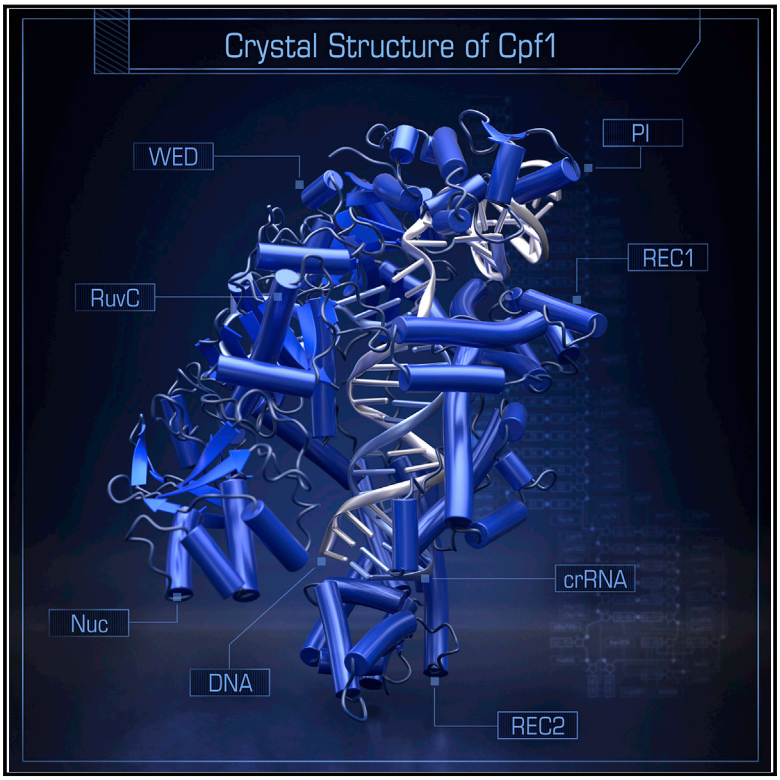
Article

Cell

Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA

Graphical Abstract



Authors

Takashi Yamano, Hiroshi Nishimasu, Bernd Zetsche, ..., Ryuichiro Ishitani, Feng Zhang, Osamu Nureki

Correspondence

zhang@broadinstitute.org (F.Z.), nureki@bs.s.u-tokyo.ac.jp (O.N.)

In Brief

The structure of Cpf1, a type V CRISPR-Cas effector nuclease, in complex with crRNA and its target DNA provides mechanistic insights into RNA-guided DNA cleavage by Cpf1 and establishes a framework for rational engineering of the CRISPR-Cpf1 toolbox.

Highlights

- Crystal structure of *Acidaminococcus sp.* Cpf1 in complex with crRNA and target DNA
- Mechanistic insights into Cpf1-induced, staggered DNA double-strand breaks
- Recognition of the 5′-TTTN-3′ PAM via base and shape readout mechanisms
- Striking similarity and major differences between the structures of Cpf1 and Cas9

Accession Numbers

5B43

Yamano et al., 2016, Cell *165*, 949–962
May 5, 2016 ©2016 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.04.003

CellPress

Cell

Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA

Takashi Yamano,[1,11] Hiroshi Nishimasu,[1,2,11] Bernd Zetsche,[3,4,5,6,7] Hisato Hirano,[1] Ian M. Slaymaker,[3,4,5,6] Yinqing Li,[3,4,5,6] Iana Fedorova,[3,4,5,6,8,9] Takanori Nakane,[1] Kira S. Makarova,[10] Eugene V. Koonin,[10] Ryuichiro Ishitani,[1] Feng Zhang,[3,4,5,6,*] and Osamu Nureki[1,*]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo 113-0032, Japan
[2]JST, PRESTO, Tokyo 113-0032, Japan
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]McGovern Institute for Brain Research
[5]Department of Brain and Cognitive Sciences
[6]Department of Biological Engineering
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Department of Developmental Pathology, Institute of Pathology, Bonn Medical School, 53127 Bonn, Germany
[8]Peter the Great St. Petersburg Polytechnic University, St. Petersburg, 195251, Russia
[9]Skolkovo Institute of Science and Technology, Skolkovo, 143026, Russia
[10]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA
[11]Co-first author
*Correspondence: zhang@broadinstitute.org (F.Z.), nureki@bs.s.u-tokyo.ac.jp (O.N.)
http://dx.doi.org/10.1016/j.cell.2016.04.003

SUMMARY

Cpf1 is an RNA-guided endonuclease of a type V CRISPR-Cas system that has been recently harnessed for genome editing. Here, we report the crystal structure of *Acidaminococcus sp.* Cpf1 (AsCpf1) in complex with the guide RNA and its target DNA at 2.8 Å resolution. AsCpf1 adopts a bilobed architecture, with the RNA-DNA heteroduplex bound inside the central channel. The structural comparison of AsCpf1 with Cas9, a type II CRISPR-Cas nuclease, reveals both striking similarity and major differences, thereby explaining their distinct functionalities. AsCpf1 contains the RuvC domain and a putative novel nuclease domain, which are responsible for cleaving the non-target and target strands, respectively, and for jointly generating staggered DNA double-strand breaks. AsCpf1 recognizes the 5′-TTTN-3′ protospacer adjacent motif by base and shape readout mechanisms. Our findings provide mechanistic insights into RNA-guided DNA cleavage by Cpf1 and establish a framework for rational engineering of the CRISPR-Cpf1 toolbox.

INTRODUCTION

The microbial adaptive immune system CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) helps bacteria and archaea defend themselves against the invasion of foreign nucleic acids (Marraffini, 2015; Wright et al., 2016). The CRISPR-Cas systems encompass arrays of direct repeats that are separated by unique spacers derived from foreign DNA. The repeat arrays are transcribed into long transcripts (precursors of CRISPR RNAs), which are then processed to yield small CRISPR RNAs (crRNAs), consisting of a spacer and a portion of the adjacent direct repeat. The crRNAs form a complex with Cas endonucleases, and in some cases with accessory Cas proteins as well, and serve as guides to target and cleave the cognate foreign nucleic acid, thus achieving interference. DNA recognition by Cas-crRNA complexes requires the presence of a protospacer adjacent motif (PAM) near the target site, which contributes to self versus non-self discrimination (Westra et al., 2013). The diverse spectrum of the CRISPR-Cas systems is broadly divided into two classes, depending on the architecture of the interference module (Makarova et al., 2015): class 1 systems use a complex of several Cas proteins, as exemplified by Cascade (Brouns et al., 2008; Redding et al., 2015), and class 2 systems use a single enzyme, such as Cas9 (Jinek et al., 2012; Gasiunas et al., 2012). Cas9 is a dual RNA-guided endonuclease that recognizes, binds, and cleaves target DNA, and it has been harnessed to create precision genome engineering tools (Cong et al., 2013; Mali et al., 2013).

Following the initial demonstration of the feasibility of using Cas9 to edit mammalian genomes, there was a burst of efforts to further adapt this endonuclease for a range of applications, from high-throughput gain-of-function screening (Gilbert et al., 2014; Konermann et al., 2015) to targeted modulation of histone marks (Hilton et al., 2015; Kearns et al., 2015). The development of these applications has been furthered, in part, through rational engineering, made possible by extensive biochemical and biophysical studies and the availability of several crystal structures of Cas9 (Nishimasu et al., 2014; Jinek et al., 2014; Anders et al., 2014; Nishimasu et al., 2015; Jiang et al., 2015, 2016; Hirano et al., 2016). Structure-guided engineering and direct evolution approaches have led to variants of Cas9 with enhanced target specificity (Slaymaker et al., 2016; Kleinstiver et al., 2016) or altered PAM requirements (Kleinstiver et al., 2015a, 2015b).

Recently, a second class 2 (type V) effector protein, Cpf1, has been harnessed for genome editing (Zetsche et al., 2015). Similar to Cas9, Cpf1 can be reprogrammed to target DNA sites of interest through complementarity to a guide RNA. However, Cpf1 possesses several unique features that distinguish it from Cas9 and could provide for a substantial expansion of the genome editing toolbox. First, Cpf1 is guided by a single crRNA, whereas Cas9 uses a crRNA and a second small RNA species, a *trans*-activating crRNA (tracrRNA) (Deltcheva et al., 2011). Second, Cpf1 recognizes a T-rich PAM, in contrast to the G-rich PAM favored by Cas9 (Fonfara et al., 2014; Karvelis et al., 2015). Third, Cpf1 generates staggered ends in its PAM-distal target site (Zetsche et al., 2015), whereas Cas9 creates blunt ends within the PAM-proximal target site (Garneau et al., 2010). Fourth, Cpf1 contains the RuvC domain but lacks a detectable second endonuclease domain (Zetsche et al., 2015), whereas Cas9 uses the HNH and RuvC endonuclease domains to cleave the target and non-target DNA strands, respectively (Jinek et al., 2012; Gasiunas et al., 2012). Together, these observations imply major differences in the target DNA recognition and cleavage mechanisms between Cas9 and Cpf1.

