Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems

Citation: Yamada, Mari et al. "Crystal Structure of the Minimal Cas9 from Campylobacter Jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems." Molecular Cell 65, 6 (March 2017): 1109–1121 © 2017 Elsevier Inc

As Published: https://doi.org/10.1016/j.molcel.2017.02.007

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/114953

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Article

Crystal structure of the minimal Cas9 from *Campylobacter jejuni* reveals the molecular diversity in the CRISPR-Cas9 systems

Mari Yamada[1,9], Yuto Watanabe[1,9], Jonathan S. Gootenberg[2,3,4,5,6], Hisato Hirano[1], F. Ann Ran[2,7], Takanori Nakane[1], Ryuichiro Ishitani[1], Feng Zhang[2,3,4,5], Hiroshi Nishimasu[1,8,*], Osamu Nureki[1,10,*]

[1] Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[2] Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[3] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[6] Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[7] Society of Fellows, Harvard University, Cambridge, MA 02138, USA
[8] JST, PRESTO, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[9] These authors equally contributed to this work
[10] Lead Contact

* Correspondence:
nisimasu@bs.s.u-tokyo.ac.jp (H.N.) and nureki@bs.s.u-tokyo.ac.jp (O.N.).

The RNA-guided endonuclease Cas9 generates a double-strand break at DNA target sites complementary to the guide RNA, and has been harnessed for the development of a variety of new technologies, such as genome editing. Here, we report the crystal structures of *Campylobacter jejuni* Cas9 (CjCas9), one of the smallest Cas9 orthologs, in complex with an sgRNA and its target DNA. The structures provided insights into a minimal Cas9 scaffold, and revealed the remarkable mechanistic diversity of the CRISPR-Cas9 systems. The CjCas9 guide RNA contains an unexpected triple-helix structure, which is distinct from known RNA triple helices, thereby expanding the natural repertoire of RNA triple helices. Furthermore, unlike the other Cas9 orthologs, CjCas9 contacts the nucleotide sequences in both the target and non-target DNA strands, and recognizes the 5′-NNNVRYM-3′ as the protospacer adjacent motif. Collectively, these findings improve our mechanistic understanding of the CRISPR-Cas9 systems, and may facilitate Cas9 engineering.

Introduction

Bacteria and Archaea utilize CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated) adaptive immune systems to defend themselves against foreign genetic elements, such as plasmids and phages (Marraffini, 2015; Barrangou and Doudna, 2016; Mohanraju et al., 2016; Wright et al., 2016). The CRISPR loci in the genome comprise a *cas* operon and a CRISPR array, consisting of short repetitive sequences (direct repeats) separated by non-repetitive sequences (spacers) derived from foreign genetic elements. The CRISPR array is transcribed and processed into CRISPR RNAs (crRNAs), which associate with single or multiple Cas proteins to form effector ribonucleoprotein complexes responsible for the destruction of invading nucleic acids (Makarova et al., 2015; Nishimasu and Nureki, 2016). In the type II CRISPR-Cas system, the Cas9 effector nuclease associates with dual guide RNAs (crRNA and *trans*-activating crRNA (tracrRNA)), and cleaves double-stranded (ds) DNA targets complementary to the crRNA guide (Garneau et al., 2010; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012). In addition to the crRNA-target DNA complementarity, Cas9-mediated target recognition requires a PAM (protospacer adjacent motif), a short nucleotide sequence adjacent to the target site (Deveau et al., 2008; Garneau et al., 2010). Importantly, a single-guide RNA (sgRNA), in which crRNA and tracrRNA are fused with an artificial tetraloop, can also direct Cas9 to the target cleavage (Jinek et al., 2012). Thus, the two component Cas9-sgRNA system has been harnessed for a variety of new technologies, including genome editing (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013).

Structural studies of *Streptococcus pyogenes* Cas9 (SpCas9) have provided mechanistic details of the RNA-guided DNA cleavage by the Cas9 enzyme. The crystal structure of SpCas9 in its apo form revealed a bilobed architecture comprising an α-helical recognition (REC) lobe and a nuclease (NUC) lobe (Jinek et al., 2014). The crystal structure of SpCas9 bound to the sgRNA and its single-stranded DNA target clarified the recognition mechanism of the sgRNA and the target DNA (Nishimasu et al., 2014). Subsequently, the crystal structure of SpCas9 bound to the sgRNA and a PAM-containing DNA revealed the recognition mechanism of the 5′-NGG-3′ PAM by SpCas9 (Anders et al., 2014). Moreover, the crystal structures of SpCas9 bound to the sgRNA (Jiang et al., 2015) and SpCas9 bound to an R-loop (Jiang et al., 2016) demonstrated the structural rearrangements in the Cas9 protein accompanying the guide RNA binding and R-loop formation, respectively.

The Cas9 orthologs from different microbes have highly divergent sequences, function with their cognate crRNA:tracrRNA guides, and recognize a variety of PAM sequences (Chylinski et al., 2013; Fonfara et al., 2014; Hsu et al., 2014; Karvelis et al., 2015; Ran et al., 2015). SpCas9 (1,368 aa) recognizes 5′-NGG-3′ as the PAM (Mojica et al., 2009), whereas *Staphylococcus aureus* Cas9 (SaCas9, 1,053 aa) and *Francisella novicida* Cas9 (FnCas9, 1,629 aa) recognize 5′-NNGRRT-3′ and 5′-NGG-3′ as the PAMs, respectively (Ran et al., 2015; Hirano et al., 2016). A structural comparison of SpCas9 (Anders et al., 2014; Nishimasu et al., 2014) with SaCas9 (Nishimasu et al., 2015) and FnCas9 (Hirano et al., 2016) revealed that, while they share the conserved RuvC and HNH nuclease domains, their REC and Wedge (WED) domains are structurally divergent, and recognize distinct structural features in their cognate RNA guides. In addition, their PAM-interacting (PI) domains share a conserved core fold, but recognize distinct PAM sequences, using a specific set of amino-acid residues.

Campylobacter jejuni Cas9 (CjCas9) has several unique features (Fonfara et al., 2014). First, CjCas9 consists of 984 residues, and is one of the smallest Cas9 orthologs. Second, the nucleotide sequences of the crRNA:tracrRNA guides for CjCas9 and the other Cas9 orthologs differ substantially. Third, CjCas9 recognizes the 5′-NNNNACA-3′ PAM, whereas most Cas9 orthologs, as exemplified by SpCas9 (5′-NGG-3′) (Gasiunas et al., 2012; Jinek et al., 2012), recognize G-rich PAMs. However, the functional mechanism of CjCas9 remains elusive, due to the lack of structural information.

Here, we performed functional and structural characterizations of CjCas9. *In vitro* cleavage experiments revealed that CjCas9 recognizes the 5′-NNNVRYM-3′ PAM, which is more promiscuous than the previously reported PAMs. The crystal structure of the CjCas9-sgRNA-target DNA complex highlighted the remarkable mechanistic diversity of the CRISPR-Cas9 systems. Unlike the tracrRNAs for the other Cas9 orthologs, the CjCas9 tracrRNA has an unanticipated triple-helix structure, which is distinct from known RNA triple helices. Furthermore, CjCas9 recognizes the PAM nucleotides on both the target and non-target DNA strands, whereas the other Cas9 orthologs recognize the PAM nucleotides on the non-target DNA strand.

Results and Discussion

CjCas9 PAM specificity

Although a previous study reported that CjCas9 recognizes the 5′-NNNNACA-3′ PAM (Fonfara et al., 2014), the CjCas9 PAM has not been fully characterized. To determine the CjCas9 PAM, we performed the PAM discovery assay, using purified CjCas9, an sgRNA and a library of plasmid DNA targets with a degenerate 7-bp PAM sequence, as described previously (Ran et al., 2015; Zetsche et al., 2015). The result revealed that CjCas9 recognizes the 5′-NNNVRYM-3′ PAM (V is A/G/C; R is A/G; Y is T/C; M is A/C) (Figures 1A and S1), which is more promiscuous than the previously reported 5′-NNNNACA-3′ PAM (Fonfara et al., 2014). Using purified CjCas9 and an sgRNA, we further examined the cleavage of 13 plasmid DNA targets with either 5′-AGANACC-3′, 5′-AGAANCA-3′, 5′-AGAAANA-3′ or 5′-AGAACN-3′ as the PAMs. The results confirmed that CjCas9 efficiently recognizes the 5′-NNNVRYM-3′ PAM, with the preference for T and C at positions 6 and 7, respectively (Figures 1B and 1C).

Crystal structure of the CjCas9-sgRNA-target DNA complex

To clarify the RNA-guided DNA recognition mechanism of CjCas9, we attempted to determine the crystal structure of CjCas9 in complex with an sgRNA and its target DNA. However, we failed to obtain diffraction-quality crystals. Previous studies revealed that the HNH nuclease domain of SpCas9 is mobile and dispensable for RNA-guided DNA recognition (Nishimasu et al., 2014; Jiang et al., 2015; Sternberg et al., 2015), suggesting that the flexibility of the HNH domain may hamper the crystallization. We thus prepared the CjCas9-ΔHNH variant lacking the HNH domain (residues 481–640), in which Leu480 (RuvC-II) and Tyr641 (RuvC-III) are connected by a GGGSGG linker (Figure 2A). After extensive crystallization screening, we determined the crystal structure of CjCas9-ΔHNH in complex with a 93-nt sgRNA, a 28-nt target DNA strand, and an 8-nt non-target DNA strand containing the 5′-AGAAACC-3′ PAM, at 2.4 Å resolution (Figures 2A–2C and Table 1).

The crystal structure revealed that CjCas9 adopts a bilobed architecture comprising the α-helical REC lobe and a NUC lobe, as in the other Cas9 orthologs (Anders et al., 2014; Nishimasu et al., 2014; Nishimasu et al., 2015) (Figure 2C), indicating that the truncation of the HNH domain does not substantially affect the overall structure. The REC lobe can be divided into the REC1 (residues 77–234) and REC2 (residues 235–426) domains. The NUC lobe comprises the RuvC (residues 1–44, 427–480 and 641–777), WED (residues 792–827), and PI domains (residues 828–984) (the HNH domain was truncated for crystallization). The REC and NUC lobes are connected by an arginine-rich “bridge” helix (residues 45–76), while the WED and RuvC domains are connected by a “phosphate lock” loop (residues 778–791), as in other Cas9

orthologs (Anders et al., 2014; Nishimasu et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016). The three residues (GGS) in the GGGSGG linker between the RuvC-II and RuvC-III motifs are disordered in the present structure.

The sgRNA comprises the guide segment (G1–C20), the repeat region (G21–U32), the tetraloop (G33–A36), the antirepeat region (A37–U48), and the tracrRNA scaffold (A49–C93) (Figure 2B). The guide segment (G1–C20) and the target DNA strand (dG1–dC20) form an RNA-DNA heteroduplex (Figures 2B and 2C). The repeat and antirepeat regions form the A-form-like duplex (referred to as the repeat-antirepeat duplex), which consists of a wobble base pair (G21•U48) and eleven Watson-Crick base pairs (U22-A47–U32-A37) (Figures 2B and 2C). The RNA-DNA heteroduplex is bound within the central channel between the REC and NUC lobes, while the repeat-antirepeat duplex is sandwiched between the REC1 and WED domains (Figure 2C). These duplexes are primarily recognized by the protein in a non-sequence-specific manner (Figure S2). The target DNA strand (dC(−8)–dT(−1)) and the non-target DNA strand (dA1*–dG8*) form a PAM-containing duplex, which is recognized by the WED and PI domains (Figures 2B and 2C).

A recent study showed that the deletion of the HNH domain in SpCas9 impairs the non-target strand cleavage by the RuvC domain, thereby suggesting that the HNH domain is required for the activation of the RuvC domain (Sternberg et al., 2015). We thus examined the effect of the deletion of the HNH domain on the CjCas9 activity, using *in vitro* cleavage assays. Our results revealed that, like the D8A RuvC-inactive nickase and the H559A HNH-inactive nickase, the CjCas9-ΔHNH variant functions as a nickase (Figure 2D). Importantly, the CjCas9-ΔHNH variant exhibited lower cleavage activity, as compared with the H559A nickase, indicating that the deletion, but not the inactivating point mutation, of the HNH domain reduces the non-target strand cleavage by the RuvC domain. These observations suggested the allosteric communication between the RuvC and HNH nuclease domains in CjCas9, as observed in SpCas9 (Sternberg et al., 2015).

