Doctoral Dissertation

博士論文

Structure and Engineering of Cas9 from *Francisella novicida*

（*Francisella novicida* 由来 Cas9 の構造解析および構造に基づく機能改変)

A Dissertation Submitted for the Degree of Doctor of Philosophy

December 2019

令和元年 12 月博士(理学)申請

Department of Biological Sciences, Graduate School of Science,

The University of Tokyo

東京大学大学院 理学系研究科

生物科学専攻

Hisato Hirano

平野 央人

1

Abstract

The CRISPR-associated protein Cas9 cleaves double-stranded DNA complementary to the guide RNA and has been harnessed for programmable genome editing. Cas9-tageting DNA sequences are restricted by the requirement of a protospacer adjacent motif (PAM) just downstream of the target sequence. The Cas9 orthologs from different microbes have highly divergent lengths and sequences and recognize diverse PAM sequences. This study presents the crystal structure of *Francisella novicida* Cas9 (FnCas9) in complex with a guide RNA and its PAM-containing target DNA. The structure revealed striking conserved and divergent structural features in the orthologous CRISPR-Cas systems. The author found that FnCas9 recognizes the 5′-NGG-3′ PAM and used the structural information to create the FnCas9 variant that recognizes more relaxed 5′-YG-3′ PAM, thereby expanding the targetable sequence of the CRISPR-Cas9 toolbox.

Content

Table of nucleic acid abbreviations

Abrreviation	Full name
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
U	Uracil
R	Adenine or guanine
Y	Thymine or cytosine

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 using conserved RuvC and HNH nuclease domains (Garneau et al., 2010; Deltcheva et al., 2011; Jinek et al., 2012; Gasiunas et al., 2012) (Figures 1A and 1B). 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 are divided into three clades called type II-A, type II-B, and type II-C (Cylinski et al., 2013) (Figure 2). The type II-B Cas9 from *Francisella novicida* 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, Ratner et al., 2019). However, the mechanisms by which FnCas9 executes its bifunctionality remain unknown due to the lack of structural information of type II-B Cas9 and low sequence homology between FnCas9 and structure-known Cas9, such as type II-A SpCas9 and SaCas9 (Appendix 1). In addition, the potential use of FnCas9 in genome editing applications has not been explored.

In this study, the author 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. The author 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, the author 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, the author 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 3A). 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 3B). Taken together, the author 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, the author 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 4A–4D and Table 1). To prevent the potential cleavage of the target DNA during crystallization, the author 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 4A–4D). 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 (Figure 17G and 17H).

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 4E and 4F and Figure 5). SpCas9 and SaCas9 adopt bilobed architectures comprising the REC and NUC lobes (Figures 4E and 4F and Figures 5B and 5C). 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 4C and Figure 5A). 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 6). Accordingly, the 3′ tracrRNA tail of the FnCas9 sgRNA is primarily recognized by the REC2 and REC3 domains (Figure 5A). 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 4C–4F and Figure 5). 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 4C–4F). 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 7A and 7B). 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 7A and Figure 8A). 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 7C–7E). Stem loop 1 consists of two base pairs (A54:G62 and U55:A61) and five unpaired nucleotides (U56–A60) (Figure 7 and Figure 8B). 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 8B). Stem loops 1 and 2 are connected by a 9-nt linker (SL1–SL2 linker), which contains a Watson-Crick base pair (A64:U68) and adopts a U-shaped structure (Figures 7A and 7C). 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 7A and Figure 8C).

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 7C–7E). 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 7C–7E). Consequently, stem loop 2 in the FnCas9 sgRNA is uniquely directed toward the REC1 and REC3 domains, unlike the SpCas9 and SaCas9 sgRNAs (Figures 7C–7D and Figure 5). These structural differences can explain the observed orthogonality between these CRISPR-Cas9 systems (Fonfara et al., 2014) (Figure 7F). A structural comparison also revealed the presence of a structurally conserved core region in their sgRNAs (Figure 7G). 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 7A and 7G). 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 9 and 10A–10C). 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 11A), 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 10A and 10B). In contrast, the REC3 domain of FnCas9 adopts a new fold comprising 20 α-helices and a β-hairpin (Figure 10C), with a structural zinc ion coordinated by Cys460, Cys657, Cys814 and Cys817 (Figure 10D). 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 9 and 10D). 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

13

phosphate lock loop interact with the +1 phosphate in the target DNA strand (Figure 11B), 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 10A–10C and Figures

12A–12C). 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 10C). The FnCas9 WED domain interacts

with the REC1 and REC2 domains to form a tunnel, which encloses the repeat:antirepeat duplex

(Figure 12A). 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 12D). The 3′

tracrRNA tail is also recognized by FnCas9 in a manner distinct from those of SpCas9 and SaCas9

(Figure 5). 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 12E). 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 12F).

The sgRNA core region is recognized by the bridge helix, the REC1 domain and the phosphate lock

loop (Figure 10E). 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 10E). Notably, SpCas9 and SaCas9 recognize the core regions of their cognate sgRNAs in similar manners (Figure 13), 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 13).

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 14A and 14B). dT1* (* represents the residue number of the non-target DNA strand) does not contact the protein (Figure 14C). 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 14C). In the 5′-TGG-3′ PAM complex, the O6 and N7 of dG3* form bidentate hydrogen bonds with Arg1556 (Figure 14C), whereas in the 5′-TGA-3′ PAM complex, the N7 of dA3* forms only a single hydrogen bond with Arg1556 (Figure 14D), 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 14C). The mutations of these residues reduced the *in vitro* DNA cleavage activity of FnCas9 (Figure 14E), 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 14F and 14G). 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 14H). 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 14G and 14H). 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 14G). In contrast, in FnCas9, the Arg1556 side chain does not form such a contact with the proximal residues (Figure 14H), 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, the author 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, the author 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 15A). The author hypothesized that the reduced activity of the R1556A variant could be recovered by additional non-base-specific interactions between the protein and the PAM duplex. The author thus introduced several mutations into the R1556A variant, which would potentially form additional interactions with the backbone phosphates of the PAM duplex. The author 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 15A). However, unexpectedly, the PAM discovery analyses indicated that RHA FnCas9 recognizes 5′-YG-3′, but not 5′-NG-3′, as the PAM (Figure 15B). Consistently, *in vitro* cleavage assays revealed that RHA FnCas9 has a stronger preference for the 1st Y, than wild-type FnCas9 (Figure 15C). 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, the author 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 15D and 15E, Table 1). 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 15D). 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 15D). 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 15D), confirming that these non-base-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, the author examined whether FnCas9 can be harnessed for genome editing in mammalian cells. The author microinjected the pre-assembled FnCas9 RNP complex targeted to the mouse Tet1EX4 locus with the 5′-TGN-3′ PAMs 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 15E), 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 15E). 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, the author presents 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 the PI domain of FnCas9 shares a similar core fold with those of SpCas9 and SaCas9, it 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 staggered cleavage functionality of FnCas9 (Chen et al., 2016). Unlike SpCas9 and SaCas9, the RuvC domain of FnCas9 does not interact with PI domain and is located farther away from the PAM. This would cause the FnCas9 to cleaves the target DNA at the downstream position relative to that of

SpCas9 and SaCas9, resulting in the enzymatic property of FnCas9 to form 5' overhang staggered end. 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. Furthermore, the HNH domain of FnCas9 is located further away from the cleavage site determined by sequencing, and the electron densities for the HNH domains are relatively weak (Figure 17G)(Chen et al., 2017). These results reflect the flexible nature of HNH domain, consistent with the previous studies that shows conformational change of HNH is critical for DNA cleavage (Sternberg et al., 2015; Shibata and NIshimasu et al., 2016). Altogether, the RNA-guided DNA cleavage mechanism is conserved among the CRISPR-Cas9 systems.

The structure of FnCas9 revealed the conserved and divergent features among the Type II-B CRISPR-Cas systems. The residues interacting with the guide:target heteroduplex in FnCas9 shows high similarity among type II-B Cas9 orthologs, whereas the regions responsible for the recognition of stem loop 2 (residues 504–529 and β-hairpin in REC3) (Figure 12E and 12F) are FnCas9 specific (Figure 9 and Appendix 2). These results indicate that Type II-B Cas9 recognizes its corresponding sgRNA in an orthogonal manner, whereas the recognition mechanism of guide:target heteroduplex is conserved among type II-B Cas9 orthologs. Recent study shows that FnCas9 does not tolerate only single mismatch at the PAM distal end, thereby achieving high DNA targeting specificity compared to SpCas9 (Chen et al., 2016). Thus, the other Type II-B Cas9 orthologs, that has similar guide:target recognition mechanism to FnCas9, may also have high specific cleavage property for target DNA. Moreover, the sequences correspond to PI domain in FnCas9 has no similarity among Type II-B Cas9, suggesting that the PAM sequence of Type II-B Cas9 are highly divergent. Taken together, type II-B Cas9 orthologs have the possibility to expand the target space of high-fidelity genome editing, although further functional analysis is needed.

The present structure of the FnCas9-sgRNA-target DNA complex provides mechanistic insight into scaRNA:tracrRNA dependent gene regulation. The 3' end of the scaRNA base pairs with the 5' end of the tracrRNA to form scaRNA:tracrRNA duplex, as in the case of crRNA:tracrRNA duplex (Ratner et al., 2019) (Figure 16A and 16B). The lower stem region is conserved between crRNA:tracrRNA and scaRNA:tracrRNA duplex, indicating that these two duplex are recognized by REC and WED domain in a same manner. Moreover, the just upstream region of the lower stem region of scaRNA functions as guide. These results indicate that the underlying mechanism of crRNA:tracrRNA and scaRNA:tracrRNA dependent DNA binding is identical, consistent with the experimental result that scaRNA:tracrRNA mediated gene regulation need the PAM sequence on the target DNA (Ratner et al., 2019). The scaRNA has 11 to 15 bases guide which is sufficient for DNA binding. Previous study shows that the incomplete RNA-DNA complementarity mediates SpCas9 to bind to its target but does not cause cleavage, and that base pairing at the PAM distal end causes a conformational change in the L1 and L2 linker region to alter the position of the HNH domain, leading to cleavage of the target DNA (Sternberg et al., 2015; Jiang et al., 2016). Thus, the author concluded that FnCas9-scaRNA-tracrRNA complex binds to the target DNA in a pre-catalytic state, as observed in SpCas9. However, the mechanism that links scaRNA:tracrRNA mediated DNA binding to gene regulation remain unknown, thus further functional study is required to unveil it.

The author showed that FnCas9 recognizes the 5′-NGG-3′ PAM, and rationally designed the RHA variant that recognizes the 5′-YG-3′ PAM, thereby expanding the target space in FnCas9-mediated genome editing. Furthermore, the author 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 shows lower activity when expressed in human cells.

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, non-base-specific interactions, thereby achieving altered PAM specificities. Recent study shows that SpCas9 can be engineered to recognize 5′-NG-3′ PAM by introducing seven mutations (R1335V/L1111R/D1135V/G1218R/E1219F/A1322R/T1337R: referred to as VRVRFRR SpCas9) with the same concept as RHA mutations of FnCas9 (Nishimasu et al., 2018). The crystal structure of VRVRFRR SpCas9 shows that introduced arginine residues and hydrophobic residues forms electrostatic interaction with phosphate backbone and hydrophobic interaction with ribose moiety of PAM duplex, respectively, whereas R1335V mutation eliminated the interaction between R1335 and third G nucleotide of original 5′-NGG-3′ PAM, thereby achieving the 5′-NG-3′ PAM recognition. Moreover, the E782K/N968K/R1015H variant of SaCas9 (referred to as the KKH SaCas9), which is engineered by bacterial-selection based method, exhibit altered 5′-NNNRRT-3′ PAM specificity. 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. Taken together, the strategy to compensate for the loss of base-specific interactions with the PAM nucleotides, by including additional non-base-specific

interactions to alter Cas9 PAM specificities, may be generally applicable to other Cas9 orthologs.

In addition to SpCas9, SaCas9 and FnCas9, recent studies reported the structure of type II-A Cas9 from *Streptococcus thermophilus* and type II-C Cas9 from *Corynebacterium diphtheriae*, *Campylobacter jejuni* and *Neisseria meningitidis* in complex with its sgRNA and target DNA (Fuchsbauer et al., 2019, Hirano et al., 2018, Yamada et al., 2017 and Sun et al., 2019)(Figure 18). The overall domain organization of these Cas9 is similar to that of the SpCas9, SaCas9 and FnCas9. These Cas9 comprises REC, RuvC, HNH, WED and PI domains, with the guide RNA-target DNA heteroduplex is bound within the central channel formed by REC and RuvC domains, and the PAM duplex is bound to the groove formed by WED and PI domains. The HNH domain is not visible in cryo-EM map of StCas9, and the deletion of HNH domain facilitated the crystallization of CdCas9 and CjCas9. Furthermore, two distinct conformations of HNH domain are observed in NmCas9 structure, and one of that represents the catalytic state, with the catalytic amino acid residues are located close to the cleavage site of the target DNA strand. These results reinforce the notion that the guide RNA-target DNA heteroduplex formation drives conformational change of HNH nuclease domain, thereby enabling the catalytic residues of HNH domain to access to the scissile phosphate of the cleavage site. Moreover, the bridge helix, phosphate lock loop and catalytic residues of RuvC domain are structurally conserved between these Cas9 orthologs. Taken together, these results indicate that the CRISPR-Cas9 system adopts a similar mechanism to generate RNA-guided DNA cleavage.

A structural comparison of CjCas9, CdCas9, NmCas9 and StCas9 illuminated the structural divergence of the sgRNA recognition mechanism among Cas9 orthologs including type II-C Cas9. The 3' sgRNA tail of these Cas9 comprises two stem loops (SL1 and SL2). In StCas9 and CdCas9, the upper region of SL1 is recognized by REC1 and WED domains, respectively, whereas the basal region

of SL2 is analogously recognized by the groove formed by RuvC and PI domains (Figure 18A and 18B). In contrast, the SL2 of NmCas9 and CjCas9 laterally interact with RuvC and PI domains (Figure 18C and 18D). Moreover, the extended 3' single strand region of CjCas9 sgRNA formed triple helix structure with SL1. These sgRNA structure and its interactions with Cas9 are specific for each CRISPR-Cas9 system, explaining the orthogonality between sgRNA and Cas9.

The author speculates that the highly divergent structural features of CRISPR-Cas9 is derived from the arm race between phage and bacteria. CRISPR-Cas system functions as bacterial defense system against phages by cleaving their nucleic acid. In contrast, phages produce an Anti-CRISPR protein (Acr) that inhibits CRISPR nucleases, thereby escaping from repression by CRISPR-Cas system (Bondy-Denomy et al., 2013). To date, 44 families of Acrs have been discovered, 15 of which have been biochemically characterized (Zhang et al., 2019). Their functions can be roughly divided into three types: crRNA loading interference, DNA binding blockage and DNA cleavage prevention. These inhibition mechanisms could impose strong selective pressure on Cas9 to evolve to escape it by introducing mutations in the Acr-interacting interfaces of Cas9 or guide RNA. For example, AcrIIC3 binds to and inhibit HNH domain of NmCas9, but does not inhibit that of SpCas9 or CjCas9, due to the diversity of the surface structure of HNH domain (Kim et al., 2019). The bacteria bearing the evolved CRISPR-Cas9 shows resistant to phages, reversely imposing a selective pressure on Acr proteins. These counterdefence strategies may drive the coevolution of CRISPR-Cas9 and Acr, resulting the highly divergent structural properties of orthologous CRISPR-Cas9 system.

Materials and methods

Sample preparation

The gene encoding full-length FnCas9 (residues 1–1,629) was cloned between the NdeI and XhoI 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) for four hours, and was purified by chromatography on Ni-NTA Superflow resin (QIAGEN) eluted with a buffer consisting of 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 300 mM imidazole and 3 mM 2-mercaptoethanol. The eluted protein was incubated overnight at 4°C with TEV protease to remove the His6-SUMO-tag during dialysis in buffer consisting of 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole and 3 mM 2-mercaptoethanol, and was purified by chromatography on Ni-NTA. The protein was loaded into Mono S (GE Healthcare) in buffer consisting of 10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM DTT, and eluted with linear NaCl gradient. The eluted protein was further purified by HiLoad Superdex 200 16/600 (GE Healthcare) columns in buffer consisting of 10 mM Tris-HCl, pH 8.0, 150 mM NaCl and 1 mM DTT. 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 17A). 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 (Figure 17B and 17C). For *in vitro* cleavage assays, the His6-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 (A260 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 (Figure 17D). 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 (A260 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 (Figure 17E). 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 (Figure 17F–17H).

In vitro cleavage assay

In vitro plasmid DNA cleavage experiments were performed, essentially as described previously (Nishimasu et al., 2015). The BamHI-linearized pUC119 plasmid (100 ng, 510 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

In vitro PAM screen was performed in collaboration with Dr. Gootenberg, J.S. and Dr. Abudayyeh, O.O. (Broad Institute of MIT and Harvard). 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

Microinjection and typing the blastocyst embryos was performed in collaboration with Dr. Horii, T. Ms. Kimura and Dr. Hatada, I. (Gunma University). 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.