To clarify how Cpf1 recognizes and cleaves DNA targets, we determined the crystal structure of *Acidaminococcus sp.* Cpf1 (AsCpf1) in complex with the crRNA and its double-stranded DNA target containing the 5′-TTTN-3′ PAM. AsCpf1 adopts a bilobed architecture that accommodates the crRNA-target DNA heteroduplex in the central channel. AsCpf1 recognizes the crRNA scaffold and the 5′-TTTN-3′ PAM in structure- and sequence-dependent manners. AsCpf1 contains a RuvC endonuclease domain and a putative novel nuclease domain, which are located at positions suitable to induce staggered DNA double-strand breaks. The structural comparison of AsCpf1 with Cas9 reveals both striking structural similarity and substantial differences between the two class 2 effector proteins, thus explaining their distinct functionalities and suggesting their functional convergence.

RESULTS

Overall Structure of the AsCpf1-crRNA-Target DNA Complex

We solved the 2.8-Å resolution crystal structure of the full-length AsCpf1 (residues 1–1307) in complex with a 43-nt crRNA, a 34-nt target DNA strand, and a 10-nt non-target DNA strand containing a 5′-TTTN-3′ PAM, by the single-wavelength anomalous diffraction (SAD) method (Figures 1 and S1 and Table S1). The structure revealed that AsCpf1 adopts a bilobed architecture consisting of an α-helical recognition (REC) lobe and a nuclease (NUC) lobe, with the crRNA-target DNA heteroduplex bound to the positively charged, central channel between the two lobes (Figures 1C, 1D, and S2). The REC lobe consists of the REC1 and REC2 domains, whereas the NUC lobe consists of the RuvC domain and three additional domains, denoted A, B, and C (Figure 1C).

A Dali search (Holm and Rosenström, 2010) detected no structural similarity between the REC1, REC2, and the A, B, and C domains and any of the available protein structures. Sequence database searches using PSI-BLAST (Altschul et al., 1997) and HHPred (Söding et al., 2005) also failed to detect significant similarity between these domains and any protein sequences in the current databases. Thus, these Cpf1 domains have no detectable homologs outside the Cpf1 protein family and appear to adopt novel structural folds (Figures 1C and S3). The REC1 domain comprises 13 α helices, and the REC2 domain comprises ten α helices and two β strands that form a small antiparallel sheet (Figures S3A and S3B). Domains A and B play functional roles similar to those of the WED (Wedge) and PI (PAM-interacting) domains of Cas9 (Anders et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016), respectively, although the two domains of AsCpf1 are structurally unrelated to the WED and PI domains (described below). Domain C is involved in DNA cleavage (described below). Thus, domains A, B, and C are referred to as the WED, PI, and Nuc domains, respectively. The WED domain is assembled from three separate regions (WED-I-III) in the Cpf1 sequence (Figures 1A, S3A, and S3C). The WED domain can be divided into a core subdomain comprising a nine-stranded, distorted antiparallel β sheet (β1-β8 and β11) flanked by seven α helices (α1-α6 and α9) and a subdomain comprising two β strands (β9 and β10) and two α helices (α7 and α8) (Figures S3A and S3C). Examination of the Cpf1 sequence alignment revealed that helices α7 and α8 are not conserved among Cpf1 homologs (Zetsche et al., 2015) (Figure S4). The PI domain comprises seven α helices (α1-α7) and a β hairpin (β1 and β2) and is inserted between the WED-II and WED-III regions, whereas the REC lobe is inserted between the WED-I and WED-II regions (Figures 1A, S3A, and S3B). As discussed previously (Zetsche et al., 2015), the RuvC domain contains the three motifs (RuvC-I-III) that form the endonuclease active center. A characteristic helix (referred to as the bridge helix) is located between the RuvC-I and RuvC-II motifs and connects the REC and NUC lobes (described below) (Figures 1A, 1C, and 1D). The Nuc domain is inserted between the RuvC-II and RuvC-III motifs.

Structure of the crRNA and Target DNA

The crRNA consists of the 24-nt guide segment (G1–C24) and the 19-nt scaffold (A(−19)–U(−1)) (referred to as the 5′ handle) (Figures 2A and 2B). The nucleotides G1–C20 in the crRNA and dC1–dG20 in the target DNA strand form the 20-bp RNA-DNA heteroduplex (Figures 2A and 2B). The nucleotide A21 in the crRNA is flipped out and adopts a single-stranded conformation. No electron density was observed for the nucleotides A22–C24 in the crRNA and dT21–dG24 in the target DNA strand, suggesting that these regions are flexible and disordered in the crystal structure. The nucleotides dG(−10)–dT(−1) in the target DNA strand and dC(−10*)–dA(−1*) in the non-target DNA strand form a duplex structure (referred to as the PAM duplex) (Figures 2A and 2B).

The crystal structure revealed that the crRNA 5′ handle adopts a pseudoknot structure, rather than a simple stem-loop structure predicted from its nucleotide sequence (Zetsche et al., 2015) (Figures 2A and 2C). Specifically, the G(−6)–A(−2) and U(−15)–C(−11) in the 5′ handle form a stem structure, via five Watson-Crick base pairs (G(−6):C(−11)–A(−2):U(−15)), whereas C(−9)–U(−7) in the 5′ handle adopt a loop structure. U(−1) and U(−16) form a non-canonical U•U base pair (Figure 2D). U(−10) and A(−18) form a reverse Hoogsteen A•U

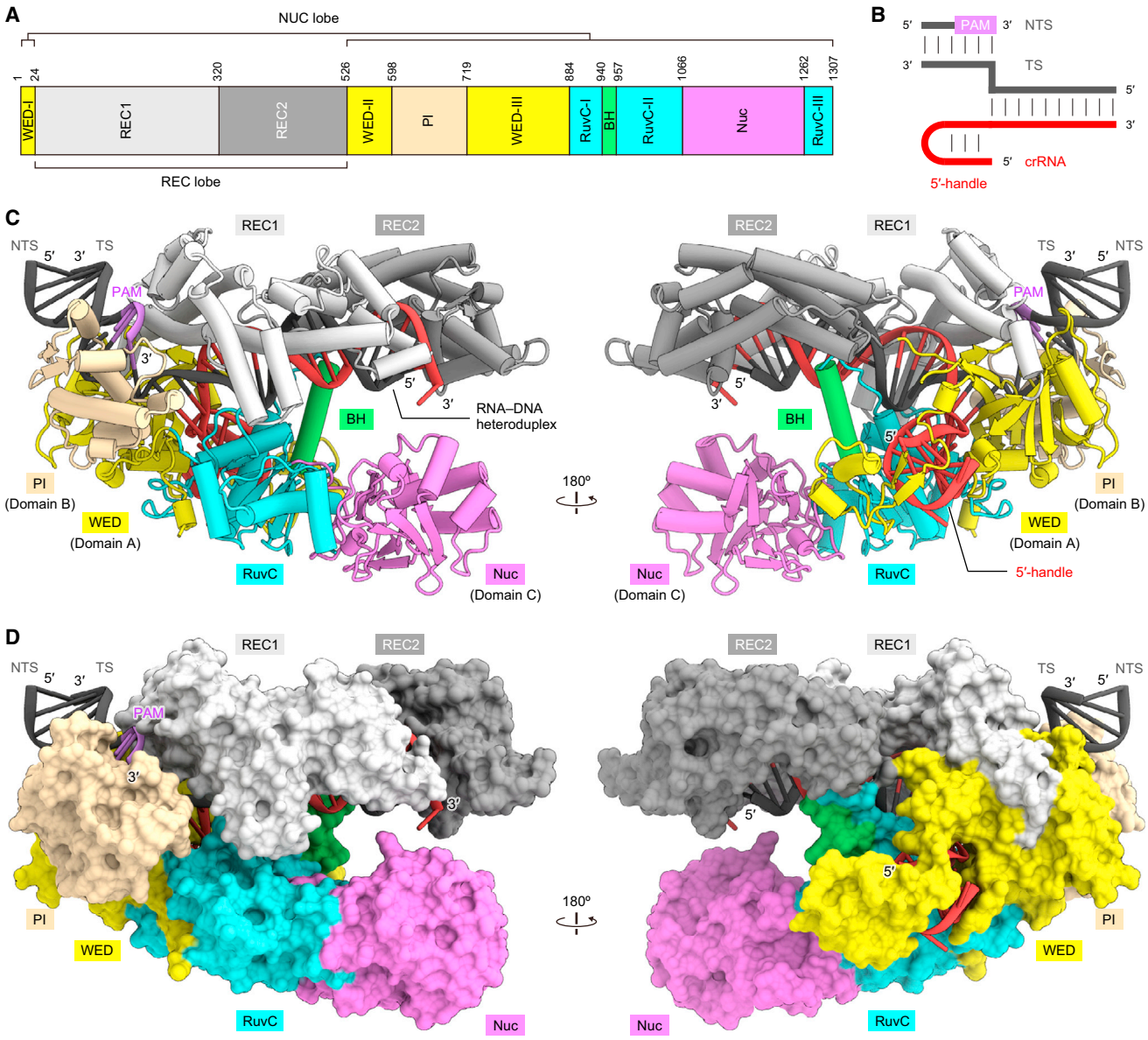


Figure 1. Overall Structure of the AsCpf1-crRNA-Target DNA Complex

(A) Domain organization of AsCpf1. BH, bridge helix.

