



Published in final edited form as:

Science. 2016 February 19; 351(6275): 867–871. doi:10.1126/science.aad8282.

Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage

Fuguo Jiang^{1,*}, David W. Taylor^{1,2,*}, Janice S. Chen¹, Jack E. Kornfeld³, Kaihong Zhou³, Aubri J. Thompson⁴, Eva Nogales^{1,2,3,5,†}, and Jennifer A. Doudna^{1,2,3,4,5,6,†}

¹Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA

²California Institute for Quantitative Biosciences, University of California, Berkeley, CA 94720, USA

³Howard Hughes Medical Institute, University of California, Berkeley, CA 94720, USA

⁴Department of Chemistry, University of California, Berkeley, CA 94720, USA

⁵Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

⁶Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Abstract

Bacterial adaptive immunity and genome engineering involving the CRISPR (clustered regularly interspaced short palindromic repeats)–associated (Cas) protein Cas9 begin with RNA-guided DNA unwinding to form an RNA-DNA hybrid and a displaced DNA strand inside the protein. The role of this R-loop structure in positioning each DNA strand for cleavage by the two Cas9 nuclease domains is unknown. We determine molecular structures of the catalytically active *Streptococcus pyogenes* Cas9 R-loop that show the displaced DNA strand located near the RuvC nuclease domain active site. These protein-DNA interactions, in turn, position the HNH nuclease domain adjacent to the target DNA strand cleavage site in a conformation essential for concerted DNA cutting. Cas9 bends the DNA helix by 30°, providing the structural distortion needed for R-loop formation.

Bacteria and archaea defend themselves against infection using adaptive immune systems comprising CRISPR (clustered regularly interspaced short palindromic repeats) loci and their associated *cas* genes (1–4). CRISPR-Cas systems use Cas proteins in complex with small CRISPR RNAs (crRNAs) to identify and cleave complementary target sequences in foreign DNA (5, 6). A defining feature of type I and type II CRISPR-Cas systems is R-loop

[†]Corresponding author. doudna@berkeley.edu (J.A.D.); enogales@lbl.gov (E.N.).

^{*}These authors contributed equally to this work.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/351/6275/867/suppl/DC1

Materials and Methods

Figs. S1 to S11

Tables S1 and S2

Movies S1 to S3

References (22–44)

formation in which the guide RNA segment of crRNAs invades double-helical DNA to form an RNA-DNA hybrid helix with the target DNA strand while displacing the opposing nontarget strand (7–10). In type II CRISPR-Cas systems, the Cas9 endonuclease together with crRNA and a trans-activating crRNA (tracrRNA) or an engineered single-guide RNA (sgRNA), which is widely used in Cas9-based genome engineering, is sufficient to form this R-loop (8, 11). The R-loop interaction occurs at target DNA sequences bearing guide RNA segment complementarity and a short protospacer adjacent motif (PAM) leading to double-stranded DNA (dsDNA) cleavage 3 base pairs (bp) upstream of the PAM via the HNH and RuvC nuclease domains (11, 12). Comparison of crystal structures of Cas9 alone (13), bound to guide RNA (14), or to a single target strand of DNA (15–17) revealed substantial conformational rearrangement of Cas9 to form a DNA recognition-competent structure. However, in all available structures, the HNH domain active site is more than 30 Å away from the target-strand DNA cleavage site, and the active DNA-bound state of the RuvC domain has not been observed due to the lack of an intact nontarget strand in these complexes.

To determine the structural basis for R-loop formation and concerted DNA cleavage by the two Cas9 endonuclease domains, we solved the 3.4-Å resolution crystal structure of wild-type *Streptococcus pyogenes* Cas9 bound to sgRNA and a 30-bp target dsDNA containing a canonical 5'-TGG-3' PAM (11, 18) (Fig. 1, A and B, and table S1). Metal ion chelation prevented DNA cleavage during complex assembly and crystallization. The Cas9-sgRNA-dsDNA ternary complex, representing the cleavage-competent state of the enzyme, has a bi-lobed architecture in which the bound target DNA resides within the central channel between the alpha-helical recognition (REC) and the nuclease (NUC) lobes (Fig. 1, C and D). The dsDNA substrate is trapped in an unwound but precleaved state, with the target DNA strand engaged in a pseudo-A-form RNA-DNA hybrid spanning the length of the central protein channel.

