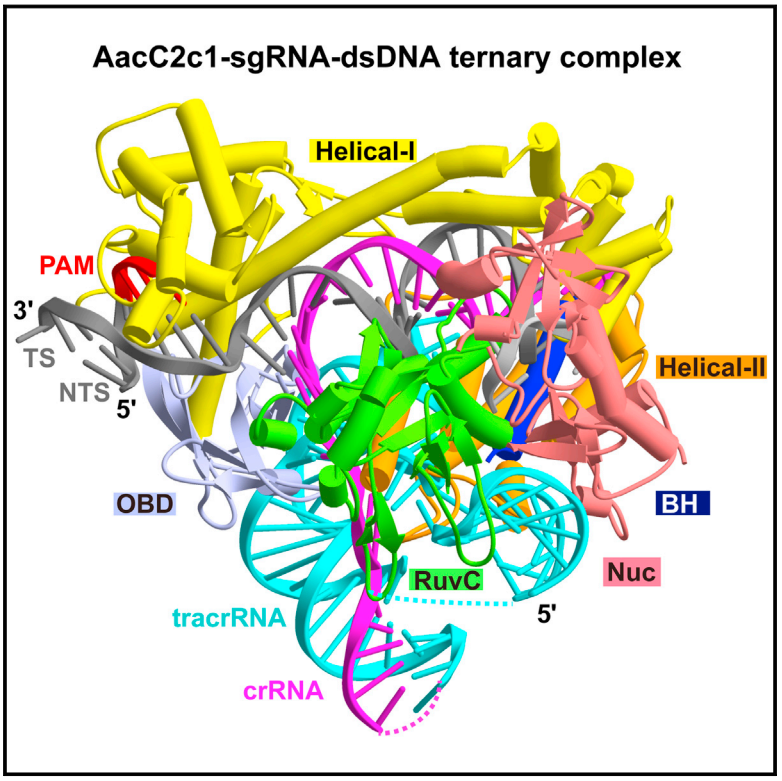
Article

Cell

PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease

Graphical Abstract



Authors

Hui Yang, Pu Gao, Kanagalaghatta R. Rajashankar, Dinshaw J. Patel

Correspondence

yangh3@mskcc.org (H.Y.), pateld@mskcc.org (D.J.P.)

In Brief

Structural analyses of a C2c1 RNA-guided DNA endonuclease reveal a distinctive mode for recognizing and cleaving target DNA, pointing to mechanistic differences from other CRISPR effectors like Cas9 and, to a lesser extent, Cpf1 that may have implications for new genome editing applications.

Highlights

- Structures of *A. acidoterrestris* C2c1 in pre-target- and target-bound states
- Block to crRNA guide alignment released on addition of target DNA
- A single RuvC active site model reflecting sequential dsDNA cleavage
- C2c1 exhibits cleavage properties similar to Cpf1 and distinct from Cas9

Yang et al., 2016, Cell *167*, 1814–1828
December 15, 2016 © 2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.11.053

CellPress

Cell

Article

PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease

Hui Yang,[1,*] Pu Gao,[1,2] Kanagalaghatta R. Rajashankar,[3,4] and Dinshaw J. Patel[1,5,*]
[1]Structurel Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA
[2]Key Laboratory of Infection and Immunity, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
[3]Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853, USA
[4]NE-CAT, Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60349, USA
[5]Lead Contact
*Correspondence: yangh3@mskcc.org (H.Y.), pateld@mskcc.org (D.J.P.)
http://dx.doi.org/10.1016/j.cell.2016.11.053

SUMMARY

C2c1 is a newly identified guide RNA-mediated type V-B CRISPR-Cas endonuclease that site-specifically targets and cleaves both strands of target DNA. We have determined crystal structures of *Alicyclobacillus acidoterrestris* C2c1 (AacC2c1) bound to sgRNA as a binary complex and to target DNAs as ternary complexes, thereby capturing catalytically competent conformations of AacC2c1 with both target and non-target DNA strands independently positioned within a single RuvC catalytic pocket. Moreover, C2c1-mediated cleavage results in a staggered seven-nucleotide break of target DNA. crRNA adopts a pre-ordered five-nucleotide A-form seed sequence in the binary complex, with release of an inserted tryptophan, facilitating zippering up of 20-bp guide RNA:target DNA heteroduplex on ternary complex formation. Notably, the PAM-interacting cleft adopts a "locked" conformation on ternary complex formation. Structural comparison of C2c1 ternary complexes with their Cas9 and Cpf1 counterparts highlights the diverse mechanisms adopted by these distinct CRISPR-Cas systems, thereby broadening and enhancing their applicability as genome editing tools.

INTRODUCTION

Bacteria and archaea have developed a set of defense mechanisms to protect themselves against invaders such as phages and plasmids. Central among these are the R-M (restriction-modification) and CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR associated) systems, which directly target the incoming phage/plasmid DNA (Dupuis et al., 2013; Hille and Charpentier, 2016; Wright et al., 2016). Unlike the R-M systems, which provide generalized or innate protection against any invaders not possessing countermeasures, the CRSIPR-Cas system functions as the only adaptive immune system in prokaryotes by generating records of previous infections (Hille and Charpentier, 2016; Wright et al., 2016). CRISPR-Cas systems are found in almost all archaea and about half of bacteria, showing extreme diversity of Cas protein composition, as well as genomic loci architecture (Makarova et al., 2011, 2015). The defense mechanism of CRISPR-Cas systems can be divided into three stages: (1) spacer acquisition, (2) crRNA (CRISPR RNA) biogenesis, and (3) target interference (van der Oost et al., 2009; Wright et al., 2016). Based on the architecture of the interference modules, the CRISPR-Cas systems can be broadly grouped into two classes: class 1 systems (including type I, III, and IV) possess multi-subunit protein complexes, whereas class 2 systems (including type II, V, and VI) rely on single effector proteins (Makarova et al., 2015).

Two class 2 CRISPR-Cas effectors, Cas9 (type II) (Jiang and Marraffini, 2015; Sternberg and Doudna, 2015; Wright et al., 2016) and Cpf1 (type V) (Zetsche et al., 2015) have been successfully harnessed for genome editing. The cleavage activity of Cas9 requires dual guide RNAs, namely crRNA and tracrRNA (*trans*-activating crRNA), or a synthetic covalently linked sgRNA (single-guide RNA) (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012), as well as a short G-rich PAM (protospacer adjacent motif) sequence near the target site (Deveau et al., 2008; Mojica et al., 2009). Cas9 contains two nuclease domains, an HNH domain, which cleaves the target strand base paired with the guide RNA, and a RuvC domain, which cleaves the non-base-paired non-target strand (Barrangou et al., 2007; Deltcheva et al., 2011; Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). As an alternative tool for genome editing, Cpf1 possesses several unique features that distinguish it from Cas9 (Dong et al., 2016; Fonfara et al., 2016; Gao et al., 2016; Yamano et al., 2016; Zetsche et al., 2015): (1) Cpf1 contains only one RuvC domain but lacks the HNH domain; (2) Cpf1 is guided by a single crRNA; (3) Cpf1 recognizes a distal 5′-T-rich PAM, in contrast to a proximal 3′-G-rich PAM recognition by Cas9; and (4) Cpf1 makes a staggered double-strand break, whereas Cas9 generates blunt ends.

C2c1 (Figure 1A), a newly identified class 2 type V-B effector protein containing one RuvC domain but no HNH domain, has also been shown to be active for target DNA cleavage in vivo and in vitro (Shmakov et al., 2015). The overall domain architecture of C2c1, in particular the organization of the RuvC domain,

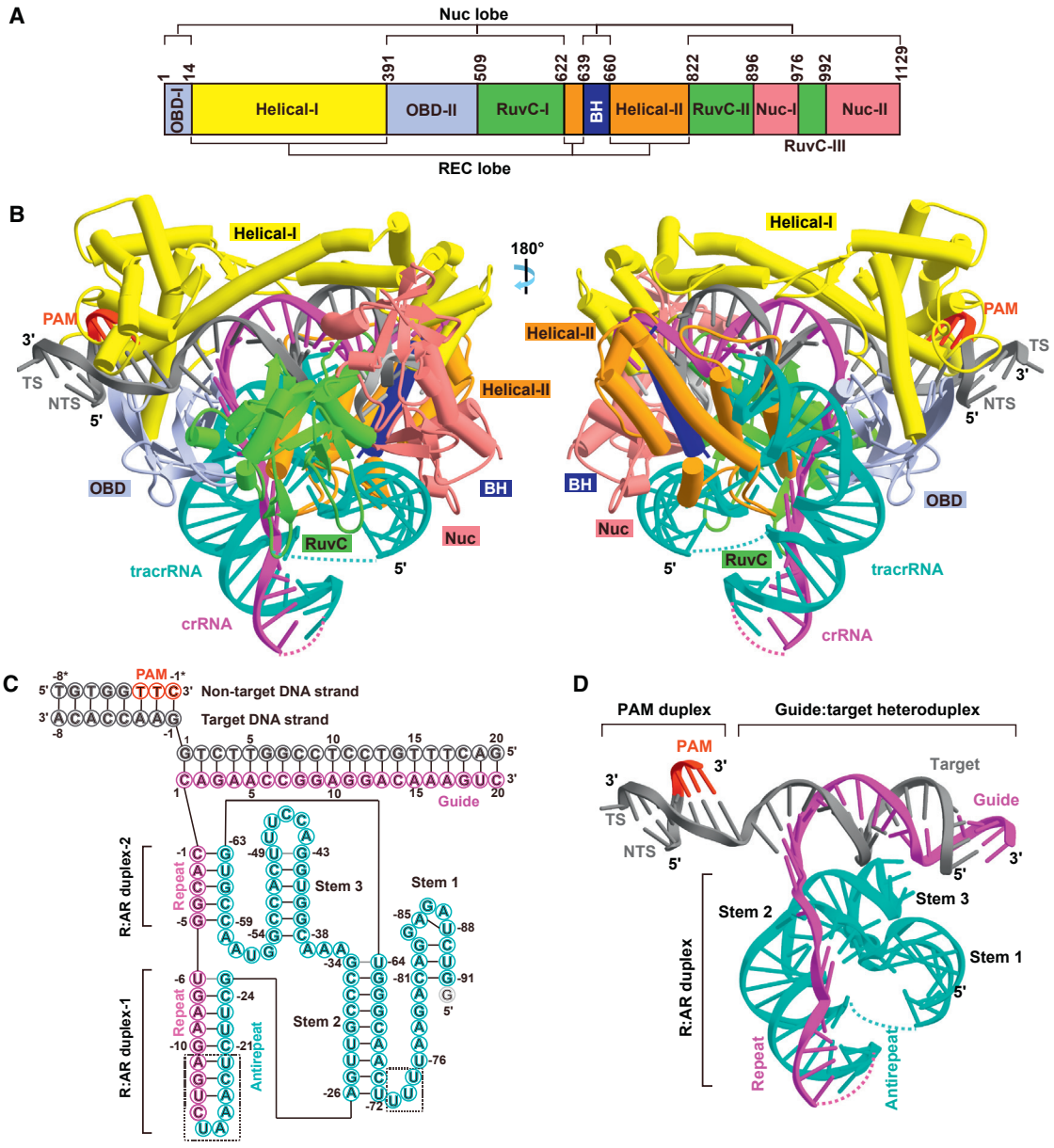


Figure 1. Overall Structure of AacC2C1-sgRNA-DNA Ternary Complex

(A) Domain organization of C2c1.

(B) Two views of a ribbon representation of C2c1-sgRNA-DNA ternary complex, color-coded as defined (A) and (C). TS, target DNA strand; NTS, non-target DNA strand.

(C) Schematic representation of the sgRNA-target DNA scaffold.

(D) Structure of the sgRNA-target DNA in the ternary complex.

See also Figures S1 and S2 and Table S1.

resembles Cpf1, but is distinct from Cas9. Similar to Cpf1, C2c1 recognizes a distal 5′-T-rich PAM sequence, in contrast to the proximal 3′-G-rich PAM favored by Cas9 (Shmakov et al., 2015). However, the cleavage activity of C2c1 requires both crRNA and tracrRNA, a feature that is in sharp contrast to Cpf1, which only requires crRNA. Similar to Cas9, an engineered sgRNA generated by covalently fusing crRNA and tracrRNA can also guide C2c1 to cleave target DNA. An evolutionary study of class 2 systems suggests that Cas9/Cpf1/C2c1 proteins originated from distinct recombination events (Shmakov et al., 2015), in line with the low similarity of the primary sequence among these proteins. Compared to the extensive structural and functional studies of the Cas9 and Cpf1 systems, little is known about the C2c1 system as regards the molecular mechanisms underlying crRNA and tracrRNA recognition, as well as the cleavage pattern of target DNA. Based on previous research on

structures of binary (Dong et al., 2016) and ternary (Gao et al., 2016; Yamano et al., 2016) complexes of Cpf1, we showed that Cpf1 employs a unique mechanism for target recognition in terms of both the "disordered-to-ordered" guide RNA seed sequence and the "open-to-closed" PAM-interacting cleft on ternary complex formation (Gao et al., 2016), which is distinct from that reported for Cas9 (Jiang et al., 2015). Whether C2c1 uses a similar strategy to Cpf1 or Cas9, or is distinct from both of them, is still unknown.

To better understand how C2c1 recognizes crRNA, tracrRNA, and target DNA, as well the principles underlying cleavage of both DNA strands, we determined the crystal structures of the *Alicyclobacillus acidoterrestris* C2c1 (AacC2c1) bound to sgRNA as a binary complex and to added target DNA containing the 5′-TTC-3′ PAM as ternary complexes. More importantly, we observed that both target and non-target DNA strands independently insert into the RuvC catalytic pocket in ternary complexes. The structural comparison of C2c1 with Cas9 and Cpf1 reveals that C2c1 employs a unique strategy for RNA and DNA recognition and cleavage. Cleavage by C2c1 generates a staggered double-strand break resulting in seven-nucleotide 5′ overhangs distal to the PAM site. Further, biochemical results show that C2c1 is active in the physiological temperature range. The structural and biochemical characterization of the C2c1 system reported in this study provides guidelines for future applications in genome engineering and related biotechnology applications.

RESULTS

Structure Determination of the AacC2c1-sgRNA-DNA Ternary Complex

To avoid potential cleavage of the target DNA during crystallization, we mutated three conserved C2c1 RuvC domain acidic residues (Asp570, Glu848, and Asp977) to alanine, based on previous bioinformatics analysis of catalytic residue assignments in Cpf1 and C2c1 (Shmakov et al., 2015). We have successfully crystallized the ternary complex of the triple catalytic mutant of full-length AacC2c1 in complex with a 111-nt sgRNA, a 28-nt target DNA, and a 8-nt non-target DNA strand containing a 5′-TTC-3′ PAM sequence (in excess) and solved the structure at 2.9 Å resolution (X-ray statistics summarized in Table S1). The overall structure of the ternary complex is shown in two views in Figure 1B, with the protein and nucleic acid domains color-coded as shown in Figures 1A and 1C, respectively.