(B) Schematic representation of the crRNA and target DNA. TS, target DNA strand; NTS, non-target DNA strand.

(C and D) Cartoon (C) and surface (D) representations of the AsCpf1-crRNA-DNA complex. Molecular graphic images were prepared using CueMol (http://www.cuemol.org).

See also Figures S1, S2, and S3 and Table S1.

base pair, and participate in pseudoknot formation (Figure 2E). The O4 and the 2′-OH of U(−10) hydrogen bond with the 2′-OH and the N1 of A(−19), respectively (Figure 2E). In addition, the N3 and the O4 of U(−17) hydrogen bond with the O4 of U(−13) and the N6 of A(−12), respectively, thereby stabilizing the pseudoknot structure (Figure 2F). Importantly, U(−1), U(−10), U(−16) and A(−18) in the crRNA are conserved among the CRISPR-Cpf1 systems (Zetsche et al., 2015), indicating that Cpf1 crRNAs form similar pseudoknot structures.

Recognition of the 5′ Handle of the crRNA

The 5′ handle of the crRNA is bound at the groove between the WED and RuvC domains (Figure 2G). The U(−1)•U(−16) base pair in the 5′ handle is recognized by the WED domain in a base-specific manner. U(−1) and U(−16) hydrogen bond with His761 and Arg18/Asn759, respectively, while U(−1) stacks on His761 (Figure 2H). These interactions explain the previous finding that the U•U base pair at this position is critical for the Cpf1-mediated DNA cleavage (Zetsche et al., 2015). The N6 of

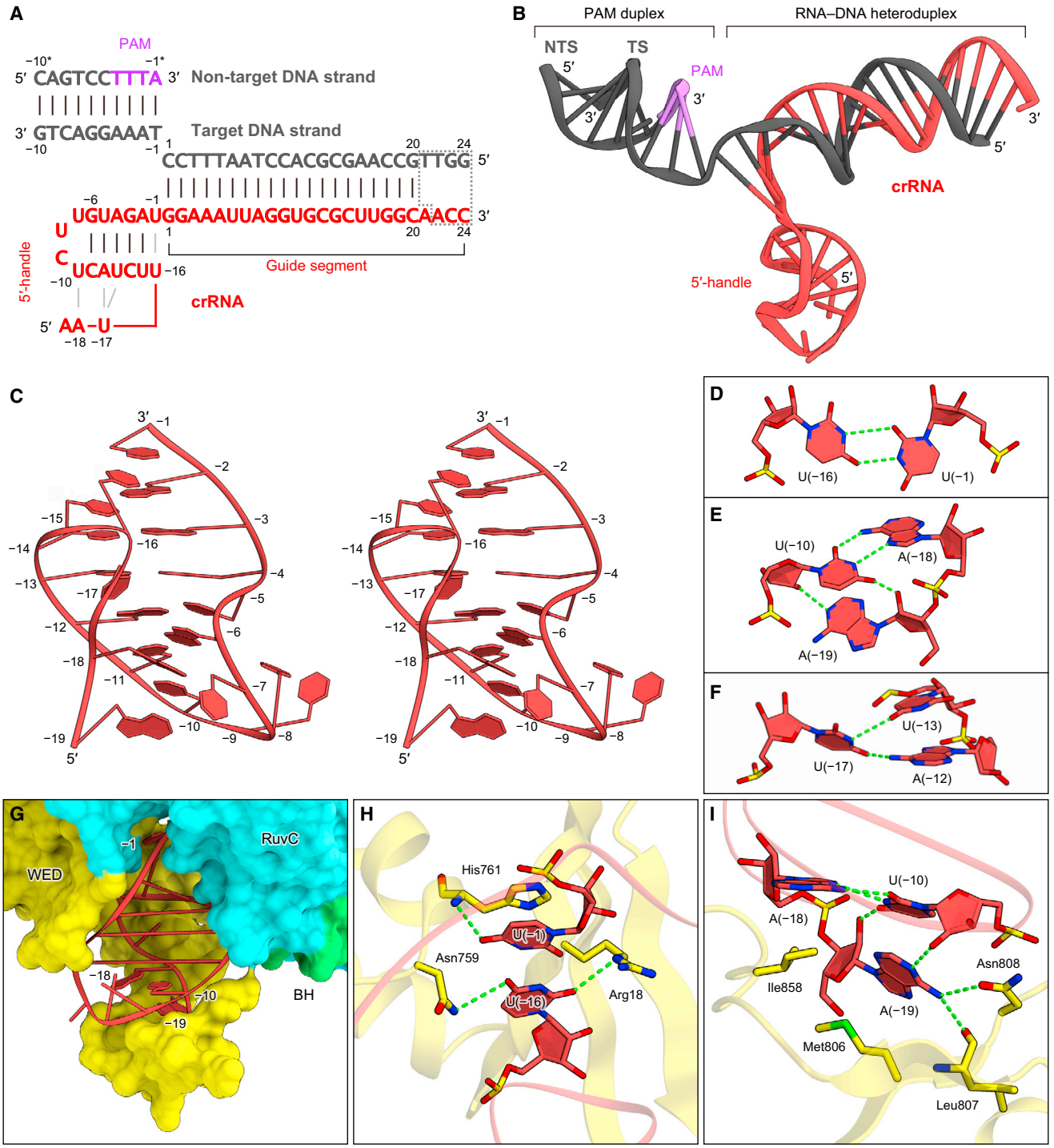


Figure 2. Structure of the crRNA and Target DNA

(A) Schematic representation of the AsCpf1 crRNA and the target DNA. The disordered region is surrounded by dashed lines.

(B) Structure of the AsCpf1 crRNA and the target DNA.

(C) Structure of the crRNA 5′ handle (stereo view).

(D–F) Close-up view of the U(−1)•U(−16) base pair (D), the reverse Hoogsteen U(−10)•A(−18) base pair (E), and the U(−13)-U(−17)-U(−12) base triple (F). Hydrogen bonds are shown as dashed lines.

(G) Binding of the crRNA 5′ handle to the groove between the WED and RuvC domains.

(H and I) Recognition of the 3′ end (H) and the 5′ end (I) of the crRNA 5′ handle. Hydrogen bonds are shown as dashed lines.

A(−19) hydrogen bonds with Leu807 and Asn808, while the base moieties of A(−18) and A(−19) form stacking interactions with Ile858 and Met806, respectively (Figure 2I). Moreover, the phosphodiester backbone of the 5′ handle forms an extensive network of interactions with the WED and RuvC domains (Figure 3). The residues involved in the crRNA 5′ handle recognition are largely conserved in the Cpf1 protein family (Zetsche et al., 2015) (Figure S4), highlighting the functional relevance of the observed interactions between AsCpf1 and the crRNA.

Recognition of the crRNA-Target DNA Heteroduplex

The crRNA-target DNA heteroduplex is accommodated within the positively charged, central channel formed by the REC1, REC2, and RuvC domains and is recognized by the protein in a sequence-independent manner (Figures 3, 4A, 4B, and S2). The PAM-distal and PAM-proximal regions of the heteroduplex are recognized by the REC1-REC2 domains and the WED-REC1-RuvC domains, respectively (Figures 3 and 4A–4C). Arg951 and Arg955 in the bridge helix, which interact with the sugar-phosphate backbone of the target DNA strand (Figure 4B), are conserved among the Cpf1 family members (Zetsche et al., 2015) (Figure S4). Notably, the sugar-phosphate backbone of the nucleotides G1–A8 in the crRNA forms multiple contacts with the WED and REC1 domains (Figures 3 and 4C), and the base pairing within the 5-bp PAM-proximal "seed" region is important for Cpf1-mediated DNA cleavage (Zetsche et al., 2015). These observations suggest that, in the Cpf1-crRNA complex, the seed of the crRNA guide is preordered in a nearly A-form conformation and serves as the nucleation site for pairing with the target DNA strand, as observed in the Cas9-sgRNA complex (Jiang et al., 2015). In addition, the backbone phosphate group between dT(−1) and dC1 of the target DNA strand (referred to as the +1 phosphate) is recognized by the side chain of Lys780 and the main-chain amide group of Gly783 (Figure 4C). This interaction results in the rotation of the +1 phosphate group, thereby facilitating base pairing between dC1 in the target DNA strand and G1 in the crRNA, as also observed in the Cas9-sgRNA-target DNA complexes (Anders et al., 2014; Nishimasu et al., 2015). The residues involved in the heteroduplex recognition are conserved in most members of the Cpf1 family (Zetsche et al., 2015) (Figure S4), and the R176A, R192A, G783P, and R951A mutants exhibited reduced activities (Figure 4D), confirming their functional relevance. Together, these observations reveal the RNA-guided DNA recognition mechanism of Cpf1.