The displaced nontarget single-stranded DNA (ssDNA) runs parallel to the RNA-DNA heteroduplex but threads into a tight side tunnel located within the NUC lobe (Fig. 1D and fig. S1). Unambiguous electron density is observed for 9 nucleotides (nt) of nontarget DNA upstream of the PAM (Fig. 2A). Analysis of crystals containing dsDNA with 5-iododeoxyuridine (5-IdU) in place of three thymidines in the nontarget strand shows two strong peaks of iodine anomalous difference density at the expected positions (fig. S2). Lack of density for the third 5-IdU-substituted nucleotide, located at the 5' end of the nontarget strand, indicates high mobility of the PAM-distal end of this strand. These crystallographic results are consistent with footprinting data showing that only 9 nt of the PAM-proximal displaced DNA strand are protected from P1 nuclease cleavage (13).

The ordered nontarget DNA exhibits an extended, distorted helical conformation in which the first nucleotide upstream of the PAM (referred to as position -1) stacks onto the PAM-proximal DNA duplex (Fig. 2A). Whereas the target DNA strand kinks at the +1 phosphodiester linkage, as observed previously (16, 17), the nontarget DNA strand undergoes a sharp kink at the -1 phosphate position. In contrast to the +1 phosphate on the target strand, there are no direct protein contacts to the -1 phosphate on the nontarget strand. Instead, Cas9 makes extensive interactions with the flipped nucleotides at the -2 and -3

positions to stabilize the kinked DNA configuration of the nontarget strand (Fig. 2B). The nontarget strand kinks again, with the base orientation at -4 twisted by $\sim 90^\circ$ with respect to -3 and stabilized by stacking with two conserved Cas9 residues, Phe⁹¹⁶ and Leu⁹²¹. After position -5 , the nontarget DNA strand extrudes laterally through the narrow side tunnel formed by the NUC lobe without many direct protein-DNA interactions.

The observed orientation of the nontarget DNA strand reveals its position in the RuvC catalytic center. The four catalytically essential residues (Asp¹⁰, Glu⁷⁶², His⁹⁸³, and Asp⁹⁸⁶) of this nuclease domain are positioned near the scissile phosphate between positions -3 and -4 (Fig. 2C). RuvC, which employs a carboxylate-chelated two-metal-ion cleavage mechanism (13, 15), adopts the same structure and position as observed in the Mn²⁺-bound apo-Cas9 structure (13). Notably, the distances between the superimposed metal ions and the nonbridging oxygen of the scissile phosphate (~ 5.5 Å) are slightly longer than the typical Mg-O coordination distance (2.1 Å). We speculate that binding of divalent cations in the active site may facilitate movement of the scissile phosphate in the nontarget strand toward the bound metal ions for catalysis.

The presence of the nontarget strand induces a more compact (closed) conformation of Cas9, relative to previous ssDNA- or PAM-containing partial duplex-bound structures, in which both the REC and NUC lobes move toward the central nucleic acid-binding channel (fig. S3). This repositioning drives helical domain 2 of the REC lobe, which does not contact the RNA-DNA heteroduplex in previously reported structures (15, 16), to interact directly with the phosphate backbone in the target DNA strand through residues Ser²⁶⁷ and Asp²⁶⁹ (Fig. 2B and fig. S4). Thermal shift assays show that Cas9 exhibits greater stability upon dsDNA substrate binding relative to association with a single target DNA strand or a PAM-containing partial duplex (fig. S5 and table S2). Together, these observations show how binding to both strands of a dsDNA target coordinates Cas9 conformational changes required for efficient catalysis.