Topology of C2c1

The structure of C2c1 in the ternary complex adopts an overall "Crab Claw" fold consisting of two lobes: an α-helical recognition (REC) lobe, including Helical-I and Helical-II, and a NUC lobe, including OBD, RuvC, and Nuc domains (Figure 1B). The bridge helix (BH) motif is inserted into the Helical-II domain in close proximity to the PAM-distal region of the guide:target heteroduplex.

Helical-I domain of C2c1 (Figures 1B and S1A) resembles a "dumbbell" shape and consists of 14 α helices and 2 short β strands forming a small antiparallel β sheet between helix α1 and helix α2. Helices α1–α7 and helices α9–α14 form two separate regions that are located at the PAM-proximal and PAM-distal regions of the guide:target heteroduplex, respectively. The long helix α8, which connects the two helical bundles, is located in the minor groove of PAM-proximal region and extends to the PAM-distal region. The Helical-II domain of C2c1 (Figures 1B and S1A) consists of 6 α helices and interacts with PAM-distal regions of the guide:target heteroduplex at the opposite side to Helical-I domain. Unlike the BH motif in Cpf1, which forms a separate region and connects the REC and NUC lobes from the middle of the whole complex (Gao et al., 2016; Yamano et al., 2016), the BH motif in C2c1 is inserted between helix α1 and helix α2 of Helical-II and forms a helix bundle together with the Helical-II domain (Figures 1B and S1).

The OBD domain of C2c1 consists of a nine-stranded β sheet and two α helices (Figure S1B), which closely resembles the core region of the OBD domain of Cpf1 (Figure S1E). The OBD and Helical-I domains contact the PAM duplex. The RuvC domain of C2c1 (Figure S1C) is composed of an 8-stranded β sheet that is surrounded by 6 α helices and a β-hairpin formed by β7 and β8, which shows a closely similar conformation to the RuvC domain in Cpf1 (Figure S1F). The Nuc domain of C2c1 contains two β sheets and six α helices (Figure S1C), showing low structural similarity to the Nuc domain in Cpf1 (Figure S1F).

Topologies of sgRNA and Target DNA

The 111-nt sgRNA was derived by connecting the 3′ end of 78-nt tracrRNA to the 5′ end of 42-nt crRNA based on the predicted repeat:anti-repeat (R:AR) duplex as described previously (Shmakov et al., 2015). The crRNA sequence (in magenta) contains a 20-nt guide region and a 14-nt repeat region, whereas the tracrRNA sequence (in cyan) contains a 16-nt anti-repeat region and three stem loops (Figure 1C). The sgRNA and target DNA (in gray) form a T-shaped architecture, consisting of the guide:target heteroduplex, the R:AR duplex, three stem loops, and the PAM duplex (Figures 1C and 1D). The nucleotides dG(20) to dG(1) segment (all sequences are labeled from 5′ to 3′ ends) of the target DNA strand and C1 to C20 segment of the crRNA form the guide:target heteroduplex. The dG(−1) to dA(−8) segment of the target DNA strand and the dT(−8*) to dC(−1*) segment of the non-target DNA strand (PAM in red) form the PAM duplex (Figures 1C and 1D).

In the crystal structure, the R:AR duplex comprises two regions: the predicted R:AR duplex-1 formed by nucleotides U(−13) to U(−6) of crRNA and G(−25) to A(−18) of tracrRNA (Shmakov et al., 2015) and an unanticipated R:AR duplex-2 formed by G(−5) to C(−1) of crRNA and G(−63) to C(−59) of tracrRNA (Figures 1C and 1D). The segment of the R:AR duplex-1 formed by U(−13) to A(−11) of crRNA and U(−20) to A(−18) of tracrRNA and the tetraloop A(−17) to C(−14) cannot be traced in the structure (dotted segment in Figures 1C and 1D). In addition, we are unable to trace nucleotides U(−75) to U(−73) between stem 1 and stem 2.

Recognition of Repeat:Anti-Repeat Duplex

The sgRNA is recognized by OBD, Helical-II, RuvC, and Nuc domains via extensive interactions, including several sequence-dependent intermolecular contacts (Figures 2A–2C, S2, and S3). The R:AR duplex-2 starts from G(−5) and ends at C(−1) of crRNA, with adjacent C1 of crRNA orienting in an opposite direction so as to pair with dG1 of the target DNA strand to form the

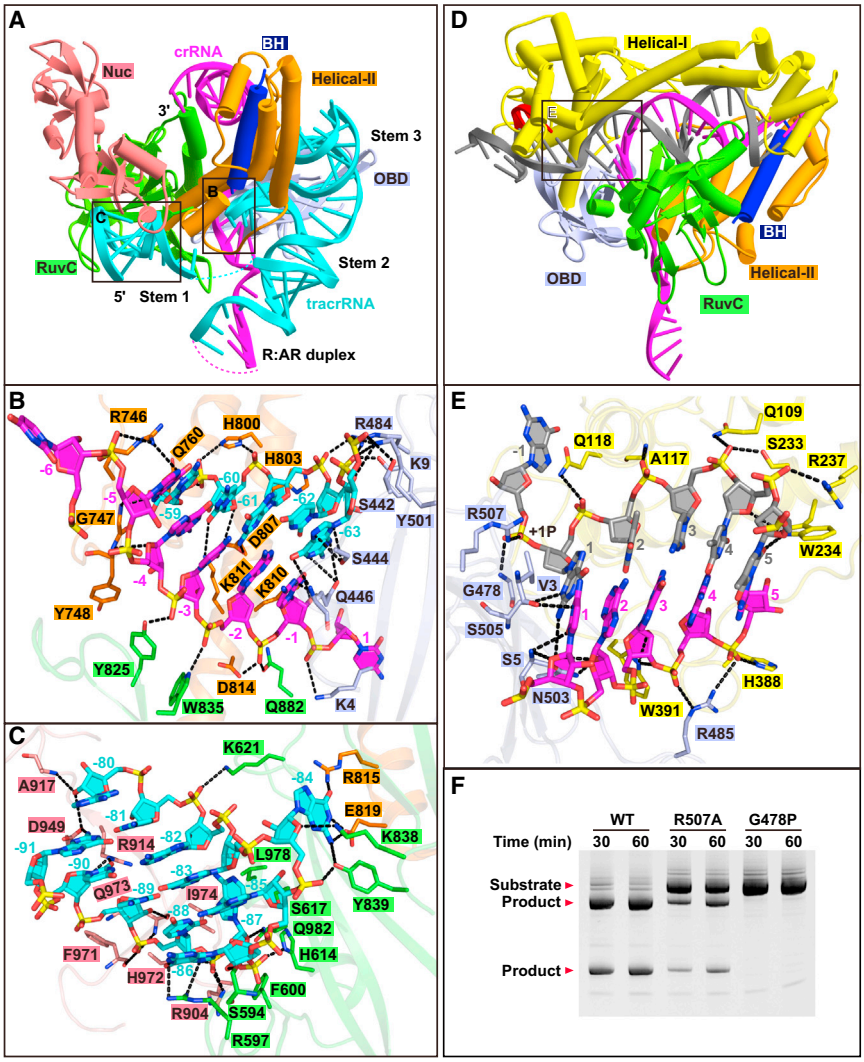


Figure 2. Detailed Interactions of C2c1 with sgRNA and Guide RNA-Target DNA Heteroduplex

(A) Recognition the sgRNA by the OBD, Helical-II, RuvC, and Nuc domains.

(B) Recognition the repeat:anti-repeat RNA duplex by the Helical-II and RuvC domains.

(C) Recognition stem 1 of sgRNA by RuvC and Nuc domains.

(D) Recognition of the guide:target heteroduplex by the REC and NUC lobes.

(E) Recognition of the +1 phosphate group (+1P) and the proximal-PAM region of the dG1-dT6/C1-A6 base pairs of the guide:target heteroduplex by the OBD and Helical-I domains.

(F) Mutational analysis of the residues involving in the binding of +1 phosphate group.

See also Figures S2, S3, and S4.

guide:target heteroduplex (Figures 1C and 2B). In addition, U(−6) of crRNA orients and pairs with G(−25) of tracrRNA to form R:AR duplex-1 (Figures 1C and 2B). The phosphates of C1 and G(−5) are stabilized by interaction with Lys4 and Arg746, respectively (Figure 2B and S2). The R:AR duplex-2 is recognized mainly by the Helical-II and OBD domains involving both non-sequence-specific interactions with the sugar-phosphate backbone and sequence-specific interactions with the nucleobases (Figures 1C, 2B, and S2). The nucleobase of G(−63) hydrogen bonds with the side chain of Ser444 and both the side chain and main-chain of Gln446 in the OBD domain. The nucleobase G(−5) stacks with Arg746 and hydrogen bonds with the main chain of Gly747 in the Helical-II domain. The N2 atom of G(−61) and O2 atom of C(−60) hydrogen bond with the side chains of Asp807 and Lys811 in the Helical-II domain, respectively (Figure 2B). In contrast to the R:AR duplex-2, which forms extensive interactions with C2c1, the R:AR duplex-1 protrudes away from the whole complex and has almost no contacts with the protein. One end of R:AR duplex-1 and the adjacent tetraloop cannot be traced in the structure, indicating that these regions may be flexible due to lack of interactions with C2c1. Similarly, no contacts were observed between the end of R:AR duplex with the protein in the Cas9 system (Deltcheva et al., 2011; Jinek et al., 2012).

Recognition of tracrRNA

Stem 1 of the tracrRNA interacts with Helical-II, RuvC, and Nuc domains primarily by non-sequence-specific hydrogen bonds (Figures 2C and S2). The atoms N1, N2, and O6 of the flipped-out nucleobase G(−84) hydrogen bond with the side chains of Glu819 and Arg815 in the Helical-II domain and with Tyr839 in the RuvC domain. Another unpaired nucleobase, G(−86), is sandwiched by A(−85):U(−88) base pair and the side chain of Arg904 in the Nuc domain and, in addition, forms hydrogen bonds with Arg904 (Figure 2C and S2). Stem-loop 3 is recognized by hydrogen-bonding interactions between nucleotides G(−42) to C(−46) and the Helical-II domain (Figures S2 and S3C). The nucleobases of C(−46) and C(−45) form two hydrogen bonds with the side chains of Arg724 and Trp774, respectively (Figures S2 and S3C). No interaction was observed between stem 2 and the protein in the structure.

The linker region A(−58) to G(−55) between R:AR duplex-2 and stem 3 of the tracrRNA is recognized by the Helical-II domain involving both sequence-dependent and backbone-binding interactions (Figures S2 and S3B). The nucleobases A(−57) and U(−56) hydrogen bond with the main chain of Ile745 and the side chains of Arg734 and Gln767, respectively. The nucleobase A(−58) flips out and is sandwiched by the side chains of Pro743 and Val753 (Figures S2, S3A, and S3B).

Recognition of Guide:Target Heteroduplex

The 20-bp guide:target heteroduplex is recognized primarily within a channel formed by REC lobe, BH motif, OBD domain,

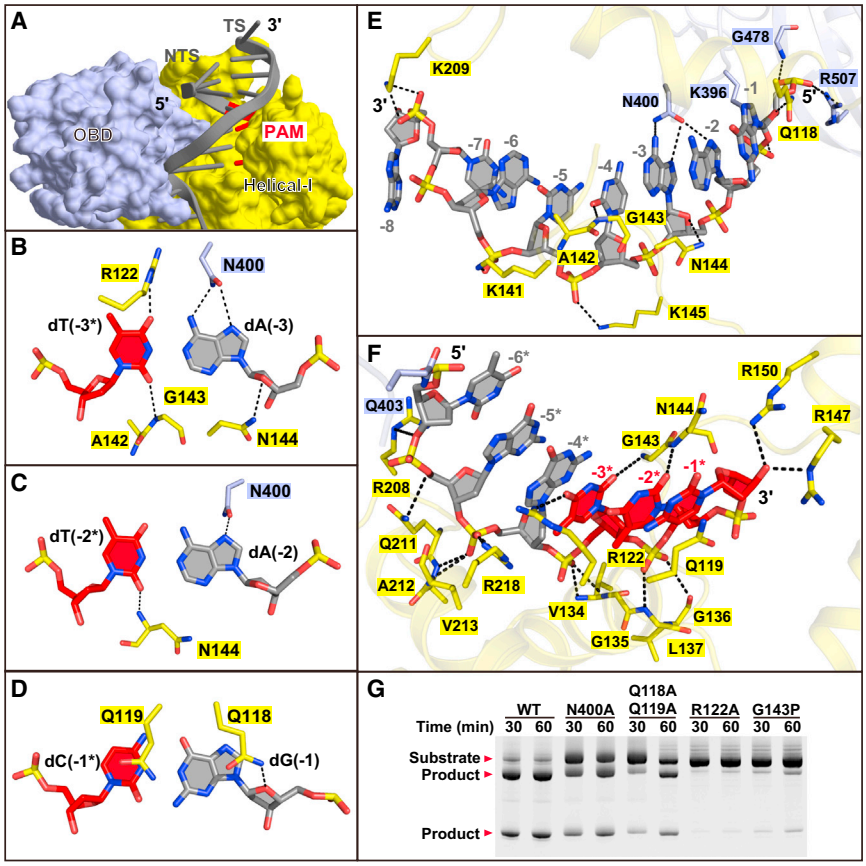


Figure 3. Recognition of the PAM Duplex

(A) Recognition of the PAM duplex by the OBD and Helical-I domains.

(B–D) Recognition of the dA(−3):dT(−3*) (panel B), dA(−2):dT(−2*) (panel C), and dG(−1):dC(−1*) (panel D) base pairs.

(E) Recognition of the target DNA strand in the PAM duplex.

(F) Recognition of the non-target DNA strand in the PAM duplex.

(G) In vitro cleavage assay of the PAM-interacting residues.

See also Figure S5.

and RuvC domain in a sequence-independent way (Figures 2D, 2E, and S4). The PAM-proximal region of the heteroduplex interacts with the OBD, Helical-I, Helical-II, and RuvC domains (Figures 2E and S2). The +1 phosphate group between dG(−1) and dG1 of the target DNA strand is stabilized by the side chain of Arg507 and the main-chain amide group of Gly478 in the OBD domain, which facilitates the unwinding of +1 phosphate group and base pairing between dG1 of target DNA and C1 of crRNA. Arg507 is conserved in some species and Gly478 is highly conserved in the C2c1 family. Such "phosphate lock" type interactions have also been observed in Cas9 (Anders et al., 2014; Jiang et al., 2016; Nishimasu et al., 2014) and Cpf1 (Gao et al., 2016; Yamano et al., 2016). Notably, the R507A and G478P mutants show reduced cleavage activities, indicating the importance of the "phosphate lock" (Figure 2F). The PAM-distal region of the heteroduplex is recognized mainly by Helical-I and Helical-II domains (Figures S2 and S4B–S4D). Two conserved residues Arg653 and Arg657 in the BH motif interact with the phosphate groups of C14 to A17 of crRNA (Figure S4C).

