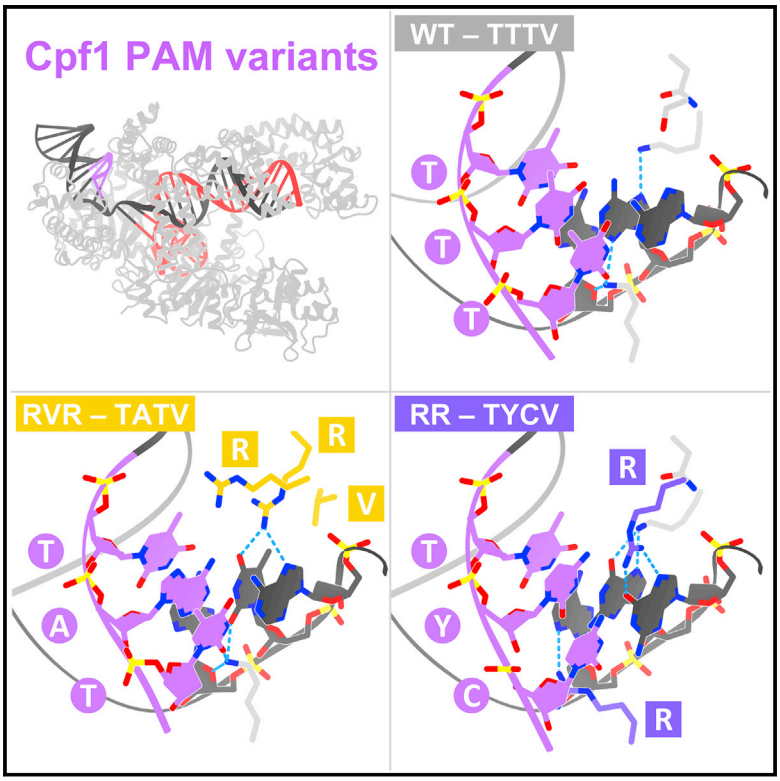
Short Article

Molecular Cell

Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1

Graphical Abstract



Authors

Hiroshi Nishimasu, Takashi Yamano, Linyi Gao, Feng Zhang, Ryuichiro Ishitani, Osamu Nureki

Correspondence

nisimasu@bs.s.u-tokyo.ac.jp (H.N.), nureki@bs.s.u-tokyo.ac.jp (O.N.)

In Brief

Variants of AsCpf1 with altered PAM specificities have been derived. Nishimasu et al. determined the high-resolution crystal structures of two of these variants bound to a crRNA and its target DNA, thereby explaining their altered PAM specificities.

Highlights

- Crystal structures of the AsCpf1 RVR and RR variants at 2.0 Å resolution
- Structural basis for the TATV PAM recognition by the RVR variant
- Structural basis for the TYCV PAM recognition by the RR variant

Nishimasu et al., 2017, Molecular Cell *67*, 1–9
July 6, 2017 © 2017 Elsevier Inc.
http://dx.doi.org/10.1016/j.molcel.2017.04.019

CellPress

Please cite this article in press as: Nishimasu et al., Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1, Molecular Cell (2017), http://dx.doi.org/10.1016/j.molcel.2017.04.019

Molecular Cell
Short Article

Structural Basis for the Altered PAM Recognition by Engineered CRISPR-Cpf1

Hiroshi Nishimasu,[1,2,7,*] Takashi Yamano,[1,7] Linyi Gao,[3,4] Feng Zhang,[3,4,5,6] Ryuichiro Ishitani,[1] and Osamu Nureki[1,8,*]
[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[2]JST, PRESTO, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan
[3]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Department of Biological Engineering
[5]McGovern Institute for Brain Research
[6]Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]These authors contributed equally
[8]Lead Contact
*Correspondence: nisimasu@bs.s.u-tokyo.ac.jp (H.N.), nureki@bs.s.u-tokyo.ac.jp (O.N.)
http://dx.doi.org/10.1016/j.molcel.2017.04.019

SUMMARY

The RNA-guided Cpf1 nuclease cleaves double-stranded DNA targets complementary to the CRISPR RNA (crRNA), and it has been harnessed for genome editing technologies. Recently, *Acidaminococcus sp. BV3L6* (AsCpf1) was engineered to recognize altered DNA sequences as the protospacer adjacent motif (PAM), thereby expanding the target range of Cpf1-mediated genome editing. Whereas wild-type AsCpf1 recognizes the TTTV PAM, the RVR (S542R/K548V/N552R) and RR (S542R/K607R) variants can efficiently recognize the TATV and TYCV PAMs, respectively. However, their PAM recognition mechanisms remained unknown. Here we present the 2.0 Å resolution crystal structures of the RVR and RR variants bound to a crRNA and its target DNA. The structures revealed that the RVR and RR variants primarily recognize the PAM-complementary nucleotides via the substituted residues. Our high-resolution structures delineated the altered PAM recognition mechanisms of the AsCpf1 variants, providing a basis for the further engineering of CRISPR-Cpf1.

INTRODUCTION

In the CRISPR-Cas prokaryotic immune systems, the effector ribonucleoprotein complexes consisting of Cas protein(s) and a CRISPR RNA (crRNA) are responsible for the degradation of foreign genetic elements (Marraffini, 2015; Barrangou and Doudna, 2016; Wright et al., 2016; Mohanraju et al., 2016). The CRISPR-Cas effector nucleases cleave the target nucleic acids complementary to the crRNA guide, and the crRNA-guided target DNA unwinding initiates with the recognition of a specific nucleotide sequence near the target sites, called the protospacer adjacent motif (PAM) (Jinek et al., 2012; Sternberg et al., 2014).

The CRISPR-Cas systems are classified into two classes, which are further divided into six types (Makarova et al., 2015). The effector nucleases in the class 2 CRISPR-Cas system, such as Cas9 in the type II system and Cpf1 in the type V system, contain a single Cas protein, and they can cleave double-stranded DNA targets complementary to the crRNA guide (Nishimasu and Nureki, 2016; Shmakov et al., 2017). Several Cas9 orthologs, such as Cas9 from *Streptococcus pyogenes* (SpCas9) and *Staphylococcus aureus* (SaCas9), exhibit robust DNA cleavage activities in numerous cell types and organisms, and they have been harnessed for a variety of genome engineering technologies, as exemplified by genome editing (Cong et al., 2013; Mali et al., 2013; Jinek et al., 2013; Ran et al., 2015). The two Cpf1 orthologs from *Acidaminococcus sp. BV3L6* (AsCpf1) and *Lachnospiraceae bacterium ND2006* (LbCpf1) also exhibit robust activities in mammalian cells, and they can be utilized as precise genome-editing tools (Zetsche et al., 2015; Kim et al., 2016, 2017; Kleinstiver et al., 2016). Cpf1 has several distinct properties from Cas9, and, thus, it can serve as a useful alternative to Cas9 in genome editing. Unlike Cas9, Cpf1 can process the crRNA array into mature crRNAs, thereby enabling the simultaneous targeting of multiple target genes (Fonfara et al., 2016; Zetsche et al., 2017). In addition, whereas SpCas9 recognizes NGG as the PAM (Jinek et al., 2012), AsCpf1 recognizes TTTV (V is A, G, or C) as the PAM, thereby extending the range of the target sequences in genome-editing applications (Zetsche et al., 2015).