TracrRNA architecture

Notably, the present structure revealed that the CjCas9 tracrRNA scaffold contains a triple-helix structure within a pseudoknot comprising three stem regions, which was not predicted from its primary sequence (Figures 3A–3C). Stem 1 consists of four canonical base pairs (A51-U67–G54-C64) and a non-canonical A50•G68 base pair, while stem 2 consists of four canonical base

pairs (G58-C93–G61-C90) and a non-canonical G62•A89 base pair (Figures 3B and 3C). Stem 3 consists of four canonical base pairs (G70-C84–G74-C81) (Figures 3B and 3C). Nucleotides 56/57 and 86/87 form base pairs with stem 2 and stem 1, respectively, thereby contributing to the triple-helix formation. U65 in stem 1 and A89 in stem 2 base pair with A86 and U56, forming an A53-U65•A86 minor-groove triple and a G62-A89•U56 major-groove triple, respectively (Figures 3D and 3E). U57 base pairs with C59 and C90, forming a G92-C59•U57•C90-G62 base quintuple (Figure 3F). A87 and A88 base pair with G54 and U55/C64, respectively, forming a U55•A88•C64-G54•A87 base quintuple (Figure 3G).

The tracrRNA scaffold is extensively recognized by CjCas9 (Figures 4A and S2). In particular, A63, A76 and U80 are flipped out, and recognized by the protein in base-specific manners. The nucleobase and ribose moieties of A63 form stacking interactions with the side chains of His70 and His67, respectively, while the N1 of A63 hydrogen bonds with the side chain of Asn74 (Figure 4B). A76 and U80 are accommodated within specific pockets in the PI domain (Figure 4A). The nucleobase of A76 is sandwiched between the side chains of Ile964 and Arg977, while the N6 and N7 of A76 hydrogen bond with the main-chain carbonyl and amide groups of Glu975, respectively (Figure 4C). The nucleobase of U80 is sandwiched between the side chains of Phe854 and Glu980, while the N3 and O4 of U80 hydrogen bond with the main-chain carbonyl group of Asp981 and the side chain of Arg832, respectively (Figure 4C). Indeed, the single mutations (A63U, A76U or U80A) reduced CjCas9-mediated DNA cleavage (Figure 4D), confirming the functional importance of the three flipped-out nucleotides. Moreover, a 4-nt substitution (nucleotides 90–93) or an 8-nt deletion (nucleotides 86–93) in the tracrRNA 3′ tail abolished the CjCas9-mediated DNA cleavage (Figure 4D), indicating that the triple-helix structure of the tracrRNA is critical for the activity. In addition, the SpCas9 sgRNA did not support the CjCas9-mediated DNA cleavage (Figure 4D). Together, these results demonstrated that CjCas9 specifically recognizes its cognate RNA guide.

5′-NNNVRYM-3′ PAM recognition

In the present structure, the 5′-AGAAACC-3′ PAM-containing DNA duplex is bound to the cleft between the WED and PI domains (Figure 5A). The nucleobases of dA1*–dA3* do not directly contact the protein (Figures 5B and 5C), consistent with the lack of specificity for positions 1–3 in the 5′-NNNVRYM-3′ PAM. The N7 of dA4* in the non-target strand forms a water-mediated hydrogen bond with the side-chain hydroxyl group of Thr913 (Figures 5B and 5C). Modeling suggested that a steric clash could occur between the methyl group of dT4* and

the side chain of Thr913 (Figures S3A and S3B), consistent with the preference of CjCas9 for the fourth V (A/G/C). The N7 of dA5* in the non-target strand forms a hydrogen bond with the side-chain hydroxyl group of Ser915 (Figures 5B and 5C). Since N7 is common among the purine nucleotides, the interaction can explain the requirement for the fifth R (A/G). Notably, the nucleobase of dC6* in the non-target strand is not recognized by the protein (Figures 5B and 5C). Instead, the N7 of dG(−6) in the target strand forms a hydrogen bond with the side-chain hydroxyl group of Ser951 (Figures 5B and 5C). These structural findings revealed that CjCas9 does not recognize the Y (T/C) nucleotides at position 6 in the non-target strand as the PAM, but detects their complementary R (A/G) nucleotides in the target strand. Similarly, the nucleobase of dC7* in the non-target strand is not recognized by the protein, whereas the O6 and N7 of dG(−7) in the target strand form bidentate hydrogen bonds with the side chain of Arg866 (Figures 5B and 5C). In addition to the 5′-AGAAACC-3′ PAM complex, we determined the crystal structure of CjCas9-ΔHNH in complex with the sgRNA and the DNA target containing the 5′-AGAAACA-3′ PAM (Table 1). In the 5′-AGAAACA-3′ PAM complex, the dT(−7):dA7* pair in the PAM duplex undergo a slight displacement toward the PI domain, as compared with the dG(−7):dC7* pair in the 5′-AGAAACC-3′ PAM complex (Figure S3C). This displacement in the PAM duplex allows Arg866 to form a hydrogen bond with the O4 of dT(−7) in the target strand (Figure S3D). These observations revealed that CjCas9 does not recognize the M (A/C) nucleotides at position 7 in the non-target strand as the PAM, but detects their complementary K (T/G) nucleotides in the target strand. The preference of CjCas9 for C over A at position 7 can be explained by the bidentate hydrogen-bonding interaction between dG(−7) and Arg866, in contrast to the single hydrogen-bonding interaction between dT(−7) and Arg866. The single mutations of Arg866, Thr913, Ser915, and Ser951 reduced or abolished the *in vitro* cleavage activity (Figure 5D), confirming their functional importance. Together, our structural and functional data revealed that CjCas9 forms sequence-specific contacts with both the target and non-target DNA strands, to achieve the recognition of the 5′-NNNVRYM-3′ PAM.

Structural comparison between the Cas9 orthologs

A structural comparison of CjCas9 with the other Cas9 orthologs highlighted the structural similarities and differences between the CRISPR-Cas9 systems, and provided insights into a minimal Cas9 scaffold (Figure 6). Unlike SpCas9 (Anders et al., 2014; Nishimasu et al., 2014), the smaller SaCas9 (Nishimasu et al., 2015) and CjCas9 lack the insertion subdomains within the REC1 and PI domains (Figures 6A–6C). Furthermore, the WED domain of CjCas9 (36

amino acids) is smaller than that of SaCas9 (122 amino acids) (Figures 6A and 6C). These structural differences contribute to the miniaturization of CjCas9. The REC and WED domains of FnCas9, one of the largest Cas9 orthologs, adopt protein folds distinct from those of CjCas9, SpCas9 and SaCas9 (Hirano et al., 2016) (Figure 6D), reinforcing the notion that FnCas9 is distantly related to the other Cas9 orthologs.

Despite the structural differences in these individual domains, CjCas9 adopts the conserved bilobed architecture, and accommodates the RNA-DNA heteroduplex in similar manners to those of the other Cas9 orthologs (Figures 6A–6D). The sugar-phosphate backbone of the sgRNA "seed" region (C13–C20) is extensively recognized by conserved arginine residues in the bridge helix (Figure S4A). In addition, the backbone phosphate group between dG1 and dT(−1) in the target DNA strand (referred to as the +1 phosphate (Anders et al., 2014)) interacts with the main-chain amide groups of Glu790 and Thr791 and the side-chain hydroxyl group of Thr791 in the phosphate lock loop, thereby facilitating target DNA unwinding (Figure S4B). Indeed, the T791A mutation abolished the *in vitro* DNA cleavage activity (Figure S4C), confirming the functional importance of the interaction between the +1 phosphate and Thr791. These observations confirmed that the RNA-guided DNA targeting mechanism is highly conserved among the CRISPR-Cas9 systems.

The present structure also illuminated the structural diversity of the crRNA:tracrRNA guides in the CRISPR-Cas9 systems (Figures 6E–6H). The repeat-antirepeat duplexes for SpCas9, SaCas9 and FnCas9 contain several unpaired nucleotides, and thus adopt distorted, distinct structures (Nishimasu et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016) (Figures 6F–6H). In contrast, the CjCas9 repeat-antirepeat duplex adopts an A-form-like conformation (Figure 6E). According to these structural differences, the repeat-antirepeat duplexes are recognized by the structurally divergent REC1 and WED domains in species-specific manners (Figure S5). The CjCas9-REC1 adopts a conserved core fold, but has two unique loops (loops 1 and 2) that interact with the repeat-antirepeat duplex. The repeat-antirepeat duplex is further recognized by the WED domain, which is structurally distinct from those of the other Cas9 orthologs. Furthermore, the present structure revealed the notable architectural differences in the tracrRNA scaffolds. The SpCas9 and SaCas9 tracrRNA scaffolds contain three and two stem loops, respectively, and the first and second stem loops are connected by a single-stranded linker (although stem loop 2 of SaCas9 was truncated for crystallization) (Nishimasu et al., 2014; Nishimasu et al., 2015) (Figures 6F and 6G). The FnCas9 tracrRNA scaffold contains two stem

loops, which are connected by a U-shaped linker (Hirano et al., 2016) (Figure 6H). In stark contrast, the CjCas9 tracrRNA scaffold contains a more complicated triple-helix structure, as described above (Figure 6E).

Mechanistic diversity in PAM recognition

Despite their limited sequence similarity, the PI domains of the Cas9 orthologs share a similar core fold comprising two distorted β-sheets (β1–β3 and β4–β9) (Figures 7A–7D). In SpCas9, SaCas9 and FnCas9, distinct sets of amino-acid residues in the β5–β7 region form sequence-specific contacts with the PAM nucleotides on the non-target DNA strand (Anders et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016). In SpCas9, Arg1333 and Arg1335 form bidentate hydrogen bonds with the second and third Gs in the 5′-NGG-3′ PAM, respectively (Anders et al., 2014) (Figure 7B). In SaCas9, Arg1015 forms a bidentate hydrogen bond with the third G in the 5′-NNGRRT-3′ PAM, while Asn985, Asn986 and Arg991 form a hydrogen-bonding network with the RRT nucleotides (Nishimasu et al., 2015) (Figure 7C). In FnCas9, Arg1585 and Arg1556 form bidentate hydrogen bonds with the second and third Gs in the 5′-NGG-3′ PAM, respectively (Hirano et al., 2016) (Figure 7D). In contrast to these Cas9 orthologs, CjCas9 forms sequence-specific contacts with the PAM nucleotides on the non-target strand and the PAM-complementary nucleotides on the target strand (Figure 7A), illuminating the mechanistic diversity of Cas9-mediated PAM recognition. Intriguingly, a recent study showed that the mutations of the PAM-complementary nucleotides on the target strand abolished the cleavage activity of *Neisseria meningitidis* Cas9 (Zhang et al., 2015), suggesting that other Cas9 orthologs, such as *N. meningitidis* Cas9, also form sequence-specific interactions with both the target and non-target DNA strands in the PAM duplex, as observed in CjCas9. Further studies will be required to fully elucidate the mechanistic diversity in the PAM recognition by Cas9 orthologs.

Comparison of the cleavage activities of CjCas9 and SpCas9

A recent study showed that the type II-C Cas9 from *Corynebacterium diphtheriae* (CdCas9), which consists of 1084 residues and shares 21% sequence identity with CjCas9, has limited unwinding and cleavage activities toward dsDNA targets, as compared with SpCas9 (Ma et al., 2015). This result suggested that the inefficiency of most of the type II-C Cas9 orthologs for genome editing results from their limited dsDNA cleavage activities (Ma et al., 2015). To examine the differences in the catalytic features of CjCas9 and SpCas9, we compared their *in vitro* dsDNA cleavage activities. Our data revealed that, like CdCas9, CjCas9 cleaves the target

dsDNA less efficiently, as compared with SpCas9 (Figure S6). These results support the notion that the type II-C Cas9 enzymes, such as CdCas9 and CjCas9, have not been harnessed for genome editing at least partly because of their relatively poor activities. Thus, it is possible that an engineered CjCas9 variant with improved dsDNA cleavage activity could be used for eukaryotic genome editing. Although CjCas9 and CdCas9 commonly exhibit relatively weak dsDNA cleavage activities, they may have distinct specificities for their cognate RNA guides. In contrast to CjCas9, which is specific to its cognate sgRNA, CdCas9 promiscuously recognizes the SpCas9 sgRNA as well as its cognate sgRNA (Ma et al., 2015). Further structural studies will provide insights into the mechanistic diversity among the type II-C CRISPR-Cas9 systems.

Structural comparison of the CjCas9 tracrRNA and other known RNA triplexes

A structural comparison of the CjCas9 tracrRNA with previously characterized RNA triplexes, such as the telomerase RNA subunit TER (Theimer et al., 2005), the SAM-II riboswitch (Gilbert et al., 2008), and the long noncoding RNA MALAT1 (Brown et al., 2014), revealed notable differences between the CjCas9 tracrRNA and the other RNA triplexes (Figure S7). Notably, the CjCas9 tracrRNA lacks a canonical U•A-U triple, whereas the other three RNA triplexes contain successive U•A-U triples (Figure S7). TER and SAM-II have three and seven U•A-U triples in their core regions, respectively. In particular, MALAT1 forms a bipartite triple helix containing stacks of five and four U•A-U triples. Moreover, in the CjCas9 tracrRNA, the L1 region between the S1 and S2 regions consists of three nucleotides and is shorter than those of the other three RNA triplexes (Figure S7). Thus, the CjCas9 tracrRNA adopts a distorted structure, which is stabilized by the two base triples and the two base quintuples. Taken together, the CjCas9-sgRNA-DNA structure revealed the previously unrecognized biological role of an RNA triple helix.

