

学位論文

Crystal structure of the minimal Cas9 and generation of mutants
with enhanced catalytic activity

(最小 Cas9 の結晶構造解析と活性向上変異体の創出)

平成 29 年 12 月博士（理学）申請

東京大学大学院理学系研究科

生物科学専攻

山田 真理

Abstract

The RNA-guided DNA endonuclease Cas9 derived from the microbial type II CRISPR-Cas system associates with dual guide RNAs, crRNA (CRISPR RNA) and tracrRNA (trans-activating crRNA), or a synthetic sgRNA (single-guide RNA), and generates a DNA double-strand break at DNA target sites. The CRISPR-Cas9 system has been harnessed for a variety of new technologies, exemplified by genome editing. Cas9-mediated DNA targeting requires a PAM (protospacer adjacent motif), a short nucleotide sequence adjacent to the target site. Cas9 proteins and their cognate guide RNAs from different microbial species are remarkably divergent in sequences and lengths. Crystal structures of Cas9 orthologs, such as *Streptococcus pyogenes* Cas9 (SpCas9, 1,368 aa) (Nishimasu et al., Cell 2014; Anders et al., Nature 2014), *Staphylococcus aureus* Cas9 (SaCas9, 1,053 aa) (Nishimasu et al., Cell 2015) and *Francisella novicida* Cas9 (FnCas9, 1,629 aa) (Hirano et al., Cell 2016), provided mechanistic insights into the RNA-guided DNA cleavage mechanism of Cas9, and facilitated engineering of CRISPR-Cas9 to create new tools, such as Cas9 variants with enhanced fidelity (Slaymaker et al., Science 2015; Kleinstiver et al., Nature 2016) or altered PAM specificities (Kleinstiver et al., Nature 2015; Kleinstiver et al., Nat Biotech 2016), and a potent transcriptional activator (Konermann et al., Nature 2015). However, the action mechanisms of divergent Cas9 orthologs have not been fully understood.

Here, the author reports the crystal structures of *Campylobacter jejuni* Cas9 (CjCas9, 984 aa), one of the smallest Cas9 orthologs, in complex with sgRNA and its target DNA. A structural comparison of CjCas9 with the other Cas9 orthologs provided insights into a minimal functional Cas9 scaffold. Furthermore, the CjCas9 structure revealed an unexpected diversity in orthologous CRISPR-Cas9 systems, such as the structural diversity in tracrRNA scaffolds and the mechanistic diversity in PAM recognition. Whereas other characterized tracrRNA scaffolds contain several stem loops, the CjCas9 tracrRNA scaffold contains a triple-helix architecture, which is distinct from other known RNA triple helices, such as the telomerase RNA subunit TER, the SAM-II riboswitch and long noncoding RNA MALAT1. Thus, these structural findings highlight the structural diversity of the tracrRNA scaffold, and expand the natural repertoire of RNA triple helices. In addition, the previously characterized Cas9 orthologs “read” nucleotide sequences in the non-target DNA strand within the PAM-containing duplex to achieve PAM recognition (for instance, SpCas9 recognizes the 5'-NGG-3' in the non-target DNA strand). In stark contrast, our structural and functional analyses revealed that CjCas9 recognizes the promiscuous 5'-NNNVR YM-3' PAM by reading nucleotide sequences not only in the non-target strand, but also in the

target strand. The crystal structure of CjCas9 revealed an unexpected diversity of recognition mechanism of guide RNA and PAM in CRISPR-Cas9 system.

CjCas9 has been thought to be unavailable for genome editing due to weak DNA cleavage activity in eukaryotic cells (Ran et al., 2015). However, very recently, it was reported that genome editing using CjCas9 was successful by examining the sequence of the guide RNA and PAM (Kim et al., 2017). Since small CjCas9 has advantages such as high efficiency of introduction into viral vectors, it is expected to be applied to genome editing. However, CjCas9 has weak DNA cleavage activity in eukaryotic cells because cleavage efficiency was lower than that of SpCas9. From this background, the author conducted a search for activity enhancing mutants and succeeded in creating mutants with twice as much activity intensity as WT CjCas9.

Overall, these findings improved our mechanistic understanding of the CRISPR-Cas9 systems, and provided a framework for further Cas9 engineering for its applications.

Contents

Abstract	i
Contents	iv
Table of Abbreviations	viii
Table of Amino Acid Abbreviations.....	ix
Table of Nucleic Acid Base Abbreviations	x
Chapter 1 General Introduction.....	1
1.1 CRISPR-Cas system	1
1.2 Cas9	3
1.3 Development of application tool using Cas9	4
1.4 Mechanistic details of SpCas9	4
1.5 Structural studies of Cas9 orthologs	5
1.6 Correlation between the evolution and the cleavage activity of Cas9 orthologs....	6
1.7 Purpose of this study	7
Figures and Tables of Chapter 1	9
Figure 1-1 Schematic diagram of class 2 CRISPR-Cas system	9
Figure 1-2 Schematic diagram of class 2 CRISPR-Cas system	10
Figure 1-3 Cas9 targeting using crRNA-tracrRNA or a single-guide RNA (sgRNA). ..	11
Figure 1-4 Engineering nuclease-deactivated Cas9 (dCas9)	12
Figure 1-5 Conformation change of Cas9 in DNA targeting	13
Figure 1-6 Structures of known Cas9 orthologs, their sgRNA and PAM sequences .	14
Table 1-1 Summary of the available structures of Cas9.....	15
Chapter 2 PAM specificity of <i>Campylobacter jejuni</i> Cas9	16
2.1 Introduction.....	16
2.2 <i>In vivo</i> PAM discovery assay.....	16
2.2.1 Materials and Methods.....	16
2.2.2 Results	17
2.3 <i>In vitro</i> PAM discovery assay.....	17
2.3.1 Materials and Methods.....	18
2.3.2 Results	18
Figures of Chapter 2	19
Figure 2-1 CjCas9 PAM Specificity.....	19
Chapter 3 X-ray crystallographic analysis of <i>Campylobacter jejuni</i> Cas9	20
3.1 Introduction.....	20

3.2 Materials and Methods.....	20
3.2.1 Plasmid construction	20
3.2.2 CjCas9-ΔHNH protein expression and purification.....	21
3.2.3 <i>In vitro</i> transcription and purification of sgRNA	22
3.2.4 DNA preparation	22
3.2.5 Complex reconstruction.....	22
3.2.6 Crystallization of the CjCas9-ΔHNH-sgRNA-target DNA complex	22
3.2.7 Expression and purification of the SeMet-substituted CjCas9-ΔHNH protein	23
3.2.8 Crystallization of the SeMet-substituted CjCas9-ΔHNH-sgRNA-target DNA complex.....	24
3.2.9 X-ray diffraction analysis and data processing.....	24
3.2.10 Model building and structure refinement	24
3.2.11 Expression and purification of WT CjCas9 protein	25
3.3 Results	25
3.3.1 Expression, purification and construct optimization.....	25
3.3.2 Complex reconstruction.....	26
3.3.3 Crystallization	26
3.3.4 X-ray diffraction analysis.....	26
Figures and Tables of Chapter 3.....	28
Figure 3-1 Schematic diagram of the CjCas9 expression constructs.....	28
Figure 3-2 Size-exclusion chromatograms and SDS-PAGE	29
Figure 3-3 Purification of the CjCas9 quaternary complex.....	30
Figure 3-4 Crystals of CjCas9-sgRNA-target DNA.....	31
Figure 3-5 Schematic diagram of the Cj sgRNA and target DNA constructs for crystallization.....	32
Table 3-1 Macromolecule-production information	33
Table 3-2 Oligonucleotides	38
Table 3-3 Buffer composition	40
Table 3-4 Data collection and refinement statistics	41
Table 3-5 Reagents or resources information.....	42
Chapter 4 Crystal structures of the minimal Cas9 from <i>Campylobacter jejuni</i>	45
4.1 Introduction.....	45
4.2 Materials and Methods.....	45
4.3 Results and Discussion.....	45
4.3.1 Crystal structure of the CjCas9-sgRNA-target DNA complex	45

4.3.2 The effect of the deletion of the HNH domain on the CjCas9 activity.....	47
4.3.3 TracrRNA architecture.....	48
4.3.4 TracrRNA scaffold recognition.....	48
4.3.5 The 5' -NNNVRYM-3' PAM recognition.....	49
4.3.6 Structural comparison between the Cas9 orthologs	52
4.3.7 Mechanistic diversity in PAM recognition	54
4.3.8 Structural comparison of the CjCas9 tracrRNA and other known RNA Triplexes	56
4.4 Discussion and Perspectives	57
Figures and Tables of Chapter 4.....	58
Figure 4-1 Overall structures.....	58
Figure 4-2 <i>In vitro</i> cleavage activity of CjCas9-ΔHNH.....	59
Figure 4-3 Schematic of the nucleic-acid recognition	60
Figure 4-4 tracrRNA structure	61
Figure 4-5 TracrRNA scaffold recognition.....	62
Figure 4-6 PAM recognition	63
Figure 4-7 PAM recognition	64
Figure 4-8 Structural comparison of the Cas9 orthologs.....	65
Figure 4-9 PAM recognition by the Cas9 orthologs	66
Figure 4-10 RNA-guided DNA targeting.....	67
Figure 4-12 Structures of the RNA triple helices	69
Chapter 5 Generation of mutants with enhanced catalytic activity of CjCas9.....	70
5.1 Introduction.....	70
5.1.1 Comparison of the Cleavage Activities of CjCas9 and SpCas9	70
5.1.2 Genome editing using CjCas9	71
5.2 Materials and Methods.....	72
5.2.1 Preparation of mutants	72
5.2.2 <i>In vitro</i> DNA cleavage assay	72
5.2.3 Purification of mut1/mut3/mut4 mutant.....	73
5.2.4 <i>In vitro</i> DNA cleavage assay by MultiNA	74
5.3 Results and Discussion.....	74
5.3.1 Discovery of activity enhancement mutants	74
5.3.2 Comparison of the PAM preference between WT and mut1 / mut3 / mut4 mutant.....	76
5.3.3 Purification of the mut1/mut3/mut4 mutant	76
5.3.4 Comparison of the activity between WT and mut1/mut3/mut4 mutant	77

5.3.5 Comparison of the DNA cleavage activities among the different PAM sequence	77
5.4 Discussion and Perspectives	78
Figures of Chapter 5	80
Figure 5-1 Comparison of the DNA cleavage activities of CjCas9 and SpCas9	80
Figure 5-2 Discovery of the activity enhancement mutants	81
Figure 5-3 Comparison of the DNA cleavage activities of WT and mutants.....	82
Figure 5-4 Comparison of the PAM preference between WT and the mut1/mut3/mut4 mutant.....	83
Figure 5-5 Purification of the CjCas9 mutant (mut1/mut3/mut4).....	84
Figure 5-6 Comparison of the cleavage activities between WT and mutant (mut1/mut3/mut4).....	85
Figure 5-7 Comparison of the cleavage activities for the different PAM sequences..	86
Chapter 6 General Discussion.....	87
6.1 Summary of this study	87
6.2 Future prospects	88
References	91
Original Paper.....	96
Conference presentations and awards.....	97
Acknowledgement	98

Table of Abbreviations

Abbreviation	Full name
β-ME	2-mercaptoethanol
Cas9	CRISPR-associated protein9
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DNA	deoxyribonucleic acid
DTT	dithiothreitol
IPTG	isopropyl-β-D-thiogalactopyranoside
LB	Luria Bertani
MWCO	molecular weight cut off
Ni-NTA	Nickel-nitrilotriacetic acid
RNA	ribonucleic acid
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
SDS-PAGE	Sodium laurylsulfate-polyacrylamide gel electrophoresis
tracrRNA	trans-activating-crRNA
Tris	Tris (hydroxymethyl) aminomethane
TEV	Tobacco etch virus

Table of Amino Acid Abbreviations

Abbreviation Full name

A, Ala	alanine
C, Cys	cysteine
D, Asp	aspartic acid
E, Glu	glutamic acid
F, Phe	phenylalanine
G, Gly	glycine
H, His	histidine
I, Ile	isoleucine
K, Lys	lysine
L, Leu	leucine
M, Met	methionine
N, Asn	asparagine
P, Pro	proline
Q, Gln	glutamine
R, Arg	arginine
S, Ser	serine
T, Thr	threonine
V, Val	valine
W, Trp	tryptophan
Y, Tyr	tyrosine

Table of Nucleic Acid Base Abbreviations

Abbreviation	Full name
A	adenine
T	thymine
G	guanine
C	cytosine
U	uracil

Chapter 1 General Introduction

1.1 CRISPR-Cas system

Bacteria and Archaea utilize CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated) adaptive immune systems to defend themselves against foreign genetic elements, such as plasmids and phages (Marraffini, 2015; Barrangou and Doudna, 2016; Mohanraju et al., 2016; Wright et al., 2016). The CRISPR loci in the genome comprise a *cas* operon and a CRISPR array, consisting of short repetitive sequences (direct repeats) separated by non-repetitive sequences (spacers) derived from foreign genetic elements. The CRISPR array is transcribed and processed into CRISPR RNAs (crRNAs), which associate with single or multiple Cas proteins to form effector ribonucleoprotein complexes responsible for the destruction of invading nucleic acids (Makarova et al., 2015; Nishimasu and Nureki, 2016).

The CRISPR-Cas system is roughly divided into three stages: (1) Adaptation stage (2) Expression stage and (3) Interference stage.

(1) Adaptation stage: It is a stage to acquire a new spacer. The part of the exogenous DNA that has invaded from the outside via phage is cut out and incorporated as a new spacer sequence in the CRISPR region. The sequence of foreign DNA incorporated in the CRISPR region is called a protospacer sequence (Mojica et al., 2009; Yosef et al., 2012; Datsenko et al., 2012).

(2) Expression stage: It is the stage of generating crRNA. CRISPR region is transcribed and Pre-CRISPR RNA (Pre-crRNA) is transcribed. The repeat region of Pre-crRNA is cleaved via processing and crRNA of spacer unit is generated.

(3) Interference stage: It is the stage of decomposing the target nucleic acid. The crRNA binds to certain Cas nuclease and forms a complex (Cas-crRNA complex). Upon a second infection, the crRNA in the Cas-crRNA complex binds complementarily to the target DNA. Then Cas nuclease in the complex recognizes and cleaves the target DNA (Deltcheva et al., 2011; Marraffini et al., 2008) (Figure 1-1).

The CRISPR-Cas system is classified into two classes based on the structure of the effector complex. In the class 1 CRISPR-Cas system, a multi subunit complex composed of multiple Cas protein is involved (Figure 1-1). In contrast, class 2 CRISPR-Cas system involves a single Cas protein. Class 1 is classified into type I, type III, type IV, whereas class 2 is classified into types II and V, which target DNA, and type VI, which targets RNA (Figure 1-1). The type V is further divided into four subtypes (V-A, V-B, V-C and V-U). In the class 2, each type of effector protein is Cas9 (types II), C2C1 (type V-B), Cpf1 (type V-A), and C2C2 (type VI), respectively (Figure 1-2).

1.2 Cas9

In the type II CRISPR-Cas system, the Cas9 effector nuclease associates with dual guide RNAs (crRNA and *trans*-activating crRNA (tracrRNA)), and cleaves double-stranded (ds) DNA targets complementary to the crRNA guide (Garneau et al., 2010; Deltcheva et al., 2011; Gasiunas et al., 2012; Jinek et al., 2012). In addition to the crRNA-target DNA complementarity, Cas9-mediated target recognition requires a PAM (protospacer adjacent motif), a short nucleotide sequence adjacent to the target site (Deveau et al., 2008; Garneau et al., 2010). Importantly, a single-guide RNA (sgRNA), in which crRNA and tracrRNA are fused with an artificial tetraloop, can also direct Cas9 to the target cleavage (Jinek et al., 2012). Thus, the two component Cas9-sgRNA system has been harnessed for a variety of new technologies, including genome editing (Cong et al., 2013; Jinek et al., 2013; Mali et al., 2013) (Figure 1-3).

Cas9 is composed of five domains: REC, HNH, RuvC, WED, PI. The guide RNA is recognized by REC and WED, whereas PAM is recognized by PI. The target DNA strand is cleaved by HNH, and the non target DNA strand is cleaved by RuvC (Figure 1-3).

1.3 Development of application tool using Cas9

In recent years, a method of using Cas9 as a scaffolding protein has been developed. This uses an inactive type Cas9 (dCas9), which does not have DNA cleavage activity. It is proposed to regulate the transcription efficiency of downstream genes by fusing transcription regulatory proteins to the dCas9 (Gilbert et al., 2013; Konermann et al., 2013; Maeder et al., 2013; Perez-Pinera et al., 2013; Mali et al., 2013; Konermann et al., 2014) and epigenetic modification at any position on the genome by fusing the enzyme to dCas9 (Konermann et al., 2013). Also, there is a method in which specific DNA sequences are live-imaged by adding a reporter fluorescent protein to dCas9 (Chen et al., 2013) (Figure 1-4).

1.4 Mechanistic details of SpCas9

Structural studies of *Streptococcus pyogenes* Cas9 (SpCas9) have provided mechanistic details of the RNA-guided DNA cleavage by the Cas9 enzyme. The crystal structure of SpCas9 in its apo form revealed a bilobed architecture comprising an α -helical recognition (REC) lobe and a nuclease (NUC) lobe (Jinek et al., 2014). The crystal structure of SpCas9 bound to the sgRNA and its single-stranded DNA target clarified the recognition mechanism of the sgRNA and the target DNA (Nishimasu et al., 2014). Subsequently, the crystal structure of SpCas9 bound to the sgRNA and a

PAM-containing DNA revealed the recognition mechanism of the 5'-NGG-3' PAM by SpCas9 (Anders et al., 2014). Moreover, the crystal structures of SpCas9 bound to the sgRNA (Jiang et al., 2015) and SpCas9 bound to an R-loop (Jiang et al., 2016) demonstrated the structural rearrangements in the Cas9 protein accompanying the guide RNA binding and R-loop formation, respectively (Figure 1-5 and Table1-1).

1.5 Structural studies of Cas9 orthologs

The Cas9 orthologs from different microbes have highly divergent sequences, function with their cognate crRNA:tracrRNA guides, and recognize a variety of PAM sequences (Chylinski et al., 2013; Fonfara et al., 2014; Hsu et al., 2014; Karvelis et al., 2015; Ran et al., 2015). SpCas9 (1,368 aa) recognizes 5'-NGG-3' as the PAM (Mojica et al., 2009), whereas *Staphylococcus aureus* Cas9 (SaCas9, 1,053 aa) and *Francisella novicida* Cas9 (FnCas9, 1,629 aa) recognize 5'-NNGRRT-3' and 5'-NGG-3' as the PAMs, respectively (Ran et al., 2015; Hirano et al., 2016) (Figure1-6). A structural comparison of SpCas9 (Anders et al., 2014; Nishimasu et al., 2014) with SaCas9 (Nishimasu et al., 2015) and FnCas9 (Hirano et al., 2016) (Figure1-6) revealed that, while they share the conserved RuvC and HNH nuclease domains, their REC and Wedge (WED) domains are structurally divergent, and recognize distinct structural features in their cognate RNA guides (Figure 1-3, 1-6 and Table1-1). In addition, their PAM-interacting (PI) domains

share a conserved core fold, but recognize distinct PAM sequences, using a specific set of amino-acid residues (Figure 1-3, 1-6 and Table1-1).

1.6 Correlation between the evolution and the cleavage activity of Cas9 orthologs

According to the previous studies, it is possible that the type II-C species that possess a relatively small Cas9 is ancestor of the type II-A species which possess relatively large Cas9, such as SpCas9 (Fonfara et al., 2014; Chylinski et al., 2015). The finding that Cas9 proteins belonging to type II-C are suitable for ssDNA break, because of their limited ability for dsDNA breaks may suggest that they evolved to recognize the substrate such as ssDNA bacteriophage (Ma et al., Mol Cell, 2015). Because the dsDNA cleavage is required the unwinding DNA to penetrate guide RNA and form the R-loop (helicase activity), small ancestral type II-C Cas9 proteins may limit their DNA unwinding ability. Therefore, it was initially thought that Cas9 derived from typeII-C subtype could not be suitable for the genome editing *in vivo*. However, very recently, it was reported that genome editing using *Campylobacter jejuni* Cas9 (CjCas9) was successful by examining the sequence of the guide RNA and PAM (Kim et al., 2017).

1.7 Purpose of this study

CjCas9 has several unique features (Fonfara et al., 2014). First, CjCas9 consists of 984 residues, and is one of the smallest Cas9 orthologs. Second, the nucleotide sequence of the crRNA:tracrRNA guides for CjCas9 and the other Cas9 orthologs differ substantially. Third, CjCas9 recognizes the 5'-NNNNACA-3' PAM, whereas most Cas9 orthologs, as exemplified by SpCas9 (5'-NGG-3') (Gasiunas et al., 2012; Jinek et al., 2012), recognize G-rich PAMs. However, the functional mechanism of CjCas9 remains elusive, due to the lack of structural information. To understand the RNA-guided DNA recognition mechanism of CjCas9, the author attempted to determine the crystal structure of CjCas9 in complex with an sgRNA and its target DNA.

Here, the author performed functional and structural characterizations of CjCas9. *In vitro* cleavage experiments revealed that CjCas9 recognizes the 5'-NNNVRYM-3' PAM, which is more promiscuous than the previously reported PAMs. The crystal structure of the CjCas9-sgRNA-target DNA complex highlighted the remarkable mechanistic diversity of the CRISPR-Cas9 systems. Unlike the tracrRNAs for the other Cas9 orthologs, the CjCas9 tracrRNA has an unanticipated triple-helix structure, which is distinct from known RNA triple helices. Furthermore, CjCas9 recognizes the PAM nucleotides on both the target and non-target DNA strands, whereas the other Cas9 orthologs recognize the PAM nucleotides on the non-target DNA strand.