The most prominent Cas9 conformational rearrangement upon dsDNA binding occurs in the position of the HNH nuclease domain, which rotates by $\sim 180^\circ$ and translates by ~ 20 Å toward the RNA-DNA heteroduplex as a rigid body from its location in the PAM-containing partial duplex-bound complex (Fig. 3A). In contrast to previous structures, in which the HNH active site was neither pointed toward nor located near the target DNA strand cleavage site (fig. S6), this rotated HNH position locates the active-site residue His⁸⁴⁰ adjacent to the scissile phosphate (Fig. 3B). The observed structural states of the HNH domain agree with intramolecular Förster resonance energy transfer (FRET) experiments showing that the HNH domain exists in conformational equilibrium between inactive and active states, favoring the active state observed here only upon on-target dsDNA binding (19). In our present dsDNA-bound structure, the distance between the catalytic residue His⁸⁴⁰ and the scissile phosphate is slightly farther apart (~ 10 Å) than required for catalysis, probably owing to omission of divalent metal ions during crystallization. Given the high mobility exhibited by the HNH domain, it is reasonable to speculate that introduction of metal ions would drive the HNH active site to the target DNA strand for cleavage (13, 15).

Double-stranded DNA binding induces additional changes in the Cas9 structure that explain both HNH domain repositioning and R-loop stabilization. The two hinge regions connecting the HNH domain with the RuvC domain—L1 (residues 765 to 780) and L2 (residues 906 to 918)—are switched in the dsDNA-bound structure relative to their positions in the PAM-containing partial duplex-bound state (Fig. 3A). This rearrangement involves local changes in protein secondary structure. L1, which is completely disordered in prior structures, refolds into a well-ordered loop and a short α helix and lies between the target DNA strand and the displaced nontarget DNA strand (Fig. 3C). The refolding of L1 appears to be critical for stabilizing the R-loop structure, owing to multiple contacts made with the RNA-DNA hybrid minor groove and the distal region of the nontarget DNA strand (Fig. 2B). In contrast, L2 unfolds from its previously observed α helical structure into an extended loop in which many residues, including Phe⁹¹⁶ and Leu⁹²¹, contact the nontarget DNA strand, thereby directing the scissile phosphate toward the RuvC active center (Fig. 2, B and C, and Fig. 3B). This dsDNA-induced unfolding of L2 is reminiscent of conformational changes observed in other protein–DNA complexes, such as the restriction enzyme BamHI–DNA complex (20) and the T7 RNA polymerase elongation complex (21). The structural rearrangements displayed by L1 and L2 upon dsDNA binding, in concert with the reorientation of the HNH domain, provide direct structural evidence for allostery between the HNH and RuvC nuclease domains through these hinge regions.

To visualize how Cas9 holds both ends of unwound dsDNA within a longer helix, we used cryo-electron microscopy (cryo-EM) to determine the structure of a wild-type *S. pyogenes* Cas9–sgRNA complex bound to a 40-bp dsDNA (table S2). We obtained structures of Cas9–sgRNA (Fig. 4, A and B) and Cas9–sgRNA–dsDNA (Fig. 4C) at 4.5 Å and 6.0 Å resolution, respectively (figs. S7 to S11). The 4.5-Å structure of Cas9–sgRNA perfectly accommodates the crystal structure of the preorganized Cas9–guide RNA complex (14). In contrast to the crystal structure, in which the 5′ end of the guide RNA (nucleotides 1 to 10) were disordered, density was visible for the entire 20-nt 5′ end of the guide RNA segment. The 5′ end of the guide lies inside the cavity formed between the HNH and RuvC nuclease domains (Fig. 4B). In the dsDNA-bound EM structure, the HNH domain and helical domain 2 are observed at lower resolution than the overall structure (~8 to 10 Å), indicating conformational plasticity of these two domains (Fig. 4C). This is consistent with a recent report showing that the HNH domain populates both active and inactive states (19). Density for dsDNA is observed on either side of the Cas9 protein complex, revealing a bend angle of ~30° in the DNA double helix as it traverses the protein (Fig. 4, C to E). The PAM-distal end duplex protrudes from the protein between the RNA–DNA heteroduplex and the RuvC nuclease domain. This end of the DNA is held rigidly on opposite sides through contacts with the RuvC and REC-lobe helical domain 3, stabilizing the distortion of the target DNA from a linear helical axis (Fig. 4F).

