAM-containing duplex, CCR density is poorly resolved (Fig. 3a and Extended Data Fig. 3b), reflecting conformational heterogeneity; however, even at low resolution, its distinct helical shape (Figs. 2b and 3a) enabled construction of an atomic model that adheres to B-form DNA constraints between the PAM-distal tip and position +3 of the CCR (Fig. 3b). Connection of this B-form helix to the PAM-containing helix requires backbone distortion and helical underwinding at precisely the position from which an R-loop would be initiated if the sgRNA were complementary to the CCR[5] (Fig. 3b,c). Underwinding is accomplished through major groove compression (Fig. 3b), as observed for other protein-induced bends[16–23].

At the distorted bending vertex, a missing wedge of cryo-EM density appears across from nontarget-strand nucleotides Ade(+1)
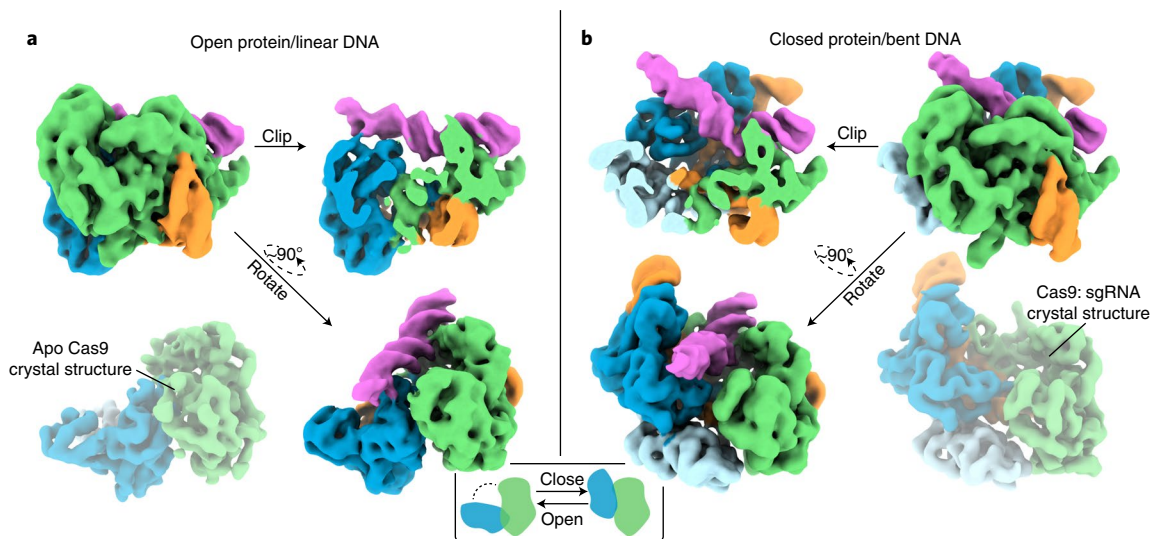
**Fig. 2 | Cryo-EM structures of the Cas9 interrogation complex, compared to previously determined crystal structures. a**, Unsharpened cryo-EM map (threshold 4σ) of Cas9 interrogation complex in an open-protein/linear-DNA conformation, alongside apo Cas9 crystal structure (PDB 4CMP, 2F₀–F꜀, threshold 1.5σ). **b**, Closed-protein/bent-DNA conformation (threshold 5σ), alongside Cas9–sgRNA crystal structure (PDB 4ZT0, 2F₀–F꜀, threshold 1.5σ). Green, NUC lobe; blue, REC lobe domains 1/2; light blue, REC lobe domain 3; orange, guide RNA and magenta, DNA. Maps produced in the present work are displayed with full opacity. REC lobe domain 3 does not appear in the cryo-EM structure in **a** (Supplementary Information). Additional classes observed in the closed-protein state are shown in Extended Data Fig. 3a.
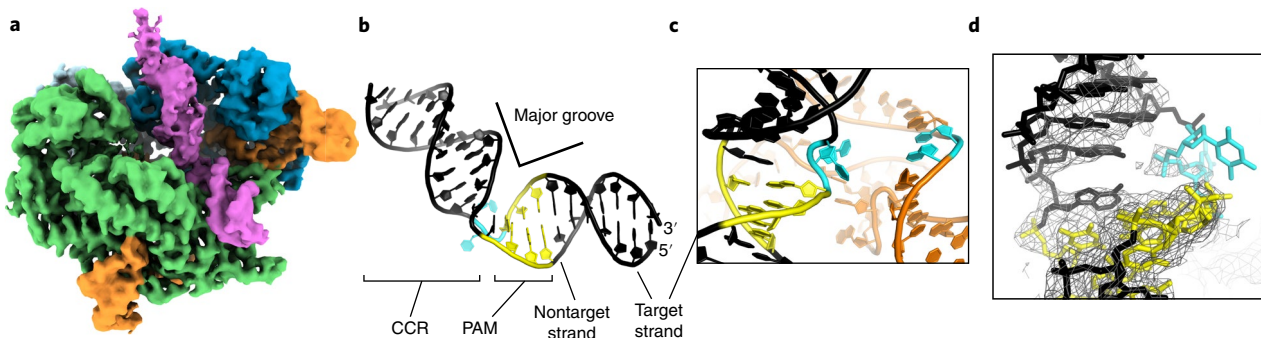


**Fig. 3 | DNA conformation at the site of bending. a**, Unsharpened cryo-EM map (threshold 5σ) of Cas9 interrogation complex in closed-protein/bent-DNA conformation. Green, NUC lobe; blue, REC lobe; orange, guide RNA and magenta, DNA. **b**, Bent-DNA model. Yellow, PAM; cyan, target-strand Thy(+1) and Thy(+2). **c**, DNA (black) and RNA (orange) models, demonstrating the proximity of the first DNA–RNA base pairs (cyan) that would form if they were complementary. **d**, DNA model within sharpened cryo-EM density (threshold 7.5σ). The break in density in the target strand indicates dramatic conformational heterogeneity. The modeled conformation of Thy(+1) and Thy(+2) (cyan), which represents just one possible conformation within a diverse ensemble, was chosen based on permanganate reactivity data and geometric constraints imposed by neighboring nucleotides.

and Ade(+2), suggesting that target-strand nucleotides Thy(+1) and Thy(+2) have become unpaired from their partners (Fig. 3d). The overall weakness of density for Thy(+1) and Thy(+2) suggests dramatic mobility, and the modeled conformation of those nucleotides represents a physically plausible member of a diverse conformational ensemble (which also agrees with solution experiments to be discussed shortly). Therefore, in the bent-DNA conformation, two helical arms join at an underwound hinge whose target-strand nucleotides are heterogeneously positioned. DNA disorder is consistent with the function of the Cas9 interrogation complex, which is to flip target-strand nucleotides from the DNA duplex toward the sgRNA to test base pairing potential.

**Cas9–sgRNA bends DNA in noncross-linked complexes.** To determine whether unmodified Cas9 can bend DNA, we produced interrogation complexes that lacked the cross-link and

tested them in a DNA cyclization assay[24] (Fig. 4a). We created a series of 160-bp double-stranded DNA substrates that all bore a 'J' shape due to the inclusion of a special A-tract sequence that forms a protein-independent 108° bend[25]. Each substrate also included two PAMs spaced by a near-integral number of B-form DNA turns (31 bp). In 11 versions of this substrate, we varied the number of base pairs between the A-tract and the proximal PAM from 21 to 31 bp, effectively rotating the Cas9 binding sites around an entire turn of a B-form DNA helix (Extended Data Fig. 5). If Cas9 bends the DNA, each additional base pair added to the variable (21–31 bp) region will turn the Cas9-induced bend by roughly 34° with respect to the fixed A-tract bend. The relative direction of the two bends can be discerned from each substrate's ligase-catalyzed cyclization efficiency, which should increase when the two bends point in the same direction (DNA assumes a 'C' shape) and decrease when the two bends point in opposite directions (DNA assumes an 'S'
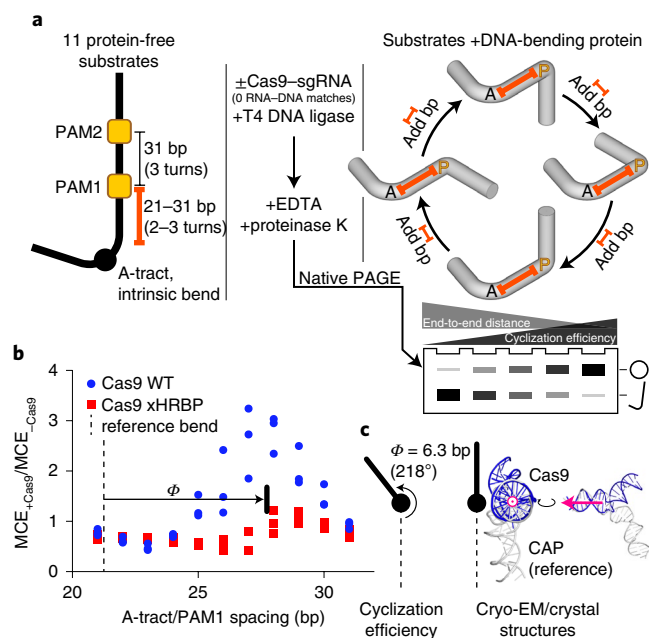
**Fig. 4 | DNA cyclization efficiency experiments. a**, Substrate structure and experimental pipeline. Black A, A-tract; yellow P, PAM and orange spacer, A-tract/PAM1 distance (21–31 bp). For simplicity, bending is only depicted at a single PAM in the cylindrical volume illustrations. Substrates that are more S-shaped (left side of the cycle diagram/gel) cyclize more slowly than substrates that are C-shaped (right) due to changes in the end-to-end distance of the molecules. When base pairs are added to the spacer, the cyclization efficiency is expected to rise as the Cas9-induced bend becomes aligned with the A-tract bend, then fall as the bends become misaligned again, in a roughly sinusoidal pattern. **b**, Cas9-dependent cyclization enhancement of 11 substrate variants. Three replicates are depicted. MCE, monomolecular cyclization efficiency; xHRBP, mutated helix-rolling basic patch (K233A/K234A/K253A/K263A) and $\Phi$, phase difference from reference bend to the Cas9 wild type (WT) peak. A protein that does not bend the DNA at all would yield the line $y = 1$. A protein that bends DNA in a different direction would yield an $x$-shifted sinusoid that peaked at a different value of spacer length. **c**, Comparison of bending phase difference in the cyclization experiments versus cryo-EM/crystal structures of DNA bends introduced by Cas9 or CAP (PDB 1CGP). See Supplementary Information for discussion of the CAP-based reference bend analysis. The magenta vector superposed on the aligned helices would point toward the A-tract in the cyclization substrates; in the larger structural diagram, it points out of the page.

shape), as a function of the proximity of the DNA ends to be sealed (Fig. 4a). We measured the cyclization efficiency of each substrate in the absence and presence of Cas9 and an sgRNA lacking homology to either of the two CCR sequences. Consistent with expectations for a bend, the Cas9-dependent enhancement (or reduction) of cyclization efficiency tracked a sinusoidal shape when plotted against the A-tract/PAM spacing, reflecting phase-dependent variation in the end-to-end distance of different substrates (Fig. 4b and Extended Data Fig. 6a,b). Additionally, by interpreting the absolute phase of the cyclization enhancement curve (that is, the spacing value at which the peak occurs, where the two bends point in the same direction) in the context of the known direction of the A-tract bend[24,26], we conclude that the bending direction observed in this experiment is the same as that observed in the bent-DNA cryo-EM structure (Fig. 4c, Extended Data Fig. 6b and Supplementary Information).

Next, we wondered whether local DNA conformations observed in the cross-linked interrogation complex resemble those in the native complex. To characterize DNA distortion with single-nucleotide resolution, we measured the permanganate reactivity of individual thymines in the target DNA strand of a noncross-linked interrogation complex (Fig. 5a). As anticipated for protein-induced base unstacking[27,28], we detected a PAM- and Cas9-dependent increase in permanganate reactivity at Thy(+1) and Thy(+2) (Fig. 5a–c and Extended Data Fig. 7a,b). The relationship between permanganate reactivity and Cas9–sgRNA concentration at these thymines suggests that the affinity of Cas9–sgRNA for this sequence is weak ($10\,\mu M$), as expected for this necessarily transient interaction with off-target DNA (Fig. 5b and Extended Data Fig. 7a). Thy(+1) and Thy(+2) are precisely the nucleotides that appeared to be unpaired in the bent-DNA cryo-EM map of the cross-linked Cas9 interrogation complex (Fig. 3d), which shared the same DNA sequence as the permanganate substrate. These results indicate that Cas9 bends DNA through a backbone distortion that exposes target-strand nucleobases +1 and +2 to solvent and, more generally, that cryo-EM analysis of the cross-linked complex captured meaningful structural features of the native complex.