The crystal structure of the AsCpf1-crRNA-target DNA complex provided mechanistic insights into the crRNA-guided DNA recognition and cleavage (Yamano et al., 2016; Gao et al., 2016). AsCpf1 adopts a bilobed architecture that accommodates the crRNA-target DNA heteroduplex. The PAM-containing DNA duplex adopts a distorted conformation characteristic of a T-rich DNA duplex, and it is recognized by AsCpf1 via the base and shape readout mechanisms (Yamano et al., 2016). The RuvC and Nuc domains are located at positions suitable to induce staggered DNA double-strand breaks at the PAM-distal

positions. A structural comparison of the AsCpf1-crRNA-DNA ternary complex (Yamano et al., 2016; Gao et al., 2016) with the LbCpf1-crRNA binary complex (Dong et al., 2016) indicated a structural rearrangement accompanying the crRNA-target DNA heteroduplex formation. Furthermore, a structural comparison of Cpf1 with Cas9 explained their distinct functionalities, and it suggested the functional convergence between the class 2 CRISPR-Cas effector nucleases (Yamano et al., 2016).

Recently, a structure-guided mutagenesis screen identified two AsCpf1 variants with altered PAM specificities (Gao et al., 2017). The RVR variant contains three substitutions (S542R/K548V/N552R), and it efficiently cleaves target sites with the non-canonical TATV PAM, in addition to those with the canonical TTTV PAM. In contrast, the RR variant contains two substitutions (S542R/K607R), and it cleaves target sites with the non-canonical TYCV PAM, including the T-less CCCC PAM. Importantly, these two AsCpf1 variants showed robust activities in human cells, thus contributing to the expansion of target spaces in Cpf1-mediated genome editing; however, the altered PAM recognition mechanisms of these variants remained unknown. It is particularly interesting to determine how the K607R substitution contributes to the altered PAM recognition by the RR variant, since Lys607 in the PAM-interacting (PI) domain is critical for the TTTV PAM recognition by wild-type (WT) AsCpf1 (Yamano et al., 2016). It is also unknown how the S542R substitution functions in the distinct PAM recognition mechanisms of the RVR and RR variants. In addition, the means by which the K548V and N552R substitutions participate in the altered PAM recognition by the RVR variant are not readily predictable.

Here we present the high-resolution crystal structures of the RVR and RR variants of AsCpf1 in complexes with the crRNA and its target DNA with the altered PAMs. A structural comparison of the two RVR and RR variants with WT AsCpf1 revealed that they achieve the altered PAM recognition mainly via newly formed base-specific interactions with the altered PAM-complementary nucleotides. Furthermore, a structural comparison between the AsCpf1 and SpCas9 variants revealed similarities and differences in their altered PAM recognition mechanisms.

RESULTS

In Vitro Cleavage Activities of the AsCpf1 Variants

In a previous study, the PAM specificities of the two AsCpf1 variants were determined by a PAM identification assay, in which AsCpf1-expressing HEK293T cell lysates were incubated with the crRNA and a library of plasmids containing a constant target sequence and a degenerate PAM (Gao et al., 2017). We thus evaluated the in vitro cleavage activities of the purified RVR and RR variants toward plasmid DNA substrates containing a 24-nt target sequence and different potential PAMs.

Since the RVR variant efficiently cleaved target sites with the TATV PAM (Gao et al., 2017), we examined the ability of the RVR variant to cleave 13 plasmid DNA targets with NATA, TNTA, TANA, or TATN as the potential PAM (Figures 1A and S1A). The RVR variant efficiently cleaved the TATA target site, as compared with the VATA (V is A, G, or C) target sites (Figures 1A and S1A), confirming the preference for the first T in the TATV PAM. In addition to the altered TATA PAM, the RVR variant efficiently recognized the canonical TTTA PAM (Figures 1A and S1A), consistent with a previous study (Gao et al., 2017) (also described below). The RVR variant was almost inactive toward the TSTA (S is G or C) and TARA (R is A or G) sites, and less active toward the TACA site (Figures 1A and S1A). These results confirmed the strong preference of the second A and the third T for the TATV PAM recognition by the RVR variant. The RVR variant was less active toward the TATT site, as compared with the TATV sites (Figures 1A and S1A), indicating the preference for the fourth V. These results demonstrated that the RVR variant efficiently recognizes TATV and TTTV as the PAM, consistent with a previous study (Gao et al., 2017).

Since the RR variant efficiently cleaved target sites with the TYCV PAM (Gao et al., 2017), we examined the RR variant for the ability to cleave 13 plasmid DNA targets with NCCC, TNCC, TCNC, or TCCN as the potential PAM (Figures 1B and S1B). The RR variant efficiently cleaved the TCCC site, as compared to the VCCC sites (Figures 1B and S1B), indicating the preference for the first T in the TYCV PAM. The RR variant efficiently cleaved the TYCC (Y is T or C) sites, but not the TRCC sites (Figures 1B and S1B), confirming the preference of the second Y for the PAM recognition by the RR variant. The RR variant was almost inactive toward the TCDC (D is A, T, or G) sites (Figures 1B and S1B), confirming the strong preference of the third C for the PAM recognition by the RR variant. The RR variant was much less active toward the TCCT site, as compared with the TCCV sites (Figures 1B and S1B), indicating the preference for the fourth V. These results confirmed that the RR variant efficiently recognizes TYCV as the PAM, consistent with earlier observations (Gao et al., 2017).

The RVR variant recognizes the TTTV PAM as well as the TATV PAM, whereas the RR variant is less active toward the TTTV PAM (Gao et al., 2017). We thus examined the in vitro cleavage activities of WT AsCpf1 and the RVR and RV variants toward four plasmid targets with the TTTN PAM (Figures 1C and S1C). WT AsCpf1 and the RVR variants efficiently cleaved the TTTV sites, as compared with the TTTT site (Figures 1C and S1C), consistent with previous studies (Zetsche et al., 2015; Kim et al., 2017). The RR variant was less active toward the TTTV sites, as compared with WT AsCpf1 and the RVR variant (Figures 1C and S1C). In stark contrast to the RVR and RR variants, WT AsCpf1 exhibited no or little activity toward the TATV and TYCV sites (Figure S2), highlighting the substantial differences in the PAM specificities between WT AsCpf1 and the two AsCpf1 variants. Together, our in vitro cleavage experiments confirmed that, unlike WT AsCpf1, the RVR and RR variants efficiently recognize the TATV and TYCV PAMs, respectively.