CjCas9 has been thought to be unavailable for genome editing due to weak DNA cleavage activity in eukaryotic cells (Ran et al., 2015). However, very recently, it was reported that genome editing using CjCas9 was successful by examining the sequence of the guide RNA and PAM (Kim et al., 2017). Since small CjCas9 has advantages such as high efficiency of introduction into viral vectors, it is expected to be applied to genome editing. However, CjCas9 has weak DNA cleavage activity in eukaryotic cells because cleavage efficiency was lower than that of SpCas9. From this background, the author conducted a search for activity enhancing mutants and succeeded in creating mutants with twice as much activity intensity as WT CjCas9.

Overall, these findings improved our mechanistic understanding of the CRISPR-Cas9 systems, and provided a framework for further Cas9 engineering for its applications.

Figures and Tables of Chapter 1

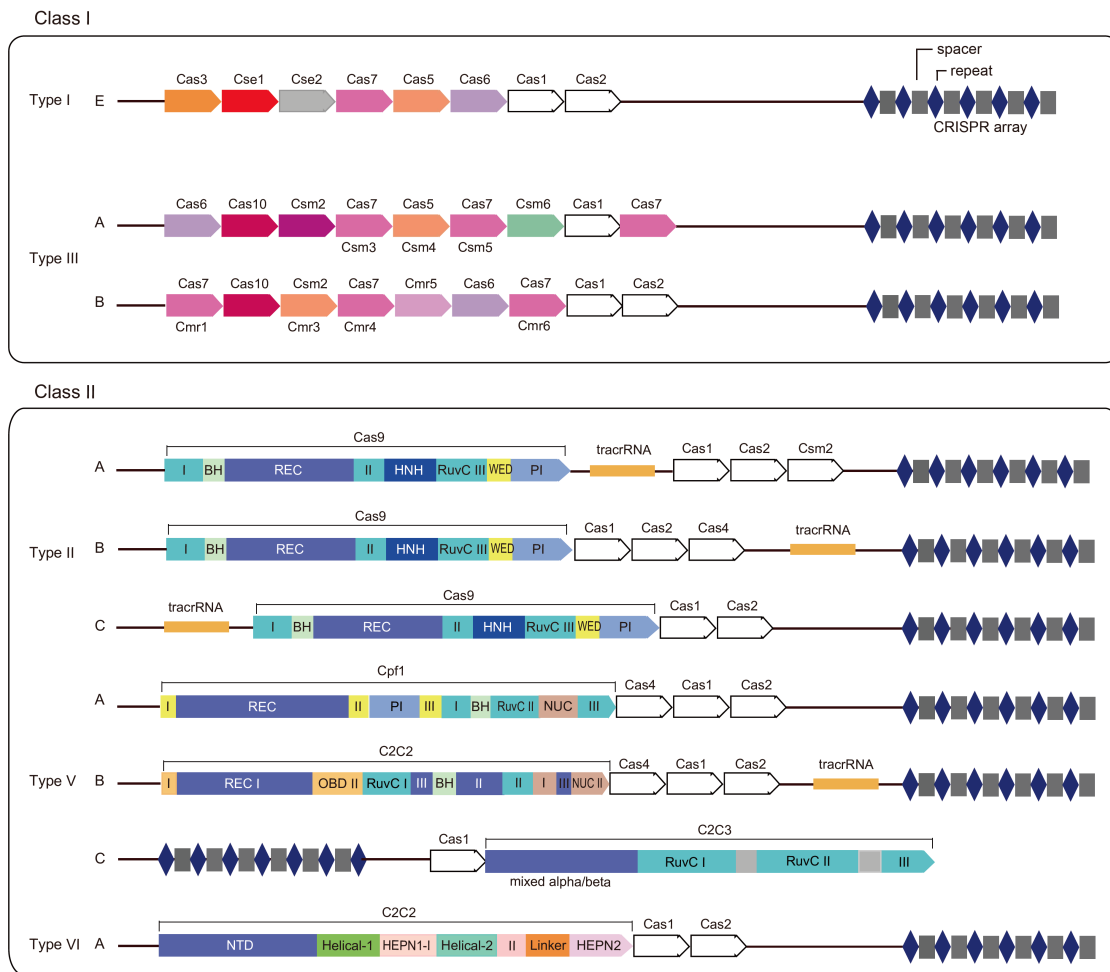


Figure 1-1 Schematic diagram of class 2 CRISPR-Cas system

Class 1 CRISPR-Cas systems use multi protein effector complexes, while class 2 CRISPR-Cas systems use single protein effectors. The majority of the Class 2 genomic loci encode adaptation module protein, Cas1 and Cas2, and accessory proteins, such as Cas4. Type II and type V-B loci also include tracrRNA, which is partially complementary to the repeats and is responsible for crRNA processing and interference. However, some Class 2 systems, particularly those of type VI, consist solely of a CRISPR array and an effector protein. See also Figure1-2.

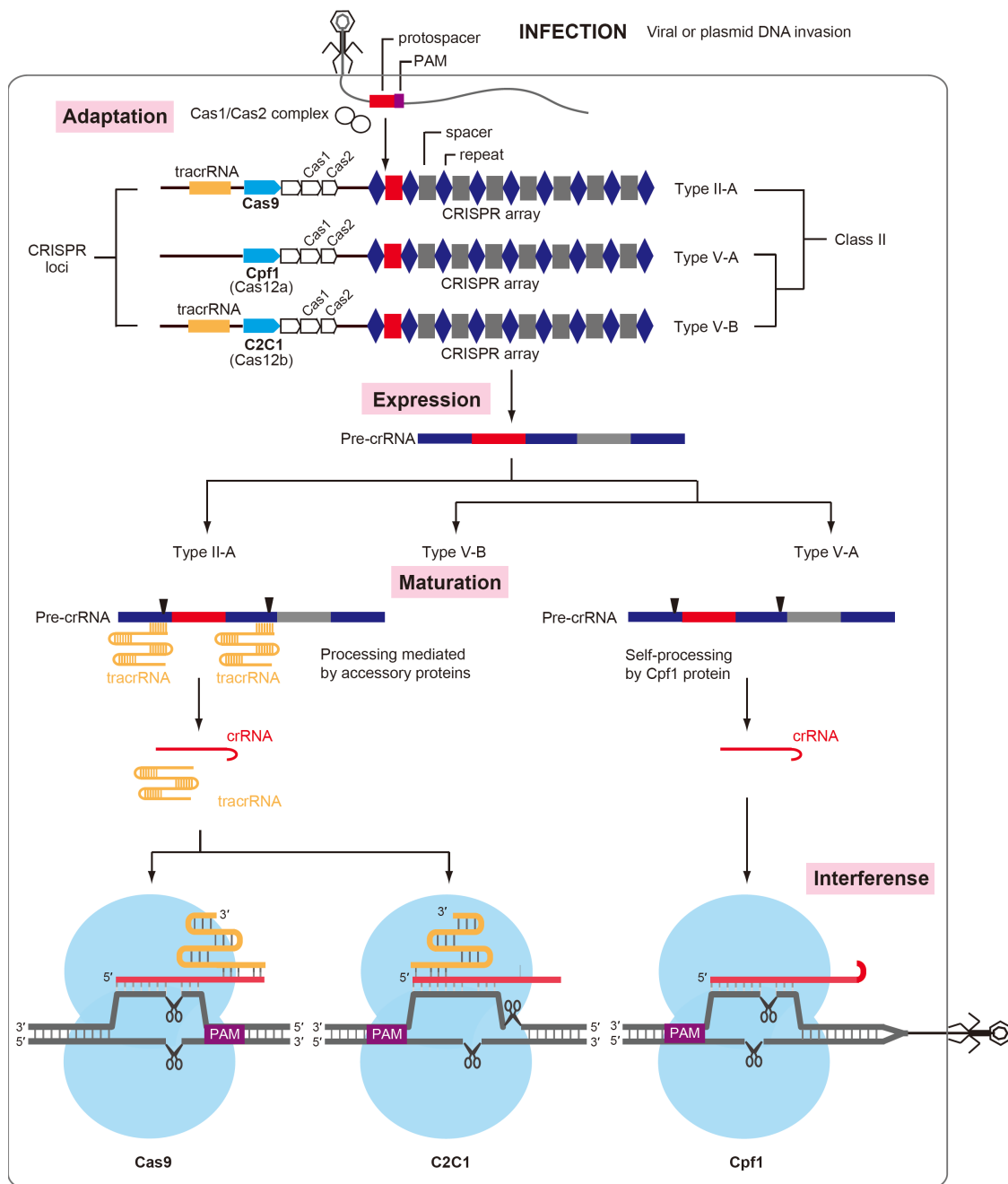
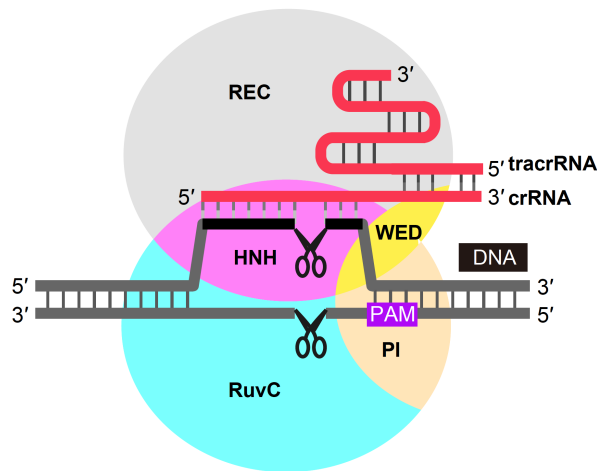


Figure 1-2 Schematic diagram of class 2 CRISPR-Cas system

The CRISPR-Cas system is roughly divided into three stages: Adaptation stage, Expression stage, and Interference stage. As a feature of the class 2 CRISPR-Cas system, a single Cas is involved. Figure 1-1 shows the mechanism of action of Cas9, C2c1 and Cpf1 (Cas12a) related to DNA targeting in class 2.

A Cas9 programmed by crRNA:tracrRNA duplex



B Cas9 programmed by single guide RNA

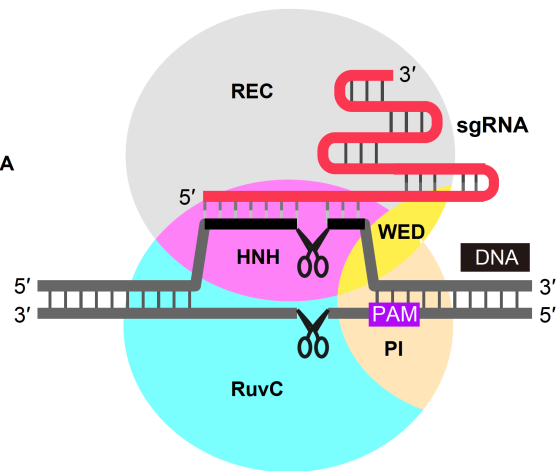
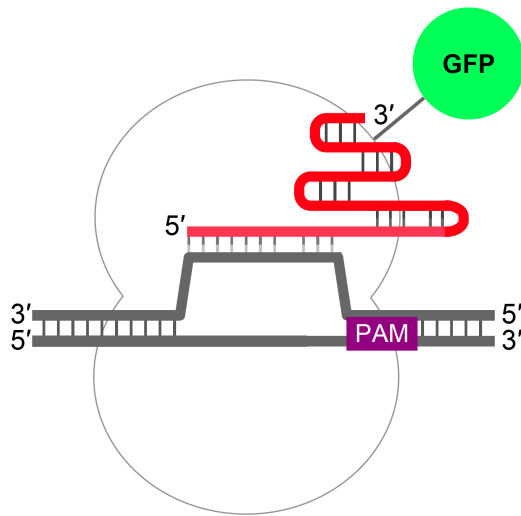


Figure 1-3 Cas9 targeting using crRNA-tracrRNA or a single-guide RNA (sgRNA)

- (A) The RNA-guided DNA endonuclease Cas9 associates with either dual guide RNAs, crRNA (CRISPR RNA) and tracrRNA (*trans*-activating crRNA) and generates a DNA double-strand break at DNA target sites complementary to the guide RNA.
- (B) A synthetic single-guide RNA (sgRNA) is also able to target and cleave double-stranded DNA.

A DNA imaging



B Genetic /Epigenetic modulation

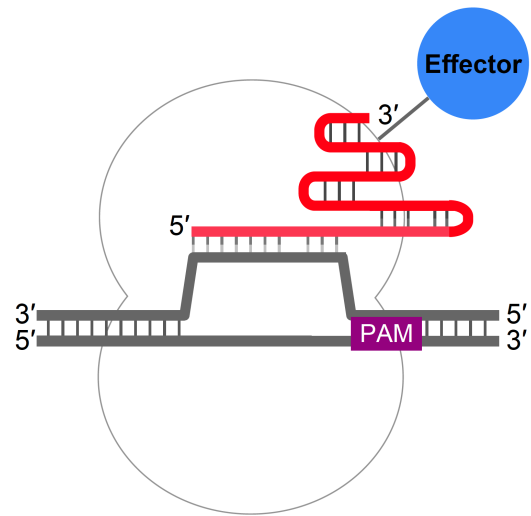


Figure 1-4 Engineering nuclease-deactivated Cas9 (dCas9)

- (A) DNA imaging tool by fusing GFP to dCas9 for labeling genomic loci and visualizing dynamic changes of chromosomes.
- (B) Genetic or epigenetic modulation tool by fusing appropriate effector to dCas9 for gene expression.

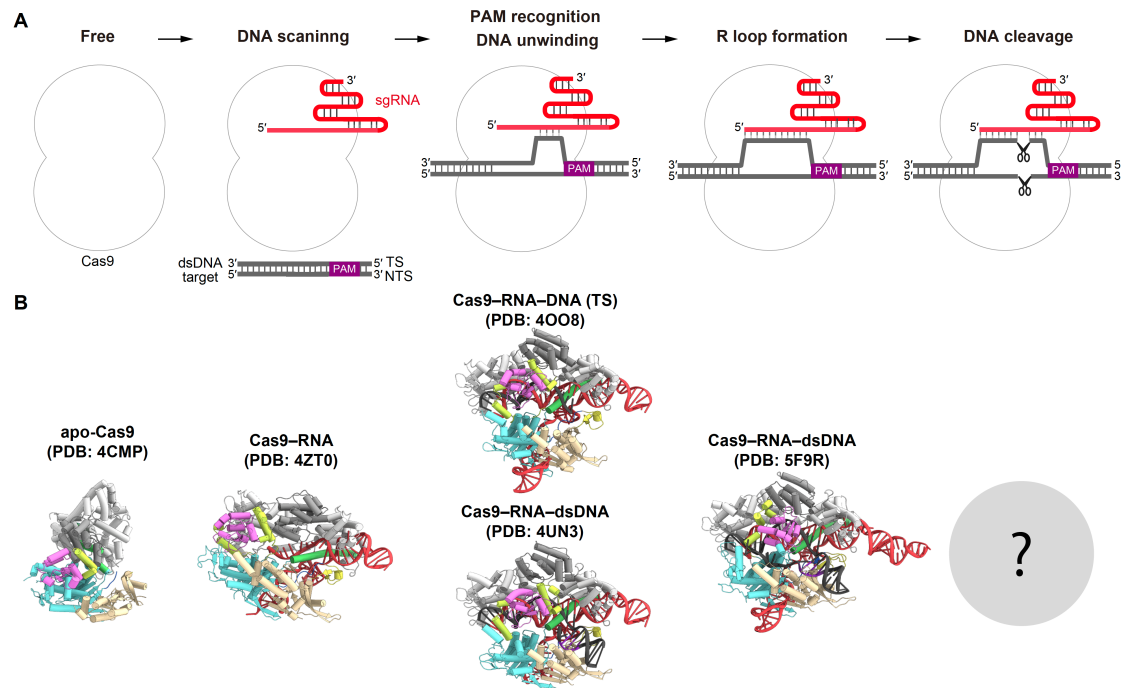


Figure 1-5 Conformation change of Cas9 in DNA targeting

(A) Model for target DNA unwinding and R-loop formation in Cas9.

(B) Crystal structures of apo-SpCas9 (PDB ID: 4CMP), SpCas9-RNA (PDB ID: 4ZT0), SpCas9-RNA bound to its target strand (TS) DNA (PDB ID: 4OO8), SpCas9-RNA bound to its dsDNA (a partial DNA duplex) (PDB ID: 4UN3) and SpCas9-RNA bound to its dsDNA (a Cas9 R-loop complex) (PDB ID: 5F9R). The guide RNA is colored red and the target DNA is colored black, respectively. The PAM is colored purple. See also Table1-1.

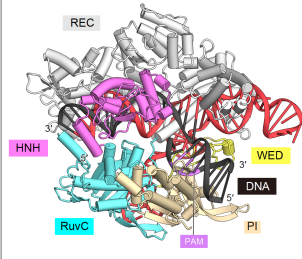
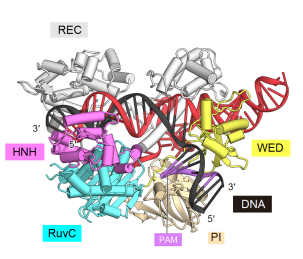
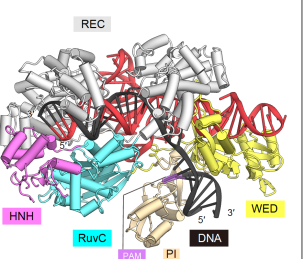
	<i>Streptococcus pyogenes</i> Cas9 (SpCas9)	<i>Streptococcus pyogenes</i> Cas9 (SaCas9)	<i>Francisella novicida</i> Cas9 (FnCas9)
structure			
sgRNA	<pre> 5' GUUUUAGA--GCUAUG A U C-GGAAUAAAAUUGAACGAUAA A A GUCCGUUAUACACUUG A AGCCACGGUGAAA A G UCGGUGCUUUUU 3' </pre>	<pre> 5' GUUUUAGUA C UCUGG A A CGGAACAAAAUCAU CUA AGACA A A A GCCGUGUUUAUCUCGUAAC U U 3' UAGAGCGGUUG U </pre>	<pre> 5' GUUUCAGUUGCUG-AAUUAG A GAAUUUACUAAUGUC-UCGUAUUUAUA A U AUUUUGAACGGACCUCUGUU U 3' CAAUAAGUCUG-C-ACAG </pre>
PAM	5'-NGG-3'	5'-NNGRRT-3'	5'-NGR-3'

Figure 1-6 Structures of known Cas9 orthologs, their sgRNA and PAM sequences

The Cas9 orthologs from different microbes have highly divergent sequences, function with their cognate crRNA:tracrRNA guides, and recognize a variety of PAM sequences. The structure of sgRNA is predicted by secondary structure prediction. See also Table1-1. SpCas9 (PDB ID:4UN3), SaCas9 (PDB ID: 5CZZ), and FnCas9 (PDB ID: 5B2O).

Table 1-1 Summary of the available structures of Cas9

Conformation	Orthologs	PDB code	Reference
Apo	<i>Streptococcus pyogenes</i>	4CMP, 4CMQ	Jinek <i>et al.</i> , <i>Science</i> , 2014
	<i>Actinomyces naeslundii</i>	4OGC, 4OGE	Jinek <i>et al.</i> , <i>Science</i> , 2014
RNA bound	<i>Streptococcus pyogenes</i>	4ZT0, 4ZT9	Jiang <i>et al.</i> , <i>Science</i> , 2015
		EMD3276	Jiang <i>et al.</i> , <i>Science</i> , 2016
ssDNA-bound	<i>Streptococcus pyogenes</i>	4OO8	Nishimasu <i>et al.</i> , <i>Cell</i> , 2014
PAM bound	<i>Streptococcus pyogenes</i>	4UN3, 4UN4, 4UN5	Anders <i>et al.</i> , <i>nature</i> , 2014
		5B2R, 5B2S, 5B2T,	Hirano <i>et al.</i> , <i>Mol Cell</i> , 2016
		5FW1, 5FW2, 5FW3,	Anders <i>et al.</i> , <i>Mol Cell</i> , 2016
	<i>Staphylococcus aureus</i>	5AXW, 5CZZ	Nishimasu <i>et al.</i> , <i>Cell</i> , 2015
	<i>Francisella novicida</i>	5B2O, 5B2P, 5B2Q	Hirano <i>et al.</i> , <i>Cell</i> , 2016
	<i>Campylobacter jejuni</i>	5X2G, 5X2H	Yamada <i>et al.</i> , <i>Mol Cell</i> , 2017 (This study)
R-loop before cleavage	<i>Streptococcus pyogenes</i>	5F9R	Jiang <i>et al.</i> , <i>Science</i> , 2016
		EMD3277	

Chapter 2 PAM specificity of *Campylobacter jejuni* Cas9

2.1 Introduction

Although a previous study reported that CjCas9 recognizes the 5'-NNNNACA-3' PAM (Fonfara et al., 2014), the CjCas9 PAM has not been fully characterized. To determine the CjCas9 PAM, I performed the PAM discovery assay, using purified CjCas9, an sgRNA and a library of plasmid DNA targets with a degenerate 7-bp PAM sequence, as described previously (Ran et al., 2015; Zetsche et al., 2015) both *in vitro* and *in vivo*.