**A Cas9 conformational rearrangement accompanies DNA bending.** The described linear- and bent-DNA conformations present a new model for Cas9 function in which open-protein Cas9–sgRNA first associates with the PAM on linear DNA, then engages a switch to the closed-protein state to bend the DNA and expose its PAM-adjacent nucleobases for interrogation (Supplementary Video 1). Because this transition involves energetically unfavorable base unstacking, we wondered how the unstacked state is stabilized.

The 'phosphate lock loop' (Lys1107-Ser1109), which was proposed to support R-loop nucleation by tugging on the target-strand phosphate between the PAM and nucleotide +1 (ref. [13]), is disordered in the linear-DNA structure but stably bound to the target strand in the bent-DNA structure (Extended Data Fig. 8a,b), highlighting this contact as a potential energetic compensator for the base unstacking penalty. In the permanganate assay, mutation of the phosphate lock loop decreased activity to the level observed without Cas9 or with a Cas9 mutant deficient in PAM recognition (which lacks the PAM-binding arginines, 'xPBA') (Fig. 5c and Extended Data Fig. 7b), indicating that the loop may play a role in DNA bending. However, a negative result in this assay could be attributed either to weakened DNA bending activity or to an overall destabilization of the protein–DNA interaction[4].

Another notable structural element is a group of lysines (Lys233/Lys234/Lys253/Lys263, termed here the 'helix-rolling basic patch') on REC2 (REC lobe domain 2) that contact the DNA phosphate backbone (at bp +8 to +13) in both the linear- and bent-DNA structures, an interaction that has not been observed before (Extended Data Fig. 8a,c). Mutation of these lysines attenuated anisotropy in the cyclization assay (Fig. 4b and Extended Data Fig. 6a) and abolished Cas9's permanganate sensitization activity (Fig. 5c and Extended Data Fig. 7b). Structural modeling of the linear-to-bent transition (Supplementary Video 2) suggests that the helix-rolling basic patch may couple DNA bending to inter-lobe protein rotations similar to those observed in multi-body refinements[29] of the cryo-EM images (Supplementary Videos 3 and 4). Consensus EM reconstructions also revealed large segments of the REC lobe and guide RNA that become ordered on lobe closure (Supplementary Video 1 and Supplementary Information), indicating that Cas9 can draw on diverse structural transitions across the complex to regulate DNA bending.

**The bent-DNA state makes R-loop nucleation structurally accessible.** We propose that the function of the bent-DNA conformation is to promote local base flipping that can lead to R-loop nucleation. To probe the structure of a complex that has already proceeded to
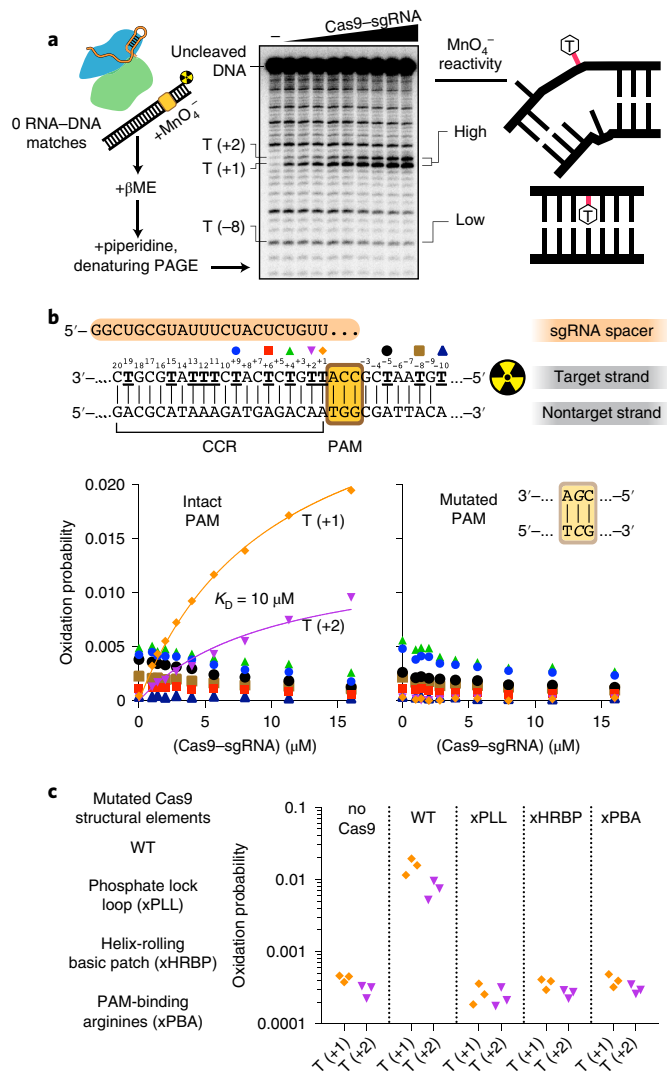
**Fig. 5 | Permanganate reactivity measurements. a**, Experimental pipeline. The autoradiograph depicts the raw data used to produce the 'intact PAM' graph in **b**. **b**, Oxidation probability of select thymines as a function of [Cas9–sgRNA]. Data depict a single replicate. Model information and additional replicates/thymines are presented in Extended Data Fig. 7a. **c**, Oxidation probability of T(+1) and T(+2) in the presence of the indicated Cas9–sgRNA variant ([Cas9–sgRNA] = 16 μM). Three replicates depicted. xPLL, KES(1107-1109)GG; xHRBP, K233A/K234A/K253A/K263A and xPBA, R1333A/R1335A.

the R-loop nucleation step, we used the same cross-linking strategy with adjusted RNA and DNA sequences that allow partial R-loop formation (Fig. 1c). Cryo-EM analysis of this construct revealed nucleotides +1 to +3 of the DNA target strand hybridized to the sgRNA spacer (Table 1, Fig. 6a and Extended Data Fig. 9). In contrast to the disorder that characterized this region in the bent-DNA map, all three nucleotides are well-resolved, apparently stabilized by their hybridization to the A-form sgRNA spacer. The increase in resolution extends to the nontarget strand and to the more PAM-distal regions of the CCR, suggesting that the DNA becomes overall more ordered in response to R-loop nucleation. The ribonucleoprotein architecture resembles that of the bent-DNA structure except for slight tilting of REC2, which accommodates a repositioning of the CCR duplex toward the newly formed RNA–DNA base pairs (Fig. 6a). Therefore, unstacked nucleotides in the bent-DNA

state can hybridize to the sgRNA spacer with minimal global structural changes, further supporting the bent-DNA structure as a gateway to R-loop nucleation.

Our structures outline a model for a poorly understood aspect of Cas9 function that is fundamental to CRISPR target search and capture (Fig. 6b). First, open-conformation Cas9–sgRNA associates with the PAM of a linear-DNA target. By engaging the open-to-closed protein conformational switch, Cas9 bends and twists the DNA to locally unwind the base pairs next to the PAM. If target-strand nucleotides are unable to hybridize to the sgRNA spacer, the candidate target is released and Cas9 proceeds to the next candidate. If the target strand is sgRNA-complementary, unwound nucleotides initiate an RNA–DNA hybrid that can expand through strand invasion to a full 20-bp R-loop, activating DNA cleavage.

Due to its energetic linkage to base flipping[30], DNA bending provides a viable mechanical solution to any biological problem that requires unrestricted access to nucleobases[31,32], including the sequence interrogation challenge faced by all DNA-targeting CRISPR systems[4,33–36]. In our key structural snapshot of this process, Cas9 specifically uses a bending mode that involves underwinding (Fig. 3), which may be a topological necessity for the downstream propagation of flipping events, and which could underlie some features of Cas9's mechanical sensitivity[37–39]. In contrast, certain methyltransferases that flip only one nucleotide at a time can afford to do so without gross helical distortion[40,41].

Other proteins define the vertex of an underwound DNA bend using intimate contacts to the distorted nucleotides, either to intact base pairs in the case of transcription factors[18–20,23] or to flipped bases and their estranged partners in the case of base excision repair enzymes[16,17,21,22,42]. During initial DNA interrogation, Cas9 appears to make no such contacts, instead straddling the bending vertex and relying on mechanical strain to stabilize extrahelical nucleotide conformations without restricting their motion. To catalyze RNA-programmable strand exchange, Cas9 must distort DNA independent of its sequence, likening its functional constraints to those of the filamentous recombinase RecA[43], which nonspecifically destabilizes candidate DNA via longitudinal stretching[44–46].

The mechanism illustrated here reveals the individual steps that comprise the slowest phase of Cas9's genome editing function[11]. The energetic tuning of binding, bending and RNA–DNA hybrid nucleation dictates the speed of target capture—bent/unstacked states must be stable enough to promote fast transitions to RNA-hybridized states, but not so stable that Cas9 wastes undue time on off-target DNA. Thus, the energetic landscape surrounding the states identified in this work will be a crucial subject of study to understand the success of current state-of-the-art genome editors and to inform the engineering of faster ones. Finally, DNA in eukaryotic chromatin is rife with bends, due either to intrinsic structural features of the DNA sequence[47] or to interactions with looping proteins[48]. Because the Cas9-induced DNA bend described here has a well-defined direction that may either match or antagonize incumbent bends, it will be important to test how local chromatin geometry affects Cas9's efficiency in both dissociating from off-target sequences and opening R-loops on real target sequences.

## 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/s41594-022-00756-0.
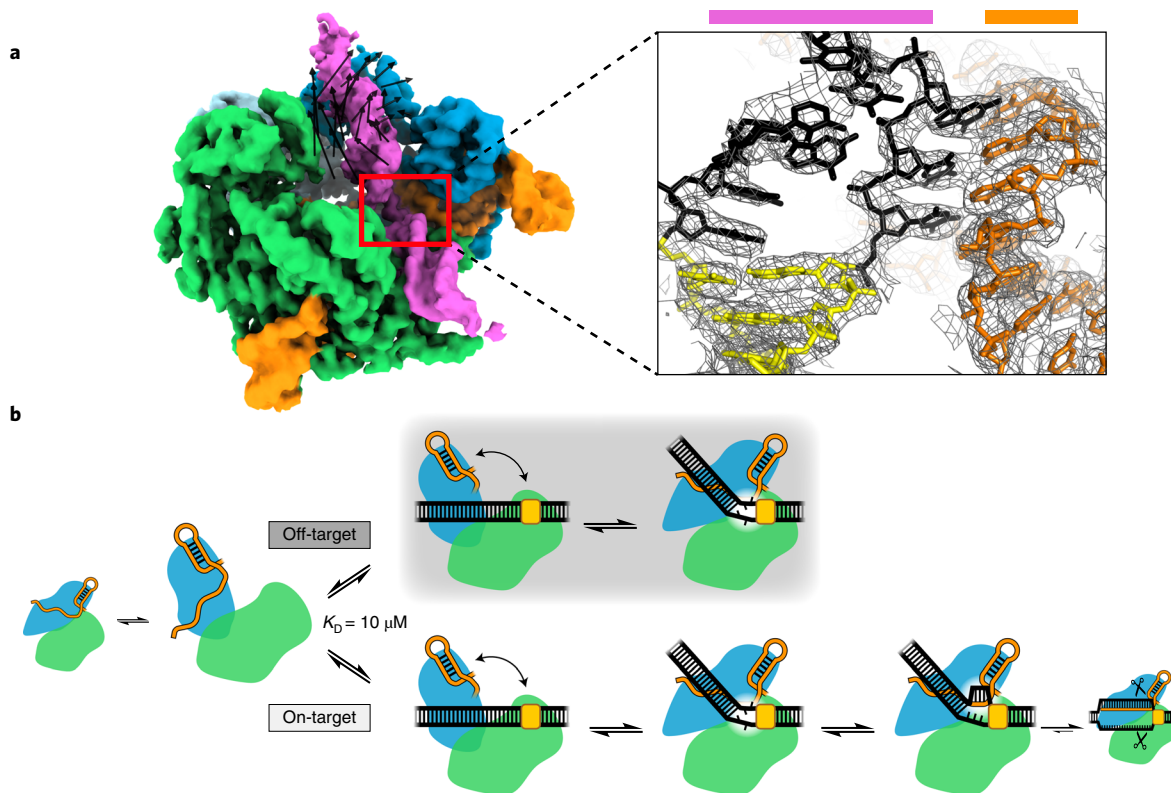
**Fig. 6 | Structure of a nascent R-loop and overview model. a**, Left, unsharpened cryo-EM map (threshold 5σ) of Cas9–sgRNA–DNA with 3 RNA–DNA bp. Green, NUC lobe; blue, REC lobe; orange, guide RNA and magenta, DNA. Black vectors indicate differences from the bent-DNA structure with 0 RNA–DNA matches. The primary difference is a rigid-body rotation of REC lobe domain 2. Inset, sharpened cryo-EM map (threshold 8σ) and model. For clarity, only DNA and RNA are shown. Black, CCR; yellow, PAM. **b**, Model for bend-dependent Cas9 target search and capture. Large diagrams depict states structurally characterized in the present work.