The structures of the Cas9–sgRNA–dsDNA complex presented here show how Cas9 holds unwound dsDNA to enable RNA–DNA hybrid formation without requiring an adenosine triphosphate-dependent helicase activity. DNA binding at a sequence complementary to the 20-nt guide RNA segment in the Cas9–RNA complex induces protein structural rearrangements that accommodate both the RNA–DNA helix and the displaced nontarget DNA strand. Those protein–nucleic acid interactions in turn direct the nontarget DNA strand

into the RuvC domain active site, favoring local conformational changes that position the HNH domain active site near the scissile phosphate of the target DNA strand. Cas9 interacts with both ends of the open DNA helix, conferring a 30° helical bend angle that provides the structural distortion required for R-loop stabilization. Collectively, these findings explain how Cas9, an individual nonpolymerase enzyme able to catalyze R-loop formation, achieves this structural change leading to accurate, precise, and programmable DNA cleavage.

Acknowledgments

The atomic coordinate of Cas9-sgRNA-dsDNA structure has been deposited in the Protein Data Bank with accession code 5F9R. The cryo-EM structures of guide RNA-bound Cas9 and target DNA-bound Cas9 have been deposited into the EMDataBank with accession codes EMD-3276 and EMD-3277, respectively. We thank G. Meigs, J. Holton (beamline 8.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory) for assistance with data collection; Y. Cheng, R. Louder, S. Howes, E. Kellogg, R. Zhang, T. Houweling, Z. Yu, and M. J. de la Cruz for expert electron microscopy assistance; and members of the Doudna and Nogales laboratories for critical reading of the manuscript. F.J. is a Merck Fellow of the Damon Runyon Cancer Research Foundation (DRG-2201-14). D.W.T. is a Damon Runyon Fellow of the Damon Runyon Cancer Research Foundation (DRG-2218-15). J.A.D. and E.N. are Howard Hughes Medical Institute Investigators.

REFERENCES AND NOTES

1. Wiedenheft B, Sternberg SH, Doudna JA. *Nature*. 2012; 482:331–338. [PubMed: 22337052]
2. Sorek R, Lawrence CM, Wiedenheft B. *Annu Rev Biochem*. 2013; 82:237–266. [PubMed: 23495939]
3. Barrangou R, Marraffini LA. *Mol Cell*. 2014; 54:234–244. [PubMed: 24766887]
4. Marraffini LA. *Nature*. 2015; 526:55–61. [PubMed: 26432244]
5. van der Oost J, Westra ER, Jackson RN, Wiedenheft B. *Nat Rev Microbiol*. 2014; 12:479–492. [PubMed: 24909109]
6. Jiang F, Doudna JA. *Curr Opin Struct Biol*. 2015; 30:100–111. [PubMed: 25723899]
7. Jore MM, et al. *Nat Struct Mol Biol*. 2011; 18:529–536. [PubMed: 21460843]
8. Szczelkun MD, et al. *Proc Natl Acad Sci USA*. 2014; 111:9798–9803. [PubMed: 24912165]
9. Rutkauskas M, et al. *Cell Reports*. 2015; 10:1534–1543.
10. Blosser TR, et al. *Mol Cell*. 2015; 58:60–70. [PubMed: 25752578]
11. Jinek M, et al. *Science*. 2012; 337:816–821. [PubMed: 22745249]
12. Gasiunas G, Barrangou R, Horvath P, Siksnys V. *Proc Natl Acad Sci USA*. 2012; 109:E2579–E2586. [PubMed: 22949671]
13. Jinek M, et al. *Science*. 2014; 343:1247997. [PubMed: 24505130]
14. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. *Science*. 2015; 348:1477–1481. [PubMed: 26113724]
15. Nishimasu H, et al. *Cell*. 2014; 156:935–949. [PubMed: 24529477]
16. Anders C, Niewoehner O, Duerst A, Jinek M. *Nature*. 2014; 513:569–573. [PubMed: 25079318]
17. Nishimasu H, et al. *Cell*. 2015; 162:1113–1126. [PubMed: 26317473]
18. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C. *Microbiology*. 2009; 155:733–740. [PubMed: 19246744]
19. Sternberg SH, LaFrance B, Kaplan M, Doudna JA. *Nature*. 2015; 527:110–113. [PubMed: 26524520]
20. Newman M, Strzelecka T, Dorner LF, Schildkraut I, Aggarwal AK. *Science*. 1995; 269:656–663. [PubMed: 7624794]
21. Yin YW, Steitz TA. *Science*. 2002; 298:1387–1395. [PubMed: 12242451]