Unexpectedly, the present structure revealed that the 24-nt crRNA guide and the target DNA strand form a 20-bp, rather than 24-bp, RNA-DNA heteroduplex (Figure 4A). The side chain of Trp382 in the REC2 domain forms a stacking interaction with the C20:dG20 base pair in the heteroduplex and thus prevents base pairing between A21 and dT21 (Figure 4E). Indeed, the W382A mutant showed reduced activity (Figure 4D), highlighting its functional importance. Trp382 is conserved in some members of the Cpf1 family, whereas others contain aromatic residues in this position (Zetsche et al., 2015) (Figure S4). These observations indicate that Cpf1 recognizes the 20-bp RNA-DNA heteroduplex and can explain the previous finding that the *Francisella novicida* Cpf1 (FnCpf1) cleaved the target DNA in a similar manner, using either the 20- or 24-nt guide-containing crRNA (Zetsche et al., 2015).

Recognition of the 5′-TTTN-3′ PAM

The PAM duplex adopts a distorted conformation with a narrow minor groove, as often observed in AT-rich DNA (Rohs et al., 2009), and is bound to the groove formed by the WED, REC1, and PI domains (Figures 5A and S5A). The PAM duplex is recognized by the WED-REC1 and PI domains from the major and minor groove sides, respectively (Figure 5B). The dT(−1):dA(−1*) base pair in the PAM duplex does not form base-specific contacts with the protein (Figure 5B), consistent with the lack of specificity in the fourth position of the 5′-TTTN-3′ PAM. Lys607 in the PI domain is inserted into the narrow minor groove and plays critical roles in the PAM recognition (Figure 5B). The O2 of dT(−2*) forms a hydrogen bond with the side chain of Lys607, whereas the nucleobase and deoxyribose moieties of dA(−2) form van der Waals interactions with the side chains of Lys607 and Pro599/Met604, respectively (Figure 5C). Modeling of the dG(−2):dC(−2*) base pair indicated that a steric clash exists between the N2 of dG(−2) and the side chain of Lys607 (Figure S5B), suggesting that dA(−2):dT(−2*), but not dG(−2):dC(−2*), is accepted at this position. These structural observations can explain the requirement of the third T in the 5′-TTTN-3′ PAM. The 5-methyl group of dT(−3*) forms a van der Waals interaction with the side-chain methyl group of Thr167, whereas the N3 and N7 of dA(−3) form hydrogen bonds with Lys607 and Lys548, respectively (Figure 5D). Modeling of the dG(−3):dC(−3*) base pair indicated that a steric clash exists between the N2 of dG(−3) and the side chain of Lys607 (Figure S5C). These observations are consistent with the requirement of the second T in the PAM. The 5-methyl group of dT(−4*) is surrounded by the side-chain methyl groups of Thr167 and Thr539, whereas the O4′ of dA(−4) forms a hydrogen bond with the side chain of Lys607 (Figure 5E). Notably, the N3 and O4 of dT(−4*) form hydrogen bonds with the N1 of dA(−4) and the N6 of dA(−3), respectively (Figure 5E). Modeling indicated that dA(−3) would sterically clash with the modeled base pairs dT(−4):dA(−4*), dG(−4):dC(−4*), and dC(−4):dG(−4*) (Figure S5D). These structural observations are consistent with the requirement of the first T in the PAM. The K548A and M604A mutants exhibited reduced activities (Figure 5F), confirming that Lys548 and Met604 participate in the PAM recognition. More importantly, the K607A mutant showed almost no activity (Figure 5F), indicating that Lys607 is critical for the PAM recognition. Together, these results demonstrate that AsCpf1 recognizes the 5′-TTTN-3′ PAM via a combination of base and shape readout mechanisms. Thr167 and Lys607 are conserved throughout the Cpf1 family, and Lys548, Pro599, and Met604 are partially conserved (Zetsche et al., 2015) (Figure S4). These observations indicate that the Cpf1 homologs from diverse bacteria recognize their T-rich PAMs in similar manners, although the fine details of the interaction could vary.

The RuvC-like Endonuclease and a Putative Second Nuclease Domain

The RuvC domain comprises a typical RNase H fold, consisting of a five-stranded mixed β sheet (β1–β5) flanked by three

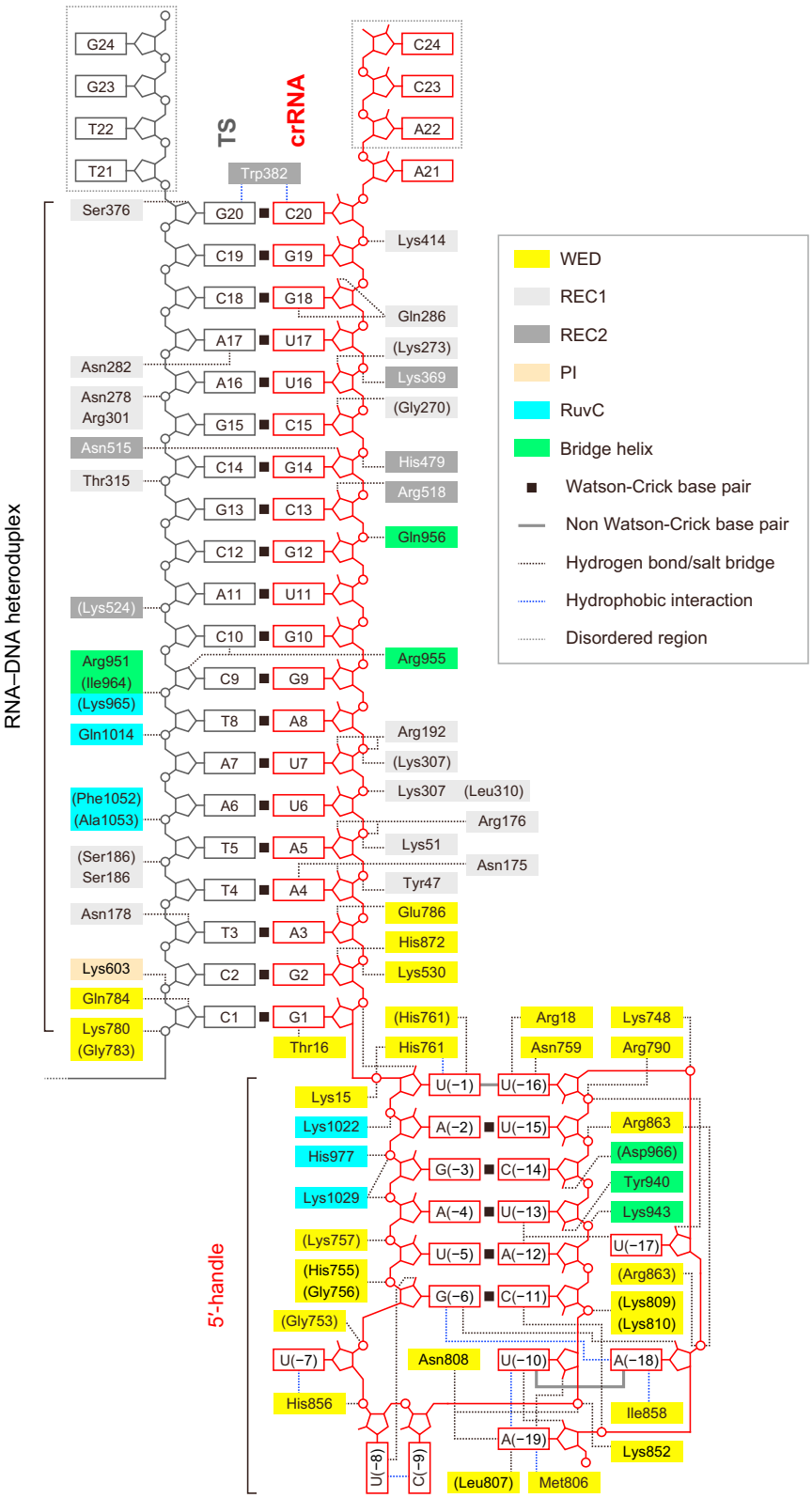


Figure 3. Schematic of Nucleic Acid Recognition by Cpf1

AsCpf1 residues that interact with the crRNA and the target DNA via their main chain are shown in parentheses. Water-mediated hydrogen-bonding interactions are omitted for clarity.
See also Figure S4.