2.2 *In vivo* PAM discovery assay

2.2.1 Materials and Methods

Plasmid libraries containing randomized PAM sequences were constructed using synthesized oligonucleotides (Integrated DNA Technologies), consisting of seven randomized nucleotides 3' of a 20-nt target sequence, as previously described (Zetsche et al., 2015). The randomized PAM plasmid library was cleaved *in vitro* using purified CjCas9 with sgRNAs targeting the PAM library, and the cleavage products were separated on 2% agarose E-gels (Life Technologies). The band corresponding to the uncleaved target plasmid was isolated with a Zymoclean Gel DNA Recovery Kit (Zymo Research), and the region surrounding the randomized PAM region was PCR-amplified and sequenced using a MiSeq (Illumina) with 150 single-end cycles. To analyze the

resulting sequence data, the seven nucleotide PAM region was extracted, individual PAMs were counted, and PAM counts were normalized to total reads for each sample. For a given PAM sequence, enrichment was measured as the log₂ ratio as compared to a no-protein control, with a 0.01 pseudocount adjustment. PAMs above an enrichment threshold set to 3.5 were compiled and used to generate sequence logos (Crooks et al., 2004). For PAM wheel generation, abundances were used to generate wheels with Krona (Ondov et al., 2011), as previously described (Leenay et al., 2016).

2.2.2 Results

CjCas9 recognized the fourth PAM specifically A/G/C, the fifth PAM specifically A/G, T/C in the sixth PAM, and A/C in the seventh PAM. The result revealed that CjCas9 recognizes the 5'-NNNVR YM-3' PAM (V is A/G/C; R is A/G; Y is T/C; M is A/C) (Figure 2-1A), which is more promiscuous than the previously reported 5'-NNNNACA-3' PAM (Fonfara et al., 2014). (Figure 2-1A).

2.3 *In vitro* PAM discovery assay

Using purified CjCas9 and an sgRNA, the author further examined the cleavage of 13 plasmid DNA targets with either 5'-AGANACC-3', 5'-AGAANACA-3', 5'-AGAAANA-3' or 5'-AGAACNA-3' as the PAMs.

2.3.1 Materials and Methods

In vitro plasmid DNA cleavage experiments were performed essentially as described previously (Nishimasu et al., 2015). The *Eco*RI-linearized pUC119 plasmid (100 ng, 4.7 nM), containing the 20-nt target sequence and the PAMs, was incubated at 37°C for 5 min with the CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5), in 10 µl of reaction buffer, containing 20 mM HEPES, pH 7.5, 100 mM KCl, 2 mM MgCl₂, 1 mM DTT and 5% glycerol. Reaction products were resolved on a 1% agarose gel, stained with ethidium bromide, and then visualized using a Typhoon FLA 9500 imager (GE Healthcare).

2.3.2 Results

The results confirmed that CjCas9 efficiently recognizes the 5'-NNNVR YM-3' PAM, with the preference for T and C at positions 6 and 7, respectively (Figures 2-1B and 1C).

Figures of Chapter 2

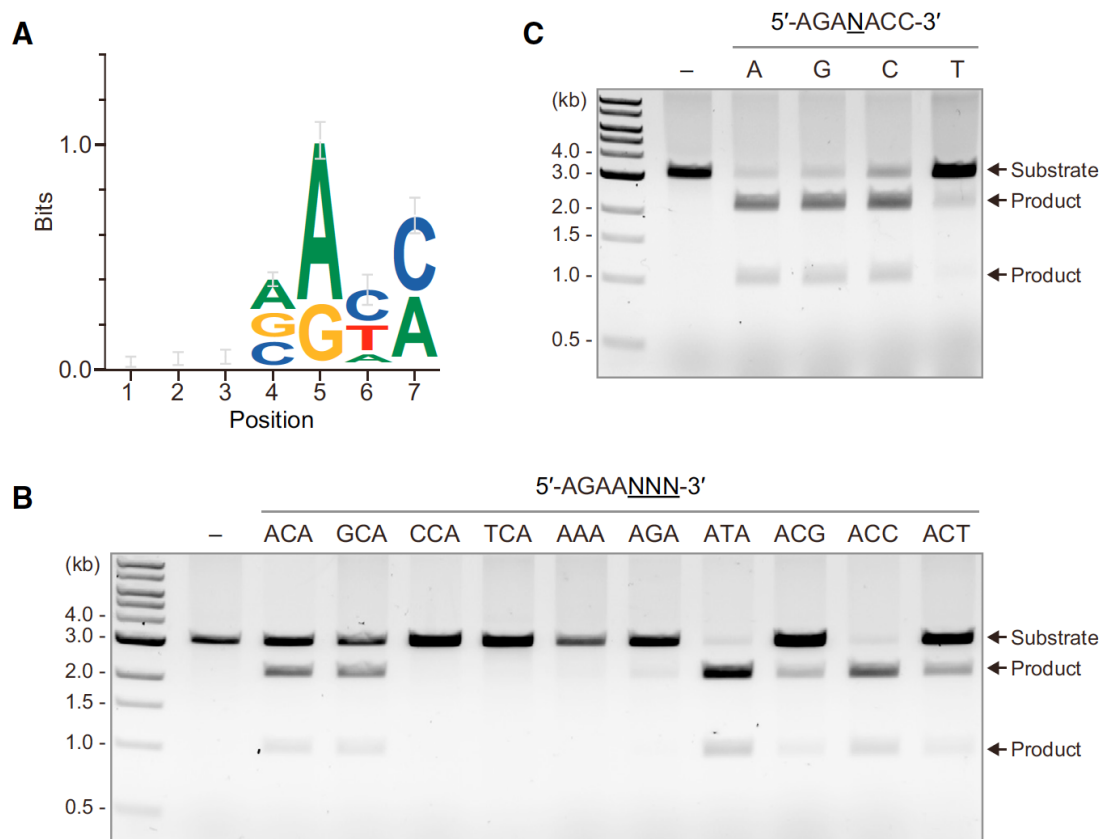


Figure 2-1 CjCas9 PAM Specificity.

(A) Motif obtained from the *in vitro* PAM discovery assay.

(B and C) *In vitro* cleavage assays for DNA targets with different PAMs. The linearized plasmid targets with either the 5'-AGAANCA-3', 5'-AGAAANA-3' or 5'-AGAACN-3' PAM (B), or the 5'-AGANACC-3' PAM (C) were incubated with CjCas9-sgRNA, and then analyzed by agarose gel electrophoresis.

Chapter 3 X-ray crystallographic analysis of *Campylobacter jejuni* Cas9

3.1 Introduction

To understand the RNA-guided DNA recognition mechanism of CjCas9, the author attempted to determine the crystal structure of CjCas9 in complex with an sgRNA and its target DNA. In this chapter, the crystallization and X-ray diffraction analysis of CjCas9, lacking the HNH domain, in complex with a 93-nt sgRNA, a 28-nt target DNA strand, and an 8-nt non-target DNA strand were described.

3.2 Materials and Methods

3.2.1 Plasmid construction

The gene encoding full-length CjCas9 (residues 1–984) was codon optimized, synthesized (Genscript), and cloned between the *Nde*I and *Xho*I sites of the modified pE-SUMO vector (LifeSensors). To improve the stability of the protein and the quality of the crystals, I prepared the CjCas9- Δ HNH variant lacking the HNH domain (residues 481–640), in which Leu480 (RuvC-II) and Tyr641 (RuvC-III) are connected by a GGGSGG linker. The CjCas9- Δ HNH variant was created by a PCR-based method, using the vector encoding the full-length CjCas9 as the template (Table 3-1 and 3-2).

3.2.2 CjCas9-ΔHNH protein expression and purification

The CjCas9-ΔHNH protein was over-expressed in *Escherichia coli* Rosetta 2 (DE3) (Novagen) in LB medium. The cells were grown at 37°C until the absorbance at 600 nm (A_{600}) reached 0.8 and then induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 20 C for 20 hours. Cell pellets was resuspended in buffer A, lysed by ultrasonic disruption, and centrifuged at 40000 g for 30 min at 4°C in a JA-14 fixed angle rotor (Beckman Coulter). The supernatant containing CjCas9 protein was purified by chromatography on Ni-NTA Superflow resin (QIAGEN) equilibrated in buffer A. After collecting the flow-through fraction, the column was washed with buffer A, then washed with buffer B to remove nucleic acids. Thereafter, the column was washed with buffer A, and then protein was eluted with buffer C.

The eluted protein was loaded to a HiTrap Heparin HP column (GE Healthcare) pre-equilibrated in a mixture of buffer D and eluted by a linear gradient from 300 mM to 2 M NaCl. The eluted protein was dialyzed overnight at 20°C with TEV protease, to remove the N-terminal His₆-SUMO-tag, then separated by Ni-NTA Superflow resin (QIAGEN) equilibrated with buffer A. The CjCas9-ΔHNH protein was further purified by chromatography on HiLoad 16/600 Superdex 200 (GE Healthcare) columns equilibrated in buffer F. Buffer compositions are summarized in Table 3-3.

3.2.3 *In vitro* transcription and purification of sgRNA

The sgRNA was transcribed *in vitro* with T7 RNA polymerase, using a PCR-amplified dsDNA template. The transcribed RNA was purified by 8% denaturing (7 M urea) polyacrylamide gel electrophoresis.

3.2.4 DNA preparation

The target and non-target DNA strands were purchased from Sigma-Aldrich.

3.2.5 Complex reconstruction

The purified CjCas9- Δ HNH protein was mixed with the sgRNA, the target DNA strand, and the non-target DNA strand (containing either the 5'-AGAAACA-3' PAM or the 5'-AGAAACC-3' PAM) (molar ratio, 1:1.5:2.3:3.4), and then the CjCas9-sgRNA-DNA complex was purified by gel filtration chromatography on a Superdex 200 Increase column (GE Healthcare).

3.2.6 Crystallization of the CjCas9- Δ HNH–sgRNA–target DNA complex

The purified CjCas9-sgRNA-DNA complex solution (containing either the 5'-AGAAACA-3' PAM or the 5'-AGAAACC-3' PAM) ($A_{260\text{ nm}} = 15$) was grown at 20°C. The initial crystallization screening was performed using the screening kits by the

sitting-drop vapor diffusion method. The screening kits used for the initial screening were PACT Suite (QIAGEN), JCSG+ Suite (QIAGEN), Crystal Screen 1 & 2 (Hampton Research), PEG/Ion (Hampton Research), JB Screen Classic 1, 2, 4, 5 (Jena Bioscience), Mem Gold, Mem Gold2 (Molecular Dimensions) and Wizard I & II (Emerald Biosystems). For crystallization conditions under Initial screening, crystallization condition was optimized by changing the type and concentration of precipitant, pH of buffer, or screening additives using Additive Screen (Hampton Research). Crystallization was performed by the sitting-drop vapor diffusion method or hanging-drop vapor diffusion method. The crystallization drops were formed by mixing 1 μ l of complex solution and 1 μ l of reservoir solution, and then incubated against 0.5 ml of reservoir solution. The crystals were improved by micro-seeding using Seed Bead (Hampton Research). The crystallization robot Mosquito (TTP Labtech) was used for the sitting-drop vapor diffusion.

3.2.7 Expression and purification of the SeMet-substituted CjCas9- Δ HNH protein

The selenomethionine (SeMet)-substituted CjCas9- Δ HNH was expressed in *E. coli* B834 (DE3) (Novagen), and was purified using a similar protocol to that for the native protein.

3.2.8 Crystallization of the SeMet-substituted CjCas9- Δ HNH-sgRNA-target DNA complex

The SeMet-labeled complex (containing the 5'-AGAAACA-3' PAM) was crystallized under similar conditions.

3.2.9 X-ray diffraction analysis and data processing

X-ray diffraction data were collected at 100 K on beamlines BL41XU at SPring-8 and PXI at the Swiss Light Source. The crystals were cryoprotected in reservoir solution supplemented with 25% ethylene glycol. X-ray diffraction data were processed using DIALS (Waterman et al., 2013) and AIMLESS (Evans and Murshudov, 2013).

3.2.10 Model building and structure refinement

The structure was determined by the Se-SAD method, using PHENIX AutoSol (Adams et al., 2010). The 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 5'-AGAAACA-3' PAM complex (2.3 Å resolution) and the 5'-AGAAACC-3' PAM complex (2.4 Å resolution) were refined using native data sets. Data collection statistics

are summarized in Table 3-3. Structural figures were prepared using CueMol (<http://www.cuemol.org>).

3.2.11 Expression and purification of WT CjCas9 protein

For *in vitro* cleavage assays, the wild type and mutants of full-length CjCas9 were expressed and purified, using a protocol similar to that for CjCas9- Δ HNH.

3.3 Results

3.3.1 Expression, purification and construct optimization

First the author attempted crystallization with full-length of CjCas9 in complex with an sgRNA and its target DNA . However, the author failed to obtain diffraction-quality crystals. Previous studies revealed that the HNH nuclease domain of SpCas9 is mobile and dispensable for RNA-guided DNA recognition (Nishimasu et al., 2014; Jiang et al., 2015; Sternberg et al., 2015), suggesting that the flexibility of the HNH domain may hamper crystallization. The author thus prepared the CjCas9- Δ HNH variant lacking the HNH domain (residues 481–640), in which Leu480 (RuvC-II) and Tyr641 (RuvC-III) are connected by a GGGSGG linker (Figure 3-1 and 3-2).

3.3.2 Complex reconstruction

After gel filtration chromatography purification of the reconstituted sample, each peak were fractionated and subjected to SDS-PAGE. Since the presence of the target protein was present at the main peak, the formation of a quaternary complex was confirmed (Figure 3-3).

3.3.3 Crystallization

After extensive screening, crystals were obtained by mixing 1 mL of complex solution ($A_{260\text{ nm}} = 15$) and 1 mL of reservoir solution (12.0%–14.5% PEG 2,000, 0.4 M ammonium acetate). The SeMet-labeled complex (containing the 5'-AGAAACA-3' PAM) was crystallized under similar conditions (Figure 3-4).

3.3.4 X-ray diffraction analysis

The author determined the crystal structures of CjCas9, lacking the HNH domain (residues 482–638), in complex with a 93-nt sgRNA, a 28-nt target DNA strand, and an 8-nt non-target DNA strand containing either the 5'-AGAAACC-3' PAM or the 5'-AGAAACA-3' PAM, respectively (Figure 3-5). The author will mainly describe the 5'-AGAAACC-3' PAM complex structure, which is relatively ordered, as compared to

the 5'-AGAAACA-3' PAM complex. Data collection statistics are summarized in Table 3-4.

Figures and Tables of Chapter 3

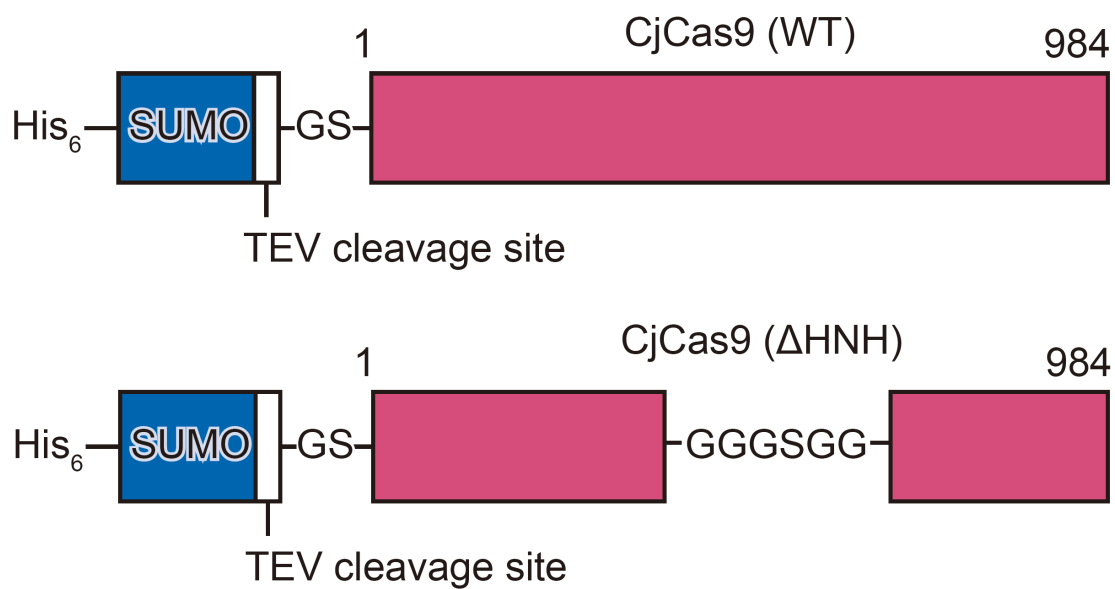


Figure 3-1 Schematic diagram of the CjCas9 expression constructs

Expression constructs of full-length CjCas9 and HNH domain truncated CjCas9. The TEV protease cleavage site and a His₆-tag were fused into the N-terminus of the inserted gene.

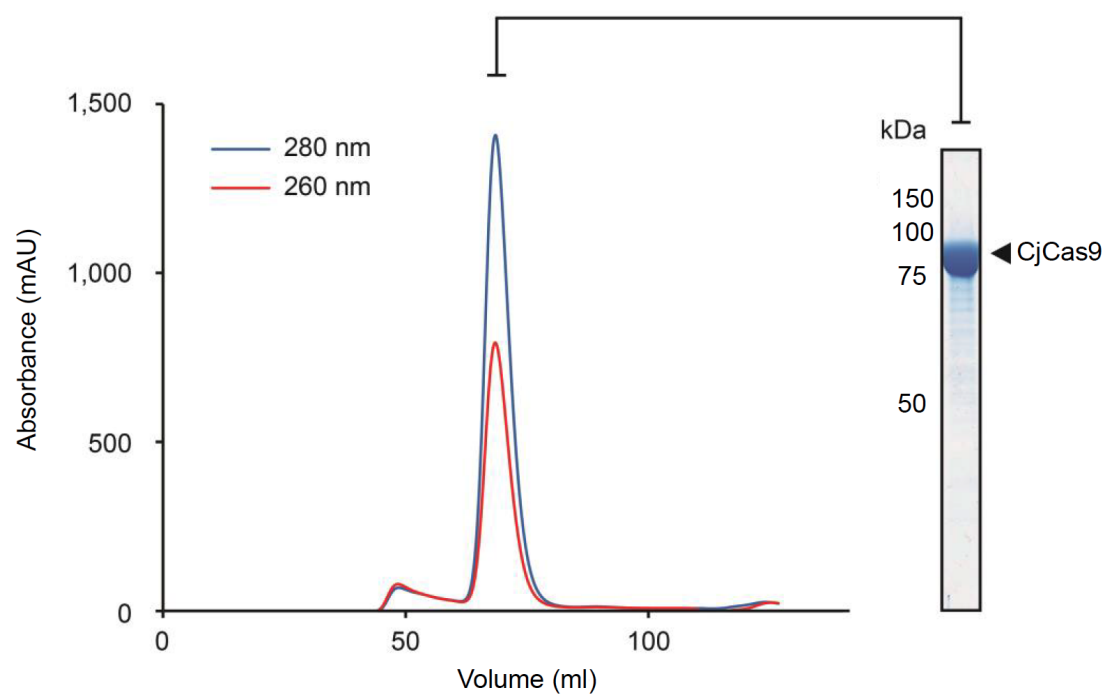


Figure 3-2 Size-exclusion chromatograms and SDS-PAGE

Chromatograms of Δ HNNH-CjCas9 and SDS-PAGE analysis with coomassie brilliant blue staining. A highly purified CjCas9- Δ HNNH for crystallization was obtained.