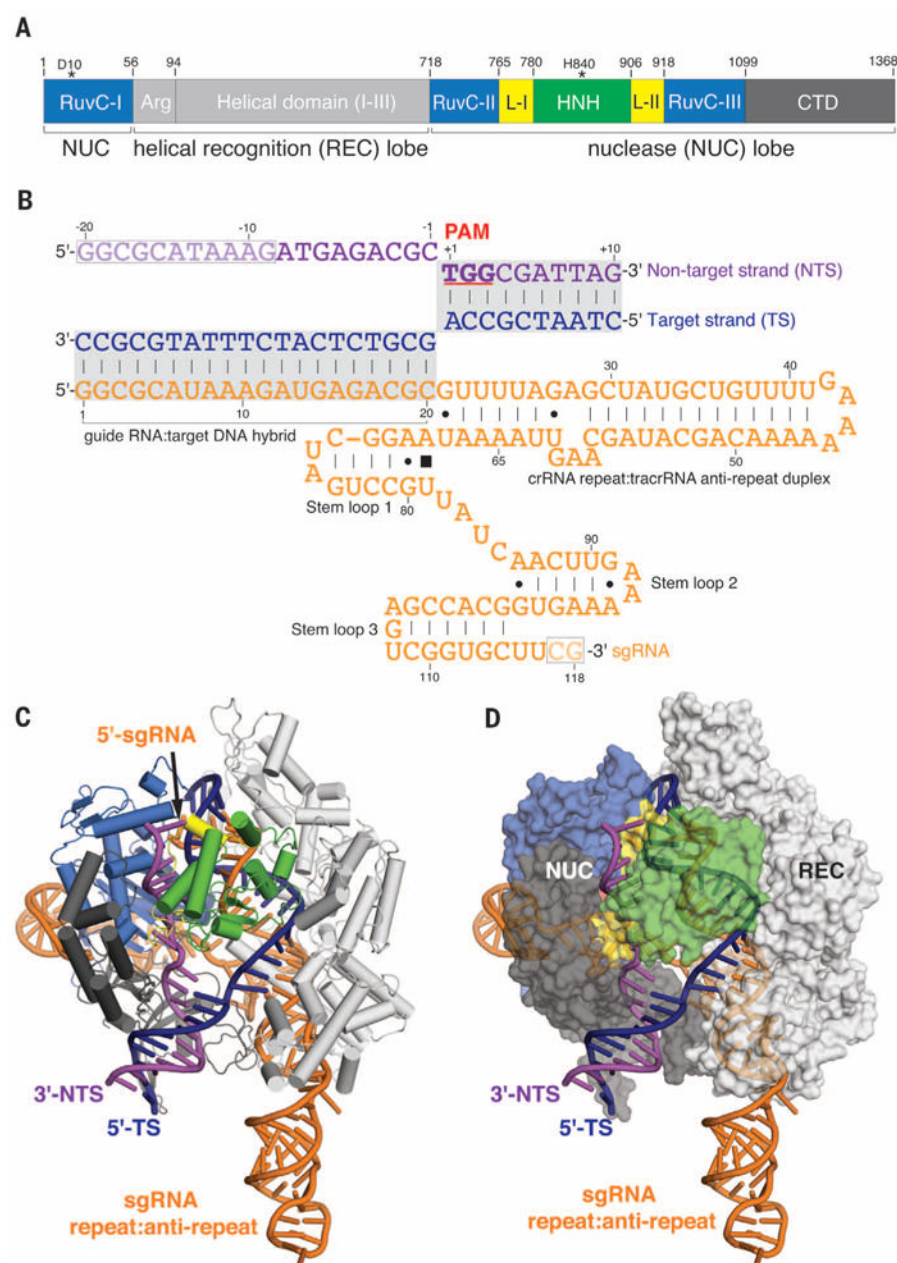


Fig. 1. Crystal structure of Cas9-sgRNA-dsDNA ternary complex

(A) Domain organization of *S. pyogenes* Cas9. (B) Schematic diagram of the sgRNA-target DNA complex. The target DNA strand and displaced nontarget strand are colored dark blue and purple, respectively. The PAM sequence is underlined, and the sgRNA sequence is highlighted in orange. Dashes and dots represent canonical and noncanonical Watson-Crick base pairs, respectively. Filled square denotes the base stacking interaction. Dashed boxes outline the portions of sgRNA and DNA that are not visible in the electron density map. (C and D) Ribbon (C) and surface (D) representations of the Cas9-sgRNA-dsDNA structure, color-coded as defined in (A) and (B).

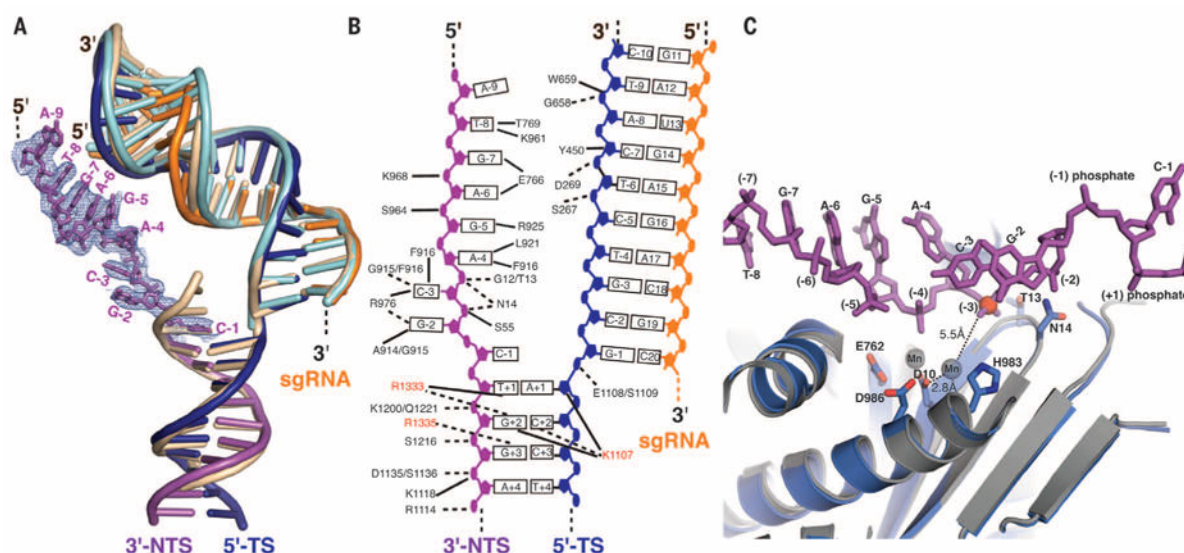


Fig. 2. Protein contacts that open and kink the nontarget DNA strand

(A) Structural comparison of sgRNA-DNA complexes in existing DNA-bound Cas9 structures (with sgRNA scaffold omitted for clarity). The ordered nontarget DNA strand in the dsDNA-bound structure is shown with a simulated-annealing omit $F_o - F_c$ electron density map contoured at 1.5σ . The sgRNA-DNA complexes in the ssDNA-bound (PDB ID 4OO8) and PAM-containing partial duplex-bound (PDB ID 4UN3) structures are colored in cyan and beige, respectively. (B) Schematic showing key Cas9-dsDNA interactions. For clarity, only the unwound nontarget strand and the seed RNA-target DNA heteroduplex are shown. Residues specifically interacting with the PAM motif are highlighted in red. Hydrogen bonds or electrostatic interactions are indicated by dashed lines; hydrophobic contacts and van der Waals interactions are shown as solid lines. (C) Position of nontarget strand in RuvC active site. The red sphere depicts the scissile phosphate. Notably, the RuvC domain in the dsDNA-bound structure (marine) superimposes well with that in the Mn²⁺-bound apo-Cas9 form (gray), with two Mn²⁺ ions fit snugly in the active center.