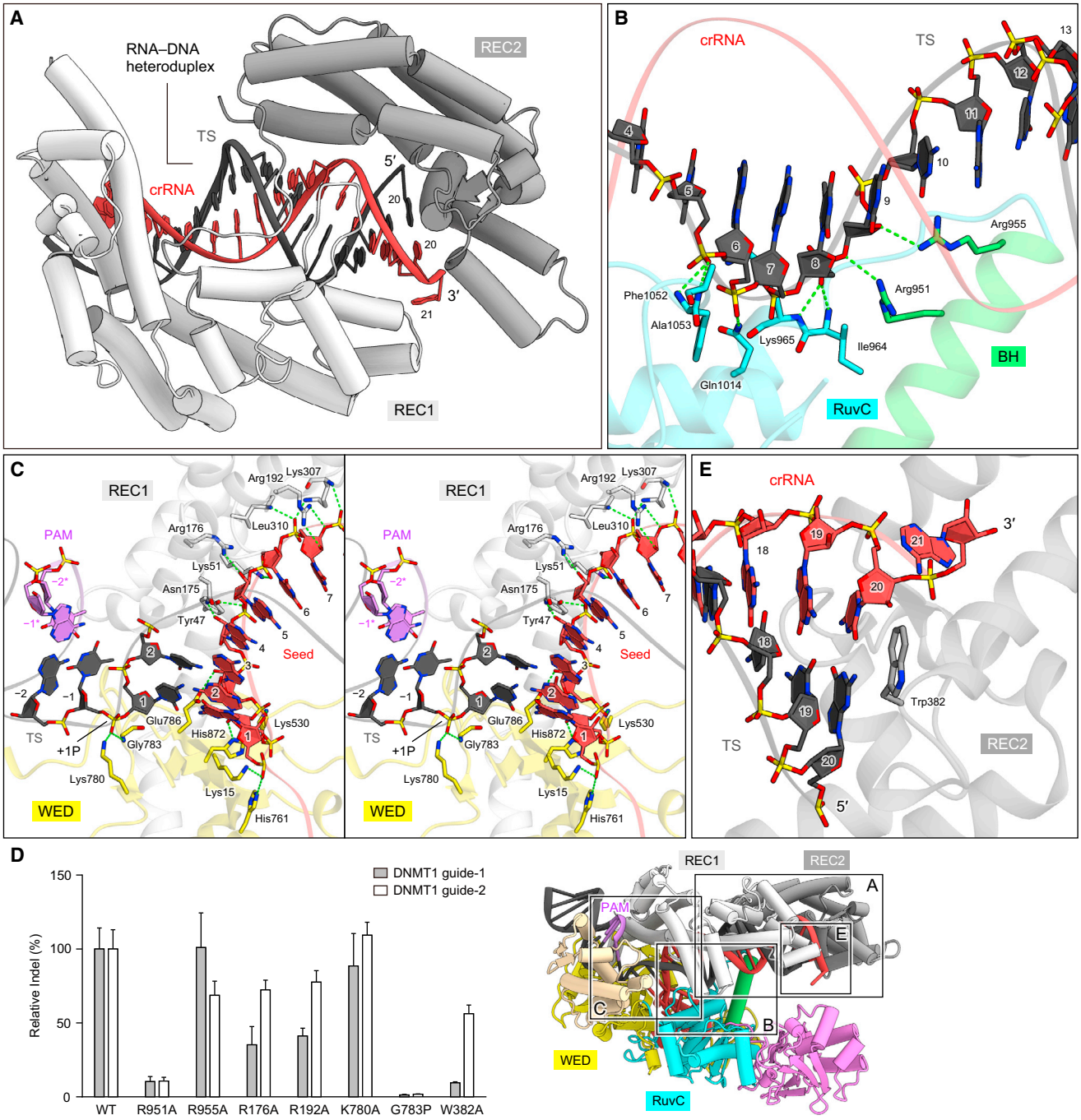


Figure 4. Recognition of the crRNA-Target DNA Heteroduplex

(A) Recognition of the crRNA-target DNA heteroduplex by the REC1 and REC2 domains.

(B) Recognition of the target DNA strand by the bridge helix and the RuvC domain. Hydrogen bonds are shown as dashed lines.

(C) Recognition of the crRNA seed region and the +1 phosphate group (+1P) (stereo view). Hydrogen bonds are shown as dashed lines.

(D) Mutational analysis of the nucleic-acid-binding residues. Effects of mutations on the ability to induce indels at two *DNMT1* targets were examined (n = 3, error bars show mean ± SEM).

(E) Stacking interaction between the 20th base pair in the heteroduplex and Trp382 of the REC2 domain.

α helices (α1–α3), and two additional α helices and three β strands (Figure 6A). The conserved, negatively charged residues Asp908, Glu993, and Asp1263 form an active site similar to that of the Cas9 RuvC domain (Nishimasu et al., 2014; Anders et al., 2014) (Figure 6B). As observed in FnCpf1 (Zetsche et al., 2015), the D908A and E993A mutants had almost no activity, whereas

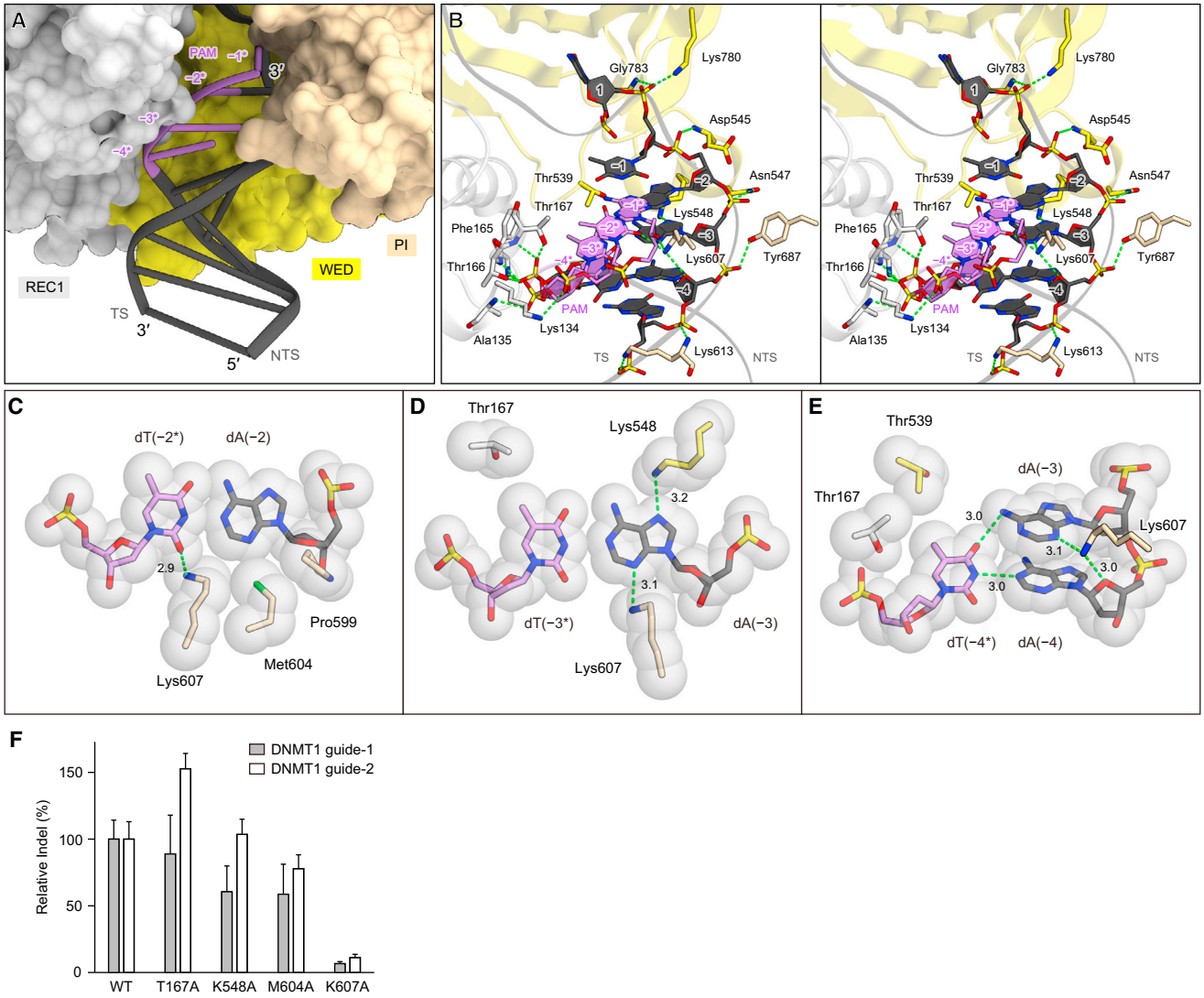


Figure 5. Recognition of the 5′-TTTN-3′ PAM

(A) Binding of the PAM duplex to the groove between the WED, REC1, and PI domains.

(B) Recognition of the 5′-TTTN-3′ PAM (stereo view). Hydrogen bonds are shown as dashed lines.

(C–E) Recognition of the dA(−2):dT(−2*) (C), dA(−3):dT(−3*) (D), and dA(−4):dT(−4*) (E) base pairs.

(F) Mutational analysis of the PAM-interacting residues. Effects of mutations on the ability to induce indels at two *DNMT1* targets were examined (n = 3, error bars show mean ± SEM).

