HHS Public Access

Author manuscript

Cell. Author manuscript; available in PMC 2016 June 09.

Published in final edited form as:

Cell. 2016 February 25; 164(5): 950–961. doi:10.1016/j.cell.2016.01.039.

Structure and Engineering of *Francisella novicida* Cas9

Hisato Hirano[1], **Jonathan S. Gootenberg**[2,3,4,5,6], **Takuro Horii**[7], **Omar O. Abudayyeh**[2,3,4,5], **Mika Kimura**[7], **Patrick D. Hsu**[2,3,4,5], **Takanori Nakane**[1], **Ryuichiro Ishitani**[1], **Izuho Hatada**[7], **Feng Zhang**[2,3,4,5], **Hiroshi Nishimasu**[1,8,*], and **Osamu Nureki**[1,*]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

[3]McGovern Institute for Brain Research, Cambridge, MA 02139, USA

[4]Department of Brain and Cognitive Sciences, Cambridge, MA 02139, USA

[5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[6]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

[7]Laboratory of Genome Science, Biosignal Genome Resource Center, Institute for Molecular and Cellular Regulation, Gunma University, 3-39-15 Showa-machi, Maebashi, Gunma 371-8512, Japan

[8]JST, PRESTO, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

Summary

The RNA-guided endonuclease Cas9 cleaves double-stranded DNA targets complementary to the guide RNA, and has been applied to programmable genome editing. Cas9-mediated cleavage requires a protospacer adjacent motif (PAM) juxtaposed with the DNA target sequence, thus constricting the range of targetable sites. Here, we report the 1.7 Å resolution crystal structures of Cas9 from *Francisella novicida* (FnCas9), one of the largest Cas9 orthologs, in complex with a guide RNA and its PAM-containing DNA targets. A structural comparison of FnCas9 with other Cas9 orthologs revealed striking conserved and divergent features among distantly related CRISPR-Cas9 systems. We found that FnCas9 recognizes the 5′-NGG-3′ PAM, and used the structural information to create a variant that can recognize the more relaxed 5′-YG-3′ PAM. Furthermore, we demonstrated that pre-assembled FnCas9 ribonucleoprotein complexes can be

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

Accession Numbers

The atomic coordinates of the wild-type (the 5′-TGG-3′ PAM or the 5′-TGA-3′ PAM) and RHA (the 5′-TGG-3′ PAM) FnCas9 complexes have been deposited in the Protein Data Bank, with the PDB codes 5AAA, 5BBB and 5CCC, respectively.

Supplemental Information

Supplemental Information includes seven figures and two tables.

Author Contributions

H.H. solved the crystal structures and performed *in vitro* cleavage experiments, with assistance from H.N.; J.S.G., O.O.A. and P.D.H performed PAM screens and functional analysis; T.H. and I.H. performed genome editing analyses in mouse zygotes; T.N. and R.I. assisted with the structural determination; and H.H., H.N., 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.

microinjected into mouse zygotes to edit endogenous sites with the 5′-YG-3′ PAMs, thus expanding the target space of the CRISPR-Cas9 toolbox.

Introduction

The RNA-guided DNA endonuclease Cas9 from the CRISPR (clustered regularly interspaced short palindromic repeat)-Cas (CRISPR associated) systems associates with the dual RNA guides (CRISPR RNA (crRNA) and *trans*-activating RNA (tracrRNA)), or a synthetic single-guide RNA (sgRNA), and cleaves double-stranded DNA targets complementary to the guide RNA (Garneau et al., 2010; Deltcheva et al., 2011; Jinek et al., 2012; Gasiunas et al., 2012). Several Cas9 orthologs, such as *Streptococcus pyogenes* Cas9 (SpCas9) (Cong et al., 2013; Mali et al., 2013) and *Staphylococcus aureus* Cas9 (SaCas9) (Ran et al., 2015), have been harnessed for genome editing in eukaryotic cells. Besides the RNA–DNA complementarity, DNA recognition and cleavage by Cas9 also require the presence of a PAM (protospacer adjacent motif) immediately downstream of the target DNA sequence (Deveau et al., 2008; Garneau et al., 2010), thereby constraining the range of the targetable sequences in Cas9-mediated genome editing. Cas9 orthologs from different microbes recognize diverse PAM sequences, and SpCas9 (Mojica et al., 2009) and SaCas9 (Ran et al., 2015) recognize the 5′-NGG-3′ and 5′-NNGRRT-3′ PAMs, respectively.

The crystal structures of SpCas9 and SaCas9 have provided mechanistic insights into the RNA-guided DNA recognition and cleavage by Cas9 (Jinek et al., 2014; Nishimasu et al., 2014; Anders et al., 2014; Nishimasu et al., 2015; Jiang et al., 2015; Jiang et al., 2016). SpCas9 and SaCas9 adopt a bilobed architecture comprising recognition (REC) and nuclease (NUC) lobes, in which the guide RNA–target DNA heteroduplex is bound within the central channel formed between the two lobes. The PAM-containing, double-stranded DNA (PAM duplex) is accommodated between the Wedge (WED) and PAM-interacting (PI) domains, where the PAM nucleotides are recognized by a specific combination of amino-acid residues in the PI domain (Anders et al., 2014; Nishimasu et al., 2015). Furthermore, a structural comparison between SpCas9 and SaCas9 illuminated both the conserved and divergent structural features among the orthologous CRISPR-Cas9 systems (Nishimasu et al., 2015).

The Cas9 orthologs have highly divergent lengths and sequences, ranging from ~900 to ~1,600 amino acid residues, and the Cas9 from *Francisella novicida* (FnCas9) is one of the largest members (Chylinski et al., 2013; Hsu et al., 2014). FnCas9 consists of 1,629 amino acids and is significantly larger than other Cas9 orthologs, such as SpCas9 (1,368 amino acids) and SaCas9 (1,053 amino acids). Notably, a previous study reported that FnCas9 can mediate not only crRNA:tracrRNA-dependent DNA cleavage, but also scaRNA (small CRISPR/Cas-associated RNA):tracrRNA-dependent gene expression regulation (Sampson et al., 2013). However, the mechanisms by which FnCas9 executes its bifunctionality remain unknown. In addition, the potential use of FnCas9 in genome editing applications has not been explored.

In this study, we solved the high-resolution crystal structures of the 240 kDa FnCas9–sgRNA–target DNA complex, thus providing insights into the RNA-guided DNA recognition mechanism. The present structures enabled a comparison of FnCas9 with

SpCas9 and SaCas9, which revealed unexpected structural divergence among the distantly related CRISPR-Cas9 systems. We found that FnCas9 recognizes the 5′-NGG-3′ PAM, and used the structural information to create an engineered FnCas9 variant that recognizes the 5′-YG-3′ PAM. Furthermore, we demonstrated that pre-assembled FnCas9–sgRNA ribonucleoprotein (RNP) complexes can be injected into mouse zygotes to facilitate genome editing, thus expanding the target space in Cas9-mediated genome engineering.

Results

PAM specificity of FnCas9

Although a previous study indicated that FnCas9 recognizes the 5′-NG-3′ PAM (Fonfara et al., 2014), the FnCas9 PAM has not been fully characterized. To identify the FnCas9 PAM, we performed the PAM discovery assay, using a library of plasmid DNA targets with a degenerated 7-bp PAM sequence, as described previously (Ran et al., 2015; Zetsche et al., 2015). The results showed that the FnCas9 recognizes the 5′-NGG-3′ PAM (Figure 1A). Consistently, our *in vitro* cleavage assay, using purified FnCas9, an sgRNA and a plasmid containing a 20-bp target site with 5′-TNN-3′ PAMs, revealed that FnCas9 efficiently cleaves a plasmid target with the 5′-TGG-3′ PAM, while it exhibits slight activities toward those with the 5′-TGA-3′ and 5′-TAG-3′ PAMs (Figure 1B). Taken together, we concluded that the FnCas9 PAM is 5′-NGG-3′, with a slight tolerance for A at positions 2 and 3.

Overall structure of the FnCas9–sgRNA–DNA complex

To clarify the RNA-guided DNA cleavage mechanism, we solved the crystal structures of full-length FnCas9 (residues 1–1,629; N995A) in complex with a 94-nt sgRNA, a 30-nt target DNA strand and a 9-nt non-target DNA strand (containing either the 5′-TGG-3′ PAM or the 5′-TGA-3′ PAM) at 1.7 Å resolutions (Figures 2A–2D, Figure S1, Table S1). To prevent the potential cleavage of the target DNA during crystallization, we replaced the conserved catalytic residue (Asn995) in the HNH domain with alanine. Since the two quaternary complex structures are virtually identical, the following discussions are based on the 5′-TGG-3′ PAM complex structure, unless otherwise stated.