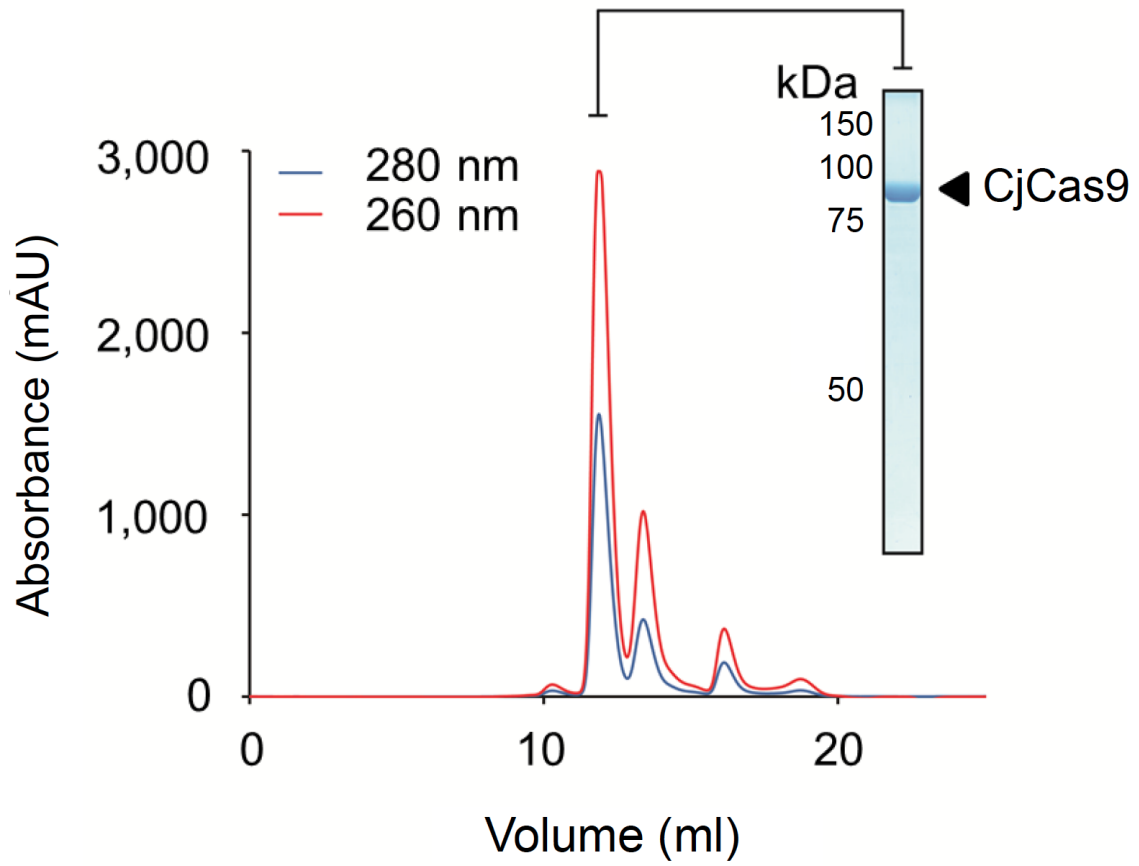


Figure 3-3 Purification of the CjCas9 quaternary complex

Size-exclusion chromatograms and SDS-PAGE of elution peak confirmed the presence of CjCas9 at the main peak. The purified CjCas9- Δ HNH protein was mixed with the sgRNA, the target DNA strand, and the non-target DNA strand (molar ratio, 1:1.5:2.3:3.4)

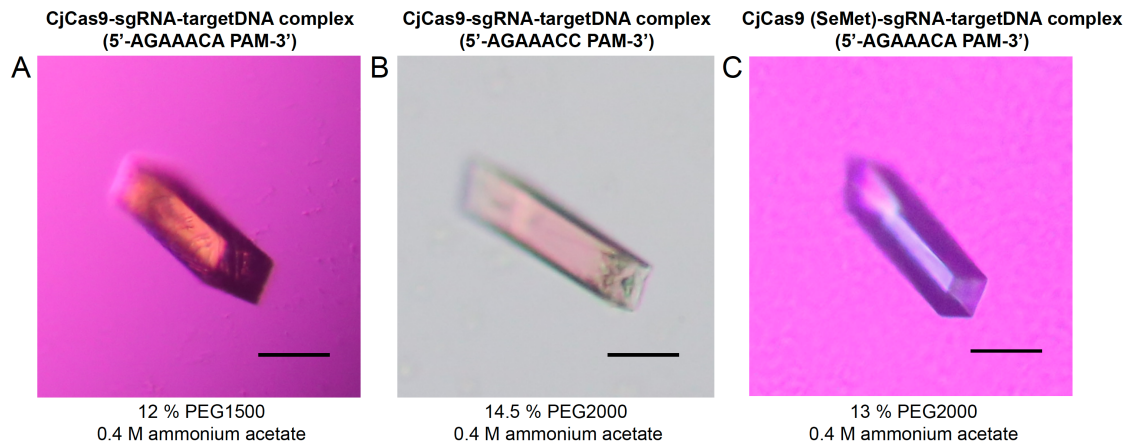


Figure 3-4 Crystals of CjCas9-sgRNA-target DNA

(A) Crystals of CjCas9-sgRNA-target DNA complex including 5'-AGAAACA-3' PAM.

(B) Crystals of CjCas9-sgRNA-target DNA complex including 5'-AGAAACC-3' PAM.

(C) Crystals of SeMet labeled CjCas9-sgRNA-target DNA complex including 5'-AGAAACA-3' PAM.

The reservoir solution obtained these crystals is shown under the figures. The scale bars represent 100 μm .

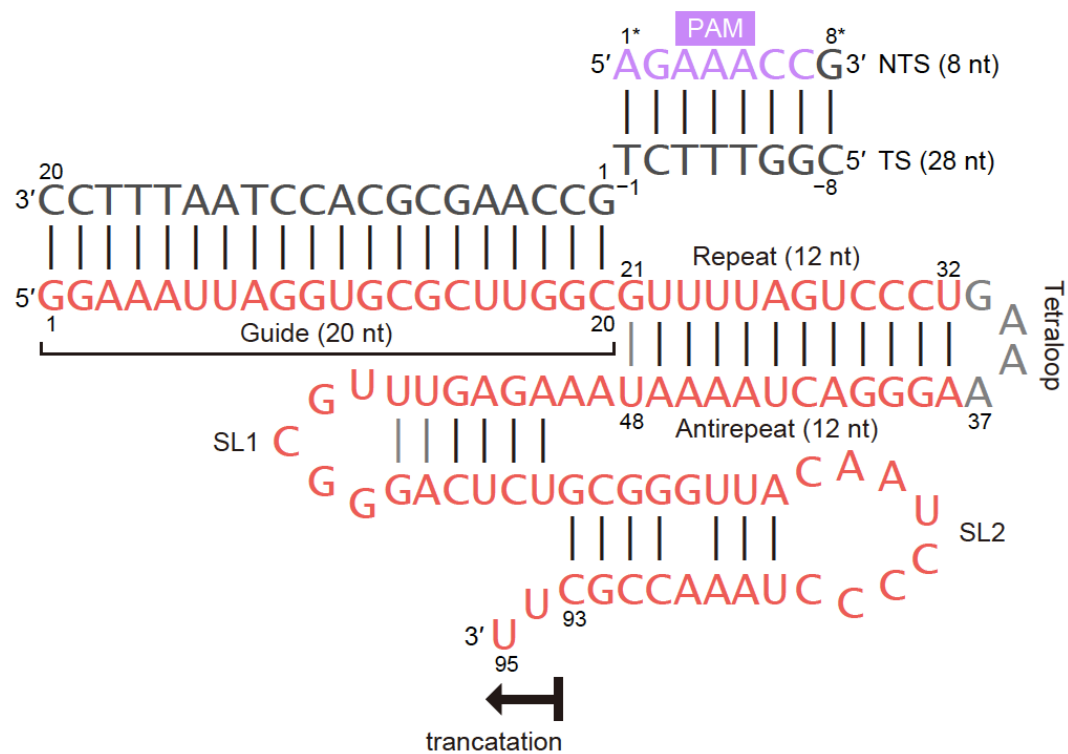


Figure 3-5 Schematic diagram of the Cj sgRNA and target DNA constructs for crystallization

The structure of RNA is based on secondary structure prediction. The guide RNA is colored red and the target DNA is colored black, respectively. The PAM is colored purple. TS, target strand; NTS, non-target strand.

Table 3-1 Macromolecule-production information

Source organism	<i>Campylobacter jejuni</i>
DNA source	Codon optimized, synthesized (Genscript)
Expression vector	Modified pESUMO-TEV
Expression host	<i>E. coli</i> Rosetta2 (DE3)
Nucleic acid sequence of the construct	AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGAT AACAATCCCCTCTAGAAATAATTTTGTTAACTTTAAGAAGGAGATATACCATGGG T CATCACCATCATCATCAC GGGTCCCTGCAGGACTCAGAAGTCAATCAAGAAGCT Green : 6×His AAGCCAGAGGTCAAGCCAGAAGTCAAGCCTGAGACTCACATCAATTTAAAGGTGT RED: SUMO site CCGATGGATCTTCAGAGATCTTCTTCAAGATCAAAAAGACCACTCCTTTAAGAAGG Orange : TEV site CTGATGGAAGCGTTTCGTAAAAGACAGGGTAAGGAAATGGACTCCTTAAGATTCT Gray :SSGS GTGATCGACGGTATTAGAATTCAAGCTGATCAGGCCCTGAAGATTGGACATGGA linker GGATAACGATATTATTGAGGCTCACCGCGAACAGATTGGAGGT GAA AACTTGTAC Blue : CjCas9 (WT) TTCCAAGGAT CT CTCTCCGGATCCGGACATATGGCCCGCATCCTCGCTTT CGACA TCGGAATCTCTAGTATCGGATGGGCCTTCTCTGAAAACGACGAAGTGAAGACTG CGGCGTGAGAATCTTCACAAAGGTTGAAAACCTAAACAGGCGAGTCTTTAGCT CTGCCACGTAGGTTGGCCCGCTCCGCCGAAAAAGGCTGGCTCGGCGGAAGGC TCGCCTCAACCACTTGAAGCATTGATAGCTAATGAGTTCAAAGTGAAGTACGAAG ATTACAGTCTTCGACGAGTCATTGGCAAAAGCCTACAAAGGCAGCCTTATCAG TCCTTATGAGTTGAGATTCGCGCACTCAACGAAGTCTTTCTAAGCAAGACTTTG CTAGGGTCATTCTGCACATCGCAAAACGGCGAGGTTATGACGATATCAAGAAGT CGACGATAAAGAAAAGGGAGCCATTCTCAAGGCGATCAAACAGAATGAGGAAAAA TTGGCAAACTACCAGAGTGTGGGCGAGTATCTGTATAAAGAGTATTTCCAGAAGT TTAAGGAAAACAGCAAGGAGTTTACAAACGTCAGAAATAAAAGGAGTCTTACGA GAGATGCATCGCGCAGTCATTCTCAAAGATGAGCTGAAGCTGATATTTAAGAAG CAACGCGAATTTGGTTTCTCATTCTCTAAGAAGTTCGAAGAGGAGGTTCTTTCCGT GGCGTTTACAAGAGGGCGCTCAAAGACTTCTCCACCTGGTTGGTAACTGTAGT TTCTTACGGATGAGAAGCGAGCTCCCCAAAATTCTCCCCTGGCTTTCATGTTTGT TGCCCTGACTCGGATCATTAACTGCTGAACAACCTGAAAAATACTGAAGGGATC TTGTATACGAAGGACGACCTAAATGCACTCCTGAATGAAGTGTCAAAAACGGAA CTCTAACCTATAAACAGACCAAGAAATTACTGGGGCTCTCTGACGACTACGAGTT CAAGGGCGAGAAGGGTACTTATTTTATCGAATTCAAAAAGTATAAGGAGTTCATTA AAGCATTGGGGGAACACAACCTCAGCCAGGACGATCTCAATGAAATTGCCAAGGA CATCACGCTGATTAAAGACGAGATAAACTGAAAAAGGCACTGGCCAAGTATGAC

CTCAACCAGAACCAGATCGACTCTCTGTCCAAGCTGGAGTTCAAAGACCACCTAA
ACATATCCTTCAAAGCCCTGAAACTGGTCACCCCTCTAATGCTCGAAGGAAAAAA
TACGACGAGGCGTGTAACTGAATCTTAAGGTGGCCATCAATGAGGATAAGA
AGGACTTTCTCCAGCCTTTAACGAGACATATTACAAAGACGAGGTCACAAACCC
GGTTGTGCTGAGGGCCATAAAAGAGTATCGGAAGGTTCTGAATGCCCTCCTGAAG
AAGTACGGCAAAGTGCACAAAATAAATATCGAATTGGCTAGGGAGGTGGGGAAGA
ACCATTCTCAGCGAGCAAAGATCGAGAAAGAGCAGAATGAGAACTACAAAGCCAA
GAAAGACGCCGAAGTGGAGTGCGAAAAGCTGGGGCTTAAATAAACAGTAAAAAC
ATCCTGAAATTAAGATTGTTCAAAGAGCAAAAGGAGTTTTGCGCCTACTCAGGGG
AAAAATCAAATATCAGACCTGCAGGACGAGAAAATGCTGGAGATCGACCACAT
CTATCCGTATAGCAGGTCATTTGACGATTCCTACATGAACAAAGTCTTGTGTTTA
CCAAACAGAACCAAGAAAAGCTGAACCAAAACCCCTTTGAGGCTTTCGGAACGA
CTCAGCCAAGTGGCAGAAAATCGAAGTCCTAGCCAAGAATCTGCCTACAAAAAA
CAAAGAGGATTCTTGATAAGAACTATAAGGACAAGGAACAGAAAACTTTAAAGA
CAGGAACCTGAATGACACGAGGTACATTGCGCGACTGGTTCTAACTATACCAA
GACTACCTGGATTTCCTCCCTCTGAGCGACGACGAGAATACTAACTGAATGATA
CCCAGAAAGGCTCAAAGGTCCACGTTGAGGCTAAGTCCGGGATGCTGACTAGCG
CCCTCCGCCACACGTGGGGCTTCAGCGCCAAAGATCGGAATAATCATCTTCATCA
CGCTATTGATGCAGTAATCATAGCCTACGCTAACAAACAGCATCGTGAAAGCCTTCT
CCGATTTCAAGAAAGAACAGGAGTCTAATAGCGCCGAGTTGTACGCCAAGAAAAT
TTCCGAATTGGACTATAAAAAATAAGAGAAAATTCTTCGAACCCTTCTCCGGGTTTC
GCCAAAAGGTCTTAGATAAGATCGACGAGATTTTCGTTTCAAGCCCCGAAAGAA
AAAGCCTTCAGGGGCACTGCACGAAGAGACATTCCGCAAGGAAGAGGAATTTTAC
CAATCTTACGGTGGTAAAGAGGGAGTTCTGAAGGCTCTGGAGCTTGGGAAGATC
CGCAAGGTAAACGGGAAAATCGTGAAAAACGGGGACATGTTTCAGGGTGGATATC
TTCAAGCACAAAAGACCAACAAGTTCTACGCAGTACCCATCTACACTATGGATTT
CGCTTTAAAGGTTCTCCCAAATAAGGCGGTGGCTCGATCGAAGAAAGGAGAGATC
AAGGACTGGATCTTAATGGATGAAAATTACGAGTTTTGCTTCTCGCTCTACAAAGA
TAGCCTGATTCTGATCCAGACAAAAGACATGCAGGAACCAGAATTTGTTTATTATA
ACGCCTTCACGAGCAGTACAGTGTCCCTGATTGTGAGCAAGCATGATAACAAGTT
CGAGACTCTGTCTAAGAATCAGAAAATCCTTTTCAAGAACGCCAACGAGAAGGAG
GTCATCGCAAAGTCAATTGGCATCCAAAACCTGAAGGTGTTGAGAAATACATAG
TGTCCGCACTCGGTGAAGTAACTAAAGCCGAATTTGACAGCGCGAGGATTTTAA
GAAAtaa
TCGAGCACCACCACCACCACCTGAGATCCGGCTGCTAACAAAGCCCCGAAAGG
AAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG

CCTCTAAACGGGTCTTGAGGGGTTTTTGTCTGAAAGGAGGAACTATATCCGGATT
GGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGT
TACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTTCGC
TTTCTTCCCTTCCTTTCTCGCCACGTTGCGCCGGCTTTCCCGTCAAGCTCTAAATC
GGGGGCTCCCTTTAGGGTTCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAA
ACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTTTT
CGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGA CTCTTGTTCCAAACTGG
AACAACACTCAACCCTATCTCGGTCTATTCTTTTGATTATAAGGGATTTTGCCGAT
TTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAA
CAAAATATTAACGTTTACAATTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAA
CCCTATTTGTTTATTTTCTAAATACATTCAAATATGTATCCGCTCATGAATTAATT
CTTAGAAAACTCATCGAGCATCAAATGAACTGCAATTTATTCATATCAGGATTAT
CAATACCATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGG
CAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAA
CATCAATACAACCTATTAATTTCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAAT
CACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAAGTTTATGCATTTCTTTC
CAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAA
CCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCT
GTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCC
AGCGCATCAACAATATTTTACCTGAATCAGGATATTCTTCTAATACCTGGAATGC
TGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATA
AAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCA
TCTCATCTGTAACATCATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCT
GGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCGACAT
TATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGC
GGCCTAGAGCAAGACGTTTCCCGTTGAATATGGCTCATAACACCCCTTGATTACT
GTTTATGTAAGCAGACAGTTTTATTGTTTCATGACCAAAATCCCTTAACGTGAGTTTT
CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGG
TGTTTTGTTTGCCGGATCAAGAGTACCAACTCTTTTTCCGAAGGTAAGTGGCTTC
AGCAGAGCGCAGATACCAAACTGTCTTCTAGTGTAGCCGTAGTTAGGCCACC
ACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCA
GTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA CTCAAGACGA
TAGTTACCGGATAAGGCGCAGCGGTGCGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCT
ATGAGAAAGCGCCACGTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG

CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGAAACGCCT
GGTATCTTTATAGTCCTGTGCGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTG
TGATGCTCGTCAGGGGGGCGGAGCCTATGAAAAACGCCAGCAACGCGGCCTTT
TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATC
CCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC
CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCG
CCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTACACCCGCATATATG
GTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTC
CGCTATCGCTACGTGACTGGGTCTGCTGCTGCGCCCCGACACCCGCCAACACCCG
CTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTG
TGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTACCGTCATCACCGAAACG
CGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTACAGAT
GTCTGCCTGTTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAGAAGCGTTAATGTC
TGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTCTGTTTGGTCACTG
ATGCCTCCGTGTAAGGGGGATTTCTGTTTCATGGGGGTAATGATACCGATGAAACG
AGAGAGGATGCTCACGATACGGGTACTGATGATGAACATGCCCGGTTACTGGAA
CGTTGTGAGGGTAAACAACCTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATC
ACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTA
GCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACT
TCCGCGTTTCCAGACTTTACGAAACACGGAACCGAAGACCATTTCATGTTGTTGC
TCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTGCTCGCGTATCGGT
GATTCACTCTGCTAACAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCTCAAC
GACAGGAGCACGATCATGCGCACCCGTGGGGCCGCCATGCCGGCGATAATGGC
CTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAG
GGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCA
GCGAAAGCGGTCTCGCCGAAATGACCCAGAGCGCTGCCGGCACCTGTCTACT
GAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCG
CGCCACCGGAAGGAGCTGACTGGGTGAAGGCTCTCAAGGGCATCGGTGAG
ATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCC
CGCTTTCAGTCGGGAAACCTGTCTGCGCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTTCGTATTGGGCGCCAGGGTGGTTTTTCTTTTACCAG
TGAGACGGGCAACAGCTGATTGCCCTTACCGCCTGGCCCTGAGAGAGTTGCAG
CAAGCGGTCCACGCTGGTTTGCCCCAGCAGGCGAAAATCCTGTTTGATGGTGGT
TAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGATCCCACTACCGAG
ATATCCGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGC
GCCATCTGATCGTTGGCAACCAGCATCGCAGTGGAACGATGCCCTCATTCAGC

ATTTGCATGGTTTGTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCG
CTATCGGCTGAATTTGATTGCGAGTGAGATATTTATGCCAGCCAGCCAGACGCAG
ACGCGCCGAGACAGAACTTAATGGGCCCCGCTAACAGCGCGATTTGCTGGTGACC
CAATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATA
ATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAAATAACGCCGGAACATTAG
TGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGAT
CAGCCCCTGACGCGTTGCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTC
GACGCCGCTTCGTTCTACCATCGACACCACACGCTGGCACCCAGTTGATCGGC
GCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGA
GGTGGCAACGCCAATCAGCAACGACTGTTTGCCCCGCCAGTTGTTGTGCCACGCG
GTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTTC
GCAGAAACGTGGCTGGCCTGGTTCACCACGCGGAAACGGTCTGATAAGAGACA
CCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCTGAA
TTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGTTTTGCGCCATTTCG
ATGGTGTCCGGGATCTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAG
CCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGC
AAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCAC
GCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGT
GATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGG
CCACGATGCGTCCGGCGTAGAGGATCG

Table 3-2 Oligonucleotides

Oligonucleotides used to generate the CjCas9 expression vector (pE-SUMO-CjCas9, with the TEV recognition site between His ₆ -SUMO and CjCas9)		
PCR template	Forward primer	Reverse primer
pE-SUMO (modified)	ACTCGAGCACCAACCACC	CATATGGGATCCTTGGAAGTACAAG
pColdGST-CjCas9	CAAGGATCCCATATGGCCCGCATCCTCGCTTTC	TGGTGGTGCTCGAGTTTATTTCTTAAATCCTCGCGCTG
Oligonucleotides used to introduce the CjCas9 mutations (pE-SUMO-CjCas9)		
Mutation	Forward primer	Reverse primer
ΔHNH	TACATTGCGCGACTGGTTC	CCCCCGGACCCCCCAATTCGATATTTATTTGTGCAC
D8A	GCGATCGGAATCTCTAGTATCGGATG	GAAAGCGAGGATGCGGGC
H559A	GCGATCTATCCGTATAGCAGGTCATTGAC	GTCGATCTCCAGCATTTTCTCGTC
T791A	GAAGAGGCATTCCGAAGGAAGAGG	GCGGAATGCCTTCTCGTCAGTGC
R866A	GTGGCTGCGTCGAAGAAAGGAGAGATC	TCTTCGACGAGCCACCGCCTTATTTGG
T913A	AGCAGTACAGTGTCCCTGATTGTGAGCAAGC	CGCGAAGGCGTTATAATAACAATCTGGTTCC
S915A	ACAGTGTCCCTGATTGTGAGCAAGC	CGCGCTCGTGAAGGCGTTATAATAACAATCTG
S951A	GCAAAGGCAATTGGCATCCAAAACCTG	GCCTAATGCCTTTGCGATGACCTCCTT
DNA oligonucleotides used for crystallization		
PAM sequence	Target DNA strand	Non-target DNA strand
AGAAACA	CTGTTTCTGCCAAGCGCACCTAATTTC	AGAAACAG
AGAAACC	CGGTTTCTGCCAAGCGCACCTAATTTC	AGAAACCG
sgRNA		
Cj sgRNA 1-93	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCTAA AACCGC	
Cj sgRNA A63U	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGTCTCTGCGGGGTTACAATCCCTAA AACCGC	
Cj sgRNA A76U	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTGCGGGGTTCAATCCCTAA AACCGC	
Cj sgRNA U80A	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTGCGGGGTTACAAACCCCTAA AACCGC	
Cj sgRNA GGCG	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCTAA AAGGCG	
Cj sgRNA 1-85	GGAAATTAGGTGCGCTTGGCGTTTTAGTCCCTGAAAAGGGACTAAAATAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCT	
Sp sgRNA	GGAAAUUAGGUGCGCUUGGCGUUUAGAGCUAGAAUAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUG GCACCGAGUCGGUGCUU	