Figures
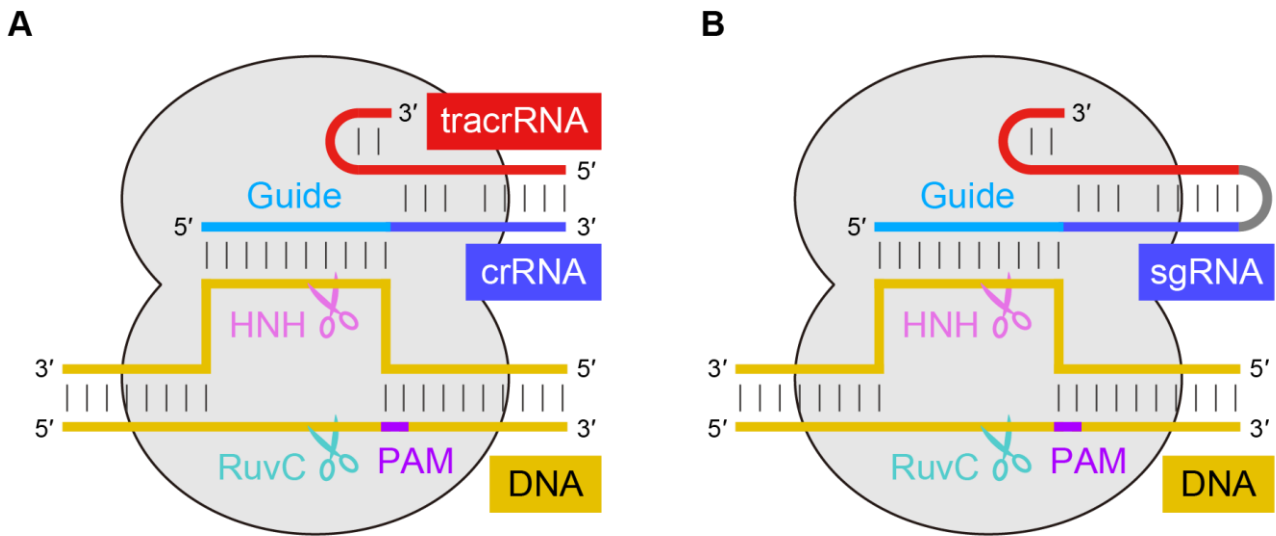
Figure 1. RNA-guided DNA cleavage mechanism of CRISPR-Cas9

Schematic representation of crRNA:tracrRNA (A) and sgRNA (B) mediated DNA cleavage.
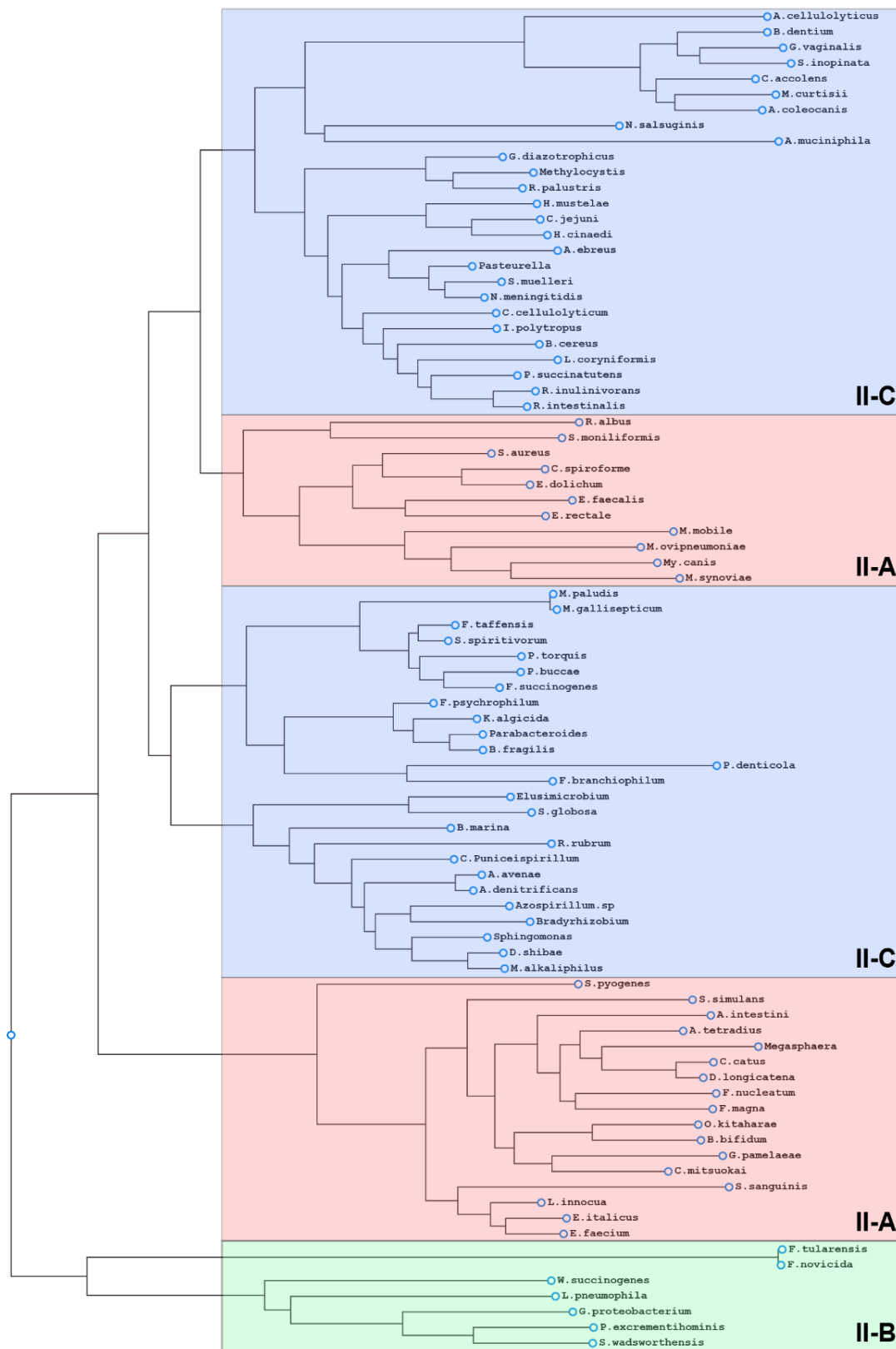
Figure 2. Phylogenetic tree of representative Cas9 orthologs

The Cas9 sequences were aligned using COBALT (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) and the phylogenetic tree were constructed by the approximately maximum likelihood based FastTree program.
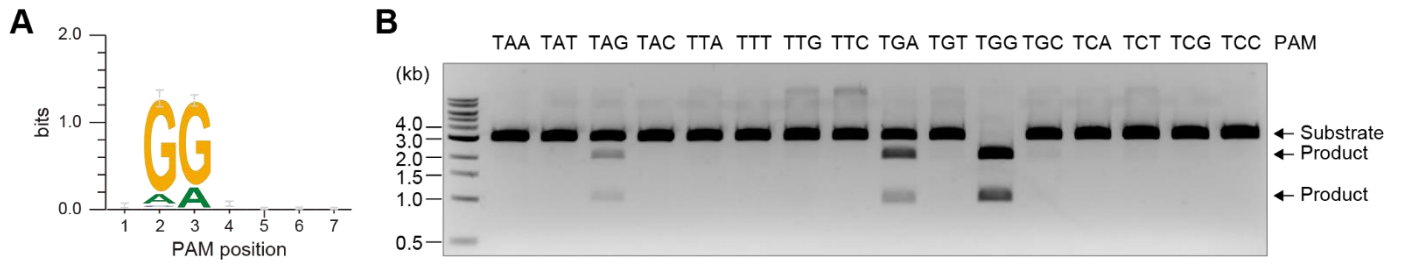
Figure 3. PAM specificity of FnCas9

(A) Motif obtained from the PAM discovery assay for FnCas9 (This data was provided by Feng Zhang).

(B) *In vitro* DNA cleavage by FnCas9. The linearized plasmid targets with the 5′-TNN-3′ PAMs were incubated with the purified FnCas9–sgRNA complex.
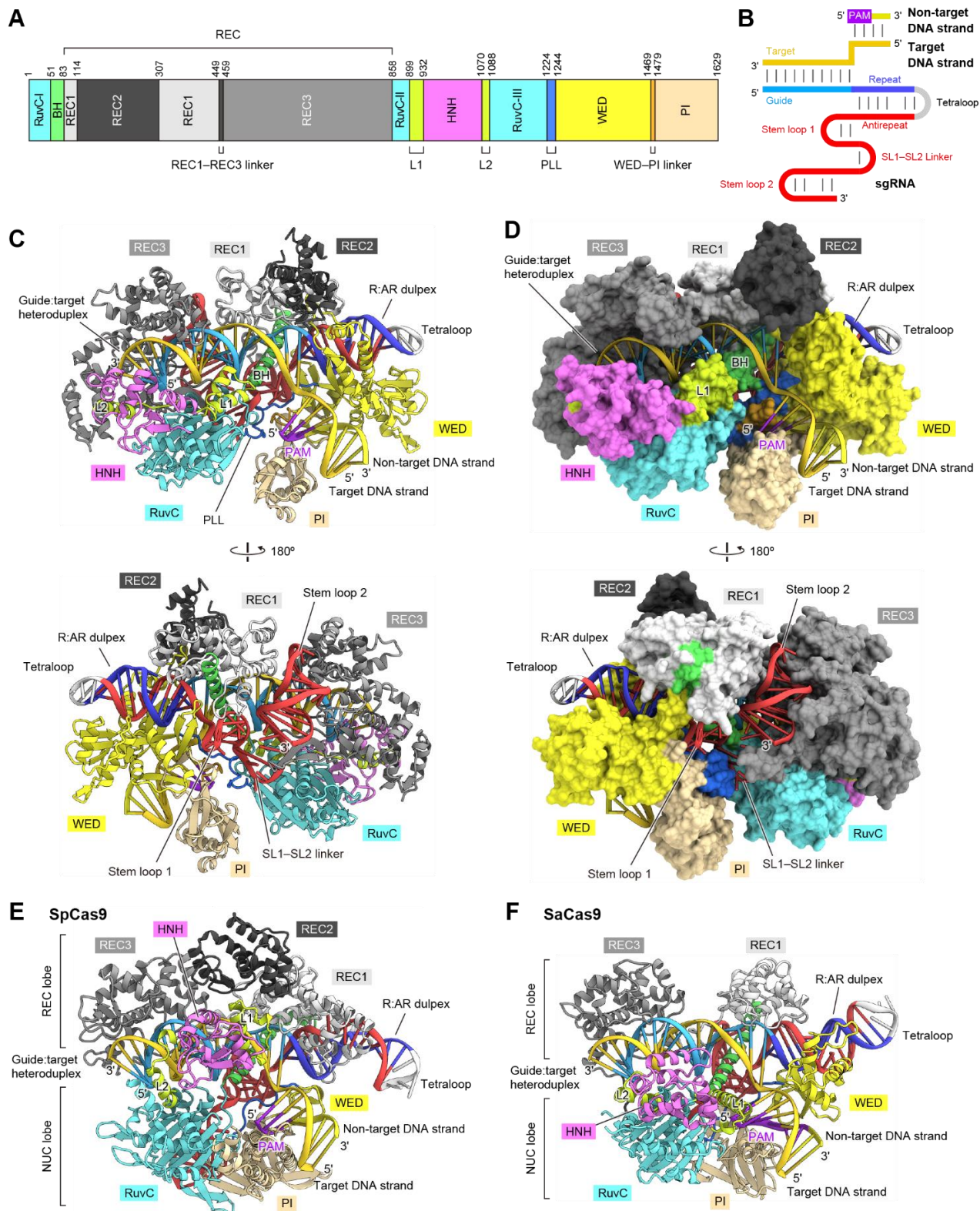
Figure 4. Overall structure of the FnCas9–sgRNA–DNA complex.

(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.

R:AR duplex, repeat:antirepeat duplex.

(E and F) Crystal structures of SpCas9 (PDB: 4UN3) (E) and SaCas9 (PDB: 5CZZ) (F).
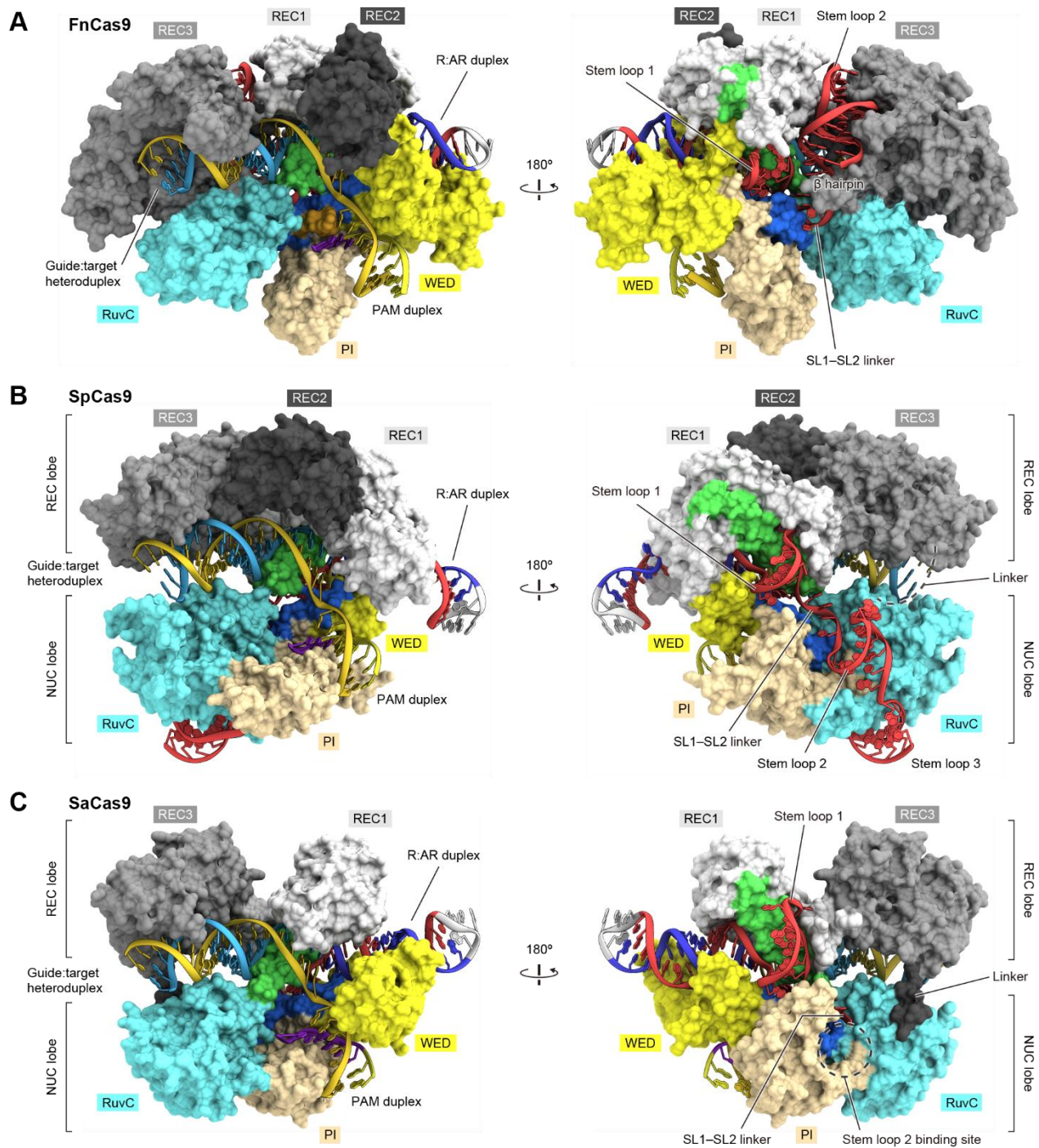
Figure 5. Comparison of the overall structures of the Cas9 orthologs

(A–C) Overall structures of FnCas9 (A), SpCas9 (PDB: 4UN3) (B) and SaCas9 (PDB: 5CZZ) (C) in complexes with the sgRNA and the target DNA. The HNH domains are omitted for clarity. The sgRNA in the SpCas9 quaternary complex (PDB: 4UN3) contains stem loops 1–2 and lacks stem loop 3, whereas the sgRNA in the SpCas9 ternary complex (PDB: 4OO8) contains stem loops 1–3. Thus, in (B), the sgRNA (PDB: 4OO8) is docked into the SpCas9–DNA complex (PDB: 4UN3).
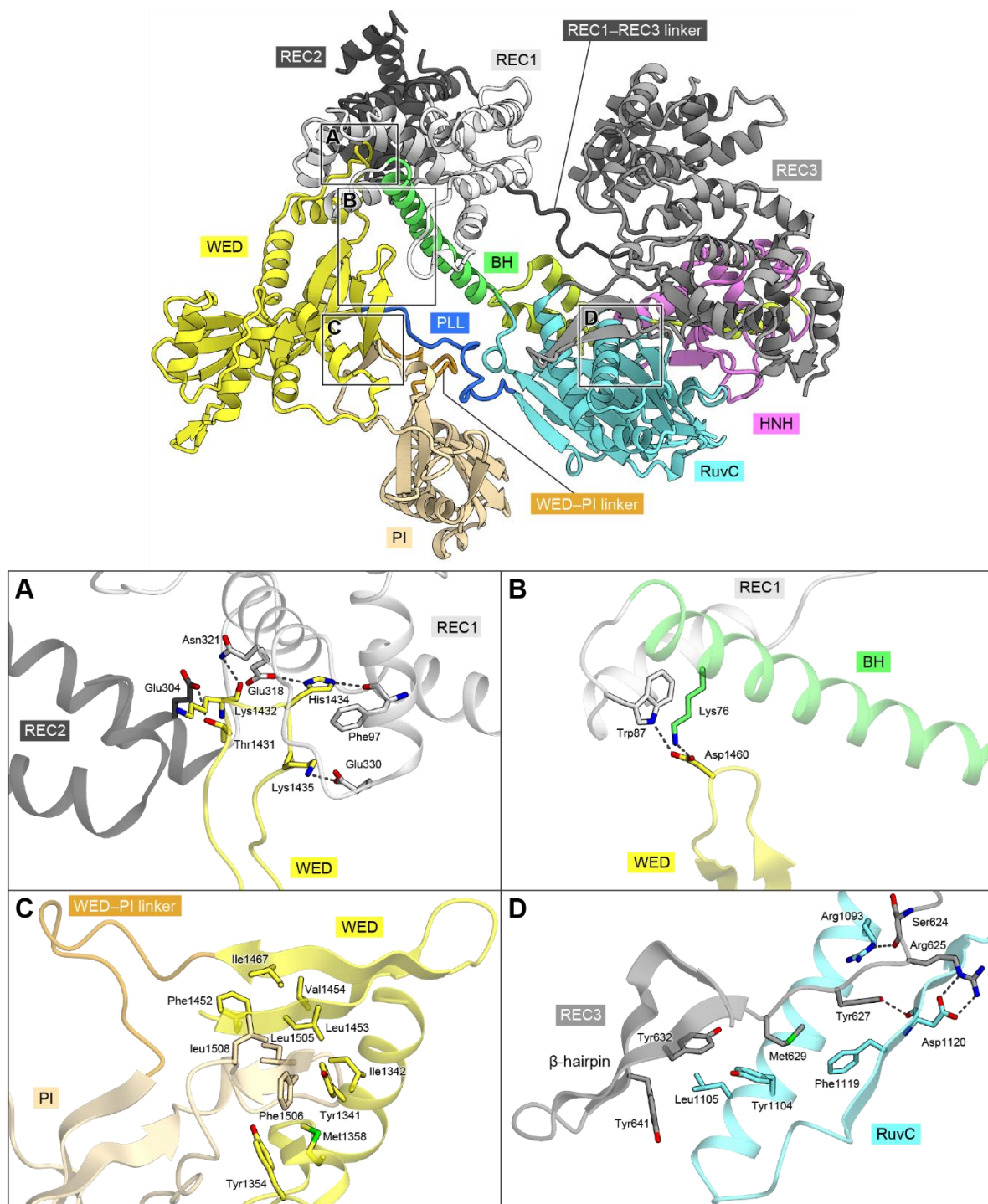
Figure 6. Inter-domain interactions in FnCas9

(A) Interaction between the WED and REC1/REC2 domains.