The crystal structure revealed that FnCas9 comprises seven domains—the REC1–3, RuvC, HNH, WED and PI domains (Figures 2A–2D). The REC2 domain is inserted into the REC1 domain, and the REC1 and REC3 domains are connected by a linker loop (referred to as the REC1–REC3 linker). The RuvC domain is composed of the three RuvC motifs (RuvC I–III). As in SpCas9 and SaCas9, RuvC-I and RuvC-III are connected to the REC1 and WED domains via the bridge helix and the phosphate lock loop, respectively. The HNH domain is connected to RuvC-II and RuvC-III via the α-helical linkers, L1 and L2, respectively. The WED and PI domains are connected by a linker loop (referred to as the WED–PI linker). The electron densities for the REC2 and HNH domains are relatively weak, indicating that the two domains are mobile.

Comparison of the overall structures of the Cas9 orthologs

A structural comparison of FnCas9 with SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) revealed unanticipated structural differences

(Figures 2E and 2F and Figure S2). SpCas9 and SaCas9 adopt bilobed architectures comprising the REC and NUC lobes (Figures 2E and 2F and Figures S2B and S2C). In the NUC lobe, the RuvC domain interacts with the PI domain, to form a platform responsible for the binding of the 3′ tracrRNA tail. In contrast, in FnCas9, the RuvC domain does not interact with the PI domain (Figure 2C and Figure S2A). Instead, the RuvC domain interacts with the REC3 domain, while the PI domain interacts with the WED domain, which contacts the REC1 and REC2 domains. These inter-domain interactions are mediated by the FnCas9-specific structural features (Figure S3). Accordingly, the 3′ tracrRNA tail of the FnCas9 sgRNA is primarily recognized by the REC2 and REC3 domains (Figure S2). Although FnCas9 and SpCas9 commonly have the REC2 domain, the FnCas9 REC2 domain adopts a new fold, and is structurally unrelated to the SpCas9 REC2 domain.

In addition to these divergent structural features, there are conserved structural features among these Cas9 orthologs. The guide:target heteroduplex is accommodated in the central channel between the RuvC and REC3 domains, while the PAM duplex is bound between the WED and PI domains (Figures 2C–2F and Figure S2). Moreover, similar to SpCas9 and SaCas9, the RuvC and HNH domains of FnCas9 have the RNase H and ββα-metal folds, respectively. These structural findings confirmed that the RNA-guided DNA cleavage mechanisms are highly conserved among the CRISPR-Cas9 systems. In the FnCas9 structure, the HNH domain is connected to the RuvC domain via the L1 and L2 linkers, and is distant from the target DNA strand, as in SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) (Figures 2C–2F). These structural observations suggest that, upon the binding of the double-stranded DNA target, the HNH domain approaches and cleaves the target DNA strand via drastic conformational changes in the L1 and L2 linkers, as observed in SpCas9 (Sternberg et al., 2015; Jiang et al., 2016).

Structures of the sgRNA and the target DNA

The sgRNA comprises the guide region, the repeat:antirepeat duplex, tetraloop, stem loop 1, the SL1–SL2 linker region, and stem loop 2 (Figures 3A and 3B). The guide region (G1–G21) and the target DNA strand (dC1–dC21) form the guide:target heteroduplex, while the target DNA strand (dC(−9)–dA(−1)) and the non-target DNA strand (dT1*–dG9*) form the PAM duplex. The repeat:antirepeat duplex consists of ten Watson-Crick base pairs (U23:A51, U24:A50, C26:G48–G28:C44 and G31:C44–C35:G40), three non-canonical base pairs (G22:U52, U25:U49 and U30:U45), and the U29 bulge, which interacts with G28 and U47 (Figure 3A and Figure S4A). The repeat:antirepeat duplex and stem loop 1 are connected by C53, which is equivalent to A51 in the SpCas9 sgRNA and A55 in the SaCas9 sgRNA (Figures 3C–3E). Stem loop 1 consists of two base pairs (A54:G62 and U55:A61) and five unpaired nucleotides (U56–A60) (Figure 3 and Figure S4B). The basal region of stem loop 1 is stabilized by a hydrogen-bonding network between G62 and C53/A54 and a stacking interaction between C53 and U63 (Figure S4B). Stem loops 1 and 2 are connected by a 9-nt linker, which contains a Watson-Crick base pair (A64:U68) and adopts a U-shaped structure (Figures 3A and 3C). Stem loop 2 consists of five Watson-Crick base pairs (C72:G94, G74:C92, A75:U91, G81:C87 and U82:A86), a wobble base pair (G73:U93), five unpaired nucleotides (U78, U83–G85 and C89), and two base triples (C76:G90•C79 and U80:A88•C77) (Figure 3A and Figure S4C).

Comparison of the orthogonal sgRNA scaffolds

SpCas9 and SaCas9 have the structurally diverse REC1 and WED domains, which recognize distinct structural features (the repeat:antirepeat duplex and stem loop 1) of their cognate sgRNAs, thereby defining the orthogonality between cognate Cas9–sgRNA pairs (Nishimasu et al., 2014; Anders et al., 2014; Nishimasu et al., 2015). The present structure revealed that the repeat:antirepeat duplex and stem loop 1 of the FnCas9 sgRNA have structural features distinct from those of the SpCas9 and SaCas9 sgRNAs (Figures 3C–3E). Furthermore, there are notable structural differences in their 3′ tracrRNA tails. The stem loops in the SpCas9 and SaCas9 sgRNAs adopt an A-form helix, whereas stem loop 2 in the FnCas9 sgRNA contains the two base triples and adopts a distorted structure. In addition, stem loops 1 and 2 are connected by a single-stranded linker in the SpCas9 and SaCas9 sgRNAs, whereas stem loops 1 and 2 are connected by the U-shaped linker in the FnCas9 sgRNA (Figures 3C–3E). Consequently, stem loop 2 in the FnCas9 sgRNA is uniquely directed toward the REC1 and REC3 domains, unlike the SpCas9 and SaCas9 sgRNAs (Figures 3C–3D and Figure S2). These structural differences can explain the observed orthogonality between these CRISPR-Cas9 systems (Fonfara et al., 2014) (Figure 3F). A structural comparison also revealed the presence of a structurally conserved core region in their sgRNAs (Figure 3G). In the FnCas9 sgRNA, the core region consists of the basal stem regions in the repeat:antirepeat duplex (G22:U52, U23:A51 and U24:A50) and stem loop 1 (C53, A54:G62 and U55:A61, U63) (Figures 3A and 3G). The sgRNA core regions are recognized by their cognate Cas9s in a similar manner (described later).

Recognition of the guide:target heteroduplex

A structural comparison of FnCas9 with SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) revealed that their REC1 domains share a 4-helix bundle core, consisting of the bridge helix and three α-helices (α1–α3). In these CRISPR-Cas9 systems, the PAM-proximal sugar-phosphate backbone of the heteroduplex is recognized by the 4-helix bundle core in a similar manner (Figures 4 and 5A–5C). Notably, the backbone phosphate in the PAM-proximal, 8-nt “seed” region in the sgRNA is extensively recognized by a conserved arginine cluster in the bridge helix (Figure S5A), consistent with the functional significance of the complementarity in the “seed” region in the heteroduplex (Jinek et al., 2012; Hsu et al., 2013; Ran et al., 2015). In SpCas9 and SaCas9, the PAM-distal region in the heteroduplex is recognized by the REC3 domain, which adopts a similar fold comprising 11 α-helices (Figures 5A and 5B). In contrast, the REC3 domain of FnCas9 adopts a new fold comprising 20 α-helices and a β-hairpin (Figure 5C), with a structural zinc ion coordinated by Cys460, Cys657, Cys814 and Cys817 (Figure 5D). Despite the lack of structural similarity, the REC3 domain of FnCas9 also recognizes the PAM-distal region in the heteroduplex, primarily in a sequence-independent manner, as in SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) (Figures 4 and 5D). Together, these structural observations explain the RNA-guided DNA targeting mechanism of FnCas9.

In SpCas9 (Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015), the backbone phosphate group between nucleotides at the +1 and −1 positions in the target DNA strand (referred to as the +1 phosphate) interacts with the phosphate lock loop between the RuvC

and WED domains, thereby facilitating the unwinding of double-stranded DNA targets. In FnCas9, Asp1242 and Gly1243 in the phosphate lock loop interact with the +1 phosphate in the target DNA strand (Figure S5B), indicating that the DNA unwinding mechanism is conserved among the CRISPR-Cas9 systems.

Recognition of the sgRNA scaffold

The REC1 and WED domains of FnCas9 are structurally distinct from those of SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015), and their REC1 and WED domains recognize the repeat:antirepeat duplex in species-specific manners (Figures 5A–5C and Figures S6A–S6C). Notably, the WED domain of FnCas9 (225 residues) is larger than those of SpCas9 (27 residues) and SaCas9 (121 residues), and adopts a new fold consisting of 3- and 4-stranded anti-parallel β-sheets, a β-hairpin and seven α-helices (Figure 5C). The FnCas9 WED domain interacts with the REC1 and REC2 domains to form a tunnel, which encloses the repeat:antirepeat duplex (Figure S6C). In the tunnel, U29, U24/A51 and G43 in the repeat:antirepeat duplex are recognized by Gln93/Gly331, Gln1466 and Glu1401 in base-specific manners, respectively (Figure S6D). The 3′ tracrRNA tail is also recognized by FnCas9 in a manner distinct from those of SpCas9 and SaCas9 (Figure S2). The SL1–SL2 linker interacts with the REC3/RuvC domains and the phosphate lock loop, while stem loop 2 interacts with the REC1/REC3 domains and the REC1–REC3 linker. In particular, the flipped-out C89 and the two base triples in stem loop 2 form hydrogen bonds with Asn454 and Asn454/Gln522/Lys660, respectively (Figure S6E). In addition, the β-hairpin in the REC3 domain interacts with the sugar-phosphate backbone in the U-shaped linker and stem loop 2 (Figure S6F).