Recognition of the 5′-TTN-3′ PAM Segment

The PAM duplex is buried within the cleft formed by the OBD and Helical-I domains (Figures 3A–3F and S2). The loop between helices α4 and α5 in Helical-I domain inserts into the minor groove of the PAM duplex and interacts with the dC(−4) to dA(−8) region of the target DNA strand via backbone interactions (Figure 3E). The head of helix α8 and the loop between helices α7 and α8 in the Helical-I domain also interact with the region dT(−8*) to dG(−4*) of the non-target DNA strand through the minor groove in a sequence-independent manner (Figure 3F). The recognition of the 5′-TTC-3′ PAM segment involved nucleobases on both strands of the duplex (Figures 3B–3D), a feature also observed in Cpf1 (Gao et al., 2016; Yamano et al., 2016). The nucleobases of the dA(−3):dT(−3*) and dA(−2):dT(−2*) base pairs in the PAM duplex are recognized by several hydrogen bonds (Figures 3B, 3C, and S2). The O4 and O2 of dT(−3*) hydrogen bond with the side chain of Arg122 and the main-chain amide group of Gly143, respectively. In addition, the N6 and N7 of dA(−3) hydrogen bond with the side chain of Asn400 (Figure 3B). We modeled the dG(−3):dC(−3*) base pair and found steric clashes between the N4 of dC(−3*) with the side chain of Arg122 and between the N2 of dG(−3) and the backbone of Gly143 (Figure S5A). The O2 of dT(−2*) and N7 of dA(−2) hydrogen bond with the main-chain amide group of Asn144 and the side chain of Asn400 (Figure 3C). We also modeled the dG(−2):dC(−2*) base pair and found steric clashes between the N2 of dG(−2) with the side chain of Asn144 (Figure S5B). Our findings explain the requirement for recognition of the two A-T base pairs in the PAM. The dG(−1):dC(−1*) base pair in the PAM duplex stacks with Gln118 and Gln119 without hydrogen bond formation (Figure 3D), consistent with the absence of sequence specificity at the 5′ position of the PAM (Shmakov et al., 2015). The observed stacking involving these two Gln residues may contribute to D-loop formation adjacent to the PAM sequence involving separation of target and non-target DNA strands. Asn400 and Gly143 are highly conserved in C2c1 homologs, whereas Gln118, Gln119, and Arg122 are conserved in some species. The mutations of these residues exhibit reduced cleavage activity (Figure 3G), with the Q118A/Q119A mutation potentially destabilizing the junction between adjacent PAM and guide:target heteroduplex.

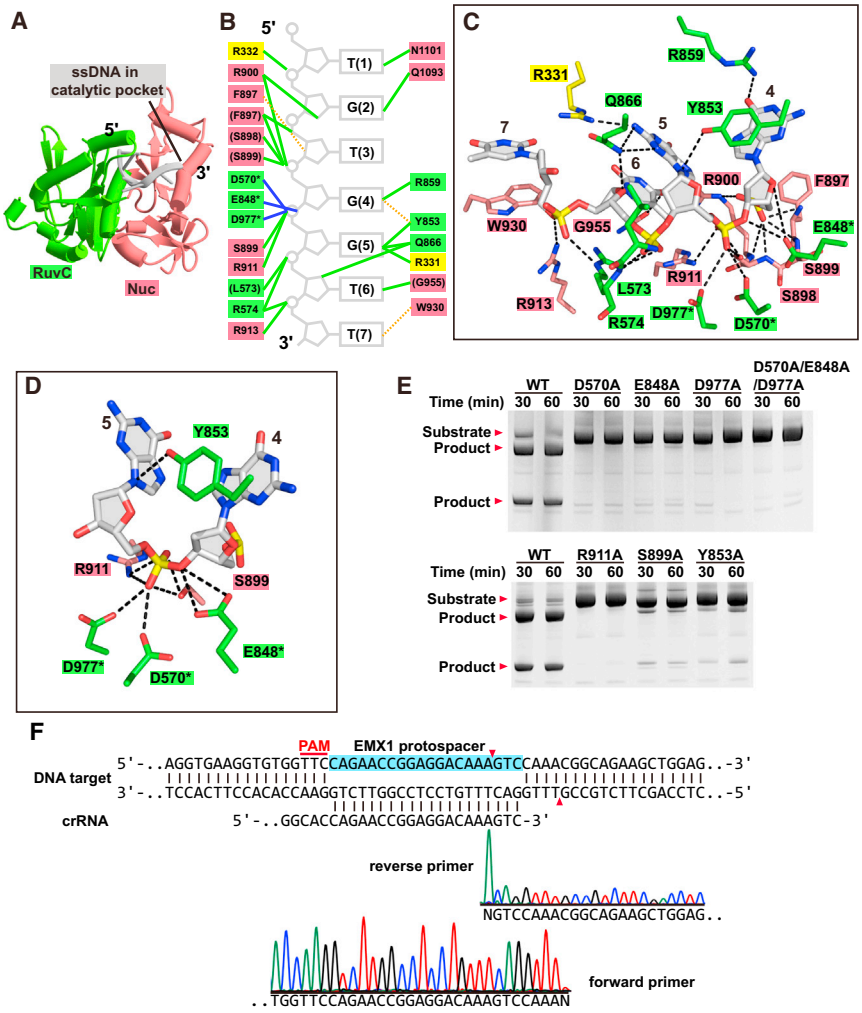


Figure 4. Recognition of Excess 8-mer Substrate DNA Strand and Positioning Relative to Cleavage Site

(A) Recognition of substrate DNA by the RuvC and Nuc domains.

(B) Schematic of detailed interactions of substrate DNA recognition by RuvC and Nuc domains. Hydrogen bonds and salt bridges are indicated with green lines. Hydrophobic and stacking interactions are shown by dashed orange lines.

(C) Recognition of substrate DNA strand in the catalytic pocket.

(D) Recognition of potential scissile phosphate of substrate DNA strand in the catalytic pocket.

The modeled side chains of three Ala-mutated catalytic residues are marked by stars in (B)–(D).

(E) Mutational analysis of key residues lining the catalytic pocket.

(F) Schematic of the engineered AacC2c1 sgRNA targeting the *EMX1* DNA. The cleavage sites are indicated by red triangles. Sanger-sequencing traces are shown below. The additional non-templated adenines are denoted as N, which resulted from the polymerase used in sequencing (Clark, 1988).

Recognition of Substrate DNA in the RuvC Catalytic Pocket

In the structure of the ternary complex, we observed a 8-nt ssDNA (reflecting excess added 8-nt non-target DNA strand, colored in silver) in the RuvC catalytic pocket, which represents the substrate DNA bound state of the RuvC domain (Figure 4A). The ssDNA is positioned on the binding surface formed by the RuvC and Nuc domains, with complex formation stabilized mainly by sequence-independent interactions (Figures 4B, 4C, and S2). The dT1 to dT3 segment is recognized by the Nuc domain, while the dG4 to dT7 is recognized by both RuvC and Nuc domains (Figures 4A and 4B). The nucleotide dT7 is located at the edge of the RuvC domain and points toward the cavity formed by the RuvC and Helical-II domains, which is close to the guide:target heteroduplex. The last nucleotide dC8 cannot be traced in the structure, indicating that it is flexible and likely forms few interactions with the protein.

The three Ala-mutated catalytic residues Asp570, Glu848, and Asp977 of the RuvC domain are positioned close to the phosphate group between dG4 and dG5, indicating this as the scissile phosphate position. We then modeled the side chains of the three catalytic residues and found that they could be positioned for hydrogen bond formation with the scissile phosphate group of dG5 (Figures 4B–4D). In addition, the side chains of Ser899 and Arg911 of the Nuc domain also hydrogen bond with this scissile phosphate. It is also worth noting that the side chain of Tyr853 inserts between dG4 and dG5, thereby stacking with the nucleobase of dG4. This π-π interaction breaks the stacking between dG4 and dG5 and makes the nucleobase of dG5 twist by ~90° and point in an alternate direction (Figures 4C and 4D). Notably, the related aromatic ring of Phe916 in SpyCas9 was also found to stack with the corresponding nucleobase at the scissile phosphate site (Jiang et al., 2016). All the above residues are highly conserved in the C2c1 family, indicating that they play essential roles in the cleavage reaction. We propose that these residues participate in the divalent metal-ion mediated catalytic cleavage of the phosphodiester bond between dG4 and dG5. Mutations of these acidic catalytic resides (Asp570, Glu848, and Asp997) exhibit complete loss of cleavage activity (top, Figure 4E), while others lining the catalytic pocket (Tyr853, Ser899, and Arg911) exhibit either very low or abolished cleavage activity (bottom, Figure 4E), which is consistent with our proposal.

The phosphate groups adjacent to scissile phosphate are also stabilized by several intermolecular hydrogen bonds (Figures 4B and 4C). Notably, the phosphate oxygens between dT3 and dG4 are hydrogen bonded with the main-chain amide groups of Phe897, Ser898, Ser899 and the side chain of Arg900. The phosphate oxygens between dG5 and dT6 are hydrogen bonded with Arg574 and the main-chain amide group of Leu573. These

interactions are likely to fix and stabilize the single-stranded DNA in the catalytic pocket. The nucleobase of dT7 flips out the catalytic pocket and stacks with the side chain of Trp930, which could potentially stabilize the kink associated with this nucleotide.

We found a sulfate group bound adjacent to the 5′ end of target DNA strand (Figure 5A). The sulfate group is stabilized by coordination to Arg766 in the Helical-II domain and Arg643 and Arg646 in the BH motif. In addition, we also found that the sulfate group bridges the nucleotide of dT7 and the 5′ end of the target DNA strand (Figure 5A). It is likely that this sulfate group mimics a phosphate group and that the target DNA strand could be connected to the ssDNA positioned in the catalytic pocket.

AacC2c1 Generates a Staggered Cleavage Site

It was reported previously that AacC2c1 was able to cleave both supercoiled and linear target DNA in presence of crRNA, tracrRNA, and Mg^{2+} in vitro (Shmakov et al., 2015). When supplied with an engineered sgRNA, AacC2c1 also showed comparable cleavage activity to both supercoiled and linear target DNA. We therefore investigated the optimal cleavage temperature for AacC2c1 in presence of sgRNA. We purified wild-type AacC2c1 and assayed its DNA cleavage activity in vitro at different temperatures. We generated in-vitro-transcribed sgRNA consisting of a 20-nt spacer targeting a sequence from the human *EMX1* locus as described previously (Shmakov et al., 2015) and used PCR to amplify a 646 bp linear fragment containing the same DNA target site. We found that AacC2c1, along with sgRNA, was able to cleave the target DNA between 37°C and 60°C and showed temperature-dependent cleavage activity (Figure S5C). The cleavage ability at 37°C is lower than when compared to that at higher temperature. Our results at higher temperatures are similar to those reported previously for the system containing crRNA and tracrRNA (Shmakov et al., 2015).

We also mapped the cleavage site of the target DNA using Sanger sequencing. We found that AacC2c1 generated a staggered double-strand break resulting in 7-nt 5′ overhangs (Figure 4F). The cleavage site in the non-target DNA strand is located after the 17th base, while the cleavage site in the target DNA strand is located after the 24th base, with the latter position close to the guide:target heteroduplex (Figure 4F).

Recognition of an Extended Target DNA Strand in the RuvC Catalytic Pocket

We therefore designed an extended 38-nt target DNA strand containing a 10-nt extension beyond the guide:target heteroduplex (sequence in Figure 5C) and determined the crystal structure of the ternary complex of AacC2c1, sgRNA, and extended target DNA at 2.9 Å resolution (X-ray statistics in Table S1). The overall structure of the extended target DNA ternary complex (Figure 5B) is similar to the non-extended target DNA ternary complex reported above. As we anticipated, the 5′ of the extended target DNA strand kinks and inserts into the RuvC catalytic pocket (Figure 5C), with the scissile phosphate positioned between dT24 and dG25, such that the cleavage site in the target DNA strand is located after the 24th base. The interactions between the catalytic pocket and extended target DNA strand are nearly identical to those observed in the substrate DNA strand ternary complex (Figure 5A), with the sulfate group replaced by a phosphate group between dG20 and dG21 (Figure 5C). The backbone of dG22 in the extended DNA strand (Figure 5C) is located in a similar position to dT7 in the substrate strand (Figure 5A), while the two bases point in different directions. The side chain of Trp930 also stacks with the nucleobase of dG21, while the nucleobase of dT22 hydrogen bonds with Arg331 in the Helical-I domain and Gln866 in the RuvC domain (Figure 5D). All these intermolecular interactions are likely to help in unwinding the target DNA strand so as to precisely position it in the catalytic pocket and also stabilize the kink of the target DNA strand.

We carried out mutagenesis assays for residues interacting with the substrate strand (Figure S5D) and found that single mutations of Arg574 and Arg900 reduced catalytic activity, while other single mutations had minimal impact.

Recognition of an Extended Non-target $(dT)_n$ DNA Strand in the RuvC Catalytic Pocket

We also designed an extended non-target strand consisting of $(dT)_{20}$ linked to the PAM (Figure 5E) and determined the crystal structure of the ternary complex of AacC2c1, sgRNA, and extended non-target DNA to 3.8 Å resolution (X-ray statistics in Table S1). Given the modest resolution, we could trace the PAM and an adjacent d(T-T) segment, as well as a d(T-T-T-T) segment positioned in the RuvC catalytic pocket and spanning the cleavage site (Figure 5E). The intervening $(dT)_n$ segment was disordered in the structure of the ternary complex. Given the sequence redundancy associated with the $(dT)_{20}$ used in this study, the structural data do not identify the position of the cleavage site on the non-target strand.

Structure Determination of the AacC2c1-sgRNA Binary Complex

To clarify how AacC2c1 assembles with sgRNA prior to the target DNA binding, we have also determined the crystal structure of the binary complex of AacC2c1(E848A) bound to sgRNA at 3.3 Å resolution (X-ray statistics summarized in Table S1). The overall structure of the AacC2c1-sgRNA binary complex (two views in Figure 6A) also adopts the bilobed architecture observed for the AacC2c1-sgRNA-DNA ternary complex (Figure 1B).