Crystal Structures of the AsCpf1 Variants

To elucidate the altered PAM recognition mechanisms of the AsCpf1 variants, we determined the crystal structures of (1) the RVR variant bound to the crRNA and its target DNA with the TATA PAM at 2.0 Å resolution, and (2) the RR variant bound to the crRNA and its target DNA with the TCCA PAM at 2.0 Å resolution (Figures 2A–2C; Table 1). The overall structures of the RVR and RR variants are essentially identical to that of WT AsCpf1 (Yamano et al., 2016) (root-mean-square deviations are

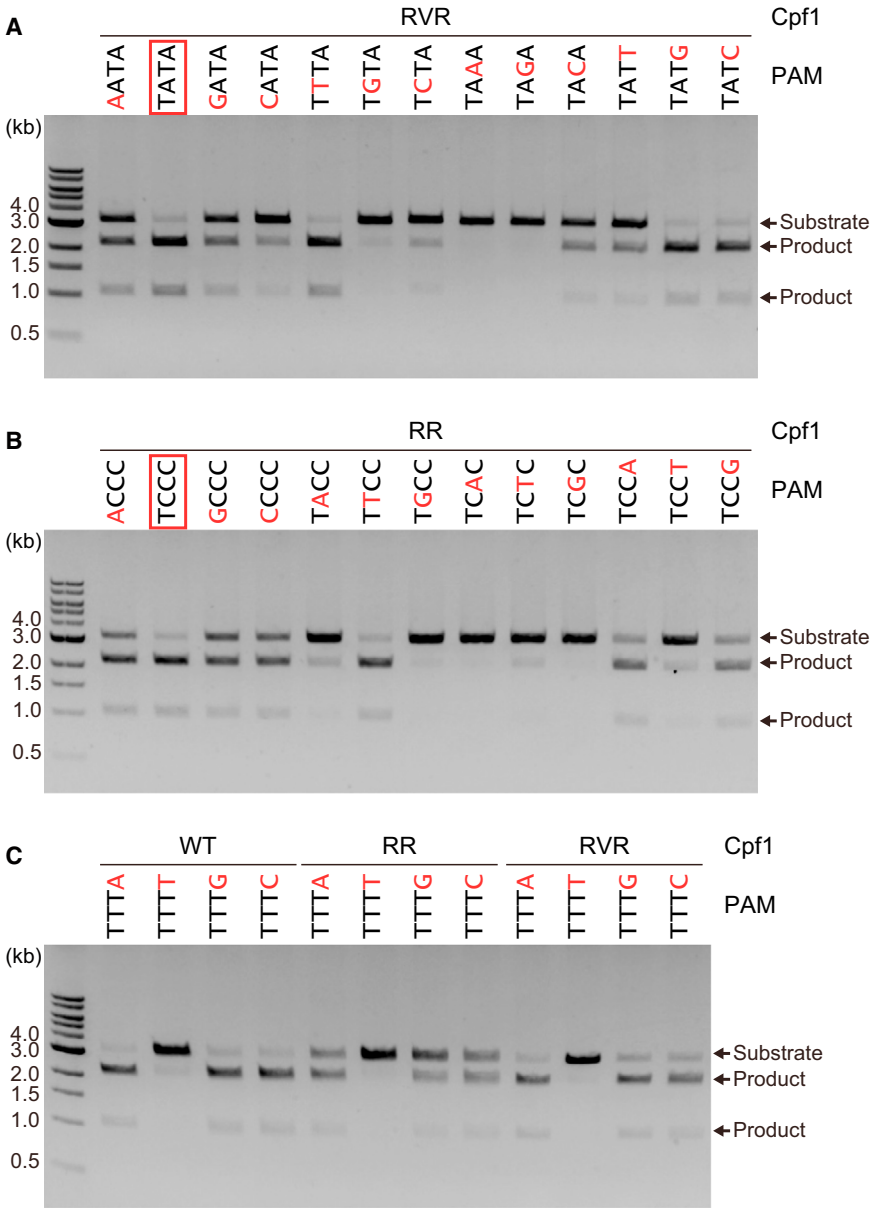


Figure 1. In Vitro Cleavage Activities of WT AsCpf1 and AsCpf1 Variants

(A and B) PAM specificities of the RVR (A) and RR (B) variants. The AsCpf1-crRNA complex (100 nM) was incubated at 37°C for 5 min with a linearized plasmid target with the different PAMs. The favorable PAMs for the RVR (TATA) and RR (TCCC) variants are boxed in red. The substituted nucleotides are colored red.

(C) Fourth PAM nucleotide preferences of WT AsCpf1 and the RVR and RR variants. The AsCpf1-crRNA complex (100 nM) was incubated at 37°C for 5 min with a linearized plasmid target with the TTTN PAM. The substituted nucleotides are colored red.

See also Figures S1 and S2.

0.52/0.60 Å for the equivalent Cα atoms between WT AsCpf1 and the RVR/RR variants) (Figure 2C). The AsCpf1 variants adopt a bilobed architecture consisting of a recognition (REC) lobe and a nuclease (NUC) lobe, in which the crRNA-target DNA heteroduplex is bound to the central channel between the two lobes (Figure 2C). In the two structures, the target DNA strand (nucleotides −10 to −1) and the PAM-containing non-target DNA strand (nucleotides −10* to −1*) form the PAM duplex, which is bound to the narrow channel formed by the wedge (WED), REC1, and PI domains (Figure 2C). The S542R/K548V/N552R and K607R substitutions are located in the WED and PI domains, respectively (Figure 2D). Lys548 and Lys607 are conserved among the Cpf1 family proteins, and they participate in the PAM recognition in the WT AsCpf1 structure (Yamano et al., 2016) (Figure 2D). In contrast, Ser542 and Asn552 are not well conserved, and they do not directly contact the PAM duplex in the WT AsCpf1 structure.

TTTA PAM Recognition by WT AsCpf1

In the WT AsCpf1 structure with the TTTA PAM, the PAM DNA duplex adopts a distorted conformation with a narrow minor groove, and it is recognized by the WED, REC1, and PI domains (Yamano et al., 2016) (Figures 3A and 3B). Notably, the conserved Lys607 residue in the PI domain forms multiple interactions with the PAM duplex from the minor groove side. Lys607 forms hydrogen bonds with the O4′ of dA(−4), the N3 of dA(−3), and the O2 of dT(−2*) (Figures 3A and 3B). Moreover, Lys548 in the WED domain hydrogen bonds with the N7 of dA(−3) from the major groove side (Figure 3A). In the WT AsCpf1 structure, the dT(−1):dA(−1*) base pair does not form base-specific contacts with the AsCpf1 protein (Yamano et al., 2016); nonetheless, our in vitro cleavage data and previous studies (Zetsche et al., 2015; Kim et al., 2017) showed that AsCpf1 prefers the V (V is A, G, or C) nucleotides at the fourth PAM position (Figure 1C). To clarify the structural basis for the fourth V preference, we modeled a T nucleotide at the fourth PAM position (dT(−1*)) in the WT AsCpf1 structure. The modeling suggested that the fourth PAM nucleotide adopts a distinct conformation, due to the interaction with the PI domain (Figure S3A), and that the 5-methyl group of dT(−1*) in the non-target strand is located closer to the neighboring backbone phosphate group, as compared with those of dT(−2*), dT(−3*), and dT(−4*) (Figure S3B). In addition, the modeling indicated that the dA(−1) in the target strand does not form unfavorable interactions with the protein. These observations suggested that AsCpf1 disfavors the fourth T in the PAM, likely due to the relatively