Conclusions

The present structure of the CjCas9-sgRNA-target DNA complex unveiled the remarkable diversity among the CRISPR-Cas9 systems. First, unlike the other tracrRNAs, the CjCas9 tracrRNA contains a triple-helix architecture, which is distinct from other known RNA triplexes, thereby expanding the natural repertoire of RNA triplexes. Second, unlike other Cas9 orthologs, CjCas9 reads the nucleotide sequences in both the target and non-target DNA strands as the PAM. Although CjCas9 has not been harnessed for genome editing applications, our structural findings may provide clues for Cas9 engineering, such as miniaturization of Cas9 and alteration of its PAM specificities.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569-573.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat Biotechnol *34*, 933-941.

Brown, J.A., Bulkley, D., Wang, J., Valenstein, M.L., Yario, T.A., Steitz, T.A., and Steitz, J.A. (2014). Structural insights into the stabilization of MALAT1 noncoding RNA by a bipartite triple helix. Nat Struct Mol Biol *21*, 633-640.

Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol *10*, 726-737.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819-823.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr *62*, 1002-1011.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res *14*, 1188-1190.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. Nature *471*, 602-607.

Deveau, H., Barrangou, R., Garneau, J.E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol *190*, 1390-1400.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Evans, P.R., and Murshudov, G.N. (2013). How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr *69*, 1204-1214.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lecrivain, A.L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res *42*, 2577-2590.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67-71.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A *109*, E2579-2586.

Gilbert, S.D., Rambo, R.P., Van Tyne, D., and Batey, R.T. (2008). Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. Nat Struct Mol Biol *15*, 177-182.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F.*, et al.* (2016). Structure and engineering of *Francisella novicida* Cas9. Cell *164*, 950-961.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867-871.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477-1481.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. Elife *2*, e00471.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S.*, et al.* (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol *16*, 253.

Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. Mol Cell *62*, 137-147.

Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. Nucleic Acids Res *30*, 3497-3531.

Ma, E., Harrington, L.B., O'Connell, M.R., Zhou, K., and Doudna, J.A. (2015). Single-stranded DNA cleavage by divergent CRISPR-Cas9 enzymes. Mol Cell *60*, 398-407.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H.*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat Rev Microbiol *13*, 722-736.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55-61.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147.

Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733-740.

Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., and Nureki, O. (2015). Crystal structure of *Staphylococcus aureus* Cas9. Cell *162*, 1113-1126.

Nishimasu, H., and Nureki, O. (2016). Structures and mechanisms of CRISPR RNA-guided effector nucleases. Curr Opin Struct Biol *43*, 68-78.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935-949.

Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics *12*, 385.

Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S.*, et al.* (2015). *In vivo* genome editing using *Staphylococcus aureus* Cas9. Nature *520*, 186-191.

Sternberg, S.H., LaFrance, B., Kaplan, M., and Doudna, J.A. (2015). Conformational control of DNA target cleavage by CRISPR-Cas9. Nature *527*, 110-113.

Theimer, C.A., Blois, C.A., and Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. Mol Cell *17*, 671-682.

Waterman, D.G., Winter, G., Parkhurst, J.M., Fuentes-Montero, L., Hattne, J., Brewster, A., Sauter, N.K., Evans, G., and Rosenstrom, P. (2013). The DIALS framework for integration software. CCP4 Newsletter *49*, 16-19.

Wright, A.V., Nunez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. Cell *164*, 29-44.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A.*, et al.* (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759-771.

Zhang, Y., Rajan, R., Seifert, H.S., Mondragon, A., and Sontheimer, E.J. (2015). DNase H activity of *Neisseria meningitidis* Cas9. Mol Cell *60*, 242-255.

Author Contributions

M.Y. performed *in vitro* cleavage experiments and crystallized the complexes, with assistance from H.H. and H.N.; Y.W. initially obtained diffraction-quality crystals; J.S.G. and F.A.R. performed PAM screens; M.Y., T.N., R.I. and H.N. determined the crystal structures; H.N. conceived the crystallization strategy; M.Y., H.H., H.N. and O.N. wrote the manuscript with help from all authors. F.Z., H.N. and O.N. supervised all of the research.

Acknowledgements

We thank the beamline scientists at PXI at the Swiss Light Source and BL32XU and BL41XU at SPring-8 for assistance with data collection. J.S.G. is supported by a D.O.E. Computational Science Graduate Fellowship. F.A.R. is a Junior Fellow at the Harvard Society of Fellows. F.Z. is supported by the National Institutes of Health through NIMH (5DP1-MH100706 and 1R01-MH110049) and NIDDK (5R01DK097768-03), the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee Foundations; and B. Metcalfe. F.Z. is a New York Stem Cell Foundation Robertson Investigator. F.Z. is a founder of Editas Medicine and a scientific advisor for Editas Medicine and Horizon Discovery. H.N. is supported by JST, PRESTO, JSPS KAKENHI Grant Numbers 26291010 and 15H01463. O.N. is supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Japan Agency for Medical Research and Development, AMED, and the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for creating next-generation agriculture, forestry and fisheries" (funding agency: Bio-oriented Technology Research Advancement Institution, NARO), and the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

Accession Numbers

The accession numbers for the atomic coordinates of the CjCas9-sgRNA-DNA complexes reported in this paper are Protein Data Bank: 5X2G (5′-AGAAACC-3′ PAM) and 5X2H (5′-AGAAACA-3′ PAM).

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Requests for further information and reagents should be directed to the Lead Contact, Osamu Nureki (nureki@bs.s.u-tokyo.ac.jp).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The plasmid DNAs were amplified in *Escherichia coli* Mach (Thermo Fisher Scientific), cultured in LB medium (Nacalai Tesque) at 37°C overnight. The recombinant proteins were overexpressed in *E. coli* Rosetta 2 (DE3) (Novagen). The *E. coli* cells were cultured at 37°C in LB medium (containing 20 mg/l kanamycin) until the OD_{600} reached 0.8, and then protein expression was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (Nacalai Tesque) and an incubation at 20°C for 20 h.

Sample preparation

The gene encoding full-length CjCas9 (residues 1–984) was codon optimized, synthesized (Genscript), and cloned between the *Nde*I and *Xho*I sites of the modified pE-SUMO vector (LifeSensors). For crystallization, we prepared the CjCas9-ΔHNH variant lacking the HNH domain (residues 481–640), in which Leu480 (RuvC-II) and Tyr641 (RuvC-III) are connected by a GGGSGG linker. The CjCas9-ΔHNH variant was created by a PCR-based method, using the vector encoding the full-length CjCas9 as the template. The CjCas9-ΔHNH protein was expressed at 20°C in *E. coli* Rosetta 2 (DE3), and was purified by chromatography on Ni-NTA Superflow resin (QIAGEN). The eluted protein was purified by chromatography on a HiTrap Heparin HP column (GE Healthcare), and was then dialyzed overnight at 20°C with TEV protease, to remove the N-terminal His_6-SUMO-tag. The CjCas9-ΔHNH protein was further purified by chromatography on NiNTA and HiLoad 16/600 Superdex 200 (GE Healthcare) columns. The selenomethionine (SeMet)-substituted CjCas9-ΔHNH was expressed in *E. coli* B834 (DE3) (Novagen), and was purified using a similar protocol to that for the native protein. The sgRNA was transcribed *in vitro* with T7 RNA polymerase, using a PCR-amplified dsDNA template. The transcribed RNA was purified by 8% denaturing (7 M urea) polyacrylamide gel electrophoresis. The target and non-target DNA strands were purchased from Sigma-Aldrich. The purified CjCas9-ΔHNH protein was mixed with the sgRNA, the target DNA strand, and the non-target DNA strand (containing either the 5′-AGAAACA-3′ PAM or the 5′-AGAAACC-3′ PAM) (molar ratio, 1:1.5:2.3:3.4), and then the CjCas9-sgRNA-DNA complex was purified by

gel filtration chromatography on a Superdex 200 Increase column (GE Healthcare). For *in vitro* cleavage assays, the wild type and mutants of full-length CjCas9 were expressed and purified, using a protocol similar to that for CjCas9-ΔHNH.

Crystallography

The purified CjCas9-sgRNA-DNA complex (containing either the 5′-AGAAACA-3′ PAM or the 5′-AGAAACC-3′ PAM) was grown at 20°C, using the hanging-drop vapor diffusion method. Crystals were obtained by mixing 1 μl of complex solution ($A_{260\ nm}$ = 15) and 1 μl of reservoir solution (12.0–14.5% PEG 2,000, 0.4 M ammonium acetate). The SeMet-labeled complex (containing the 5′-AGAAACA-3′ PAM) was crystallized under similar conditions. X-ray diffraction data were collected at 100 K on beamlines BL41XU at SPring-8 and PXI at the Swiss Light Source. The crystals were cryoprotected in reservoir solution supplemented with 25% ethylene glycol. X-ray diffraction data were processed using DIALS (Waterman et al., 2013) and AIMLESS (Evans and Murshudov, 2013). The structure was determined by the Se-SAD method, using PHENIX AutoSol (Adams et al., 2010). The model was automatically built using Buccaneer (Cowtan, 2006), followed by manual model building using COOT (Emsley and Cowtan, 2004) and structural refinement using PHENIX (Adams et al., 2010). The final models of the 5′-AGAAACA-3′ PAM complex (2.3 Å resolution) and the 5′-AGAAACC-3′ PAM complex (2.4 Å resolution) were refined using native data sets. Data collection statistics are summarized in Table 1. Structural figures were prepared using CueMol (http://www.cuemol.org).

PAM discovery assay

To generate plasmid libraries containing randomized PAM sequences, synthesized ssDNA oligonucleotides (Integrated DNA Technologies), consisting of seven randomized nucleotides 3′ of a 20-nt target sequence, were used to generate dsDNAs through annealing to a short primer and extension by the large Klenow fragment (New England Biolabs). The dsDNAs were subsequently cloned into linearized pUC19 with Gibson cloning (New England Biolabs). To propagate and purify the cloned plasmids, the products were used to transform $>10^7$ competent Stbl3 *E. coli* cells (Invitrogen), which were pooled and harvested with a Maxi-prep kit (QIAGEN) after overnight growth. The randomized PAM plasmid library was cleaved *in vitro*, using purified CjCas9 with sgRNAs targeting the PAM library, and the cleavage products were separated on 2% agarose E-gels (Life Technologies). The band corresponding to the un-cleaved target plasmid was isolated with a Zymoclean Gel DNA Recovery Kit (Zymo Research), and the region surrounding the randomized PAM region was PCR-amplified and sequenced using a

MiSeq sequencer (Illumina) with 150 single-end cycles. To analyze the resulting sequence data, the seven nucleotide PAM region was extracted, the individual PAMs were counted, and the PAM counts were normalized to the total reads for each sample. For a given PAM sequence, the enrichment was measured as the $\log_2$ ratio as compared to a no-protein control, with a 0.01 pseudocount adjustment. PAMs above an enrichment threshold set to 3.5 were compiled and used to generate sequence logos (Crooks et al., 2004). For the PAM wheel generation, abundances were used to generate wheels with Krona (Ondov et al., 2011), as described in the previous report (Leenay et al., 2016). The ratios of PAM abundances as compared to a no-protein control, with a 0.01 pseudocount adjustment, were used directly as input for Krona.

***In vitro* cleavage assay**

In vitro plasmid DNA cleavage experiments were performed essentially as described previously (Nishimasu et al., 2015). The *Eco*RI-linearized pUC119 plasmid (100 ng, 4.7 nM), containing the 20-nt target sequence and the PAMs, was incubated at 37°C for 5 min with the CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5), in 10 μl of reaction buffer containing 20 mM HEPES, pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 1 mM DTT and 5% glycerol. Reaction products were resolved on a 1% agarose gel, stained with ethidium bromide, and then visualized using a Typhoon FLA 9500 imager (GE Healthcare).

METHOD DETAILS QUANTIFICATION AND STATISTICAL ANALYSES

In vitro cleavage experiments were performed at least three times, and representative results were shown.

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates of the CjCas9-sgRNA-DNA complexes have been deposited in the Protein Data Bank, with the accession numbers PDB: 5X2G (5′-AGAAACC-3′ PAM) and 5X2H (5′-AGAAACA-3′ PAM). Data of *in vitro* cleavage experiments have been deposited in the Mendeley Data repository (http://dx.doi.org/doi:10.17632/6v2dwwcgs3.1). The CueMol program is available at http://www.cuemol.org.

Figure Legends

Figure 1 CjCas9 PAM specificity

(A) Motif obtained from the *in vitro* PAM discovery assay.

(B and C) *In vitro* cleavage assays for DNA targets with different PAMs. The linearized plasmid targets with either the 5′-AGAANCA-3′, 5′-AGAAANA-3′ or 5′-AGAACN-3′ PAM (B), or the 5′-AGANACC-3′ PAM (C) were incubated with CjCas9-sgRNA, and then analyzed by agarose gel electrophoresis.

See also Figure S1.

