

Crystal structures of CRISPR RNA-guided Cas9 and Cpf1 nucleases

Bacteria and Archaea utilize the CRISPR-Cas adaptive immune system for defense against invading genetic elements such as phages and plasmids. The CRISPR arrays in the microbial genome comprise direct repeats separated by variable spacer sequences derived from foreign genetic elements. The CRISPR arrays are transcribed and then processed into small CRISPR RNAs (crRNAs), which bind to Cas endonucleases to form effector ribonucleoprotein complexes. The Cas effector complexes cleave the foreign nucleic acids complementary to the crRNA guide. Based on the architecture of the effector module, the CRISPR-Cas systems can be divided into two classes. Class 1 systems use effector complexes comprising multiple Cas proteins, whereas class 2 systems employ a single effector nuclease, such as Cas9 (type II) or Cpf1 (type V). Cas9 uses dual RNA guides (a crRNA and a *trans*-activating crRNA) or a chimeric single-guide RNA (sgRNA), and cleaves a double-stranded DNA target complementary to the crRNA guide (Fig. 1(a)). In contrast, Cpf1 is guided by a single crRNA and cleaves a double-stranded DNA target (Fig. 1(b)). In addition to the RNA-DNA complementarity, Cas9 and Cpf1 require a short nucleotide sequence adjacent to the target site, called a protospacer adjacent motif (PAM). Cas9 and Cpf1 have been harnessed for a variety of new technologies, exemplified by genome editing in various cell types and organisms. Previous structural studies of Cas9 from *Streptococcus pyogenes* (SpCas9) and *Staphylococcus aureus* (SaCas9) provided mechanistic insights into RNA-guided DNA cleavage by Cas9 and enhanced molecular engineering of Cas9 variants with improved functionality [1,2]. Furthermore, a structural comparison between the two Cas9 orthologs revealed

both divergent and convergent structural features among orthologous CRISPR-Cas9 systems.

Cas9 orthologs have diverse amino-acid sequences and recognize distinct guide RNA and PAM sequences. Cas9 from *Francisella novicida* (FnCas9), one of the largest Cas9 orthologs, consists of 1629 residues and is significantly larger than SpCas9 (1368 residues) and SaCas9 (1053 residues). To elucidate the RNA-guided DNA cleavage mechanism of FnCas9, we crystallized FnCas9 in a complex with the sgRNA and its DNA targets containing either TGG or TGA as the PAM, collected X-ray diffraction data at SPRING-8 BL41XU, and determined the structures at 1.7 Å resolution [3] (Fig. 2(a)). The structures revealed that FnCas9 recognizes the guide RNA-target DNA heteroduplex within the central channel between the REC and RuvC domains. The HNH and RuvC endonuclease domains are located at positions suitable for cleaving the target and non-target DNA strands, respectively. The PAM DNA duplex is bound to the groove between the WED and PI domains, and the TGR PAM is recognized by Arg1556 and Arg1585 in the PI domain. A structural comparison of FnCas9 with SpCas9 and SaCas9 revealed striking conserved and divergent features among the orthologous CRISPR-Cas9 systems. Furthermore, we used the structural information to engineer a variant that can recognize the altered YG PAM. The FnCas9 variant modified endogenous target sites with the YG PAM in mouse zygotes, thereby expanding the target space in Cas9-mediated genome editing.

Using molecular evolution, SpCas9 has been engineered to exhibit altered PAM specificities. Whereas wild-type SpCas9 recognizes the NGG PAM, the engineered VQR (D1135V/R1335Q/T1337R), EQR (D1135E/R1335Q/T1337R), and VRER (D1135V/G1218R/R1335E/T1337R) variants recognize the NGA, NGAG, and NGCG PAMs, respectively. These SpCas9 variants thus contribute to expanding the target space in Cas9-mediated genome editing applications. To elucidate the altered PAM recognition mechanisms of the SpCas9 variants, we crystallized the three SpCas9 variants in complexes with the sgRNA and its PAM-containing DNA targets, collected X-ray diffraction data at BL41XU, and determined their structures at 2.0–2.2 Å resolutions [4] (Fig. 2(b)). Whereas the third G in the PAM is recognized by Arg1335 in wild-type SpCas9, the third A in the PAM is recognized by R1335A in the VQR and EQR variants. The third C in the PAM is recognized by R1335E in the VRER

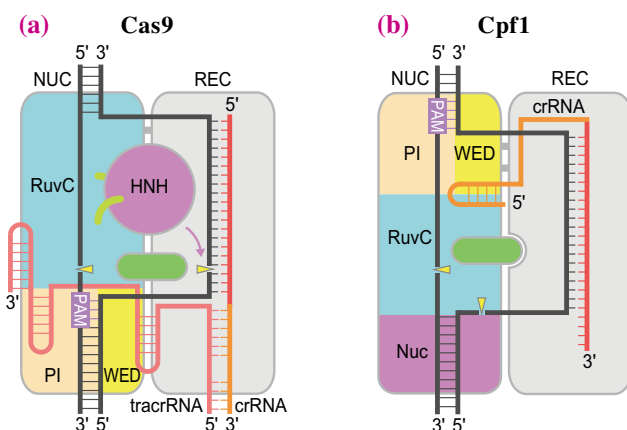


Fig. 1. RNA-guided DNA cleavage by Cas9 (a) and Cpf1 (b).

variant. A structural comparison of the three SpCas9 variants with wild-type SpCas9 revealed that the multiple mutations induce an unexpected structural displacement in the sugar-phosphate backbone of the PAM duplex, thereby enabling direct base-specific hydrogen-bonding interactions between the PAM-interacting residue and the altered PAM nucleotides. Our findings explain the altered PAM specificities of the SpCas9 variants and establish a framework for the further engineering of CRISPR-Cas9.