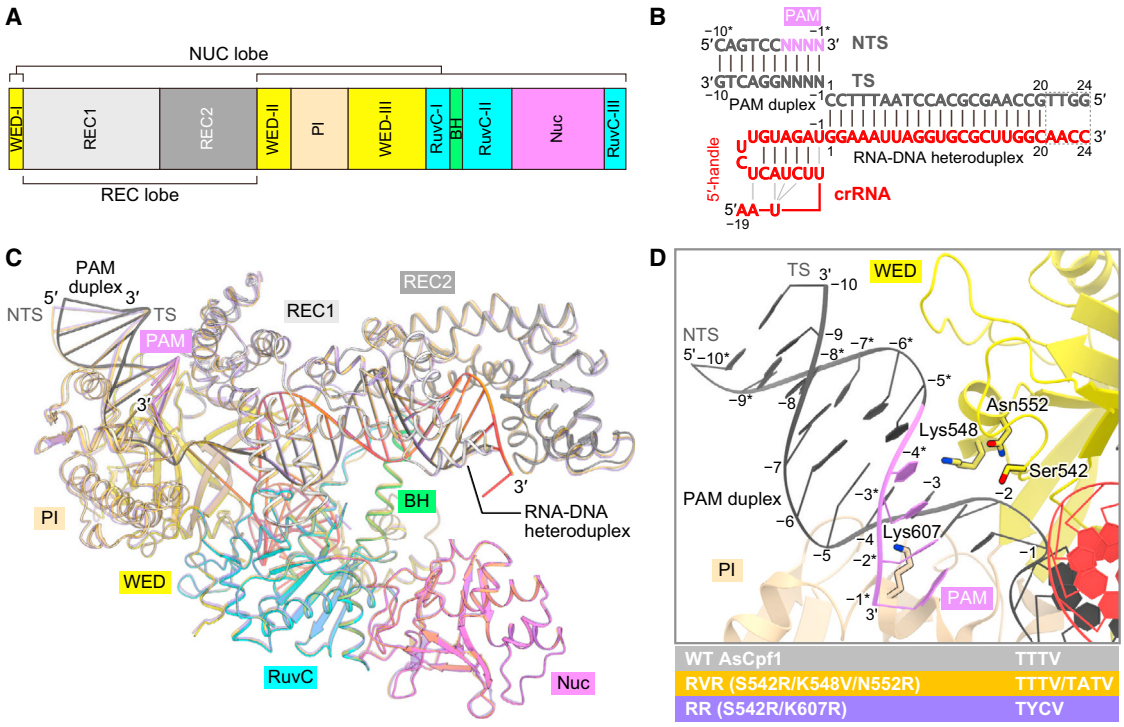


Figure 2. Overall Structure of the AsCpf1 Variant

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

(B) Nucleotide sequences of the crRNA and the target DNA. The PAM nucleotides are TATA and TCCA in the RVR and RR variant structures, respectively. The disordered nucleotides in the variant structures are surrounded by dashed lines. TS, target DNA strand; NTS, non-target DNA strand.

(C) Superimposition of the crystal structures of WT AsCpf1 (Yamano et al., 2016) (PDB: 5B43) (colored as in A and B) and the RVR (orange) and RR (purple) variants.

(D) PAM duplex in the WT AsCpf1 structure (Yamano et al., 2016) (PDB: 5B43). The substituted residues are shown as stick models.

shorter distance between its 5-methyl group and the backbone phosphate group.

TATA PAM Recognition by the RVR Variant

In the RVR (S542R/K548V/N552R) variant structure with the TATA PAM, the O4 of dT(−4*) forms a water-mediated hydrogen bond with Arg552 (N552R) (Figure 3C), explaining the preference for the first T in the TATV PAM. The N6 and N7 of dA(−3*) are recognized by Thr167 and Thr539 via a water-mediated hydrogen-bonding network (Figure 3C). Notably, the PAM-complementary dT(−3) is extensively recognized by the protein (Figures 3C and S3C). The O2 and O4 of dT(−3) form hydrogen bonds with Lys607 and Arg552 (N552R), respectively. In addition, the 5-methyl group of dT(−3) forms a hydrophobic interaction with the side chain of Val548 (K548V) (Figure 3C). These structural findings can explain the strong preference of the second A for the TATV PAM recognition by the RVR variant. As in the WT AsCpf1 structure, the O2 of dT(−2*) hydrogen bonds with Lys607 (Figure 3D). In addition, the N7 of dA(−2) hydrogen bonds with Arg552 (N552R) (Figures 3D and S3D). These structural observations are consistent with the preference for the third T in the TATV PAM. In the RVR variant, Arg542 (S542R) does not contact the PAM duplex. Together, these structural findings explain the mechanism of TATV PAM recognition by the RVR variant.

TCCA PAM Recognition by the RR Variant

In the RR (S542R/K607R) variant structure with the TCCA PAM, the O4 of dT(−4*) forms a water-mediated hydrogen bond with Lys548, while the N3 of dA(−4) hydrogen bonds with Arg607 (K607R) (Figure 3E). These observations explain the preference of the RR variant for the first T in the TYCV PAM. It is likely that, in WT AsCpf1, Lys548 forms a similar water-mediated interaction with dT(−4*) and contributes to the preference for the first T in the TTTV PAM, although such a water molecule was not resolved in the previous WT AsCpf1 structure at a lower resolution (2.8 Å) (Yamano et al., 2016). dC(−3*) does not directly contact the protein. Instead, the O6 and N7 of dG(−3) hydrogen bond with Lys548, while the N3 of dG(−3) forms a water-mediated hydrogen bond with Arg607 (K607R) (Figure 3E). It is likely that the N3 and N7 of the A nucleotide at this position are recognized by Arg607 (K607R) and Lys548, respectively. These observations explain the preference of the second Y for the TYCV PAM recognition by the RR variant. Notably, the O6 and N7 of dG(−2) are recognized by Arg542 (S542R) via bidentate hydrogen-bonding interactions, whereas dC(−2*) does not directly contact the protein (Figures 3F and S3E). These structural findings can explain the strong preference of the third C in the TYCV PAM recognition by the RR variant. Moreover, the side chain of Arg607 (K607R) inserts into the minor groove of the PAM duplex, and it interacts with the ribose moieties of dA(−4), dC(−2*), and

Table 1. Data Collection and Refinement Statistics

AsCpf1 Variant	RVR (TATA PAM)	RR (TCCA PAM)
Data Collection		
Beamline	SPring-8 BL41XU	SPring-8 BL41XU
Wavelength (Å)	1.0000	1.0000
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	81.2, 133.7, 199.6	80.7, 133.3, 200.0
α, β, γ (degrees)	90, 90, 90	90, 90, 90
Resolution (Å)[a]	40.0–2.0 (2.03–2.00)	40.0–2.0 (2.03–2.00)
R_{pim}	0.021 (0.396)	0.034 (0.466)
$I/\sigma I$	17.1 (2.0)	10.5 (1.5)
Completeness (%)	98.2 (98.3)	99.9 (99.9)
Multiplicity	6.6 (6.6)	6.4 (6.4)
CC(1/2)	0.999 (0.813)	0.999 (0.841)
Refinement		
Resolution (Å)	40.0–2.0 (2.02–2.00)	40.0–2.0 (2.02–2.00)
No. of reflections	143,851 (4,771)	145,320 (4,787)
R_{work}/R_{free}	0.175/0.210 (0.272/0.333)	0.183/0.214 (0.256/0.304)
No. of atoms		
Protein	10,455	10,431
Nucleic acid	1,639	1,639
Ion	6	6
Solvent	731	747
B-factors (Å^2)		
Protein	51.8	49.5
Nucleic acid	50.2	48.2
Ion	50.1	45.2
Solvent	52.3	49.6
Root-mean-square deviations (RMSDs)		
Bond lengths (Å)	0.007	0.006
Bond angles (degrees)	0.843	0.834
Ramachandran plot (%)		
Favored region	98.27	98.12
Allowed region	1.65	1.80
Outlier region	0.08	0.08
MolProbity score		
Clashscore	2.68	2.48
Rotamer outlier	2.20	2.30

[a]Values in parentheses are for the highest resolution shell.

dA(−1*) (Figure S3F). Together, these structural findings explain the mechanism of TYCV PAM recognition by the RR variant.