Figure 2 Overall structure

(A) Domain structure of CjCas9. The HNH nuclease domain was truncated for crystallization. BH, bridge helix; PLL, phosphate lock loop.

(B) Schematics of the sgRNA and the target DNA. TS, target strand; NTS, non-target strand.

(C) Overall structure of CjCas9-ΔHNH in complex with an sgRNA and its target DNA. The predicted location of the HNH domain is indicated by the pink circle.

(D) *In vitro* cleavage activity of CjCas9-ΔHNH. The circular and linearized plasmid targets with the 5′-AGAAACC-3′ PAM were incubated with wild-type CjCas9 or the three CjCas9 variants (D8A, H559A and ΔHNH), and then were analyzed by agarose gel electrophoresis. The D8A and H559A variants of CjCas9 correspond to the D10A and H840A nickases of SpCas9, respectively.

See also Figure S2 and Table 1.

Figure 3 TracrRNA scaffold

(A) $2mF_O - DF_C$ electron density map for the tracrRNA scaffold (contoured at 2σ).

(B) Schematics of the tracrRNA scaffold. Non-Watson-Crick base pairs are indicated with Leontis-Westhof notations (Leontis et al., 2002). The base triples and quintuples are highlighted in gray and orange backgrounds, respectively.

(C) Structure of the CjCas9 tracrRNA scaffold (stereo view). The 3′ nucleotides involved in the triple-helix formation are colored blue. Nucleotides involved in the formation of the base triple and quintuple are highlighted in gray and orange backgrounds, respectively.

(D–G) Base triples (D and E) and base quintuples (F and G) in the triple helix. Hydrogen bonds between canonical and non-canonical base pairs are depicted with green and gray dashed lines, respectively.

Figure 4 TracrRNA scaffold recognition

(A) Binding of the tracrRNA scaffold to CjCas9 (stereo view).

(B and C) Specific recognition of A63 (B) and A76/U80 (C). Hydrogen bonds are depicted by dashed lines.

(D) Functional importance of the tracrRNA scaffold. The linearized plasmid target with the 5′-AGAAACC-3′ PAM was incubated with CjCas9, together with either the full-length sgRNA (nucleotides 1–93), the sgRNA variants, or the SpCas9 sgRNA. 1–93, the full-length sgRNA; GGCG, the sgRNA variant, in which nucleotides 90–93 (CCGC) were replaced with GGCG; 1–85, the sgRNA variant, in which nucleotides 86–93 were truncated; Sp, the SpCas9 sgRNA.

Figure 5 PAM recognition

(A) Binding of the PAM duplex to CjCas9.

(B) Schematics of the PAM recognition by CjCas9. Hydrogen bonds are depicted by dashed lines. Water molecules are shown as red spheres. Water-mediated hydrogen bonds between the protein and the sugar-phosphate backbone are omitted for clarity.

(C) Recognition of the 5′-AGAAACC-3′ PAM (stereo view). Water molecules are shown as red spheres. Hydrogen bonds are depicted by dashed lines.

(D) Functional importance of the PAM-interacting residues. The linearized plasmid target with the 5′-AGAAACC-3′ PAM was incubated with either the wild type or the mutants of CjCas9, together with the sgRNA.

See also Figure S3.

Figure 6 Structural comparison of the Cas9 orthologs

(A–D) Structures of CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D) in complexes with their cognate sgRNAs and target DNAs. SpCas9 and FnCas9 have structurally distinct subdomains inserted within their REC1 domains (previously referred to as the REC2 domains (Nishimasu et al., 2014; Hirano et al., 2016)).

(E–H) Structures of the sgRNAs for CjCas9 (E), SpCas9 (PDB: 4OO8) (F), SaCas9 (PDB: 5CZZ) (G), and FnCas9 (PDB: 5B2O) (H). The guide segments are omitted for clarity.

See also Figures S4–S7.

Figure 7 PAM recognition by the Cas9 orthologs

(A–D) PAM recognition by CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D). In (B), the subdomain inserted between β6 and β7 is omitted for clarity. The PAMs are highlighted in purple. The conserved core β-strands are numbered. Close-up views for the PAM recognition are shown in right panels.

Table 1. Data collection and refinement statistics

	5′-AGAAACC-3′ PAM (Native)	5′-AGAAACA-3′ PAM (Native)	5′-AGAAACA-3′ PAM (SeMet)
Data collection			
Beamline	SLS PX1	SPring-8 BL41XU	SPring-8 BL41XU
Wavelength (Å)	1.0000	1.0000	0.9791
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	104.3, 105.1, 136.5	102.0, 103.9, 134.5	102.0, 104.2, 134.6
α, *β*, *γ* (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)*	136.5–2.4 (2.47–2.40)	103.9–2.3 (2.36–2.30)	49.5–2.2 (2.25–2.20)
R_{merge}	0.150 (0.882)	0.105 (0.783)	0.075 (0.967)
R_{pim}	0.043 (0.529)	0.032 (0.316)	0.031 (0.396)
$I/\sigma I$	13.8 (1.6)	12.4 (2.4)	14.2 (2.6)
Completeness (%)	100 (100)	99.8 (99.8)	100 (100)
Multiplicity	12.9 (13.4)	10.4 (7.0)	6.8 (6.9)
CC (1/2)	0.998 (0.757)	0.986 (0.875)	0.999 (0.866)
Refinement			
Resolution (Å)	52.6–2.4	72.8–2.3	
No. reflections	59,303	63,923	
R_{work} / R_{free}	0.190 / 0.221	0.200 / 0.231	
No. atoms			
Protein	5,929	5,925	
Nucleic acid	2,715	2,715	
Solvent	323	279	
B-factors (Å^2)			
Protein	59	65	
Nucleic acid	49	54	
Solvent	40	45	
R.m.s. deviations			
Bond lengths (Å)	0.0023	0.0024	
Bond angles (°)	0.498	0.515	
Ramachandran plot (%)			
Favored region	97.28	96.58	
Allowed region	2.58	3.01	
Outlier region	0.14	0.41	

*Values in parentheses are for the highest resolution shell.

1

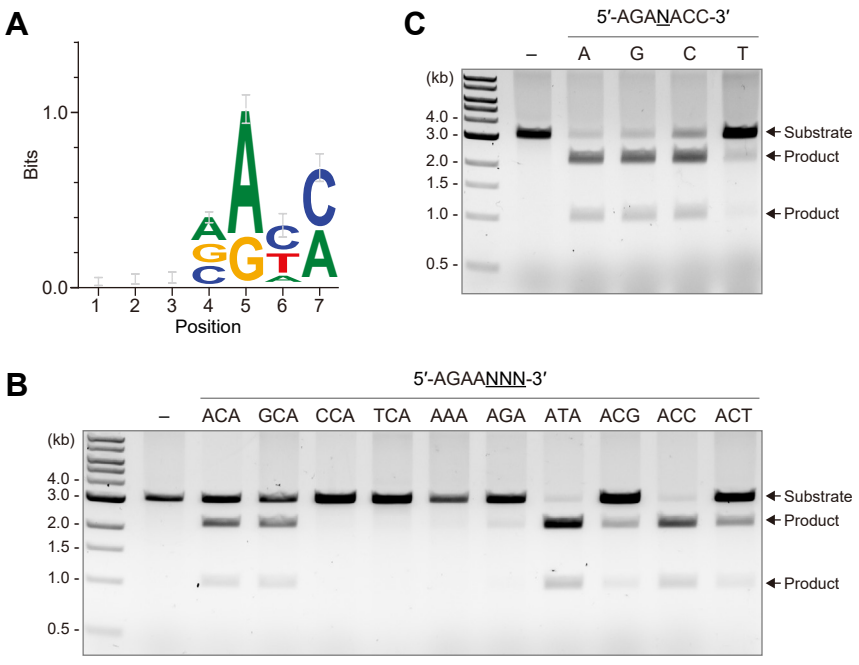
A
1.0
0.0
Bits
1 2 3 4 5 6 7
Position
C
5′-AGANACC-3′
– A G C T
(kb)
4.0 -
3.0 -
2.0 -
1.5 -
1.0 -
0.5 -
Substrate
Product
Product
B
5′-AGAANNN-3′
– ACA GCA CCA TCA AAA AGA ATA ACG ACC ACT
(kb)
4.0 -
3.0 -
2.0 -
1.5 -
1.0 -
0.5 -
Substrate
Product
Product

Figure 1

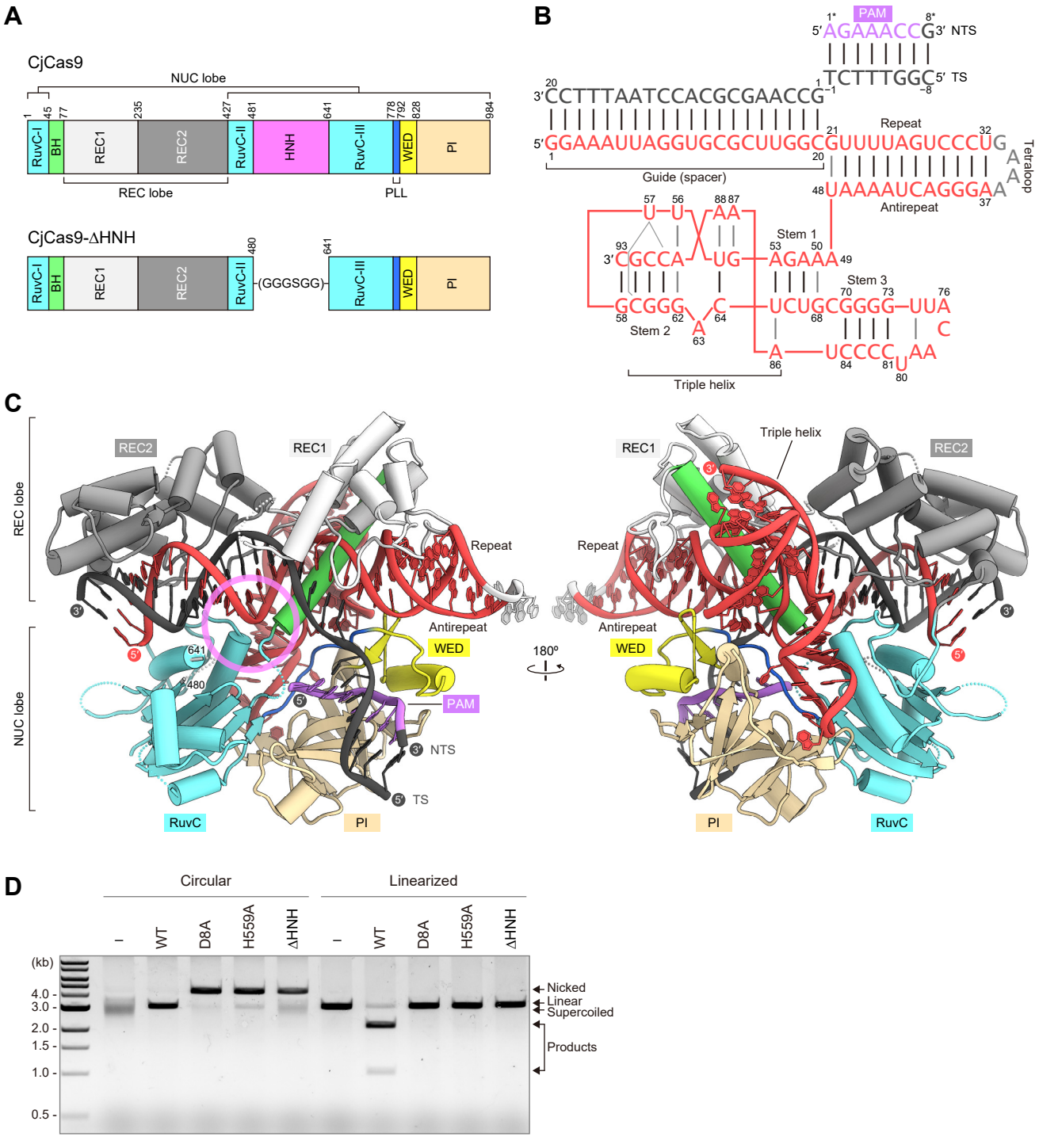
A
CjCas9
NUC lobe
RuvC-I
BH
REC1
REC2
RuvC-II
HNH
RuvC-III
WED
PI
REC lobe
PLL
CjCas9-ΔHNH
(GGGSGG)
B
PAM
NTS
TS
Repeat
Guide (spacer)
Antirepeat
Tetraloop
Stem 1
Stem 2
Stem 3
Triple helix
C
REC lobe
NUC lobe
REC2
REC1
Repeat
Antirepeat
WED
PAM
NTS
TS
RuvC
PI
180°
Triple helix
D
Circular
Linearized
–
WT
D8A
H559A
ΔHNH
(kb)
4.0
3.0
2.0
1.5
1.0
0.5
Nicked
Linear
Supercoiled
Products