The sgRNA core region is recognized by the bridge helix, the REC1 domain and the phosphate lock loop (Figure 5E). The backbone phosphate groups of A50–U52 in the repeat:antirepeat duplex interact with Gln69/Lys72 (bridge helix) and Ser96/Asn100 (REC1), while the 2′-OH of U23 hydrogen bonds with the main-chain carbonyl group of Ile1244 (phosphate lock loop). The backbone phosphate groups of A61–A64 in stem loop 1 interact with Arg55/Arg62/Arg63 (bridge helix) and Arg1237 (phosphate lock loop). The C53-U63 stacking pair is sandwiched between the side chains of Arg58 and Met1239, while C53 forms base-specific contacts with Arg1237, Met1239 and Thr1240 (Figure 5E). Notably, SpCas9 and SaCas9 recognize the core regions of their cognate sgRNAs in similar manners (Figure S7), consistent with the notion that the core regions of the crRNA:tracrRNA guides are highly conserved among the CRISPR-Cas9 systems, and are important for Cas9-mediated DNA cleavage (Briner et al., 2014). Intriguingly, the adenine nucleotides between the repeat:antirepeat duplex and stem loop 1 (A51 in the SpCas9 sgRNA and A55 in the SaCas9 sgRNA), which are equivalent to C53 in the FnCas9 sgRNA, adopt the *syn* conformation, and form analogous interactions with the proteins (Figure S7).

Recognition of the 5′-NGG-3′ PAM

In the present structure, the PAM duplex is sandwiched between the WED and PI domains, and the PAM sequences are read by the PI domain (Figures 6A and 6B). dT1* does not contact the protein (Figure 6C). The O6 and N7 of dG2* form bidentate hydrogen bonds with Arg1585 in the PI domain, while the N3 of dG2* forms a hydrogen bond with Ser1473

in the WED–PI linker (Figure 6C). In the 5′-TGG-3′ PAM complex, the O6 and N7 of dG3* form bidentate hydrogen bonds with Arg1556 (Figure 6C), whereas in the 5′-TGA-3′ PAM complex, the N7 of dA3* forms only a single hydrogen bond with Arg1556 (Figure 6D), consistent with the higher activity of FnCas9 with the 5′-NGG-3′ PAM than the 5′-NGA-3′ PAM. In addition, dA(−1) in the target DNA strand forms a stacking interaction with Arg1474 in the WED–PI linker (Figure 6C). The mutations of these residues reduced the *in vitro* DNA cleavage activity of FnCas9 (Figure 6E), confirming the functional significance of Ser1473, Arg1474, Arg1556 and Arg1585. In addition to these direct interactions, dC(−2), dG2* and dG3* form water-mediated hydrogen bonds with Glu1449, Asp1470 and Lys1451 in the WED domain, respectively. Together, these structural findings explain the mechanism of the 5′-NGG-3′ PAM recognition by FnCas9.

The PI domains of SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) share a similar core fold comprising two distorted, anti-parallel β-sheets (β1–β3 and β4–β9), with the β5–β7 region responsible for the PAM recognition (Figures 6F and 6G). In SpCas9, the 5′-NGG-3′ PAM is recognized by Arg1333/Arg1335 in the β7 loop (Anders et al., 2014), whereas in SaCas9, the 5′-NNGRRT-3′ PAM is recognized by Asn985/Asn986/Arg991/Arg1015 in the β5–β7 region (Nishimasu et al., 2015). The FnCas9 structures revealed that, despite the lack of sequence homology, the PI domain of FnCas9 adopts a similar core fold to those of SpCas9 and SaCas9 (Figure 6H). Whereas, in SpCas9 and SaCas9, the β8 and β9 strands in the PI domain are responsible for the interaction with the RuvC domain, the FnCas9 PI domain lacks the equivalent strands, consistent with the structural observation that the RuvC and PI domains do not interact in FnCas9. In FnCas9, the 5′-NGG-3′ PAM is recognized by Arg1556 in the β5–β6 loop and Arg1585 in the β6–β7 loop. Although both SpCas9 and FnCas9 recognize the 5′-NGG-3′ PAM with a pair of arginine residues (Arg1333/Arg1335 in SpCas9 and Arg1585/Arg1556 in FnCas9), these arginine pairs are located at different positions, due to the substantial difference in their relative arrangement between the PI domain and the PAM duplex (Figures 6G and 6H). In SpCas9, the 3rd G in the 5′-NGG-3′ PAM is recognized by the Arg1335 side chain, which is anchored by a salt bridge with Glu1219, consistent with the specific recognition of the 3rd G by SpCas9 (Figure 6G). In contrast, in FnCas9, the Arg1556 side chain does not form such a contact with the proximal residues (Figure 6H), explaining why, unlike SpCas9, FnCas9 can also recognize the 3rd A in the PAM, albeit with low efficiency. Together, these structural findings reinforced the notion that the Cas9 orthologs recognize diverse PAM sequences using distinct sets of PAM-interacting residues in the PI domains.

Structure-guided engineering of the FnCas9 PAM specificity

To expand the target space in genome engineering, we sought to rationally design the FnCas9 variant that can recognize the 5′-NG-3′ PAM. To eliminate the interaction between Arg1556 and the 3rd G in the 5′-NGG-3′ PAM, we first prepared the R1556A variant, in which Arg1556 is substituted with an alanine. *In vitro* cleavage experiments confirmed the decreased activities of the R1556A variant for the 5′-TGA-3′ and 5′-TGG-3′ PAMs (Figure 7A). We hypothesized that the reduced activity of the R1556A variant could be recovered by additional base-non-specific interactions between the protein and the PAM duplex. We thus introduced several mutations into the R1556A variant, which would potentially form new

interactions with the backbone phosphates of the PAM duplex. We then examined their effects on the *in vitro* cleavage activity for the 5′-TGN-3′ PAMs, and found that the E1369R/E1449H/R1556A triple mutant (referred to as the RHA FnCas9 variant) cleaves the target sites with the 5′-TGN-3′ PAMs with efficiencies comparable to that of the wild-type FnCas9 for the target sites with the 5′-TGA-3′ PAM (Figure 7A). However, unexpectedly, the PAM discovery analyses indicated that RHA FnCas9 recognizes 5′-YG-3′, but not 5′-NG-3′, as the PAM (Figure 7B). Consistently, *in vitro* cleavage assays revealed that RHA FnCas9 has a stronger preference for the 1st Y, as compared with wild-type FnCas9 (Figure 7C). Together, these results demonstrated that FnCas9 can be engineered to recognize the 5′-YG-3′ PAM by the E1369R/E1449H/R1556A substitutions.

To elucidate the 5′-YG-3′ PAM recognition mechanism, we solved the crystal structure of RHA FnCas9 in complex with the sgRNA and the DNA target with the 5′-TGG-3′ PAM, at 1.7 Å resolution (Figures 7D and 7E, Table S1). As in the original 5′-TGG-3′ PAM complex, dG2* is recognized by Ser1473 and Arg1585, while dA(−1) forms a stacking interaction with Arg1474 (Figure 7D). As the purine bases are larger than the pyrimidine bases, the purine nucleotides at the −1 position in the target DNA strand would form a favorable stacking interaction with Arg1474, thereby explaining the preference of RHA FnCas9 for the 1st Y in the 5′-YG-3′ PAM. In contrast to the original structure, dG3* is not recognized by the protein, due to the R1556A substitution (Figure 7D). Notably, the newly incorporated Arg1369 and His1449 interact with the backbone phosphate group between dC(−2) and dA(−1) in the target DNA strand (Figure 7D), confirming that these base-non-specific interactions contribute to compensate for the loss of the base-specific interactions between Arg1556 and the 3rd G in the 5′-NGG-3′ PAM. Unlike wild-type FnCas9, RHA FnCas9 requires the 1st Y in the 5′-YG-3′ PAM. The difference in their 1st PAM nucleotides suggested that the interactions between Arg1369/His1449 and the PAM duplex in RHA FnCas9 are not sufficient to fully compensate for the loss of the interactions between Arg1556 and the 3rd PAM nucleotides. The requirement of the 1st Y by the RHA FnCas9 may be eliminated by additional substitutions that enhance the PAM duplex binding, thereby achieving the recognition of the 5′-NG-3′ PAM. Together, our structural data explain the 5′-YG-3′ PAM recognition mechanism of the RHA FnCas9 variant.