Conformational Differences in the PAM Duplex

A structural comparison between WT AsCpf1 and the variants revealed the conformational differences in their PAM duplexes (Figure 4A). In the WT AsCpf1 structure with the TTTA PAM, the PAM duplex adopts a distorted conformation characteristic of a T-rich DNA duplex, in which Lys607 forms multiple interactions with the minor-groove edge of the PAM duplex (Yamano et al., 2016) (Figure 4A). In contrast, in the structures of the RVR (with the TATA PAM) and RR (with the TCCA PAM) variants, the PAM duplexes adopt B-form-like conformations (Figure 4A), supporting the notion that the distorted conformation of the PAM duplex in the WT AsCpf1 structure is due to the three successive T nucleotides. Unlike the RR variant, the RVR variant efficiently recognizes the TTTV PAM (Figure 1C), and the location of the Lys607 residue is similar to that in the WT AsCpf1 structure (Figure 4A). These observations suggested that the RVR variant recognizes the TTTV PAM in a similar manner to that of WT AsCpf1, and they highlighted the importance of Lys607 for the TTTV PAM recognition.

Cooperative Structural Rearrangements Induced by the Substitutions

A structural comparison between WT AsCpf1 and the two variants also revealed conformational differences in the AsCpf1 proteins. In the structures of the RVR and RR variants, Arg542 (S542R) adopts distinct conformations and plays different functional roles (Figure 4B). In the RR variant structure, Arg542 forms bidentate hydrogen bonds with dG(−2) in the target DNA strand, and it plays a critical role in the TYCV PAM recognition (Figure 4B). In contrast, in the RVR variant structure, Arg542 in the WED domain interacts with Thr167 and Ser170 in the REC1 domain (Figure 4B). Our in vitro cleavage experiments revealed that the VR (K548V/N552R) variant exhibits reduced activities, as compared with the RVR (S542R/K548V/N552R) variant (Figure S4), indicating the functional importance of the Arg542-mediated inter-domain interaction. Given that Arg542 is located far away from the PAM duplex in the RVR structure, it is likely that Arg542 does not interact directly with the PAM duplex and contributes to the structural maintenance of the PAM-duplex channel, thereby enhancing the PAM recognition.

Our high-resolution structures further revealed unexpected conformational rearrangements induced by the N552R substitution in the RVR variant (Figure 4C). In the structures of WT AsCpf1 and the RR variant, the side chain of Asn552 hydrogen bonds with the side chain of Thr539 (Figure 4C). In the RR variant structure, the side chain of Asn552 also interacts with the backbone phosphate group between dA(−2) and dT(−1) (Figure 4C). In contrast, in the RVR variant structure, the side chains of Thr539 and Asn551 adopt distinct conformations, as compared with those in the WT AsCpf1 and RR variant structures, and they interact with the side chain of Arg552 (N552R) (Figure 4C). Arg552 (N552R) forms a water-mediated interaction with the backbone phosphate group between dA(−2) and dT(−1), while Asn551 interacts with the backbone phosphate group between dC(−6*) and dC(−5*).

DISCUSSION

The present high-resolution structures reveal the altered PAM recognition mechanisms of the RVR and RR variants, and they also provide detailed insights into the functional mechanism of

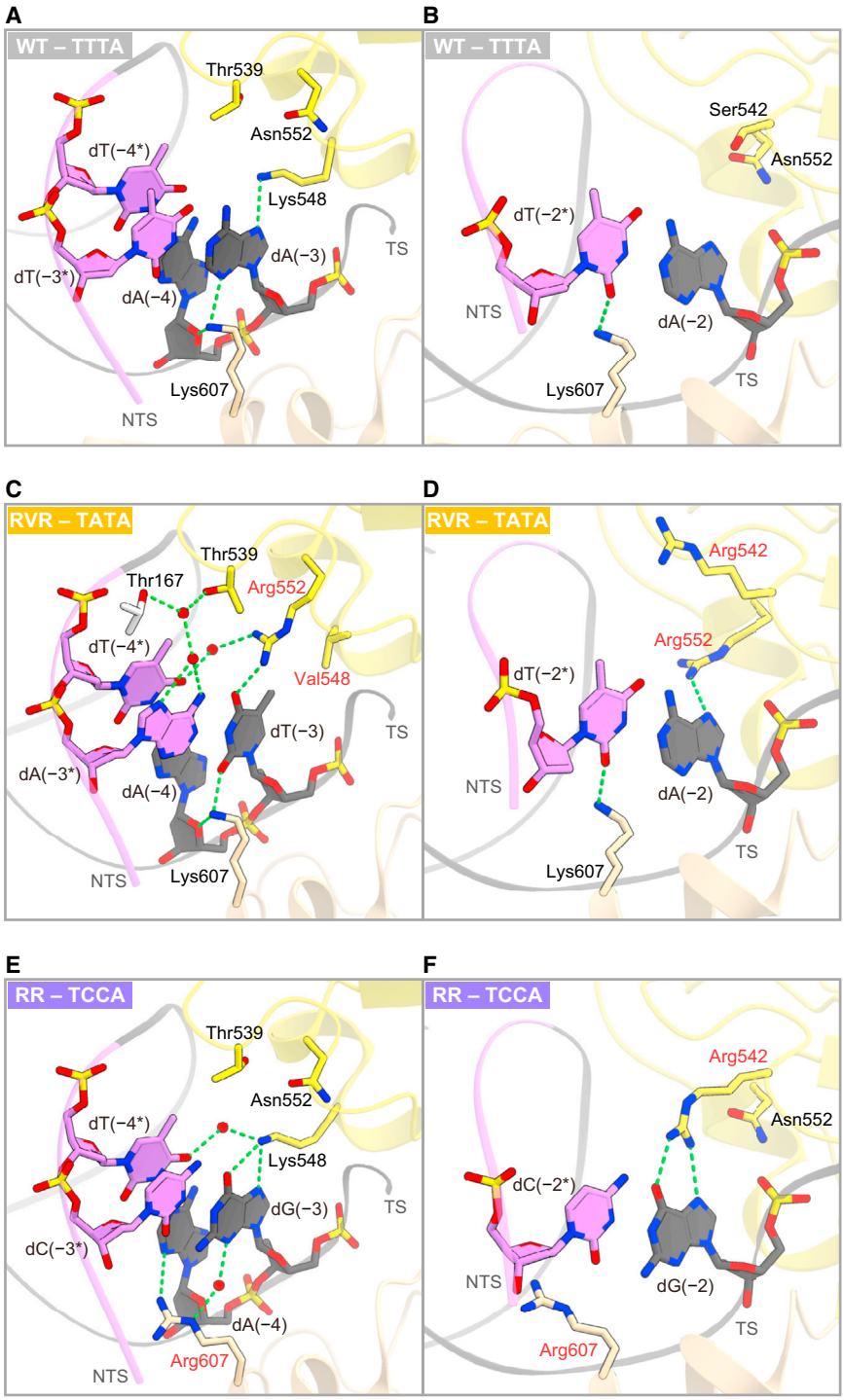


Figure 3. PAM Recognition by WT AsCpf1 and AsCpf1 Variants

(A and B) TTTA PAM recognition by WT AsCpf1 (Yamano et al., 2016) (PDB: 5B43).