Figure 2

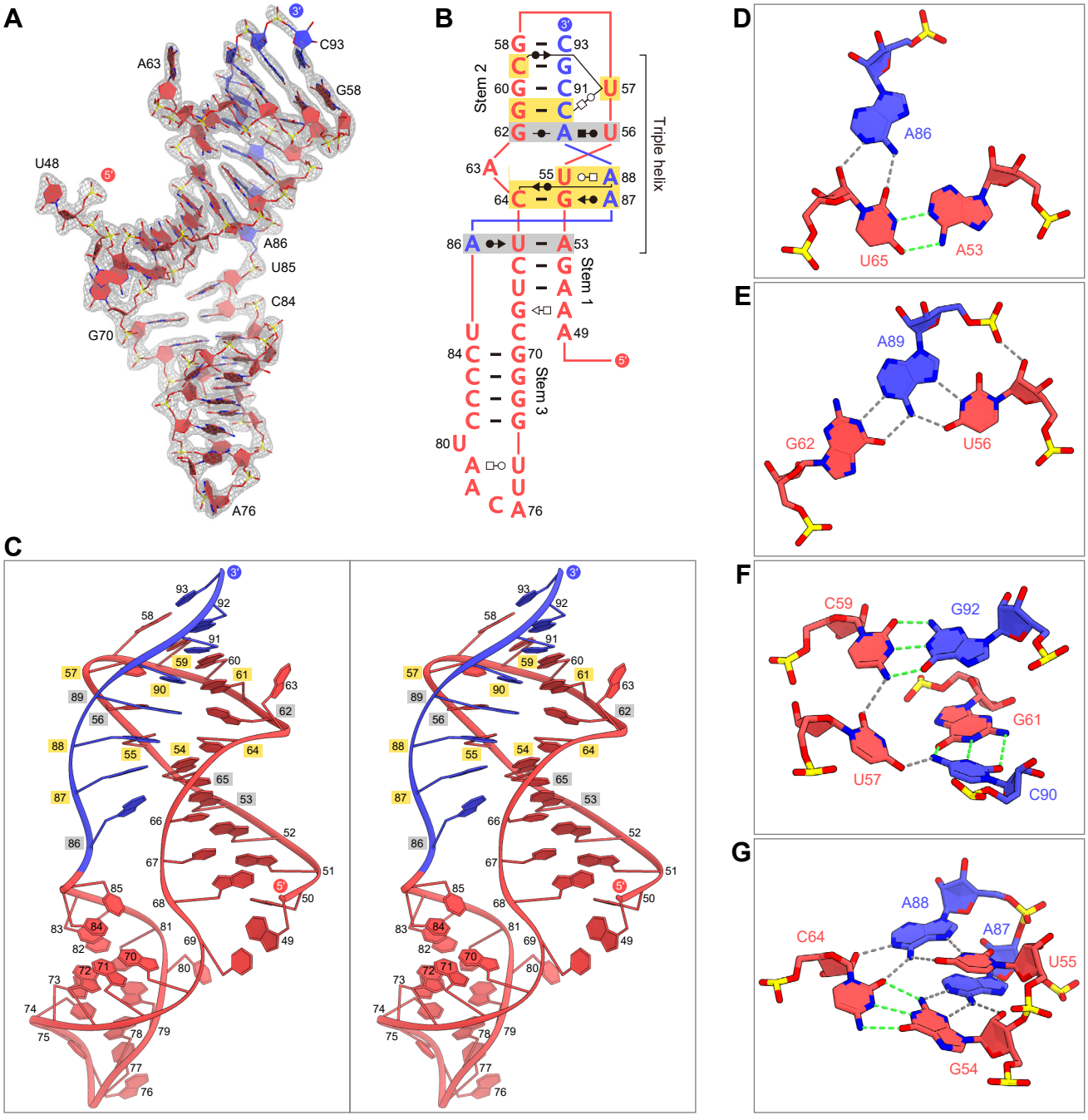
A
C93
A63
G58
U48
A86
U85
C84
G70
A76
B
Stem 2
Triple helix
Stem 1
Stem 3
C
D
A86
U65
A53
E
A89
G62
U56
F
C59
G92
G61
U57
C90
G
A88
C64
A87
U55
G54

Figure 3

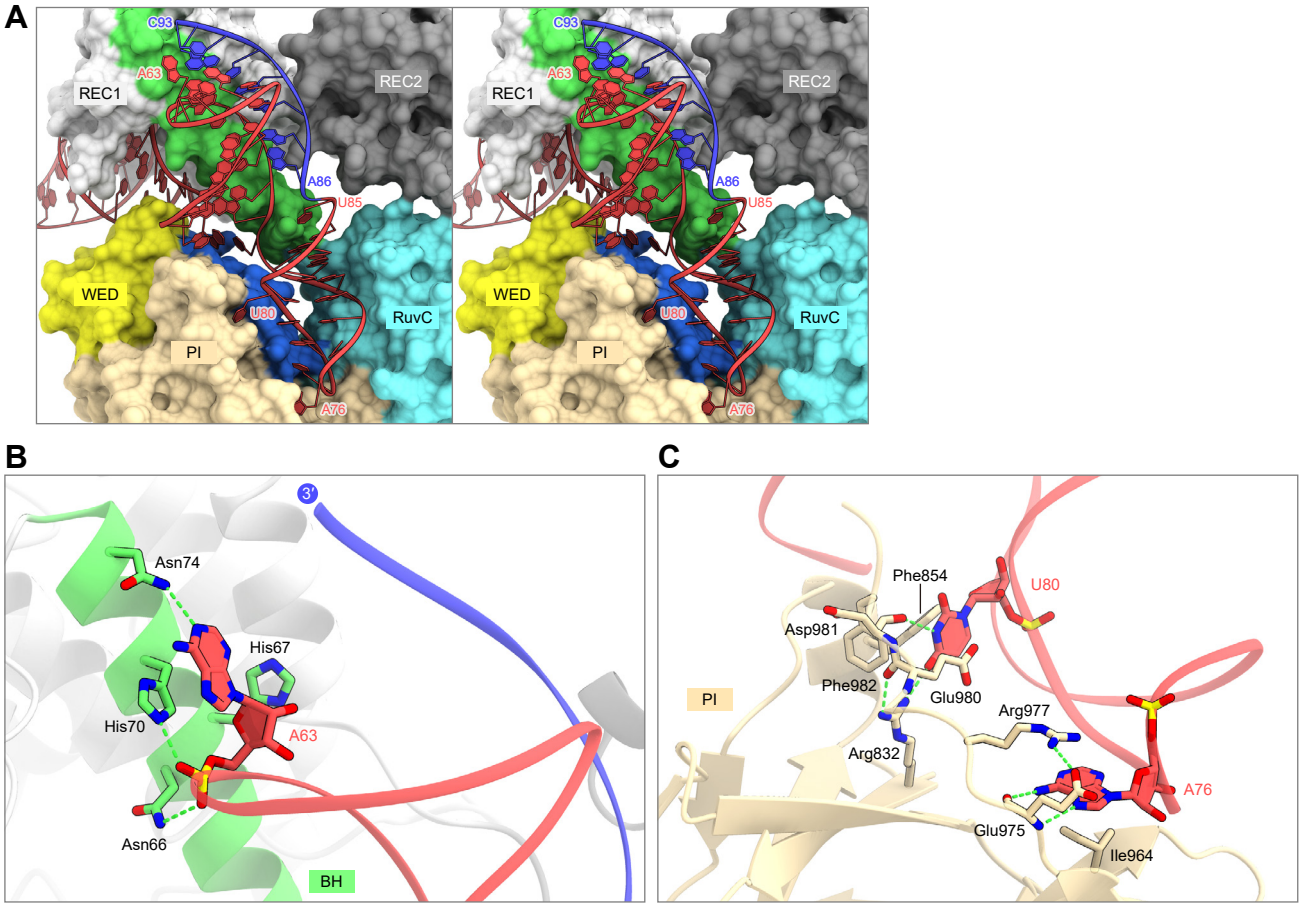
A
C93
A63
REC1
REC2
A86
U85
WED
U80
RuvC
PI
A76
B
3'
Asn74
His67
His70
A63
Asn66
BH
C
Phe854
U80
Asp981
PI
Phe982
Glu980
Arg977
Arg832
A76
Glu975
Ile964
D
–
1–93
A63U
A76U
U80A
GGCG
1–85
Sp
(kb)
4.0 -
3.0 -
2.0 -
1.5 -
1.0 -
0.5 -
Substrate
Product
Product

Figure 4

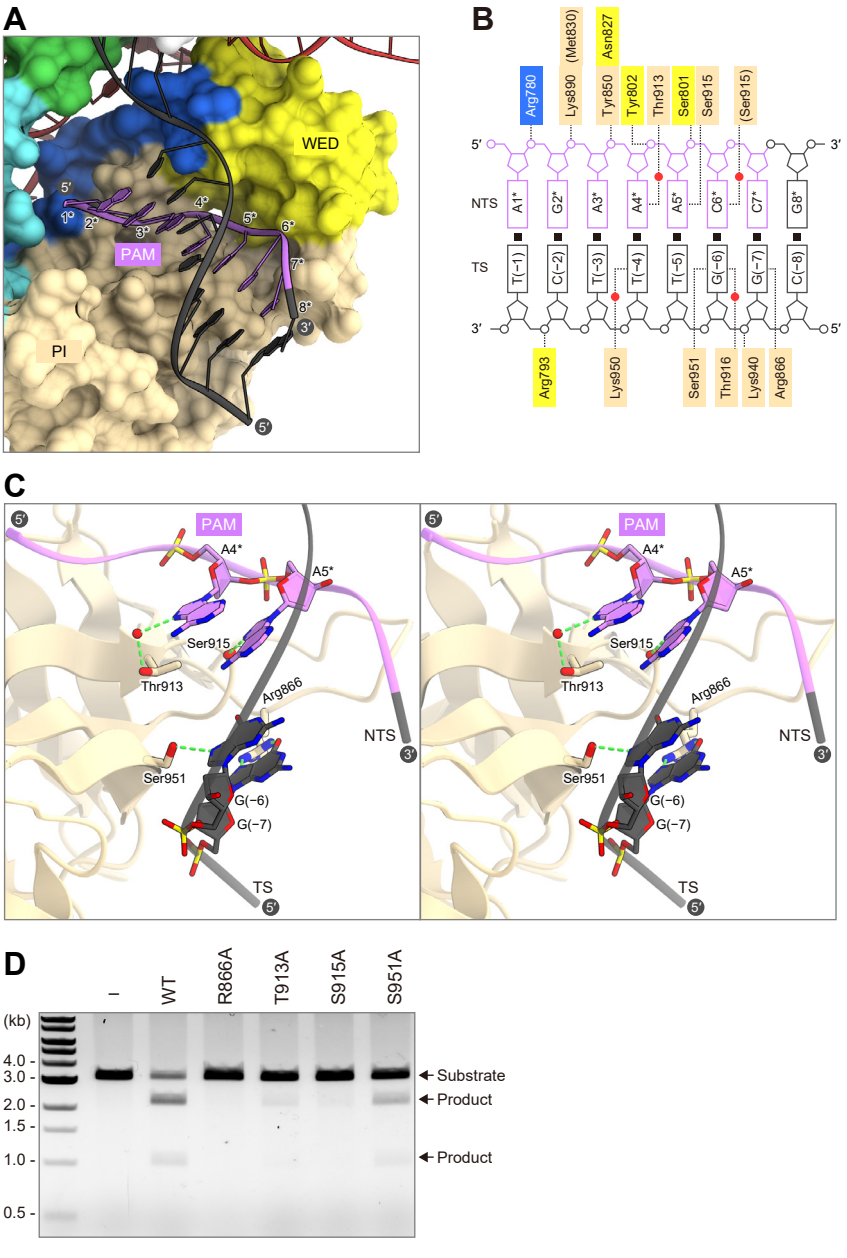
A
WED
PAM
PI
5′
1*
2*
3*
4*
5*
6*
7*
8*
3′
5′
B
Arg780
Lys890
(Met830)
Asn827
Tyr850
Tyr802
Thr913
Ser801
Ser915
(Ser915)
5′
3′
NTS
A1*
G2*
A3*
A4*
A5*
C6*
C7*
G8*
TS
T(−1)
C(−2)
T(−3)
T(−4)
T(−5)
G(−6)
G(−7)
C(−8)
3′
5′
Arg793
Lys950
Ser951
Thr916
Lys940
Arg866
C
PAM
A4*
A5*
Ser915
Thr913
Arg866
NTS
Ser951
G(−6)
G(−7)
TS
D
–
WT
R866A
T913A
S915A
S951A
(kb)
4.0
3.0
2.0
1.5
1.0
0.5
Substrate
Product
Product