FnCas9-mediated genome editing in mouse zygotes

Finally, we examined whether FnCas9 can be harnessed for genome editing in mammalian cells. FnCas9 failed to induce indels efficiently, when expressed in human embryonic kidney 293FT cells (data not shown), as in the cases of many Cas9 orthologs, except for a few such as SpCas9 (Cong et al., 2013; Mali et al., 2013) and SaCas9 (Ran et al., 2015). We reasoned that one possibility is the inefficient assembly of the vector-expressed FnCas9 and the sgRNA in human cells. We thus mixed the purified FnCas9 protein, the 60-nt crRNA and the 114-nt tracrRNA, to assemble an FnCas9–crRNA:tracrRNA RNP complex targeted to the mouse *Tet1EX4* locus with the 5′-TGN-3′ PAMs. We microinjected the pre-assembled FnCas9 RNP complex into mouse zygotes, and monitored FnCas9-mediated indel formation four days after microinjection. The FnCas9 RNP complex was able to induce indels at the *Tet1EX4* target sites with 5′-TGA-3′ and 5′-TGG-3′ PAMs, but not at those with 5′-TGT-3′ and 5′-TGC-3′ PAMs (Figure 7E), while FnCas9 showed *in vitro* preference for the 5′-

NGG-3′ PAM over the 5′-NGA-3′ PAM. Notably, unlike wild-type FnCas9, RHA FnCas9 was able to induce indels at the *Tet1EX4* sites with the 5′-TGN-3′ PAMs (Figure 7E). In contrast, RHA FnCas9 failed to edit the *Tet1EX4* sites with the 5′-GGN-3′ PAMs (data not shown), consistent with the requirement of the 1st Y in the PAM by RHA FnCas9. Together, these results demonstrated that the wild-type and RHA FnCas9 RNP complexes can be microinjected into mouse zygotes, to facilitate genome editing in target sites with the 5′-NGG-3′ and 5′-YG-3′ PAMs, respectively.

Discussion

In this study, we present the high-resolution structures of the FnCas9–sgRNA–DNA complex. A structural comparison of FnCas9 with SpCas9 (Nishimasu et al., 2014; Anders et al., 2014) and SaCas9 (Nishimasu et al., 2015) enhanced our understanding of the divergence in orthologous CRISPR-Cas9 systems. The present structure revealed that the WED domain of FnCas9 adopts a new fold and is structurally distinct from those of SpCas9 and SaCas9, thereby reinforcing the notion that the WED domains are highly divergent and critical for defining the orthogonality among the CRISPR-Cas9 systems. Although it shares a similar core fold with those of SpCas9 and SaCas9, the PI domain of FnCas9 recognizes the 5′-NGG-3′ PAM in a unique manner, revealing the new repertoire of diverse PAM recognition mechanisms. Furthermore, the present structure revealed unexpected structural divergence in the CRISPR-Cas9 systems. First, unlike SpCas9 and SaCas9, FnCas9 does not adopt a bilobed architecture. Second, the REC domain of FnCas9 has distinct structural features, as compared with those of SpCas9 and SaCas9. Third, there are notable structural differences in their sgRNA scaffolds. Stem loop 2 of the FnCas9 sgRNA contains the base triples and is recognized by the REC domain, whereas those of the SpCas9 and SaCas9 sgRNAs adopt canonical A-form structures and are recognized by the RuvC and PI domains. These striking structural differences may be related to the FnCas9-specific scaRNA:tracrRNA-mediated RNA targeting (Sampson et al., 2013). The 5′ and 3′ regions of the tracrRNA are complementary to the scaRNA and the target RNA, respectively (Sampson et al., 2013). In the present structure, the 5′ region of tracrRNA base pairs with crRNA to form the repeat:antirepeat duplex, suggesting that tracrRNA and scaRNA form a similar duplex structure, which is recognized by the REC and WED domains. In contrast, the 3′ region of tracrRNA forms the stem loops, and is not available for base pairing with the target RNA. It is thus possible that the 3′ region of tracrRNA and the target RNA may form a distinct, scaRNA-dependent duplex structure, which is recognized by the REC lobe. Further studies will be required to elucidate the underlying mechanism of the scaRNA:tracrRNA-mediated RNA targeting. Moreover, the present structure illuminated the highly conserved features across the CRISPR-Cas9 systems. Similar to SpCas9 and SaCas9, FnCas9 has the bridge helix and the phosphate lock loop, indicating that the RNA-guided DNA cleavage mechanism is conserved among the CRISPR-Cas9 systems.

We showed that FnCas9 recognizes the 5′-NGG-3′ PAM, and rationally designed the RHA variant that recognizes the 5′-YG-3′ PAM. Furthermore, we demonstrated that in mouse zygotes, pre-assembled RNP complexes of wild-type and RHA FnCas9 can edit endogenous genomic loci with the 5′-NGG-3′ and 5′-YG-3′ PAMs, respectively, although FnCas9 failed to facilitate genome editing when expressed in human cells. Since the other Cas9 orthologs

characterized so far do not have PAM specificities for either 5′-NGR-3′ or 5′-YG-3′, both the wild-type and RHA FnCas9s will contribute to expanding the target space in Cas9-mediated genome editing. Previous studies showed that the delivery of pre-assembled Cas9–sgRNA RNP complexes enables genome editing with improved efficiency and specificity in human cells (Lin et al., 2014; Kim et al., 2014; Zuris et al., 2015), mouse and zebrafish embryos (Sung et al., 2014) and plants (Woo et al., 2015), as compared to the transfection of plasmids encoding Cas9 and sgRNA. Our results suggested that, in addition to these advantages, the delivery of pre-assembled Cas9–sgRNA RNP complexes might provide a general means to rescue the *in vivo* cleavage activities of some Cas9 orthologs that fail to function in mammalian cells.

The RHA FnCas9 structure demonstrated that the loss of base-specific interactions with the PAM can be partly compensated by newly incorporated, base-non-specific interactions, thereby achieving altered PAM specificities. Recent studies showed that SpCas9 (Kleinstiver et al., 2015a) and SaCas9 (Kleinstiver et al., 2015b) can be engineered by molecular evolution strategies to exhibit altered PAM specificities. The VQR and VRER SpCas9 variants recognize the 5′-NGA-3′ and 5′-NGCG-3′ PAMs, respectively, whereas the KKH SaCas9 variant recognizes the 5′-NNNRRT-3′ PAM. Intriguingly, the G1218R substitution in VRER SpCas9 and the E782K/N968K substitutions in KKH SaCas9 are located close to the phosphate backbone in the PAM duplex, suggesting that these newly incorporated, positively charged residues interact with the phosphate backbone in the PAM duplex, as in the case of the E1369R/E1449H substitutions in RHA FnCas9. Thus, our strategy to compensate for the loss of base-specific interactions with the PAM nucleotides, by including additional base-non-specific interactions to alter Cas9 PAM specificities, may be generally applicable to other Cas9 orthologs, such as the widely used SpCas9 and SaCas9.

Experimental Procedures

Sample Preparation

The gene encoding full-length FnCas9 (residues 1–1,629) was cloned between the *Nde*I and *Xho*I sites of the modified pE-SUMO vector (LifeSensors), and the N995A mutation was introduced by a PCR-based method. The FnCas9 N995A mutant protein was expressed at 37°C in *Escherichia coli* Rosetta 2 (DE3) (Novagen), and was purified by chromatography on Ni-NTA Superflow resin (QIAGEN). The eluted protein was incubated overnight at 4°C with TEV protease to remove the His_6-SUMO-tag, and was further purified by chromatography on Ni-NTA, Mono S (GE Healthcare) and HiLoad Superdex 200 16/600 (GE Healthcare) columns. The SeMet-labeled FnCas9 N995A mutant and the RHA FnCas9 N995A mutant were expressed in *E. coli* B834 (DE3) (Novagen) and *E. coli* Rosetta2 (DE3) respectively, and were purified using a similar protocol to that for the native protein. The 94-nt sgRNA was transcribed *in vitro* with T7 RNA polymerase, using a PCR-amplified DNA template, and was purified by 10% denaturing polyacrylamide gel electrophoresis. To facilitate crystallization, the internal loop in the repeat:antirepeat duplex was replaced by G:C base pairs (Figure S1A). The 30-nt target DNA strand and the 9-nt non-target DNA strand were purchased from Sigma-Aldrich. The purified FnCas9 protein was mixed with the sgRNA, the target DNA strand and the non-target DNA strand (containing either the 5′-

TGG-3′ PAM or the 5′-TGA-3′ PAM) (molar ratio, 1:1.5:2.3:4), and then the reconstituted FnCas9–sgRNA–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. For *in vitro* cleavage assays, the His_6-tagged FnCas9 proteins were expressed at 37°C in *E. coli* Rosetta 2 (DE3), and were purified by chromatography on Ni-NTA and HiTrap SP HP (GE Healthcare) columns. The purified SpCas9 and SaCas9 proteins and their cognate sgRNAs were prepared as described previously (Nishimasu et al., 2014; Nishimasu et al., 2015). For microinjection experiments, wild-type and RHA FnCas9 were prepared using a similar protocol to that for the N995A mutant used for crystallization.