Oligonucleotides used to introduce the sgRNA mutations (pUC119-sgRNA)		
Mutation	Forward primer	Reverse primer
A63U	TCTCTGCGGGTTACAATCCCCTAAAACCGC	CCCGCAAACCTTTATTTTAGTCCCTTTTC
A76U	TCAATCCCCTAAAACCGCTTGTACCGAGCTCG	AACCCCGCAGAGTCCCGCAAACCTTTATTTAG
U80A	ACCCCTAAAACCGCTTGTACCGAGCTCGAATTCAC	TTGTAACCCCGCAGAGTCCCGCAAACCTC
GGCG	GGCGTTGTACCGAGCTCGAATTCAGTGGCCGTCG	TTTAGGGGATTGTAACCCCGCAGAGTCCCGC
Oligonucleotides used to generate the dsDNA templates for <i>in vitro</i> transcription		
sgRNA	Forward primer	Reverse primer
1-93, A63U, A76U	AGCGCCCAATACGCAAAC	GCGGTTTTAGGGGATTG
U80A	AGCGCCCAATACGCAAAC	GCGGTTTTAGGGGTTTG
GGCG	AGCGCCCAATACGCAAAC	CGCCTTTTAGGGGATTG
1-85	AGCGCCCAATACGCAAAC	AGGGGATTGTAACCCCGCAGAGTCCC

Table 3-3 Buffer composition

Name	Composition
Buffer A	20 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole-HCl pH 8.0, 3 mM β -ME
Buffer B	20 mM Tris-HCl pH 8.0, 1 M NaCl, 20 mM Imidazole-HCl pH 8.0, 3 mM β -ME
Buffer C	20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM Imidazole-HCl pH 8.0, 3 mM β -ME
Buffer D	20 mM Tris-HCl pH 8.0, 1 mM DTT
Buffer E	20 mM Tris-HCl pH 8.0, 2 M NaCl, 1 mM DTT
Buffer F	10 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT

Table 3-4 Data collection and refinement statistics

	5'-AGAAACC-3' PAM (Native)	5'-AGAAACA-3' PAM (Native)	5'-AGAAACA-3' PAM (SeMet)
Data collection			
Beamline	SLS PX1	SPring-8 BL41XU	SPring-8 BL41XU
Wavelength (Å)	1.0000	1.0000	0.9791
Space group	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	104.3, 105.1, 136.4	102.0, 103.9, 134.5	102.0, 104.2, 134.6
<i>α</i> , <i>β</i> , <i>γ</i> (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)*	136.5–2.4 (2.47–2.40)	103.9–2.3 (2.36–2.30)	49.5–2.2 (2.25–2.20)
<i>R</i> _{merge}	0.150 (0.882)	0.105 (0.783)	0.075 (0.967)
<i>R</i> _{pim}	0.043 (0.529)	0.032 (0.316)	0.031 (0.396)
<i>I</i> / <i>σI</i>	13.8 (1.6)	12.4 (2.4)	14.2 (2.6)
Completeness (%)	100 (100)	99.8 (99.8)	100 (100)
Multiplicity	12.9 (13.4)	10.4 (7.0)	6.8 (6.9)
CC (1/2)	0.998 (0.757)	0.986 (0.875)	0.999 (0.866)
Refinement			
Resolution (Å)	52.6–2.4	72.8–2.3	
No. reflections	113,324	122,025	
<i>R</i> _{work} / <i>R</i> _{free}	0.187 / 0.223	0.223 / 0.268	
No. atoms			
Protein	6,135	6,051	
Nucleic acid	2,715	2,712	
Solvent	353	95	
<i>B</i> -factors (Å ²)			
Protein	53.9	58.5	
Nucleic acid	41.2	48.6	
Solvent	39.1	44.1	
R.m.s. deviations			
Bond lengths (Å)	0.0004	0.0085	
Bond angles (°)	0.729	1.113	
Ramachandran plot (%)			
Favored region	95.9	95.1	
Allowed region	4.1	4.9	
Outlier region	0	0	

*Values in parentheses are for the highest resolution shell.

Table 3-5 Reagents or resources information

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
CjCas9	This paper	N/A
CjCas9-ΔHNNH	This paper	N/A
CjCas9, various mutants	This paper	N/A
SpCas9	Nishimasu <i>et al.</i> , 2014	N/A
T7 RNA polymerase	Home made	N/A
Ammonium acetate - 1.0 M solution	Hampton Research	HR2-565
NeXtalStock 50% (w/v) PEG 2,000	QIAGEN	N/A
Critical Commercial Assays		
Zymoclean Gel DNA Recovery Kit	Zymo Research	D4007
MiSeq Reagent Kit v3 (150-cycle)	Illumina	MS-102-3001
QIAprep Spin Miniprep Kit	QIAGEN	27106
Prepacked Disposable PD-10 Columns	GE Healthcare	17085101
Deposited Data		
Atomic coordinates, CjCas9 (AGAAACA PAM)	This paper	PDB: 5X2G
Atomic coordinates, CjCas9 (AGAAACC PAM)	This paper	PDB: 5X2H
<i>In vitro</i> DNA cleavage experiments	This paper; Mendeley Data	https://data.mendeley.com/datasets/n3bjs35rjy/draft?a=489ca4d1-3db3-4576-ae4c-2e439f30ec2c
Experimental Models: Organisms/Strain		
<i>E. coli</i> Mach	Thermo Fisher Scientific	C862003
<i>E. coli</i> B834(DE3)	Novagen	69041
<i>E. coli</i> Rosetta2(DE3)	Novagen	71397
Recombinant DNA		
pE-SUMO-CjCas9	This paper	N/A
pE-SUMO-CjCas9-ΔHNNH	This paper	N/A
pE-SUMO-CjCas9, various mutants	This paper	N/A
pUC119-Cj sgRNA	This paper	N/A

pUC119-Cj sgRNA, various mutants	This paper	N/A
pUC119-T20, various PAMs	This paper	N/A
Sequence-Based Reagents		
DNA primers	This paper	Table 3-2
DNA oligos (for crystallization)	This paper	Table 3-2
CjCas9 sgRNA	This paper	Table 3-2
CjCas9 sgRNA, various mutants	This paper	Table 3-2
SpCas9 sgRNA	Nishimasu <i>et al.</i> , 2014	Table 3-2
Software and Algorithms		
DIALS	Waterman <i>et al.</i> , 2013	http://dials.lbl.gov
AIMLESS	Evans and Murshudov, 2013	http://www.ccp4.ac.uk/html/aimless.html
MOLREP	Vagin and Teplyakov, 2010	http://www.ccp4.ac.uk/html/molrep.html
Buccaneer	Cowtan, 2006	http://www.ysbl.york.ac.uk/~cowtan/buccaneer/buccaneer.html
COOT	Emsley and Cowtan, 2004	http://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot
PHENIX	Adams <i>et al.</i> , 2010	https://www.phenix-online.org
CueMol	N/A	http://www.cuemol.org
Krona	Ondov <i>et al.</i> , 2011	https://github.com/marbl/Krona/wiki
WebLogo	Crooks <i>et al.</i> , 2004	http://weblogo.threepiusone.com/
Other		
Amicon® Ultra-4 Centrifugal Filter Units - 10,000 NMWL	Millipore	UFC801024
Ni-NTA Superflow	QIAGEN	30450

HiTrap Heparin HP	GE Healthcare	17040701
HiLoad 16/600 Superdex 200 pg	GE Healthcare	28989335
Superdex 200 Increase 10/300 GL	GE Healthcare	28990944
Whatman Elutrap electroelution system	Sigma-Aldrich	WHA10447705

Chapter 4 Crystal structures of the minimal Cas9 from *Campylobacter jejuni*

4.1 Introduction

Here, the author performed functional and structural characterizations of CjCas9. By comparing the structure of this minimal Cas9 with these of the more elaborate orthologs, we gain a deeper understanding of the function of each structural element. Each Cas9 appears to recognize the tracrRNA with a unique mechanism. CjCas9 recognizes an unexpected triple-helix in its tracrRNA. The PAM recognition mechanism by CjCas9 is also quite unique, involving a set of rather weak contacts to both strands of DNA.

4.2 Materials and Methods

The final models of the 5'-AGAAACA-3' PAM complex (2.3 Å resolution) and the 5'-AGAAACC-3' PAM complex (2.4 Å resolution) were refined using native data sets. Data collection statistics are summarized in Table 3-4.

4.3 Results and Discussion

4.3.1 Crystal structure of the CjCas9-sgRNA-target DNA complex

The crystal structure revealed that CjCas9 adopts a bilobed architecture comprising the α -helical REC lobe and a NUC lobe, as in the other Cas9 orthologs (Anders et al., 2014;

Nishimasu et al., 2014; Nishimasu et al., 2015) (Figure 4-1C), indicating that the truncation of the HNH domain does not substantially affect the overall structure. The REC lobe can be divided into the REC1 (residues 77–234) and REC2 (residues 235–426) domains. The NUC lobe comprises the RuvC (residues 1–44, 427–480 and 641–777), WED (residues 792–827), and PI domains (residues 828–984) (the HNH domain was truncated for crystallization). The REC and NUC lobes are connected by an arginine-rich “bridge” helix (residues 45–76), while the WED and RuvC domains are connected by a “phosphate lock” loop (residues 778–791), as in other Cas9 orthologs (Anders et al., 2014; Nishimasu et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016). The three residues (GGS) in the GGSGGG linker between the RuvC-II and RuvC-III motifs are disordered in the present structure.

The sgRNA comprises the guide segment (G1–C20), the repeat region (G21–U32), the tetraloop (G33–A36), the antirepeat region (A37–U48), and the tracrRNA scaffold (A49–C93) (Figure 2B). The guide segment (G1–C20) and the target DNA strand (dG1–dC20) form an RNA-DNA heteroduplex (Figures 4-1B and C). The repeat and antirepeat regions form the A-form-like duplex (referred to as the repeat-antirepeat duplex), which consists of a wobble base pair (G21•U48) and eleven Watson-Crick base pairs (U22-A47–U32-A37) (Figures 4-1B and C). The RNA-DNA heteroduplex is bound within the central channel between the REC and NUC lobes, while the

repeat-antirepeat duplex is sandwiched between the REC1 and WED domains (Figure 4-1C). These duplexes are primarily recognized by the protein in a non-sequence-specific manner (Figure S2). The target DNA strand (dC(−8)–dT(−1)) and the non-target DNA strand (dA1*–dG8*) form a PAM-containing duplex, which is recognized by the WED and PI domains (Figures 4-1B and C).

4.3.2 The effect of the deletion of the HNH domain on the CjCas9 activity

A recent study showed that the deletion of the HNH domain in SpCas9 impairs the non-target strand cleavage by the RuvC domain, thereby suggesting that the HNH domain is required for the activation of the RuvC domain (Sternberg et al., 2015). The author thus examined the effect of the deletion of the HNH domain on the CjCas9 activity, using *in vitro* cleavage assays. The results revealed that, like the D8A RuvC-inactive nickase and the H559A HNH-inactive nickase, the CjCas9-ΔHNH variant functions as a nickase (Figure 4-2). Importantly, the CjCas9-ΔHNH variant exhibited lower cleavage activity, as compared with the H559A nickase, indicating that the deletion, but not the inactivating point mutation, of the HNH domain reduces the non-target strand cleavage by the RuvC domain. These observations suggested the allosteric communication between the RuvC and HNH nuclease domains in CjCas9, as observed in SpCas9 (Sternberg et al., 2015).

4.3.3 TracrRNA architecture

Notably, the present structure revealed that the CjCas9 tracrRNA scaffold contains a triple-helix structure within a pseudoknot comprising three stem regions, which was not predicted from its primary sequence (Figures 4-4A–C). Stem 1 consists of four canonical base pairs (A51-U67–G54-C64) and a non-canonical A50•G68 base pair, while stem 2 consists of four canonical base pairs (G58-C93–G61-C90) and a non-canonical G62•A89 base pair (Figures 4-4B and C). Stem 3 consists of four canonical base pairs (G70-C84–G74-C81) (Figures 4-4B and C). Nucleotides 56/57 and 86/87 form base pairs with stem 2 and stem 1, respectively, thereby contributing to the triple-helix formation. U65 in stem 1 and A89 in stem 2 base pair with A86 and U56, forming an A53-U65•A86 minor-groove triple and a G62-A89•U56 major-groove triple, respectively (Figures 4-4D and E). U57 base pairs with C59 and C90, forming a G92-C59•U57•C90-G62 base quintuple (Figure 4-4F). A87 and A88 base pair with G54 and U55/C64, respectively, forming a U55•A88•C64-G54•A87 base quintuple (Figure 4-4G).

4.3.4 TracrRNA scaffold recognition

The tracrRNA scaffold is extensively recognized by CjCas9 (Figures 4-5A and 4-3). In particular, A63, A76 and U80 are flipped out, and recognized by the protein in base-specific manners. The nucleobase and ribose moieties of A63 form stacking

interactions with the side chains of His70 and His67, respectively, while the N1 of A63 hydrogen bonds with the side chain of Asn74 (Figure 4-5B). A76 and U80 are accommodated within specific pockets in the PI domain (Figure 4-5A). The nucleobase of A76 is sandwiched between the side chains of Ile964 and Arg977, while the N6 and N7 of A76 hydrogen bond with the main-chain carbonyl and amide groups of Glu975, respectively (Figure 4-5C). The nucleobase of U80 is sandwiched between the side chains of Phe854 and Glu980, while the N3 and O4 of U80 hydrogen bond with the main-chain carbonyl group of Asp981 and the side chain of Arg832, respectively (Figure 4-5C). Indeed, the single mutations (A63U, A76U or U80A) reduced CjCas9-mediated DNA cleavage (Figure 4-5D), confirming the functional importance of the three flipped-out nucleotides. Moreover, a 4-nt substitution (nucleotides 90–93) or an 8-nt deletion (nucleotides 86–93) in the tracrRNA 3' tail abolished the CjCas9-mediated DNA cleavage (Figure 4-5D), indicating that the triple-helix structure of the tracrRNA is critical for the activity. In addition, the SpCas9 sgRNA did not support the CjCas9-mediated DNA cleavage (Figure 4-5D). Together, these results demonstrated that CjCas9 specifically recognizes its cognate RNA guide.

4.3.5 The 5'-NNNVRYM-3' PAM recognition

In the present structure, the 5'-AGAAACC-3' PAM-containing DNA duplex is bound to the cleft between the WED and PI domains (Figure 4-6A). The nucleobases of dA1*–

dA3* do not directly contact the protein (Figures 4-6B and C), consistent with the lack of specificity for positions 1–3 in the 5'-NNNVR YM-3' PAM. The N7 of dA4* in the non-target strand forms a water-mediated hydrogen bond with the side-chain hydroxyl group of Thr913 (Figures 4-6B and C). Modeling suggested that a steric clash could occur between the methyl group of dT4* and the side chain of Thr913 (Figures S3A and S3B), consistent with the preference of CjCas9 for the fourth V (A/G/C). The N7 of dA5* in the non-target strand forms a hydrogen bond with the side-chain hydroxyl group of Ser915 (Figures 4-6B and C). Since N7 is common among the purine nucleotides, the interaction can explain the requirement for the fifth R (A/G). Notably, the nucleobase of dC6* in the non-target strand is not recognized by the protein (Figures 4-6B and C). Instead, the N7 of dG(–6) in the target strand forms a hydrogen bond with the side-chain hydroxyl group of Ser951 (Figures 4-6B and C). These structural findings indicated that CjCas9 does not recognize the Y (T/C) nucleotides at position 6 in the non-target strand as the PAM, but detects their complementary R (A/G) nucleotides in the target strand. Similarly, the nucleobase of dC7* in the non-target strand is not recognized by the protein, whereas the O6 and N7 of dG(–7) in the target strand form bidentate hydrogen bonds with the side chain of Arg866 (Figures 4-6B and C).

In addition to the 5'-AGAAACC-3' PAM complex, the author determined the crystal structure of CjCas9- Δ HNH in complex with the sgRNA and the DNA target containing the 5'-AGAAACA-3' PAM (Table 3-3). In the 5'-AGAAACA-3' PAM complex, the dT(-7):dA7*-dC(-8):dG8* pairs in the PAM duplex undergo a slight displacement toward the PI domain, as compared with the dG(-7):dC7*-dC(-8):dG8* pairs in the 5'-AGAAACC-3' PAM complex (Figure S3C). This displacement in the PAM duplex allows Arg866 to form a hydrogen bond with the O4 of dT(-7) in the target strand (Figure 4-9D). These observations revealed that CjCas9 does not recognize the M (A/C) nucleotides at position 7 in the non-target strand as the PAM, but their complementary K (T/G) nucleotides in the target strand. The preference of CjCas9 for C over A at position 7 can be explained by the bidentate hydrogen-bonding interaction between dG(-7) and Arg866, in contrast to the single hydrogen-bonding interaction between dT(-7) and Arg866. The single mutations of Arg866, Thr913, Ser915, and Ser951 reduced or abolished the *in vitro* cleavage activity (Figure 4-6D), confirming their functional importance. Together, these structural and functional data revealed that CjCas9 forms sequence-specific contacts with both the target and non-target DNA strands, to achieve the recognition of the 5'-NNNVR YM-3' PAM.

4.3.6 Structural comparison between the Cas9 orthologs

A structural comparison of CjCas9 with the other Cas9 orthologs highlighted the structural similarities and differences between the CRISPR-Cas9 systems and provided insights into a minimal Cas9 scaffold (Figure 4-8). Unlike SpCas9 (Anders et al., 2014; Nishimasu et al., 2014), the smaller SaCas9 (Nishimasu et al., 2015) and CjCas9 lack the insertion subdomains within the REC1 and PI domains (Figures 4-8A–C). Furthermore, the WED domain of CjCas9 (36 amino acids) is smaller than that of SaCas9 (122 amino acids) (Figures 4-8A and C). These structural differences contribute to the miniaturization of CjCas9. The REC and WED domains of FnCas9, one of the largest Cas9 orthologs, adopt protein folds distinct from those of CjCas9, SpCas9 and SaCas9 (Hirano et al., 2016) (Figure 4-8D), reinforcing the notion that FnCas9 is distantly related to the other Cas9 orthologs.

Despite the structural differences in these individual domains, CjCas9 adopts the conserved bilobed architecture and accommodates the RNA-DNA heteroduplex in similar manners to those of the other Cas9 orthologs (Figures 4-8A–D). The sugar-phosphate backbone of the sgRNA “seed” region (C13– C20) is extensively recognized by conserved arginine residues in the bridge helix (Figure 4-10A). In addition, the backbone phosphate group between dG1 and dT(–1) in the target DNA

strand (referred to as the +1 phosphate; Anders et al., 2014) interacts with the main-chain amide groups of Glu790 and Thr791 and the side-chain hydroxyl group of Thr791 in the phosphate lock loop, thereby facilitating target DNA unwinding (Figure 4-10B). Indeed, the T791A mutation abolished the in vitro DNA cleavage activity (Figure 4-10C), confirming the functional importance of the interaction between the +1 phosphate and Thr791. These observations confirmed that the RNA-guided DNA targeting mechanism is highly conserved among the CRISPR-Cas9 systems.

The present structure also illuminated the structural diversity of the crRNA:tracrRNA guides in the CRISPR-Cas9 systems (Figures 4-8E–H). The repeat-antirepeat duplexes for SpCas9, SaCas9, and FnCas9 contain several unpaired nucleotides, and thus adopt distorted, distinct structures (Nishimasu et al., 2014, 2015; Hirano et al., 2016) (Figures 4-8F–H). In contrast, the CjCas9 repeat-antirepeat duplex adopts an A-form-like conformation (Figure 4-8E). According to these structural differences, the repeat-antirepeat duplexes are recognized by the structurally divergent REC1 and WED domains in species-specific manners (Figure 4-11). The CjCas9-REC1 adopts a conserved core fold but has two unique loops (loops 1 and 2) that interact with the repeat-antirepeat duplex. The repeat-antirepeat duplex is further recognized by the WED domain, which is structurally distinct from those of the other Cas9 orthologs.

Furthermore, the present structure revealed the notable architectural differences in the tracrRNA scaffolds. The SpCas9 and SaCas9 tracrRNA scaffolds contain three and two stem loops, respectively, and the first and second stem loops are connected by a single-stranded linker (although stem loop 2 of SaCas9 was truncated for crystallization) (Nishimasu et al., 2014, 2015) (Figures 6F and 6G). The FnCas9 tracrRNA scaffold contains two stem loops, which are connected by a U-shaped linker (Hirano et al., 2016) (Figure 4-8H). In stark contrast, the CjCas9 tracrRNA scaffold contains a more complicated triple-helix structure, as described above (Figure 4-8E).