Figure 5

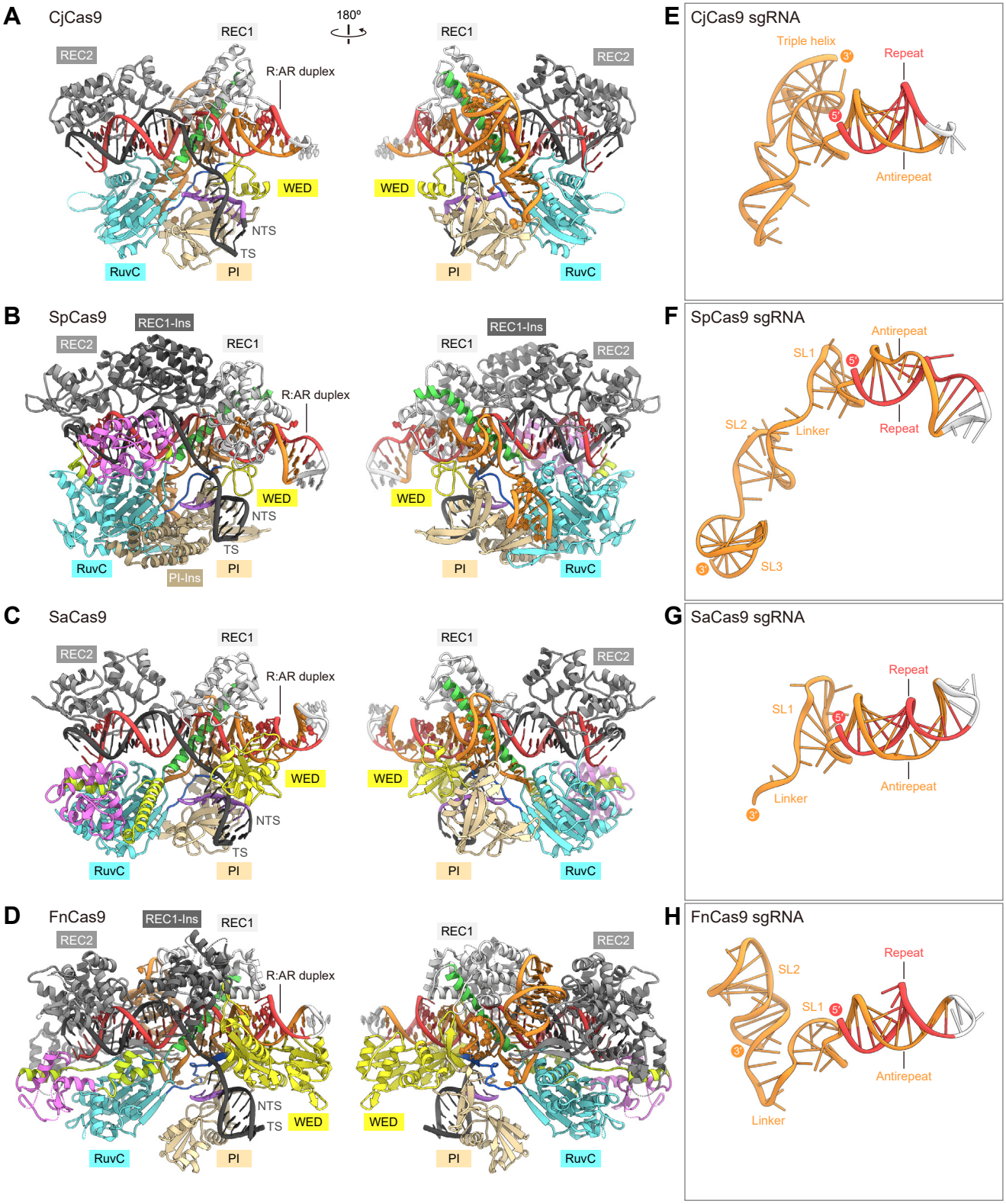
A
CjCas9
180°
REC1
REC2
R:AR duplex
WED
NTS
TS
RuvC
PI
B
SpCas9
REC1-Ins
PI-Ins
C
SaCas9
D
FnCas9
E
CjCas9 sgRNA
Triple helix
Repeat
Antirepeat
F
SpCas9 sgRNA
SL1
SL2
Linker
SL3
G
SaCas9 sgRNA
H
FnCas9 sgRNA
5'
3'

Figure 6

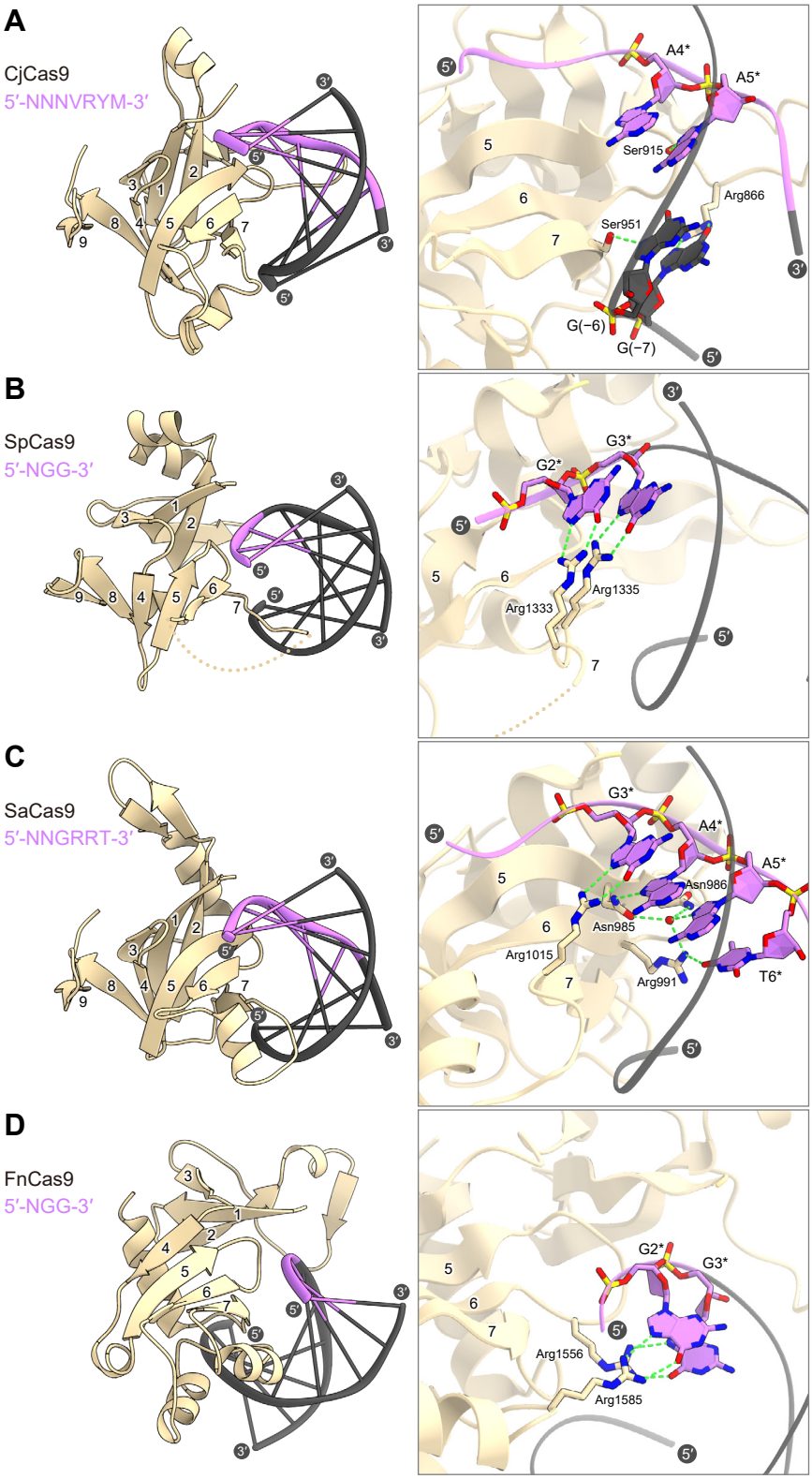
A
CjCas9
5′-NNNVRYM-3′
A4*
A5*
Ser915
Arg866
Ser951
G(−6)
G(−7)
B
SpCas9
5′-NGG-3′
G3*
G2*
Arg1335
Arg1333
C
SaCas9
5′-NNGRRT-3′
G3*
A4*
A5*
Asn986
Asn985
Arg1015
Arg991
T6*
D
FnCas9
5′-NGG-3′
G2*
G3*
Arg1556
Arg1585

Figure 7

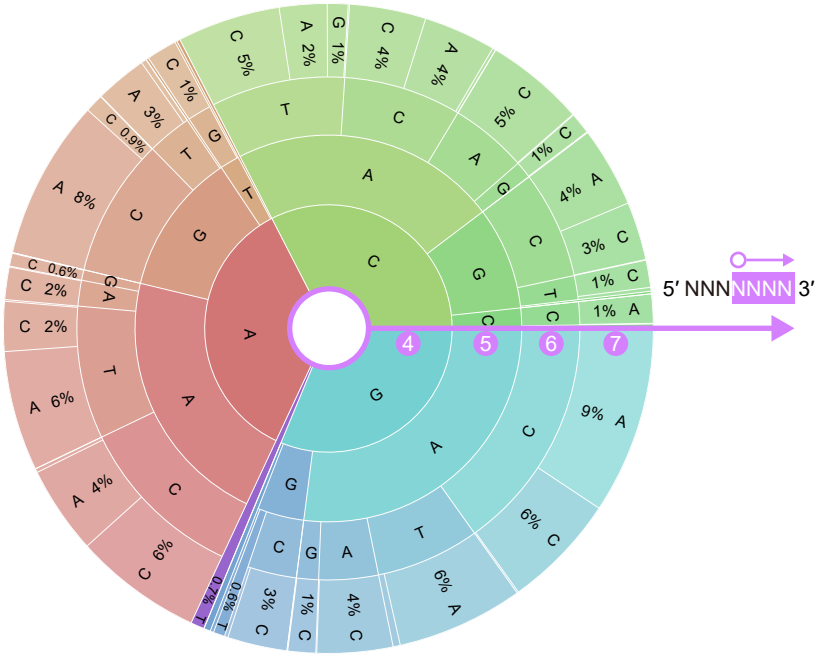

Figure S1. CjCas9 PAM specificity

The results of the *in vitro* PAM discovery assay are represented as the PAM wheel (Leenay et al., 2016). The area of the sector indicates the relative enrichment in the library.

Related to Figure 1.

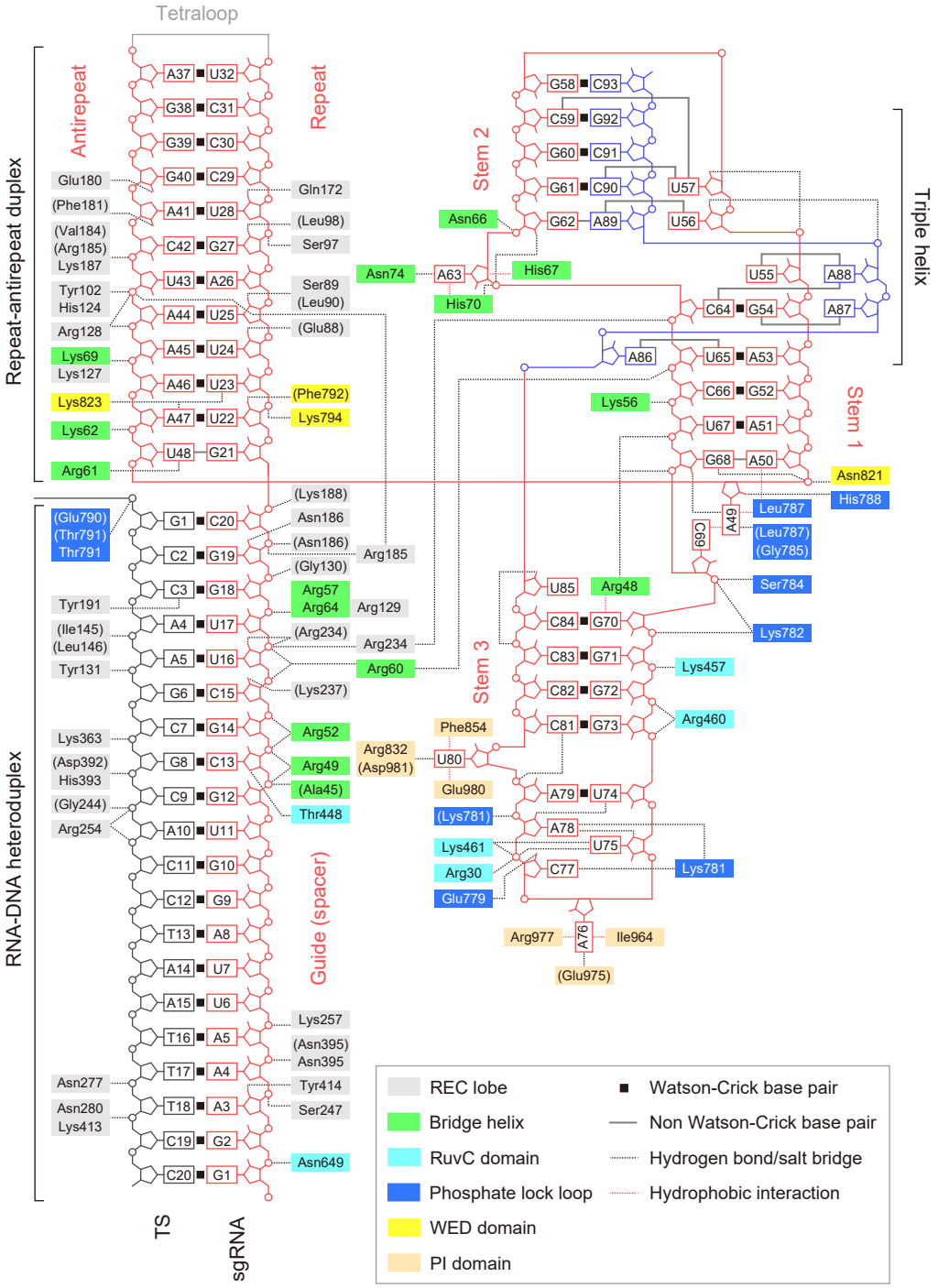

Figure S2. Schematic of the nucleic-acid recognition
Residues that interact with nucleic acids via their main chain are shown in parentheses. Water-mediated hydrogen bonds are omitted for clarity. Related to Figure 2.