See also Figure S5.

the D1263A mutant exhibited significantly reduced activity (Figure 6C), confirming the roles of Asp908, Glu993, and Asp1263 in DNA cleavage. Notably, the bridge helix is inserted between strand β3 and helix α1 in the RNase H fold and interacts with the REC2 domain (Figures 6A and 6D). The main-chain carbonyl group of Gln956 in the bridge helix forms a hydrogen bond with the side chain of Lys468 in the REC2 domain (Figure 6E). In addition, Trp958 in the RuvC domain is accommodated in the hydrophobic pocket formed by Leu467, Leu471, Tyr514, Arg518, Ala521, and Thr522 in the REC2 domain (Figure 6E). These residues, with the exceptions of Leu467 and Ala521, are highly conserved among the Cpf1 family members (Zetsche et al., 2015) (Figure S4), and the W958A mutant exhibited reduced activity (Figure 6C). These observations highlight the functional importance of the bridge helix-mediated interaction between the REC and NUC lobes.

The crystal structure revealed the presence of the Nuc domain, which is inserted between the RuvC-II (strand β5) and RuvC-III (helix α3) motifs in the RuvC domain. The Nuc domain is connected to the RuvC domain via two linker loops (referred to as L1 and L2) (Figure 6A). The Nuc domain comprises five α helices and nine β strands and lacks detectable structural or sequence similarity to any known nucleases or proteins. Notably, the conserved polar residues Arg1226 and Asp1235 and the

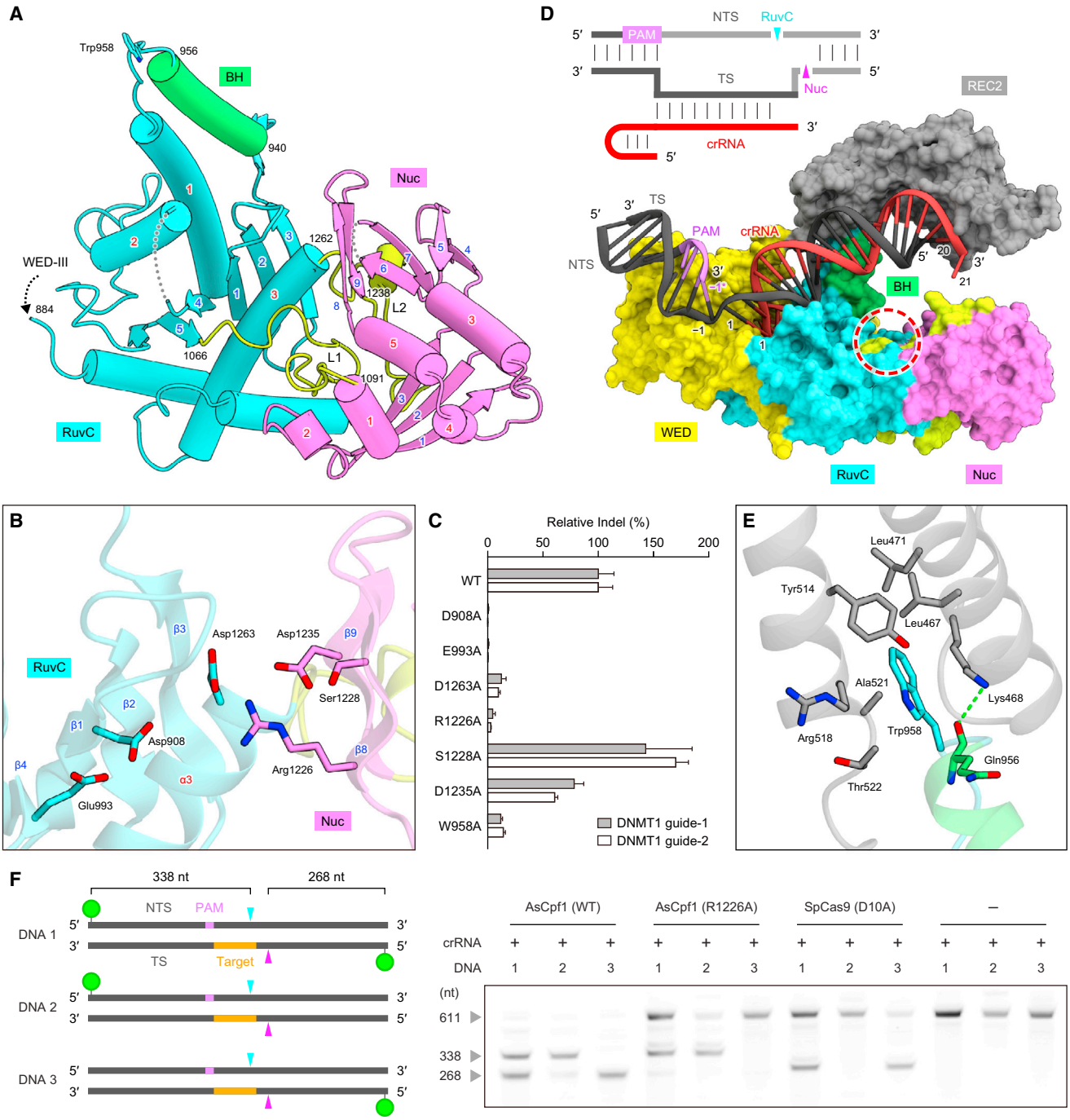


Figure 6. RuvC and Nuc Nuclease Domains

(A) Structures of the RuvC and Nuc domains. The α helices (red) and β strands (blue) in the RuvC (RNase H fold) and Nuc domains are numbered. Disordered regions are shown as dashed lines.

(B) Active site of the RuvC domain.

(C) Mutational analysis of key residues in the RuvC and Nuc domains. Effects of mutations on the ability to induce indels at two *DNMT1* targets were examined (n = 3, error bars show mean ± SEM). Indel values are normalized against wild-type AsCpf1.

(D) Spatial arrangement of the nuclease domains relative to the potential cleavage sites of the target DNA. The catalytic center of the RuvC domain is indicated by a red circle. The REC1 and PI domains are omitted for clarity. A schematic of the crRNA and target DNA is shown above the structure. The DNA strands not contained in the crystal structure are represented in light gray.

(E) Interaction between Trp958 and the hydrophobic pocket in the REC2 domain.

(legend continued on next page)

partially conserved Ser1228 are clustered in the proximity of the active site of the RuvC domain (Zetsche et al., 2015) (Figures 6B and S4). The S1228A mutant showed DNA cleavage activity comparable to that of wild-type AsCpf1 (Figure 6C). In contrast, the D1235A mutant exhibited reduced activity, whereas the R1226A mutant showed almost no activity (Figure 6C), indicating that Arg1226 is critical for DNA cleavage. Further characterization revealed that the R1226A mutant acts as a nickase that cleaves the non-target DNA strand, but not the target strand (Figure 6F), indicating that the Nuc and RuvC domains cleave the target and non-target DNA strands, respectively (Figure 6D). As in FnCpf1 (Zetsche et al., 2015), the mutations of the catalytic residues in the AsCpf1 RuvC domain abolished the cleavage of both DNA strands (Figure S6), suggesting that the cleavage of the non-target strand by the RuvC domain is a prerequisite for the target strand cleavage by the Nuc domain, presumably via a conformational change in the complex. However, further functional and structural studies are required to fully characterize the RNA-guided DNA cleavage mechanism of Cpf1.

DISCUSSION

The present structure of the AsCpf1-crRNA-target DNA complex provides mechanistic insights into RNA-guided DNA cleavage by Cpf1. The structural comparison between Cpf1 and Cas9, the only available structures of class 2 (single protein) effectors, illuminated the considerable similarity in their overall architectures, which was unanticipated given the lack of sequence similarity outside the RuvC domain (Figures 7A–7D). Both effector proteins are roughly the same size and adopt distinct bilobed structures, in which the two lobes are connected by the characteristic bridge helix and the crRNA-target DNA heteroduplex is accommodated in the central channel between the two lobes (Figures 7A and 7B). However, despite this overall similarity, only the RuvC nuclease domains of Cas9 and Cpf1 are homologous, whereas the rest of the proteins share neither sequence nor structural similarity.