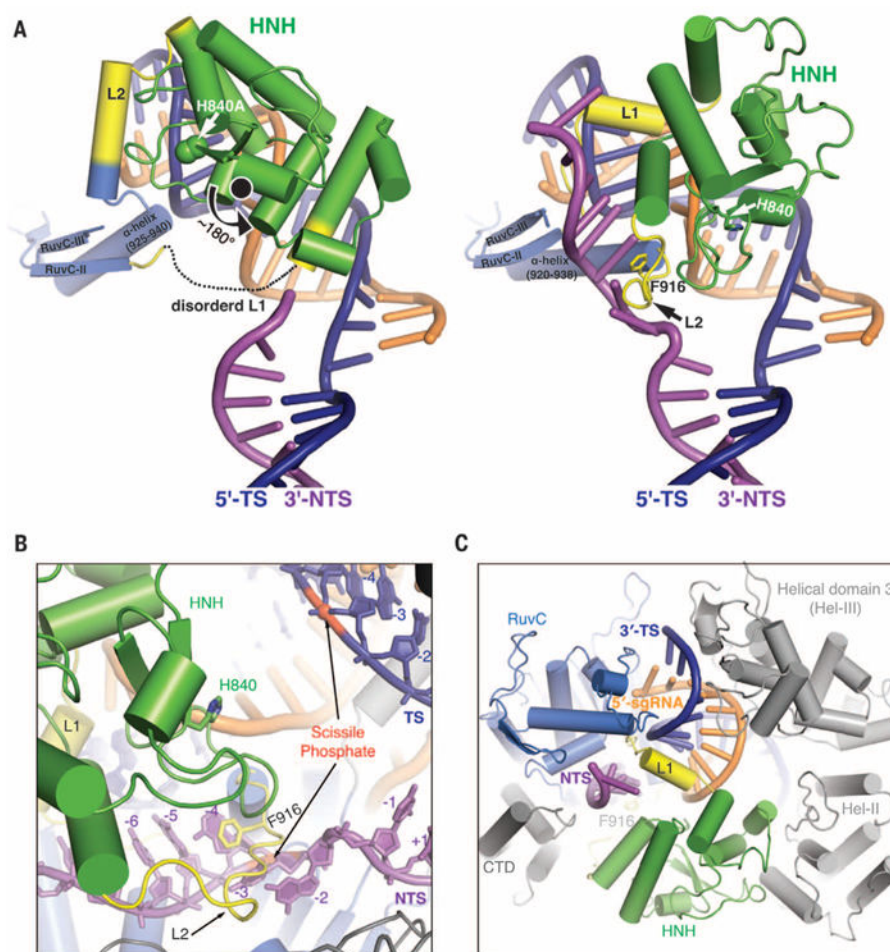


Fig. 3. dsDNA unwinding induces drastic conformational changes in Cas9's HNH domain and the linker regions connecting it to RuvC domain

(A) Comparison of HNH nuclease domain with that observed in a PAM-containing partial duplex-bound structure. The complexes are aligned through the less-flexible RuvC, helical domain 1, and PAM-interacting C-terminal domain (CTD) and presented in the same direction, but side by side. There is a $\sim 180^\circ$ rotation (marked by an arrow) of the L1-HNH-L2 region in the dsDNA-bound structure (right) along an axis perpendicular to the central channel relative to that of the partial duplex-bound state (left). (B) Close-up view of the HNH active site and L2 linker. The aromatic ring of Phe⁹¹⁶ intercalated between positions -3 and -4 is shown as a stick. (C) Close-up view of the L1 linker.