(B) Interaction between the WED and REC1 domains.

(C) Interaction between the WED and PI domains.

(D) Interaction between the RuvC and REC3 domains.

Hydrogen-bonding interactions are shown as dashed lines.
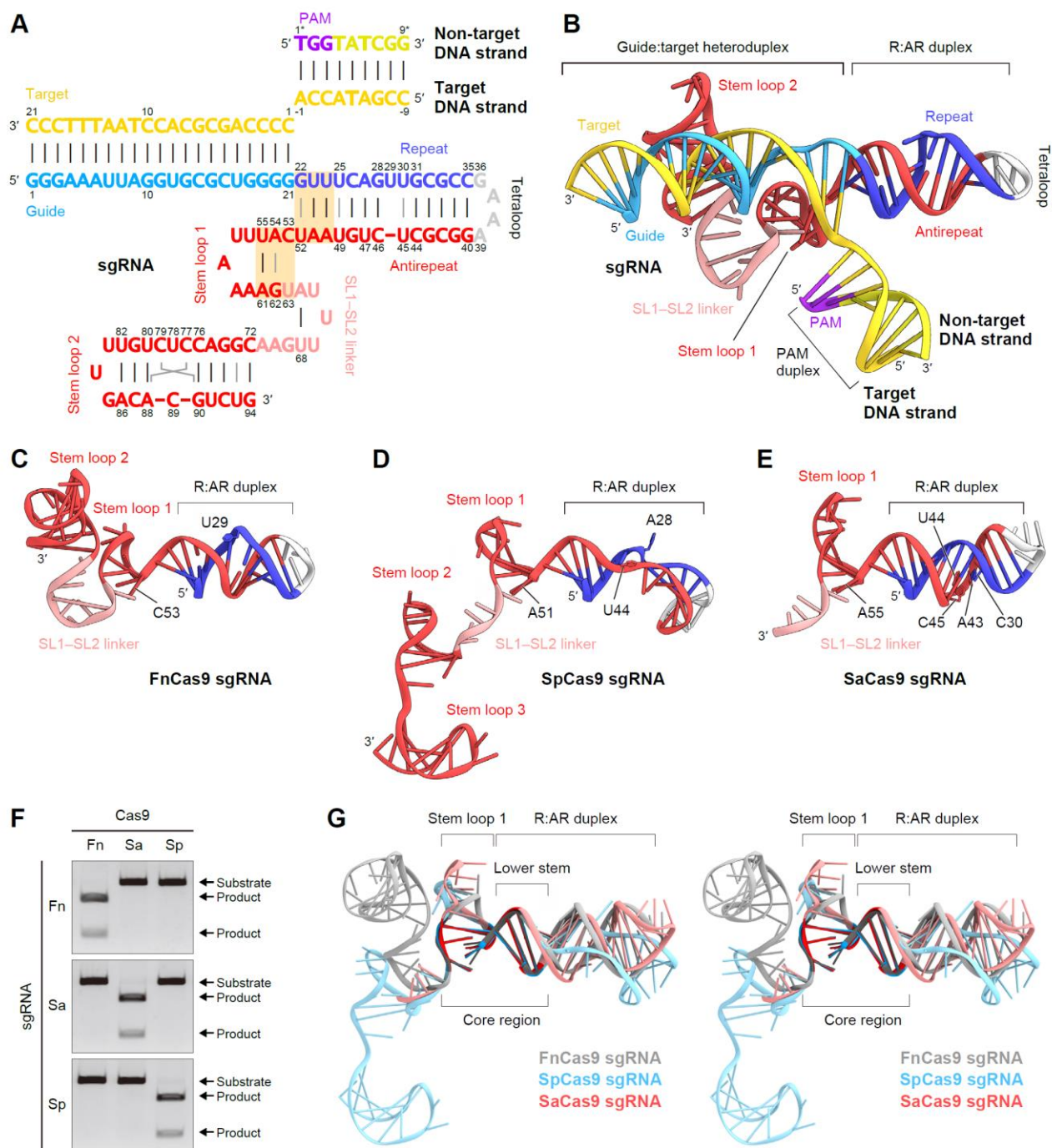
Figure 7. Structure of sgRNA–DNA

(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) Orthogonality between Cas9 and sgRNA. The linearized plasmid targets with the appropriate PAM were incubated with different combinations of the Cas9–sgRNA complexes.

(G) Comparison of the sgRNA scaffolds of FnCas9 (gray), SpCas9 (blue), and SaCas9 (red).
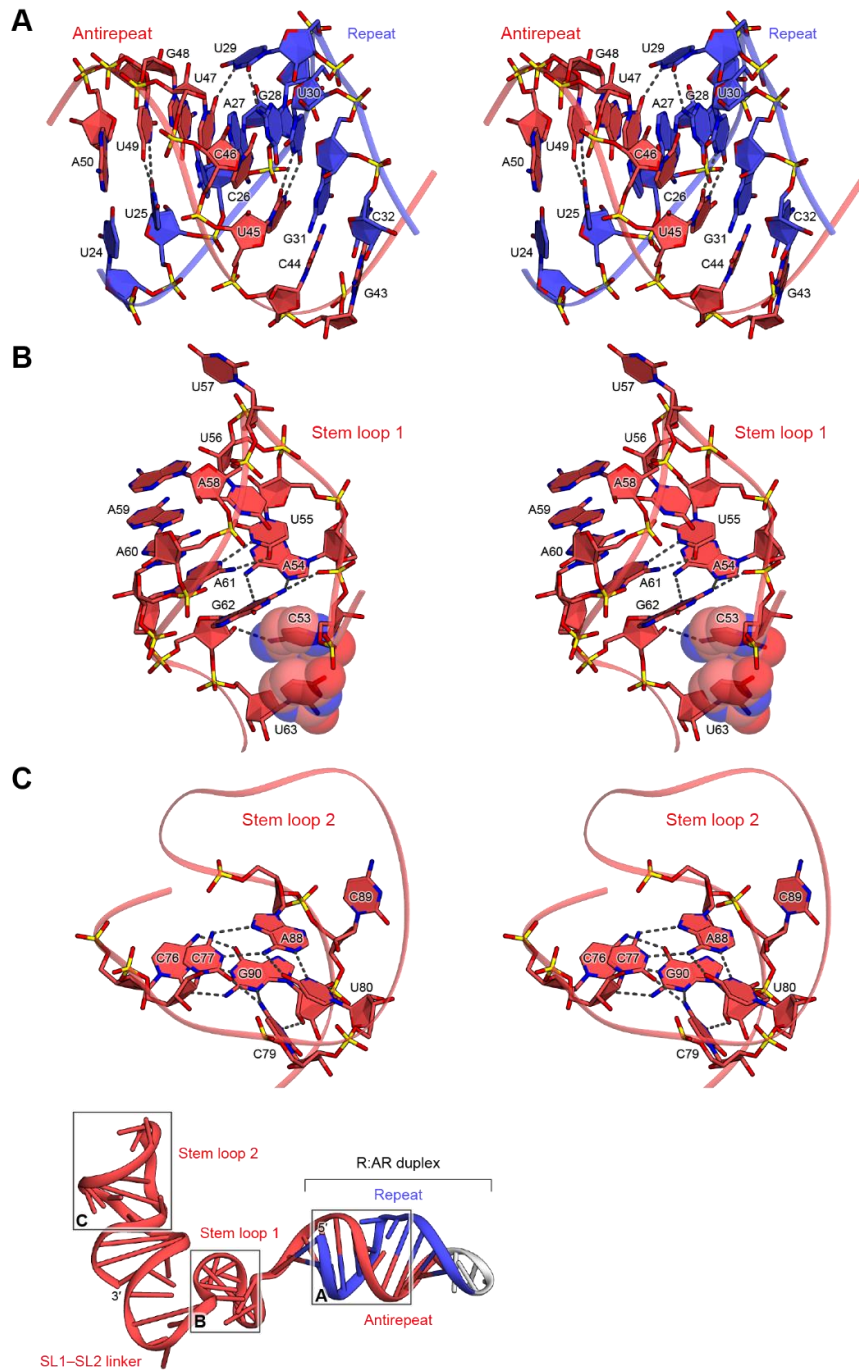
Figure 8. Structure of the FnCas9 sgRNA scaffold

(A–C) Structures of the repeat:antirepeat duplex (A), stem loop 1 (B) and stem loop 2 (C) (stereo view). Hydrogen-bonding interactions are shown as dashed lines.

In (B), C53 and U63 are depicted as semi-transparent space-filling models, highlighting the base-stacking interaction. U57 is disordered.
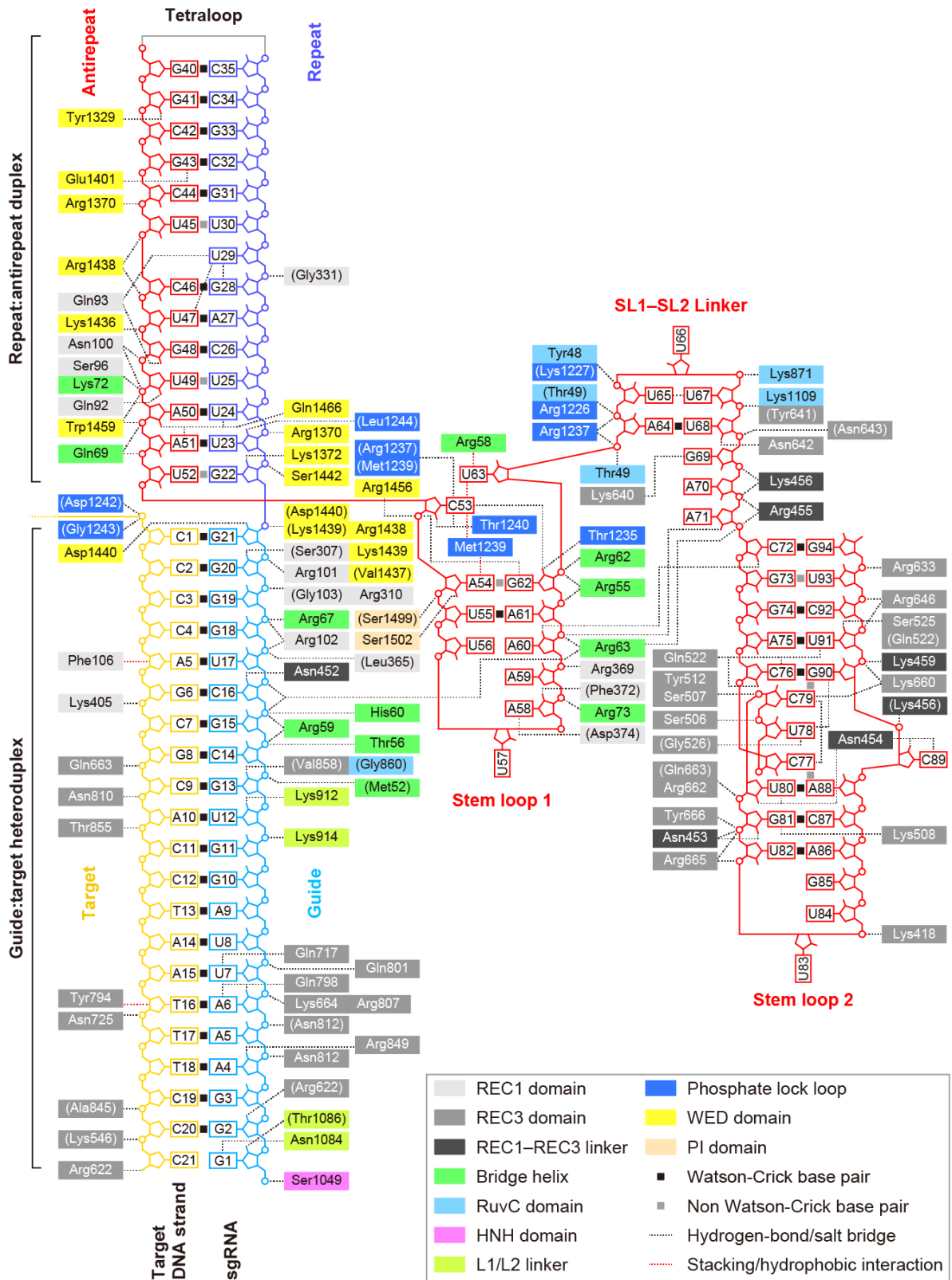
Figure 9. Schematic of the nucleic acid recognition by FnCas9

Residues that interact with nucleic acids via their main chain are shown in parentheses. Water-mediated hydrogen bonds are not shown for clarity.
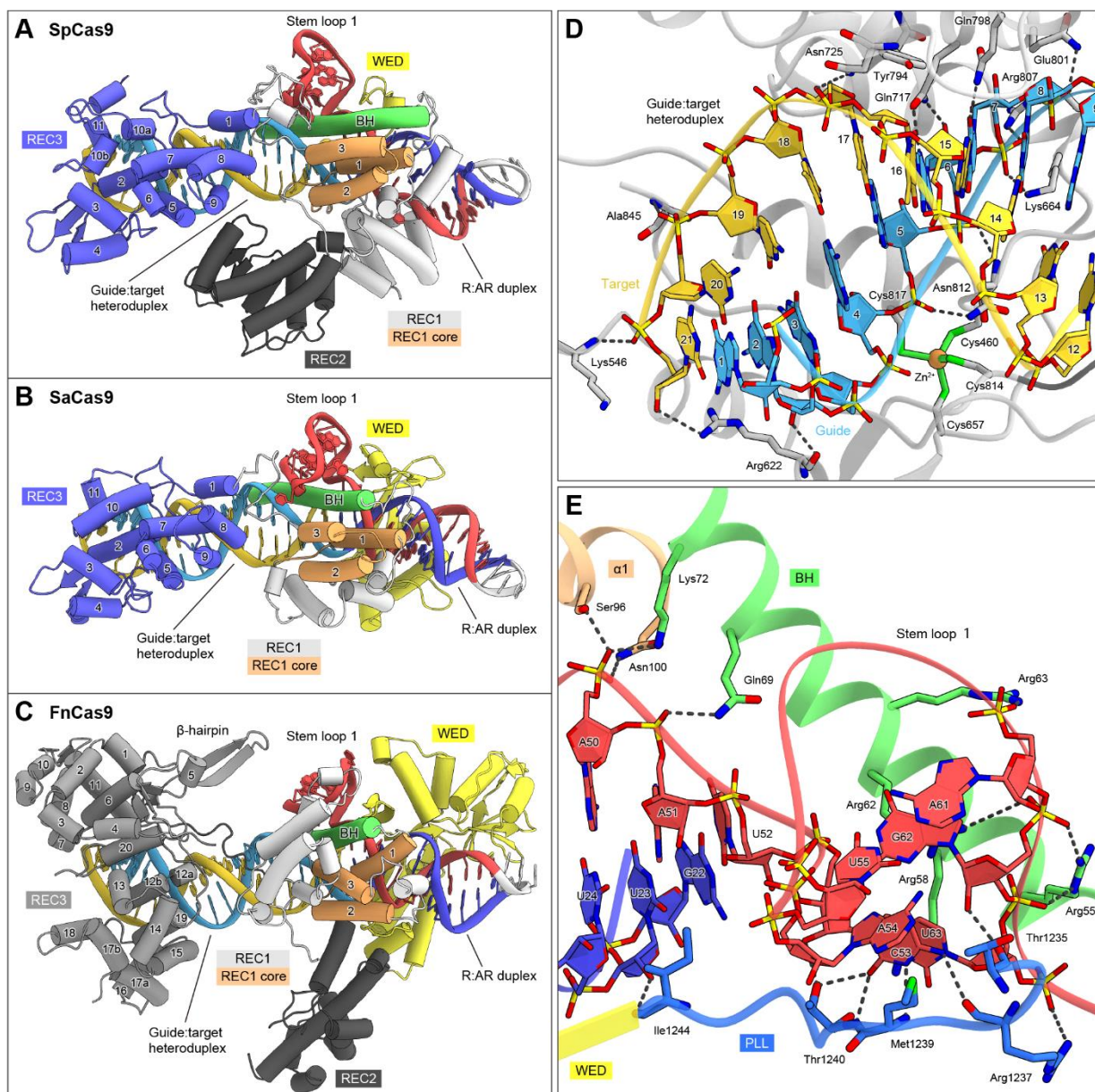
Figure 10. Recognition of the sgRNA and the target DNA by the Cas9 Orthologs

(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.
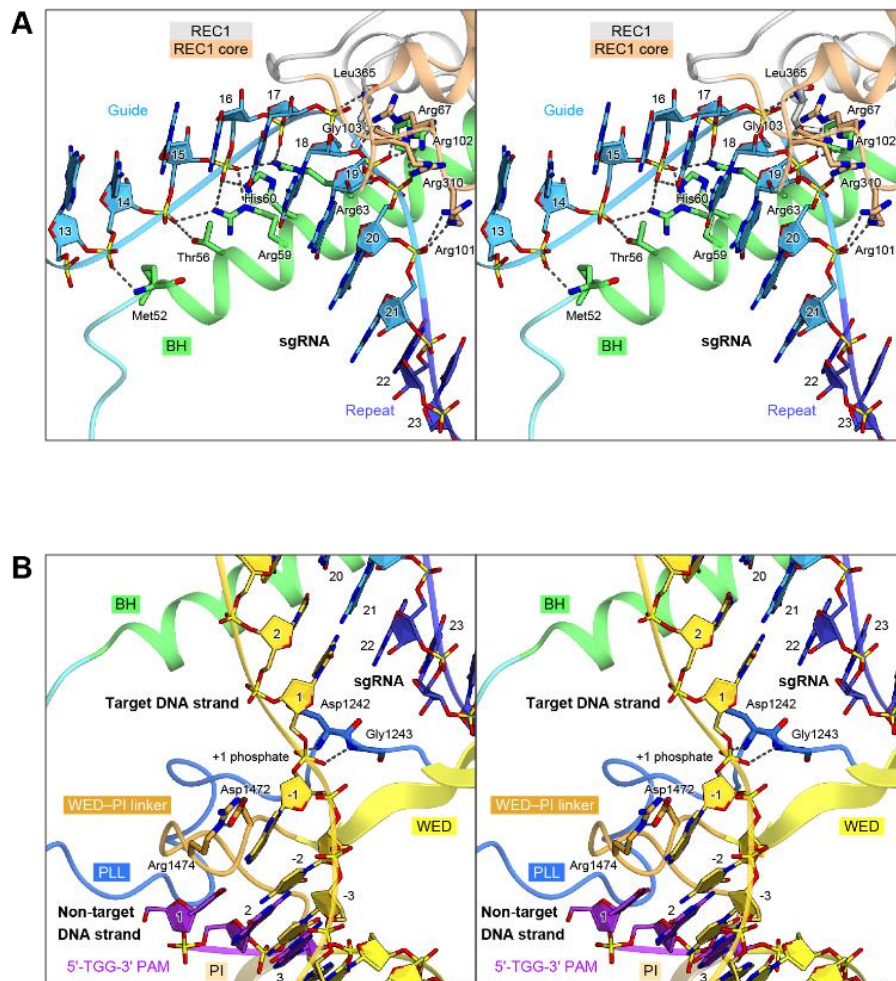
Figure 11. DNA-targeting mechanism of FnCas9

(A) Recognition of the sgRNA seed region by the bridge helix (BH) and the REC1 domain (stereo view). The target DNA strand is omitted for clarity.