## References

1. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
2. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
3. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
4. Mekler, V., Minakhin, L. & Severinov, K. Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. *Proc. Natl Acad. Sci. USA* **114**, 5443–5448 (2017).
5. 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 (2014).
6. Jiang, F. & Doudna, J. A. CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.* **46**, 505–529 (2017).
7. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
8. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
9. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871 (2016).
10. Globyte, V., Lee, S. H., Bae, T., Kim, J.-S. & Joo, C. CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion. *EMBO J.* **38**, e99466 (2019).
11. Jones, D. L. et al. Kinetics of dCas9 target search in *Escherichia coli*. *Science* **357**, 1420–1424 (2017).
12. Verdine, G. L. & Norman, D. P. G. Covalent trapping of protein-DNA complexes. *Annu. Rev. Biochem.* **72**, 337–366 (2003).
13. 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).
14. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
15. Osuka, S. et al. Real-time observation of flexible domain movements in CRISPR-Cas9. *EMBO J.* **37**, e96941 (2018).
16. Bruner, S. D., Norman, D. P. & Verdine, G. L. Structural basis for recognition and repair of the endogenous mutagen 8-oxoguanine in DNA. *Nature* **403**, 859–866 (2000).
17. Fromme, J. C. & Verdine, G. L. Structure of a trapped endonuclease III-DNA covalent intermediate. *EMBO J.* **22**, 3461–3471 (2003).
18. Kim, J. L., Nikolov, D. B. & Burley, S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527 (1993).
19. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520 (1993).
20. Love, J. J. et al. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* **376**, 791–795 (1995).
21. Slupphaug, G. et al. A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* **384**, 87–92 (1996).
22. Vassylyev, D. G. et al. Atomic model of a pyrimidine dimer excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell* **83**, 773–782 (1995).
23. Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* **81**, 705–714 (1995).
24. Kahn, J. D. & Crothers, D. M. Protein-induced bending and DNA cyclization. *Proc. Natl Acad. Sci. USA* **89**, 6343–6347 (1992).
25. Koo, H. S., Drak, J., Rice, J. A. & Crothers, D. M. Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29**, 4227–4234 (1990).
26. Schultz, S. C., Shields, G. C. & Steitz, T. A. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science* **253**, 1001–1007 (1991).
27. Bui, C. T., Rees, K. & Cotton, R. G. H. Permanganate oxidation reactions of DNA: perspective in biological studies. *Nucleosides Nucleotides Nucleic Acids* **22**, 1835–1855 (2003).
28. Cofsky, J. C. et al. CRISPR-Cas12a exploits R-loop asymmetry to form double-strand breaks. *eLife* **9**, e55143 (2020).
29. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *eLife* **7**, e36861 (2018).
30. Ramstein, J. & Lavery, R. Energetic coupling between DNA bending and base pair opening. *Proc. Natl Acad. Sci. USA* **85**, 7231–7235 (1988).

31. Allan, B. W. et al. DNA bending by EcoRI DNA methyltransferase accelerates base flipping but compromises specificity. *J. Biol. Chem.* **274**, 19269–19275 (1999).

32. Su, T.-J., Tock, M. R., Egelhaaf, S. U., Poon, W. C. K. & Dryden, D. T. F. DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.* **33**, 3235–3244 (2005).

33. Blosser, T. R. et al. Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* **58**, 60–70 (2015).

34. Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E. & Doudna, J. A. DNA targeting by a minimal CRISPR RNA-guided cascade. *Mol. Cell* **63**, 840–851 (2016).

35. Westra, E. R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).

36. Xiao, Y. et al. Structure basis for directional r-loop formation and substrate handover mechanisms in Type I CRISPR-Cas system. *Cell* **170**, 48–60.e11 (2017).

37. Ivanov, I. E. et al. Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling. *Proc. Natl Acad. Sci. USA* **117**, 5853–5860 (2020).

38. Newton, M. D. et al. DNA stretching induces Cas9 off-target activity. *Nat. Struct. Mol. Biol.* **26**, 185–192 (2019).

39. Szczelkun, M. D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl Acad. Sci. USA* **111**, 9798–9803 (2014).

40. Klimasauskas, S., Kumar, S., Roberts, R. J. & Cheng, X. HhaI methyltransferase flips its target base out of the DNA helix. *Cell* **76**, 357–369 (1994).

41. Reinisch, K. M., Chen, L., Verdine, G. L. & Lipscomb, W. N. The crystal structure of HaeIII methyltransferase convalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143–153 (1995).

42. Dalhus, B., Laerdahl, J. K., Backe, P. H. & Bjørås, M. DNA base repair–recognition and initiation of catalysis. *FEMS Microbiol. Rev.* **33**, 1044–1078 (2009).

43. Bell, J. C. & Kowalczykowski, S. C. RecA: regulation and mechanism of a molecular search engine. *Trends Biochem. Sci.* **41**, 491–507 (2016).

44. Chen, Z., Yang, H. & Pavletich, N. P. Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. *Nature* **453**, 489–484 (2008).

45. Yang, D., Boyer, B., Prévost, C., Danilowicz, C. & Prentiss, M. Integrating multi-scale data on homologous recombination into a new recognition mechanism based on simulations of the RecA-ssDNA/dsDNA structure. *Nucleic Acids Res.* **43**, 10251–10263 (2015).

46. Yang, H., Zhou, C., Dhar, A. & Pavletich, N. P. Mechanism of strand exchange from RecA-DNA synaptic and D-loop structures. *Nature* **586**, 801–806 (2020).

47. Koo, H. S., Wu, H. M. & Crothers, D. M. DNA bending at adenine·thymine tracts. *Nature* **320**, 501–506 (1986).

48. Garcia, H. G. et al. Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers* **85**, 115–130 (2007).

## Methods

**Protein expression and purification.** Cas9 was expressed and purified as described previously[28], with slight modifications. Briefly, protein was expressed from custom pET-based vectors in *Escherichia coli* BL21 Star(DE3) cells. Cells were sonicated in lysis buffer (50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 0.5 mM PMSF, 10 tablets per l cOmplete EDTA-free protease inhibitor cocktail (Roche), 0.25 mg ml⁻¹ chicken egg white lysozyme (Sigma-Aldrich)) and clarified lysate was loaded onto Ni-NTA resin, which was then washed (50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 5% glycerol, 20 mM imidazole) and eluted (50 mM HEPES (pH 7.5), 500 mM NaCl, 1 mM TCEP, 5% glycerol, 300 mM imidazole). Proteins were cleaved overnight with tobacco etch virus protease at 4 °C without dialysis. The digested protein solution was diluted with one volume of low-salt ion exchange buffer (50 mM HEPES (pH 7.5), 250 mM KCl, 1 mM TCEP, 10% glycerol). Digested protein was purified on a HiTrap Heparin HP affinity column (Cytiva), eluting with high-salt ion exchange buffer (50 mM HEPES (pH 7.5), 1 M KCl, 1 mM TCEP, 10% glycerol). Eluted protein was then purified by size exclusion in protein-purification size-exclusion buffer (20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM dithiothreitol (DTT), 10% glycerol) on a Superdex 200 Increase 10/300 GL column (Cytiva). Plasmid/protein sequences and Addgene IDs can be found in the Supplementary Information.

**Nucleic acid preparation.** All DNA oligonucleotides were synthesized by Integrated DNA Technologies except the cystamine-functionalized target strand, which was synthesized by TriLink Biotechnologies (with high-performance liquid chromatography purification). DNA oligonucleotides that were not high-performance liquid chromatography purified by the manufacturer were PAGE-purified in house (unless a downstream preparative step involved another PAGE purification), and all DNA oligonucleotides were stored in water. Duplex DNA substrates were annealed by heating to 95 °C and cooling to 25 °C over the course of 40 min on a thermocycler. Guide RNAs were transcribed and purified as described previously[28], except no ribozyme was included in the transcript. Briefly, in vitro transcription reactions included PCR-assembled DNA template, 40 mM Tris-Cl (pH 7.9 at 25 °C), 25 mM MgCl₂, 10 mM DTT, 0.01% (v/v) Triton X-100, 2 mM spermidine, 5 mM of each nucleoside triphosphate and 100 μg ml⁻¹ T7 RNA polymerase. Transcription was allowed to proceed for 2.5 h at 37 °C, after which RNA was purified by urea-PAGE, ethanol-precipitated and resuspended in RNA storage buffer (0.1 mM EDTA, 2 mM sodium citrate, pH 6.4). All sgRNA molecules were annealed (80 °C for 2 min, then moved directly to ice) in RNA storage buffer before use. For both DNA and RNA, A₂₆₀ was measured on a NanoDrop (Thermo Scientific), and concentration was estimated according to extinction coefficients reported previously[49]. Oligonucleotide sequences can be found in the Supplementary Information.

**Cryo-EM construct preparation.** DNA duplexes were pre-annealed in water at 10× concentration (60 μM target strand, 75 μM nontarget strand). Cross-linking reactions were assembled with 300 μl of water, 100 μl of 5× disulfide reaction buffer (250 mM Tris-Cl, pH 7.4 at 25 °C, 750 mM NaCl, 25 mM MgCl₂, 25% glycerol, 500 μM DTT), 50 μl of 10× DNA duplex, 25 μl of 100 μM sgRNA and 25 μl of 80 μM Cas9. Cross-linking was allowed to proceed at 25 °C for 24 h (0 RNA–DNA matches) or 8 h (3 RNA–DNA matches). Sample was then purified by size exclusion (Superdex 200 Increase 10/300 GL, Cytiva) in cryo-EM buffer (20 mM Tris-Cl, pH 7.5 at 25 °C, 200 mM KCl, 100 μM DTT, 5 mM MgCl₂, 0.25% glycerol). Peak fractions were pooled, concentrated to an estimated 6 μM, snap-frozen in 10-μl aliquots in liquid nitrogen and stored at −80 °C until grid preparation. For the Cas9–sgRNA structural construct, which lacked a cross-link, the reaction was assembled with 350 μl of water, 100 μl of 5× disulfide reaction buffer, 0.45 μl of 1 M DTT, 25 μl of 100 μM sgRNA and 25 μl of 80 μM Cas9. The complex was allowed to form at 25 °C for 30 min. The Cas9–sgRNA sample was then size-exclusion purified and processed as described for the DNA-containing constructs. For Cas9–sgRNA, cryo-EM buffer contained 1 mM DTT instead of 100 μM DTT.

**SDS–PAGE analysis.** For nonreducing SDS–PAGE, thiol exchange was first quenched by the addition of 20 mM S-methyl methanethiosulfonate (S-MMTS). Then, 0.25 volumes of 5× nonreducing SDS–PAGE loading solution (0.0625% w/v bromophenol blue, 75 mM EDTA, 30% glycerol, 10% SDS, 250 mM Tris-Cl, pH 6.8) were added, and the sample was heated to 90 °C for 5 min before loading of 3 pmol onto a 4–15% Mini-PROTEAN TGX Stain-Free Precast Gel (Bio-Rad), alongside PageRuler Prestained Protein Ladder (Thermo Scientific). Gels were imaged using the Stain-Free imaging protocol (5-min activation, 3-s exposure) of Bio-Rad Image Lab v.5.2.1 on a Bio-Rad ChemiDoc. For reducing SDS–PAGE, no S-MMTS was added and 5% β-mercaptoethanol (βME) was added along with the nonreducing SDS–PAGE loading solution. For radioactive SDS–PAGE analysis, a 4–20% Mini-PROTEAN TGX Precast Gel (Bio-Rad) was prerun for 20 min at 200 V (to allow free ATP to migrate ahead of free DNA), run with radioactive sample for 15 min at 200 V, dried (80 °C, 3 h) on a gel dryer (Bio-Rad) and exposed to a phosphor screen, subsequently imaged on an Amersham Typhoon using the Amersham Typhoon Control Software v.2.0.0.6 (Cytiva).

**Nucleic acid radiolabeling.** Standard 5′ radiolabeling was performed with T4 polynucleotide kinase (New England BioLabs) at 0.2 U μl⁻¹ (manufacturer's

units), 1× T4 PNK buffer (New England BioLabs), 400 nM DNA oligonucleotide and 200 nM [γ-³²P]-ATP (PerkinElmer) for 30 min at 37 °C, followed by a 20-min heat-killing incubation at 65 °C. Radiolabeled oligos were then buffer exchanged into water using a Microspin G-25 spin column (GE Healthcare). For 5′ radiolabeling of sgRNAs, the 5′ triphosphate was first removed by treatment with Quick CIP (New England BioLabs, manufacturer's instructions). The reaction was then supplemented with 5 mM DTT and the same concentrations of T4 polynucleotide kinase (New England BioLabs) and [γ-³²P]-ATP (PerkinElmer) used for DNA radiolabeling, and the remainder of the protocol was completed as for DNA.