One of the striking features of the Cas9 structure is the nested arrangement of the two unrelated HNH and RuvC nuclease domains, which cleave the target and non-target DNA strands, respectively (Figures 7A and 7C). In Cas9, the HNH domain is inserted between strand β4 and helix α2 of the RNase H fold in the RuvC domain (Nishimasu et al., 2014; Anders et al., 2014) (Figure 7E). In contrast, Cpf1 lacks the HNH domain and instead contains the Nuc domain, which is inserted at a different position (albeit also between the RuvC-II and RuvC-III motifs), i.e., between strand β5 and helix α3 of the RNase H fold (Figure 7F). Our mutational analysis suggested that the Nuc domain is a bona fide nuclease responsible for the target DNA strand cleavage, although the domain is relatively poorly conserved within the Cpf1 family and lacks sequence or structural similarity to any characterized nuclease (or any other protein outside the Cpf1 family). Notably, the Nuc domain of Cpf1 is located at a suitable position to cleave the single-stranded region of the target DNA strand outside the heteroduplex (Figures 7B and 7D), whereas the HNH domain of Cas9 cleaves the target DNA strand within the heteroduplex (Jinek et al., 2012; Gasiunas et al., 2012) (Figure 7C). These structural differences can explain why Cpf1 induces a staggered DNA double-strand break in the PAM-distal site, whereas Cas9 creates a blunt end in the PAM-proximal site (Zetsche et al., 2015). Unlike *Streptococcus pyogenes* Cas9 (SpCas9), in which inactivation of the RuvC nuclease turns the enzyme into a nickase that cleaves the target strand, an active RuvC domain is required for the cleavage of both strands by AsCpf1 (Figure 6F), suggesting that in Cpf1 the non-target strand cleavage by the RuvC domain is a prerequisite of the target strand cleavage by the Nuc domain. Together, these findings indicate that, despite the overall structural similarity and the apparent analogous roles of the two nuclease domains, there are substantial mechanistic differences between SpCas9 and AsCpf1. Further biochemical and structural studies with different members of the Cas9 and Cpf1 families are required to determine the generality of these distinctions between the two effector proteins and to completely elucidate the catalytic mechanism of Cpf1.

The structural comparison between Cpf1 and Cas9 revealed a striking degree of apparent structural and functional convergence between Cpf1 and Cas9, which is compatible with the previously proposed scenario of independent evolution of the effectors in the different types and subtypes of class 2 (Shmakov et al., 2015). Intriguingly, Cpf1 and Cas9 employ distinct structural features and recognize the seed region in the crRNA and the +1 phosphate group in the target DNA to achieve RNA-guided DNA targeting. In Cas9, the seed region is anchored by an arginine cluster in the bridge helix between the RuvC and REC domains, whereas the +1 phosphate group is recognized by the "phosphate lock" loop between the RuvC and WED domains (Anders et al., 2014; Nishimasu et al., 2015) (Figure S7A). In contrast, in Cpf1, the seed region is anchored by the WED and REC domains, whereas the +1 phosphate group is recognized by the WED domain (Figure S7B). Structural analyses of additional class 2 effectors, as well as the transposon-encoded TnpB proteins, which appear to be the evolutionary ancestors of the RuvC domains in the type II and type V effectors (Shmakov et al., 2015), are expected to shed further light on the evolution of this remarkable class of RNA-guided endonucleases.

The AsCpf1 structure also revealed notable differences in the PAM recognition mechanism between Cpf1 and Cas9. In Cas9, the PAM nucleotides in the non-target DNA strand are primarily read out from the major groove side, via hydrogen-bonding interactions with specific residues in the PI domain. In SpCas9, the second G and third G in the 5′-NGG-3′ PAM are recognized by Arg1333 and Arg1335 in the PI domain, via bidentate hydrogen

(F) The AsCpf1 R1226A mutant is a nickase cleaving the non-target DNA strand. The wild-type or the R1226A mutant (inactivation of the Nuc domain) of AsCpf1 was incubated with crRNA and the target DNA, which was labeled at the 5′ ends of both strands (DNA 1) or at the 5′ end of either the non-target strand (DNA 2) or the target strand (DNA 3). The cleavage products were analyzed by 10% polyacrylamide TBE-Urea denaturing gel electrophoresis. The SpCas9 D10A mutant (inactivation of the RuvC domain) is a nickase cleaving the target strand and was used as a control.
See also Figure S6.

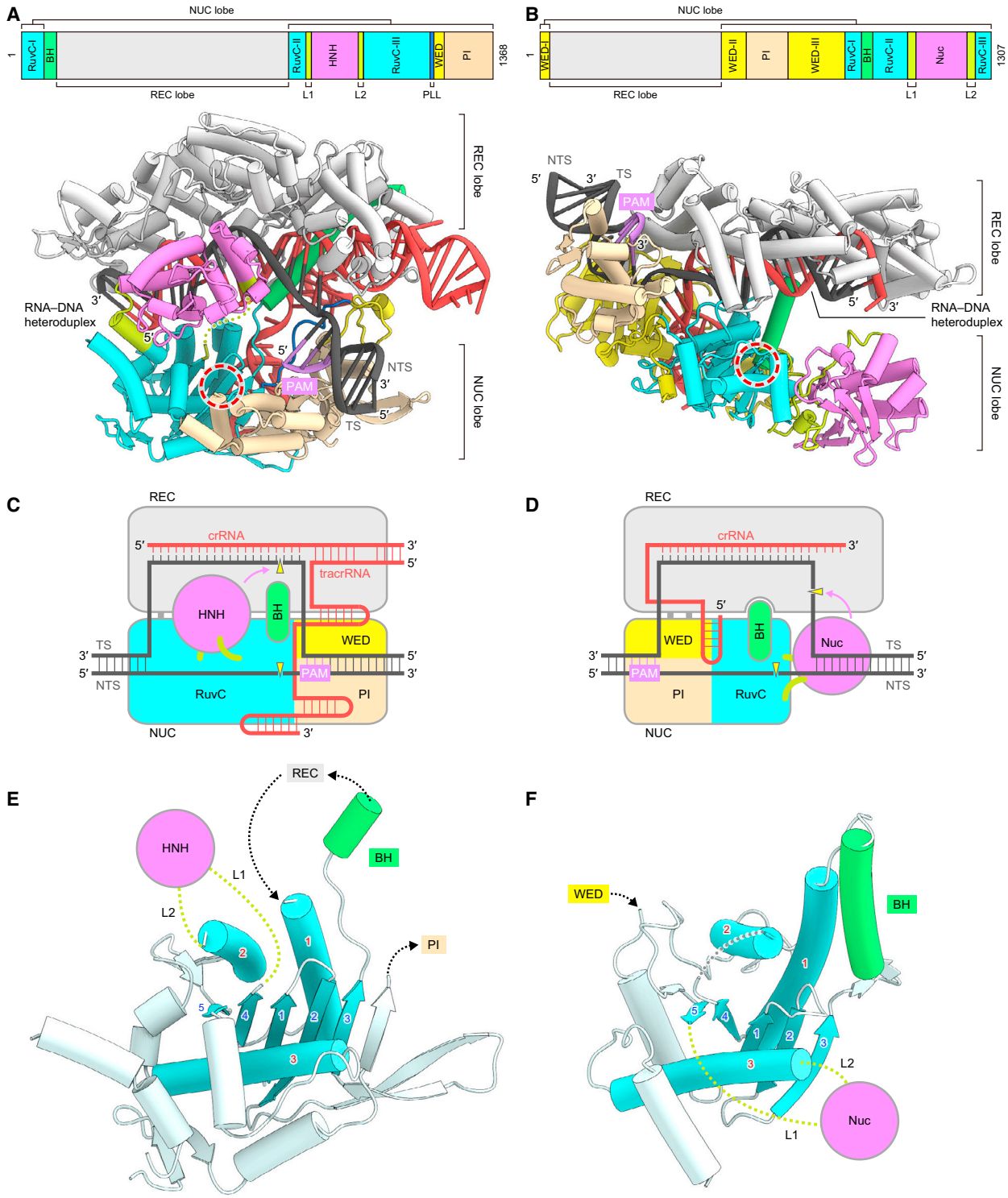


Figure 7. Comparison between Cas9 and Cpf1

(A and B) Comparison of the domain organizations and overall structures between SpCas9 (PDB: 4UN3) (A) and AsCpf1 (B). The catalytic centers of the RuvC domain are indicated by a red circle.

(C and D) Models of RNA-guided DNA cleavage by Cas9 (C) and Cpf1 (D).

(E and F) Comparison of the RuvC domains of SpCas9 (PDB: 4UN3) (E) and AsCpf1 (F). The secondary structures of the conserved RNase H fold are numbered. See also Figure S7.

bonds, respectively (Anders et al., 2014) (Figure S7A). In contrast, in AsCpf1, the PAM nucleotides in both the target and non-target DNA strands are read out by the PI domain from both the minor and major groove sides. In particular, as observed in other protein-DNA complexes (Rohs et al., 2009), the conserved lysine residue (Lys607 in AsCpf1) in the PI domain is inserted into the narrow minor groove of the PAM duplex and plays critical roles in the PAM recognition (Figure S7B). These structural observations show that, whereas Cas9 recognizes the PAM primarily via a base readout mechanism, Cpf1 combines base and shape readout to recognize the PAM. These mechanistic differences in the PAM recognition can explain why Cas9 orthologs recognize G-rich, diverse PAM sequences, whereas the widely different members of the Cpf1 family recognize similar T-rich PAMs (Zetsche et al., 2015).

In summary, the present structure of AsCpf1, combined with the mutational analysis of the two nuclease domains, provides mechanistic insights into the RNA-guided DNA recognition and cleavage by this recently discovered CRISPR-Cas effector protein and highlights the similarity and differences between the type V (Cpf1) and type II (Cas9) effectors. The structural analysis of Cas9 has enabled the design of numerous Cas9 variants with improved features and novel functions. Thus, the structural information described here will facilitate the engineering of Cpf1 and further increase the utility of the CRISPR-Cpf1 toolbox.