(B) Recognition of the +1 phosphate group by the phosphate lock loop (PLL) (stereo view). Hydrogen-bonding interactions are shown as dashed lines.
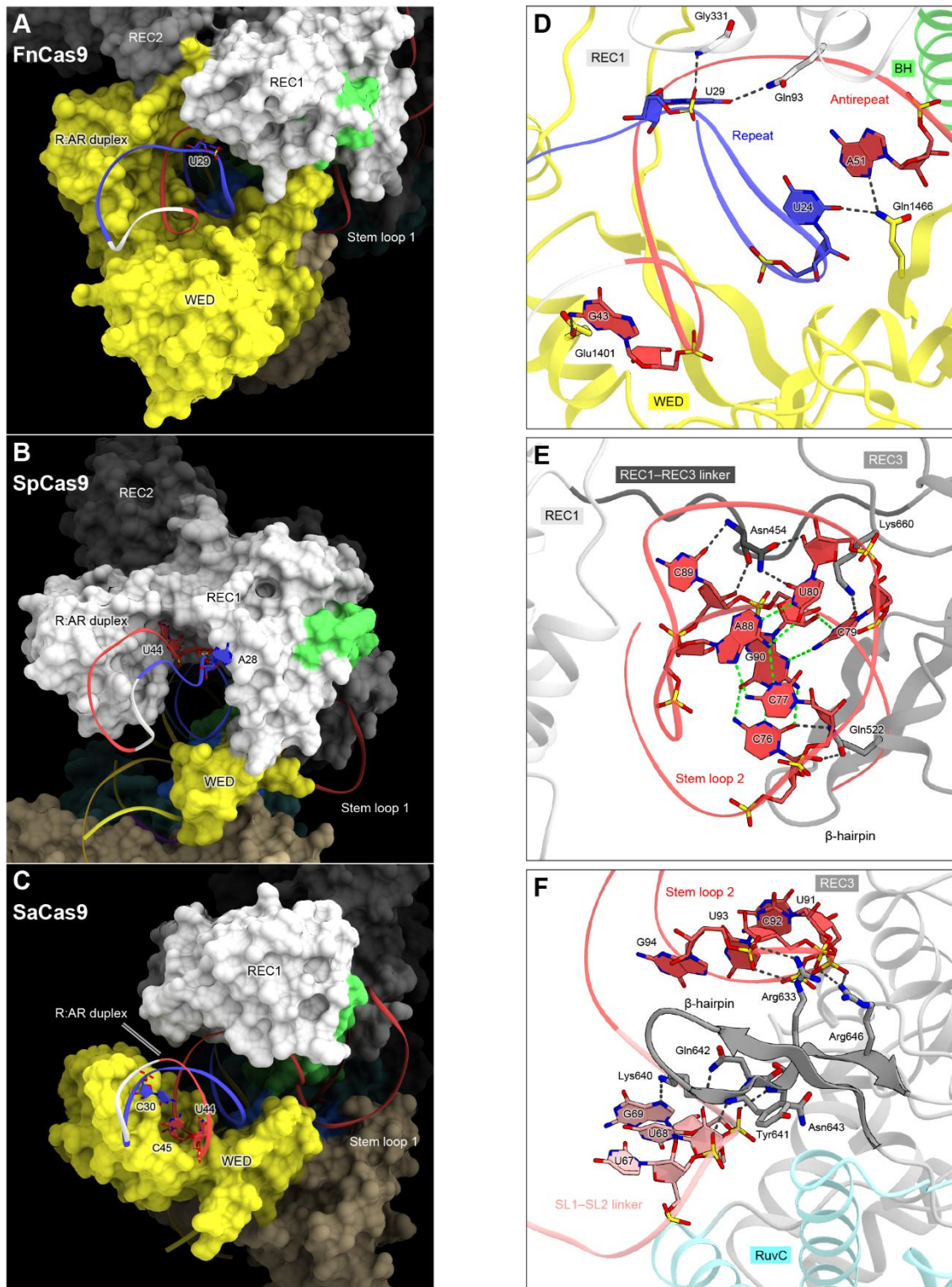
Figure 12. Recognition of sgRNA scaffolds by the Cas9 orthologs

(A–C) Recognition of the sgRNA scaffolds by FnCas9 (A), SpCas9 (B) and SaCas9 (C).

(D–F) Recognition of the repeat:antirepeat duplex (D), stem loop 2 (E), and the SL1–SL2 linker (F) by FnCas9. In (F), the REC3 β-hairpin is inserted between stem loop 2 and the SL1–SL2 linker.
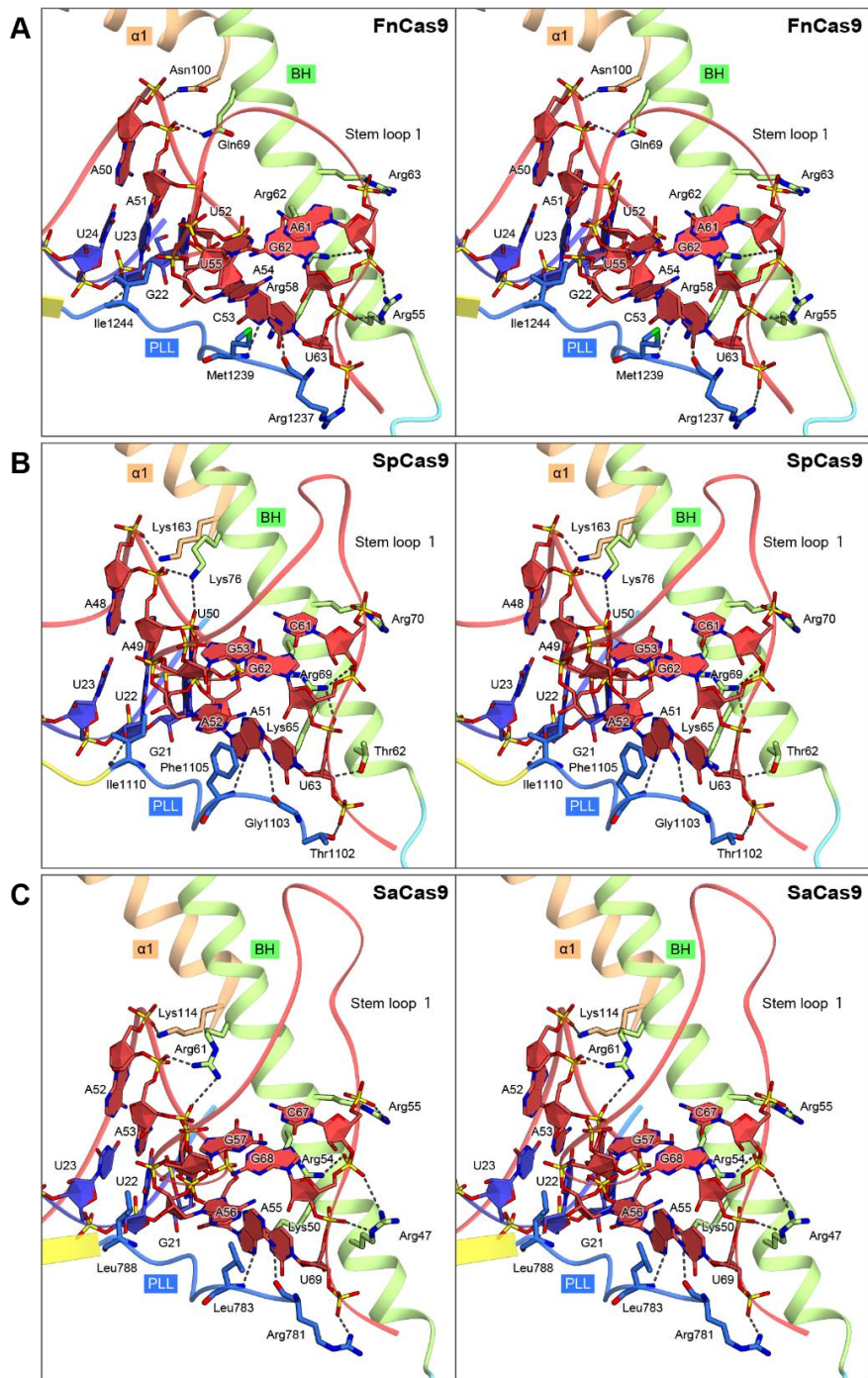
Figure 13. Recognition of the sgRNA core regions by the Cas9 orthologs

(A–C) Recognition of the sgRNA core regions by FnCas9 (A), SpCas9 (B) and SaCas9 (C) (stereo view). Hydrogen-bonding and electrostatic interactions are shown as dashed lines.
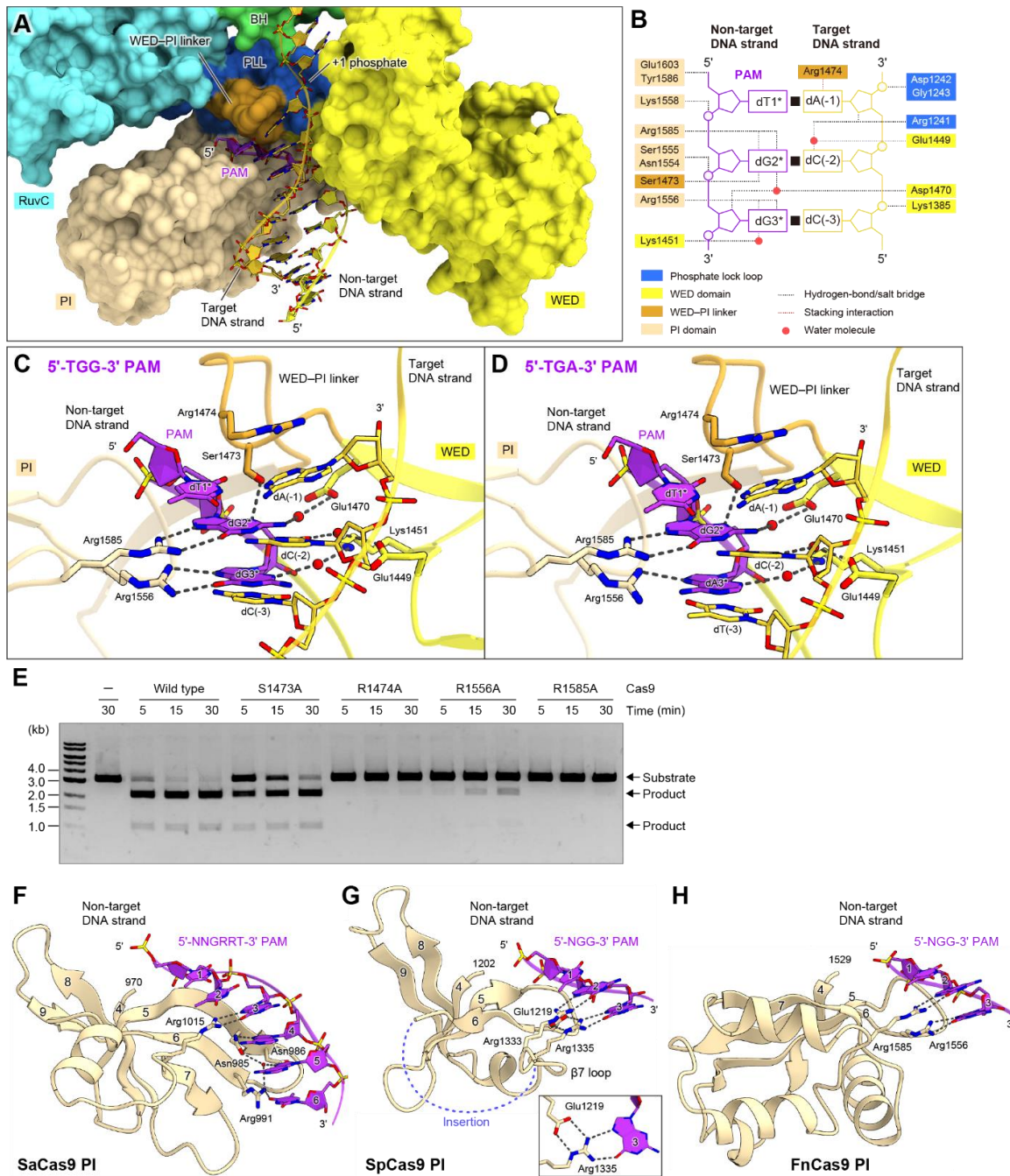
Figure 14. PAM recognition

(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′-TGG-3′ PAMs were incubated with the wild-type and mutants of FnCas9.

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

43
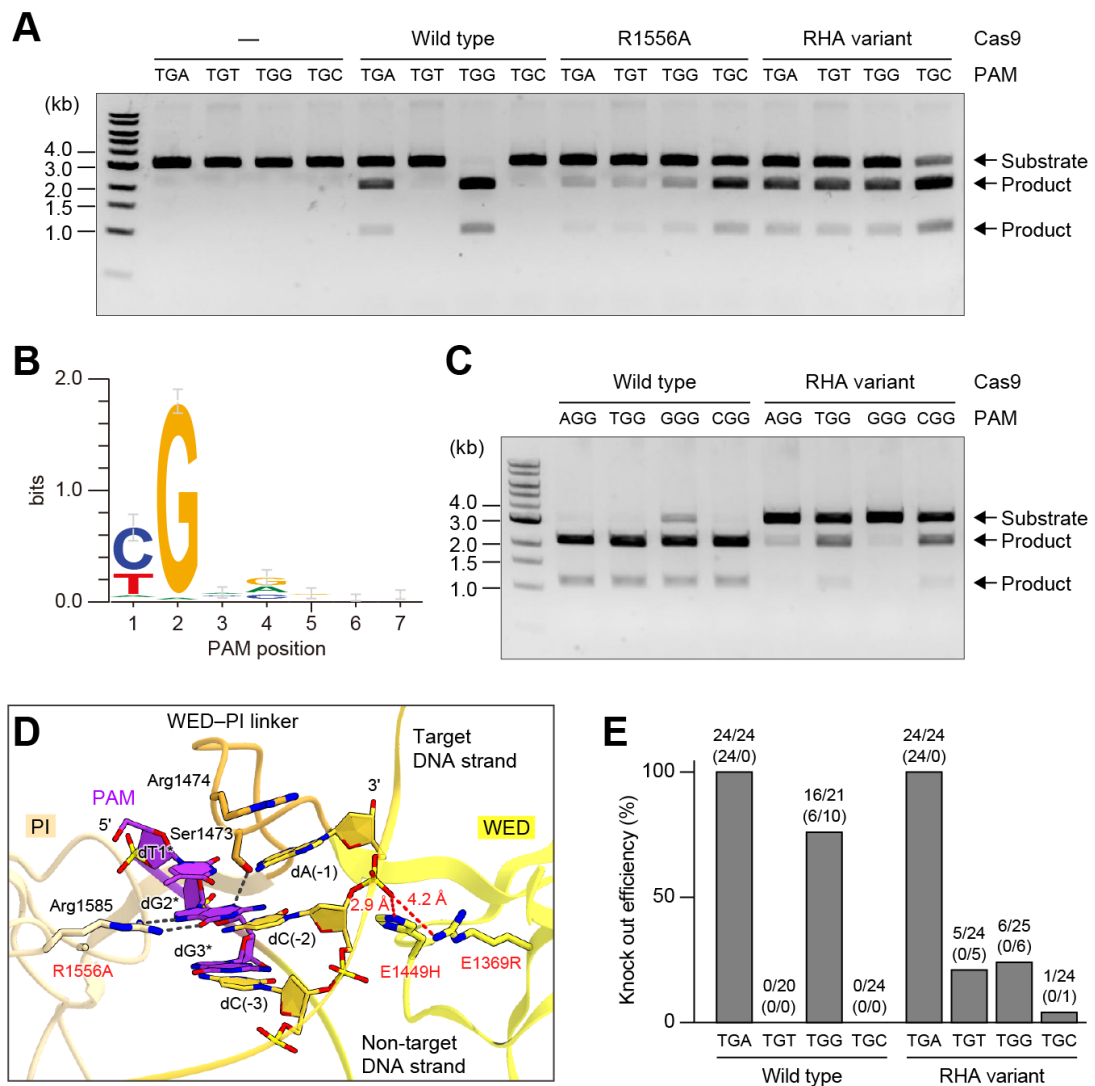
Figure 15. Structure-guided engineering and genome editing in mouse zygotes

(A) *In vitro* cleavage activity of wild-type FnCas9, and R1556A and E1369R/E1449H/R1556A (RHA) variants of 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 (This data was provided by Izuho Hatada).