4.3.7 Mechanistic diversity in PAM recognition

Despite their limited sequence similarity, the PI domains of the Cas9 orthologs share a similar core fold comprising two distorted β sheets ($\beta 1$ – $\beta 3$ and $\beta 4$ – $\beta 9$) (Figures 4-9A–D). In SpCas9, SaCas9, and FnCas9, distinct sets of amino acid residues in the b5–b7 region form sequence-specific contacts with the PAM nucleotides on the non-target DNA strand (Anders et al., 2014; Nishimasu et al., 2015; Hirano et al., 2016). In SpCas9, Arg1333 and Arg1335 form bidentate hydrogen bonds with the second and third Gs in the 5'-NGG-3' PAM, respectively (Anders et al., 2014) (Figure 4-9B). In SaCas9, Arg1015 forms a bidentate hydrogen bond with the third G in the

5'-NNGRRT-3' PAM, while Asn985, Asn986, and Arg991 form a hydrogen-bonding network with the RRT nucleotides (Nishimasu et al., 2015) (Figure 4-9C). In FnCas9, Arg1585 and Arg1556 form bidentate hydrogen bonds with the second and third Gs in the 5'-NGG-3' PAM, respectively (Hirano et al., 2016) (Figure 4-9D). In contrast to these Cas9 orthologs, CjCas9 forms sequence-specific contacts with the PAM nucleotides on the non-target strand and the PAM-complementary nucleotides on the target strand (Figure 4-9A), illuminating the mechanistic diversity of Cas9-mediated PAM recognition. Intriguingly, a recent study showed that the mutations of the PAM-complementary nucleotides on the target strand abolished the cleavage activity of *Neisseria meningitidis* Cas9 (Zhang et al., 2015), suggesting that other Cas9 orthologs, such as *N. meningitidis* Cas9, also form sequence-specific interactions with both the target and non-target DNA strands in the PAM duplex, as observed in CjCas9. Further studies will be required to fully elucidate the mechanistic diversity in the PAM recognition by Cas9 orthologs.

4.3.8 Structural comparison of the CjCas9 tracrRNA and other known RNA Triplexes

Structural Comparison of the CjCas9 tracrRNA and Other Known RNA Triplexes A structural comparison of the CjCas9 tracrRNA with previously characterized RNA triplexes, such as the telomerase RNA subunit TER (Theimer et al., 2005), the SAM-II riboswitch (Gilbert et al., 2008), and the long noncoding RNA MALAT1 (Brown et al., 2014), revealed notable differences between the CjCas9 tracrRNA and the other RNA triplexes (Figure 4-12). Notably, the CjCas9 tracrRNA lacks a canonical U • A-U triple, whereas the other three RNA triplexes contain successive U • A-U triples (Figure 4-12). TER and SAM-II have three and seven U • A-U triples in their core regions, respectively. In particular, MALAT1 forms a bipartite triple helix containing stacks of five and four U • A-U triples. Moreover, in the CjCas9 tracrRNA, the L1 region between the S1 and S2 regions consists of three nucleotides and is shorter than those of the other three RNA triplexes (Figure 4-12). Thus, the CjCas9 tracrRNA adopts a distorted structure, which is stabilized by the two base triples and the two base quintuples. Taken together, the CjCas9-sgRNA-DNA structure revealed the previously unrecognized biological role of an RNA triple helix.

4.4 Discussion and Perspectives

The present structure of the CjCas9-sgRNA-target DNA complex unveiled the remarkable diversity among the CRISPR-Cas9 systems. First, unlike the other tracrRNAs, the CjCas9 tracrRNA contains a triple-helix architecture, which is distinct from other known RNA triplexes, thereby expanding the natural repertoire of RNA triplexes. Second, unlike other Cas9 orthologs, CjCas9 reads the nucleotide sequences in both the target and non-target DNA strands as the PAM. Although CjCas9 has not been harnessed for genome editing applications, these structural findings may provide clues for Cas9 engineering, such as miniaturization of Cas9 and alteration of its PAM specificities.

Figures and Tables of Chapter 4

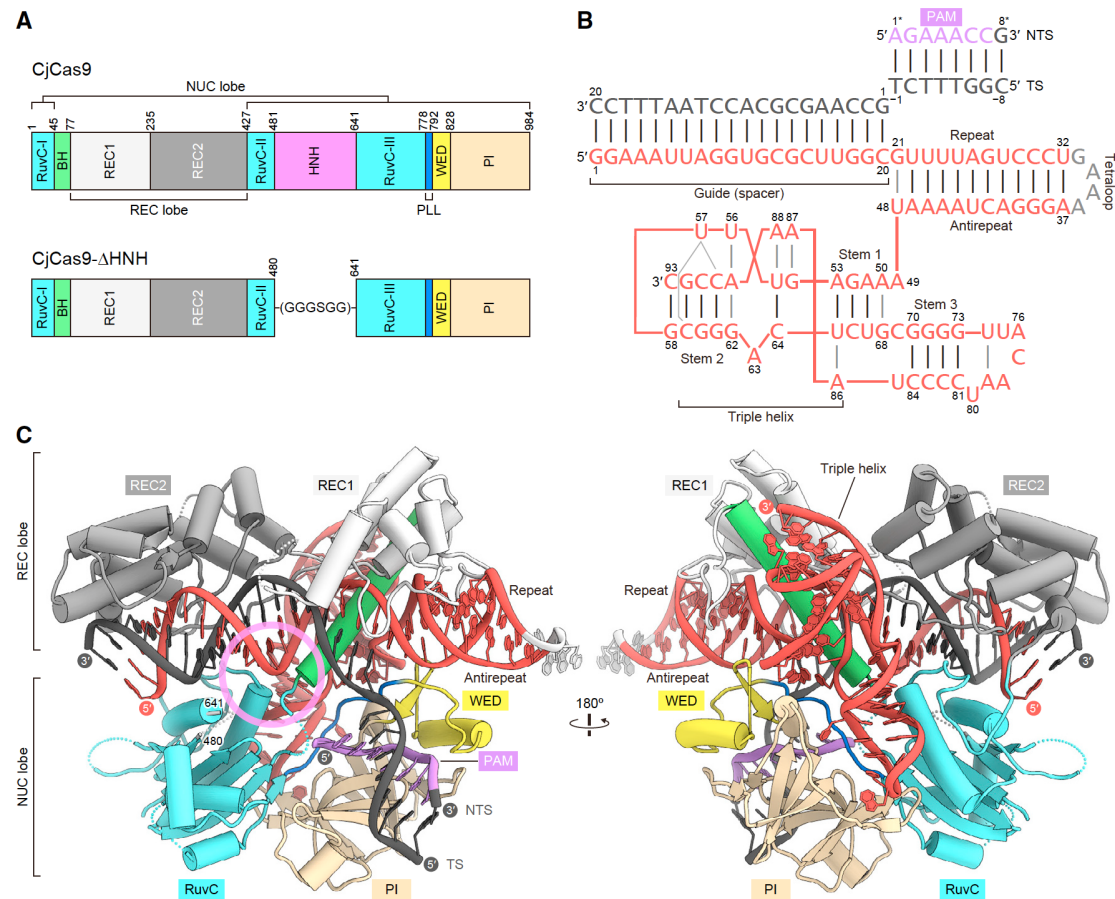


Figure 4-1 Overall structures

(A) Domain structure of CjCas9. The HNH nuclease domain was truncated for crystallization. BH, bridge helix; PLL, phosphate lock loop.

(B) Schematics of the sgRNA and the target DNA. TS, target strand; NTS, non-target strand.

(C) Overall structure of CjCas9-ΔHNH in complex with an sgRNA and its target DNA. The predicted location of the HNH domain is indicated by the pink circle.

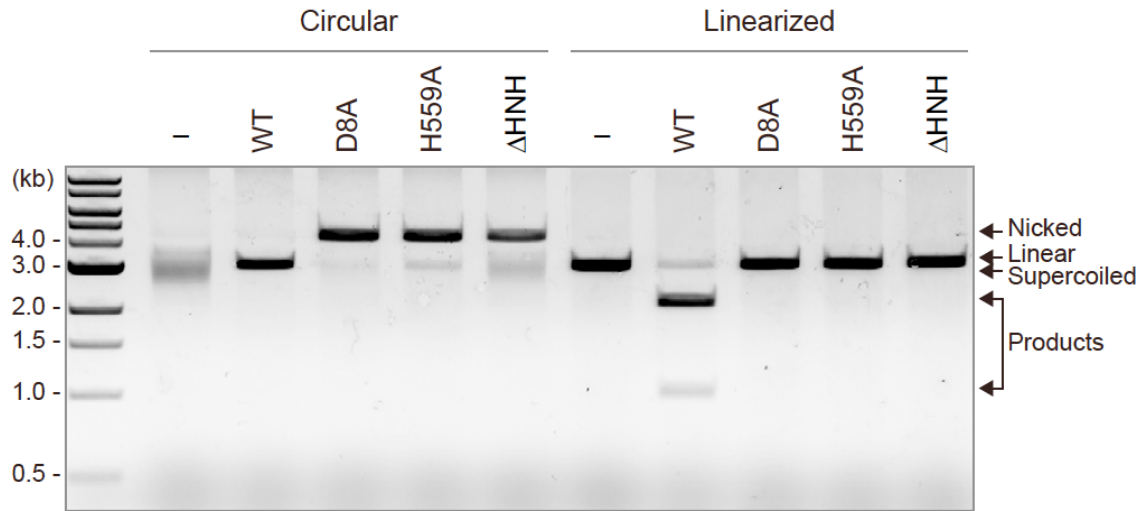


Figure 4-2 *In vitro* cleavage activity of CjCas9- Δ HNH

The circular and linearized plasmid targets with the 5'-AGAAACC-3' PAM were incubated with wild-type CjCas9 or the three CjCas9 variants (D8A, H559A, and Δ HNH), and then were analyzed by agarose gel electrophoresis. The D8A and H559A variants of CjCas9 correspond to the D10A and H840A nickases of SpCas9, respectively.

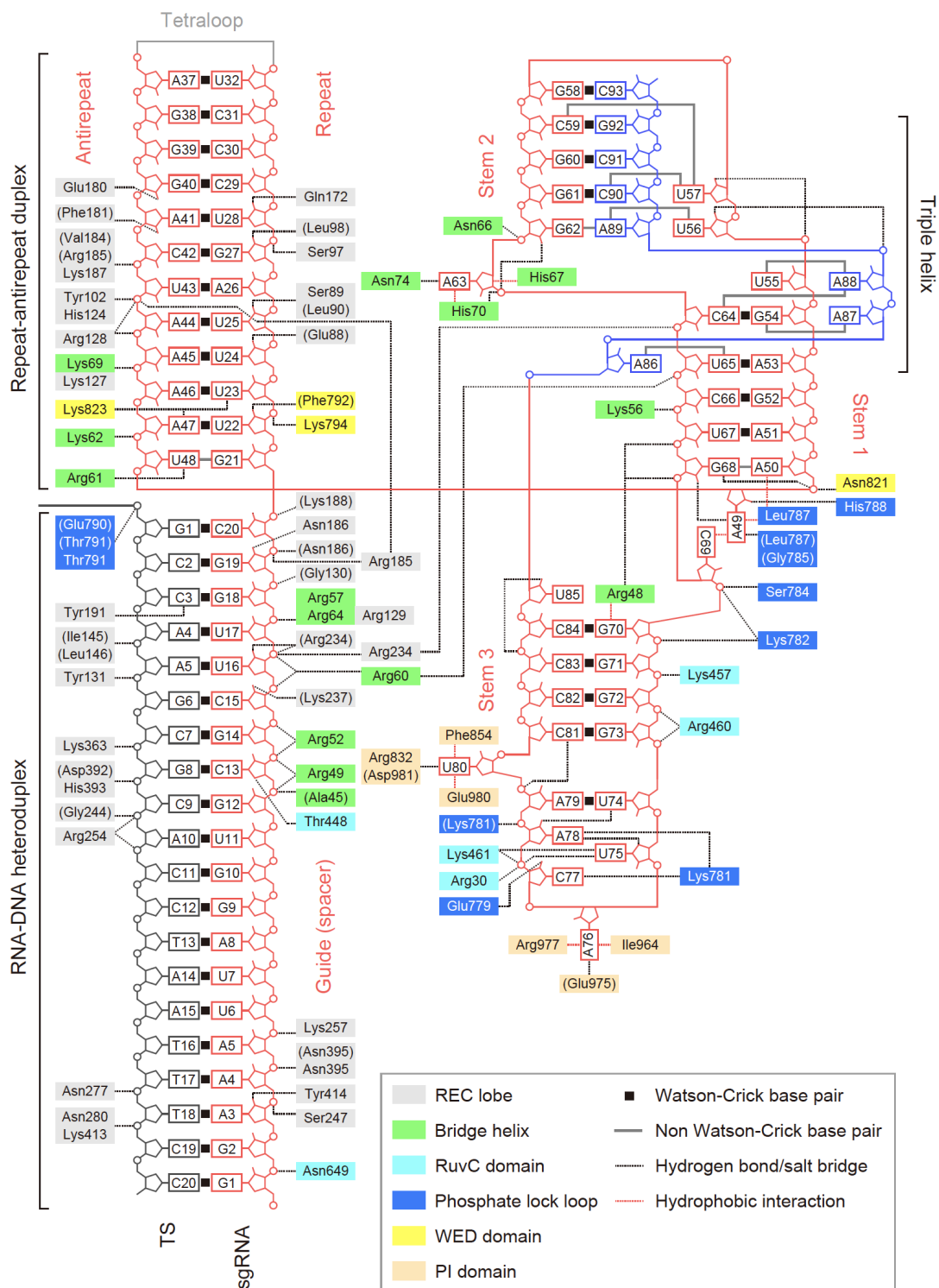


Figure 4-3 Schematic of the nucleic-acid recognition

Residues that interact with nucleic acids via their main chain are shown in parentheses.

Water-mediated hydrogen bonds are omitted for clarity.

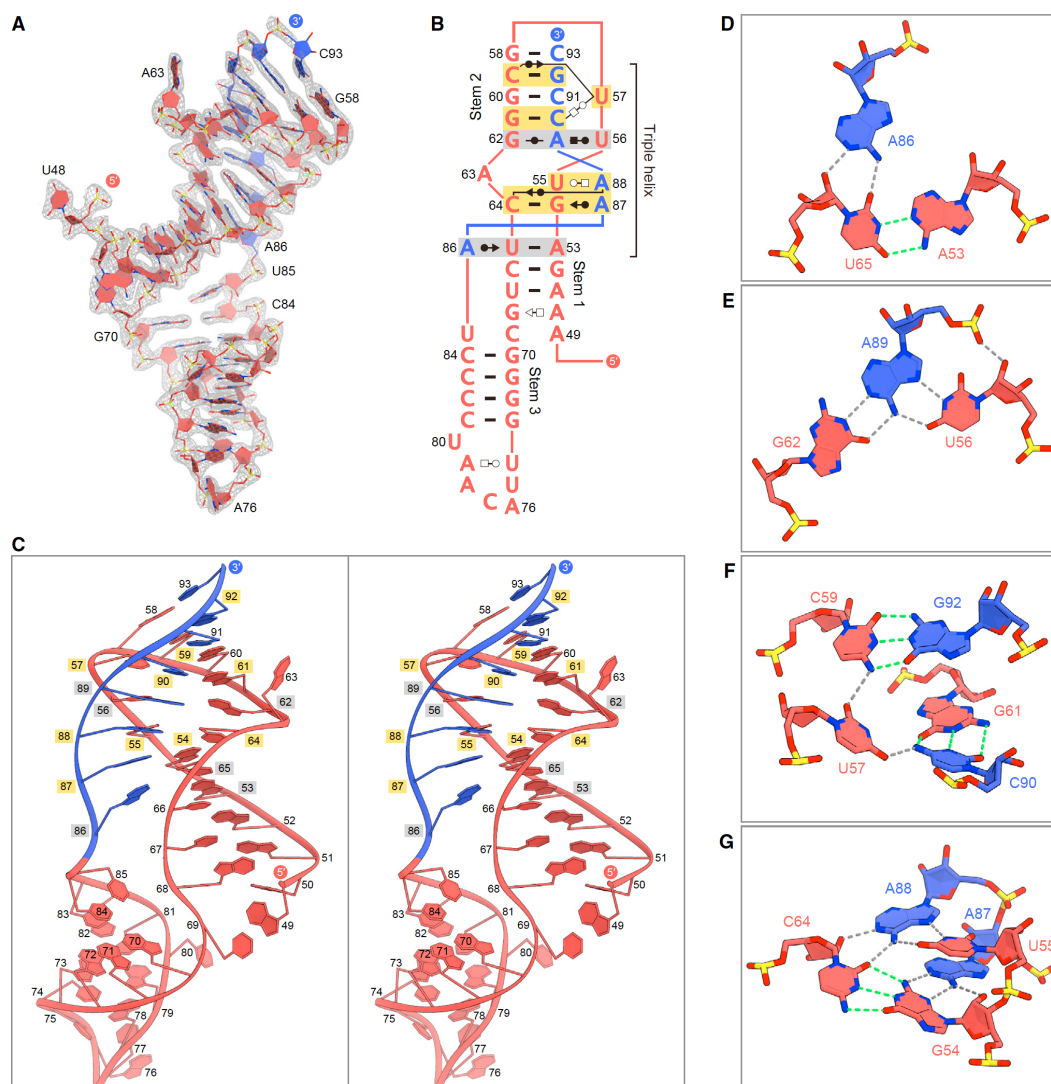


Figure 4-4 tracrRNA structure

(A) $2mF_o - DF_c$ electron density map for the tracrRNA scaffold (contoured at 2σ).

(B) Schematics of the tracrRNA scaffold. Non-Watson-Crick base pairs are indicated with Leontis-Westhof notations (Leontis et al., 2002). The base triples and quintuples are highlighted in gray and orange backgrounds, respectively.

(C) Structure of the CjCas9 tracrRNA scaffold (stereo view). The 3' nucleotides involved in the triple-helix formation are colored blue. Nucleotides involved in the formation of the base triple and quintuple are highlighted in gray and orange backgrounds, respectively.

(D–G) Base triples (D and E) and base quintuples (F and G) in the triple helix. Hydrogen bonds between canonical and non-canonical base pairs are depicted with green and gray dashed lines, respectively.

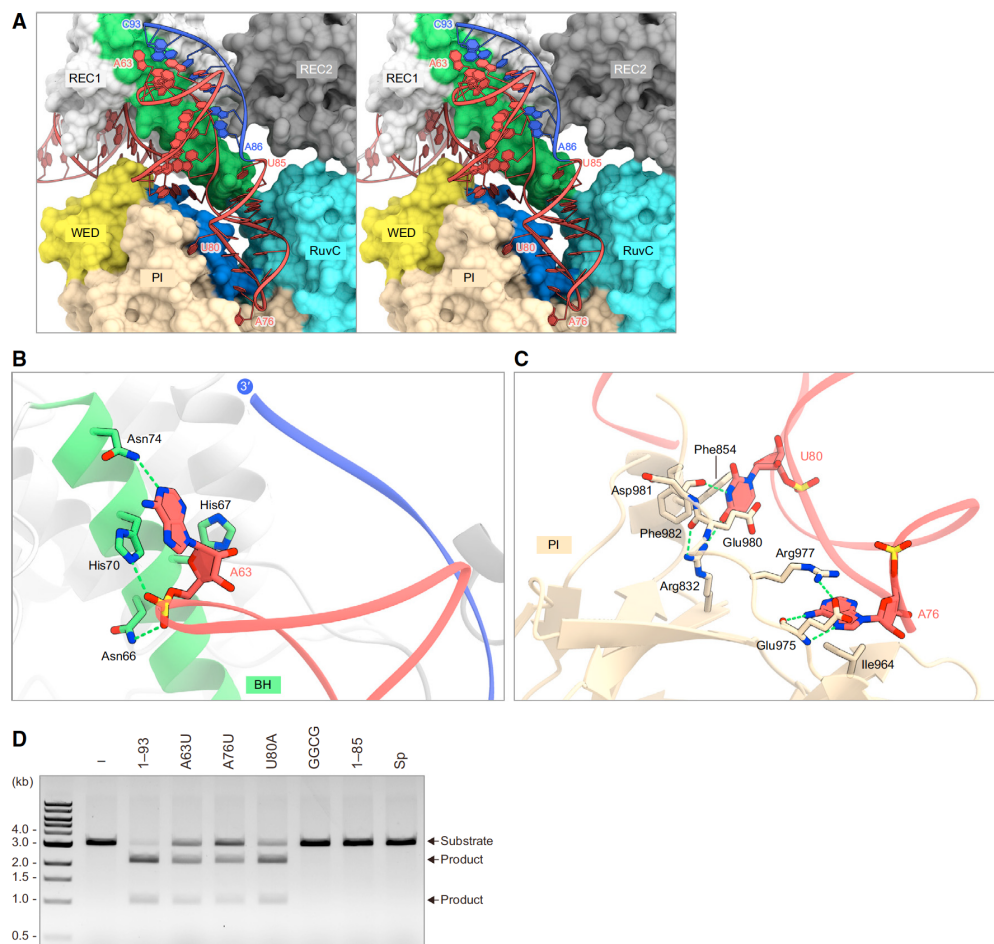


Figure 4-5 TracrRNA scaffold recognition

(A) Binding of the tracrRNA scaffold to CjCas9 (stereo view).

(B and C) Specific recognition of A63 (B) and A76/U80 (C). Hydrogen bonds are depicted by dashed lines.

(D) Functional importance of the tracrRNA scaffold. The linearized plasmid target with the 5'-AGAAACC-3' PAM was incubated with CjCas9, together with either the full-length sgRNA (nucleotides 1–93), the sgRNA variants, or the SpCas9 sgRNA. 1–93, the full-length sgRNA; GGCG, the sgRNA variant, in which nucleotides 90–93 (CCGC) were replaced with GGCG; 1–85, the sgRNA variant, in which nucleotides 86–93 were truncated; Sp, the SpCas9 sgRNA.