Modest Domain Rearrangements following DNA Recognition

A structural comparison between the AacC2c1-sgRNA binary and AacC2c1-sgRNA-DNA ternary complexes reveals a modest "rigid body" shift in the REC lobe (helical-I and helical-II) and BH motif following DNA recognition (Figures 6B, S6A, and S6B). The helical α1–α7 segment in the "dumbbell"-shaped Helical-I domain that interacts with the PAM duplex exhibits small conformational changes (Figures 6B and S6A). The helical α9–α14 segment and the long helix α8 in the Helical-I domain that interacts with the guide:target heteroduplex rotate toward the NUC lobe (Figures 6B and S6A). In addition, the helical-II domain and the BH motif move away from the NUC lobe (Figures 6B and S6B). These movements create enough space to accommodate

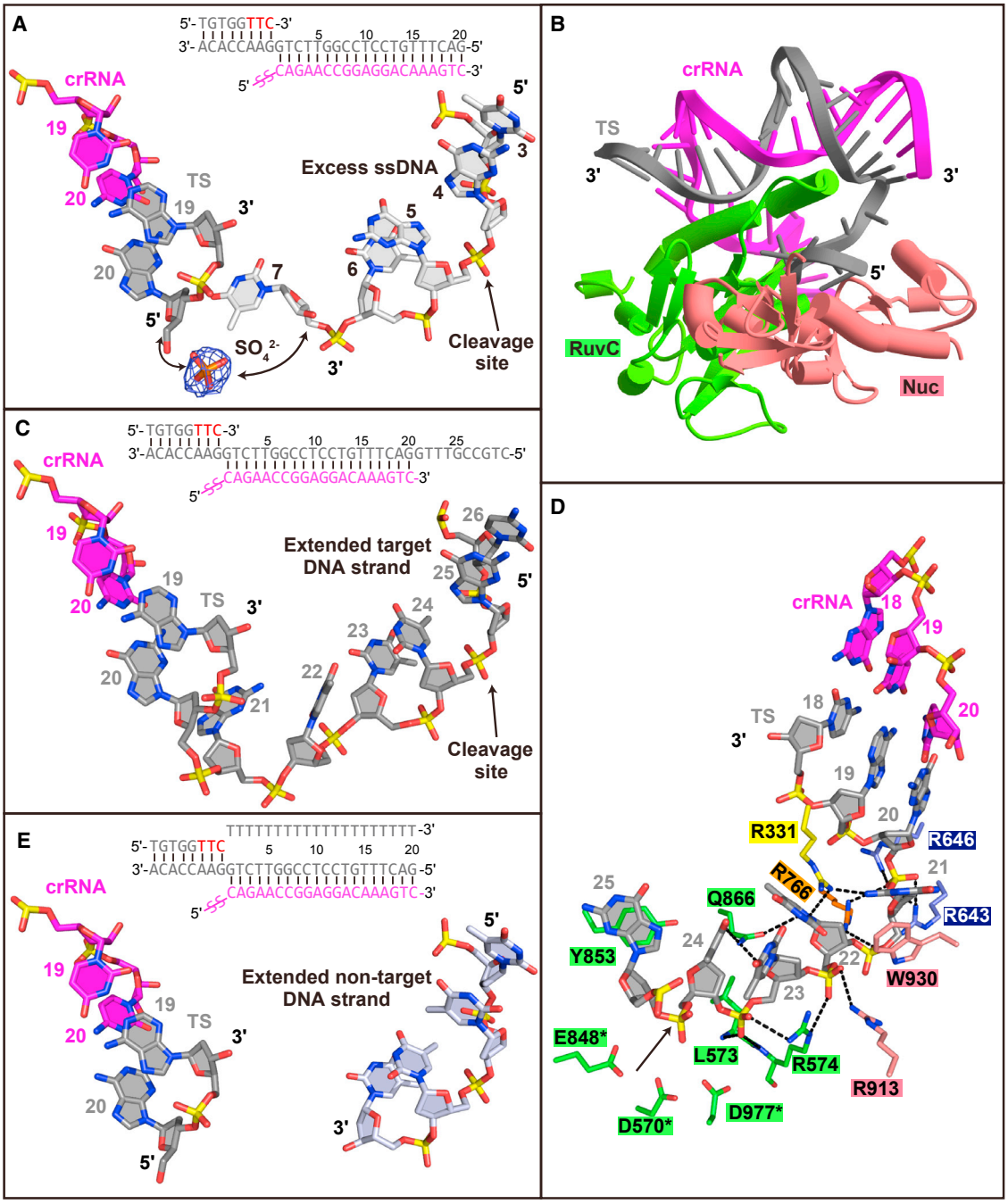


Figure 5. Recognition of the Extended Target and Non-target DNA Strands

(A) Stick representation of the excess 8-mer substrate DNA strand positioned in the catalytic pocket. The bound sulfate group mimics a phosphate group potentially connecting the target DNA strand and substrate DNA strand.

(B) Overall structure of the extended target DNA strand recognition by the RuvC and Nuc domains.

(C) Stick representation of the extended target DNA strand in the RuvC catalytic pocket.

(D) Recognition of the dG21 to dG24 segment of extended target DNA strand in the catalytic pocket. The modeled side chains of three Ala-mutated catalytic residues are marked by stars.

(E) Stick representation of the extended non-target strand d(T-T-T-T) segment in the RuvC catalytic pocket.

See also Figure S5 and Table S1.

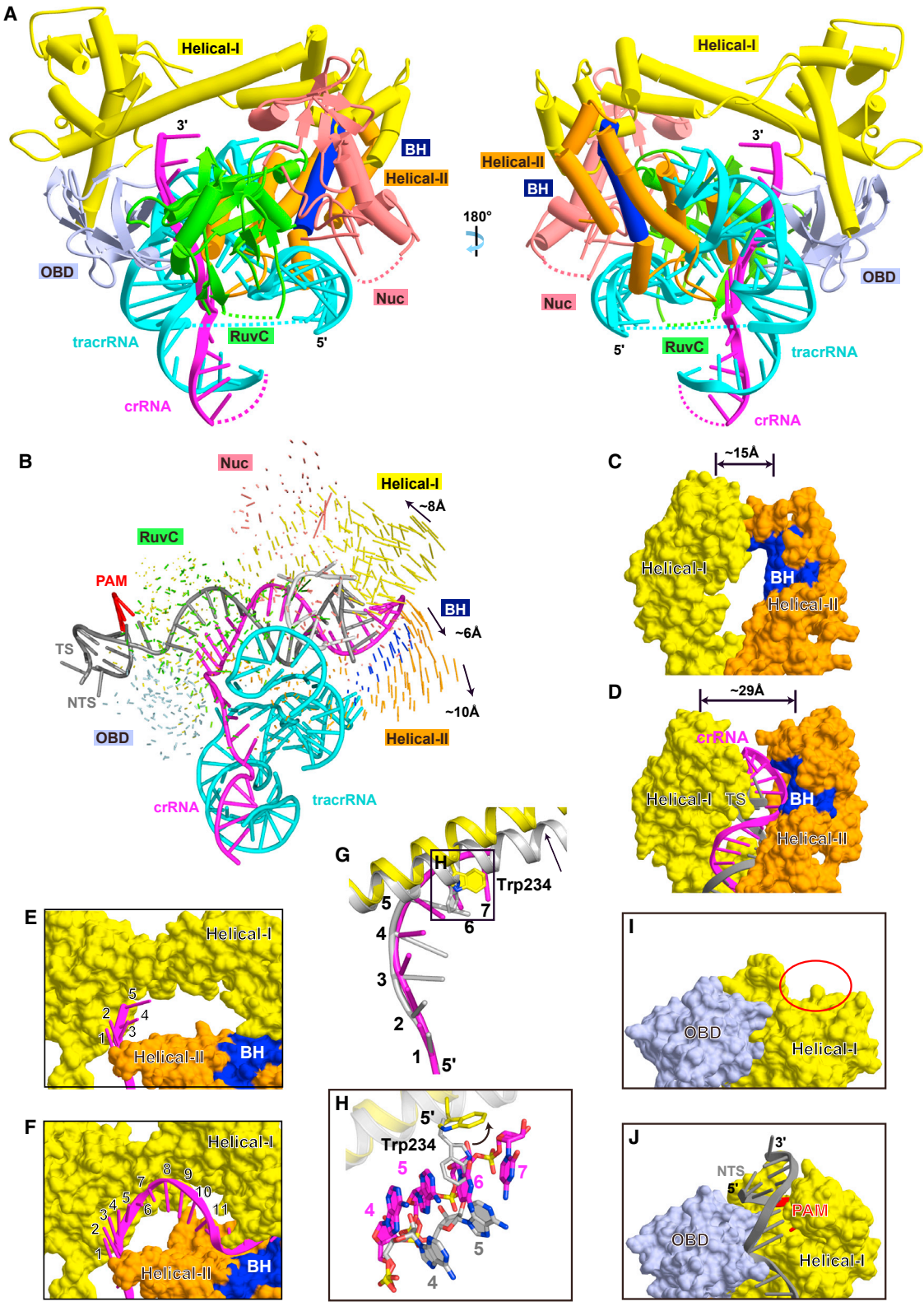


Figure 6. Domain Rearrangements on Ternary Complex Formation

(A) Two views of ribbon representation of C2c1-sgRNA binary complex.

(B) Structural comparison between C2c1-sgRNA binary and C2c1-sgRNA-DNA ternary complexes. Vector length correlates with the domain transition scale.

(legend continued on next page)

the guide:target heteroduplex (Figures 6C and 6D). The OBD, RuvC, and Nuc domains in the NUC lobe are nearly identical between the RNA-bound and RNA-DNA-bound complexes (Figure S6C).

Pre-organized A-Form crRNA Backbone in the Binary Complex

Previous structural data revealed formation of a pre-organized A-form crRNA in the seed region of the Cas9-sgRNA binary complex (Jiang et al., 2015), but not in Cpf1-crRNA binary complex (Dong et al., 2016). In the AacC2c1-sgRNA binary complex, only the first five nucleotides of the guide segment of sgRNA could be traced and located in the head of guide:target heteroduplex binding channel (Figure 6E). This region maintains a nearly A-form conformation along the sugar-phosphate backbone, with the first two nucleotides adopting a nearly identical conformation as those in AacC2c1-sgRNA-DNA ternary complex (Figure 6F). Notably, stacking is discontinuous within the 2-3-4 segment in the binary complex (Figure 6C). The side chain of conserved Trp234 in the Helical-I domain occupies the position otherwise adopted by the A6-C7 step, thereby disrupting formation of a pre-ordered A-form crRNA in the remaining A6 to C20 segment. Following DNA recognition on ternary complex formation, the Helical-I domain moves away, resulting in the rotation and movement of Trp234, thereby facilitating formation of an A-form helix along the length of the crRNA (Figures 6G and 6H). In addition, the Watson-Crick edges of bases 1 to 5 in the binary complex are exposed toward solvent (Figure 6H), facilitating access to and pairing with target DNA. Formation of a pre-organized A-form backbone for the crRNA over the first five nucleotide segment in the C2c1 binary complex may facilitate nucleation of guide:target duplex formation and allow the target DNA strand to be easily accommodated in the heteroduplex channel.

Locking the PAM Duplex on Ternary Complex Formation

Structural comparison between AacC2c1 binary (Figure 6I) and ternary (Figure 6J) complexes reveals that the OBD and Helical-I domains adopt similar conformations, except that helices α5, α6, and the loop between helices α4 and α5 in Helical-I domain are disordered in the binary complex (Figure 6I). The loop between helices α4 and α5 in the Helical-I domain inserts into the minor groove of the PAM duplex on ternary complex formation (Figure 6J) and interacts with the non-target DNA strand, implying a pre-organized PAM duplex cleft in AacC2c1. It is likely that helices α5, α6, and the loop between helices α4 and α5 act as a "switch" during the binding and recognition of the PAM duplex. Prior to PAM duplex loading, the flexible loop between helices α4 and α5 could allow helices α5 and α6 to flip out, thereby potentially allowing the PAM duplex easy access to the tunnel reflecting an "unlocked" conformation (Figure 6I). After the correct PAM duplex is bound and recognized, the PAM cleft likely transits to a "locked" conformation (Figure 6J). The loop could have the potential to flip back, thereby inserting into the minor groove so as to fix the duplex, and also make helices α5 and α6 "cover and lock" the duplex.

DISCUSSION

Class II CRISPR-Cas Endonucleases

To date, Cas9 and Cpf1 are the only known ternary complex structures of Class 2 CRISPR-Cas effectors containing bound guide RNA and target DNA. As a newly identified type V-B CRISPR-Cas effector, AacC2c1 turns out to be more like type V-A effector Cpf1 rather than type II effector Cas9 for several reasons: (1) the domain organization and architecture of C2c1 (Figure S7A) resemble that of Cpf1 (Figure S7B) (Shmakov et al., 2015); (2) both C2c1 and Cpf1 recognize a 5′-T-rich PAM (Shmakov et al., 2015), whereas Cas9 recognizes a 3′-G-rich PAM (Deveau et al., 2008; Mojica et al., 2009); (3) both C2c1 (Figure 4F) and Cpf1 generate staggered double-strand breaks on target DNA (Zetsche et al., 2015), whereas Cas9 generates blunt ends (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012). However, C2c1 also exhibits its own unique features and, in addition, includes some features, such as a requirement for tracrRNA, that are common with Cas9, but not Cpf1 (Shmakov et al., 2015).

Individual domains of C2c1 (Figure S7A) and Cpf1 (Figure S7B) share similar positions (except for the BH motif) in structures of both ternary complexes, which explains why C2c1, like Cpf1, also generates PAM-distal double-strand breaks (Zetsche et al., 2015). The OBD and RuvC domains adopt relatively similar folds in both ternary complexes, whereas other domains, including the Helical-I, Helical-II, and Nuc domains adopt distinct folds (Figure S1). Notably, C2c1 lacks the LHD domain found in Cpf1, thereby resulting in a different PAM-duplex-binding cleft (see Discussion blow).

Target DNA Recognition

In Cas9, the +1 phosphate group is recognized by the "phosphate lock" loop between the OBD and Helical-I domains (Anders et al., 2014; Jiang et al., 2016; Nishimasu et al., 2014), whereas both C2c1 and Cpf1 (Gao et al., 2016; Yamano et al., 2016) recognize the +1 phosphate group by residues in the OBD domain. The anticipated 25-bp guide:target heteroduplex in Cpf1 ternary complex is disrupted by the side chain of an inserted Trp, thereby instead forming a 20-bp heteroduplex (Gao et al., 2016; Yamano et al., 2016), which allows the target

(C and D) Binding with target DNA widens the guide RNA-target DNA heteroduplex binding channel on proceeding from the binary complex (C) to the ternary complex (D).