44
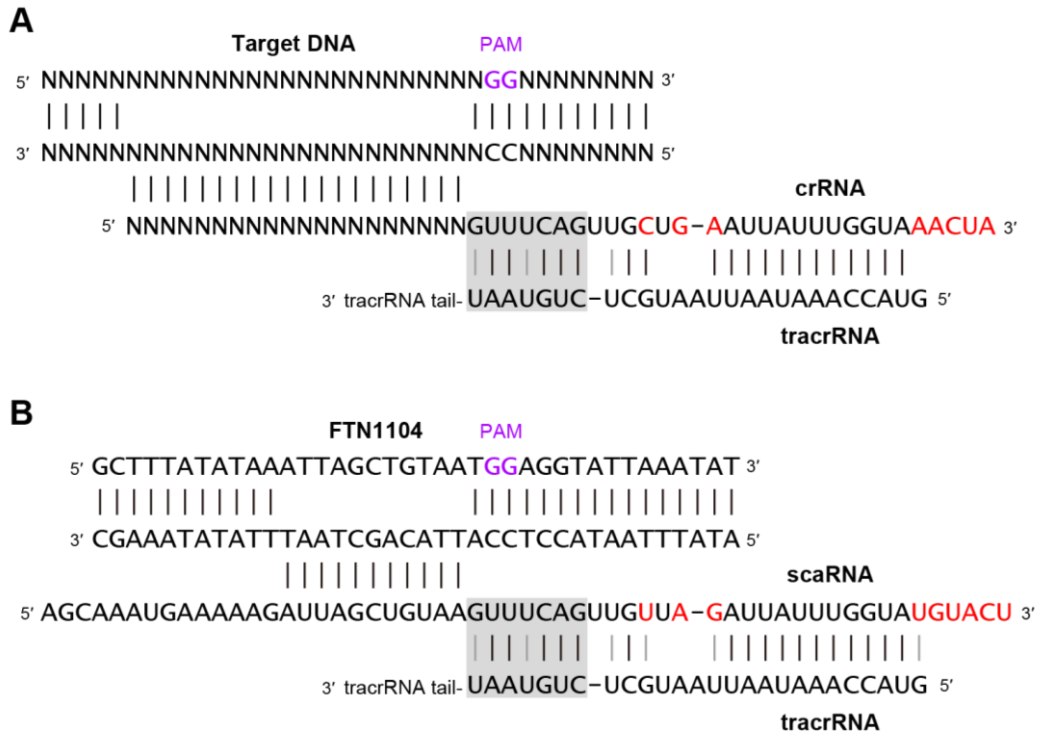
Figure 16. Bifunctionality of FnCas9

Schematic representation of crRNA:tracrRNA-mediated DNA cleavage (A) and scaRNA:tracrRNA-mediated gene expression regulation (B). The different sequences between crRNA and scaRNA are shown in red. The lower stem region of repeat:antirepeat duplex is highlighted in gray. FTN1104 is one of the gene repressed by scaRNA in *Francisella novicida*.
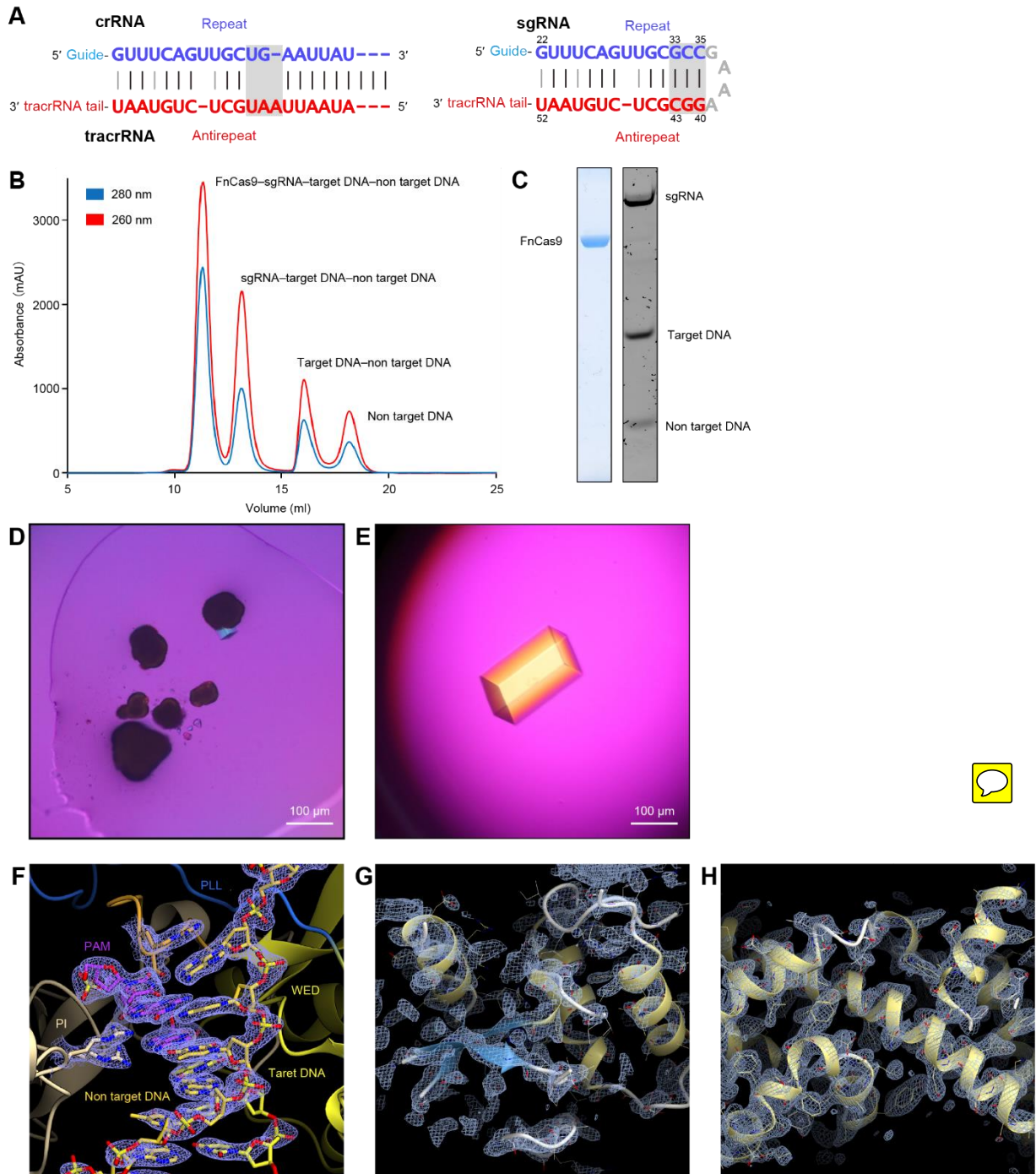
Figure 17. Structural determination

(A) The repeat:antirepeat duplex in the sgRNA for crystallization.

(B) Gel filtration profile of FnCas9 FnCas9–sgRNA–DNA complex.

(C) SDS-PAGE and Urea denaturing PAGE analysis of the eluted peak flaction corresponding to FnCas9–sgRNA–DNA complex.

(D–E) The crystals obtained by initial sitting-drop vapor diffusion (D) and microseeding method (E).

(F–H) The $2mF_O–DF_C$ electron density map around PAM (contoured at 3σ), HNH (G) and REC2 (H) domain (contoured at 1σ).
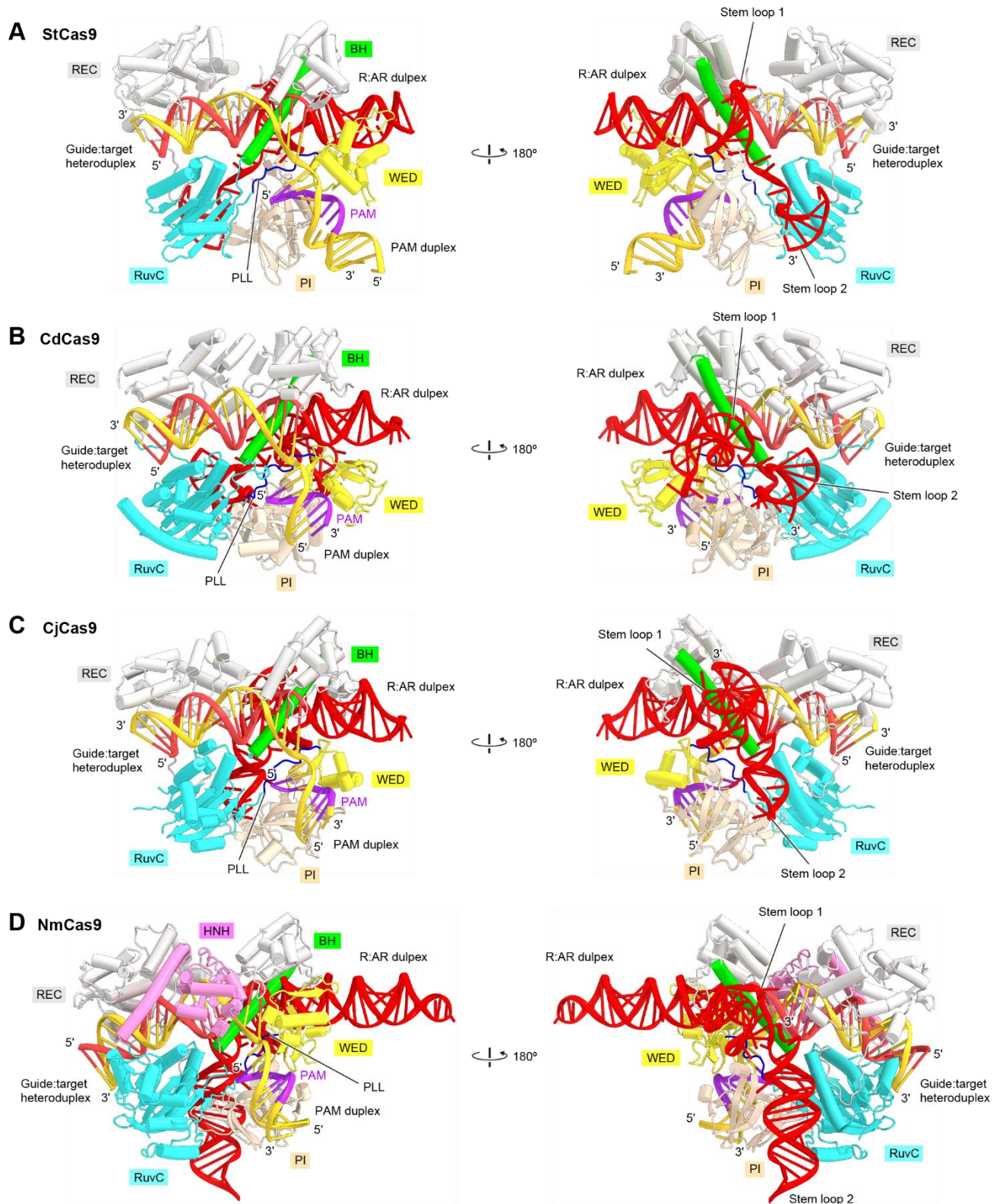
Figure 18. Structural comparison of StCas9, CdCas9, CjCas9 and NmCas9

(A–D) Crystal structures of StCas9 (PDB: 6RJD) (A), CdCas9 (PDB: 6JOO) (B), CjCas9 (PDB: 5X2H) (C) and NmCas9 (PDB:6JDV) (D)

Tables

Table 1. Data collection and refinement statistics

	TGG PAM (Native)	TGA PAM (Native)	TGG PAM (RHA variant)	TGG PAM (SeMet)
Data collection				
Beamline	SPring-8 BL41XU	SPring-8 BL41XU	SPring-8 BL41XU	SLS PXIII
Wavelength (Å)	1.0000	1.0000	1.0000	0.9780
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$
Cell dimensions				
a, b, c (Å)	81.9, 159.1, 96.8	81.6, 159.3, 96.7	81.6, 156.0, 96.7	81.3, 157.4, 96.5
$α$, $β$, $γ$ (°)	90, 107.0, 90	90, 106.9, 90	90, 106.9, 90	90, 106.8, 90
Resolution (Å)*	46.3–1.7 (1.73–1.70)	46.3–1.7 (1.73–1.70)	46.3–1.7 (1.73–1.70)	19.9–2.3 (2.34–2.30)
R_{merge}	0.062 (0.72)	0.060 (1.42)	0.039 (0.77)	0.091 (1.02)
R_{pim}	0.037 (0.44)	0.024 (0.56)	0.024 (0.47)	0.026 (0.28)
$I/σI$	14.2 (2.8)	19.6 (2.4)	20.3 (2.5)	19.1 (2.0)
Completeness (%)	96.3 (96.2)	98.8 (97.9)	98.9 (98.7)	99.7 (98.9)
Multiplicity	7.2 (7.2)	14.1 (14.3)	7.0 (7.1)	14.3 (13.8)
CC (1/2)	0.99 (0.65)	0.99 (0.86)	1.00 (0.74)	0.96 (0.86)
Refinement				
Resolution (Å)	46.3–1.7	46.2–1.7	46.3–1.7	
No. reflections	249,252	255,162	254,945	
R_{work}/ R_{free}	0.184 / 0.207	0.180 / 0.203	0.183 / 0.206	
No. atoms				
Protein	11,791	11,809	11,802	
Nucleic acid	2,771	2,771	2,771	
Ion	24	24	24	
Solvent	1,131	1,153	1,097	
B-factors (Å^2)				
Protein	50.1	49.5	51.5	
Nucleic acid	34.7	35.1	35.3	
Ion	53.5	53.7	53.5	
Solvent	41.0	41.7	41.5	
R.m.s. deviations				
Bond lengths (Å)	0.010	0.014	0.008	
Bond angles (°)	1.14	1.39	1.00	
Ramachandran plot (%)				
Favored region	97.2	97.4	97.7	
Allowed region	2.6	2.5	2.2	
Outlier region	0.2	0.1	0.1	

*Values in parentheses are for the highest resolution shell.

Table 2. Target sequences of mouse zygotes

Gene	Target sequence	PAM	Indel detection
Tet1EX4	CAGGGAGCTCATGGAGACTAGG	TGA	*Bfa*I
Tet1EX4	CACTTGGTCCTGCCCCAAGGTG	TGT	*Eco*T14 I
Tet1EX4	GTGGCTGCTGTCAGGGAGCTCA	TGG	*Sac*I
Tet1EX4	ATGGAGACTAGGTGAGGAACTC	TGC	HMA

Reference

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

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

Bondy-Denomy, J., Pawluk, A., Maxwell, K. et al. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. Nature 493, 429–432 (2013).

Briner, A.E., Donohoue, P.D., Gomaa, A.A., Selle, K., Slorach, E.M., Nye, C.H., Haurwitz, R.E., Beisel, C.L., May, A.P., and Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. Mol Cell 56, 333-339.

Chen, F., Ding, X., Feng, Y. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nat Commun 8, 14958 (2017).

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

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

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

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

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

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

thermophilus. J Bacteriol 190, 1390-1400.

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

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

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

Fuchsbauer O, Swuec P, Zimberger C, Amigues B, Levesque S, Agudelo D, Duringer A, Chaves-Sanjuan A, Spinelli S, Rousseau GM, et al. 2019. Cas9 allosteric inhibition by the anti-CRISPR protein AcrIIA6. Mol Cell 76: 922–937.

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

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

Hirano, S., Abudayyeh, O.O., Gootenberg, J.S. et al. Structural basis for the promiscuous PAM recognition by Corynebacterium diphtheriae Cas9. Nat Commun 10, 1968 (2019).

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

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

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

Jiang, F., Taylor D.W., Chen J.S., Kornfeld J.E., Zhou K., Thompson A.J., Nogales E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science (DOI: 10.1126/science.aad8282).

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

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

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr 66, 125-132.

Kim Y, Lee SJ, Yoon HJ, Kim NK, Lee BJ & Suh JY (2019) Anti-CRISPR AcrIIC3 discriminates between Cas9 orthologs via targeting the variable surface of the HNH nuclease domain. FEBS J, 286, 4661-6674.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Res 24, 1012-1019.

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

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

Lin, S., Staahl, B.T., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife 3, e04766.

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.

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

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

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

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.

Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature 497, 254-257.

Shibata, M., Nishimasu, H., Kodera, N. et al. Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. Nat Commun 8, 1430 (2017).

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

Sung, Y.H., Kim, J.M., Kim, H.T., Lee, J., Jeon, J., Jin, Y., Choi, J.H., Ban, Y.H., Ha, S.J., Kim, C.H., et al. (2014). Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res 24, 125-131.

Sun W, Yang J, Cheng Z, Amrani N, Liu C, Wang K, Ibraheim R, Edraki A, Huang X, Wang M, Wang J, Liu L, Sheng G, Yang Y, Lou J, Sontheimer EJ, Wang Y. (2019). Structures of Neisseria meningitidis Cas9 Complexes in Catalytically Poised and Anti-CRISPR-Inhibited States. Mol. Cell 76, 938-952.

Woo, J.W., Kim, J., Kwon, S.I., Corvalan, C., Cho, S.W., Kim, H., Kim, S.G., Kim, S.T., Choe, S., and Kim, J.S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol 33, 1162-1164.

M. Yamada, Y. Watanabe, J.S. Gootenberg, H. Hirano, F.A. Ran, T. Nakane, R. Ishitani, F. Zhang, H. Nishimasu, O. Nureki, Crystal structure of the minimal Cas9 from Campylobacter jejuni reveals the molecular diversity in the CRISPR-Cas9 systems. Mol. Cell, 65 (2017).

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.

Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. Nat Biotechnol 33, 73-80.

Zhang, F., Song, G., Tian, Y. (2019). Anti-CRISPRs: The natural inhibitors for CRISPR‐Cas systems. Animal Model Exp. Med. Jun; 2, 69-75.