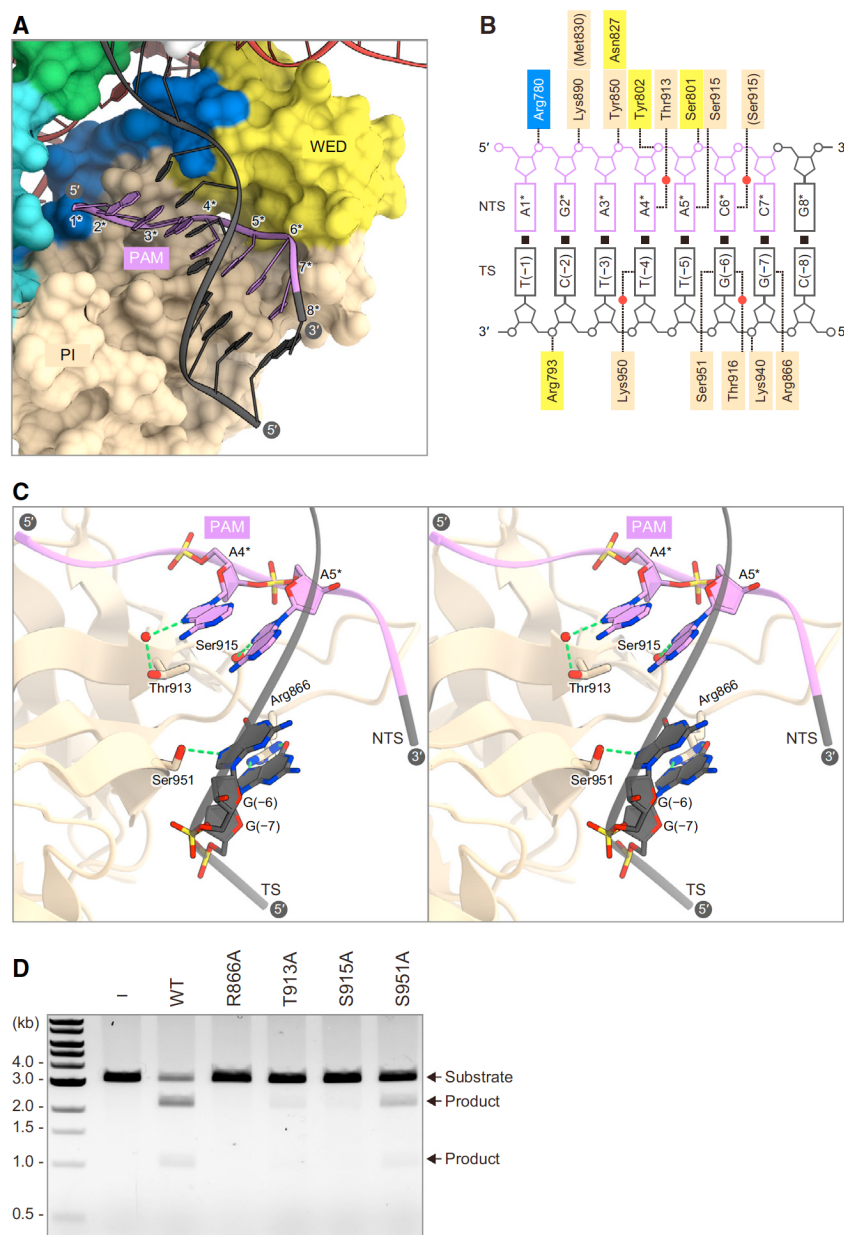


Figure 4-6 PAM recognition

(A) Binding of the PAM duplex to CjCas9.

(B) Schematics of the PAM recognition by CjCas9. Hydrogen bonds are depicted by dashed lines. Water molecules are shown as red spheres. Water-mediated hydrogen bonds between the protein and the sugar-phosphate backbone are omitted for clarity.

(C) Recognition of the 5'-AGAAACC-3' PAM by CjCas9 (stereo view). Water molecules are shown as red spheres. Hydrogen bonds are depicted by dashed lines.

(D) Functional importance of the PAM-interacting residues. The linearized plasmid target with the 5'-AGAAACC-3' PAM was incubated with either the wild type or the mutants of CjCas9 in the presence of the sgRNA.

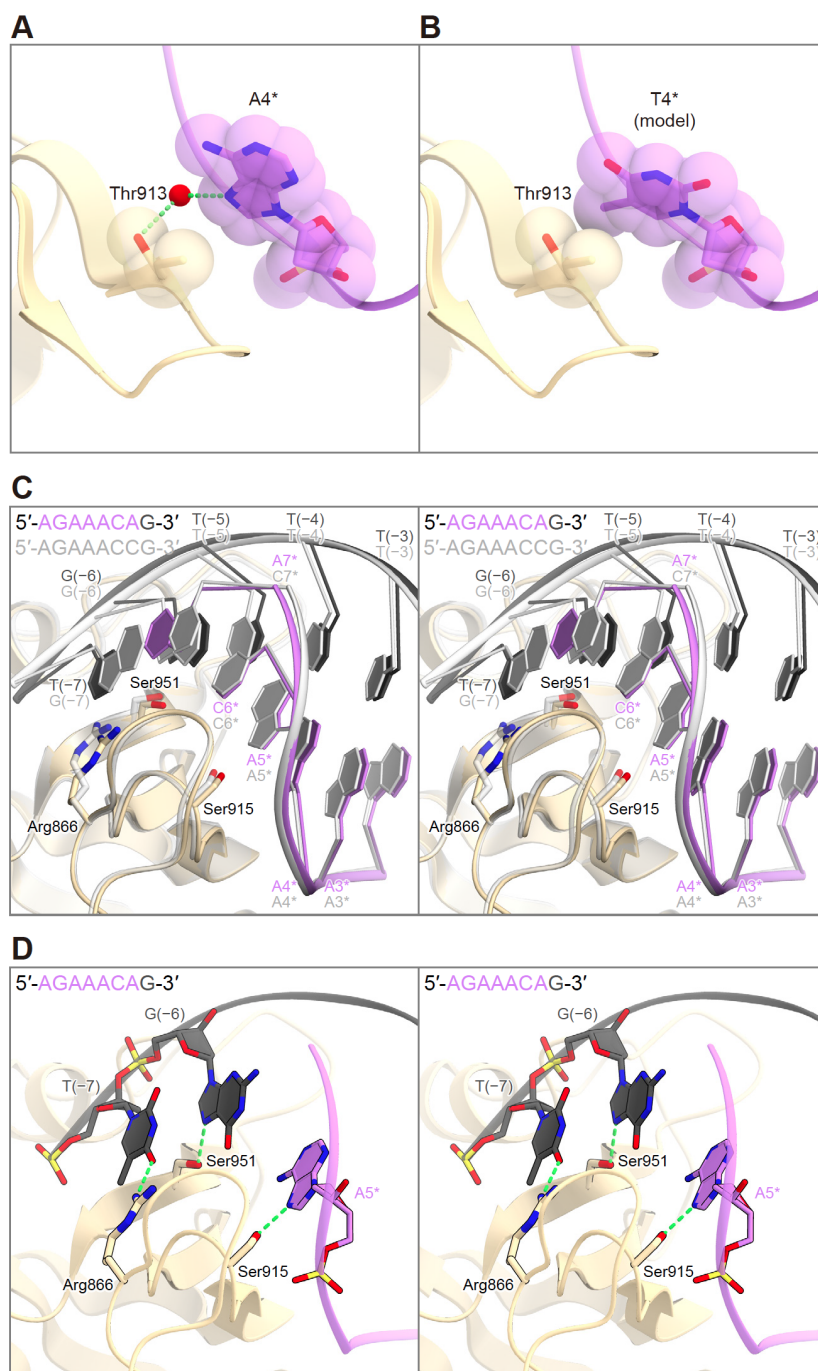


Figure 4-7 PAM recognition

(A and B) Interactions between Thr913 and dA4* (A) and modeled dT4* (B). (C) Superimposition of the 5'-AGAAACCG-3' PAM complex (semitransparent gray) and the 5'-AGAAACA-3' PAM complex (colored) (stereo view). (D) Recognition of the 5'-AGAAACA-3' PAM (stereo view). Hydrogen bonds are depicted by dashed lines. Related to Figure 4-6.

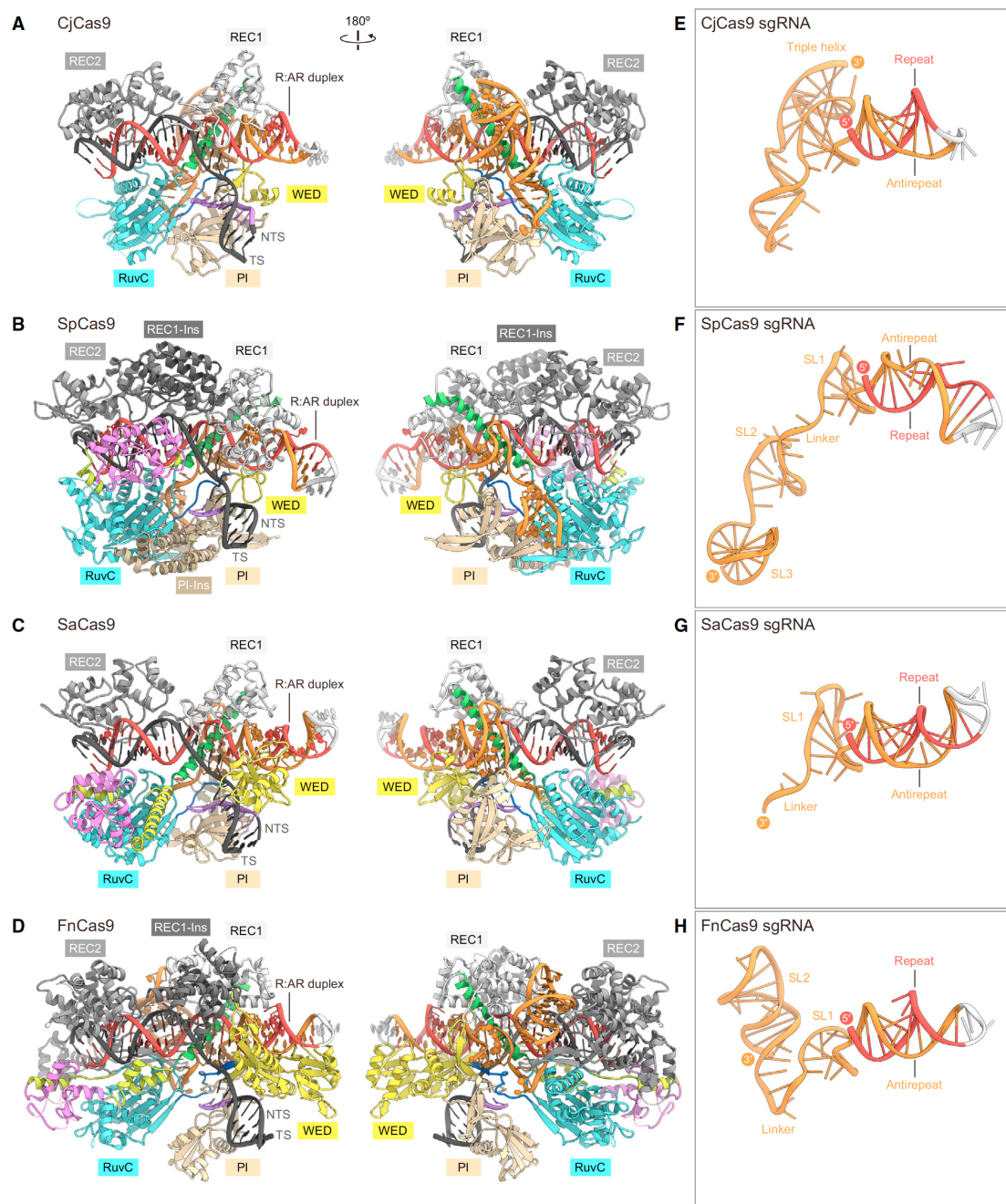


Figure 4-8 Structural comparison of the Cas9 orthologs

(A–D) Structures of CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D) in complexes with an sgRNA and its target DNA. SpCas9 and FnCas9 have structurally distinct subdomains inserted within their REC1 domains (previously referred to as the REC2 domains (Nishimasu et al., 2014; Hirano et al., 2016)).

(E–H) Structures of the sgRNAs for CjCas9 (E), SpCas9 (PDB: 4OO8) (F), SaCas9 (PDB: 5CZZ) (G), and FnCas9 (PDB: 5B2O) (H). The guide segments are omitted for clarity.

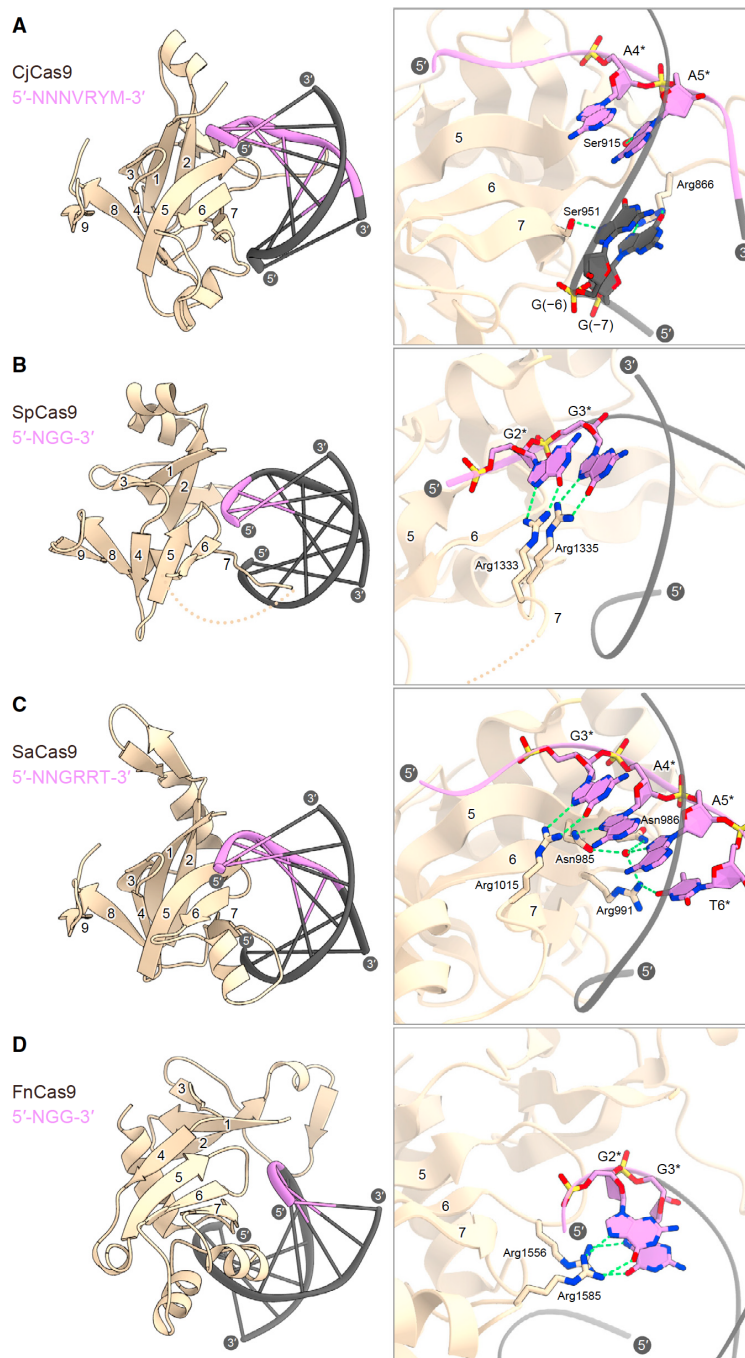


Figure 4-9 PAM recognition by the Cas9 orthologs

(A–D) PAM recognition by CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D). In (B), the subdomain inserted between $\beta 6$ and $\beta 7$ is omitted for clarity. The PAMs are highlighted in purple. The conserved core β strands are numbered. Close-up views for the PAM recognition are shown on the right.

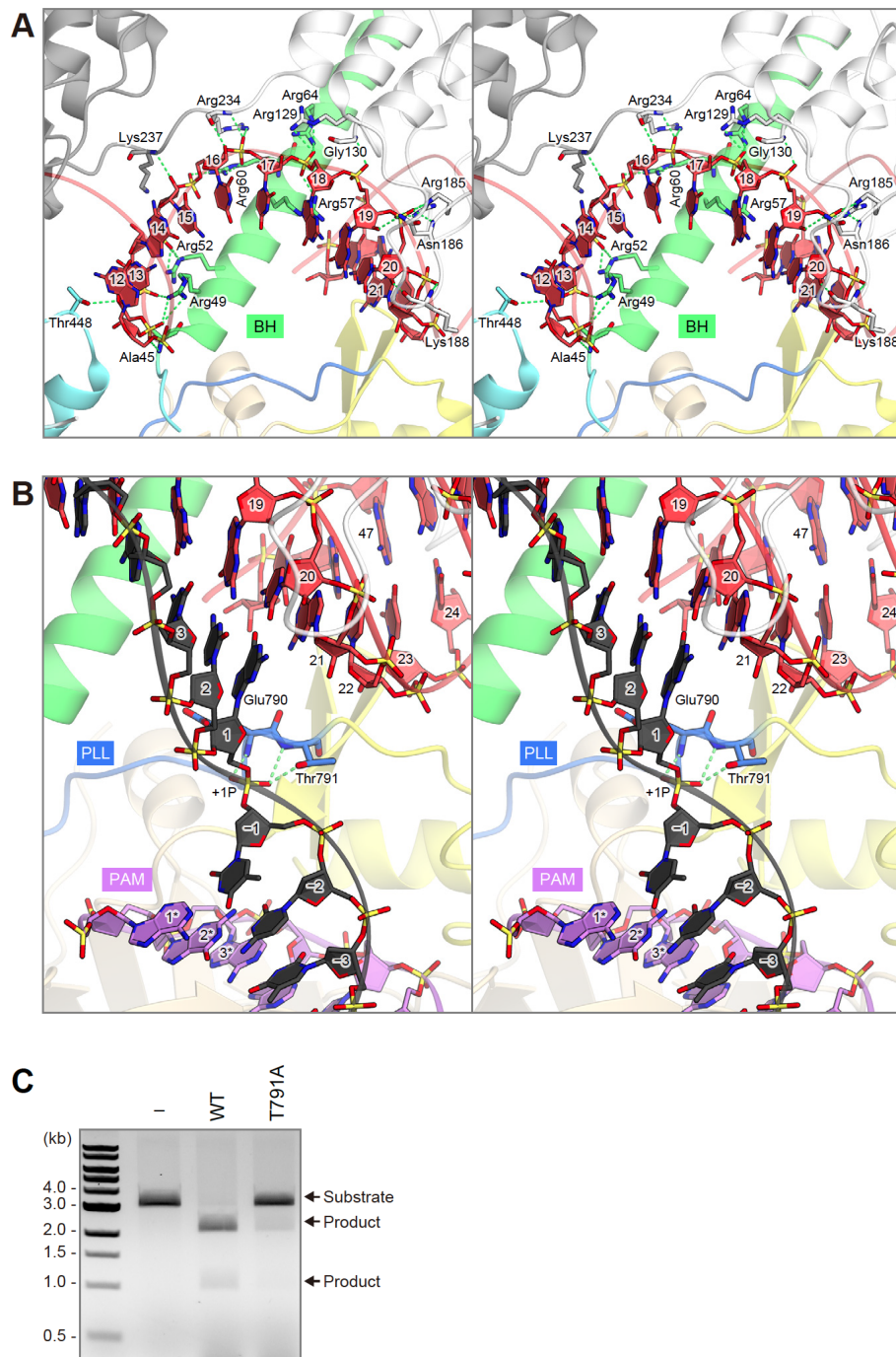


Figure 4-10 RNA-guided DNA targeting

(A) Recognition of the sgRNA seed region by CjCas9 (stereo view). The target DNA strand is omitted for clarity. (B) Recognition of the +1 phosphate by CjCas9 (stereo view). (C) Functional importance of the phosphate lock loop. The linearized plasmid target with the 5'-AGAAACC-3' PAM was incubated with either the wild type or the T791A mutant of CjCas9, together with the sgRNA. Related to Figure 4-8.