**Radiolabeled target-strand cleavage rate measurements.** DNA duplexes at 10× concentration (20 nM radiolabeled target strand, 75 μM unlabeled nontarget strand) were annealed in water with 60 μM cystamine dihydrochloride (pH 7). A 75-μl reaction was assembled from 15 μl of 5× Mg-free disulfide reaction buffer (250 mM Tris-Cl, pH 7.4 at 25 °C, 750 mM NaCl, 5 mM EDTA, 25% glycerol, 500 μM DTT), 7.5 μl of 600 μM cystamine dihydrochloride (pH 7), 37.5 μl of water, 3.75 μl of 80 μM Cas9, 3.75 μl of 100 μM sgRNA, 7.5 μl of 10× DNA duplex. The reaction was incubated at 25 °C for 2 h, at which point the cross-linked fraction had fully equilibrated. To nonreducing or reducing reactions, 5 μl of 320 mM S-MMTS or 80 mM DTT (respectively) in 1× Mg-free disulfide reaction buffer was added. Samples were incubated at 25 °C for an additional 5 min, then cooled to 16 °C and allowed to equilibrate for 15 min. One aliquot was quenched into 0.25 volumes 5× nonreducing SDS–PAGE solution and subject to SDS–PAGE analysis to assess the extent of cross-linking (for the reduced sample, no βME was added, as the DTT had already effectively reduced the sample). Another aliquot was quenched for reducing urea-PAGE analysis as timepoint 0. DNA cleavage was initiated by combining the remaining reaction volume with 0.11 volumes 60 mM MgCl₂. Aliquots were taken at the indicated timepoints for reducing urea-PAGE analysis.

**Urea-PAGE analysis.** To each sample was added 1 volume of 2× urea-PAGE loading solution (92% formamide, 30 mM EDTA, 0.025% bromophenol blue, 400 μg ml⁻¹ heparin). For reducing urea-PAGE analysis, 5% βME was subsequently added. Samples were heated to 90 °C for 5 min, then resolved on a denaturing polyacrylamide gel (10% or 15% acrylamide:bis-acrylamide 29:1, 7 M urea, 0.5× TBE). For radioactive samples, gels were dried (80 °C, 3 h) on a gel dryer (Bio-Rad), exposed to a phosphor screen and imaged on an Amersham Typhoon (Cytiva). For samples containing fluorophore-conjugated DNA, gels were directly imaged on the Typhoon without further treatment. For unlabeled samples, gels were stained with 1× SYBR Gold (Invitrogen) in 0.5× TBE before Typhoon imaging.

**Fluorescence and autoradiograph data analysis.** Band volumes in fluorescence images and autoradiographs were quantified in Image Lab v.6.1 (Bio-Rad). For fluorescence images recorded by the ChemiDoc, Image Lab's native *.scn files were used for quantification. For images recorded by the Typhoon, the Typhoon software's native *.gel files (square root encoded) were used for quantification. Data were fit by the least-squares method in Prism 7 (GraphPad Software).

**Cryo-EM grid preparation and data collection.** Cryo-EM samples were thawed and diluted to 3 μM (Cas9–sgRNA–DNA) or 1.5 μM (Cas9–sgRNA) in cryo-EM buffer. An UltrAuFoil grid (1.2/1.3-μm, 300 mesh, Electron Microscopy Sciences, catalog no. Q350AR13A) was glow-discharged in a PELCO easiGlow for 15 s at 25 mA, then loaded into an FEI Vitrobot Mark IV equilibrated to 8 °C with 100% humidity. From the sample, kept on ice up until use, 3.6 μl was applied to the grid, which was immediately blotted (Cas9–sgRNA–DNA (0 RNA–DNA matches) and Cas9–sgRNA: blot time 4.5 s, blot force 8; Cas9–sgRNA–DNA (3 RNA–DNA matches): blot time 3 s, blot force 6) and plunged into liquid nitrogen-cooled ethane. Micrographs for Cas9–sgRNA were collected on a Talos Arctica transmission electron microscope operated at 200 kV and ×36,000 magnification (1.115 Å per pixel), at −0.8 to −2 μm defocus, using the super-resolution camera setting (0.5575 Å per pixel) on a Gatan K3 Direct Electron Detector. Micrographs for Cas9–sgRNA–DNA complexes were collected on a Titan Krios G3i transmission electron microscope operated at 300 kV with energy filter, ×81,000 nominal magnification (1.05 Å per pixel), −0.8 to −2 μm defocus, using the super-resolution camera setting (0.525 Å per pixel) in CDS mode on a Gatan K3 Direct Electron Detector. All images were collected using beam shift in SerialEM v.3.8.7 software.

**Cryo-EM data processing and model building.** Details of cryo-EM data processing and model building can be found in the Supplementary Information.

**Permanganate reactivity measurements.** DNA duplexes were annealed at 50× concentration (100 nM radiolabeled target strand, 200 nM unlabeled nontarget strand) in 1× annealing buffer (10 mM Tris-Cl, pH 7.9 at 25 °C, 50 mM KCl, 1 mM EDTA), then diluted to 10× concentration in water. A Cas9 titration at 5× was prepared by diluting an 80 μM Cas9 stock solution with protein-purification size-exclusion buffer. An sgRNA titration at 5× was prepared by diluting a 100 μM sgRNA stock solution with RNA storage buffer. For all reactions, the sgRNA concentration was 1.25 times the Cas9 concentration, and the reported

ribonucleoprotein concentration is that of Cas9. Reactions were assembled with 11 µl of 5× permanganate reaction buffer (100 mM Tris-Cl, pH 7.9 at 25 °C, 120 mM KCl, 25 mM MgCl₂, 5 mM TCEP, 500 µg ml⁻¹ UltraPure BSA, 0.05% Tween-20), 11 µl of water, 11 µl of 5× Cas9, 11 µl of 5× sgRNA and 5.5 µl of 10× DNA. A stock solution of KMnO₄ was prepared fresh in water, and its concentration was corrected to 100 mM (10× reaction concentration) based on eight averaged NanoDrop readings ($\varepsilon_{526} = 2.4 \times 10^3 \, M^{-1} \, cm^{-1}$). Reaction tubes and KMnO₄ (or water, for reactions lacking permanganate) were equilibrated to 30 °C for 15 min. To initiate the reaction, 22.5 µl of Cas9–sgRNA–DNA was added to 2.5 µl of 100 mM KMnO₄ or water. After 2 min, 25 µl of 2× stop solution (2 M βME, 30 mM EDTA) was added to stop the reaction, and 50 µl of water was added to each quenched reaction. Samples were extracted once with 100 µl of 25:24:1 phenol:chloroform:isoamyl alcohol (pH 8) in 5PRIME Phase Lock Heavy tubes (Quantabio). The aqueous phase was isolated and combined with 10 µl of 3 M sodium acetate (pH 5.2), 1 µl of GlycoBlue coprecipitant (Invitrogen) and 300 µl of ethanol, and left at −20 °C for >2 h. DNA was precipitated by centrifugation, and supernatant was decanted. A second wash was performed with 500 µl of 70% ethanol. Pellets were resuspended in 70 µl of 10% piperidine and incubated at 90 °C for 30 min. Solvent was evaporated in a SpeedVac (ThermoFisher Scientific). Approximate yield was determined by measuring radioactivity of the pellet-containing tube in a benchtop radiation counter (Bioscan QC-4000), and pellets were resuspended in an appropriate volume of loading solution (50% water, 50% formamide, 0.025% w/v bromophenol blue) to normalize signal across samples before resolution by denaturing PAGE and autoradiography.

Because piperidine treatment leads to low levels of cleavage at every nucleotide, the exhaustive single-nucleotide ladder could be used to assign band identities, also confirmed by the dark/light pattern (piperidine-catalyzed cleavage at thymines is less efficient than at other nucleotides in the absence of permanganate modification and more efficient in the presence of permanganate modification). We carried out the data analysis as follows: let $v_i$ denote the volume of band $i$ in a lane with $n$ total bands (band 1 is the shortest cleavage fragment, band $n$ is the topmost band corresponding to the starting/uncleaved DNA oligonucleotide). The probability of cleavage at thymine $i$ is defined as: $P_{cleave,i} = \frac{v_i}{\sum_{j=i}^{n} v_j}$. Oxidation (ox) probability of thymine $i$ is defined as: $P_{ox,i} = P_{cleave,i,+pm} - P_{cleave,i,-pm}$, where +pm indicates the experiment that contained 10 mM KMnO₄ and −pm indicates the no-permanganate experiment. An extensive description of this type of analysis can be found in ref. [28].

**Preparation of DNA cyclization substrates.** Each variant DNA cyclization substrate precursor was assembled by PCR from two amplification primers (one of which contained a fluorescein-dT) and two assembly primers. Each reaction was 400 µl in total (split into four 100-µl aliquots) and contained 1× Q5 reaction buffer (New England BioLabs), 200 µM deoxynucleoside triphosphates, 200 nM forward amplification primer, 200 nM reverse amplification primer, 1 nM forward assembly primer, 1 nM reverse assembly primer and 0.02 U µl⁻¹ Q5 polymerase. Thermocycle parameters were as follows: 98 °C, 30 s (98 °C, 10 s; 55 °C, 20 s; 72 °C, 15 s); (98 °C, 10 s; 62 °C, 20 s; 72 °C, 15 s); (98 °C, 10 s; 72 °C, 35 s) × 25; 72 °C, 2 min; 10 °C, ∞. PCR products were phenol-chloroform-extracted, ethanol-precipitated and resuspended in 80 µl of water. To this was added 15 µl of 10× CutSmart buffer, 47.5 µl of water and 7.5 µl of ClaI restriction enzyme (10,000 U ml⁻¹, New England BioLabs), and digestion was allowed to proceed overnight at 37 °C. Samples were then combined with 0.25 volumes 5× native quench solution (25% glycerol, 250 µg ml⁻¹ heparin, 125 mM EDTA, 1.2 mg ml⁻¹ proteinase K, 0.0625% w/v bromophenol blue), incubated at 55 °C for 15 min and resolved on a preparative native PAGE gel (8% acrylamide:bis-acrylamide 37.5:1, 0.5× TBE) at 4 °C. Fluorescent bands, made visible on a blue LED transilluminator, were cut out and DNA was extracted, ethanol-precipitated and resuspended in water.

**Cyclization efficiency measurements.** Each cyclization reaction contained the following components: 1 µl of 10× T4 DNA ligase reaction buffer (New England BioLabs), 2 µl of water, 1 µl of 10× ligation buffer additives (400 µg ml⁻¹ UltraPure BSA, 100 mM KCl, 0.1% NP-40), 2 µl of 80 µM Cas9 (or protein-purification size-exclusion buffer), 2 µl of 100 µM sgRNA (or RNA storage buffer), 1 µl of 25 nM cyclization substrate and 1 µl of T4 DNA ligase (400,000 units ml⁻¹, New England BioLabs) (or ligase storage buffer). All reaction components were incubated together at 20 °C for 15 min before reaction initiation except for the ligase, which was incubated separately. Reactions were initiated by combining the ligase with the remainder of the components, allowed to proceed at 20 °C for 30 min, then quenched with 2.5 µl of 5× native quench solution. Samples were then incubated at 55 °C for 15 min, resolved on an analytical native PAGE gel (8% acrylamide:bis-acrylamide 37.5:1, 0.5× TBE) at 4 °C, and imaged for fluorescein on an Amersham Typhoon (Cytiva). Monomolecular cyclization efficiency for a given lane is defined as (band volume of circular monomers) / (sum of all band volumes). Bimolecular ligation efficiency is defined as (sum of band volumes of all linear or circular $n$-mers, for $n \geq 2$) / (sum of all band volumes). The nonspecific degradation products indicated in Extended Data Fig. 6a were not included in the analysis.

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

## Data availability

All data generated or analyzed during this study are included within this paper and its supporting information files except for the cryo-EM data/models, which can be accessed as follows: Cas9–sgRNA–DNA (*S. pyogenes*) with 0 RNA–DNA base pairs, open-protein/linear-DNA conformation (Protein Data Bank (PDB) 7S3H, EMD-24823); Cas9–sgRNA–DNA (*S. pyogenes*) with 0 RNA–DNA base pairs, closed-protein/bent-DNA conformation (PDB 7S36, EMD-24817); Cas9–sgRNA (*S. pyogenes*) in the open-protein conformation (PDB 7S37, EMD-24818); Cas9–sgRNA–DNA (*S. pyogenes*) forming a 3-base-pair R-loop (PDB 7S38, EMD-24819). The analyses and figures in this study also draw on previously determined structures with PDB codes 4ZT0, 5FQ5, 5F9R, 4CMP, 1CGP and 4UN3 as described in the figure legends and the Supplementary Information. For figures containing fluorescence images, autoradiographs or scatter plots, the original data are available as source data files. Source data are provided with this paper.

## References

49. Cavaluzzi, M. J. & Borer, P. N. Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.* **32**, e13 (2004).

## Author contributions

J.C.C. and J.A.D. conceived the study. J.C.C. produced all reagents and performed all biochemical experiments. J.C.C., K.M.S. and G.J.K. conducted structural studies including EM grid preparation, data collection and analysis, map calculation and model building and refinement. J.A.D. and E.N. provided supervision and guidance on data analysis and interpretation. J.C.C. produced the figures with assistance from K.M.S. J.C.C. wrote the manuscript with assistance from J.A.D. All authors edited and approved the manuscript.