Appendix 1. Sequence alignment of type II-A Cas9 and FnCas9

```
II-A_S.pyogenes     1  ....MDKKYSIGLDIGTNSVGWAVITDDYKVPSKKLKGLGNTDRHGIKKNLIGALLFDSG
II-A_E.faecalis     1  ....MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEG
II-A_P.duerdenii    1  ..MKNLKEYYIGLDIGTASVGWAVTDESYNIPK..........FNGKKMWGVRLFDDA
II-A_F.nucleatum    1  MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLR..........FNKKDMWGSRLFDEA
II-A_A.intestini    1  ...MGKMYYLGLDIGTNSVGYAVTDPSYHLLK..........FKGEPMWGAHVFAAG
II-A_C.mitsuokai    1  .MSQNIVDYCIGLDLGTGSVGWAVVDMNHRLMK..........RNGKHLWGSRLFSNA
II-A_L.rhamnosus    1  .MTKLNQPYGIGLDIGSNSIGFAVVDANSHLLR..........LKGETAIGARLFREG
II-B_F.novicida     1  .MNFKILPIAIDLGVKNTGVFSAFYQKGTSLER..........LDNKNGKVYELSKDSY


II-A_S.pyogenes    57  ETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEED.....K
II-A_E.faecalis    57  HTAEDRRLKRTARRRISRRRNRLRYLQAFFEEAMTDLDENFFARLQESFLVPED.....K
II-A_P.duerdenii   47  KTAEERRTQRGSRRRLNRRKERINLLQDLFATEISKVDPNFFLRLDNSDLYRED.....K
II-A_F.nucleatum   49  KTAAERRVQRNSRRRLKRRKWRLNLLEEIFSDEIMKIDSNFFRRLKESSLWLED.....K
II-A_A.intestini   45  NQSAERRSFRTSRRRLDRRQQRVKLVQEIFAPVISPIDPRFFIRLHESALWRDD.....V
II-A_C.mitsuokai   48  ETAATRRSSRSIRRRYNKRRERIRLLRAILQDMVLEKDPTFFIRLEHTSFLDEEDKAKYL
II-A_L.rhamnosus   48  QSAAERRRESRTTRRRLSRTRWRLSFLRDFFAPHITKIDPDFFLRQKYSEISPKD......
II-B_F.novicida    49  TLLMNNRTARRHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAISFLFNRRG.FSFI


II-A_S.pyogenes   112  KHERHPIFGNIVD....EVAYHEKYPTIYHLRKKLADSTDKVDLRLIYLALAHMIKFRGH
II-A_E.faecalis   112  KWHRHPIFAKLED....EVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGH
II-A_P.duerdenii  102  DEKLKSKYTLFNDKDFKDRDYHKKYPTIHHLIMDLIEDEGKKDIRLLYLACHYLLKNRGH
II-A_F.nucleatum  104  NS..KEKFTLFNDDNYKDYDFYKQYPTIFHLRDELIKNPEKKDIRLIYLALHSIFKSRGH
II-A_A.intestini  100  AETDKHIF..FNDPTYTDKEYYSDYPTIHHLIVDLMESSEKHDPRLVYLAVAWLVAHRGH
II-A_C.mitsuokai  108  GTDYKDNYNLFIDEDFNDYTYYHKYPTIYHLRKALCESTEKADPRLIYLALHHIVKYRGH
II-A_L.rhamnosus  102  ..KDRFKYEKRLFNDRTDAEFYEDYPSMYHLRLHLMTHTHKADPREIFLAIHHILKSRGH
II-B_F.novicida   108  TDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLMQKILE


II-A_S.pyogenes   168  FLIEGD.......LNPDNSDVDKLFIQLVQTYNQLFEENP......INASRVDAKAI.LS
II-A_E.faecalis   168  FLIEGK.......LSTENISVKEQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEE.LT
II-A_P.duerdenii  162  FIFEGQ.......KFDTKNSFDKSINDLKIHLRDEYN..IDLEFNNEDLIEIITDTT.LN
II-A_F.nucleatum  162  FLFEGQ.......NLKEIKNFETLYNNLISFLEDNGINKSIDKDNIEKLEKIICDSG.KG
II-A_A.intestini  158  FLNEVDK....DNIGDVLSFDAFYPEFLAFLSDNGVSPWVCESKALQATLLSRNSV.ND
II-A_C.mitsuokai  168  FLYEGQK....FNMDASNIEDKLSDVFTQFADFNNIPYEDDEKKNLEILEILKKPLSKK
II-A_L.rhamnosus  160  FLTPGAAKDFNTDKVDLEDIFPALTEAYAQVYPDLELTFDLAKADDFKAKLLDEQATPSD
II-B_F.novicida   168  FKLMKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELS


II-A_S.pyogenes   214  ARLSKSRRLENLIAQLPGEKK.....NGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSK
II-A_E.faecalis   220  EKASRTKKSEKVLQQFPQEKA.....NGLFGQFLKLMVGNKADFKKVFGLEEEAKITYAS
II-A_P.duerdenii  212  KTN.KKKELKNIVGDTKFLKA.....ISAIMIGSSQKLVDLFEDGEFEETTVKS..VDFS
II-A_F.nucleatum  214  LKD.KEKEFKGIFNSDKQLVA.....IFKLSVGSSVSLNDLFDTDEYKKEEVEKEKISFR
II-A_A.intestini  212  KYKALKSLIFGSQKPEDNFDA.....NISEDGLIQLLAGKKVKVNKLFP.QESNDASFTL
II-A_C.mitsuokai  223  AKVDEVMALISPEKEFKSAYK.....ELVTGIAGNKMNVTKMILCESIKQGDSEIKLKFS
II-A_L.rhamnosus  220  TQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAITGLKTKFNLALGTEVDEADASNWQFSM
II-B_F.novicida   228  YYHHDKYNIQEFLKRHATINDRILDTLLTDDLDIWNFNFEKFDFDKNEEKLQNQEDKDHI


II-A_S.pyogenes   269  DTYDDDLDNLLAQIGDQYADLFLAAKNLSDATLLSDILRVNSEIT..KAPLSASMIKRYD
II-A_E.faecalis   275  ESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKS..HAKLSSSMIVRFT
II-A_P.duerdenii  264  TTAFDDKYSEYEEALGDTISLLNILKSIYDSSILENLLKDADKSKDGNKYISKAFVKKFN
II-A_F.nucleatum  268  EQIYEDDKPIYYSILGEKIELLDIAKSFYDFMVLNNILSDSN......YISEAKVKLYE
II-A_A.intestini  266  NDKEDAIEEILGTLTPDECEWIAHIRRLFDWAIMKHALKDGR......TISESKVKLYE
II-A_C.mitsuokai  278  DSNYDDQFSEVENDLGEYVEFIDSLHNIYSWVELQTIMGATHTDN...ASISEAMVSRYN
II-A_L.rhamnosus  280  GQLDDKWSNIETSMTDQGTEIFEQIQEHYRARLLNGIVPAGMS......LSQAKVADYG
II-B_F.novicida   288  QAHLHHFVFAVNKIKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNFCENLHNKKYSN


II-A_S.pyogenes   327  EHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG..............ASQEEF
II-A_E.faecalis   333  EHQEDLKKFKRFIRENCPDEYDNLFKNEQKDGYAGYIAHAG..........KVSQLKF
II-A_P.duerdenii  324  KHGKDLKTLKRIIKKYLPSEYANIFRNKSINDNYVAYTKSNITSNKR.TKASKFTKQEDF
II-A_F.nucleatum  321  EHKKDLKNLKYIIRKYNKENYDKLFKDKNENNYPAYIGLNKEKDKKE.VVEKSRLKIDDL
II-A_A.intestini  319  QHHHDLTQLKYFVKTYLAKEYDDIFRNVDSETTKNYVAYSYHVKEVKGTLPKNKATQEEF
II-A_C.mitsuokai  335  KHHEDLQLLKKCIKDNVPKKYFDMFRNDSEKVKGYYNYINR......PNKAPVDEF
II-A_L.rhamnosus  333  QHKEDLELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLR..............EDF
II-B_F.novicida   348  LSVKNLVNLIGNLSNLELKPLRKYFNDKIHAKADHWDEQKFTETYCHWILGEWRVGVKDQ


II-A_S.pyogenes   373  YKFIKPILEKMDG.............TEELLAKLNREDLLRKQRTFDNGSIPYQIHLGEL
II-A_E.faecalis   381  YQYVKKIIQDIAG..........AEYFLEKIAQENFLRKQNFLRKQRTFDNGVIPHQIHLAEL
II-A_P.duerdenii  383  YKFIKKHLDTIKETKLNSSENEDLKLIDEMLTDIEFKTFIPKLKSSDNGVIPYQLKLMEL
II-A_F.nucleatum  380  IKVIKGYLP.....KPERIEEKDKTIFNEILNKIELKTILPKQRISDNGTLPYQIHEVEL
II-A_A.intestini  379  CKYVLGKVKNIECS......EADKVDFDEMIQRLTDNSFMPKQVSGENRVIPYQLYYYEL
II-A_C.mitsuokai  385  YKFVKKCIEKVDTP..........EAKQILHDIELENFLLKQNSRTNGSVPYQMQLDEM
II-A_L.rhamnosus  379  VKALTKEVTAHPN...........EVSEQLLNRMGQANFMLKQRTKANGAIPVQLQQREL
II-B_F.novicida   408  DKKDGAKYSYKDLCN.ELKQKVTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQSLILN


II-A_S.pyogenes   420  HAILRRQEDFYPFLKDN..........REKIEKILTFRIPYYVGPL.......ARG.NSR
II-A_E.faecalis   428  QAIIHRQAAYYPFLKEN..........QEKIEQLVTFRIPYYVGPL.......SKGDAST
II-A_P.duerdenii  443  KKILDNQSKYYDFLNESD...EYGTV.KDKVESIMEFRIPYYVGPLNP....DSK....
II-A_F.nucleatum  435  EKILENQSKYYDFLNYEE...NGVST.KDKLLKTFKFRIPYYVGPLNS.....YHKDKGG
II-A_A.intestini  433  KTILNKAASYLPFLTQCG...KDAISNQDKLLSIMTFRIPYFVGPLR.......KDNSE
II-A_C.mitsuokai  434  IKIIDNQAKYYPVLKEK..........REQLLSILTFRIPYYFGPLN.......ETSEHAW
II-A_L.rhamnosus  428  DQIIANQSKYYDWLAKPNPVEKHRASKPYQLDELLTFRIPYYVGPLITPKEQAESRENVF
II-B_F.novicida   467  PKFLDNQYPNWQQYLQELKKLQSIQNYLDSFETDLKVLKSSKDQPYFVEYK.SSNQQIAS


II-A_S.pyogenes   462  FAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVY
II-A_E.faecalis   471  FAWLKRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLPSEKVLPKHSLLYEKFMVF
II-A_P.duerdenii  490  YAWIKR.ENTKITPWNFEKDIVDLDSSREEFIDRLIGRCTYLKEEKVLPKASLIYNEFMVL
II-A_F.nucleatum  486  NSWIVRKEEGKILPWNFEQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDSFLYSEYIIL
II-A_A.intestini  482  HAWLER.KAGKIYPWNFNDKVDLDKSEEAFIRRMTNTCTYYPGEDVLPLDSLIYEKFMIL
II-A_C.mitsuokai  478  IKRLEGKENQRILPWNYQDTVDVDATAEGFIKRMRSYCTYFPDEEVLPKNSLIVSKYEVY
II-A_L.rhamnosus  488  AWMIRKDPSGNITPYNFDEKVDCPASANMFIQRMKNTDTYLIGEDVLPKRSLLYQKYEVL
II-B_F.novicida   526  GQRDYKDLDARILQFIFDRVKASDELLLNEIYFQAKKLKQKASSELEKLESSKKLDEVIA
```

```
                        .                   .           .             .
II-A_S.pyogenes   522  NELTKVKYVTEGMR.........KPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKI
II-A_E.faecalis   531  NELTKISYTDD.RG.........IKANFSGKEKEKIFDYLFKTRRKVKKKDIIQFYRNEY
II-A_P.duerdenii  549  NELNNLKLNEF..............LITEEMKKAIFEELFKTKKKVTLKAVSNLLKKEF
II-A_F.nucleatum  546  NELNKVQVNDE...............FLNEENKRKIIDELFKENKKVSEKKFKEYLLVNQ
II-A_A.intestini  541  NEINNIRIDGY...............PISVDVKQQVFGLFEKKRRVTVKDIQNLLLSLG
II-A_C.mitsuokai  538  NELNKIRVDDK..............LLEVDVKNDIYNELFMKNKTVTEKKLKNWLVNNQ
II-A_L.rhamnosus  548  NELNNVSIDGR...............RLGADQKQRLIREVFERHSSVTIKQVAENLVAHG
II-B_F.novicida   586  NSQLSQILKSQHTNGIFEQGTFLHLVCKYYKQRQRARDSRLYIMPEYRYDKKLHKYNNTG

                          .           .           .           .           .
II-A_S.pyogenes   573  ECFDSVEISGV..EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDRE
II-A_E.faecalis   581  NTEIVTLSGLE..EDQFNASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQ
II-A_P.duerdenii  594  NLTGDILLSGT..DGDFKQGLNSYIDFKNIIGDKVDRDDYRIKIEEIIKLIVLYEDDKTY
II-A_F.nucleatum  591  IANRTVELKGI..KDSFNSNYVSYIKFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKI
II-A_A.intestini  585  ALDKHGKLTGI..DTTIHSNYNTYHHFKSLMERG.....VLTRDDVERIVERMTYSDDTK
II-A_C.mitsuokai  583  CCRKDAEIKGFQKENQFSTSLTPWIDFTNIFGKID....QSNFDLIEKIIYDLTVFEDKK
II-A_L.rhamnosus  593  DFAKRPEIRGLADEKRFLSSLSTYHQLKKILHEAIDDPTKEHDIEN...IITWSTVFEDA
II-B_F.novicida   646  RFDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVSPNFLKDKIGSDDDLFISKWLVEHIRG

                            .           .           .           .
II-A_S.pyogenes   631  MIEERLKTYAHLFDDKVMKQLKR.RRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG..
II-A_E.faecalis   639  RIRTQLSTFKGQFSAEVLKKLER.KHYTGWGRLSKKLINGIYDKESGKTILGYLIKDDGV
II-A_P.duerdenii  652  LKKKIKSAYKNDFTDDEIKKIAA.LNYKDWGRLSKRFLTGIEGVDKTTGEKGS.IIYFMR
II-A_F.nucleatum  649  FEKKIKNEYGDILNKDEIKKINS.FKFNTWGRLSEKLLTGIEFINLETGECYSSVMEALR
II-A_A.intestini  638  RVRLWLNNSYGTLTADDVKHISR.LRKHDFGRLSKMFLTGLKGVHKETGERASILDFMWN
II-A_C.mitsuokai  639  IMKRRLKKKYALP.DDKIKQILK.LKYKDWSRLSKKLLDGIVADNRFGSSVTVLDVLEMS
II-A_L.rhamnosus  650  AIFKTELAKISWLDSDTIKALSN.IRYRGWGQFSRKLLDGLKLGNG........HTVIQE
II-B_F.novicida   706  FKKACEDSLKIQKDNRGLLNHKINIARNTKGKCEKEIFNLICKIEGSEDKKGNYKHGLAY

                                                                  .           .
II-A_S.pyogenes   688  ...........................FANRNFMQLIHDDSLTFKED..IQKA
II-A_E.faecalis   698  SK.........................HYNRNFMQLINDSQLSFKNA..IQKA
II-A_P.duerdenii  710  .........................EYNLNLMELMSG.HYTFTEE..VEKL
II-A_F.nucleatum  708  .........................RTNYNLMELLSS.KFTLQES..IDNE
II-A_A.intestini  697  ........................TNDNLMQLLSECYTFSDEITKLQEA
II-A_C.mitsuokai  697  .......................RLNLMEIINDKELGYAQMIEEASS
II-A_L.rhamnosus  701  LM.....................LSNHNLMQILTDETLKTTMA...ELN
II-B_F.novicida   766  ELGVLLFGEPNEASKPEFDRKIKKFNSIYSFAQIQQIAFAERKGNANTCAVCSADNAHRM

                            .           .           .           .           .
II-A_S.pyogenes   712  QVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQ
II-A_E.faecalis   724  QSSEHEETLSETVNELAGSPAIKKGIYQSLKIVDELVAIMG.YAPKRIVVEMARENQTTS
II-A_P.duerdenii  733  NPVENRELCYEMVDELYLSPSVKRMLWQSLRVVDEIKRIIG.KDPKKIFIEMAR..AKEA
II-A_F.nucleatum  731  NKEMNEVSYRDLIEESYVSPSLKRAILQTLKIYEEIKKITG.RVPKKVFIEMARGGDESM
II-A_A.intestini  722  YYAKAQLSLNDFLDSMYISNAVKRPIYRTLAVVNDIRKACG.TAPKRIFIEMARDGESKK
II-A_C.mitsuokai  721  CP.KDGKFTYEEVAKLAGSPALKRGIWQSLQIVEEITKVMK.CRPKYIYIEFERSEEAKE
II-A_L.rhamnosus  726  QDKLKADDIEDVINDAYTSPSNKKALRQVLRVVEDIKHAANGQDPSWLYIETADGPGTAG
II-B_F.novicida   826  QQIKITEPVEDNKDKIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVL

                            .           .           .           .           .
II-A_S.pyogenes   772  KGQKNSRERMKRIEEGIKELGSDILKEY...............PVENTQLQNEKLYLYYLQN
II-A_E.faecalis   783  TGKRRSIQRLKIVEKAMAEIGSNLLKEQ..............PTTNEQLRDTRLFLYYMQN
II-A_P.duerdenii  790  KNSRKE.SRKNKLLEFYKFGKKAFINEIGEERYNYLLNEINSEEESKFRWDNLYLYYTQL
II-A_F.nucleatum  790  KNKKIP.ARQEQLKKLYDSCGNDIANFS..IDIKEMKNSLSSYDNNSLRQKKLYLYYLQF
II-A_A.intestini  781  ...KRSVTRREQIKNLYRSIRKDFQQEVD....FLEKILENKSDGQLQSDALYLYFAQL
II-A_C.mitsuokai  779  RTESKIKKLENVYKDLDEQTKKEYKSVLEELKG.......FDNTKKISSDSLFLYFTQL
II-A_L.rhamnosus  786  ...KRTQSRQKQIQTVYANAAQELIDSAVRG......ELEDKIADKASFTDRLVLYFMQG
II-B_F.novicida   886  SAKHQLHIPIITESNAFEFEFEPALADVKGKSLKDRRKKALERISPENIFKDKNNRIKEFAK

                          .           .           .           .           .
II-A_S.pyogenes   819  GRDMYVDQELDINRL....SDYDVDHIVPQS...FLKDDSIDNKVLTRSDKNRGKSD...
II-A_E.faecalis   830  GKDMYTGDELSLHRL....SHYDIDHIIPQS...FMKDDSLDNLVLVGSTENRGKSD...
II-A_P.duerdenii  849  GRCMYSLEPIDLADLKSNN.IYDQDHIYPKS...KIYDDSLENRVLVKKNLNHEKGN...
II-A_F.nucleatum  847  GKCMYTGREIDLDRLLQNNDTYDIDHIYPRSK..VIKDDSFDNLVLVLKNENAEKSN...
II-A_A.intestini  833  GRDMYTGDPIKLEHIK.DQSFYNIDHIPQS...MVKDDSLDNKVLVQSEINGBKSS...
II-A_C.mitsuokai  831  GKCMYSGKKLDIDSLD....KYQIDHIVPQS...LVKDDSFDNRVLVLPSENQRKLD...
II-A_L.rhamnosus  837  GRDIYTGAPLNIDQLS....HYDIDHILPQS...LIKDDSLDNRVLVNATINREKNN...
II-B_F.novicida   946  GISAYSGANLTDGDFDG..AKEELDHIIPRSHKKYGTLNDEANLICVTRGDNKNKGNRIF

                                                        .           .           .
II-A_S.pyogenes   869  .............NVPSEEVVKKMKNY..WRQLLNAKLITQRKFDNLTKAERGGLSELD
II-A_E.faecalis   880  .............DVPSKEVVKDMKAY..WEKLYAAGLISQVIITLKASLTQFRSIFGL
II-A_P.duerdenii  902  ...........QYPIPEKVLNKNAYGFWKILFDKGLIGQKKYTRLTR..RTPFEERE
II-A_F.nucleatum  902  ...........EYPVKKEIQEKMKS.FWRFLKEKNFISDEKYKRLTG..KDDFELR.
II-A_A.intestini  886  ...........RYPLDAAIRNKMKP.LWDAYYNHGLISKYQRLTR..STPFTDDE
II-A_C.mitsuokai  881  ...........DLVVPFDIRDKMYR.FWKLLFDHELISPKKFYSLIK..TEYTERD
II-A_L.rhamnosus  887  ...........VFASTLFASKMAAT..WRKWHEAGLISGRKLRNLMLR..PDEIDKF
II-B_F.novicida   1004 CLRDLADNYKLKQFETTDDLEIEKKIADTIWDANKKDFKFG..NYRSFINLTPQEQKAFR

                          .           .           .           .           .
II-A_S.pyogenes   913  KVGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVRVITLKSKLVSDFRKDFQF
II-A_E.faecalis   924  KAHFIQRQLVETRQITKNVAGILDQRYNANSKE.....KKVQIITLKASLTSQFRSIFGL
II-A_P.duerdenii  946  LAEFIERQIVETRQATKETANLLKNICQD...........SEIVYSKAENASRFRQEFDI
II-A_F.nucleatum  944  ..GFMARQLVNVRQTTKEVGKILQQIEPE...........IKIVYSKAEIASSFREMFDF
II-A_A.intestini  929  KWDFINRQLVETRQSTKALAILLKRKFPD...........TEIVYSKAGLSSDFRHEFGL
II-A_C.mitsuokai  923  EERFINRQLVETRQITKNVTQIIEDHYST.......TKVAAIRANLSHEFRVKNHI
II-A_L.rhamnosus  929  AHGFVARQLVETRQIIKLTEQIAAAQYPN.........TKIIAVKAGLSHQLREELDF
II-B_F.novicida   1062 HALFLADENPIKQAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLNTDKISFDY

                                                                  .           .
II-A_S.pyogenes   973  YKVREIN...........................NYHHAHDAYLNAVVGTALIKKYP
II-A_E.faecalis   979  YKVREVN...........................DYHHGQDAYLNCVVATTLLKVYP
II-A_P.duerdenii  995  IKCRTVN...........................DLHHMHDAYLNIIVVGNVYNTKFT
II-A_F.nucleatum  991  IKVRELN...........................DTHHAKDAYLNIIVAGNVYNTKFT
II-A_A.intestini  978  VKSRNIN...........................DLHHAKDAFLAIVTGNVYHERFN
II-A_C.mitsuokai  972  YKNRDIN...........................DYHHAHDAYIVALIGGFMRDRYP
II-A_L.rhamnosus  978  PKNRDVN...........................HYHHAFDAFLAARIGTYLLKRYP
II-B_F.novicida   1122 FGIPTIGNGRGIAEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDG
```