Crystallography

The purified FnCas9–sgRNA–DNA complex (with the 5′-TGG-3′ PAM or 5′-TGA-3′ PAM) was crystallized at 20°C. The initial crystals were obtained by mixing 0.1 μl of complex solution ($A_{260\ nm}$ = 15) and 0.1 μl of reservoir solution (9–11% PEG 3,350, 0.2 M calcium acetate and 0.1 M sodium citrate, pH 5.0), using the sitting drop vapor diffusion method. The crystals were improved by the microseeding method, using Seed Bead (Hampton Research). The initial crystal was harvested in stabilization solution (9–11% PEG 3,350, 0.2 M calcium acetate and 0.1 M sodium acetate, pH 5.0), and then crushed using the Seed Bead to prepare the seed stock solution. The crystallization drops were formed by mixing 1 μl of complex solution ($A_{260\ nm}$ = 15) and 1 μl of the seed stock solution, and then were incubated against 0.5 ml of reservoir solution (9–11% PEG 3,350, 0.2 M calcium acetate and 0.1 M sodium acetate, pH 5.0), using the hanging drop vapor diffusion method. The SeMet-labeled FnCas9 complex (the 5′-TGG-3′ PAM) and the RHA FnCas9 complex (the 5′-TGG-3′ PAM) were crystallized under similar conditions, using the seed stock solution containing the wild-type crystals. X-ray diffraction data were collected at 100 K on beamlines BL41XU at SPring-8 (Hyogo, Japan) and PXI at the Swiss Light Source (Villigen, Switzerland). The crystals were cryoprotected in reservoir solution supplemented with 25% ethylene glycol. The X-ray diffraction data were processed using XDS (Kabsch, 2010) and AIMLESS (Evans and Murshudov, 2013). The structure was determined by the Se-SAD method, using PHENIX AutoSol (Adams et al., 2010). The structural model was automatically built using Buccaneer (Cowtan, 2006), followed by manual model building using COOT (Emsley and Cowtan, 2004) and structural refinement using PHENIX (Adams et al., 2010). The final models of the wild-type (the 5′-TGG-3′ PAM or the 5′-TGA-3′ PAM) and RHA (the 5′-TGG-3′ PAM) FnCas9 complexes were refined, using their native data sets.

In vitro Cleavage Assay

In vitro plasmid DNA cleavage experiments were performed, essentially as described previously (Nishimasu et al., 2015). The *Bam*HI-linearized pUC119 plasmid (100 ng, 10 nM), containing the 20-nt target sequence and the PAM sequence, was incubated at 37°C for 30 min with the FnCas9-sgRNA complex (30 nM), in 10 μl of reaction buffer, containing 20 mM Tris-HCl, pH 8.5, 100 mM KCl, 10 mM $MgCl_2$ and 1 mM DTT. Reaction products were resolved on an ethidium bromide-stained 1% agarose gel, and then visualized using an Amersham Imager 600 (GE Healthcare). To test the orthogonality between Cas9 and sgRNA, each Cas9 ortholog (250 nM) and sgRNA (250 nM) were incubated at 37°C for 30

min with the plasmid DNA, in a reaction buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl and 10 mM $MgCl_2$.

In vitro PAM Screen

Randomized PAM plasmid libraries were constructed using synthesized oligonucleotides (IDT) consisting of 7 randomized nucleotides 3′ of a 20-nt target sequence, as previously described (Zetsche et al., 2015). *In vitro* cleavage reactions using wild-type FnCas9 or RHA FnCas9 with sgRNAs targeting the PAM library were fractionated on 2% agarose E-gels (Life Technologies). Bands corresponding to the un-cleaved target were extracted from the gel, using a Zymoclean Gel DNA Recovery Kit (Zymo Research), and the target PAM region was amplified and sequenced using a MiSeq (Illumina) with single-end 150 cycles. From the sequence data, the PAM regions were extracted, counted, and normalized to total reads for each sample. For a given PAM, enrichment was measured as the log ratio as compared to no protein control, with a 0.01 pseudocount adjustment. PAMs above a 3.5 enrichment threshold were collected and used to generate sequence logos (Crooks et al., 2004).

Microinjection and Typing the Blastocyst Embryos

All animal procedures were approved by the Animal Care and Experimentation Committee at Gunma University, and performed in accordance with approved guidelines. Female B6D2F1 mice (8–10 weeks old, CLEA Japan) were superovulated by the injection of 7.5 units of pregnant mare's serum gonadotropin (PMSG; ASKA Pharmaceutical), followed by 7.5 units of human chorionic gonadotrophin (hCG; ASKA Pharmaceutical) 48 h later, and mated overnight with B6D2F1 male mice. Zygotes were collected from oviducts 21 h after the hCG injection, and pronuclei-formed zygotes were placed into the M2 medium. Microinjection was performed using a microscope equipped with a microinjector (Narishige). The FnCas9 RNP complex was assembled by mixing the purified FnCas9 protein (0.2 μM), the 115-nt tracrRNA (0.9 μM) and the 60-nt crRNA (1.1 μM) targeting the mouse *Tet1EX4* locus (Table S2), and then the FnCas9 RNP complex (1 pl) was injected into the pronuclei of the zygotes. The crRNA and the tracrRNA were prepared by *in vitro* T7 transcription. After injection, all zygotes were cultured in M16 medium for 4 days. To detect indels, the targeted *Tet1EX4* region was amplified by PCR, using genomic DNA extracted from each blastocyst and the following primers: 5′-AGAACAAAGCCCCTGTGCTA-3′ (forward) and 5′-ACCACTCCAAGCCCTTTTCT-3′ (reverse). The PCR products were digested with a specific restriction enzyme that cleaves the Cas9 target site of the unmodified genomes, and then were analyzed by agarose gel electrophoresis. For the *Tet1EX4* target site with 5′-TGC-3′, indels were detected by a heteroduplex mobility assay (HMA). Briefly, the PCR products were reannealed and fractionated by polyacrylamide gel electrophoresis to detect the heteroduplex.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the beamline scientists at PXI at the Swiss Light Source and BL41XU at SPring-8 for assistance with data collection. I.H. is supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine, from the Japan Agency for Medical Research and Development, AMED. F.Z. is supported by the National Institutes of Health through NIMH (5DP1-MH100706 and 1R01-MH110049) and NIDDK (5R01DK097768-03), 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. H.N. is supported by JST, PRESTO, JSPS KAKENHI Grant Numbers 26291010 and 15H01463, and the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology. O.N. is supported by the Basic Science and Platform Technology Program for Innovative Biological Medicine from the Japan Agency for Medical Research and Development, AMED, and the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Technologies for creating next-generation agriculture, forestry and fisheries” (funding agency: Bio-oriented Technology Research Advancement Institution, NARO), and Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010; 66:213–221. [PubMed: 20124702]

Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014; 513:569–573. [PubMed: 25079318]

Briner AE, Donohoue PD, Gomaa AA, Selle K, Slorach EM, Nye CH, Haurwitz RE, Beisel CL, May AP, Barrangou R. Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell. 2014; 56:333–339. [PubMed: 25373540]

Chylinski K, Le Rhun A, Charpentier E. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. 2013; 10:726–737. [PubMed: 23563642]

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]

Cowtan K. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr. 2006; 62:1002–1011. [PubMed: 16929101]

Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–1190. [PubMed: 15173120]

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E. CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. Nature. 2011; 471:602–607. [PubMed: 21455174]

Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J Bacteriol. 2008; 190:1390–1400. [PubMed: 18065545]

Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr. 2004; 60:2126–2132. [PubMed: 15572765]

Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallogr D Biol Crystallogr. 2013; 69:1204–1214. [PubMed: 23793146]

Fonfara I, Le Rhun A, Chylinski K, Makarova KS, Lecrivain AL, Bzdrenga J, Koonin EV, Charpentier E. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. 2014; 42:2577–2590. [PubMed: 24270795]

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature. 2010; 468:67–71. [PubMed: 21048762]

Gasiunas G, Barrangou R, Horvath P, Siksnys V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A. 2012; 109:E2579–2586. [PubMed: 22949671]

Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014; 157:1262–1278. [PubMed: 24906146]

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol. 2013; 31:827–832. [PubMed: 23873081]

Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. A Cas9-guide RNA complex preorganized for target DNA recognition. Science. 2015; 348:1477–1481. [PubMed: 26113724]

Jiang F, Taylor DW, Chen JS, Kornfeld JE, Zhou K, Thompson AJ, Nogales E, Doudna JA. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science. 2016; doi: 10.1126/science.aad8282

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012; 337:816–821. [PubMed: 22745249]

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science. 2014; 343:1247997. [PubMed: 24505130]

Kabsch W. Xds. Acta Crystallogr D Biol Crystallogr. 2010; 66:125–132. [PubMed: 20124692]

Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res. 2014; 24:1012–1019. [PubMed: 24696461]

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales AP, Li Z, Peterson RT, Yeh JR, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015a; 523:481–485. [PubMed: 26098369]

Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, Zheng Z, Joung JK. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015b; 33:1293–1298. [PubMed: 26524662]

Lin S, Staahl BT, Alla RK, Doudna JA. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife. 2014; 3:e04766. [PubMed: 25497837]