EXPERIMENTAL PROCEDURES

Sample Preparation

The gene encoding full-length AsCpf1 (residues 1–1307) was cloned between the *Nde*I and *Xho*I sites of the modified pE-SUMO vector (LifeSensors). The AsCpf1 protein was expressed at 20°C in *Escherichia coli* Rosetta2 (DE3) (Novagen) and was purified by chromatography on Ni-NTA Superflow (QIAGEN) and HiTrap SP HP (GE Healthcare) columns. The protein was incubated overnight at 4°C with TEV protease to remove the His_6-SUMO-tag and was then passed through the Ni-NTA column. The protein was further purified by chromatography on a HiLoad Superdex 200 16/60 column (GE Healthcare). The selenomethionine (SeMet)-labeled AsCpf1 protein was expressed in *E. coli* B834 (DE3) (Novagen) and purified using a protocol similar to that used for the native protein. The crRNA was purchased from Gene Design. The target and non-target DNA strands were purchased from Sigma-Aldrich. The purified AsCpf1 protein was mixed with the crRNA, the target DNA strand, and the non-target DNA strand (molar ratio, 1:1.5:2.3:3.4), and then the reconstituted AsCpf1-crRNA-target DNA complex was purified by gel filtration chromatography on a Superdex 200 Increase column (GE Healthcare), in buffer consisting of 10 mM Tris-HCl (pH 8.0), 150 mM NaCl, and 1 mM DTT.

Crystallography

The purified AsCpf1-crRNA-target DNA complex was crystallized at 20°C by the hanging-drop vapor diffusion method. The crystallization drops were formed by mixing 1 μl of complex solution ($A_{280\ nm}$ = 10) and 1 μl of reservoir solution (8%–10% PEG 3350, 100 mM sodium acetate [pH 4.5], and 10%–15% 1,6-hexanediol) and then were incubated against 0.5 ml of reservoir solution. The SeMet-labeled complex was crystallized by mixing 1 μl of complex solution ($A_{280\ nm}$ = 10) and 1 μl of reservoir solution (27%–30% PEG 400, 100 mM sodium acetate [pH 4.0], and 200 mM lithium sulfate). The native crystals were cryoprotected in a solution consisting of 11% PEG 3350, 100 mM sodium acetate [pH 4.5], 15% 1,6-hexanediol, and 30% ethylene glycol. The Se-Met-labeled crystals were cryoprotected in a solution consisting of 35% PEG 400, 100 mM sodium acetate (pH 4.0), 200 mM lithium sulfate, and 150 mM NaCl. X-ray diffraction data were collected at 100 K on the beamlines BL41XU at SPring-8 and PXI X06SA at the Swiss Light Source. The 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 structure 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).

Generation of the AsCpf1 Mutants

The human codon-optimized AsCpf1 mutants were cloned using the Golden Gate strategy (Engler et al., 2009). Briefly, wild-type AsCpf1 (pY010) was used as the template to amplify two PCR fragments, using primers containing the *Bsm*BI restriction sites. *Bsm*BI digestion results in distinct 5′ overhangs that either are compatible with the *Hind*III or *Xba*I overhangs of the recipient vector or will reconstitute the desired point mutation at the junction of the two AsCpf1 DNA pieces.

Cleavage Activity of AsCpf1 in 293FT Cells

The plasmid expressing the wild-type or mutants of AsCpf1 with N- and C-terminal nuclear localization tags (400 ng) and the plasmid expressing the crRNA (100 ng) were used to transfect human embryonic kidney 293FT cells at 75%–90% confluency in a 24-well plate, using the Lipofectamine 2000 reagent (Life Technologies). Genomic DNA was extracted using QuickExtract DNA Extraction Solution (Epicenter). Indels were analyzed by deep sequencing, as previously described (Hsu et al., 2013).

Synthesis of crRNAs

The crRNA for in vitro cleavage assay was synthesized using the HiScribe T7 High-Yield RNA Synthesis Kit (NEB). DNA oligos corresponding to the reverse complement of the target RNA sequence were synthesized from IDT and annealed to a short T7 priming sequence. T7 transcription was performed for 4 hr and then the RNA was purified using Agencourt RNAClean XP beads (Beckman Coulter).

Preparation of AsCpf1-Containing Cell Lysate

HEK293 cells, growing in six-well plates, were transfected with AsCpf1 expression plasmids (2 μg) using the Lipofectamine 2000 reagent. After 48 hr, the cells were harvested by washing with DPBS (Life Technologies) and then were resuspended in 0.25 ml of lysis buffer (20 mM HEPES [pH 7.5], 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 0.1% Triton X-100, and 1× cOmplete Protease Inhibitor Cocktail Tablets [Roche]). After 10-min sonication and 20-min centrifugation (20,000 × *g*), the supernatants were frozen for subsequent use in in vitro cleavage assays.

In Vitro Cleavage Assay

The in vitro cleavage assay was performed with a mammalian cell lysate containing either AsCpf1 or SpCas9 protein, at 37°C for 20 min in cleavage buffer (1× CutSmart buffer [NEB] and 5 mM DTT). The cleavage reaction used 500 ng of synthesized crRNA and 200 ng of target DNA. To prepare the substrate DNA, a 611-bp region containing the target sequence with the 5′-TTTA-3′ PAM was amplified by PCR, using the pUC19 vector as a template. To generate fluorescent-labeled substrates, PCR primers were labeled by the 5′ EndTag Nucleic Acid Labeling System (Vector Laboratories); the forward and reverse primers were labeled to generate the labeled non-target and target strands, respectively. Reactions were processed with a Zymoclean Gel DNA Recovery Kit (Zymo Research) and were run on a 10% polyacrylamide TBE-Urea gel. The gel was visualized using an Odyssey CLx Imaging System (Li-Cor). For the RuvC domain mutants, the processed reactions were run on TBE 6% polyacrylamide or TBE-Urea 6% polyacrylamide gels (Life Technologies), and the gels were then stained with SYBR Gold (Invitrogen).

ACCESSION NUMBERS

The accession number for the atomic coordinates of the AsCpf1-crRNA-target DNA complex reported in this paper has been uploaded to the Protein Data Bank: 5B43.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.04.003.

AUTHOR CONTRIBUTIONS

T.Y. crystallized the complex. T.Y. and H.N. performed the structural analysis. B.Z., I.M.S., Y.L., and I.F. performed the functional analysis. T.N. and R.I. assisted with the structural determination. H.H., K.S.M., and E.V.K. analyzed the data. T.Y., H.N., H.H., E.V.K., F.Z., and O.N. wrote the manuscript with help from all authors. H.N., F.Z., and O.N. directed and supervised all of the research.

ACKNOWLEDGMENTS

We thank Arisa Kurabayashi for assistance with vector construction. We thank the beamline scientists at PXI X06SA at the Swiss Light Source and BL41XU at SPring-8 for assistance with data collection. H.N. is supported by JST, PRESTO, JSPS KAKENHI (26291010 and 15H01463), and the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology. K.S.M. and E.V.K. are supported by intramural funds of the US Department of Health and Human Services (to the National Library of Medicine). F.Z. is supported by the NIH through the NIMH (5DP1-MH100706 and 1R01-MH110049); a Waterman Award from the National Science Foundation; 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. 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, the Council for Science, 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 NIH.

Received: March 18, 2016
Revised: March 31, 2016
Accepted: March 31, 2016
Published: April 21, 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.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

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.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (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.

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.

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

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. PLoS ONE *4*, e5553.

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., Lécrivain, 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., 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.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. Cell *159*, 647–661.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

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.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

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.

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.

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.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*, 401–403.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Topkar, V.V., Zheng, Z., and Joung, J.K. (2015a). Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. Nat. Biotechnol. *33*, 1293–1298.

Kleinstiver, B.P., Prew, M.S., Tsai, S.Q., Topkar, V.V., Nguyen, N.T., Zheng, Z., Gonzales, A.P., Li, Z., Peterson, R.T., Yeh, J.R., et al. (2015b). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Konermann, S., Brigham, M.D., Trevino, A.E., Joung, J., Abudayyeh, O.O., Barcena, C., Hsu, P.D., Habib, N., Gootenberg, J.S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature *517*, 583–588.

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.

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.

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.

Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and processing of foreign DNA by the Escherichia coli CRISPR-Cas system. Cell *163*, 854–865.

Rohs, R., West, S.M., Sosinsky, A., Liu, P., Mann, R.S., and Honig, B. (2009). The role of DNA shape in protein-DNA recognition. Nature *461*, 1248–1253.

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.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science *351*, 84–88.

Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. *33*, W244–W248.

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, Summer 2013. http://www.ccp4.ac.uk/newsletters/newsletter49/content.html.

Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. PLoS Genet. *9*, e1003742.

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.

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.