56

```
II-A_S.pyogenes   1003 KLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPL
II-A_E.faecalis   1009 NLAPEFVYGEYPKFQTFKENKATAKAIIYTNLLRFFTEDEPRFTK...............
II-A_P.duerdenii  1025 KNPLNFIK.DKDNVRSYNLENMFKYDVVRGSYTAWIADDSEGNVKA...............
II-A_F.nucleatum  1021 EKPYRYLQ.EIKEN..YDVKKIYNYDIKN....AWDKENS....................
II-A_A.intestini  1008 RR.....WFMVNQPYSVKTKTLFTHSIKNGNFVAWNGEEDLG.................
II-A_C.mitsuokai  1002 NMHDSKAVYSEYMKMFRKNKNDQKKWKDGFVINSMNYPYEVDGKLIWN............
II-A_L.rhamnosus  1008 KLAPFFTYGEFAKVDVKKFREFNFIGALTHAKKNIVAKDTGEIVWDKE............
II-B_F.novicida   1182 SIGLEIDKNYSLYPLDKNTGEVFTKDIFSQIKITDNEFSDKKLVRKKAIEGFNTHRQMTR

II-A_S.pyogenes   1063 IETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILP..........
II-A_E.faecalis   1054 ......DGEILWSN.SYLKTIKKELNYHQMNIVKKVEVQKGGFSKESIKP..........
II-A_P.duerdenii  1070 ...............ATIKKVKRELEGKNYRFTRMSYIGTGGLYDQNLMR..........
II-A_F.nucleatum  1054 ................LEIVKKNMEKNTVNITRFIKEEKGELFNLNPIK...........
II-A_A.intestini  1045 ................RIVKMLKQNKNNIHFTRFSFDRKEGLFDIQPLK...........
II-A_C.mitsuokai  1050 ...........PDLINEIKKCFYYKDCYCTTKLDQKSGQMFNLTVLP...........
II-A_L.rhamnosus  1056 ..............RDIQELDRIYNFKRMLITHEVYFETADLFKQTIYA..........
II-B_F.novicida   1242 DG....IYAENYLPILIHKELNEVRKGYTWKNSEEIKIFKGKKYDIQQLNNLVYCLKFVD

II-A_S.pyogenes   1113 ...KRNSD........KLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVK
II-A_E.faecalis   1097 ...KGPSN........KLIPVKNGLDPQKYGGFDSPIVAYTVLFTHEK....GKKPLIKQ
II-A_P.duerdenii  1105 ...KGKGQ.....IPQKENTNKSN..IEKYGGYNKASSAYFALIESDGKAGRERTLETIP
II-A_F.nucleatum  1087 ...KGETSNEIISIKPKLYDGKDNKLNEKYGYYTSLKAAYFIYVEHEKKNKKVKTFERIT
II-A_A.intestini  1078 ...ASTGLVP.......RKAGLDVVKYGGYDKSTAAYYLLVRFTLEDKKTQHKLMMI
II-A_C.mitsuokai  1086 ...NDAHSAKG.TTEAVIPVNKNRKDVNKYGGFSGLQYVIAAIEGTKKKGKKLVKVRKLS
II-A_L.rhamnosus  1091 ...AKDSKERG..GSKQLIPKKQGYPTQVYGGYTRENTAYLAVVKTVEKKADVYRVVRIA
II-B_F.novicida   1298 KPISIDIQISTLEELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYNTALGYKKYSKE

II-A_S.pyogenes   1162 ELLGITIMERSSFEKDPIDFLEAKGYKEVRKDLIIKLPKYSLFELENGRKRMLASAGELQ
II-A_E.faecalis   1142 EILGITIMEKTRFEQNPILFLEEK..GFLRPRVLMKLPKYTLYEFPEGRRRLLASAKEAQ
II-A_P.duerdenii  1155 IMVYNQEKYGNTEAVDKYLKDNLELQDPKILKDKIKINSLIKLDGFLYNIKGKTGDSLSI
II-A_F.nucleatum  1144 RIDSTLIKN..EKNLIKYLVSQKKLLNPKIIKKIYKEQTLIIDSYPYTFTGVDSNKKVEL
II-A_A.intestini  1125 PVEG.LYKARIDHDKEFLTDYAQTTISEILQKDKQKVINIMFPMGTRHIKLNSMISIDGF
II-A_C.mitsuokai  1142 GIPLYLKQADIKEQIEYVEK.EEKLSDVKIIKNNIPLNQLIEIDGRQYLLTSPTECVNAM
II-A_L.rhamnosus  1146 TSQVAALKEARAQSTVKERELLKKFLTTKFTKIGKNGKKAITPFDIVIPRVPREQLFHNK
II-B_F.novicida   1358 MEFLRSLAYRSERVKIKSIDDVKQVLDKDSNFIIGKITLPFKKEWQRLYREWQNTTIKDD

II-A_S.pyogenes   1222 KGNELALPSKYVNFLYLASHYEKLKGSPED..............................
II-A_E.faecalis   1200 KGNQMVLPEHLLTLLYHAKQCLLPNQS................................
II-A_P.duerdenii  1215 AGSVQLIVNKEEQKLIKKMDKFLVK.......................KKDNKDIKVTS
II-A_F.nucleatum  1202 KNKKQLYLEKKYEQILKNALKFVEDNQGETEE..........NYKFIYLKKRNNNEKNET
II-A_A.intestini  1184 YLSIGGKSSKGKSVLCHAMVPLIVPHKIECYI..........KAMESFARKFKENNKLRI
II-A_C.mitsuokai  1201 QLVLNEEQCKLIADIYNAIYKQDFDG.................................
II-A_L.rhamnosus  1206 RYGFFMVNSDTLMHNYQELWMSRSDQQILS...........................KL
II-B_F.novicida   1418 YEFLKSFFNVKSITKLHHKKVRKDFSLPISTNEGKFLVKRKTWDNNFIYQILNDSDSRADG

II-A_S.pyogenes   1252 .NEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENII
II-A_E.faecalis   1227 ...ESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFI
II-A_P.duerdenii  1251 FDNIKE...EELIKLYKTLSDKLNNGIYSNKRRNQAKN.ISEALDKFKEISIEEKIDVLN
II-A_F.nucleatum  1252 IDAVKERYNIEFNEMYDKFLEKLSKDYKNYINNKLYTNFLNSKEKFKKLKLWEKSLILR
II-A_A.intestini  1234 VEKFDKITVEDNLNLYELFLQKLQHNPYN.KFFSTQFDVLTNGRSTFTKLSPEEQVQTLL
II-A_C.mitsuokai  1227 .........LDNMLMIQLYLQLIDKLKTLYPIYMGIVEKFEKLTEDFVSISKEEKANVIK
II-A_L.rhamnosus  1238 VKSAKTVENGQVGRLFENIVEQIN.KYFPLYDINQFRKKLDQSKDIFSKLPLEGNGTTSG
II-B_F.novicida   1478 TKPFIPAFDISKNEIVEAIIDSFTSKNIFWLPKNIELQKVDNKNIFAIDTSKWFEVETPS

II-A_S.pyogenes   1311 HLFTLTNLG.....................APAAFKYFDTTIDRKRYTSTKEVLDAT
II-A_E.faecalis   1284 QLMQFNAMG.....................APSTFKFFQKDIERARYTSIKEIFDAT
II-A_P.duerdenii  1307 QIILLFQS....................YNNGCNLKSIGLSAKTGVVFIPKKLNYKECK
II-A_F.nucleatum  1312 EFLKIFNKNTYGKY.............EIKDSQTKEKLFSFPEDTGRIRLGQSSLGNNKE
II-A_A.intestini  1293 NILSIFKTCRSSG...............CDLKSINGSAQAARIMISADLTGLSKKYSDIR
II-A_C.mitsuokai  1278 QMLIIMHKGPQNG................NIIYDDFNVGKRIGRLNGRTFYLDNIE
II-A_L.rhamnosus  1297 KRETIQNILTGAG.............TGAKNGDLKNLGIKTPFGFLQVPSGIVLDDDTT
II-B_F.novicida   1538 DLRDIGIATIQYKIDNNSRPKVRVKLDYVIDDDSKINYFMNHSLLKSRYPDKVLEILKQS

II-A_S.pyogenes   1347 LIHQSITGLYETRIDLSQLGGD..........
II-A_E.faecalis   1320 IIYQSTTGLYETRRKVVD..............
II-A_P.duerdenii  1346 LINQSITGLFENEVDLLNL.............
II-A_F.nucleatum  1359 LLEESVTGLFVKKIKL................
II-A_A.intestini  1338 LVEQSASGLFVSKSQNLLEYL...........
II-A_C.mitsuokai  1318 FISQSPTGIYTKKYKL................
II-A_L.rhamnosus  1343 LIYQSPTGMLNRIVRLSEL.............
II-B_F.novicida   1598 TIIEFESSGFNKTIKEMLGMKLAGIYNETSNN
```

57

Appendix 2. Sequence alignment of type II-B Cas9

```
                                        RuvC
F.novicida          1  ....MNFKILPIAIDLGVKNTGVFSAFYQKGTSLERLDNKNGKVYELSKDSYTLLMNNRT
G.proteobacterium   1  ...MTKNYISPIAIDLGAKFTGVALYQYLEG..ADCTQEVAKGLLVDDRGNVTWSQEGRR
S.wadsworthensis    1  MTQSERRFSCSIGIDMGAKYTGVFYALFDREE.LPTNLNSKAMTLVMPETGPRYVQAQRT
W.succinogenes      1  ......MLVSPISVDLGGKNTGFFSFTDSLD......NSQSGTVIYDES.FVLSQVGRR
B.thermophilus      1  ....MSKIISPIAIDMGAKNTGVYYAHYQRN...STFQEVDKKGQVLVYGNYTPLLVSRT
```

```
                      BH                              REC1
F.novicida         57  ARRHQRRGIDRKQLVKRLFKLIWTEQLNLEWDKDTQQAIS.......FLFNRRGFSFITD
G.proteobacterium  56  GKRHQVRGYKRRKMAKRLLWLILDSEYGIKREEVTEPLLKFIN....GLLNRRGYTYISE
S.wadsworthensis   60  AVRHRLRGQKRYTLARKLAFLVVDDMMKKQEKRLTDEEWKRGREALSGLLKRRGYSRPNA
W.succinogenes     47  SKRHSKRNNLRNKLVKRLFLLILQEHHGLSIDVLPDEIR.......GLFNKRGYTYAGF
B.thermophilus     54  ANRHTRRGYTRKKLAKRLLSVVLKEYFDFPAEKHTQALG.......FLMNRRGFSFLEE
```

```
                                        REC2
F.novicida        110  GYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYN..........
G.proteobacterium 112  EVD....EESMNVSPLPFSEMMPDYFNSSAPLLEQLAKLLSD................
S.wadsworthensis  120  DGEDLTPLENVRAEVFAAHPAFSTYFSEVRSLAEQWEEFTAN...............
W.succinogenes     99  ELDEKKKDALESDTLKEFLSEKLQSIDRDSDVEDFLNQIASN................
B.thermophilus    106  EYSKEYLNNLPNDSWEALSEEVQKMLGGQENIADRLMNLATTNPSEVTKLLEAVQNIKGY
```

```
                                        REC2
F.novicida        160  KLMQKILEFKLMKLCTDIKDDKVSTKTLKEITSYEFELLADYLANYSESLKTQKFSYTDK
G.proteobacterium 150  ..........KNKLVRFRA....EGKIPSNKNEFKKLLDTALDGKYKDEKKELSEAWGNI
S.wadsworthensis  162  ..........ISNVEMFLG....DSNIPADKEFVDFAVAEGLIDKT..EKKAYLSALSTM
W.succinogenes    141  ..........AESFKDYKKGFEAVFASATHSPNKKLELKDELKSEYGENAKELLAGLRVT
B.thermophilus    166  KDFQEEDKEIKADLVYFDYIEKISKACKLVKKGESIEDDKVTKSKKDKQNLSKTANWVVA
```

```
                                        REC2
F.novicida        220  QGNLKELSYYHHDKYNIQEFLKRHATINDRILDTLLTDDLDIWNFNFEKFDFDKNEE...
G.proteobacterium 196  LIASENVL................................................
S.wadsworthensis  206  RANANVLT................................................
W.succinogenes    191  KEILDEFD................................................
B.thermophilus    226  RLNNENALNLTVATESYQTNLIEEITQINYQQIQNNLPDFEKEREVIKQRKKDNNTSDWK
```

```
                      REC2                            REC1
F.novicida        277  ..........KLQNQEDKDHIQAHLHHFVFAVNKIKSEMASGGRHRSQYFQEVTNVLDEN
G.proteobacterium 204  .................................KSTVDGHKSRSEYLANIKEDIK..
S.wadsworthensis  214  ...............................GLRQMGHKPRSEYFKAIEANLK..
W.succinogenes    199  ................................KQENQGNLPRAKYFEELGEYIA..
B.thermophilus    286  FNYSEFKFNTPNEKGLEEDVARTHLHHFCYAVYKINNEIISGGRHRSKFFDEIKDDLDNF
```

```
                                        REC1
F.novicida        327  NHQEG......YLKNFCENLHNKKYSNLSVKNLVNLIGNLSNLELKPLRKYFNDKIHAKA
G.proteobacterium 226  ...........SNEELEKQIS.SKE.IDGFYNLVGHLSNFQLRLLRKYFNDPNMSGV
S.wadsworthensis  236  ............KDSRLAKICE.AFGGADRLARLLGNLSNLQLRAERWYFNAPDIMKD
W.succinogenes    221  .............TNEKVKSFFDSNSLKLTDMTKLIGNISNYQLKELRRYFNDKEMEKG
B.thermophilus    346  TTNSFLNNKHNPEYLKDFILAINENKGLDVDKLYKLVCHISNFELKPLRKYFNTEAYKSG
```

```
                                        REC1
F.novicida        381  DHWDEQKFTETYCHWILGEWRVGVKDQDKKDGAKYSYKDLCNELKQKVTKAG....LVDF
G.proteobacterium 270  SYWDEKRLEKYFYQWVQG.WHTKG.GTDEAEKKNIILKTKGAP.............LLKT
S.wadsworthensis  281  RGWEPDRFKKTLVRAFKF.FHPAK.DQNKQHLELIKQIENSED.............IIET
W.succinogenes    267  DIWIPNKLHKITERFVRS.WHPKN.DADRQRRAELMKDLKSKE.............IMEL
B.thermophilus    406  DKLDNKNLSKITSRWFLQHWIVTK.EKDKEAKVEDYEKLKKAWRAHQKCQDDNKNNIISF
```

```
                      REC1        Linker          REC3
F.novicida        437  LLELDPCRTIPPYLDNNNRKPPKCQSLILNPKFLDNQY.PNWQQYLQELKKLQSIQDYLD
G.proteobacterium 315  LKSLSADLTIPPYEDQNNRRPPKCQSVLLSDEKLTMHY.PKWKEWVGQLVKQN.......
S.wadsworthensis  326  LCTLDPNRTIPPYEDQNNRRPPLDQTLLLSPEKLTRQYGEIWKTWSARLTSDEPTLAPAA
W.succinogenes    312  LTTTEPVMTIPPYDDMNNRGAVKCQTLRLNEEYLDKHL.PNWRDIAKRLNHGKFN.....
B.thermophilus    465  WLKTDPKLTIPPYQAMLNRHPPKCQSLLLNADYLNKHY.PHWEDWLDDLNSNEEYQTKLQ
```

```
                                        REC3
F.novicida        496  SFETDLKVLKSSKDQPYFVEYKSSNQQIASGQRDYKDLDARILQFIFDRVKASDELLLNE
G.proteobacterium 367  ...................................DNAYLNENVTLANALHRIVERSRSIDPYQLRL
S.wadsworthensis  386  EILERSTDRKSR...........VAVNGHEPQPTLAYQLSYALQRAFDRSKALDPYALRA
W.succinogenes    366  ...................................DDLADSTVKGYSEDSTLLHRLLDTSKEIDIYELRG
B.thermophilus    524  SLQRGKGRRKAK...........NGKEIDGRLLDDEKIKTRQLQFILDTAKKQDFYKLNE
```