(C and D) TATA PAM recognition by the RVR variant.

(E and F) TCCA PAM recognition by the RR variant.

The interactions with the nucleotides at the first and second PAM positions are shown in (A), (C), and (E). The interactions with the nucleotides at the third PAM position are shown in (B), (D), and (F). In (C)–(F), the substituted residues are highlighted by red labels. In (C) and (E), water molecules are depicted by red spheres.

See also Figure S3.

WT AsCpf1. WT AsCpf1 recognizes the TTTV PAM mainly via multiple interactions between Lys607 and the minor-groove edge of the PAM duplex (Yamano et al., 2016). In contrast, the RVR and RR variants achieve the altered PAM recognition via newly formed interactions with the major-groove edges of the PAM-complementary nucleotides in the target strand, rather than the altered PAM nucleotides in the non-target strand. In the RVR variant, Val548 (K548V) and Arg552 (N552R) form base-specific contacts with the T nucleotide complementary to the altered second A in the TATV PAM. In the RR variant, Arg542 (S542R) forms bidentate hydrogen bonds with the G nucleotide complementary to the altered third C nucleotide in the TYCV PAM. This Arg-G interaction is frequently observed in Cas9-mediated PAM recognition, such as those in SpCas9 (Arg1333-G2 and Arg1335-G3 in the NGG PAM) (Anders et al., 2014), SaCas9 (Arg1015-G3 in the NNGRRT PAM) (Nishimasu et al., 2015), and *Francisella novicida* Cas9 (Arg1585-G2 and Arg1556-G3 in the NGG PAM) (Hirano et al., 2016a). In addition, in the RR variant, Arg607 (K607R) donates hydrogen bonds and van der Waals contacts with the PAM duplex, thereby compensating for the loss of the interactions between Lys607 and the PAM duplex observed in WT AsCpf1.

A structural comparison of the AsCpf1 variants with the previously reported SpCas9 variants, such as VQR (D1135V/R1335Q/T1337R) and VRER (D1135V/G1218R/R1335E/T1337R) (Kleinstiver et al., 2015), reveal striking differences in their altered PAM recognition mechanisms. Whereas the third G in the NGG PAM is recognized by Arg1335 in WT SpCas9 (Anders et al., 2014), the third A in the NGA PAM and the third C in the NGCG PAM are recognized by Gln1335 (R1335Q) in the VQR variant and Glu1335 (R1335E) in the VRER variant, respectively (Anders et al., 2016; Hirano et al., 2016b). Thus, the altered PAM recognition by the SpCas9 variants mainly relies on the replacement of the Arg1335-G3 interaction in WT SpCas9 with the altered base-specific interactions (i.e., the Gln1335-A3 interaction in the VQR variant and the

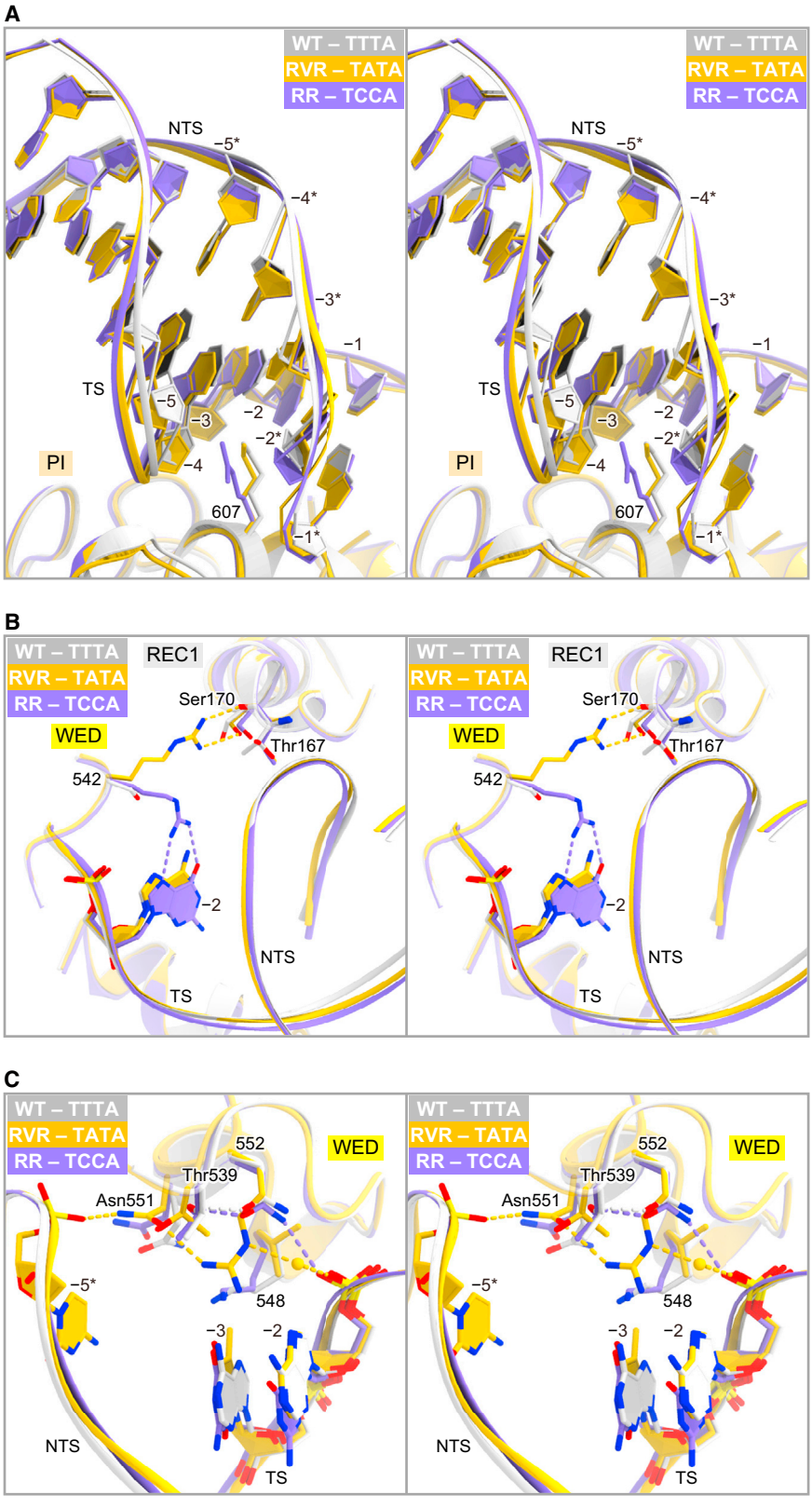


Figure 4. Structural Differences between WT AsCpf1 and AsCpf1 Variants

(A) Conformational differences in the PAM duplexes between WT AsCpf1 (PDB: 5B43) and the variants (stereo view).