## Competing interests

The Regents of the University of California have patents issued and/or pending for CRISPR technologies on which G.J.K. and J.A.D. are inventors. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Algen Biotechnologies, Synthego, Algen Biotechnologies, Felix Biosciences, The Column Group, and Inari. J.A.D. is Chief Science Advisor to Sixth Street, a Director at Johnson & Johnson, Altos and Tempus, and has research projects sponsored by Biogen, Pfizer, Apple Tree Partners, and Roche. The remaining authors declare no competing interests.

## Additional information

**Extended data** are available for this paper at https://doi.org/10.1038/s41594-022-00756-0.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41594-022-00756-0.

**Correspondence and requests for materials** should be addressed to Jennifer A. Doudna.

**Peer review information** *Nature Structural and Molecular Biology* thanks Rick Russell, 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. Sara Osman was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

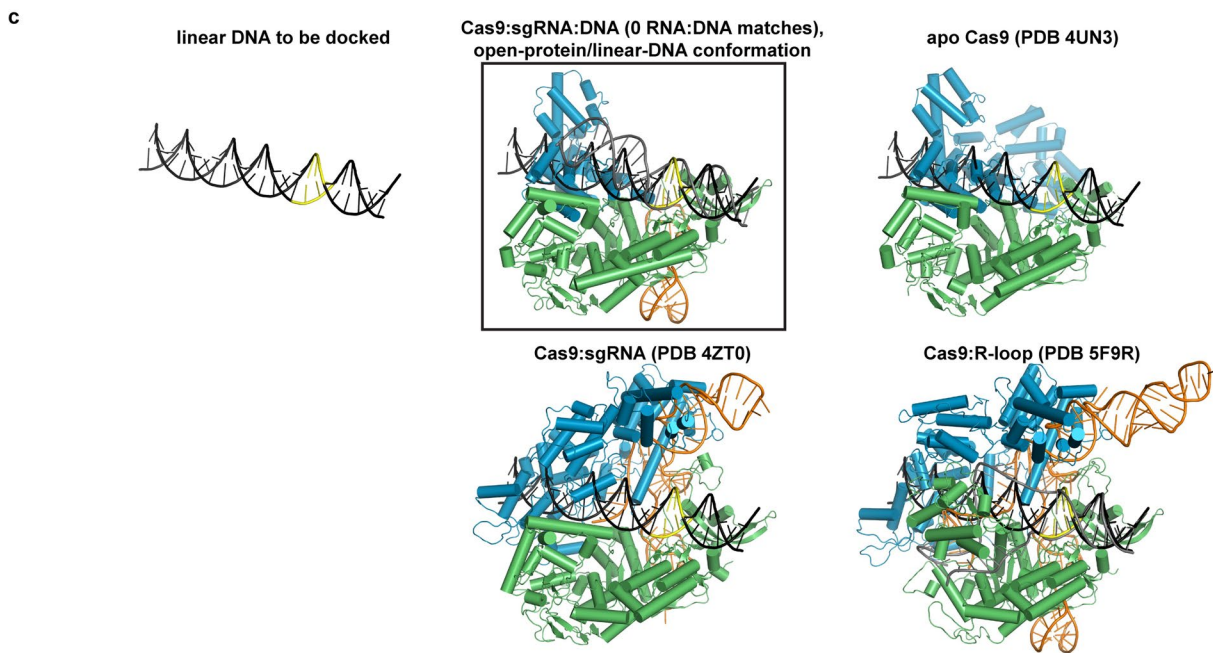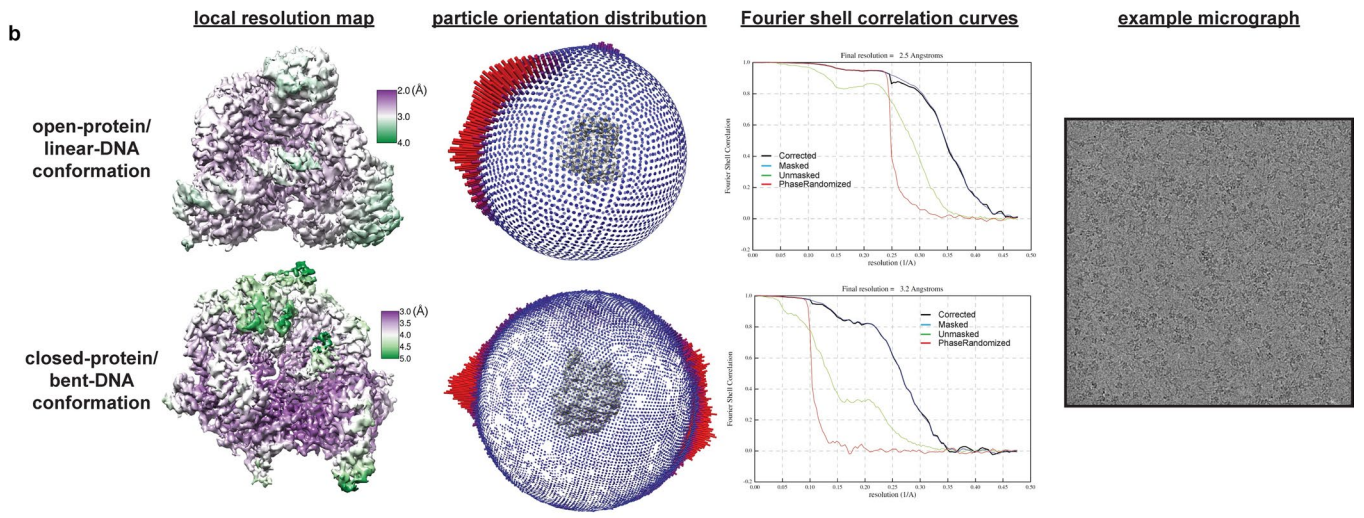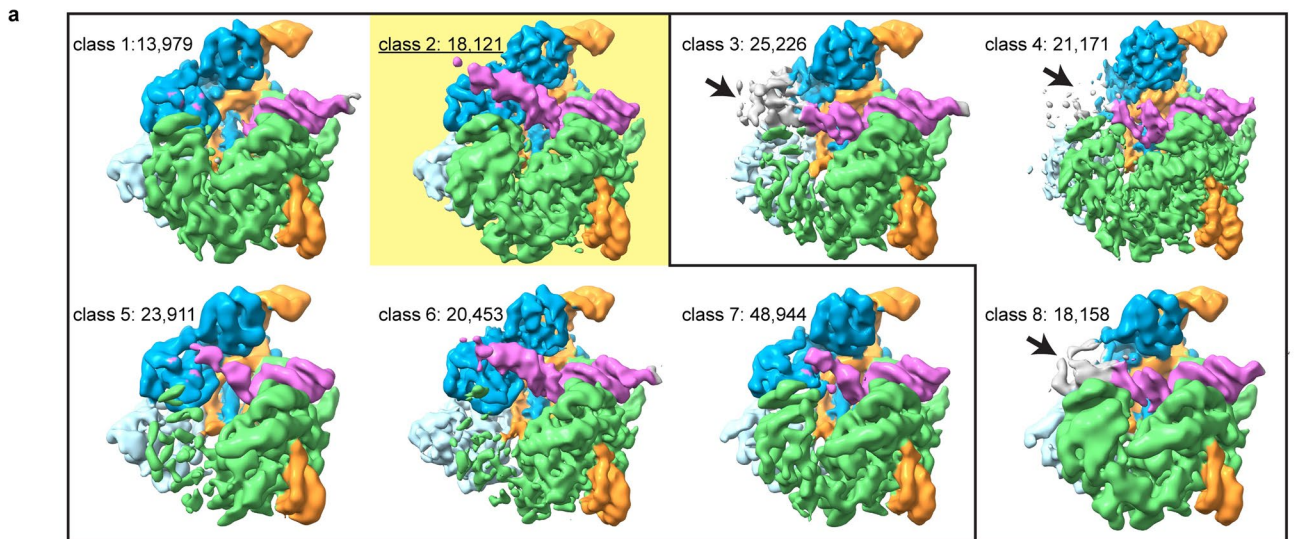**Reprints and permissions information** is available at www.nature.com/reprints.
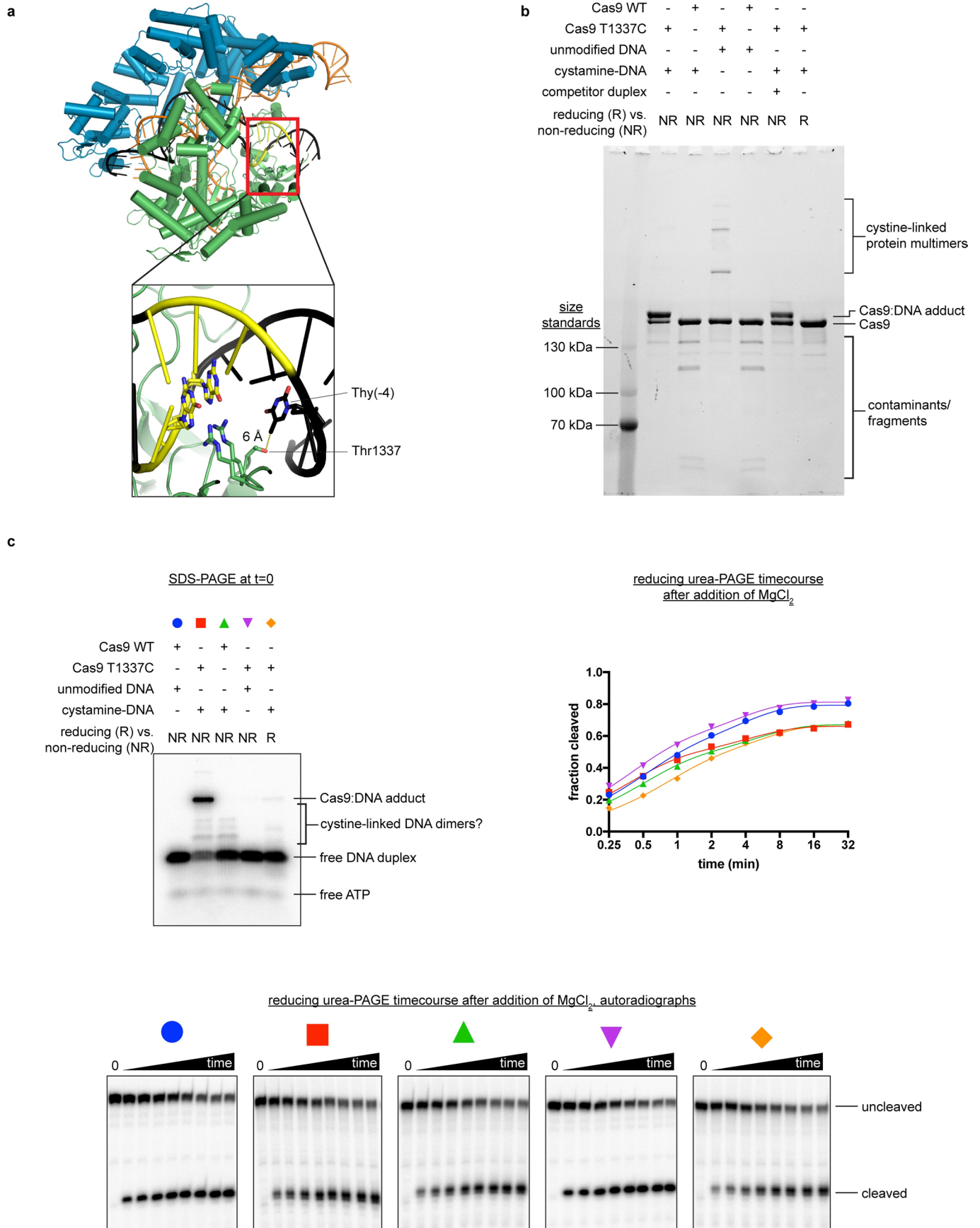
Extended Data Fig. 1 | See next page for caption.