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. RNA-guided human genome engineering via Cas9. Science. 2013; 339:823–826. [PubMed: 23287722]

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Almendros C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology. 2009; 155:733–740. [PubMed: 19246744]

Nishimasu H, Cong L, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. Crystal structure of *Staphylococcus aureus* Cas9. Cell. 2015; 162:1113–1126. [PubMed: 26317473]

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell. 2014; 156:935–949. [PubMed: 24529477]

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9. Nature. 2015; 520:186–191. [PubMed: 25830891]

Sampson TR, Saroj SD, Llewellyn AC, Tzeng YL, Weiss DS. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature. 2013; 497:254–257. [PubMed: 23584588]

Sternberg SH, LaFrance B, Kaplan M, Doudna JA. Conformational control of DNA target cleavage by CRISPR-Cas9. Nature. 2015; 527:110–113. [PubMed: 26524520]

Sung YH, Kim JM, Kim HT, Lee J, Jeon J, Jin Y, Choi JH, Ban YH, Ha SJ, Kim CH, et al. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res. 2014; 24:125–131. [PubMed: 24253447]

Woo JW, Kim J, Kwon SI, Corvalan C, Cho SW, Kim H, Kim SG, Kim ST, Choe S, Kim JS. DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol. 2015; 33:1162–1164. [PubMed: 26479191]

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell. 2015; 163:759–771. [PubMed: 26422227]

Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen ZY, Liu DR. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. Nat Biotechnol. 2015; 33:73–80. [PubMed: 25357182]

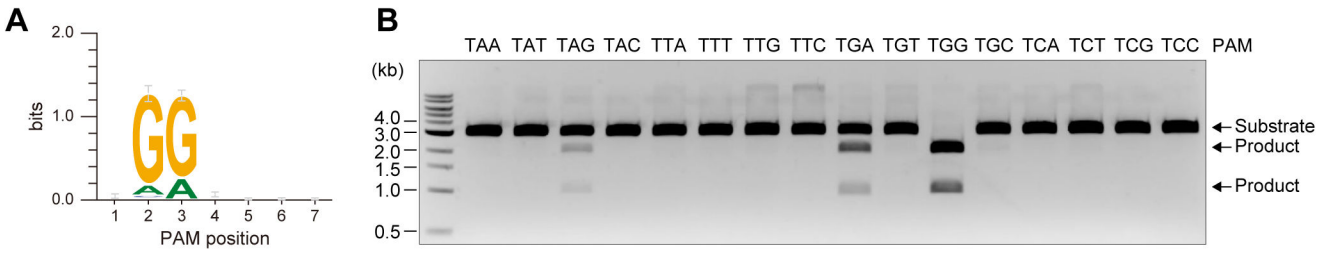

Figure 1. PAM specificity of FnCas9

(A) PAM discovery assay for FnCas9.

(B) *In vitro* DNA cleavage by FnCas9. The linearized plasmid targets with the 5′-TNN-3′ PAM were incubated with the purified FnCas9–sgRNA complex.

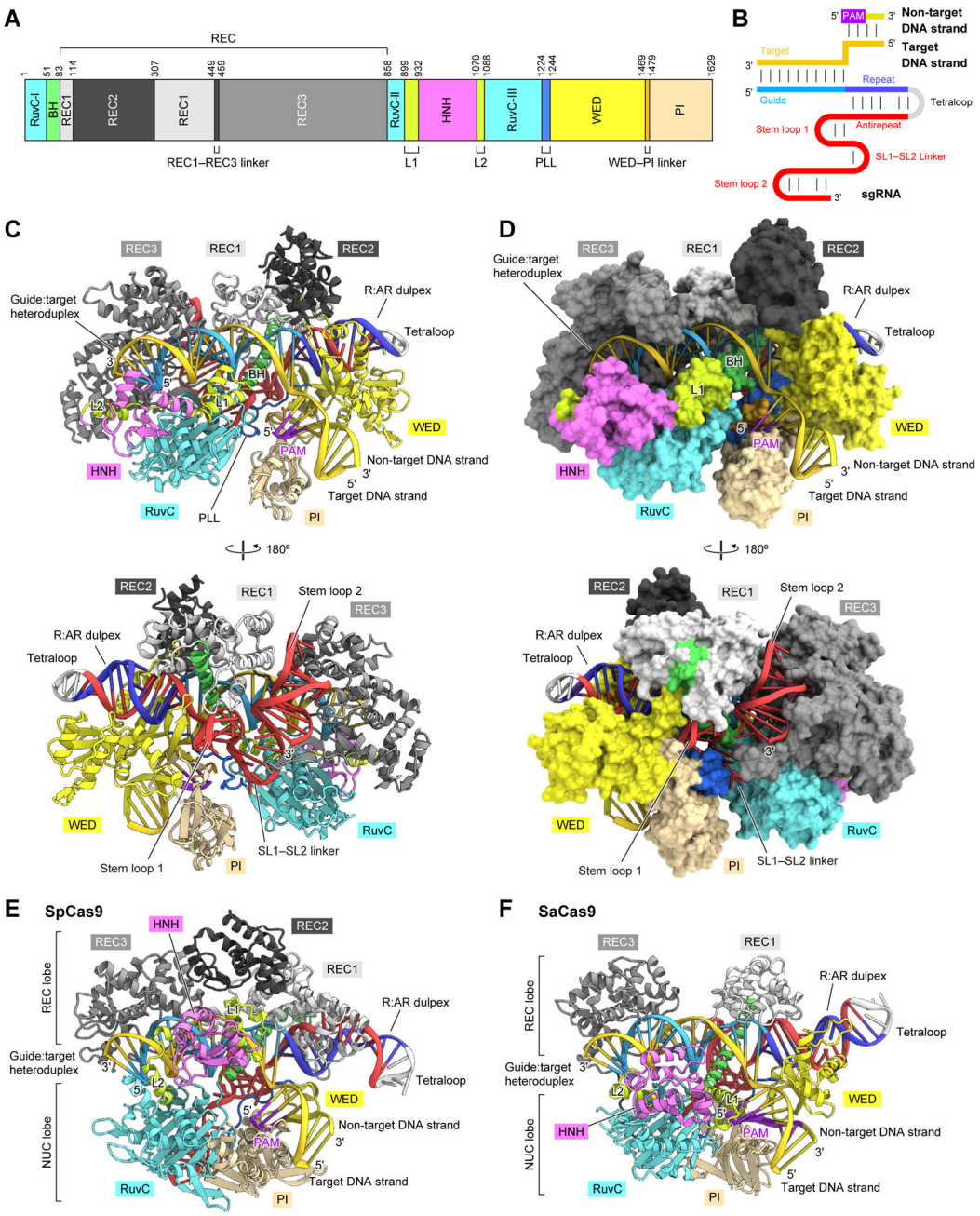

Figure 2. Overall structure of the FnCas9–sgRNA–DNA complex

(A) Domain organization of FnCas9. BH, bridge helix; PLL, phosphate lock loop.

(B) Schematic representation of the sgRNA–DNA.

(C and D) Ribbon (C) and surface (D) representations of the FnCas9–sgRNA–DNA complex.

(E and F) Crystal structures of SpCas9 (PDB: 4UN3) (E) and SaCas9 (PDB: 5CZZ) (F).