REC3

```
F.novicida         556  IYFQAKKLKQKASDELEKLESSKKLDEVIANSQLSQIIKSQHTNGIFEQGTFLHLVCKYY
G.proteobacterium  399  LISITDAEKRNDLAGYKRLKLSL.....................GSEVDEFLLLVKN.IV
S.wadsworthensis   435  LAAGSKSNKLTSAR.TALENCIG.....................GQNVEKFLDCARR.YY
W.succinogenes     401  KKPNELLVKTLGQSDANRLYGFA.................QNYYELIRQKVRAGIW
B.thermophilus     573  IWSVYHKLQKSDVDKIKWQEKLGNLIDASK.....LPKDLKQDLDFSQEESFGHFLNKYY
```

REC3

```
F.novicida         616  KQRQRARDSRLYIMPEYRYDKKLDKYKNTGRFDDNNQLLTYCNHKPRQKRYQLLNDLAGV
G.proteobacterium  437  DETKEAREGLWFET....................ENKLFFKCGKTPPRKEKLKSTLLSAV
S.wadsworthensis   472  READDAKVGLWFDN....................AEGLLERSDLHPPMKKKILPLLVANI
W.succinogenes     440  VPVKNKDDSLNLED....................NSNMLKRCNHNPPHKKNQIHNLVAGI
B.thermophilus     628  QTRRKARDGRYFLIQEK...........KEKWLTNNKLLILCTHKPRQKKHQWQLDLAAV
```

REC3

```
F.novicida         676  LQVSPNFLKDKIGSDDDLFISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIARNTK
G.proteobacterium  477  LGKNLSDDEQSSFIE...EFWKSGTPKIERRNVRGWCRLASQVQKTYGVYLKEYGLQQLH
S.wadsworthensis   512  LQTDETTG..QKFLD...EIWRKQIKGRETVASRCARIETVRKSFGGGFNIAYNTAQYRE
W.succinogenes     480  LGVKLDEAKFAEFEK...ELWSAKVGN.KKLSAYCKNIEELRKTHGNTFKIDIEELRKKD
B.thermophilus     677  LGVNADDLKAKIVGN...PED..YFKSIKGFASNCEKAAKAQKEHRGELKHKIYTELTSN
```

REC3

```
F.novicida         736  GKCEKEIFNLICKIEGSEDKKGNYKHGLAYELGVLLFGEPNEASKPEFDRKIKKFNSIYS
G.proteobacterium  534  ..............KLEAGKKLDDKPLALLYKNSGLIASKIGEALNIEPDEVSRFASPHS
S.wadsworthensis   567  ..............VNKLPRNAQDKELLIIRDRVAETADFIAANLGLSDEQKRKFANPFS
W.succinogenes     536  ..............PAELSKEEKAKLRLTDDVILNEWSQKIANFFDIDDKHRQRFNNLFS
B.thermophilus     732  ..............PIDKLGKLANKCSNLKQELIKEIKIPELLADKQTNEVNQKIALVFK
```

REC3

```
F.novicida         796  FAQIQQIAFAERKGNANTCAVCSADNAHRMQQIKITEPVEDN........KGNIILSTKA
G.proteobacterium  580  LAQIFNIIEGDVAGFNKTCRACTYENIWRMQEEKVESLLTNQLLSEIHGERKVPLKSAMC
S.wadsworthensis   613  LAQFYTLIETEVSGFSATTLAVHLENAWRMTIKDAVING.............ETVRAAQC
W.succinogenes     582  MAQLHTVIDTPRSGFSSTCKRCTAENRFRSETAFYNDETG..........EFHKKATATC
B.thermophilus     778  FAQIHNIVFKDRSGFSKTCPVCSTDNAFRMQEN....................DKGITKA
```

RuvC

```
F.novicida         848  QRLPAIPTRIIDGAVKKMATILAKNIVDDNWQNIKQVLSAKHQ.LHIPIITESNAFEFEP
G.proteobacterium  640  TRLSADSTRPFDGQMASIIEHIARKIAQHKIAQINDVPKEFS..IDIPIIIESNQFSFTA
S.wadsworthensis   660  SRLPAETARPFDGLVRRLVDRQAWEIAKRASTDIQSKVDFSNGIVDVSIFVEENKFEFSA
W.succinogenes     632  QRLPADTQRPFSGKIERYIDKLGYELAKIKAKELEGMEAKEIK...VPIILEQNAFEYEE
B.thermophilus     818  SRLPALSIRLIDGVVMRICDAIARQVAVTKWNDIKDDLQNGTK.VSVPLILEQNRFEFEP
```

L1 / HNH

```
F.novicida         907  ALADVKGKSLK........DKRKKVLERINSENIFKDKNNRIKEFAKGISAYSGDNLAS
G.proteobacterium  698  ELEEIKRGRGS........AKAKKAKELGEKSKAGWVSKTERIKTSSEGICPYTGAPLGG
S.wadsworthensis   720  SVADLKKKNK..........RVKDKMLSEAEKLETRWLNKNERIKKASRETCPYTGDRLAE
W.succinogenes     689  SLRKSKTGSNDRVINSKKDRDGKKLAKAKENAEDRLKDKDKRIKAFSSGICPYCGDTIGD
B.thermophilus     877  NLKKIKTAPKP............GVKAAEKADEIKDNRTNKYKGQKVSLYSGEGLGS
```

HNH

```
F.novicida         958  GDFDGAKEELDHIIPRS...HKKYGTLNDEANLICVTRGDNENKGNRIFCLRDLANNYKL
G.proteobacterium  750  ......SGEIDHIIPRSLTGRTKKTVFNSEANLIYCSSKGNHDKGNRVYVIEQLNDKYLK
S.wadsworthensis   770  ......GGEIDHILPRSLIKDARGIVFNAEPNLIYASSRGNQLKKNQRYSLSKLDPDYLN
W.succinogenes     749  ......DGEIDHILPRSHTLKIYGTVFNPEGNLIYVHQKCNQAKADSIYKLSDIKAGVSA
B.thermophilus     922  ......DGEIDHIIPR.....TGPYGILNDEANLIYISKSDNQNKGNDIKFLKDLHKNYKT
```

HNH

```
F.novicida         1015 KQFETTDDLKIEKKIADTIWDANKKD.FKFGNYRSFINLTPQEQKAFRHALFLADENPIK
G.proteobacterium  804  KQFSTSDVNLIKKKIKTTIQRFTEGG....EKLRSFSELSREDQKAFRHALFVP...ELK
S.wadsworthensis   824  KVFKTSNIAAITAEIEDVVTKLQQT.....HRLKFFDLLNEHEQDCVRHALFLDDGSEAR
W.succinogenes     803  QWIEEQVANIKG...............YKTFSVLSAEQQKAFRYALFLQNDNEAY
B.thermophilus     972  EIFGKKSDAEIKEFIYKRLGKDIESKDFAFDKYLSFINLGEKDKKAFRHALFLQEGDPLR
```

L2 / RuvC

```
F.novicida         1074 QAVIRAINNRNRTFVNGTQRYFAEVLANNIYLRAKKENLN.TDKISFDYFGIPTIGNGRG
G.proteobacterium  857  SEVTSLLAVKNITRVNGTQAWLAKKIASLLAEHLDKQGRDYT....LSAHQIDPWSVSK.
S.wadsworthensis   879  NAVLELLATQRRTRVNGTQIWMIKSLANKIREELQDWCRTTNNRLHFQAAATDVSTAKN.
W.succinogenes     843  KKVVDWLRTDQSARVNGTQKYLAKKIQEKLTKMLP......NKHLSFEFILADATEVSE.
B.thermophilus     1032 EKVIRALNNRNRAIVNGTQRYLAQCIADKIHRIAKKENK..HNLIEFDYLEYTARWDDP.
```

Alignment block 1 — RuvC

```
                            RuvC
F.novicida         1133  IAEIRQLYEKVDSDIQAYAKGDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEIDKN..
G.proteobacterium   912  .......QRKMLASAEPIWAKKDPQPAASHVVDAVCTFLEALEQPHTASRLK........
S.wadsworthensis    938  .......LRLKLAQNQPNFEKPDIQPIASHSIDALCSFAVG.....SADAER.......
W.succinogenes      896  .......LRRQYARQNPLLAKAEKQAPSSHAIDAVMAFVARYQKVFKDGTPPNADEVAKL
B.thermophilus     1089  .......KSTYNLRKKYDLVKSKKQPLYSHLIDAQLAFLLASEDHQNDGTMGIKFDNNQT
```

Alignment block 2 — RuvC / PLL

```
                            RuvC                                PLL
F.novicida         1191  YSLYPLDKNTGEVFTKDIFS................QIKITDNEFSDKKLVRKKATEGFN
G.proteobacterium   957  ....TISSTSFEKTGWRSA.................LIPDLIKVDALDRRPKYRRYNI
S.wadsworthensis    978  ....DQDGFDYLDGKAVLG................LYPQSCEVIRLQAKPQEEKSHF
W.succinogenes      949  AMLDSWNPASNEPLTKGLS................TNQKIEKMIKSGDYGQKNMREV
B.thermophilus     1142  IWEYETNKETGEISPSKSFNAIDVGESDLKKIDLSPKDSNQKIIDIEKNNSNKKQNLSKI
```

Alignment block 3 — PLL / WED

```
       PLL                          WED
F.novicida         1235  THRQMTRDGIYAENYLPILIHKELDEVRKGYTWKNSE........EIKIFKGKKYDIQQL
G.proteobacterium   994  GSTSLFKDGIYAERFLPILIDENGLMAGYDIDNSL.........KAKGADVVFESLSP.
S.wadsworthensis   1015  DSVAIFKEGIYAEQFLPIFTLNGRIWIGYETLDSK.........GERCGAIEVSGKKPE
W.succinogenes      990  FGKSIFGENAIGERYKPIVVQEGGYYIGYPATVKKGY........ELKNCKVVTSKNDIA
B.thermophilus     1202  FKRRMFKANAIGERYKPIVKFKDKWYLGYPMTIKGGIYNCDEYCQEVASKNDIKKIEDVV
```

Alignment block 4 — WED

```
                            WED
F.novicida         1287  NNLLYCLKFVDKPISIDIQISTLEELRNILTTNNIAATAEYYYINLKTQKLHEYYIENYN
G.proteobacterium  1043  FLLFKGEEVGAQSLSDWQERIDGRYLYMSIDKVKAFDYLQEKVGEKDIAAELLNSIHFTQ
S.wadsworthensis   1065  ELLSMLAPFFNKPVGDLSAHATYRIQKKPAYEFLAKAALQPLSAEEKRLAALLDALRYCT
W.succinogenes     1042  KLEKIIKNQDLISLKENQYIKIFSINKQTISELSNRYFNMNYKNLVERDKEIVGLLEFIV
B.thermophilus     1262  NDDKYYQLTTDHKQIKIWTIKIVDKKYKKINKDSHRYFSQLNPEYNNDEKKEVDQIKFIL
```

Alignment block 5 — WED

```
                            WED
F.novicida         1347  TALGYKKYSKEMEFLRSLAYRSERVKIKSIDDVKQVLDK...............DSNFII
G.proteobacterium  1103  RKTELRAKFSDDSGKKMKTLDAIRKSLKLTVTVNEIGKRKEKCGFSG......TIGIPAK
S.wadsworthensis   1125  SRKSLKSFFMAANGKSLKKREDLLKPKLFQLKVELKGEK..AFKLNG......SLTLPMK
W.succinogenes     1102  ENCRYYTKKVDVKFAPKYIHETKYPFYDDWRRFDEAWRYLQENQNKTSSKDRFVIDKSSL
B.thermophilus     1322  DKCRYYVAKTDVINAPKIFDKQGYPFYKNWVDFDNAWRK.................EIG
```

Alignment block 6 — WED

```
                            WED
F.novicida         1392  GKITLPFKKEWQRLYLEWQNTTIKDDYEFLKSFFNVKSITKLHKKVRKDFSLPISTNEGK
G.proteobacterium  1157  SAWENLLDEPLLETYWGTKMPPQEIWEKVYRKHFPRNIPNQAHRKVRKDFSLPVVDSVSG
S.wadsworthensis   1177  QDWLRICNSPELADAFGNPCSADELTSKLAQIWKRPVMRDLAHAPVRREFSLPVIDNPSG
W.succinogenes     1162  NEYYQPDKNEYKLDVDTQPIWDDFCRWYFLDRYKTANDKKSIRIKARKTFSLLAESGVQG
B.thermophilus     1364  DDYSVKTGEYDLSNNTTKKKWDIFCKNRFS...PPENQHKNKHQVKGKKYTMISSGTPSG
```

Alignment block 7 — WED / Linker / PI

```
                  WED            Linker              PI
F.novicida         1452  ..FLVKRKTWDNNFIYQILNDSDSRVDGTKPFIPAFDISKNEIVEAIIDSFTS...KNIF
G.proteobacterium  1217  .GFRVKRKTPNG.YNYQLLAIDGYSAVGFKKEGDNVDFKSPALVPQIAESKSVTPISSEL
S.wadsworthensis   1237  .GFRIRRTNLFGNELYQVHAINAKKYRGFASADSNVDWANGILFNELQHEN...LTECCG
W.succinogenes     1222  KVFRAKRKIPTG.YAYQALPMDNNVIAGDYANILLEANSKTLSLVPKSGISIEKQLDKKL
B.thermophilus     1421  NVFRVKRNNQNI...YQALPLDNNIILKKKSNFLIKHSKNLTLSSAVADKDLAKPIDIEE
```

Alignment block 8 — PI

```
                            PI
F.novicida         1507  WLPKNIELQKVDNKNIFAIDTSKWFEVETPSDLRDIGIATIQYKIDNNSRPKVRVKLDYV
G.proteobacterium  1275  VHLDKNEIVYFDEWRKIDISDSDLKQFVSSLELAPGSQNRFYIRFTVDEDQFERHFKSAL
S.wadsworthensis   1293  RFITSADVTPMSEWRKVVTEDN......LRIWIAPGTEGRRYVRVETTFIQASHWFEQSV
W.succinogenes     1281  DVIKKTDVRGLAIDNNSFFNADFDTHGIRLIVENTSVKVGNFPISAIDKSAKRMIFRALF
B.thermophilus     1478  KIELKNCVIPANNFFKDELKGVKVFLNNTSVDIKDFPLDKFREYFNKEIDVKKSIKIEIL
```

Alignment block 9 — PI

```
                            PI
F.novicida         1567  IDDDSKINYFMNHSLLKSRYP.........DKVLETLKQSTTIEFESSGFNKTIKEMLGM
G.proteobacterium  1335  RVNGIQDLDTVNKTFDWNR..........EIPSLLIPPRSNLFLLETG..QKITFEYIAN
S.wadsworthensis   1347  ENWAITSPLSLPASFKVDKPAEFQKAVGTELSELLGQPRSEIFIESVGNAKHIRFWYIVV
W.succinogenes     1341  EKEKGKRKKKTTISFKESGP.......VQDYLKVFLKKIVKIQLRTDGSISNIVVRKNAA
B.thermophilus     1538  TEEQAEIDLTENPNLLYAHK..........CKVDASIRYTTRDGSVTNIRVGVKERLVS
```

Alignment block 10 — PI

```
              PI
F.novicida         1618  TLAGIYNETSNN.....
G.proteobacterium  1383  GANAEVKKAYSLRRA..
S.wadsworthensis   1407  SSNKKMNESYNNASKS.
W.succinogenes     1394  DFTLSFRSEHIQKLLK.
B.thermophilus     1587  FSLPFRSCEKIYKDSQG
```

Original paper

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

Acknowledgements