**Extended Data Fig. 1 | Characterization of the Cas9:DNA cross-link. a**, Crystal structure of Cas9:sgRNA:DNA with 20-bp RNA:DNA hybrid formed (PDB 4UN3). In the inset, Arg1333 and Arg1335 recognize the two guanines of the PAM. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, DNA; yellow, PAM. **b**, Non-reducing SDS-PAGE (Stain-Free) analysis of cross-linking reactions and controls. Complexes were prepared identically to structural constructs but in smaller volumes and without size exclusion purification. Competitor duplex, where indicated, was added before the cross-linkable duplex at an equivalent concentration. The depicted experiment was performed once, and a prior optimization experiment yielded similar results. **c**, Top left, non-reducing SDS-PAGE autoradiograph to determine the fraction of DNA cross-linked to Cas9 at t = 0. The target strand is radiolabeled. Bottom, reducing urea-PAGE autoradiograph revealing target-strand cleavage kinetics; quantification depicted in top right. The depicted model is $y = C\left(1 - e^{-k_1 t}\right) + (B_{max} - C)(1 - e^{-k_2 t})$. The depicted experiments were performed once, and a prior optimization experiment yielded similar results.
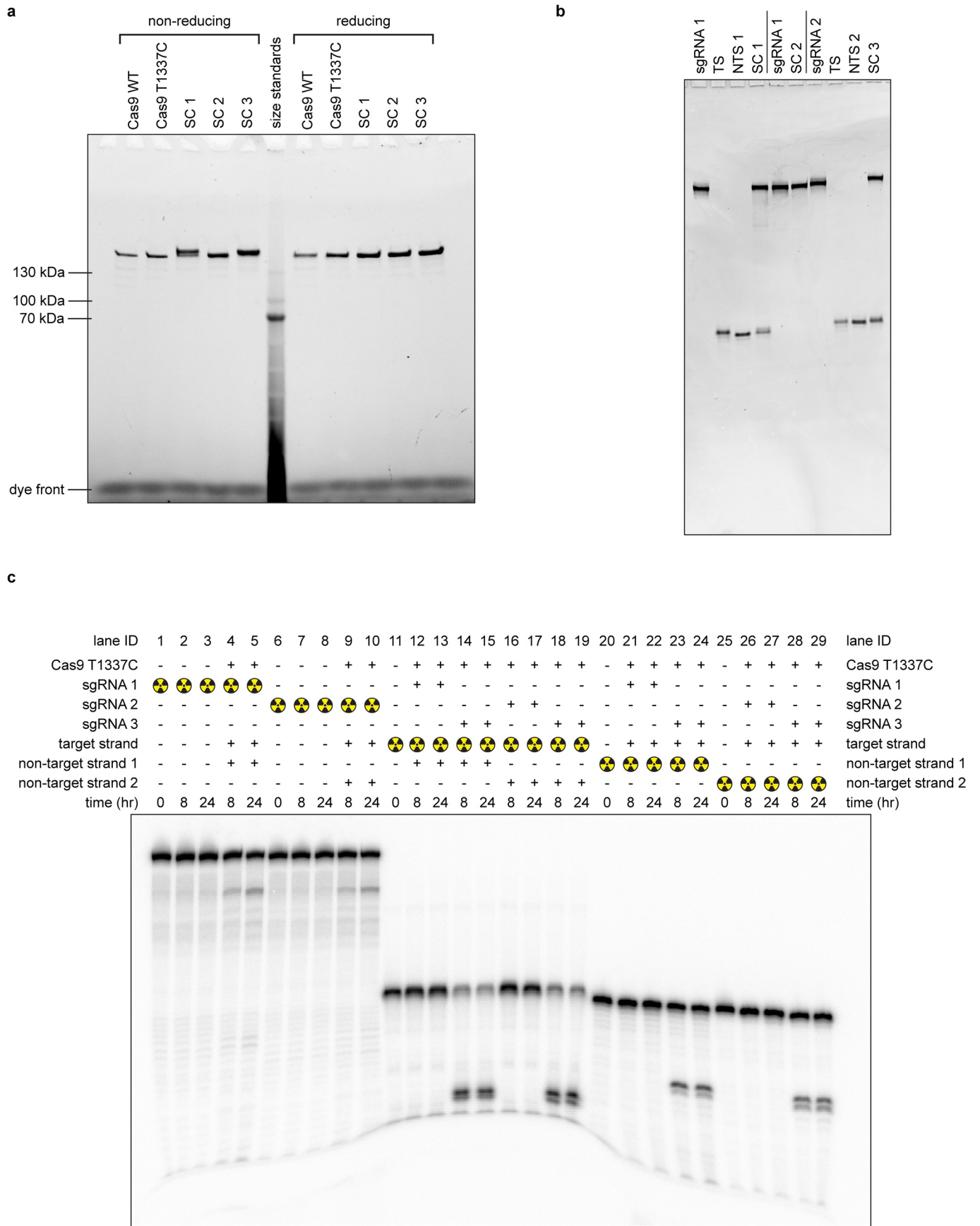
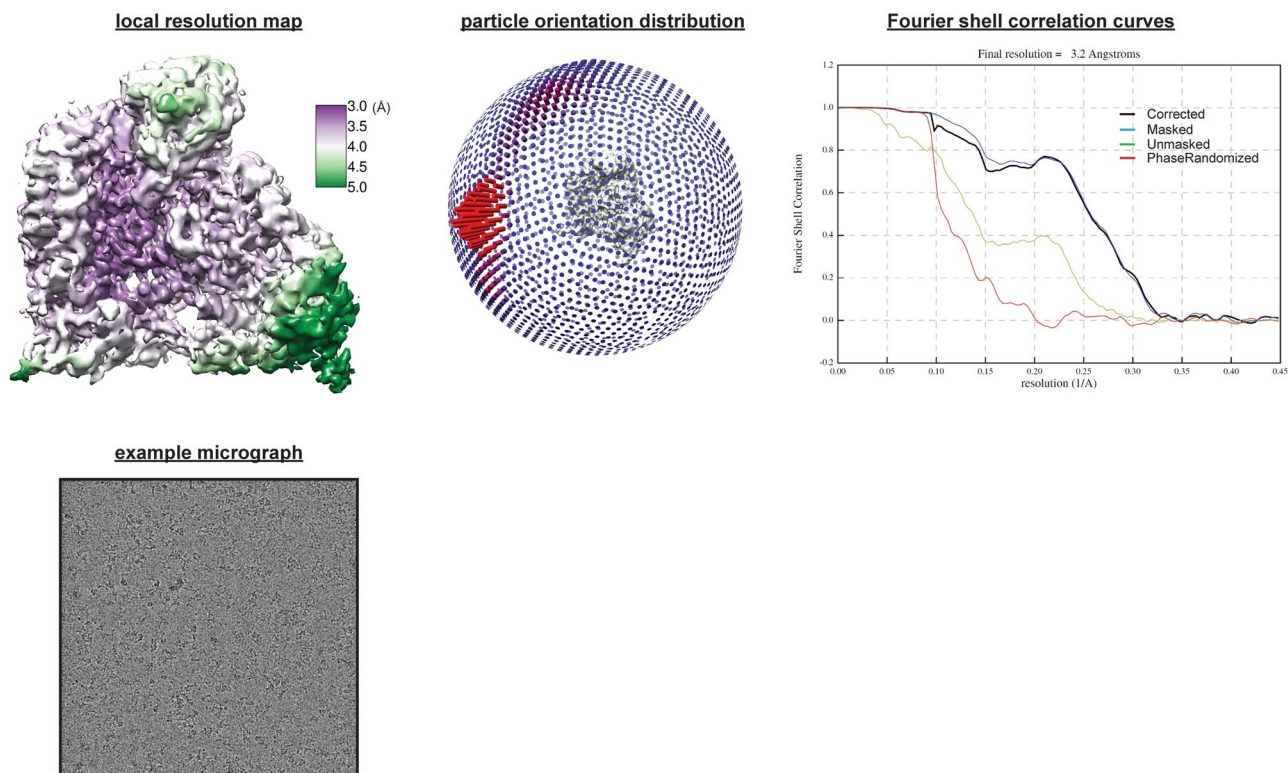Extended Data Fig. 2 | See next page for caption.

**Extended Data Fig. 2 | Cryo-EM sample quality. a**, SDS-PAGE (Stain-Free) analysis of purified proteins and cryo-EM samples. SC, structural construct; SC 1, Cas9:sgRNA:DNA with 0 RNA:DNA matches; SC 2, Cas9:sgRNA; SC 3, Cas9:sgRNA:DNA with 3 RNA:DNA matches. The depicted experiment was performed once (on the samples used for the final cryo-EM grids). **b**, Reducing urea-PAGE (SYBR-Gold-stained) analysis of purified nucleic acid components and cryo-EM samples. TS, target strand; NTS, non-target strand. The depicted experiment was performed once (on the samples used for the final cryo-EM grids). **c**, Reducing urea-PAGE autoradiograph of radioactive mimics of structural constructs. sgRNA 1 and non-target strand 1 are those used to create SC 1. sgRNA 2 and non-target strand 2 are those used to create SC 3. sgRNA 3 bears a spacer with 20 nt of complementarity to the DNA target strand. The depicted experiment was performed once, and a prior optimization experiment yielded similar results.
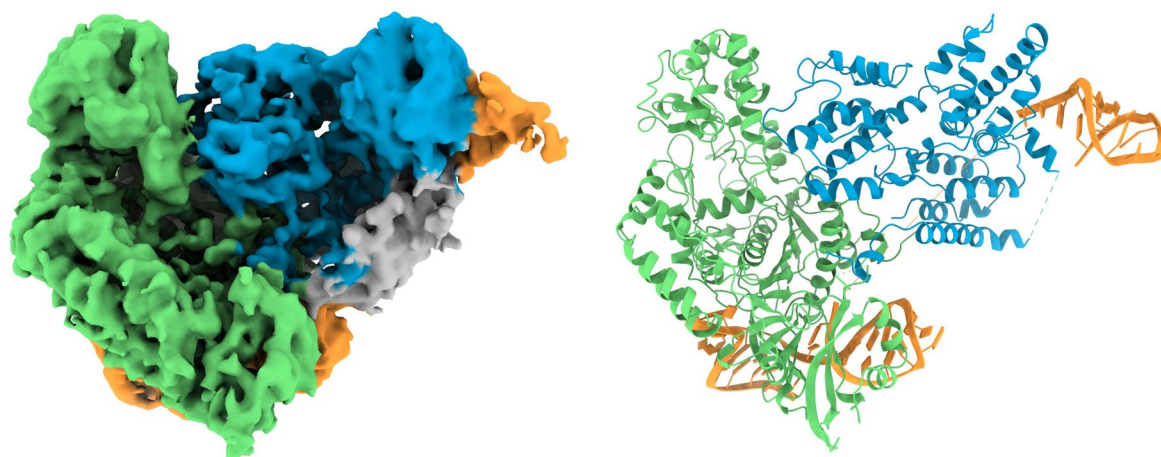
**a**

class 1:13,979 | class 2: 18,121 | class 3: 25,226 | class 4: 21,171

class 5: 23,911 | class 6: 20,453 | class 7: 48,944 | class 8: 18,158

**b**

local resolution map | particle orientation distribution | Fourier shell correlation curves | example micrograph

open-protein/
linear-DNA
conformation

closed-protein/
bent-DNA
conformation

**c**

linear DNA to be docked | Cas9:sgRNA:DNA (0 RNA:DNA matches), open-protein/linear-DNA conformation | apo Cas9 (PDB 4UN3)

Cas9:sgRNA (PDB 4ZT0) | Cas9:R-loop (PDB 5F9R)

**Extended Data Fig. 3 | See next page for caption.**

**Extended Data Fig. 3 | Cryo-EM analysis of Cas9:sgRNA:DNA with 0 RNA:DNA matches. a**, Classes from RELION 3D classification of closed-protein particles (threshold 6σ). The number of particles in each class is indicated next to the class number. In classes 1/2/5/6/7 the DNA is bent next to the PAM (visible for class 1 at lower contour). In classes 3/4/8 the DNA continues along a more linear trajectory for half a turn past the PAM, into the region normally occupied by REC2; in these classes, density in the region of the putative collision (black arrow) is uninterpretable as either protein or DNA, likely due to particle damage, and is thus colored gray. The class used for the final closed-protein/bent-DNA map is class 2, highlighted in yellow. Green, NUC lobe; blue, REC lobe domains 1/2; light blue, REC lobe domain 3; orange, guide RNA; magenta, DNA. **b**, Details of final cryo-EM maps. **c**, Linear DNA docked into open-protein/linear-DNA cryo-EM structure and previous Cas9 crystal structures. All structures were aligned to the C-terminal domain of PDB 5F9R; then, the linear DNA was aligned to the PAM-containing duplex of PDB 5F9R. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, docked DNA; yellow, PAM. The DNA truly belonging to each structure (if present) is depicted in gray.