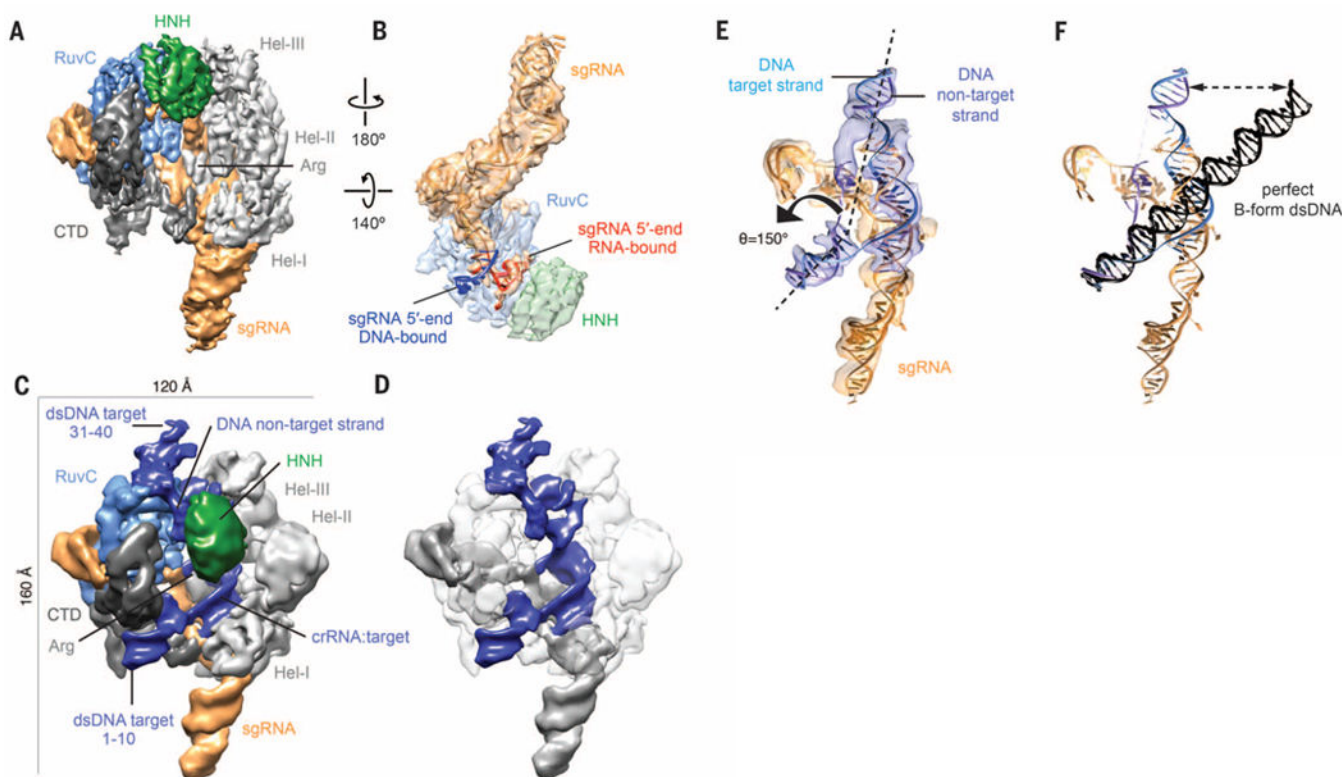


Fig. 4. Architecture of the Cas9-R loop complex

(A and C) Cryo-EM reconstructions of sgRNA-bound Cas9 (A) and Cas9-sgRNA bound to a 40-bp target dsDNA (C) at 4.5- and 6.0-Å resolution (using the 0.143 gold standard Fourier Shell Correlation criterion), respectively. Subunits are segmented and colored as indicated. (B) Close-up view of the sgRNA (orange) from our sgRNA-bound structure showing that the 5' end of the guide (nucleotides 1 to 10, red) undergoes a dramatic kink and is buried within a pocket created by the HNH (light green) and RuvC (light blue) nuclease domains of Cas9. The 5' end of the guide is reorganized upon base pairing with the target strand in the DNA-bound crystal structure presented in Fig. 1 (blue). (D) R-loop density composed of sgRNA (dark gray) and target DNA (dark blue) and Cas9 (transparent surface), with the HNH domain removed for clarity. (E and F) R-loop density composed of sgRNA (orange) and target DNA (dark blue) with the pseudo-atomic model reveals local bending of the dsDNA by ~30°, creating an angle of 150° between ends (arrow) (E) and large distortions as compared to a linear helical duplex extending from the PAM-proximal end (F).