(B) Structural differences in Arg542 (S542R) between the RVR and RR variants (stereo view).

(C) Structural rearrangements around Arg552 (N552R) in the RVR variant (stereo view). A water molecule is shown as a sphere. In (A)–(C), WT AsCpf1 and the RVR and RR variants are colored gray, orange, and purple, respectively.

See also Figure S4.

Glu1335-C3 interaction in the VRER variant). In contrast, the altered PAM recognition by the AsCpf1 variants relies on newly formed interactions between the substituted residues and the altered PAM-complementary nucleotides (i.e., Val548/Arg552-A2-complementary T2 in the RVR variant and Arg542-C3-complementary G3 in the RR variant). These differences are reflected by the distinct PAM recognition mechanisms of SpCas9 (base readout from the major-groove side) (Anders et al., 2014) and AsCpf1 (base and shape readout from the minor- and major-groove sides) (Yamano et al., 2016).

The present structures reveal that Arg542 (S542R) plays distinct roles in the RVR and RR variants. Arg542 forms the inter-domain interactions and may reinforce the PAM-duplex-binding channel in the RVR variant, whereas Arg542 forms the base-specific contacts with the PAM duplex in the RR variant. These observations demonstrate that the amino acid substitutions that do not provide interactions with the PAM duplex can contribute to the engineering of the Cpf1's PAM specificity. This contrasts with the altered PAM recognition by the SpCas9 variants, in which the substituted residues provide new contacts with the PAM duplex. These differences also highlight the mechanistic differences in the PAM recognition between SpCas9 (via the PAM-binding groove within the PI domain) (Anders et al., 2014) and AsCpf1 (via the PAM-binding channel formed by the WED, REC1, and PI domains) (Yamano et al., 2016). Furthermore, the present findings provide important clues for the Cpf1 engineering, and they suggest that amino acid substitutions that reinforce the PAM-binding channel could contribute to the alteration of Cpf1's PAM specificity.

There are also mechanistic similarities in the altered PAM recognition by the SpCas9 and AsCpf1 variants. In the SpCas9 and AsCpf1 variants, unexpected structural rearrangements play important roles in the altered PAM recognition, thus highlighting the power of structure-guided random mutagenesis approaches. In the SpCas9 variant structures, the direct hydrogen-bonding interactions between the altered third PAM nucleotides and the substituted residues (Gln1335 and Glu1335) are enabled by the unexpected displacement of the PAM duplex, which is cooperatively induced by the other substitutions (D1135V and T1337R) (Anders et al., 2016; Hirano et al., 2016b). In the AsCpf1 RR variant structure, the PAM duplex undergoes a conformational change, partly due to the replacement of Lys607 (K607R). Moreover, in the AsCpf1 RVR variant structure, the N552R substitution induces local conformational changes in Thr539 and Asn551, thus rearranging the interactions with the PAM duplex. These cooperative structural rearrangements are not readily predictable from the WT AsCpf1 structure (Yamano et al., 2016), and thus they confirm the power of the combination of structural information and molecular evolution for the engineering of the CRISPR-Cas nucleases.

In summary, our structural studies reveal the altered PAM recognition mechanisms of the recently engineered AsCpf1 variants. Furthermore, the structural comparison between the AsCpf1 and SpCas9 variants enhances our understanding of the PAM recognition mechanisms of class 2 CRISPR-Cas nucleases, and it provides a framework for the future engineering of the CRISPR-Cpf1 toolbox.

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
  - Sample preparation
  - Crystallography
  - In vitro cleavage assay
- QUANTIFICATION AND STATISTICAL ANALYSES
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.04.019.

AUTHOR CONTRIBUTIONS

H.N. conceived the project, collected diffraction data, determined the structures, and wrote the manuscript with help from all the authors. T.Y. prepared the proteins, crystallized the complex, and performed in vitro cleavage experiments. L.G. and F.Z. provided unpublished data. R.I. analyzed the structural data. H.N. and O.N. supervised the project.

ACKNOWLEDGMENTS

We thank Hisato Hirano and the beamline scientists at BL41XU at SPring-8 for assistance with data collection, Dr. Takanori Nakane for assistance with data processing, and Hisato Hirano and Seiichi Hirano for comments on the manuscript. H.N. is supported by JST, PRESTO (JPMJPR13L8), and JSPS KAKENHI (grants 26291010 and 15H01463). F.Z. is supported by the NIH through the National Institute of Mental Health (NIMH) (5DP1-MH100706 and 1R01-MH110049); the NSF; Howard Hughes Medical Institute; the New York Stem Cell, Simons, Paul G. Allen Family, and Vallee Foundations; and James and Patricia Poitras, Robert Metcalfe, and David Cheng. 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).

Received: February 25, 2017
Revised: April 13, 2017
Accepted: April 25, 2017
Published: June 5, 2017

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.

Anders, C., Bargsten, K., and Jinek, M. (2016). Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. Mol. Cell *61*, 895–902.

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

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.

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.

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

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

Fonfara, I., 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.

Gao, L., Cox, D.B.T., Yan, W.X., Manteiga, J.C., Schneider, M.W., Yamano, T., Nishimasu, H., Nureki, O., Crosetto, N., and Zhang, F. (2017). Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. http://dx.doi.org/10.1038/nbt.3900.

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. (2016a). Structure and engineering of *Francisella novicida* Cas9. Cell *164*, 950–961.

Hirano, S., Nishimasu, H., Ishitani, R., and Nureki, O. (2016b). Structural basis for the altered PAM specificities of engineered CRISPR-Cas9. Mol. Cell *61*, 886–894.

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

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

Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H., and Kim, J.S. (2016). Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat. Biotechnol. *34*, 863–868.

Kim, H.K., Song, M., Lee, J., Menon, A.V., Jung, S., Kang, Y.M., Choi, J.W., Woo, E., Koh, H.C., Nam, J.W., and Kim, H. (2017). *In vivo* high-throughput profiling of CRISPR-Cpf1 activity. Nat. Methods *14*, 153–159.

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. (2015). Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature *523*, 481–485.

Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J., and Joung, J.K. (2016). Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat. Biotechnol. *34*, 869–874.

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

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

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

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

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

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.

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

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. *15*, 169–182.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. Acta Crystallogr. D Biol. Crystallogr. *66*, 22–25.

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

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.