**a**

local resolution map

particle orientation distribution

Fourier shell correlation curves



**b**



**Extended Data Fig. 4 | Cryo-EM analysis of Cas9:sgRNA. a**, Details of cryo-EM analysis. **b**, Unsharpened cryo-EM map (threshold 5σ) and model of Cas9:sgRNA in open-protein conformation. Green, NUC lobe; blue, REC lobe; orange, guide RNA; gray, unattributed density (REC1 or guide RNA, see Supplementary Information).
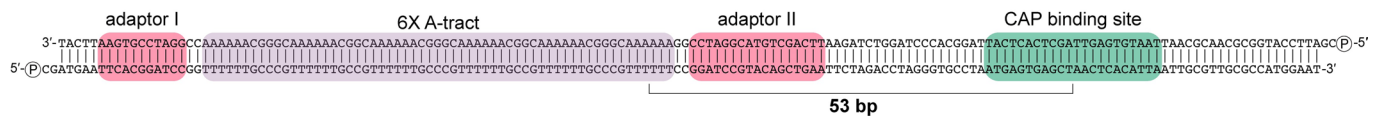
**sgRNA:**

spacer

5'-GGCUGCGUAUUUCUACUCUGUUGUUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUCG-3'
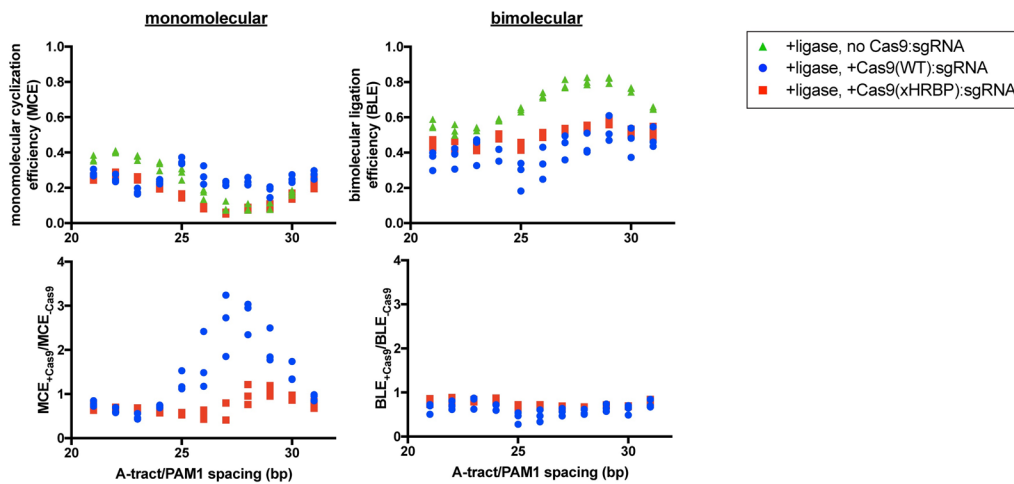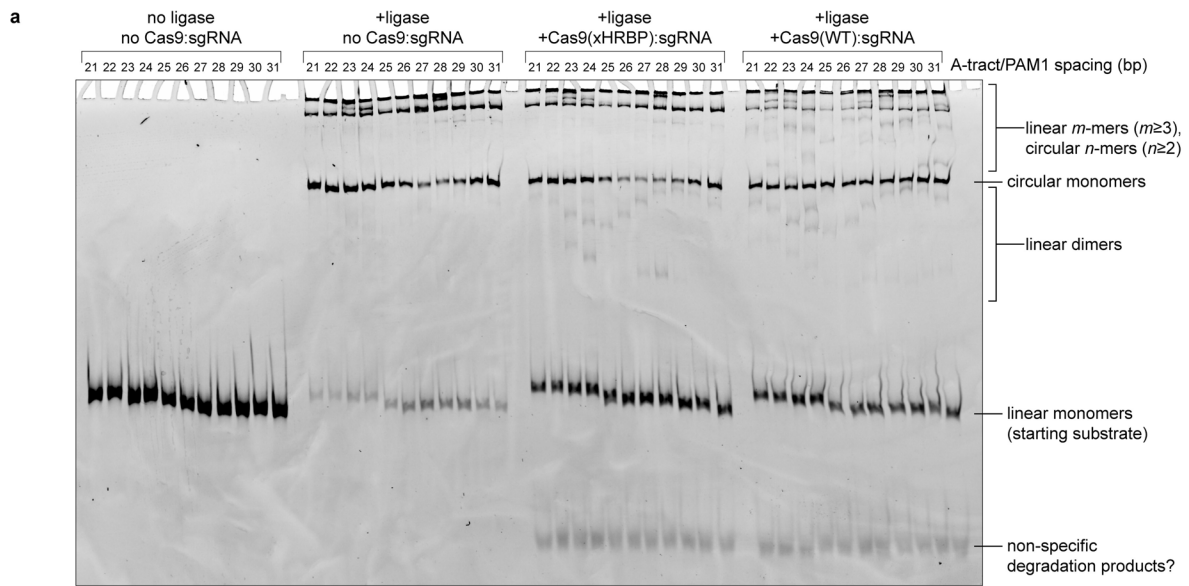
**11 PAM-containing DNA substrates:**
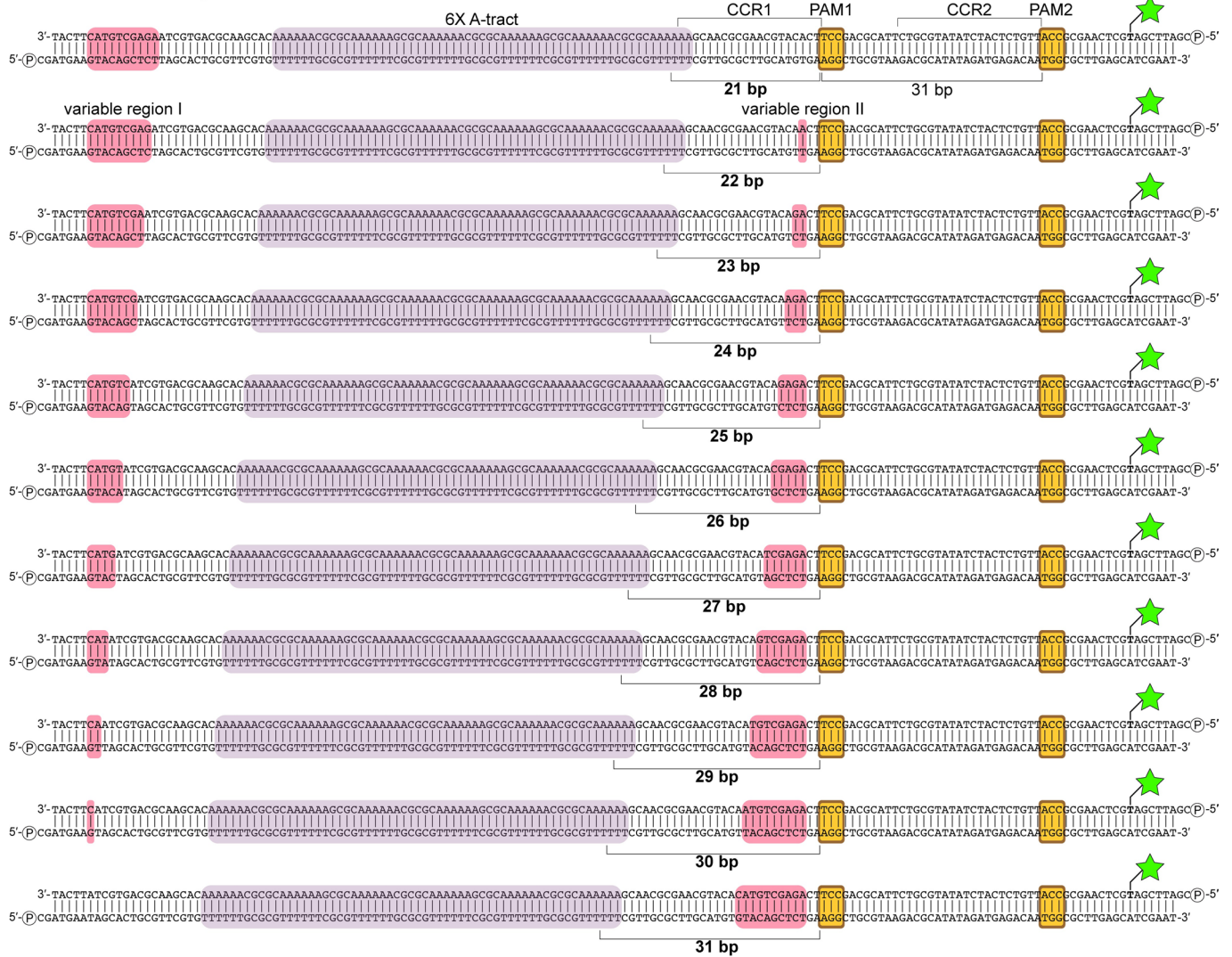