Since, except for the RuvC domain, Cpf1 shares no sequence similarity with other proteins, the crRNA-guided DNA cleavage mechanism of Cpf1 remains unknown. We crystallized *Acidaminococcus sp.* Cpf1 (AsCpf1) in a complex with the crRNA and its target DNA, collected X-ray diffraction data at BL41XU, and determined the structure at 2.8 Å resolution [5] (Fig. 2(c)). The structure revealed that AsCpf1 adopts a bilobed architecture consisting of the REC and NUC lobes. The REC lobe can be divided into the REC1 and REC2 domains, whereas the NUC lobe comprises the RuvC, WED, PI, and Nuc domains. Except for the RuvC domain, the five domains have new protein folds. The crRNA 5' handle adopts an unexpected pseudoknot and is recognized by the WED and RuvC domains.

In contrast, the crRNA guide segment and the target DNA strand form the RNA-DNA heteroduplex, which is accommodated within the central channel between the REC and NUC lobes. The PAM DNA duplex adopts a distorted double-stranded conformation and is recognized by the WED-REC1-PI domains in base- and shape-dependent manners, consistent with the recognition of TTTV PAM by AsCpf1. The RuvC domain is located at a position suitable for cleaving the non-target DNA strands. Notably, the Nuc domain is located adjacent to the RuvC domain, and mutational analysis revealed the involvement of the Nuc domain in the cleavage of the target DNA strand. These findings indicated that the RuvC and Nuc domains jointly participate in generating DNA double-strand breaks. A structural comparison of AsCpf1 with Cas9 reveals both striking similarity and major differences, thereby explaining their distinct functionalities. Although except for the RuvC domain, Cpf1 and Cas9 share no sequence similarity, they use similar bilobed architectures to recognize the RNA-DNA heteroduplex. In contrast, Cpf1 and Cas9 recognize their PAM sequence and cleave the DNA target in distinct manners. These structural findings revealed an intriguing functional convergence between the two class 2 CRISPR-Cas nucleases.

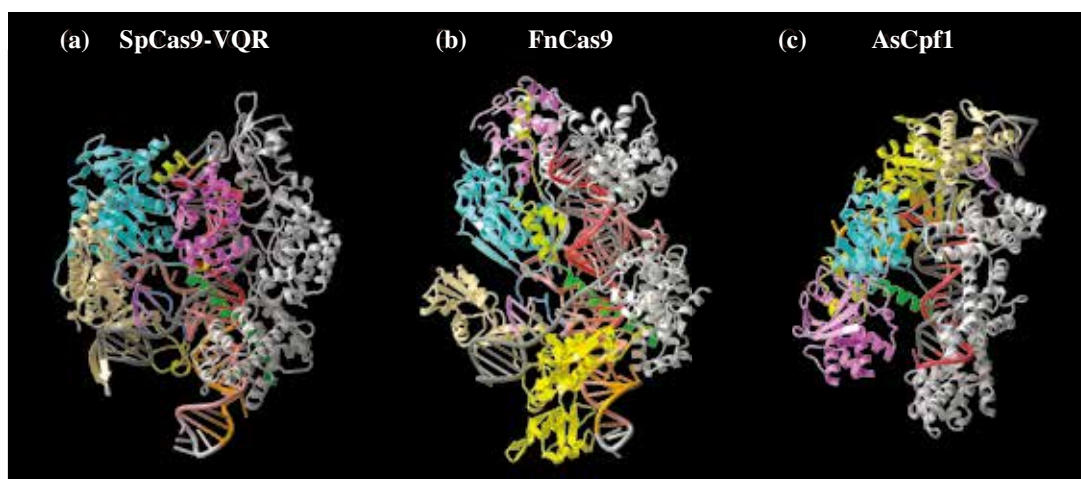


Fig. 2. (a–c) Crystal structures of FnCas9-sgRNA-DNA (a), SpCas9 VQR-sgRNA-DNA (b), and AsCpf1-crRNA-DNA (c). Structural images were prepared using CueMol (<http://www.cuemol.org>).

Hiroshi Nishimasu^{a,b,*} and Osamu Nureki^a

^aDepartment of Biological Sciences,
The University of Tokyo

^bPRESTO / JST

*Email: nisimasu@bs.s.u-tokyo.ac.jp

References

- [1] H. Nishimasu *et al.*: *Cell* **156** (2014) 935.
- [2] H. Nishimasu *et al.*: *Cell* **162** (2015) 1113.
- [3] H. Hirano, J.S. Gootenberg, T. Horii, O.O. Abudayyeh, M. Kimura, P.D. Hsu, T. Nakane, R. Ishitani, I. Hatada, F. Zhang, H. Nishimasu and O. Nureki: *Cell* **164** (2016) 950.
- [4] S. Hirano, H. Nishimasu, R. Ishitani, O. Nureki: *Mol. Cell* **61** (2016) 886.
- [5] T. Yamano, H. Nishimasu, B. Zetsche, H. Hirano, I.M. Slaymaker, Y. Li, I. Fedorova, T. Nakane, K.S. Makarova, E.V. Koonin, R. Ishitani, F. Zhang and O. Nureki: *Cell* **165** (2016) 949.