Zetsche, B., Heidenreich, M., Mohanraju, P., Fedorova, I., Kneppers, J., DeGennaro, E.M., Winblad, N., Choudhury, S.R., Abudayyeh, O.O., Gootenberg, J.S., et al. (2017). Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat. Biotechnol. *35*, 31–34.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
AsCpf1 WT	Yamano et al., 2016	N/A
AsCpf1 RR	This paper	N/A
AsCpf1 RVR	This paper	N/A
AsCpf1 VR	This paper	N/A
Polyethylene glycol 3,350 Monodisperse 50% (w/v)	Hampton Research	HR2-527
Sodium acetate trihydrate pH 4.5 - 1.0 M solution	Hampton Research	HR2-789
1,6-Hexanediol	Fluka	52800
Deposited Data		
Atomic coordinates, AsCpf1 RVR (TATA PAM)	This paper	PDB: 5XH6
Atomic coordinates, AsCpf1 RR (TCCA PAM)	This paper	PDB: 5XH7
In vitro DNA cleavage experiments	This paper; Mendeley Data	http://dx.doi.org/10.17632/5swpr7yx7d.1
Experimental Models: Organisms/Strain		
E. coli Mach	Thermo Fisher Scientific	C862003
E. coli Rosetta 2 (DE3)	Novagen	71397
Recombinant DNA		
pE-SUMO-AsCpf1-RR	This paper	N/A
pE-SUMO-AsCpf1-RVR	This paper	N/A
pE-SUMO-AsCpf1-VR	This paper	N/A
pUC119-T24, various PAMs	This paper	N/A
Sequence-Based Reagents		
DNA primers	This paper	Table S1
DNA oligos (for crystallization)	This paper	Table S1
AsCpf1 crRNA	Yamano et al., 2016	Table S1
Software and Algorithms		
DIALS	Waterman et al., 2013	http://dials.lbl.gov
AIMLESS	Evans and Murshudov, 2013	http://www.ccp4.ac.uk/html/aimless.html
MOLREP	Vagin and Teplyakov, 2010	http://www.ccp4.ac.uk/html/molrep.html
COOT	Emsley and Cowtan, 2004	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
PHENIX	Adams et al., 2010	https://www.phenix-online.org
CueMol	N/A	http://www.cuemol.org
Other		
Amicon Ultra-4 Centrifugal Filter Units - 10,000 NMWL	Millipore	UFC801024
Ni-NTA Superflow	QIAGEN	30450
HiTrap SP HP	GE Healthcare	17040701
HiLoad 16/600 Superdex 200 pg	GE Healthcare	28989335
Superdex 200 Increase 10/300 GL	GE Healthcare	28990944

CONTACT FOR REAGENT AND RESOURCE SHARING

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The plasmid DNAs were amplified in *Escherichia coli* Mach (Thermo Fisher Scientific), cultured in LB medium (Nacalai Tesque) at 37°C overnight. The recombinant proteins were overexpressed in *E. coli* Rosetta 2 (DE3) (Novagen).

Sample preparation

WT AsCpf1 and the RVR and RV variants were prepared essentially as described previously (Yamano et al., 2016). The gene encoding full-length AsCpf1 (residues 1–1307) was cloned into the modified pE-SUMO vector (LifeSensors), and the mutations (S542R, K548V, N552R and K607R) were introduced by a PCR-based method (Table S1). The AsCpf1-expressing *E. coli* Rosetta2 (DE3) cells were cultured at 37°C in LB medium (containing 20 mg/l kanamycin) until the OD_{600} reached 0.8, and protein expression was then induced by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside (Nacalai Tesque). The *E. coli* cells were further cultured at 20°C for 18 hr, and harvested by centrifugation at 5,000 g for 10 min. The *E. coli* cells were resuspended in buffer A (50 mM Tris-HCl, pH 8.0, 20 mM imidazole, 300 mM NaCl and 3 mM 2-mercaptoethanol), lysed by sonication, and then centrifuged at 40,000 g for 30 min. The supernatant was mixed with 5 mL Ni-NTA Superflow (QIAGEN), and the mixture was loaded into an Econo-Column (Bio-Rad). The resin was washed with buffer A, and the protein was eluted with buffer B (50 mM Tris-HCl, pH 8.0, 300 mM imidazole, 300 mM NaCl and 3 mM 2-mercaptoethanol). The eluted protein was loaded onto a HiTrap SP HP column (GE Healthcare) equilibrated with buffer C (20 mM Tris-HCl, pH 8.0 and 200 mM NaCl). The column was washed with buffer C, and the protein was then eluted with a linear gradient of 200–1000 mM NaCl. To remove the His_6-SUMO-tag, the protein was mixed with TEV protease, and was dialyzed at 4°C for 12 hr against buffer D (20 mM Tris-HCl, pH 8.0, 40 mM imidazole, 300 mM NaCl and 3 mM 2-mercaptoethanol). The protein was passed through the Ni-NTA column, and was then concentrated using an Amicon Ultra 10K filter (Millipore). The AsCpf1 protein was further purified by a HiLoad Superdex 200 16/60 column (GE Healthcare) equilibrated with buffer E (10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM DTT). The purified AsCpf1 proteins were stored at −80°C until use. The crRNA and target DNA were purchased from Gene Design and Sigma-Aldrich, respectively. The purified AsCpf1 protein was mixed with the crRNA, the target DNA strand and the non-target DNA strand (molar ratio, 1:1.5:2.3:2.3), and the reconstituted complex was concentrated using an Amicon Ultra 10K filter. The AsCpf1-crRNA-DNA complex was purified by gel filtration chromatography on a Superdex 200 Increase column (GE Healthcare) equilibrated with buffer E.

Crystallography

The purified AsCpf1-crRNA-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 ($A_{260\ nm}$ = 10) and 1 μL of reservoir solution (7%–10% PEG 3,350, 100 mM sodium acetate, pH 4.5, and 10% 1,6-hexanediol), and then were incubated against 0.5 mL of reservoir solution. The crystals were cryoprotected in a solution consisting of 9%–10% PEG 3,350, 100 mM sodium acetate, pH 4.5, 10%–15% 1,6-hexanediol and 30% ethylene glycol. X-ray-diffraction data were collected at 100 K on the beamline BL41XU at SPring-8, and the data were processed using DIALS (Waterman et al., 2013) and AIMLESS (Evans and Murshudov, 2013). The structures were determined by molecular replacement with Molrep (Vagin and Teplyakov, 2010), using the coordinates of WT AsCpf1 (PDB: 5B43) (Yamano et al., 2016) as the search model. The model building and structural refinement were performed using COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010), respectively. Structural figures were prepared using CueMol (http://www.cuemol.org).

In vitro cleavage assay

The target pUC119 plasmids with the different PAMs were generated by a PCR-based method (Table S1). The EcoRI-linearized pUC119 plasmid (100 ng), containing the 24-nt target sequence and the PAMs, was incubated at 37°C for 5 or 10 min with the AsCpf1-crRNA complex (100 nM), in 10 μL of reaction buffer containing 20 mM HEPES-NaOH, pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 1 mM DTT and 5% glycerol. The reaction was stopped by the addition of a solution containing EDTA (40 mM final concentration) and Proteinase K (10 μg). Reaction products were resolved on an ethidium bromide-stained 1% agarose gel, and then visualized using an Amersham Imager 600 (GE Healthcare).

QUANTIFICATION AND STATISTICAL ANALYSES

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

DATA AND SOFTWARE AVAILABILITY

The atomic coordinates of the AsCpf1 variants-crRNA-target DNA complexes have been deposited in the Protein Data Bank, with the accession numbers PDB: 5XH6 (the AsCpf1 RVR variant) and 5XH7 (the AsCpf1 RR variant). Data of in vitro cleavage experiments have been deposited in the Mendeley Data repository (http://dx.doi.org/10.17632/5swpr7yx7d.1). The CueMol program is available at http://www.cuemol.org.