(E and F) Observable guide regions of crRNA in the C2c1-sgRNA binary complex (E) and the C2c1-sgRNA-DNA ternary complex (F).

(G and H) Superposition of guide RNA seed segment in binary (nucleotides 1 to 5, in silver) and ternary (nucleotides 1 to 7, in magenta) complexes. The unblocking movement of Trp234 from the binary complex (in silver) to the ternary complex (in yellow) allows continuous stacking of guide RNA in the ternary complex. (H) is a blow up of (G).

(I and J) "Locking" of the PAM duplex on proceeding from the binary complex (I) to the ternary complex (J). The untraceable loop-α5-α6 helical I segment in the binary complex is shown by a red circle in (I).

See also Figure S6 and Table S1.

DNA strand be cleaved in the resulting single-stranded region adjacent to the shortened heteroduplex (Zetsche et al., 2015). Despite lacking an inserted Trp in the C2c1 ternary complex, 20-bp of guide:target heteroduplex are recognized and accommodated in the binding channel, with site-specific cleavage observed in the adjacent single-stranded target DNA segment (Figure 4F).

PAM Duplex Recognition

Both Cpf1 (Gao et al., 2016; Yamano et al., 2016) and C2c1 (Figures 3B–3D) recognize both target and non-target DNA strands of the PAM duplex through both minor and major grooves. In both ternary complexes, the 5′-TTN-3′ PAM duplex segment is recognized through sequence-specific interactions with the first two A-T pair bases, while the last base pair is stabilized by sequence-independent stacking interactions in C2c1 or backbone-interactions in Cpf1. In Cpf1, the LHD domain interacts with the OBD domain and, together with the Helical-I domain, forms a relatively large channel in the binary complex (Dong et al., 2016), thereby providing access to the PAM duplex. Upon ternary complex formation (Gao et al., 2016; Yamano et al., 2016), PAM duplex binding induces movements in both LHD and Helical-I domains of Cpf1, which results in an "open-to-closed" conformational transition (Gao et al., 2016). In C2c1, the PAM duplex cleft is pre-organized in the binary complex (Figure 6I) and utilizes an alternate mechanism to achieve complex formation. Helices α5, α6, and the flexible loop between helices α4 and α5 in C2c1 occupy a similar position to that of the LHD domain in Cpf1. This region cannot be traced in the C2c1 pre-target-bound state (Figure 6I), indicative of flexibility and potential for a conformational transition on addition of target DNA. Upon achieving sequence-specific base recognition, the PAM duplex-binding loop segment of C2c1 becomes ordered on ternary complex formation (Figure 6J). This leads to additional sequence-independent sugar-phosphate backbone recognition in minor groove of the PAM duplex (Figures 3B–3F), thereby further stabilizing and fixing the alignment of the PAM duplex within the PAM duplex-binding cleft. It is likely that the disordered to ordered ("locked") transition in the PAM duplex-binding cleft of C2c1 has the potential for reducing off-target effects during target DNA recognition, binding, and cleavage.

Pre-organized crRNA Seed Segment

A critical feature, namely the pre-ordered guide RNA seed segment, has been found in Class 1 CRISPR Cascade complexes (Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014) and eukaryotic Argonaute binary complexes (Elkayam et al., 2012; Nakanishi et al., 2012; Schirle and MacRae, 2012), as well as for Cas9 in the pre-target-bound state (Jiang et al., 2015). By contrast, the seed segment cannot be traced in the Cpf1 pre-target bound state (Dong et al., 2016). In the pre-target-bound state of C2c1, the first five nucleotides of the guide RNA segment form a partially stacked pre-ordered A-form sugar-phosphate backbone conformation (Figures 6E and 6G), with a similar observation reported for Cas9 in the pre-target bound state. The intermolecular backbone interactions between the first five nucleotides and C2c1 are nearly identical in both pre-target-bound and target-bound states, resulting in the anchoring of these nucleotides in both states. The side chain of Trp234 disrupts formation of the A-form conformation beyond nucleotide 5 in the C2c1 binary complex (Figures 6G and 6H). On binding the target DNA, the Trp side chain reorients and moves away, so as to provide enough space for heteroduplex accommodation along the entire length of the guide segment (Figures 6G and 6H) (see also Jiang et al., 2015).

Taken together, it appears that the first five nucleotides in guide segment adopt a pre-organized conformation, with the Watson-Crick edges of their bases exposed to solvent and available for nucleation with target DNA. In addition, such an alignment allows the target DNA ready access to the guide: target heteroduplex-binding channel, compatible with high cleavage efficiency of C2c1.

Cleavage of Target DNA Strand

One challenging question relates to the precise mechanism of type V endonuclease cleavage of target DNA strand. For Cas9, the target strand is cleaved by the HNH domain, while the non-target strand is cleaved by the RuvC domain (Jiang et al., 2015). By contrast, both Cpf1 and C2c1 contain a RuvC domain but lack the HNH domain. It has been proposed for Cpf1 that the acidic catalytic pocket residues in the RuvC domain are key to cleavage of the non-target DNA strand, while conserved Arg1226 in the adjacent Nuc domain is the primary contributor to cleavage of the target DNA strand (Yamano et al., 2016). In the absence of structures of DNA in either proposed catalytic pocket of the Cpf1 ternary complex, these conclusions were solely based on the impact of mutational studies of putative catalytic residues.

Having observed DNA in the catalytic pocket of the C2c1 ternary complex for both substrate DNA (excess 8-nt DNA) (Figure 5A) and an extended target DNA strand (Figure 5C), our ternary complex structures provide direct insights into the cleavage mechanism of the target DNA strand. In our structures, the RuvC and Nuc domains together form a large flat catalytic pocket that can accommodate around a 6-nt ssDNA. The scissile phosphate group is in a position to hydrogen bond with the three modeled catalytic Asp570, Glu848, and Asp977 residues of the RuvC domain, with additional hydrogen bonding contributions from Ser899 and Arg911 on the adjacent Nuc domain (Figures 4C and 4D). These five residues have their corresponding counterparts in AsCpf1 (Asp908, Glu993, Asp1263, Ser1071, and Arg1226) (Gao et al., 2016; Yamano et al., 2016) (Figures 7B and 7E). Another important residue in C2c1 is Tyr853, which disrupts the stacking of the nucleobases adjacent the scissile phosphate group (Figures 4C and 4D), while this region (aa 996–1,009; contains conserved Phe999) is disordered in all reported AsCpf1 ternary complexes (Gao et al., 2016; Yamano et al., 2016). It appears that the above residues most likely play a similar role in the cleavage and function of C2c1 and Cpf1 endonucleases.

Our mutagenesis assays establish that all the three acidic catalytic residues are critical for the cleavage of both strands in C2c1 (Figure 4E, top), consistent with a previous mutagenesis study of acidic resides in the RuvC domain in AsCpf1 (Yamano et al., 2016). Of additional interest, mutations of Ser899, Arg911, and Tyr853 in C2c1 also impact strongly on cleavage activity of both DNA strands (Figure 4E, bottom). These observations

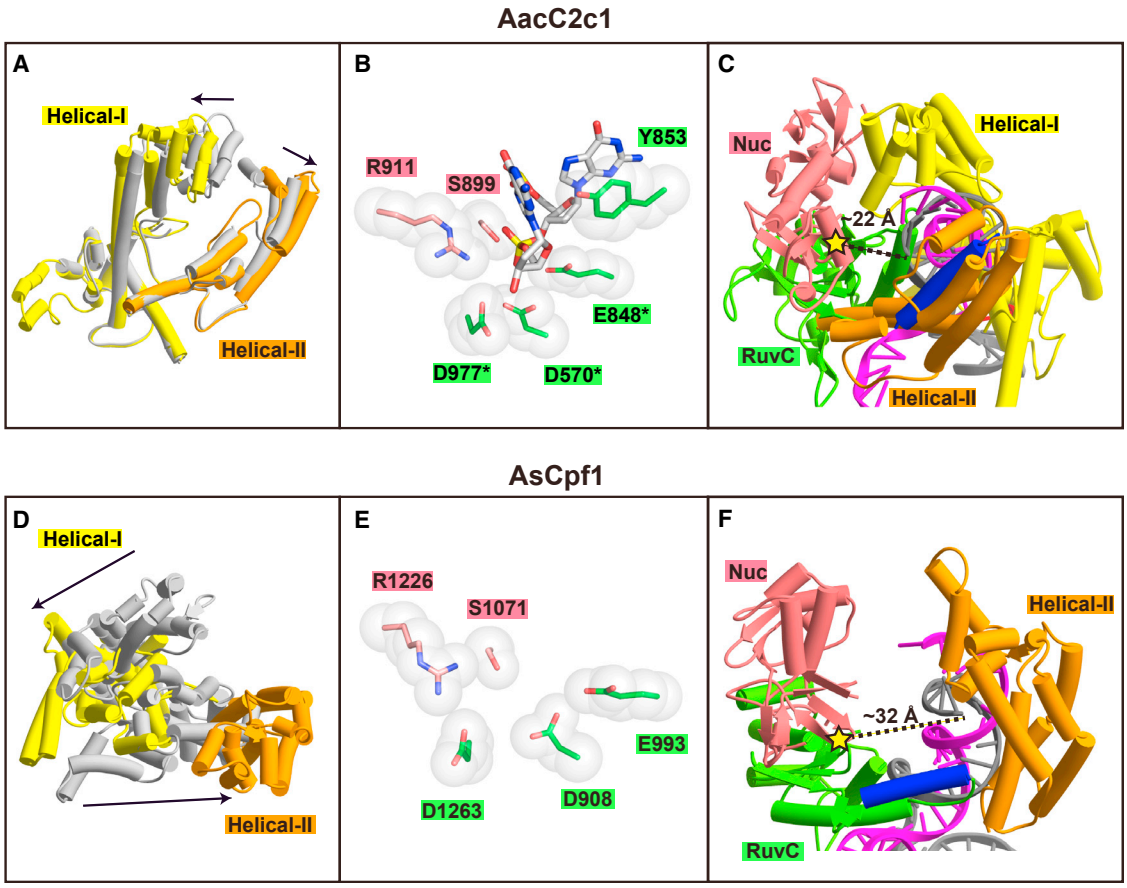


Figure 7. Structural Comparison between AacC2c1 and AsCpf1 Ternary Complexes

(A–C) AacC2c1 complex: Conformational transition in Helical-I and Helical-II domains on proceeding from the binary (in silver) to ternary (in color) complex (panel A), substrate DNA positioned in the RuvC catalytic pocket containing modeled acidic residues (panel B), and overview of the ternary complex structure emphasizing the distance between last nucleotide of guide:target heteroduplex and the catalytic pocket (panel C) .

(D–F) AsCpf1 complex: conformational transition in Helical-I and Helical-II domains on proceeding from the binary (in silver) to ternary (in color) complex (D), empty RuvC catalytic pocket (E), and overview of the ternary complex structure emphasizing the distance between last nucleotide of guide:target heteroduplex and the catalytic pocket (F).

See also Figure S7.

involving alignment of an extended target DNA strand in the RuvC pocket, together with identification of residues critical for cleavage chemistry (Figure 4E), definitively identify the RuvC pocket as the site for cleavage of the target DNA strand.

Positioning a Bent Target DNA Strand in the Catalytic Pocket

In our structure of the C2c1 ternary complex, the extended single-strand segment of the target DNA strand folds back to access the catalytic pocket. This kinked conformation of the target strand is stabilized through backbone- and stacking-interactions with residues from the BH motif, Nuc, Helical-I, and Helical-II domains (Figure 5D). In Cpf1, the PAM-distal end of the guide:target heteroduplex mainly interacts with the Helical-II domain. The residues involving in the DNA bending interactions in C2c1 are absent at the corresponding positions in Cpf1. Interestingly, the cleavage site of the target DNA strand generated by both Cpf1 (Zetsche et al., 2015) and C2c1 (Figure 4F) is between the 24th and 25th nucleotides. In addition, the length of the guide:target heteroduplexes is 20 bp for both C2c1 and Cpf1 (Gao et al., 2016; Yamano et al., 2016). Notably, in Cpf1, the distance between the last nucleotide of the guide:target heteroduplex and the catalytic pocket in the RuvC domain is considerably longer than that in C2c1 (Figures 7C and 7F). This may imply that the catalytic RuvC and Nuc domains in Cpf1 will have to undergo additional movements toward the target DNA strand to facilitate cleavage.

Although the catalytic residues that interact with the scissile phosphate in C2c1 adopt similar conformations to their counterparts in Cpf1, additional residues in RuvC and Nuc domains that interact with the backbones of nucleotides flanking the cleavage site share low sequence similarity between these two proteins. In addition, the Nuc domains in C2c1 (Figure S1C) and Cpf1 (Figure S1F) show low structural similarity, indicative of differences between C2c1 and Cpf1 for recognition and stabilization of target DNA strand in the catalytic pocket.

Cleavage of Non-target DNA Strand

Another challenging question relates to the precise mechanism of type V endonuclease cleavage of the non-target DNA strand. To address this question, in the first instance, we have solved the structure of a ternary complex containing an extended $(dT)_{20}$ sequence projecting from the PAM sequence. To this end, we observed a d(T-T) segment adjacent to the PAM and a d(T-T-T-T) segment positioned in the RuvC catalytic pocket (Figure 5E), together with a disordered looped out intervening $(dT)_n$ segment. To address the issue of target site cleavage on the non-target DNA strand, we are currently substituting a few purines at the putative cleavage site within the $(dT)_{20}$ overhang on the non-target strand (see schematic in Figure 4F) and in turn attempt to solve the structure.

A Single RuvC Pocket for Cleavage of Target and Non-target DNA Strands

A major conclusion of our study is that both target (Figure 5C) and non-target (Figure 5E) DNA strands are independently positioned in the same RuvC catalytic pocket lined by three invariant acidic and flanking catalytic residues in the C2c1 system. This result based on structures of ternary complexes of C2c1 containing either target or non-target DNA strand in the RuvC catalytic pocket reported in the current study contrasts with earlier conclusions on the related Cpf1 system, where separate binding pockets were proposed based solely on cleavage assays on catalytic mutants (Yamano et al., 2016). Specifically, it was proposed that the non-target DNA strand was cleaved within the RuvC catalytic pocket, while the target strand was cleaved in an adjacent pocket on the Nuc domain with contributions from acidic residues lining the RuvC pocket (Yamano et al., 2016). Clearly, further insights into cleavage by Cpf1 would benefit from determination of ternary complex structures that trap the target and non-target DNA strands in their respective catalytic pocket(s).

It has not escaped our attention that, if the same RuvC pocket in C2c1 is involved in independent cleavage of both target and non-target DNA strands, it will be necessary in the future to attempt elucidation of the sequential order of cleavage of DNA strands leading to double-strand breaks of target DNA.