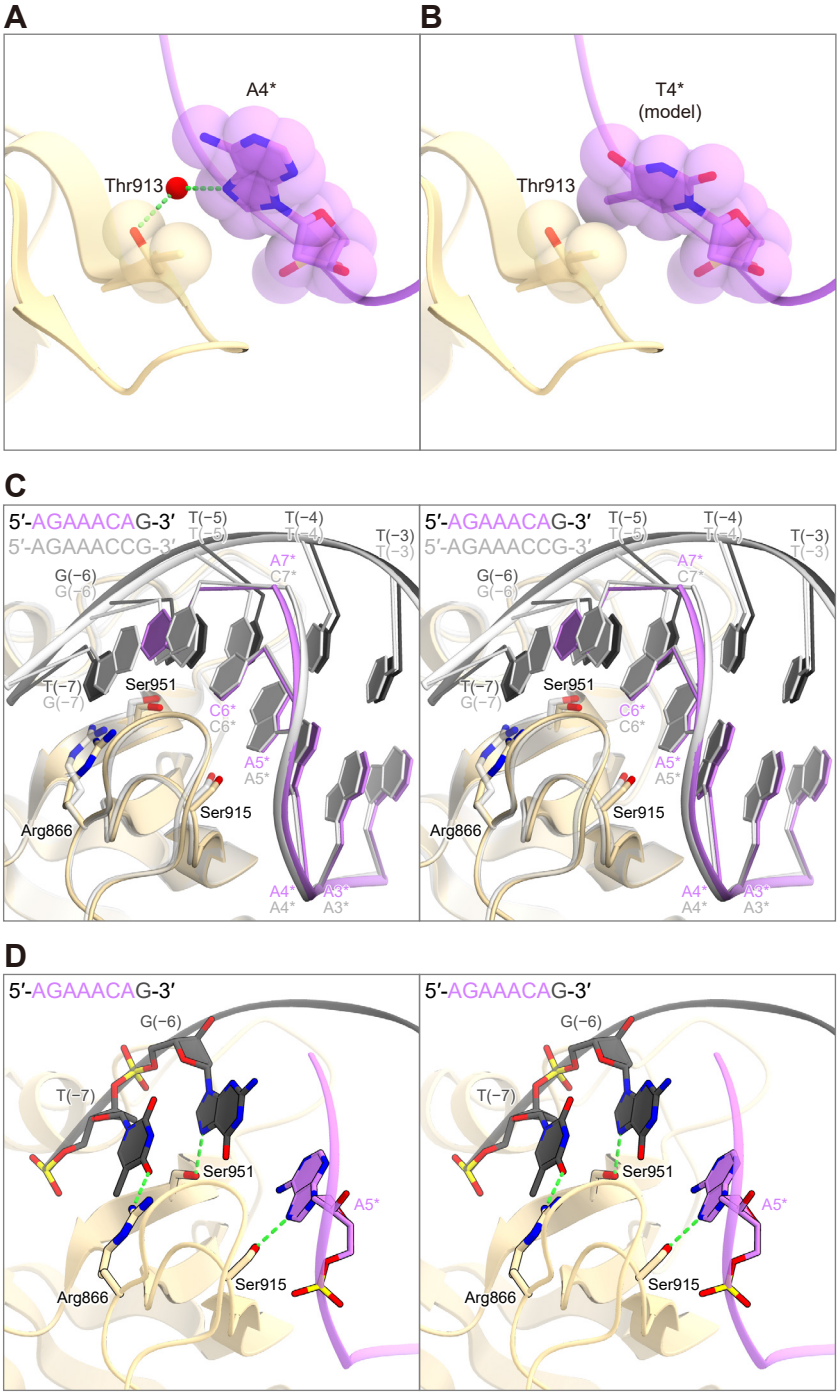

Figure S3. PAM recognition

(A and B) Interactions between Thr913 and dA4* (A) and modeled dT4* (B).

(C) Superimposition of the 5′-AGAAACC-3′ PAM complex (semitransparent gray) and the 5′-AGAAACA-3′ PAM complex (colored) (stereo view).

(D) Recognition of the 5′-AGAAACA-3′ PAM (stereo view). Hydrogen bonds are depicted by dashed lines.

Related to Figure 5.

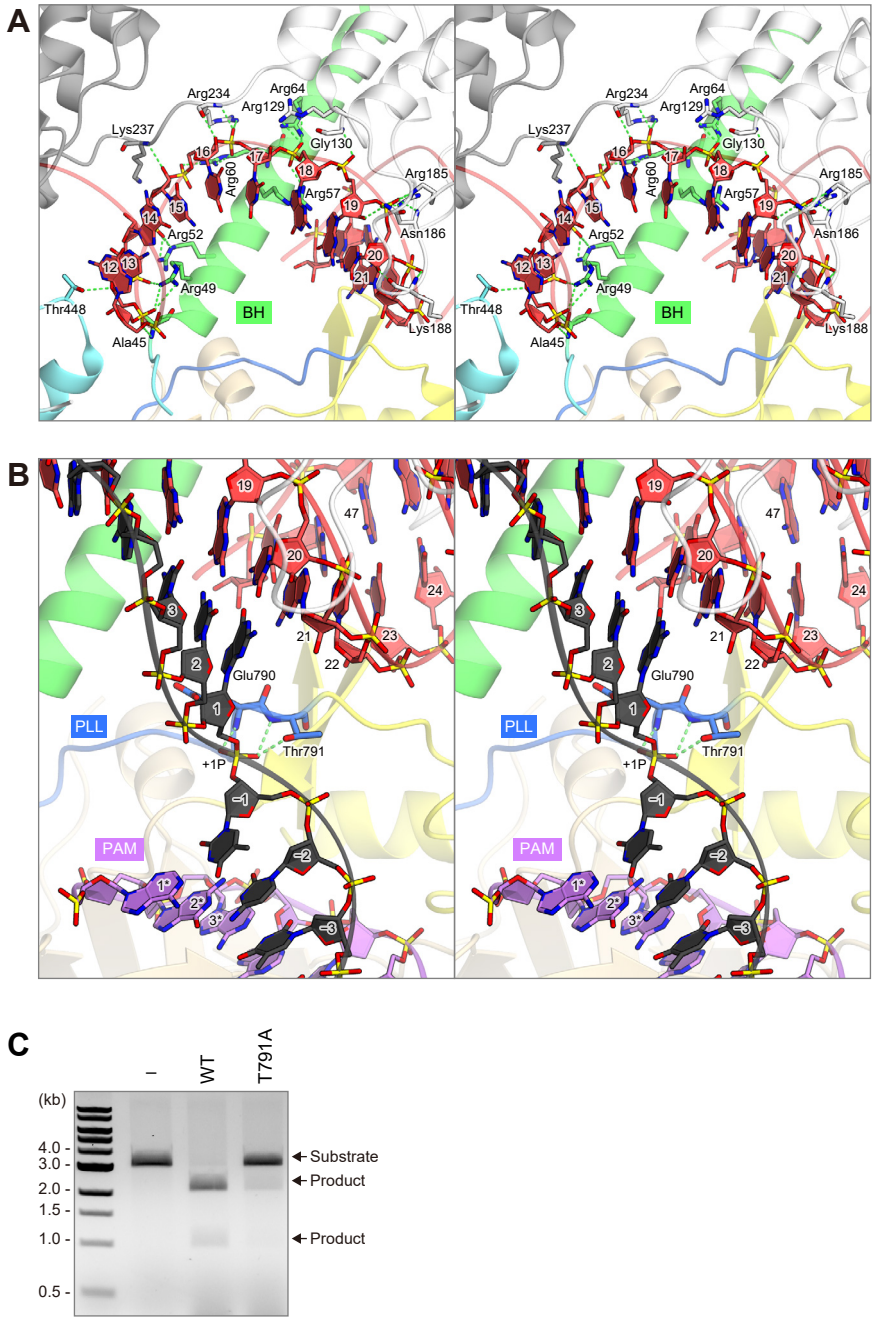

Figure S4. RNA-guided DNA targeting

(A) Recognition of the sgRNA seed region by CjCas9 (stereo view). The target DNA strand is omitted for clarity.

(B) Recognition of the +1 phosphate by CjCas9 (stereo view).

(C) Functional importance of the phosphate lock loop. The linearized plasmid target with the 5′-AGAAACC-3′ PAM was incubated with either the wild type or the T791A mutant of CjCas9, together with the sgRNA.

Related to Figure 6.

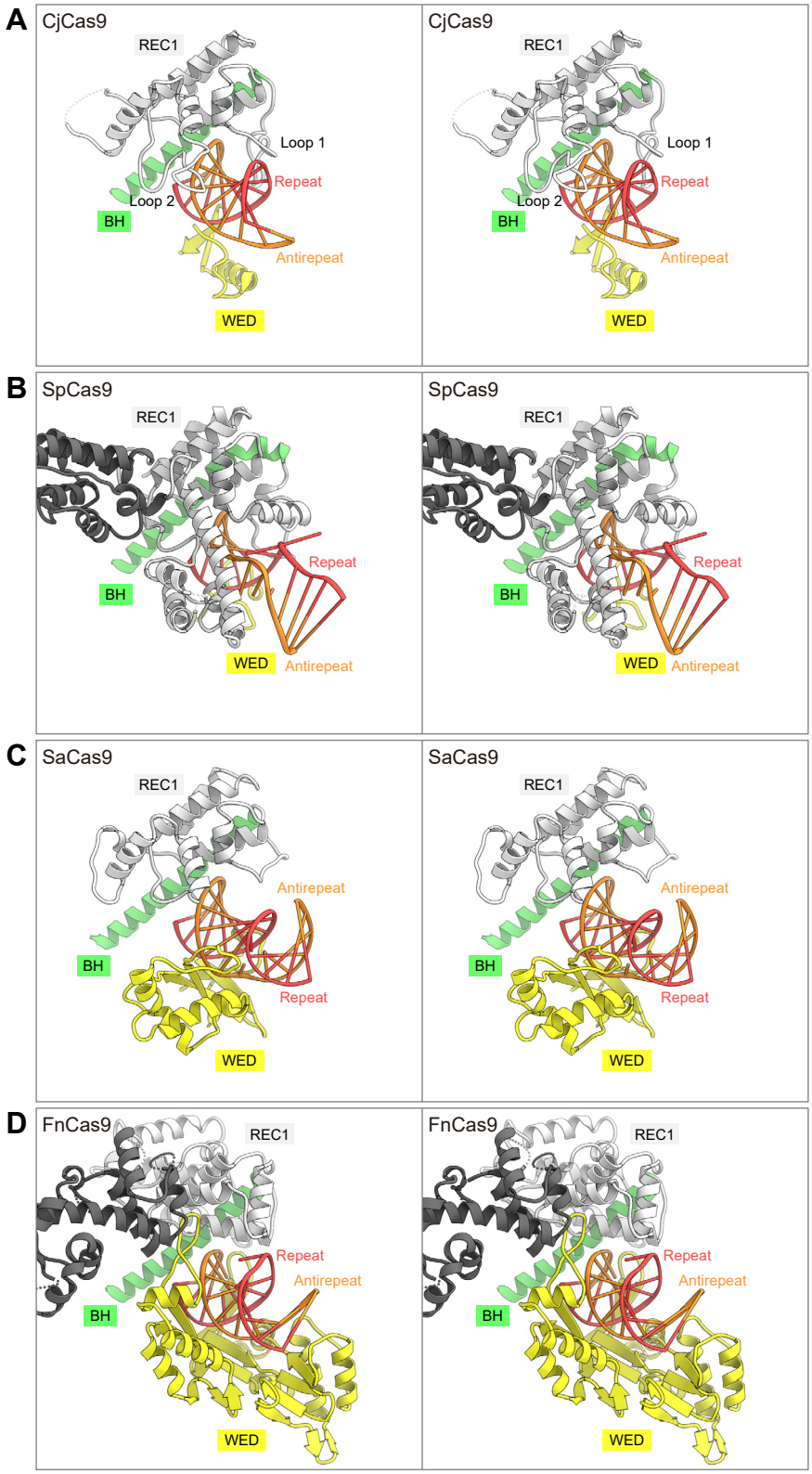

Figure S5. Recognition of the repeat-antirepeat duplex by the Cas9 orthologs

(A–D) Species-specific recognition of the repeat-antirepeat duplex by the REC1/WED domains of CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D).