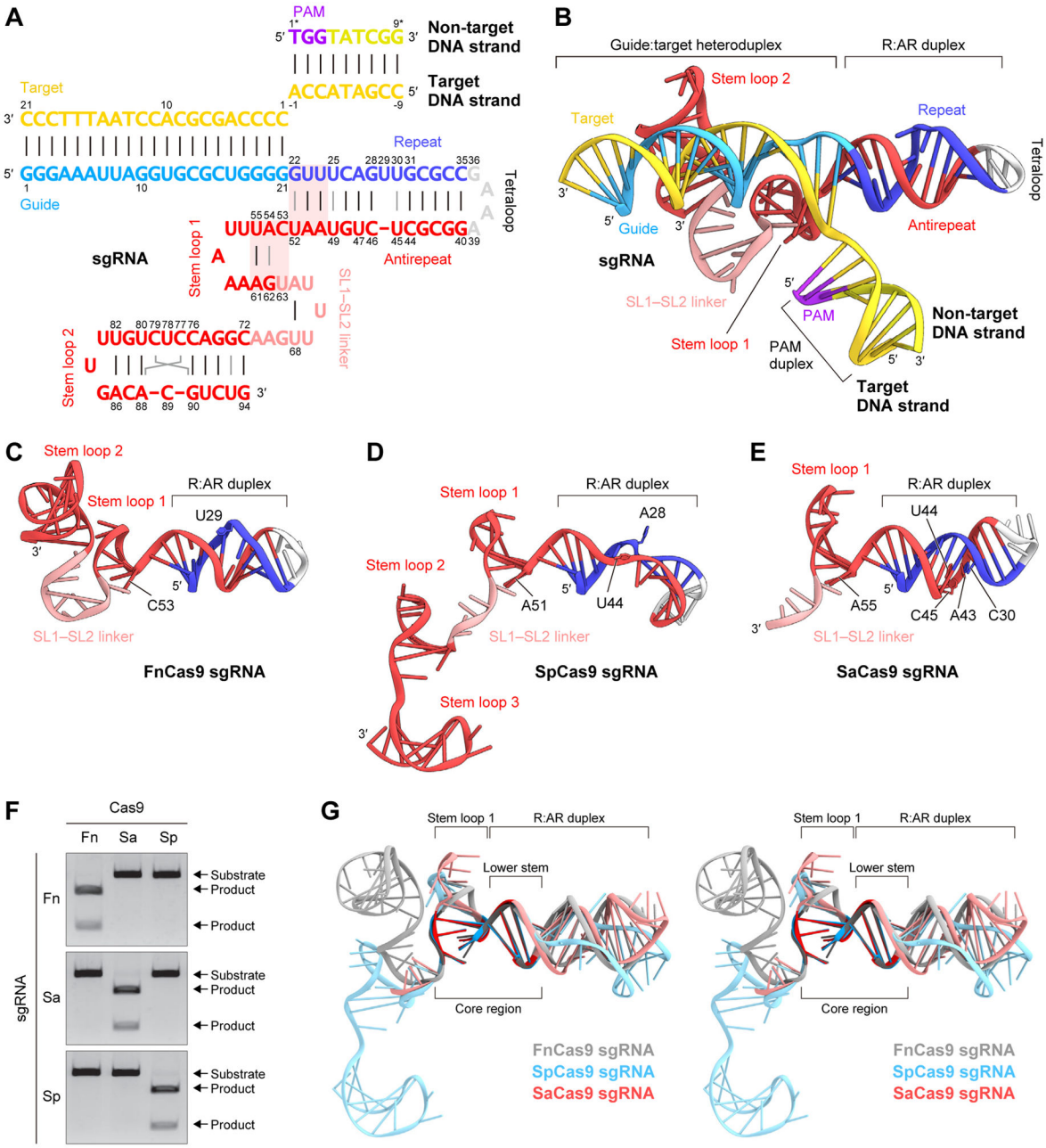

Figure 3. Structure of sgRNA–DNA

(A) Schematic representation of the FnCas9 sgRNA scaffold. The sgRNA core fold is highlighted in pink.

(B) Structure of the FnCas9 sgRNA–DNA.

(C–E) sgRNA scaffolds for FnCas9 (C), SpCas9 (D) and SaCas9 (E). The guide regions are omitted for clarity.

(F) Comparison of the sgRNA scaffolds of FnCas9 (red), SpCas9 (blue) and SaCas9 (gray).

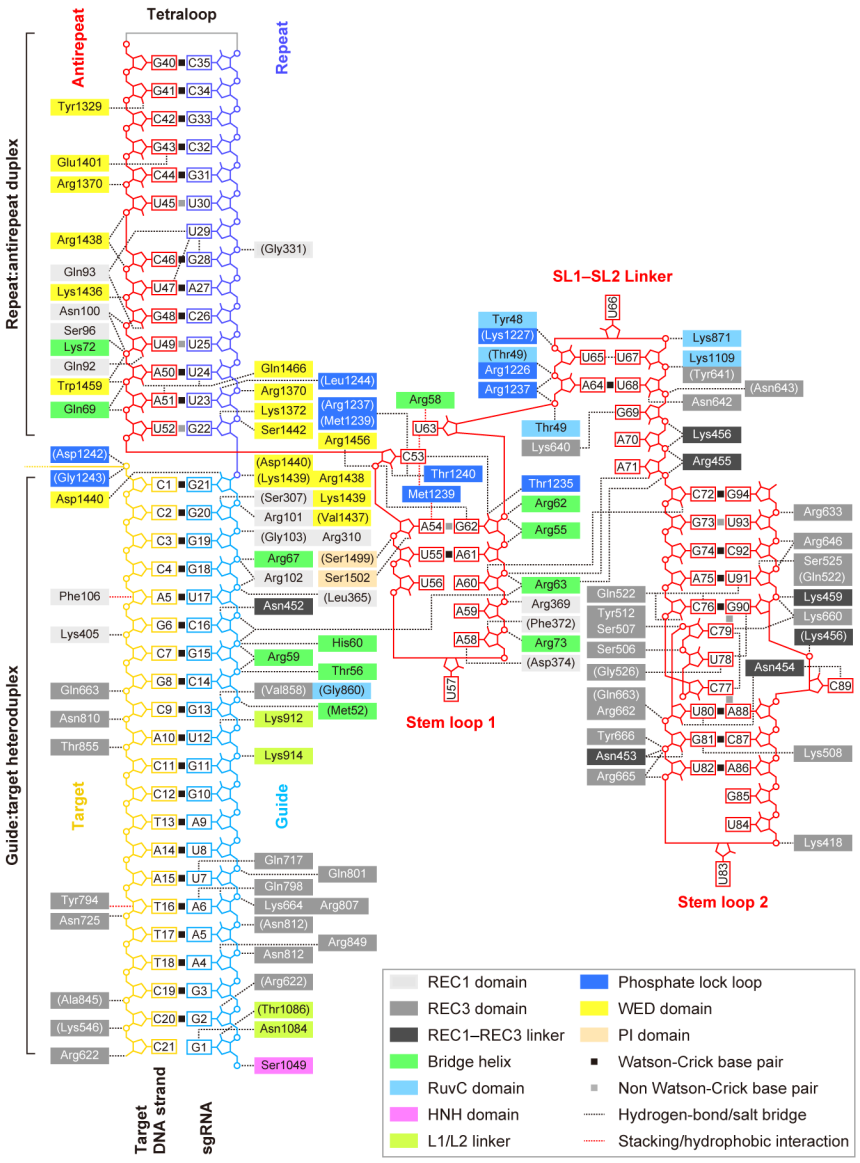

Figure 4. Schematic of the nucleic acid recognition by FnCas9

Residues that interact with nucleic acids via their main chain are shown in parentheses. Water-mediated hydrogen bonds are not shown for clarity.

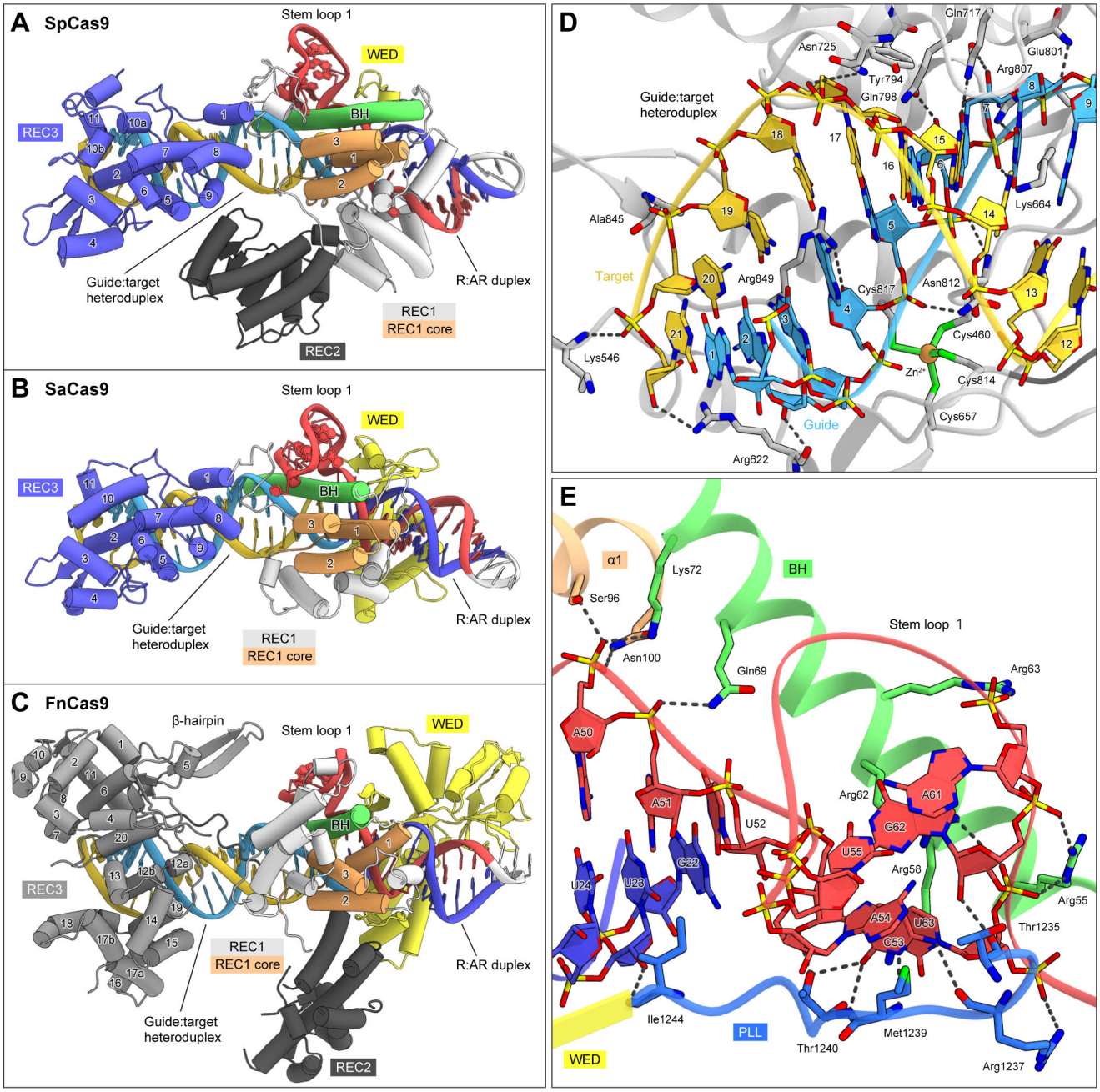

Figure 5. Recognition of the sgRNA and the target DNA by the Cas9 orthologs

(A–C) Recognition of the nucleic acids by the REC/WED domains of SpCas9 (A), SaCas9 (B) and FnCas9 (C).