Bound sgRNA Reinforces the Pre-organized Conformation of C2c1

In Cpf1, which binds crRNA, the transition from the pre-target-bound state (Dong et al., 2016) to the target-bound state (Gao et al., 2016; Yamano et al., 2016) involved large conformational changes (Figure 7D, highlighted for Helical-I and Helical-II domains) (Gao et al., 2016). By contrast, Cas9, which binds sgRNA (crRNA and tracrRNA), undergoes a modest conformational change (Jiang et al., 2016), presumably due to extensive interactions with the bound sgRNA. In our structure, the sgRNA also forms extensive interactions with C2c1, resulting in modest conformational changes (Figures 6B and 7A) for the same transition. We propose that these intermolecular interactions with sgRNA in C2c1 reinforce maintenance of the pre-organized conformation of C2c1 for target binding and cleavage. The R:AR duplex of sgRNA interacts with the OBD, Helical-II, and RuvC domains, and the adjacent stem 3 forms extensive contacts with the Helical-II domain (Figures 2B and S3B). These interactions may stabilize the OBD, Helical-II, and RucvC domains in a pre-organized conformation in the binary complex. Stem 1 interacts with the surface formed by the RuvC and Nuc domains (Figure 2C), which together may help in maintenance of the conformation of the catalytic pocket. The RuvC and Nuc domains of C2c1 show minor conformational transitions on proceeding from the pre-target-bound state to the target-bound state, likely implying that the intermolecular interactions between these domains and stem 1 maintain the catalytic pocket in a "pre-organized" state for target DNA strand recognition. The C2c1 ternary complex was generated using an excess 8-nt DNA, conditions under which this excess substrate DNA is positioned in the catalytic pocket (Figure 5A), where it adopts an identical conformation to that observed in the ternary complex for an extended target DNA strand (Figure 5C), further implying that the RuvC and Nuc domains are "pre-organized" for access of the target DNA strand.

Key Conclusions

Our structures of the binary and ternary complexes of C2c1 provide critical molecular insights into the mechanism of guide RNA-mediated targeting and cleavage of target DNA in a newly identified type V-B CRISPR-Cas system. crRNA adopts a pre-ordered five-nucleotide A-form seed sequence in the binary complex, with release of an inserted tryptophan that stacks over nucleotide 5, facilitating zippering up of 20-bp guide:target heteroduplex formation on ternary complex formation. Our structural studies on C2c1 ternary complexes containing extended DNA single strands have established that both target and non-target strands can be independently positioned for cleavage within a single acidic residue-lined RuvC catalytic pocket. Our biochemical studies identify the positions of C2c1-mediated cleavage sites resulting in a staggered 7-nucleotide cleavage of the target DNA duplex. The position of the cleavage site on the target DNA strand was validated from structural studies of a ternary complex containing an extended target DNA strand. Despite modest conformational changes associated with binary to ternary complex formation, the PAM-interacting cleft adopts a "locked" conformation on ternary complex formation. Our structural and biochemical studies on C2c1 have facilitated a comparison with related Cas9 and Cpf1, thereby highlighting the similarities and differences in cleavage mechanisms within this family of class II CRISPR-Cas endonucleases. Although C2c1 shows lower cleavage activity at 37°C than at elevated temperatures, our research provides useful insights toward future rational engineering of other C2c1 orthologs, with the potential for higher cleavage activity at ambient temperature, thereby facilitating utilization of this CRISPR-Cas endonuclease as a versatile genome editing tool.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING

- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Protein Expression and Purification
  - Crystallization, Data Collection, and Structure Determination
  - In vitro Transcription and Purification of sgRNA
  - In vitro cleavage assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.11.053.

AUTHOR CONTRIBUTIONS

H.Y. conducted all the crystallographic and biochemical experiments under the supervision of D.J.P. Both P.G. and K.R.R. were involved in assisting H.Y. in the refinement of the AacC2c1-sgRNA binary complex. H.Y., P.G., and D.J.P. wrote the paper.

ACKNOWLEDGMENTS

We thank Satoko Ishibi-Murakami for technical assistance in generation of C2c1 mutants. X-ray diffraction studies were conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team beamlines, which are supported by NIGMS grant P41 GM103403 and U.S. Department of Energy grant DE-AC02-06CH11357. The Pilatus 6M detector on 24-ID-C beam line is funded by a NIHORIP HEI grant (S10 RR029205). The research was supported by NIH grant GM104962 to D.J.P., the Memorial Sloan Kettering Cancer Center Core Grant (P30 CA008748), Cancer Research Institute Irvington Postdoctoral Fellowship, and start-up funds from the Institute of Biophysics, Beijing, China to P.G.

Received: October 29, 2016
Revised: November 18, 2016
Accepted: November 28, 2016
Published: December 15, 2016

REFERENCES

Adams, P.D., Afonine, P.V., Bunkóczi, 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., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Clark, J.M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res. *16*, 9677–9686.

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., Labonté, 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.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature *532*, 522–526.

Dupuis, M.E., Villion, M., Magadán, A.H., and Moineau, S. (2013). CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance. Nat. Commun. *4*, 2087.

Elkayam, E., Kuhn, C.D., Tocilj, A., Haase, A.D., Greene, E.M., Hannon, G.J., and Joshua-Tor, L. (2012). The structure of human argonaute-2 in complex with miR-20a. Cell *150*, 100–110.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517–521.

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res. *26*, 901–913.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, 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. USA *109*, E2579–E2586.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150496.

Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. Science *345*, 1473–1479.

Jiang, W., and Marraffini, L.A. (2015). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annu. Rev. Microbiol. *69*, 209–228.

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

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.

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.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

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.

Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. Science *345*, 1479–1484.

Nakanishi, K., Weinberg, D.E., Bartel, D.P., and Patel, D.J. (2012). Structure of yeast Argonaute with guide RNA. Nature *486*, 368–374.

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.

Schirle, N.T., and MacRae, I.J. (2012). The crystal structure of human Argonaute2. Science *336*, 1037–1040.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al.

(2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell *60*, 385–397.

Sternberg, S.H., and Doudna, J.A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. Mol. Cell *58*, 568–574.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. Trends Biochem. Sci. *34*, 401–407.

Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. Cell *164*, 29–44.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell *165*, 949–962.

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.

Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. Nature *515*, 147–150.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
ULP1	Home-made	N/A
T7 RNA polymerase	Home-made	N/A
1 M Lithium sulfate	QIAGEN	Cat#132558
1 M Ammonium acetate	QIAGEN	Cat#133011
50% w/v PEG 3350	QIAGEN	Cat#133083
1M HEPES, pH 7.5	QIAGEN	Cat#133048
Ethylene glycol	QIAGEN	Cat#133044
M9 Broth	Sigma-Aldrich	Cat#63011-500G-F
Urea	Sigma-Aldrich	CAS: 57-13-6 Cat#U5378-5KG
Diethyl pyrocarbonate	Sigma-Aldrich	CAS: 1609-47-8 Cat# D5758-50ML
40% acrylamide and bis-acrylamide solution, 29:1	Bio-Rad	Cat#161-0146
Critical Commercial Assays		
NeXtal Tubes Classics II Suite	QIAGEN	Cat#130723
High Pure PCR Product Purification Kit	Roche	Cat#11732676001
PureLink HiPure Plasmid Megaprep Kit	ThermoFisher	Cat#K210008
Deposited Data		
AacC2c1-sgRNA binary complex structure	This paper	PDB: 5U34
AacC2c1-sgRNA-8-mer substrate DNA ternary complex structure	This paper	PDB: 5U31
AacC2c1-sgRNA-extended target DNA ternary complex structure	This paper	PDB: 5U30
AacC2c1-sgRNA-extended non-target DNA ternary complex structure	This paper	PDB: 5U33
Experimental Models: Organisms/Strains		
E.coli: BL21(DE3) Competent cells	Novagen	Cat#69450-3
E.coli: DH5α Competent cells	ThermoFisher	Cat#18265017
Recombinant DNA		
pRSF-Duet-1-His_6-SUMO-AacC2c1	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R507A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-G478P	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-N400A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-Q118A/Q119A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R122A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-G143P	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-D570A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-E848A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-D977A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-D570A/E848A/D977A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-Y853A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-S899A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R911A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R331A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-Q866A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R574A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-W930A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R643A/R646A	This paper	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pRSF-Duet-1-His_6-SUMO-AacC2c1-R766A	This paper	N/A
pRSF-Duet-1-His_6-SUMO-AacC2c1-R900A	This paper	N/A
pUT7-sgRNA	This paper	N/A
Sequence-Based Reagents		
Sanger sequencing forward primer 1 GGACCCCGGCCTGGGGC	ThermoFisher	N/A
Sanger sequencing forward primer 2 CTAACCCTATGTAGCCTCAGT	ThermoFisher	N/A
Sanger sequencing reverse primer 1 GGGAGATTGGAGACACGGAGA	ThermoFisher	N/A
Sanger sequencing reverse primer 2 CTTCAAGCCCTGTGGGGCCA	ThermoFisher	N/A
8-nt non-target DNA strand TGTGGTTC	IDT	N/A
28-nt target DNA strand GACTTTGTCCTCCGGTTCTGGAACCACA	IDT	N/A
38-nt target DNA strand CTGCCGTTTGGACTTTGTCCTCCGGTTCTG GAACCACA	IDT	N/A
28-nt non-target DNA strand TGTGGTTCTTTTTTTTTTTTTTTTTTTT	IDT	N/A
Software and Algorithms		
Coot	Emsley et al., 2010	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
Phenix	Adams et al., 2010	https://www.phenix-online.org
NECAT RAPD online server		https://rapd.nec.aps.anl.gov/
PyMOL	The PyMOL Molecular Graphics System, Version 1.8.4.0, Schrodinger, LLC	http://www.pymol.org
CueMol	Version 2.2.2.397	http://www.cuemol.org
Other		
Amicon concentrators (10K)	Millipore	Cat#UFC901024
HisTrap FF (5 mL)	GE Healthcare	Cat#17-5255-01
HiTrap Heparin HP (5 mL)	GE Healthcare	Cat#17-0407-03
HiTrap Q Sepharose FF (5 mL)	GE Healthcare	Cat#17-5156-01
HiLoad 16/600 Superdex 200 pg	GE Healthcare	Cat#28989335
Superdex 200 10/300 GL	GE Healthcare	Cat#17517501
10% TBE-Urea gels	ThermoFisher	Cat#EC68752BOX
Whatman Elutrap electroelution system	Sigma-Aldrich	Cat#WHA10447700

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents could be directed to, and will be fulfilled by Lead Contact Dinshaw Patel (pateld@mskcc.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plasmid DNA for in vitro transcription was amplified in *Escherichia coli* DH5α strain in Lysogeny broth (LB) medium at 37°C overnight.

Recombinant proteins were overexpressed in *Escherichia coli* BL21 (DE3) strain in LB medium or M9 for seleonomethionine-derivatized proteins. The cells were grown at 37°C until OD600 reached 0.8 and then induced with 0.25 mM isopropyl β-D-1-thiogalactopyranoside (GoldBio) at 18°C for 20 hr.

METHOD DETAILS

Protein Expression and Purification

The gene encoding full-length *Alicyclobacillus acidoterrestris* C2c1 was synthesized and inserted into a modified pRSF-Duet-1 vector (Novagen), in which AacC2c1 was attached with N-terminal His6-SUMO tag following an ubiquitin-like protease (ULP1) cleavage site.

The fusion protein was expressed in *Escherichia coli* BL21 (DE3) strain. Cells were harvested by centrifugation and frozen at −80°C until purification. Cell pellets was resuspended in buffer A (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 20 mM imidazole, 1 mM phenylmethylsulfonyl fluoride), lysed by the EmulsiFlex-C3 homogenizer (Avestin), and centrifuged at 16,000 rpm for 1 hr in a JA-20 fixed angle rotor (Avanti J-E series centrifuge, Beckman Coulter). The supernatant containing AacC2c1 protein was loaded to 5 mL Ni-NTA Fastflow column pre-equilibrated in buffer A. The column was washed with 20 column volumes of buffer A, and the recombinant protein was eluted with buffer A supplemented with 480 mM imidazole. The His6-SUMO tag was removed by ULP1 and during dialysis against buffer A and then separated by re-loading to Ni-NTA column. The flow-through fraction containing recombinant protein was further dialysed against buffer B (20 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM β-mercaptoethanol) and loaded on 5 mL HiTrap Heparin HP Sepharose pre-equilibrated in buffer B. Elution of recombinant protein was achieved by a linear gradient from 300 mM to 1 M NaCl in 20 column volumes. Fractions containing recombinant protein were concentrated in 10 kDa molecular mass cut-off Amicon concentrators and loaded on Superdex 200 16/60 column pre-equilibrated in buffer C (20 mM HEPES, pH 7.2, 300 mM NaCl, 2 mM $MgCl_2$, 2 mM DTT). The relevant fractions were concentrated to ~16 mg/ml and flash-frozen in liquid nitrogen and stored in −80°C. For the selenomethionine (SeMet) derivative proteins, the cells were grown in M9 medium supplemented with amino acids Lys, Thr, Phe, Leu, Ile, Val, and SeMet. Different mutations were generated based on PCR-based method. The SeMet substituted protein and mutants were purified by the same method as described above.

To assemble the AacC2c1-sgRNA binary complex, the purified single mutation E848A protein was mixed with the sgRNA at the molar ratio of 1:1.1 and incubated on ice for 30 min. To assemble the different AacC2c1-sgRNA-dsDNA ternary complexes, the purified triple mutation D570A/E848A/D977A protein was mixed with the sgRNA, and the dsDNA at a molar ratio of 1:1.1:1.5 and incubated on ice for 30 min. The reconstituted binary and ternary complexes were purified by gel filtration chromatography on a Superdex 200 10/300 column pre-equilibrated in buffer D (20 mM HEPES, pH 7.2, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT). The target and non-target DNA strands are purchased from IDT (Integrated DNA Technologies) and dissolved in buffer E consisting of 20 mM Tris-HCl, pH 7.5, 50 mM NaCl. The target and non-target DNA strands were mixed together in equal molar ratio. To assemble AacC2c1 in complex with excess ssDNA, the target and non-target DNA strands were mixed together with a molar ratio of 1:1.5. The mixtures of two DNA strands were denatured at 95°C for 5 min and then annealed by slowly cooling to room temperature.