Related to Figure 6.

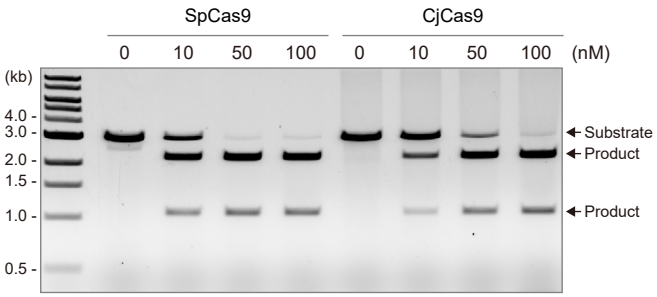

Figure S6. Comparison of the DNA cleavage activities of CjCas9 and SpCas9

The linearized plasmid targets with the 5′-TGG-3′ PAM and the 5′-AGAAACC-3′ PAM were incubated at 37°C for 5 min with SpCas9 (10–100 nM) and CjCas9 (10–100 nM) in the presence of their cognate sgRNAs, respectively. Related to Figure 6.

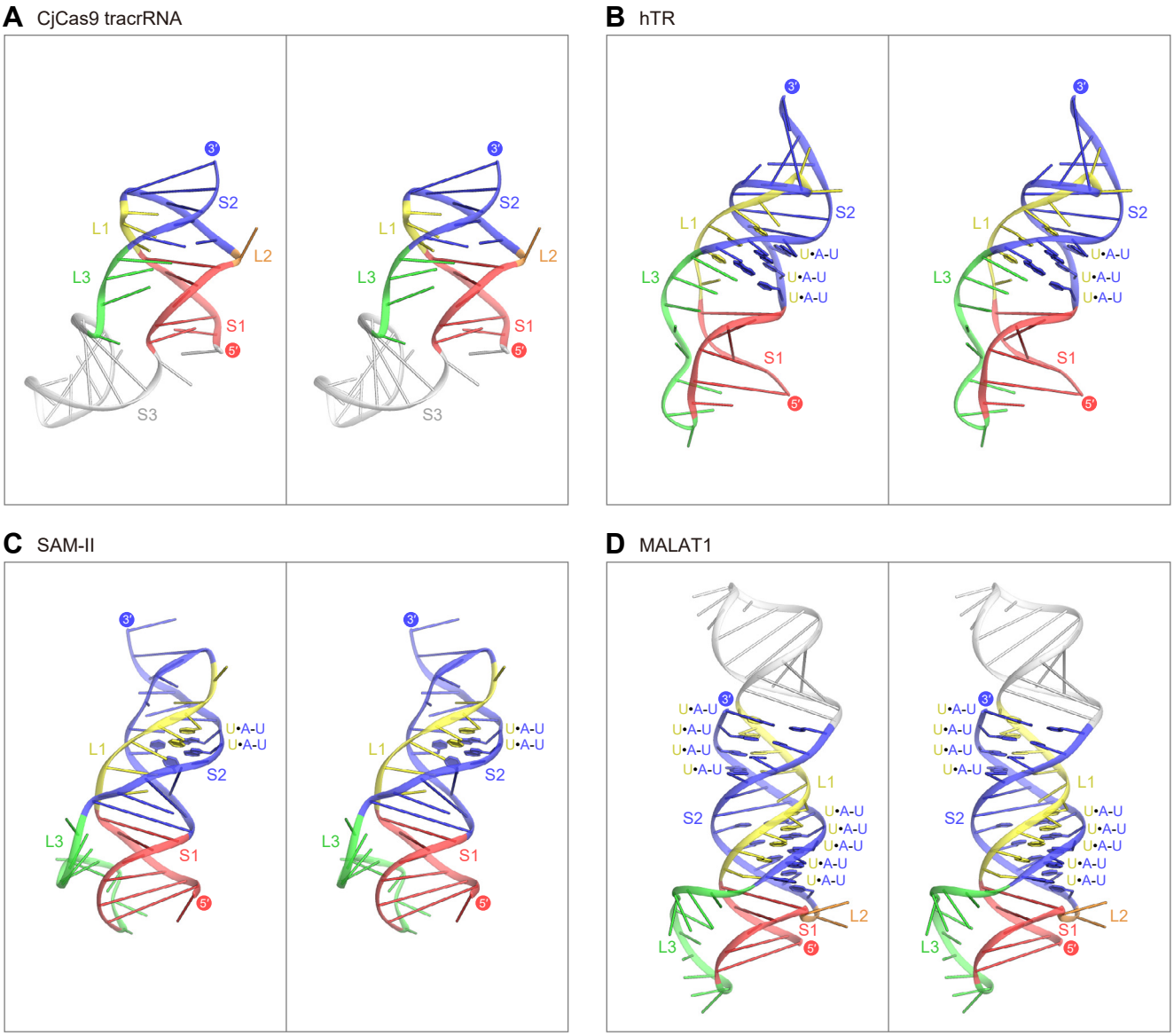

Figure S7. Structures of the RNA triple helices

(A–D) Structures of the CjCas9 tracrRNA (A), the telomerase RNA subunit TER (PDB: 1YMO) (B), the SAM-II riboswitch (PDB: 2QWY) (C), and the long noncoding RNA MALAT1 (PDB: 4PLX) (D) (stereo view). The U•A-U triples are depicted as stick models. S, stem; L, loop. Related to Figure 6.

Table S1 Oligonucleotides, related to the STAR Methods

Oligonucleotides used to generate the CjCas9 expression vector (pE-SUMO-CjCas9, with the TEV recognition site between His_6-SUMO and CjCas9)		
PCR template	Forward primer	Reverse primer
pE-SUMO (modified)	ACTCGAGCACCACCACCACC	CATATGGGATCCTTGGAAGTACAAG
pColdGST-CjCas9	CAAGGATCCCATATGGCCCGCATCCTCGCTTTC	TGGTGGTGCTCGAGTTTATTTCTTAAAATCCTCGCGCTG
Oligonucleotides used to introduce the CjCas9 mutations (pE-SUMO-CjCas9)		
Mutation	Forward primer	Reverse primer
ΔHNH	TACATTGCGCGACTGGTTC	CCCCCCGGACCCCCCCCCCAATTCGATATTTATTTTGTGCAC
D8A	GCGATCGGAATCTCTAGTATCGGATG	GAAAGCGAGGATGCGGGC
H559A	GCGATCTATCCGTATAGCAGGTCATTTGAC	GTCGATCTCCAGCATTTTCTCGTC
T791A	GAAGAGGCATTCCGCAAGGAAGAGG	GCGGAATGCCTCTTCGTGCAGTGC
R866A	GTGGCTGCGTCGAAGAAAGGAGAGATC	TCTTCGACGCAGCCACCGCCTTATTTGG
T913A	AGCAGTACAGTGTCCCTGATTGTGAGCAAGC	CGCGAAGGCGTTATAATAAACAAATTCTGGTTCC
S915A	ACAGTGTCCCTGATTGTGAGCAAGC	CGCGCTCGTGAAGGCGTTATAATAAACAAATTCTG
S951A	GCAAAGGCAATTGGCATCCAAAACCTG	GCCAATTGCCTTTGCGATGACCTCCTT
DNA oligonucleotides used for crystallization		
PAM sequence	Target DNA strand	Non-target DNA strand
AGAAACA	CTGTTTCTGCCAAGCGCACCTAATTTCC	AGAAACAG
AGAAACC	CGGTTTCTGCCAAGCGCACCTAATTTCC	AGAAACCG
sgRNA		
Cj sgRNA 1-93	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGC	
Cj sgRNA A63U	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGTCTCTGCGGGGTTACAATCCCCTAAAACCGC	
Cj sgRNA A76U	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTTCAATCCCCTAAAACCGC	
Cj sgRNA U80A	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAAACCCCTAAAACCGC	
Cj sgRNA GGCG	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAAGGCG	
Cj sgRNA 1-85	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCT	
Sp sgRNA	GGAAAUUAGGUGCGCUUGGCGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUU	
Oligonucleotides used to introduce the sgRNA mutations (pUC119-sgRNA)		
Mutation	Forward primer	Reverse primer
A63U	TCTCTGCGGGGTTACAATCCCCTAAAACCGC	CCCGCAAACTCTTTATTTTAGTCCCTTTTC
A76U	TCAATCCCCTAAAACCGCTTGTACCGAGCTCG	AACCCCGCAGAGTCCCGCAAACTCTTTATTTTAG
U80A	ACCCCTAAAACCGCTTGTACCGAGCTCGAATTCAC	TTGTAACCCCGCAGAGTCCCGCAAACTC
GGCG	GGCGTTGTACCGAGCTCGAATTCACTGGCCGTCG	TTTTAGGGGATTGTAACCCCGCAGAGTCCCGC
Oligonucleotides used to generate the dsDNA templates for *in vitro* transcription		
sgRNA	Forward primer	Reverse primer
1-93, A63U, A76U	AGCGCCCAATACGCAAAC	GCGGTTTTAGGGGATTG
U80A	AGCGCCCAATACGCAAAC	GCGGTTTTAGGGGTTTG
GGCG	AGCGCCCAATACGCAAAC	CGCCTTTTAGGGGATTG
1-85	AGCGCCCAATACGCAAAC	AGGGGATTGTAACCCCGCAGAGTCCC

TABLE FOR AUTHOR TO COMPLETE

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
CjCas9	This paper	N/A
CjCas9-ΔHNH	This paper	N/A
CjCas9, various mutants	This paper	N/A
SpCas9	Nishimasu *et al.*, 2014	N/A
T7 RNA polymerase	Home made	N/A
Ammonium acetate - 1.0 M solution	Hampton Research	HR2-565
NeXtalStock 50% (w/v) PEG 2,000	QIAGEN	N/A
Critical Commercial Assays		
Zymoclean Gel DNA Recovery Kit	Zymo Research	D4007
MiSeq Reagent Kit v3 (150-cycle)	Illumina	MS-102-3001
QIAprep Spin Miniprep Kit	QIAGEN	27106
Deposited Data		
Atomic coordinates, CjCas9 (AGAAACC PAM)	This paper	PDB: 5X2G
Atomic coordinates, CjCas9 (AGAAACA PAM)	This paper	PDB: 5X2H
In vitro DNA cleavage experiments	This paper; Mendeley Data	http://dx.doi.org/doi:10.17632/6v2dwwcgs3.1
Experimental Models: Organisms/Strain		
E. coli Mach	Thermo Fisher Scientific	C862003
E. coli B834 (DE3)	Novagen	69041
E. coli Rosetta 2 (DE3)	Novagen	71397
Recombinant DNA		
pE-SUMO-CjCas9	This paper	N/A
pE-SUMO-CjCas9-ΔHNH	This paper	N/A
pE-SUMO-CjCas9, various mutants	This paper	N/A
pUC119-Cj sgRNA	This paper	N/A
pUC119-Cj sgRNA, various mutants	This paper	N/A
pUC119-T20, various PAMs	This paper	N/A
Sequence-Based Reagents		
DNA primers	This paper	Table S1
DNA oligos (for crystallization)	This paper	Table S1
CjCas9 sgRNA	This paper	Table S1
CjCas9 sgRNA, various mutants	This paper	Table S1
SpCas9 sgRNA	Nishimasu *et al.*, 2014	Table S1
Software and Algorithms		
DIALS	Waterman *et al.*, 2013	http://dials.lbl.gov
AIMLESS	Evans and Murshudov, 2013	http://www.ccp4.ac.uk/html/aimless.html
Buccaneer	Cowtan, 2006	http://www.ysbl.york.ac.uk/~cowtan/buccaneer/buccaneer.html

COOT	Emsley and Cowtan, 2004	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
PHENIX	Adams *et al.*, 2010	https://www.phenix-online.org
CueMol	N/A	http://www.cuemol.org
Krona	Ondov *et al.*, 2011	https://github.com/marbl/Krona/wiki
WebLogo	Crooks *et al.*, 2004	http://weblogo.threeplusone.com/
Other		
Amicon Ultra-4 Centrifugal Filter Units - 10,000 NMWL	Millipore	UFC801024
Ni-NTA Superflow	QIAGEN	30450
HiTrap Heparin HP	GE Healthcare	17040701
HiLoad 16/600 Superdex 200 pg	GE Healthcare	28989335
Superdex 200 Increase 10/300 GL	GE Healthcare	28990944
Whatman Elutrap electroelution system	Sigma-Aldrich	WHA10447705