**"11A17" cyclization substrate from Kahn and Crothers, 1992:**



**Extended Data Fig. 5 | Nucleic acid sequences used in DNA cyclization experiments.** Green star/bold T, fluorescein-conjugated dT; circled P, 5' phosphate; CCR, candidate complementarity region.
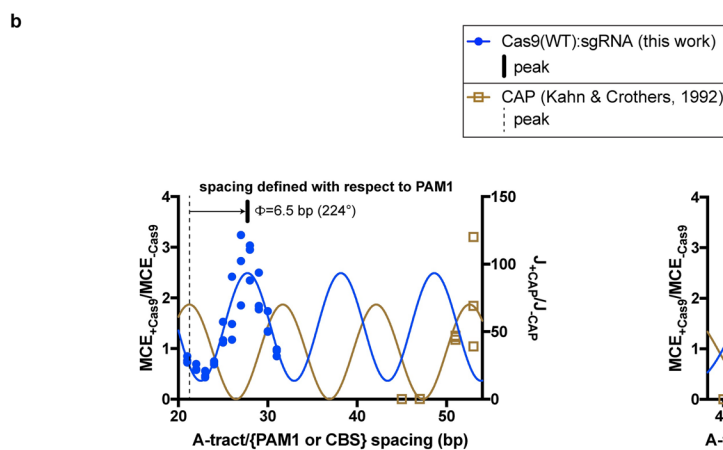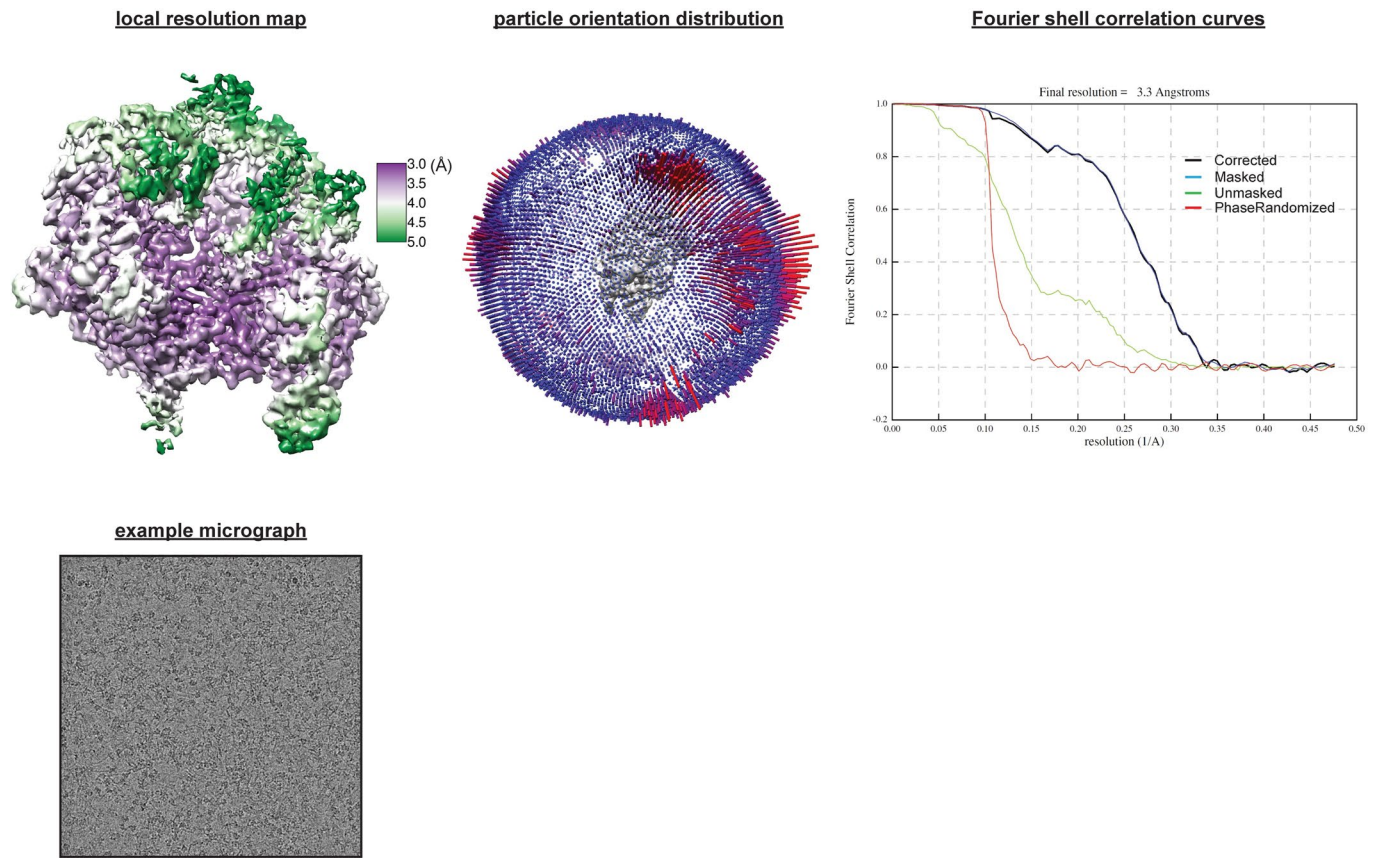
**Extended Data Fig. 6 | Details of DNA cyclization experiments. a**, Fluorescence image and analysis of native PAGE gel resolving ligation products. Gel represents one replicate. Three replicates are plotted on the graphs. The polymeric/cyclized band assignments were made by reference to the relative electrophoretic mobilities observed in Kahn & Crothers, 1992. **b**, Comparison of Cas9:DNA cyclization data to CAP:DNA cyclization data. The depicted model is $y = A \cdot \sin\left(\frac{2\pi}{10.45\,bp}\left(x + \phi_0\right)\right) + b$, with the following constraints: $A > 0$, $b > A$. The average of 224° and 212° is reported in Fig. 4c. J, J-factor (defined in Kahn & Crothers, 1992); Φ, phase difference; CBS, CAP-binding site.
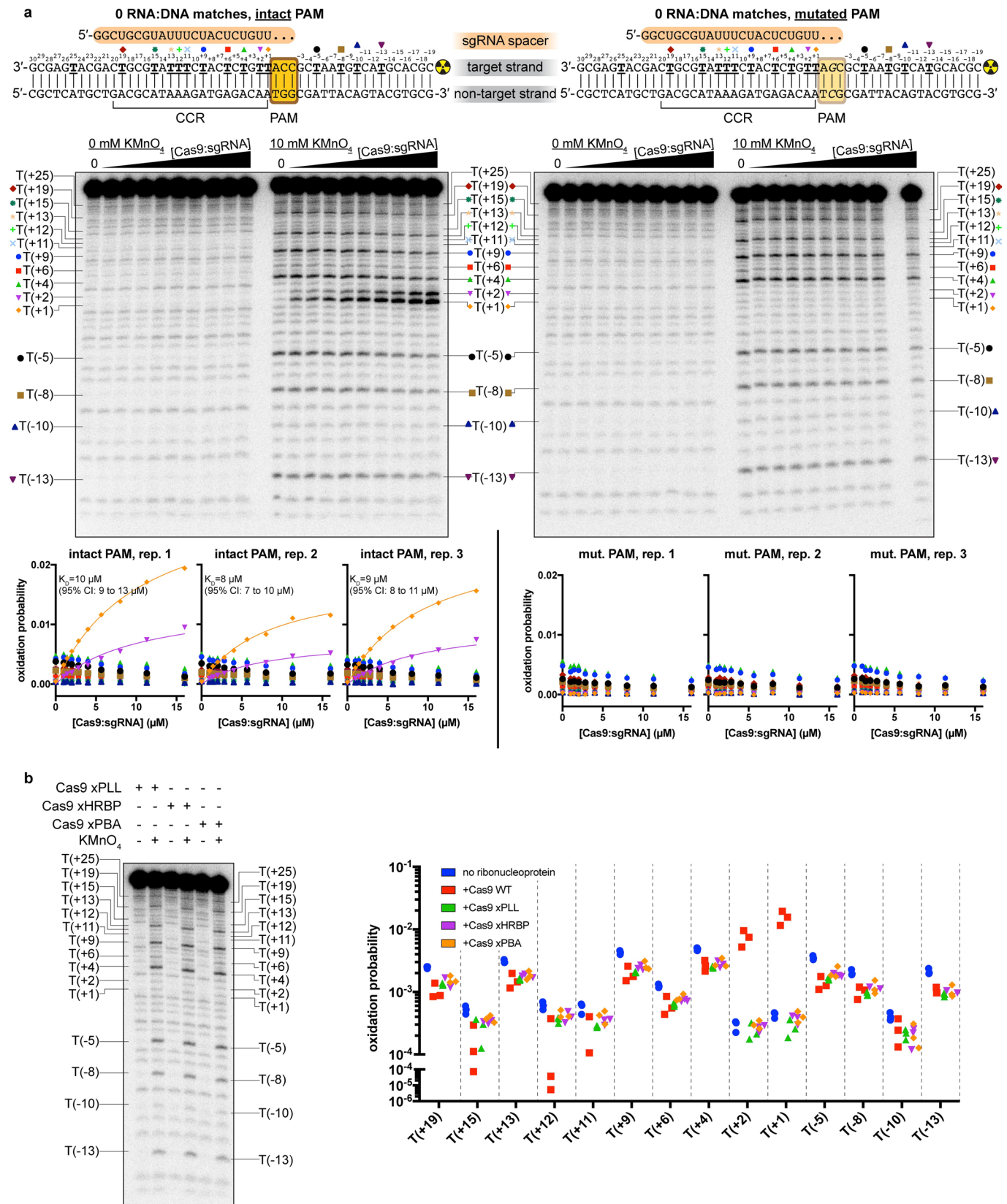
**Extended Data Fig. 7 | Details of permanganate reactivity measurements. a**, Autoradiographs and analysis of all thymines except T(+25), which was insufficiently resolved from neighboring bands. The depicted autoradiographs are replicate 1. Due to systematic variation across replicates, individual replicates are presented on separate graphs and fitted separately. The depicted model is $P_{ox} = \frac{B_{max}[Cas9:sgRNA]}{K_D + [Cas9:sgRNA]}$, with $K_D$ shared across T(+1) and T(+2). CCR, candidate complementarity region; CI, confidence interval. **b**, Autoradiograph (same gel as 'intact PAM' autoradiograph in **a**) and analysis of experiments containing variants of Cas9:sgRNA (16 μM), with the 'intact PAM' DNA substrate. Graph depicts three replicates.
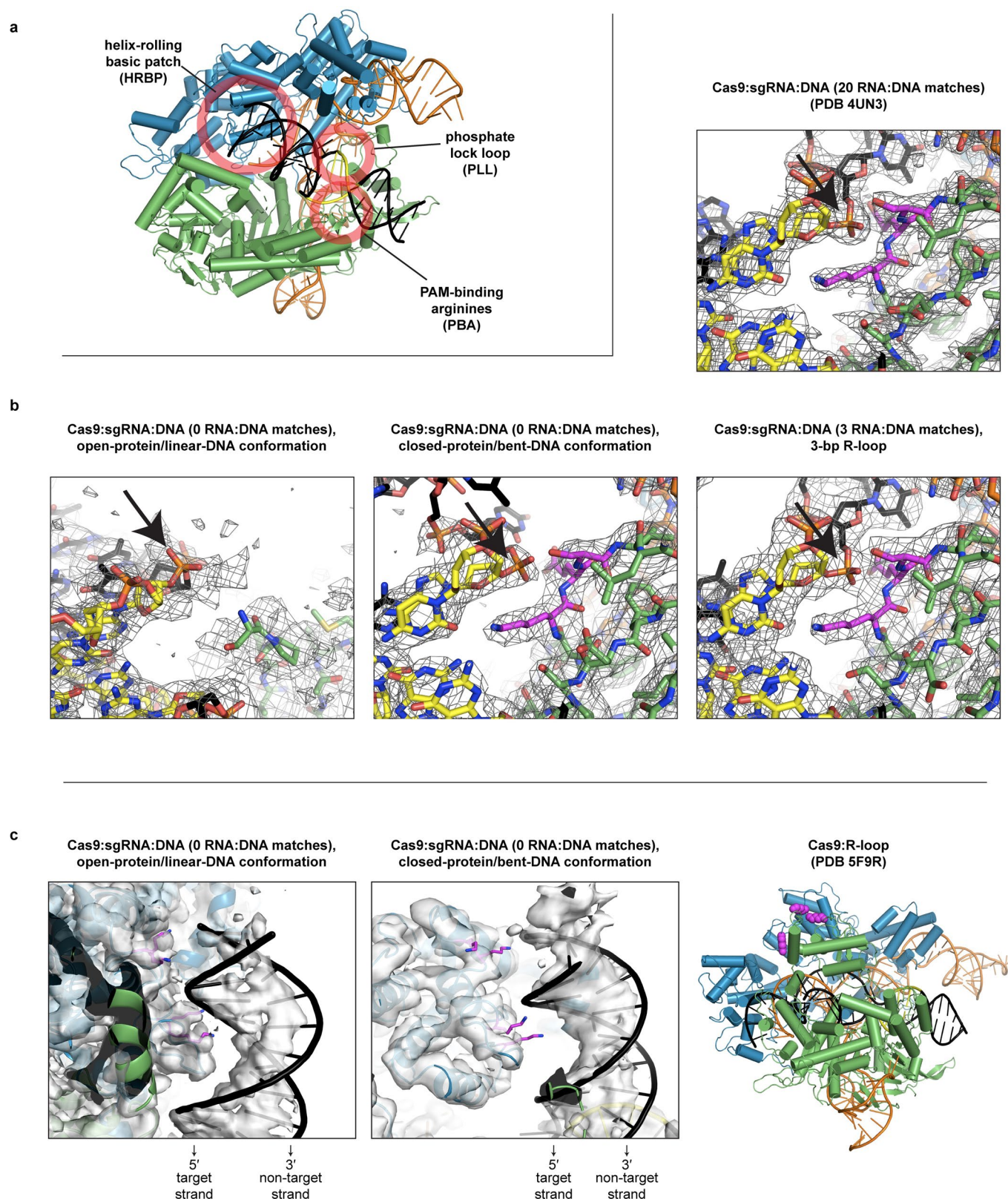
**Extended Data Fig. 8 | Structural features potentially relevant to Cas9-induced DNA bending. a**, Location of each feature in the bent-DNA structure. Green, NUC lobe; blue, REC lobe; orange, guide RNA; black, DNA; yellow, PAM. **b**, Comparison of phosphate lock loop (magenta) in various structures, within sharpened cryo-EM maps (row of three, threshold 8σ) or $2F_o$-$F_c$ map (upper right, threshold 1.5σ). Black arrow indicates the eponymous phosphate between nucleotides 0 and +1 of the target strand. **c**, Comparison of helix-rolling basic patch (magenta) in various structures, within unsharpened cryo-EM maps (first panel, threshold 5σ; second panel, threshold 6σ).

local resolution map

particle orientation distribution

Fourier shell correlation curves



example micrograph



**Extended Data Fig. 9 | Cryo-EM analysis of Cas9:sgRNA:DNA with 3 RNA:DNA matches.** Details of cryo-EM analysis.

# nature portfolio

Corresponding author(s):   Jennifer Doudna

Last updated by author(s):   Jan 26, 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 | EM images were collected using SerialEM v.3.8.7. Typhoon autoradiographs and fluorescence images were collected using the Amersham Typhoon Control Software 2.0.0.6. Stain-Free Gel Images were collected using Bio-Rad Image Lab 5.2.1. |
|---|---|
| Data analysis | EM analysis was performed using cryoSPARC (v.3.2), RELION (v.3.1.0), LocSpiral (v.1.0.0.0), LocBSharpen (v.1.0.0.0), LocScale (within CCP-EM v.1.4.1), Chimera (v.1.4), ChimeraX (v.1.1.1), Coot (v.0.9.2), PyMOL (v.2.4.1), PHENIX (v.1.19.2-4158-000), and X3DNA (v.2.4). Autoradiographs and fluorescence images were analyzed in Bio-Rad Image Lab 6.1. Model fitting was performed in GraphPad Prism 7.0. |

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

All data generated or analyzed during this study are included within this manuscript and its supporting information files except for the cryo-EM data/models, which can be accessed as follows: Cas9:sgRNA:DNA (S. pyogenes) with 0 RNA:DNA base pairs, open-protein/linear-DNA conformation (PDB 7S3H, EMD-24823); Cas9:sgRNA:DNA (S. pyogenes) with 0 RNA:DNA base pairs, closed-protein/bent-DNA conformation (PDB 7S36, EMD-24817); Cas9:sgRNA (S. pyogenes) in the open-protein conformation (PDB 7S37, EMD-24818); Cas9:sgRNA:DNA (S. pyogenes) forming a 3-base-pair R-loop (PDB 7S38, EMD-24819). The analyses and figures in

this study also draw on previously determined structures with PDB codes 4ZT0, 5FQ5, 5F9R, 4CMP, 1CGP, and 4UN3 as described in the figure legends and the Supplementary Information. For figures containing fluorescence images, autoradiographs, or scatter plots, the original data are available as Source Data files.

# 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 | Experiments were generally performed three times, and the precise approach is explained in each figure legend. Three was chosen arbitrarily, as the work contains no statistical claims that necessitate a rigorous sample size calculation. |
| Data exclusions | No data were excluded. |
| Replication | The number of experimental replicates is described in each figure legend. All attempts at replication were successful. |
| Randomization | There were no "experimental groups" associated with this study. |
| Blinding | There was no "group allocation" associated this study, so investigators were not blinded. |

# 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 |