Crystallization, Data Collection, and Structure Determination

Crystallization was performed using the hanging drop vapor diffusion method at 20°C. Crystals of AacC2c1-sgRNA binary complex were grown from drops consisting of 1 μL protein solution (about 8 mg/ml) and 1 μL reservoir solution containing 0.2 M lithium sulfate, 0.1 M HEPES (pH 7.5), and 25% PEG3350 (v/v). Crystals of SeMet substituted AacC2c1-sgRNA-DNA ternary complex with excess 8-mer non-target DNA strand were grown from drops consisting of 1 μL protein solution (about 8 mg/ml) and 1 μL reservoir solution containing 0.2 M ammonium acetate, 0.1 M HEPES (pH 7.5), and 25% PEG3350 (v/v). Crystals of native AacC2c1-sgRNA-extended target DNA ternary complex and AacC2c1-sgRNA-extended non-target DNA ternary complex were grown in the same condition as that obtained for crystals of AacC2c1-sgRNA-DNA ternary complex containing excess 8-mer non-target DNA strand. The crystals were cryopretected by the reservoir solution supplemented with 20% ethylene glycol. All the datasets were collected at 100 K at the Advanced Photo Source (APS) at the Argonne National Laboratory. The diffraction data were processed with the NECAT RAPD online server. The statistics of the diffraction data are summarized in Table S1.

The structure of AacC2c1-sgRNA-DNA ternary complex containing excess 8-mer non-target DNA strand was solved by the single-wavelength anomalous dispersion (SAD) method using Phenix (Adams et al., 2010). The other structures of AacC2c1-sgRNA binary complex and AacC2c1-sgRNA-DNA ternary complexes containing extended target and non-target DNA strands were solved by the molecular replacement (MR) method using Phenix. Model building was performed using Coot (Emsley et al., 2010). The structure model was refined using Phenix (Adams et al., 2010). The statistics of the structure refinement and the quality of the final structure model are also summarized in Table S1. All molecular graphics were generated by PyMOL (http://www.pymol.org) and CueMol (http://www.cuemol.org).

In vitro Transcription and Purification of sgRNA

The sgRNA followed by the hammerhead ribozyme was transcribed in vitro using T7 RNA polymerase. Large scale transcription reaction (20 ml) was performed in buffer 100 mM Tris-HCl, pH 7.9, 30 mM DTT, 15 mM $MgCl_2$, 2 mM spermidine, 4 mM each NTP, 50 μg/ml DNA template, 2.5 μg home-made T7 RNA polymerase. The mixture was incubated at 37°C for 3 hr and then supplemented by $MgCl_2$ to final concentration of 50 mM and incubated for another 30 min. The transcribed sgRNA was purified by 10% denaturing TBE-urea PAGE, extracted from gel by electroelution using Elutrap, and then further purified by ion-exchange using HiTrap Q Fastflow Sepharose column pre-equilibrated by buffer E (20 mM Tris-HCl, pH 7.0). Elution of sgRNA was achieved by a linear gradient from 0 mM to 1 M NaCl in 20 column volumes. The RNA was denatured at 95°C for 5 min and slowly cooling to room temperature.

Template of sgRNA for in vitro transcription (from 5′ to 3′):

GTCTAGAGGACAGAATTTTTCAACGGGTGTGCCAATGGCCACTTTCCAGGTGGCAAAGCCCGTTGAGCTTCTCAAATCTGAGAAGTGGCACCAGAACCGGAGGACAAAGTC.

In vitro cleavage assay

The 646 bp target DNA containing human EMX gene and the 5′-TTC-3′ PAM was synthesized and amplified by PCR. The cleavage reaction was performed by mixing 200 ng of target DNA, an equimolar ratio of sgRNA and purified AacC2c1 proteins, and in cleavage buffer (NEB buffer3 and 5 mM DTT). Reaction was cleaned up by PCR purification columns (Roche) and then run in 10% TBE Urea gel or is used for Sanger sequencing.

Target DNA sequence (from 5′ to 3′):

ACCCATGGGAGCAGCTGGTCAGAGGGGACCCCGGCCTGGGGCCCCTAACCCTATGTAGCCTCAGTCTTCCCATCAGGCTCTC
AGCTCAGCCTGAGTGTTGAGGCCCCAGTGGCTGCTCTGGGGGCCTCCTGAGTTTCTCATCTGTGCCCCTCCCTCCCTGGCCCAG
GTGAAGGTGTGG**TTC**CAGAACCGGAGGACAAAGTCCAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAA
GAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACTA
GGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGC
CACTCCCTGGCCAGGCTTTGGGGAGGCCTGGAGTCATGGCCCCACAGGGCTTGAAGCCCGGGGCCGCCATTGACAGAGGGACA
AGCAATGGGCTGGCTGAGGCCTGGGACCACTTGGCCTTCTCCTCGGAGAGCCTGCCTGCCTGGGCGGGCCCGCCCGCCACCGC
AGCCTCCCAGCTGCTCTCCGTGTCTCCAATCTCCCTTTTGTTTTGATGCATTTCTGTTTTAATT

PAM is highlighted in bold, and the protospacer is underlined.

QUANTIFICATION AND STATISTICAL ANALYSIS

In vitro cleavage and Sanger sequencing experiments were repeated three times, and representative results were shown.

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates have been deposited in the Protein Data Bank with accession number PDB: 5U34 (AacC2c1-sgRNA binary complex), PDB: 5U31 (AacC2c1-sgRNA-DNA ternary complex containing excess 8-mer non-target DNA strand), PDB: 5U30 (AacC2c1-sgRNA-extended target DNA ternary complex), and PDB: 5U33 (AacC2c1-sgRNA-extended non-target DNA ternary complex).

Cell

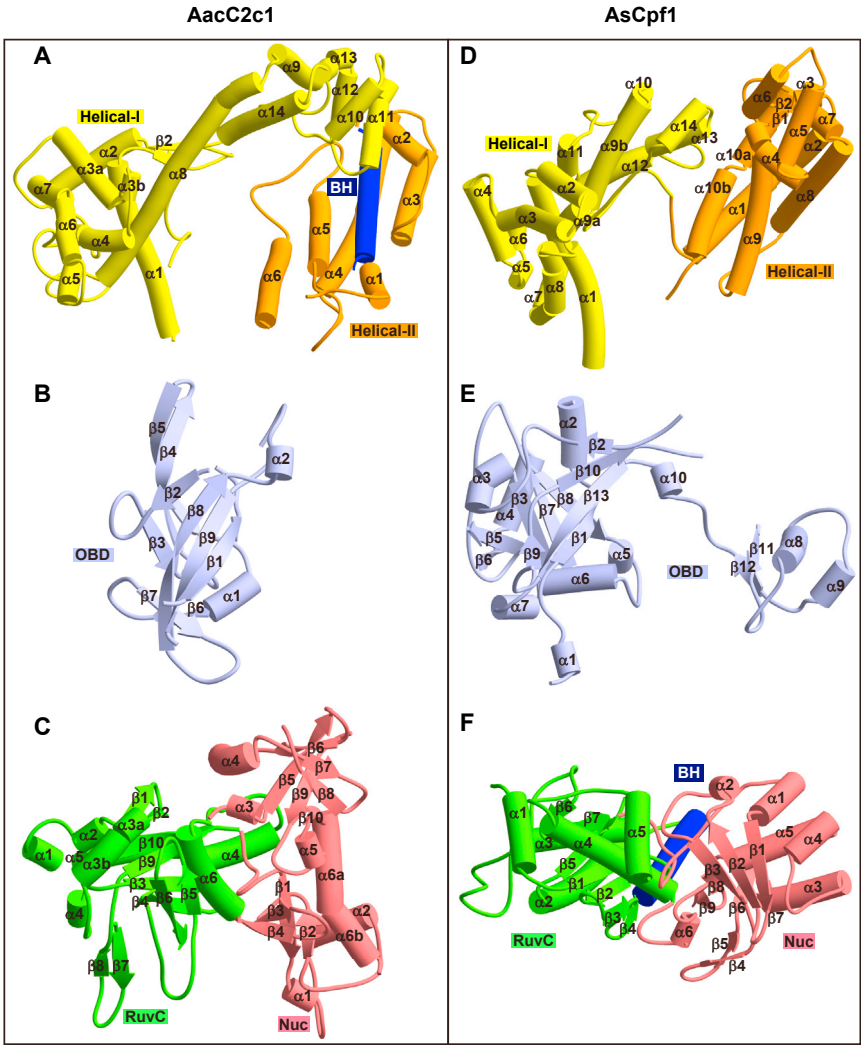


Figure S1. Structural Comparison of Helical-I, Helical-II, OBD, RuvC, and Nuc Domains in Ternary Complexes of AacC2c1 and AsCpf1, Related to Figure 1

(A–C) Ribbon diagrams of Helical-I and Helical-II domains (A), OBD domain (B), RuvC, and Nuc domains (C) in ternary complex of AacC2c1.
(D–F) Ribbon diagrams of Helical-I and Helical-II domains (D), OBD domain (E), RuvC, and Nuc domains (F) in ternary complex of in AsCpf1.
The color-code is the same as defined in Figure 1.

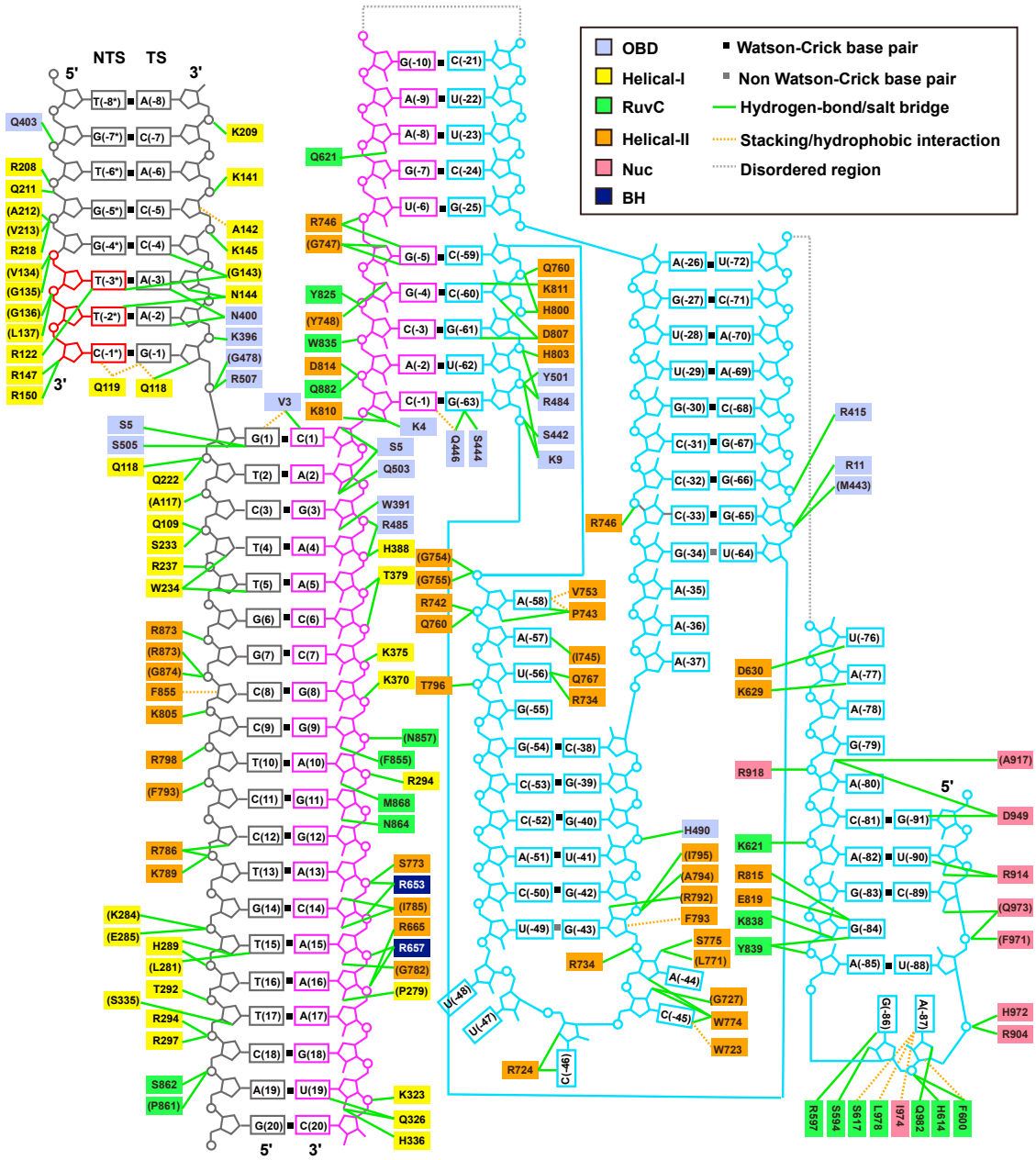


Figure S2. Schematic of Nucleic Acid Recognition by AacC2c1 in Ternary Complex, Related to Figures 2 and 3

Intermolecular contacts between C2c1 and sgRNA and target DNA strand. Hydrogen bonds and salt bridges are indicated with green lines. Hydrophobic and stacking interactions are shown by dashed orange lines. The interactions between substrate DNA strand and the RuvC and Nuc domain are described below.

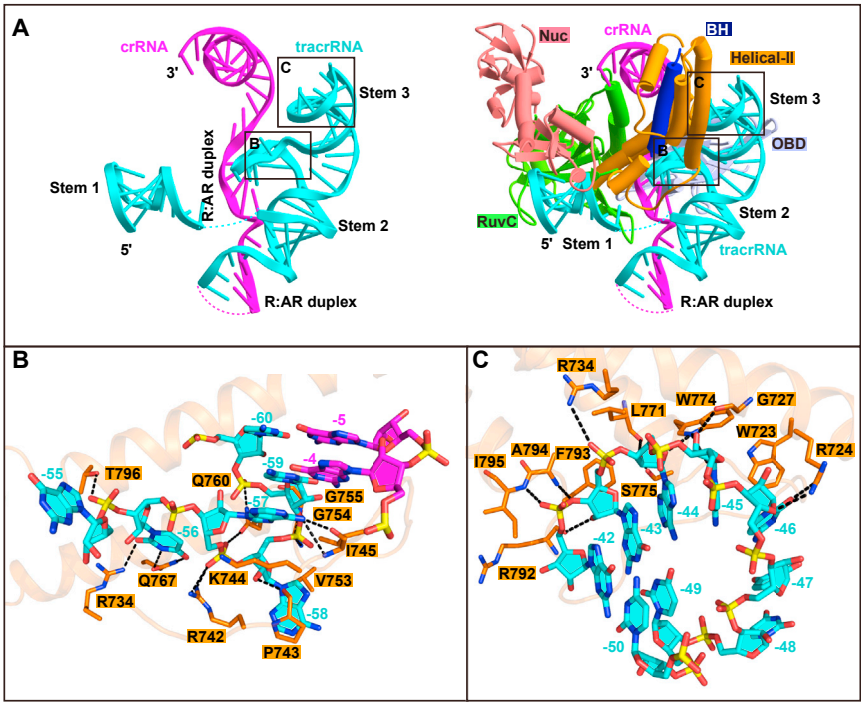


Figure S3. Detailed Interactions of C2c1 with sgRNA in Ternary Complex, Related to Figure 2

(A) Ribbon diagram of sgRNA (left panel) and interaction with OBD, Helical-II, RuvC and Nuc domains (right panel).