(D) Recognition of the RNA–DNA heteroduplex by FnCas9. Hydrogen-bonding and electrostatic interactions are indicated by gray dashed lines.

(E) Recognition of the sgRNA core fold by FnCas9.

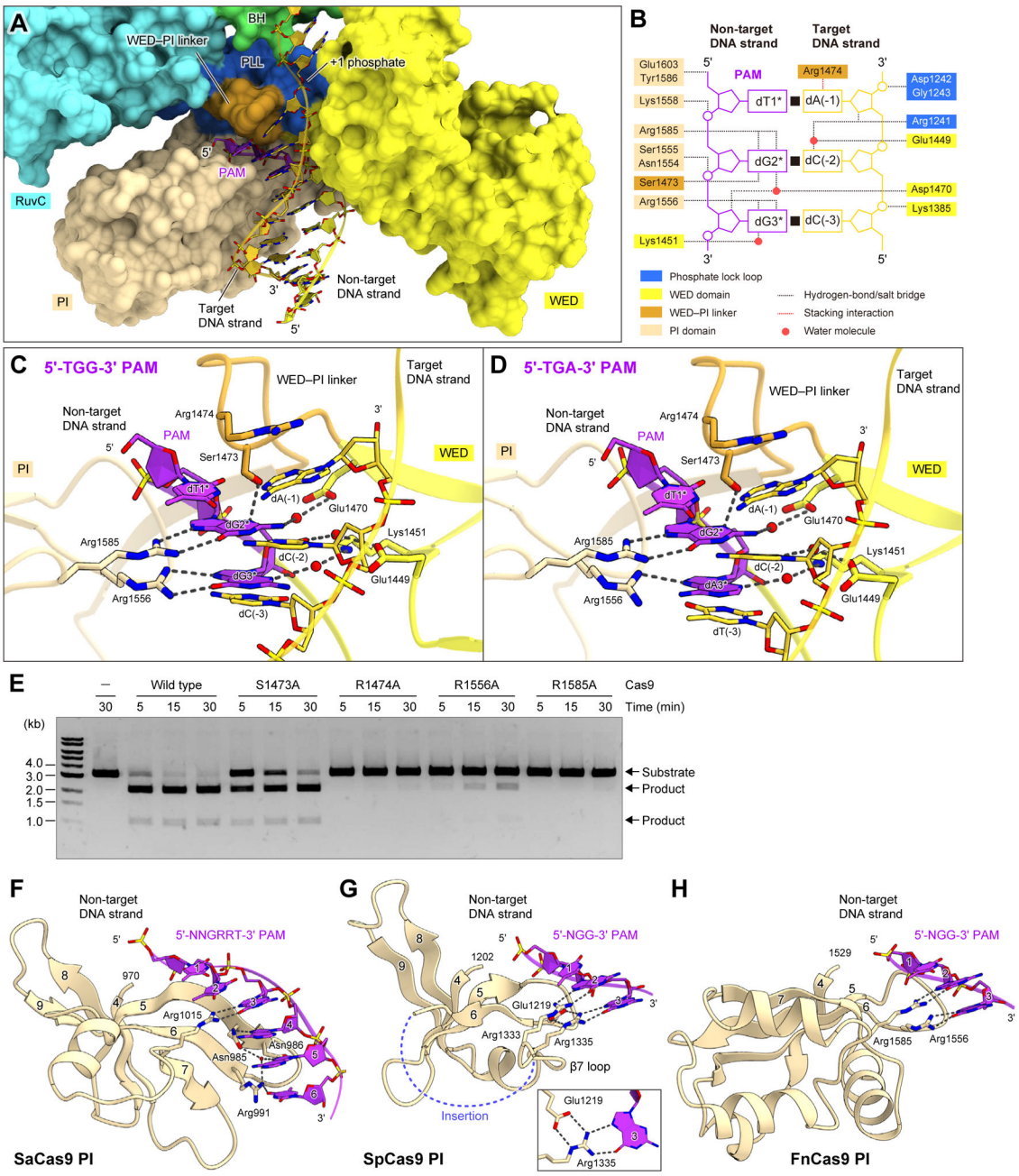

Figure 6. PAM recognition

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

(B) Schematics of the PAM duplex recognition. Water-mediated hydrogen bonds between the protein and the sugar-phosphate backbone are omitted for clarity.

(C and D) Recognition of the 5′-TGG-3′ (C) and 5′-TGA-3′ (D) PAMs. Water molecules are shown as red spheres.

(E) *In vitro* mutational analysis of the PAM-interacting residues. The linearized plasmid targets with the 5′-TGN-3′ PAMs were incubated with the wild type and mutants of FnCas9.

(F) Comparison of the PI domains of SpCas9 (F), SaCas9 (G) and FnCas9 (H)

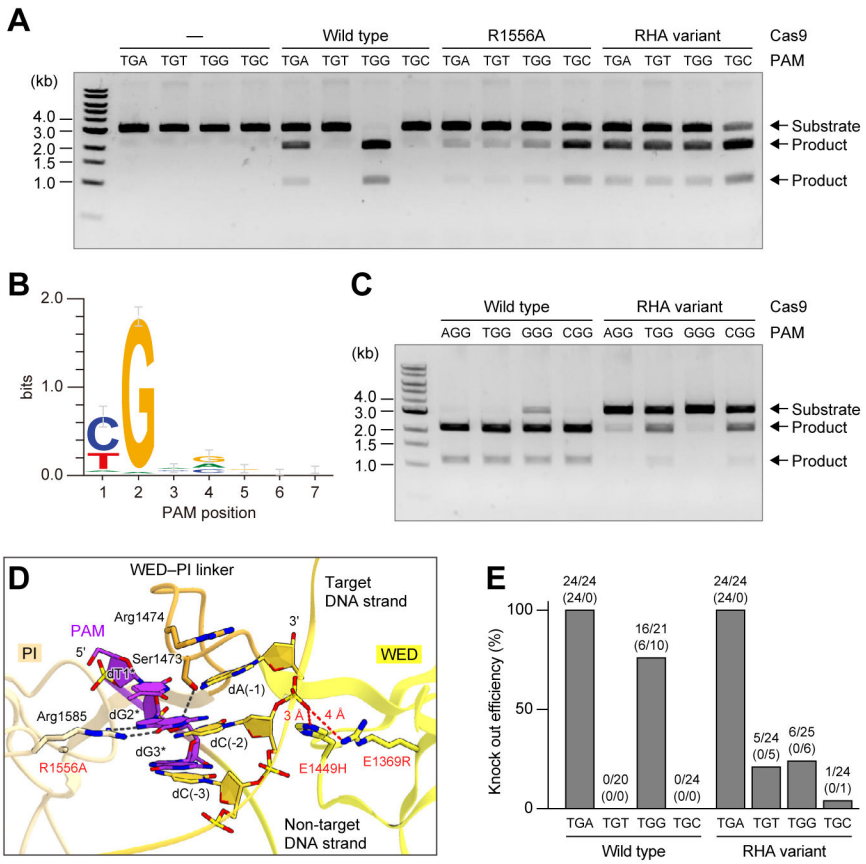

Figure 7. Structure-guided engineering and genome editing in mouse zygotes

(A) *In vitro* cleavage activity of wild-type and RHA FnCas9. The linearized plasmid targets with the 5′-TGN-3′ PAMs were incubated with the purified FnCas9–sgRNA complex.

(B) PAM discovery assay for RHA FnCas9.

(C) Preference of wild-type and RHA FnCas9 for the 1st PAM nucleotides. The linearized plasmid targets with the 5′-NGG-3′ PAMs were incubated with the FnCas9–sgRNA complex.

(D) PAM recognition mechanism of wild-type (left) and RHA (right) FnCas9.

(E) FnCas9-mediated genome editing in mouse zygotes. The pre-assembled wild-type and RHA FnCas9 RNP complexes were microinjected into mouse zygotes. The ratios between the numbers of embryos with FnCas9-mediated indels and the total numbers of injected embryos are shown above the bars. The numbers of embryos with mutations in both alleles (left) and a single allele (right) are shown in parentheses.