(B) Recognition of the linker region A(−58) to G(−55) between R:AR duplex-2 and stem 3 of the tracrRNA by the Helical II domain of C2c1.

(C) Recognition stem 2 of sgRNA by Helical-II domain of C2c1.

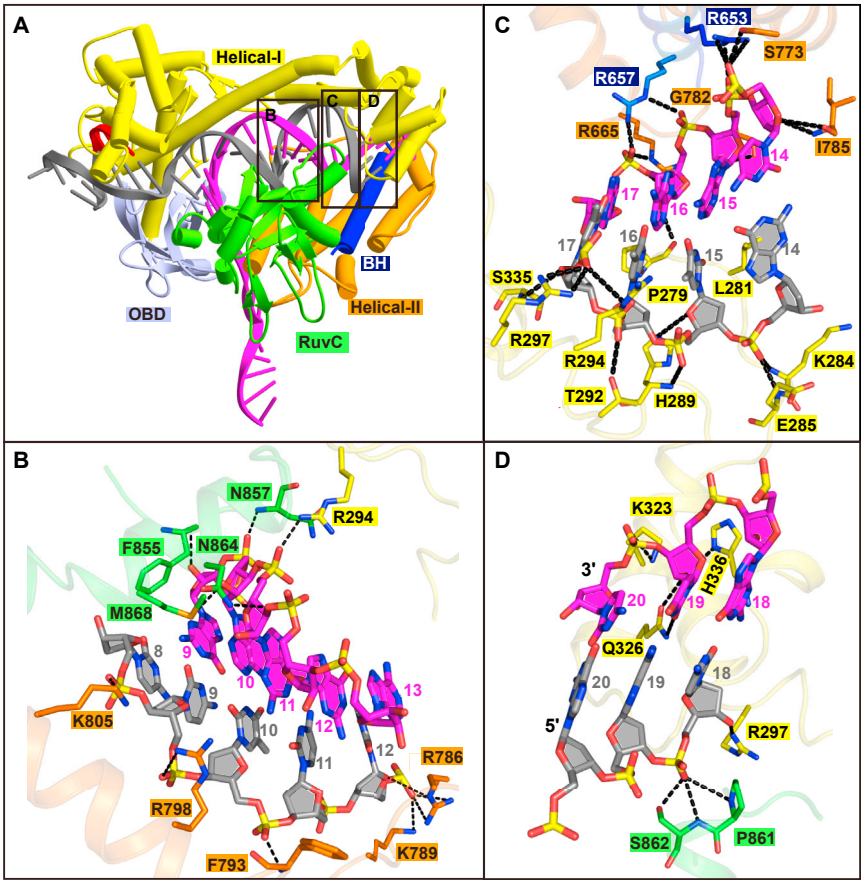


Figure S4. Detailed Interactions of C2c1 with Guide:Target Heteroduplex in Ternary Complex, Related to Figure 2

(A) Ribbon diagram of interaction of the guide:target heteroduplex by the BH motif and the OBD, Helical-I, Helical-II and RuvC domains.

(B) Recognition the dC8 to dT13/G8 to A13 base pairs of the guide:target heteroduplex by the Helical-II and RuvC domains.

(C) Recognition the dG14 to dT17/C14 to A17 base pairs of the guide:target heteroduplex by the BH motif, Helical-I and Helical-II domains.

(D) Recognition the dC18 to dG20/G18 to C20 base pairs of the guide:target heteroduplex by the Helical-I and RuvC domains.

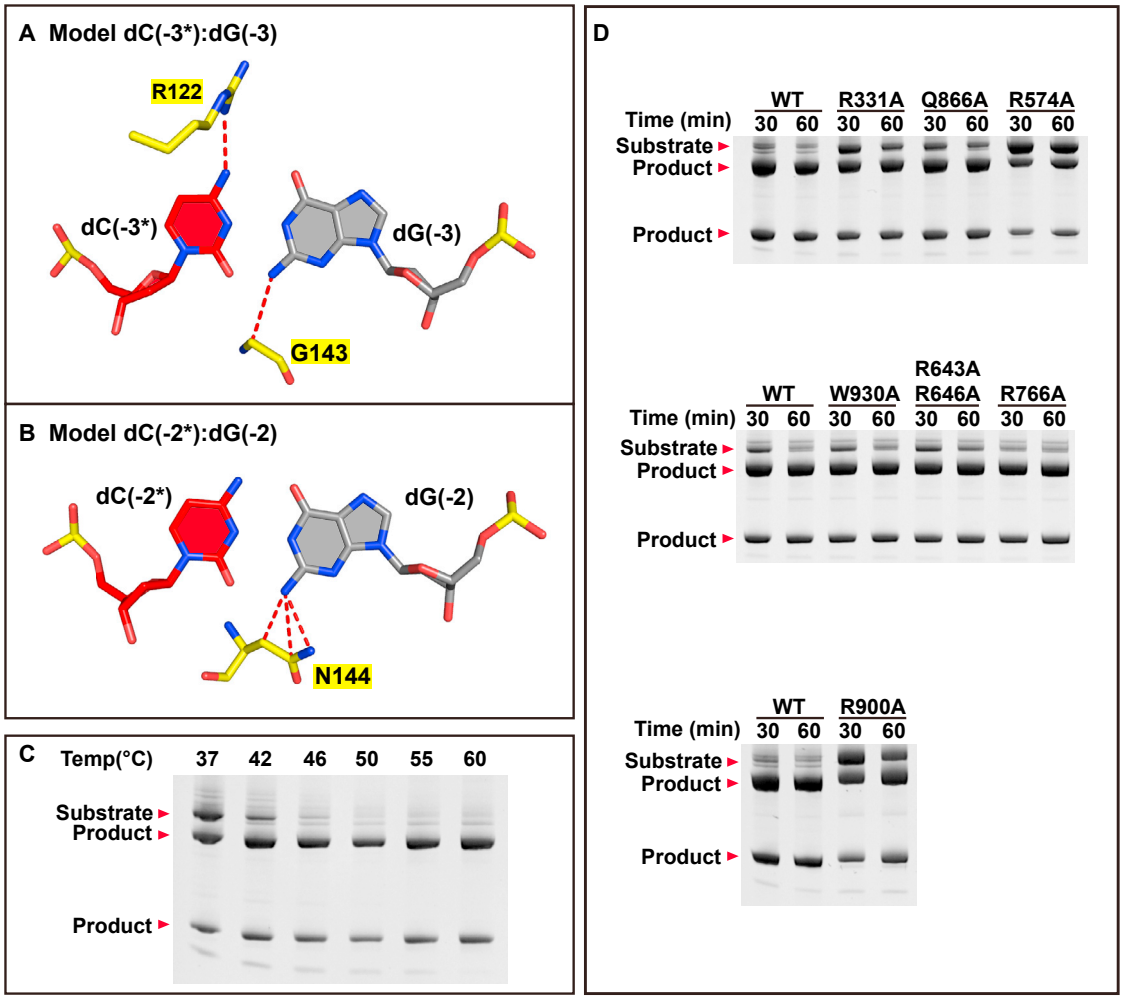


Figure S5. Recognition of the PAM Duplex and Temperature Dependence of Cleavage Activities, Related to Figure 3

(A) The modeled dC(−3*):dG(−3) base pair would result in steric clashes with Arg122 and Gly143.

(B) The modeled dC(−2*):dG(−2) base pair would result in steric clashes with Asn144.

(C) Temperature dependence of the cleavage activity of AacC2c1 associated with sgRNA in vitro.

(D) Mutational analysis of RuvC and Nuc domains positioned opposite the kinked site in the extended DNA target strand.

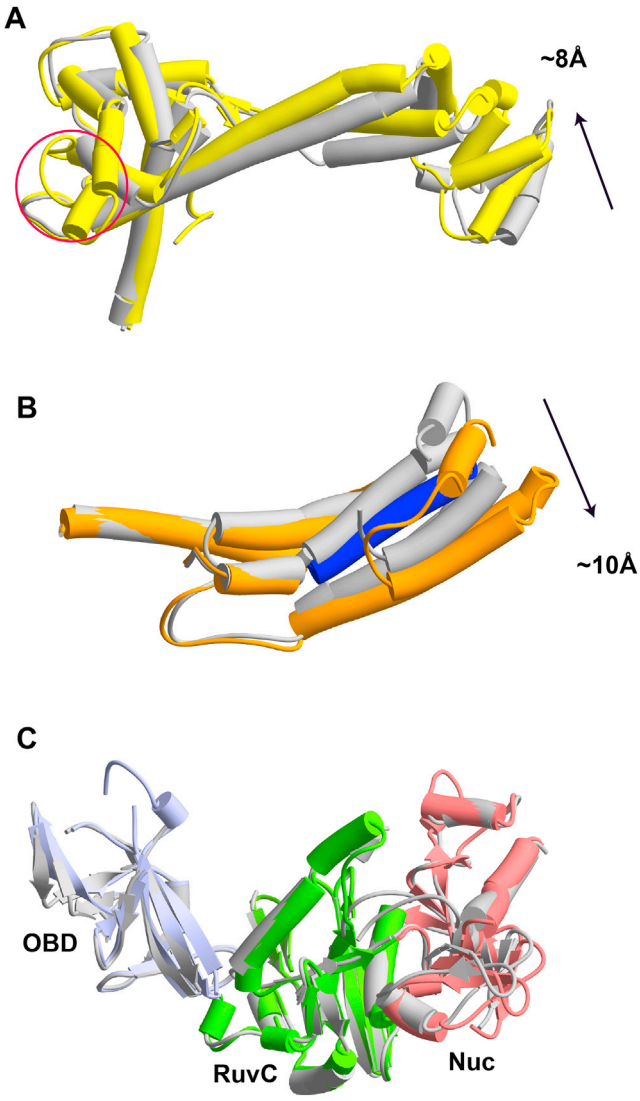
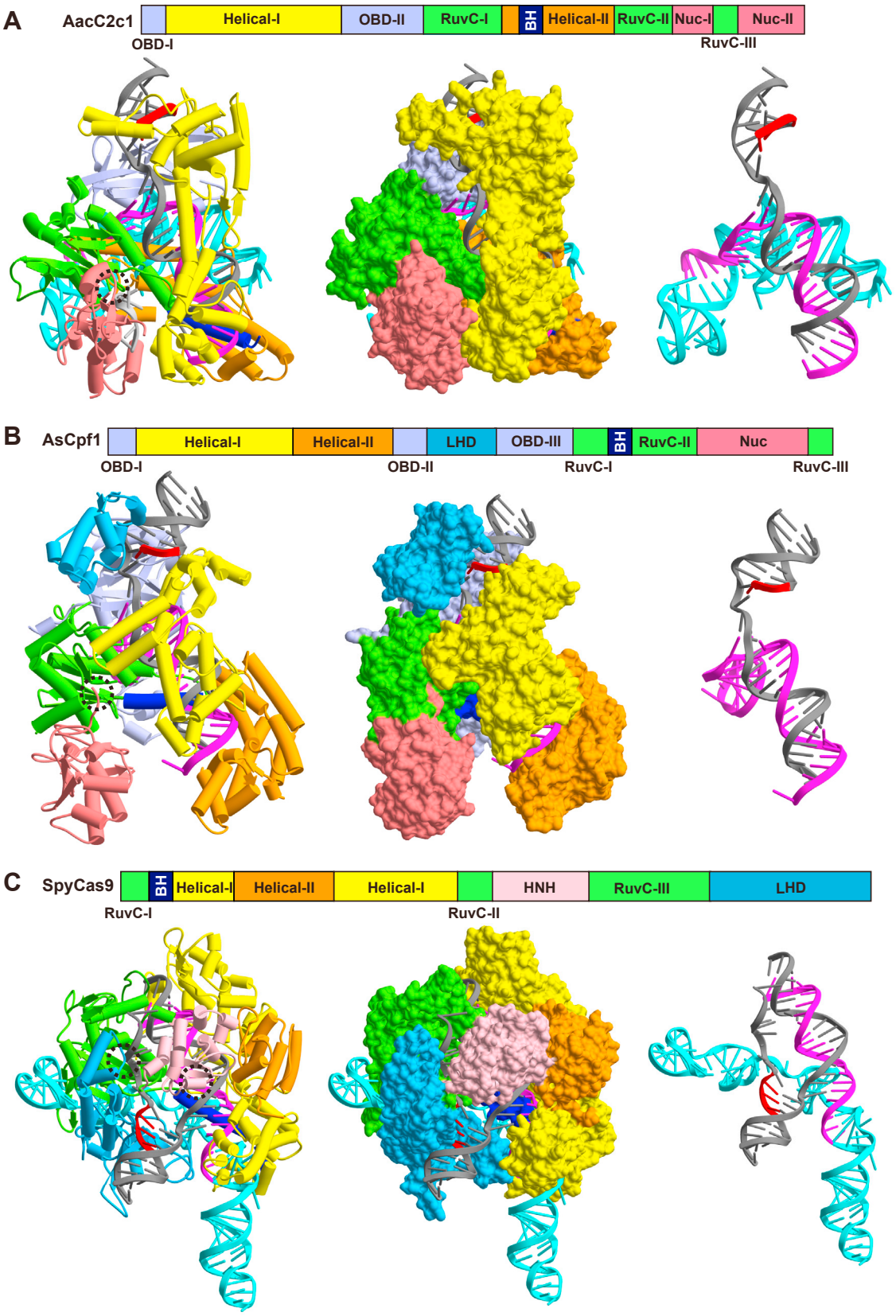


Figure S6. Domain Movements in AacC2c1-sgRNA Binary Complex upon Ternary Complex Formation with Target DNA, Related to Figure 6

(A and B) Structural movements of Helical-I (panel A) and Helical-II (panel B) between AacC2c1-sgRNA binary (in silver) and AacC2c1-sgRNA-DNA ternary (same color-code as defined in Figure 1) complexes. The domain movements are indicated by black arrows.

(C) Structural comparison of the OBD, RuvC and Nuc domains between AacC2c1-sgRNA binary (in silver) and AacC2c1-sgRNA-DNA ternary (same color-code as defined in Figure 1) complexes.



(legend on next page)

Figure S7. Comparison between AacC2c1, AsCpf1, and SpyCas9 Endonuclease Ternary Complexes, Related to Figure 7

(A–C) Domain organization, ribbon diagram, surface representation, and topology of RNA-DNA in AacC2c1 (panel A), AsCpf1 (panel B), and SpyCas9 (panel C) endonuclease ternary complexes.

See also Figure 7.