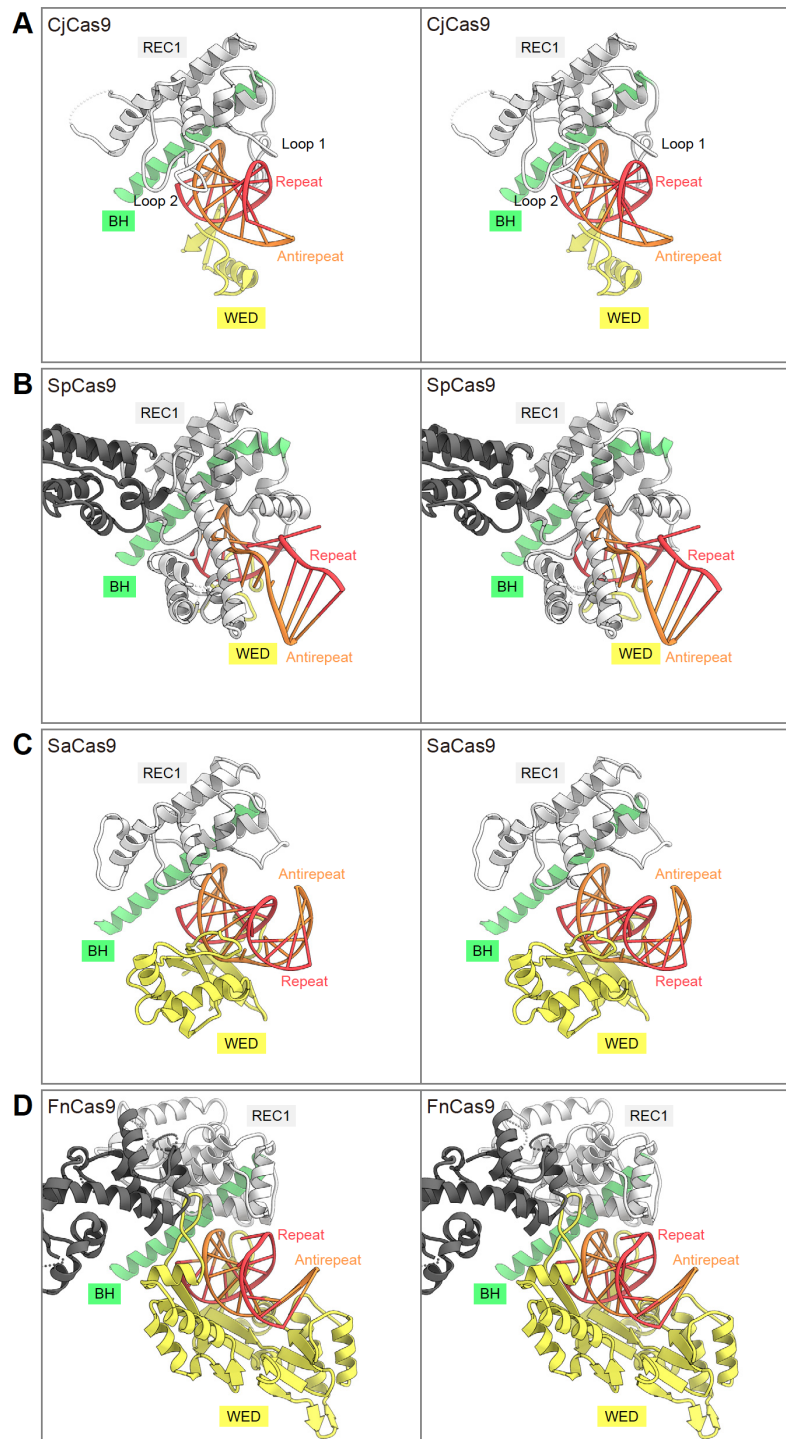


Figure 4-11 Recognition of the repeat-antirepeat duplex by the Cas9 orthologs
 (A–D) Species-specific recognition of the repeat-antirepeat duplex by the REC1/WED domains of CjCas9 (A), SpCas9 (PDB: 4UN3) (B), SaCas9 (PDB: 5CZZ) (C), and FnCas9 (PDB: 5B2O) (D). Related to Figure 4-8.

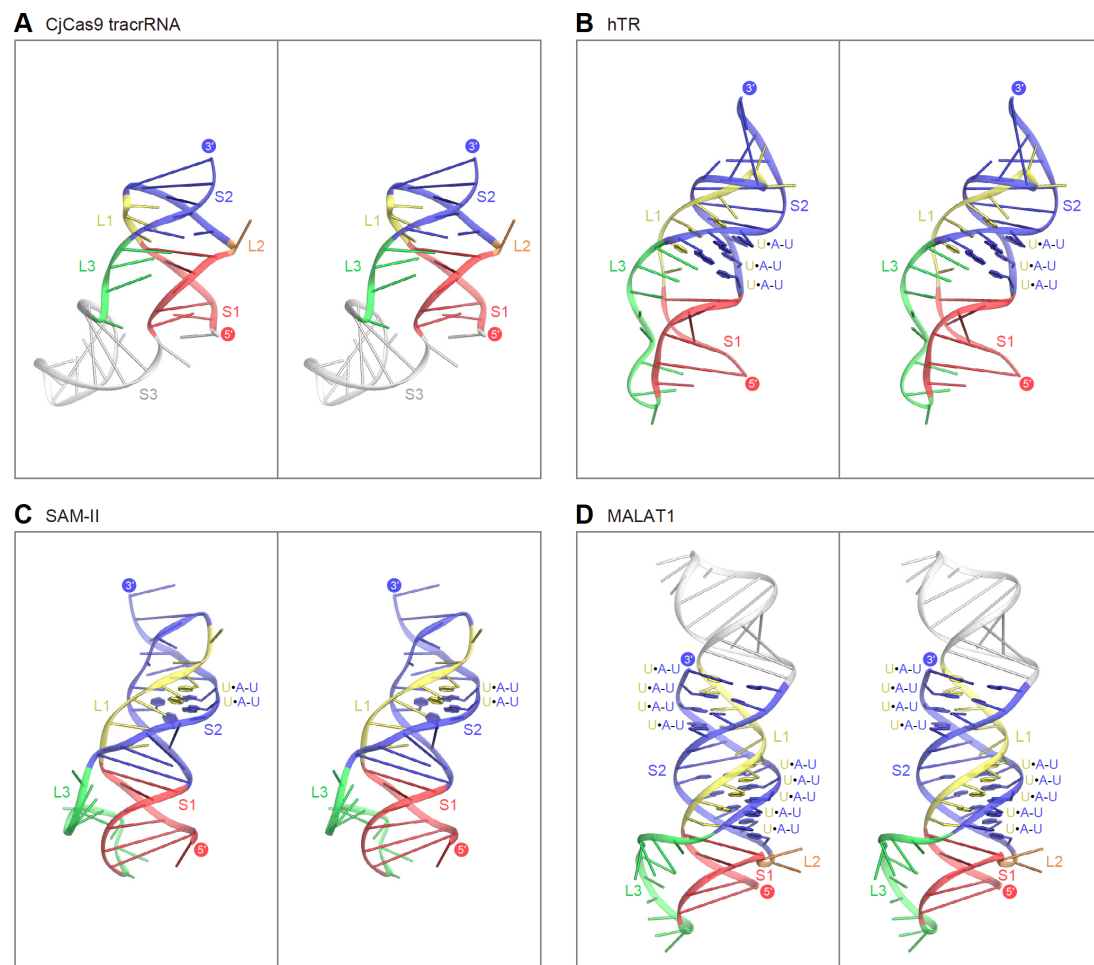


Figure 4-12 Structures of the RNA triple helices

(A–D) Structures of the CjCas9 tracrRNA (A), the telomerase RNA subunit TER (PDB: 1YMO) (B), the SAM-II riboswitch (PDB: 2QWY) (C), and the long noncoding RNA MALAT1 (PDB: 4PLX) (D) (stereo view). The U•A-U triples are depicted as stick models. S, stem; L, loop. Related to Figure 4-8.

Chapter 5 Generation of mutants with enhanced catalytic activity of CjCas9

5.1 Introduction

5.1.1 Comparison of the Cleavage Activities of CjCas9 and SpCas9

A recent study showed that the type II-C Cas9 from *Corynebacterium diphtheriae* (CdCas9), which consists of 1,084 residues and shares 21% sequence identity with CjCas9, has limited unwinding and cleavage activities toward dsDNA targets, compared with SpCas9 (Ma et al., 2015). This result suggested that the inefficiency of most of the type II-C Cas9 orthologs for genome editing results from their limited dsDNA cleavage activities (Ma et al., 2015). To examine the differences in the catalytic features of CjCas9 and SpCas9, the author compared their *in vitro* dsDNA cleavage activities. These data revealed that like CdCas9, CjCas9 cleaves the target dsDNA less efficiently, compared with SpCas9 (Figure 5-1). These results support the notion that the type II-C Cas9 enzymes, such as CdCas9 and CjCas9, have not been harnessed for genome editing at least partly because of their relatively poor activities. Thus, it is possible that an engineered CjCas9 variant with improved dsDNA cleavage activity could be used for eukaryotic genome editing. Although CjCas9 and CdCas9 commonly exhibit relatively weak dsDNA cleavage activities, they may have distinct specificities for their cognate

RNA guides. In contrast to CjCas9, which is specific to its cognate sgRNA, CdCas9 promiscuously recognizes the SpCas9 sgRNA as well as its cognate sgRNA (Ma et al., 2015). Further structural studies will provide insights into the mechanistic diversity among the type II-C CRISPR- Cas9 systems.

5.1.2 Genome editing using CjCas9

Very recently, it was reported that genome editing using CjCas9 was successful by examining the sequence of the guide RNA and PAM (Kim et al., *Nature Commun*, 2015). Small CjCas9 has advantages such as high efficiency of introduction into viral vectors, so it is expected to be applied to genome editing.

However, since the cleavage activity of CjCas9 is lower than that of SpCas9 described above, lower editing efficiency than SpCas9 having high versatility is cited as a problem. Therefore, the creation of activity enhancing mutants of CjCas9 is essential. The author aimed to create activity enhancement mutants based on the obtained structural information.

* Since this research includes content that is pending patent application, detailed information is omitted.

5.2 Materials and Methods

5.2.1 Preparation of mutants

A total of 78 mutant plasmids were prepared by inverse PCR based on the CjCas9 WT vector used in Chapter 3. These proteins of CjCas9 mutants were over-expressed in *Escherichia coli* Rosetta 2 (DE3) (Novagen) in LB medium in 5 ml small scale. The cells were grown at 37°C until the absorbance at 600 nm (A_{600}) reached 0.8 and then induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 20°C for 20 hours. Cell pellets were resuspended in buffer A, lysed by ultrasonic disruption, and centrifuged at 20000 g for 3 min at 4°C. The supernatant containing protein was purified by Magnetic beads (Thermo Fisher Scientific). After collecting the flow-through fraction, the column was washed with buffer A, and then protein was eluted with buffer C. Buffer compositions are summarized in Table 3-3.

5.2.2 *In vitro* DNA cleavage assay

In vitro plasmid DNA cleavage experiments were performed essentially as described previously (Nishimasu et al., 2015). The *Eco*RI-linearized pUC119 plasmid (100 ng, 4.7 nM), containing the 20-nt target sequence and the PAMs, was incubated at 37°C for 30 min with the CjCas9-sgRNA complex, in 10 μ l of reaction buffer, containing 20 mM

HEPES, pH 7.5, 100 mM KCl, 2 mM MgCl₂, 1 mM DTT and 5% glycerol. Reaction products were resolved on a 1% agarose gel, stained with ethidium bromide, and then visualized using a Typhoon FLA 9500 imager (GE Healthcare).

5.2.3 Purification of mut1/mut3/mut4 mutant

The CjCas9 mutant protein (mut1/mut3/mut4) was over-expressed in *Escherichia coli* Rosetta 2 (DE3) (Novagen) in LB medium. The cells were grown at 37°C until the absorbance at 600 nm (A_{600}) reached 0.8 and then induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) at 20°C for 20 hours. Cell pellets were resuspended in buffer A, lysed by ultrasonic disruption, and centrifuged at 40000 g for 30 min at 4°C in a JA-14 fixed angle rotor (Beckman Coulter). The supernatant containing CjCas9 protein was purified by chromatography on Ni-NTA Superflow resin (QIAGEN) equilibrated in buffer A. After collecting the flow-through fraction, the column was washed with buffer A, then washed with buffer B to remove nucleic acids. Thereafter, the column was washed with buffer A, and then protein was eluted with buffer C. The eluted protein was loaded to a HiTrap Heparin HP column (GE Healthcare) pre-equilibrated in a mixture of buffer D and eluted by a linear gradient from 300 mM to 2 M NaCl. The eluted protein was dialyzed overnight at 20°C with TEV protease, to

remove the N-terminal His₆-SUMO-tag, then separated by Ni-NTA Superflow resin (QIAGEN) equilibrated with buffer A. The CjCas9 mutant protein was further purified by chromatography on HiLoad 16/600 Superdex 200 (GE Healthcare) columns equilibrated in buffer F. Buffer compositions are summarized in Table 3-3.

5.2.4 *In vitro* DNA cleavage assay by MultiNA

In vitro plasmid DNA cleavage experiments were performed essentially as described previously (Nishimasu et al., 2015). The *Eco*RI-linearized pUC119 plasmid (100 ng, 4.7 nM), containing the 20-nt target sequence and the PAMs, was incubated at 37°C for 10 min with the CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5), in 10 µl of reaction buffer, containing 20 mM HEPES, pH 7.5, 100 mM KCl, 2 mM MgCl₂, 1 mM DTT and 5% glycerol. Reaction products were collected at each time (the number of corrections depends on the time course of each experiment), and visualized using MultiNA (SHIMADZU).

5.3 Results and Discussion

5.3.1 Discovery of activity enhancement mutants

In the Cas9 protein, the PLL existing between the RuvC domain and the PI domain interacts with the +1 phosphate group of the target strand DNA (Figure 4-10B)

promotes the base pairing complementary between the guide region of sgRNA and the target strand DNA. Thus, after recognition of the PAM, Cas9 induces unwinding of the target double-stranded DNA and forms the RNA: DNA heteroduplex from the vicinity of PAM (helicase activity). SpCas9 and SaCas9, which have the high helicase activity and acquire the versatility in *in vivo* genome editing experiments, recognize the PAM sequence from a position close to PLL (Figure 4-9). In contrast, in CjCas9 with weak helicase activity, the PAM sequence is recognized from the fourth position (Figure 4-9). From this fact, the author focused on the position where the residues in the PI domain begin to recognize the PAM sequence.

Based on the structural findings of CjCas9, mutations were introduced at various sites around PAM to create 78 mutants. These affinity-purified mutant proteins were used to compare the cleavage activity with WT CjCas9. The result showed that some single mutants, mut1, mut3, and mut4 have the improved activity as compared to WT (Figure 5-2). Furthermore, the triple mutants, in which these single mutants were combined, had more improved activity than these single mutants (Figure 5-3).

5.3.2 Comparison of the PAM preference between WT and mut1 / mut3 / mut4 mutant

In order to use this triple mutant as a base construct in the future, the author conducted DNA cleavage experiment to confirm whether the PAM preference has changed. the author prepared 24 PAM mutants which mutated each base from 1st to 8th base of 5'-AGAAACCC-3'. DNA cleavage experiment showed that the PAM preference of the triple mutant (mut1/mut3/mut4) was similar to that of WT (Figure5-4). These results show that the PAM preference was not changed in the triple mutant and that the mutated residues may recognize the phosphate backbone but not the nucleobases in the PAM duplex.

5.3.3 Purification of the mut1/mut3/mut4 mutant

Since a affinity-purified sample contains endogenous nucleic acids, it is necessary to use a highly purified sample to compare more accurately activities. the author performed purification of the CjCas9 mutant protein (mut1/mut3/mut4) with which has the highest activity intensity in the same way as described in chapter 2. However, it was found that the mutant binds more strongly to the endogenous nucleic acids compared to WT and it was difficult to separate of target protein and nucleic acid even through Hitrap Heparin HP column (GE Healthcare) suitable for separation of contaminating

nucleic acids. Therefore, the author examined the amount of sample loaded on the column. While, in case of WT, 5 L culture sample was loaded through a 10 ml Hitrap Heparin HP column, in case of this mutant, the 2.5 L culture sample was loaded to prevent excessive amount of nucleic acid from binding to the column. As a result, it succeeded in separating the target protein and nucleic acid. The elution sample was finally purified by gel filtration chromatography to obtain highly purified protein (Figure 5-5).

5.3.4 Comparison of the activity between WT and mut1/mut3/mut4 mutant

Comparison of the activity intensity of WT and mutant (mut1/mut3/mut4) was performed using highly purified sample. The samples were collected at each time (0.5, 1, 5, and 10 min). By measuring the concentration of cleavage products with MultiNA, the author compared the exact activities. The mutant (mut1/mut3/mut4) was found to have 2-fold higher activity as compared to WT (Figure 5-6).

5.3.5 Comparison of the DNA cleavage activities among the different PAM sequence

Recently, a study on genome editing efficiency using CjCas9 *in vivo* was published (Fujii et al., 2017). According to the result, although we reported the PAM sequence of

CjCas9 is 5'-NNNVR YM N-3', 7th and 8th PAM sequences were better AC than MN in editing efficiency *in vivo*. Therefore, the author prepared 7th and 8th PAM mutants (a total of 8 mutants) and carried out a DNA cleavage experiment with WT CjCas9 by taken time course (5 min and 10 min). Also, although the author used 5'-AGAAACCC-3' in the DNA cleavage experiment, we changed the 1st to 4th PAM sequences from AGAA to TTTG in order to avoid A rich and C rich sequences. The target DNA including 5'-AGAAACCC-3' PAM was used as a control.

When the 7th nucleotide was C, the cleavage efficiency was lower than that of the control 5'-AGAAACCC-3', irrespective of the 8th PAM in the case that the 1st to 4th PAM sequences was TTTG. In contrast, when the 7th nucleotide was A, there was a difference in cleavage efficiency independent on the 1st to 4th PAM sequences. Consistent with the results of previous studies, when the 7th and 8th PAM sequences were AC, the cleavage efficiency was highest, whereas the 7th and 8th PAM sequences are similar in AT. These results revealed that the cleavage efficiency of CjCas9 is the highest *in vitro* when the 7th is A and the 8th is Y (C or T) (Figure 5-6).

5.4 Discussion and Perspectives

Although small CjCas9 has attracted attention as a genome editing tool *in vivo*, it has not been practically available as a tool in general due to its low activity. Therefore, it

was indispensable to create an activity enhancing mutant of CjCas9. The author conducted a search for activity enhancing mutants of CjCas9 and succeeded in creating mutants with double activity intensity. Furthermore, the author found that the difference in activity intensity occurs examining PAM sequence in WT CjCas9. Furthermore, more improvement in activity is expected by examining the length of the guide sequence of sgRNA. Unfortunately, this activity enhancing mutant has been shown to be still less active than SpCas9. Recently, the author also found a single mutant with activity intensity equivalent to mut1. the author is planning to carry out activity measurement experiment using multiple mutants. In addition, the author will search genomic editing experiments, transcription repression experiments *in vivo* and crystal structure analysis. By feeding back these information, the author will search for further activity enhancing mutants.

Figures of Chapter 5

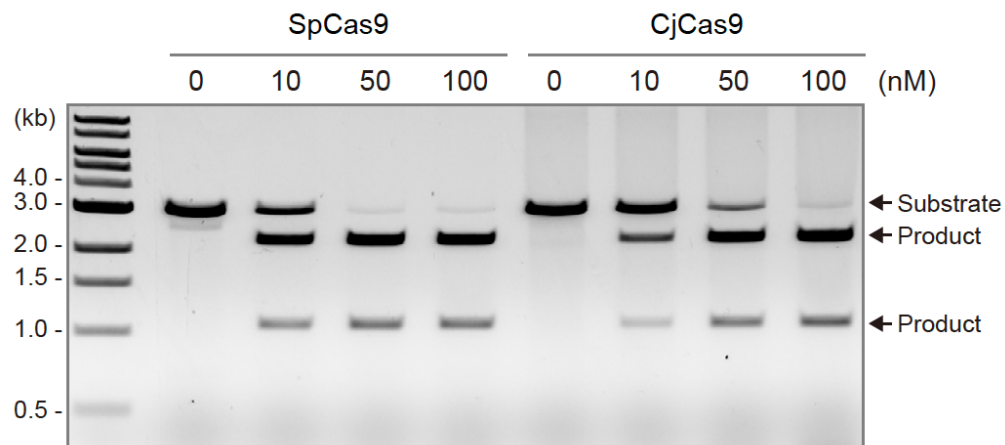


Figure 5-1 Comparison of the DNA cleavage activities of CjCas9 and SpCas9
The linearized plasmid targets with the 5'-TGG-3' PAM and the 5'-AGAAACC-3' PAM were incubated at 37°C for 5 min with SpCas9 (10–100 nM) and CjCas9 (10–100 nM) in the presence of their cognate sgRNAs, respectively.

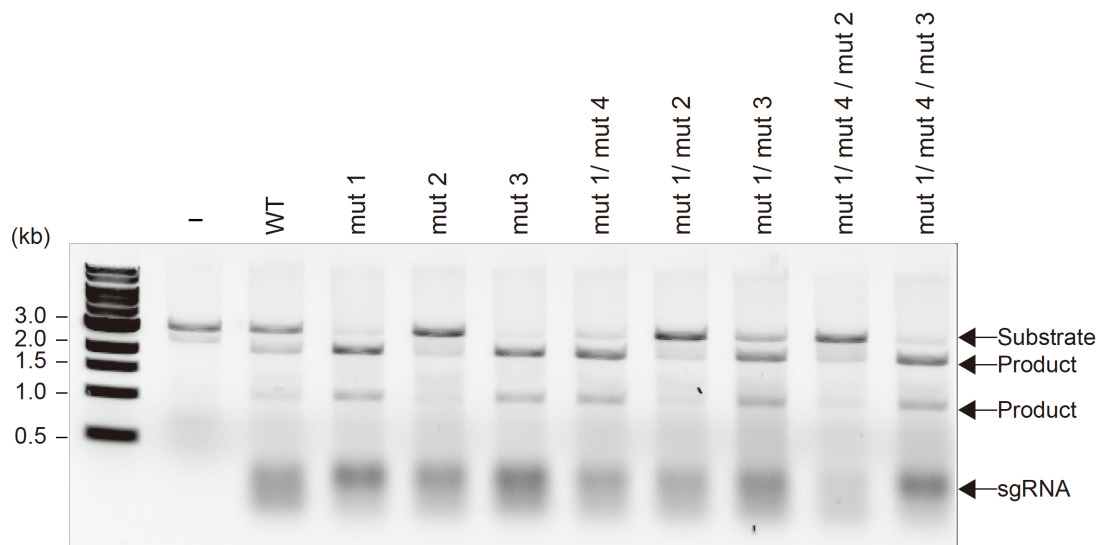


Figure 5-2 Discovery of the activity enhancement mutants

The linearized plasmid targets with the the 5'-AGAAACCC-3' PAM (100 ng) were incubated at 37°C for 30 min with the CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5) in the presence of their cognate sgRNAs, respectively in 10 μ l of reaction buffer, and then analyzed by agarose gel electrophoresis.

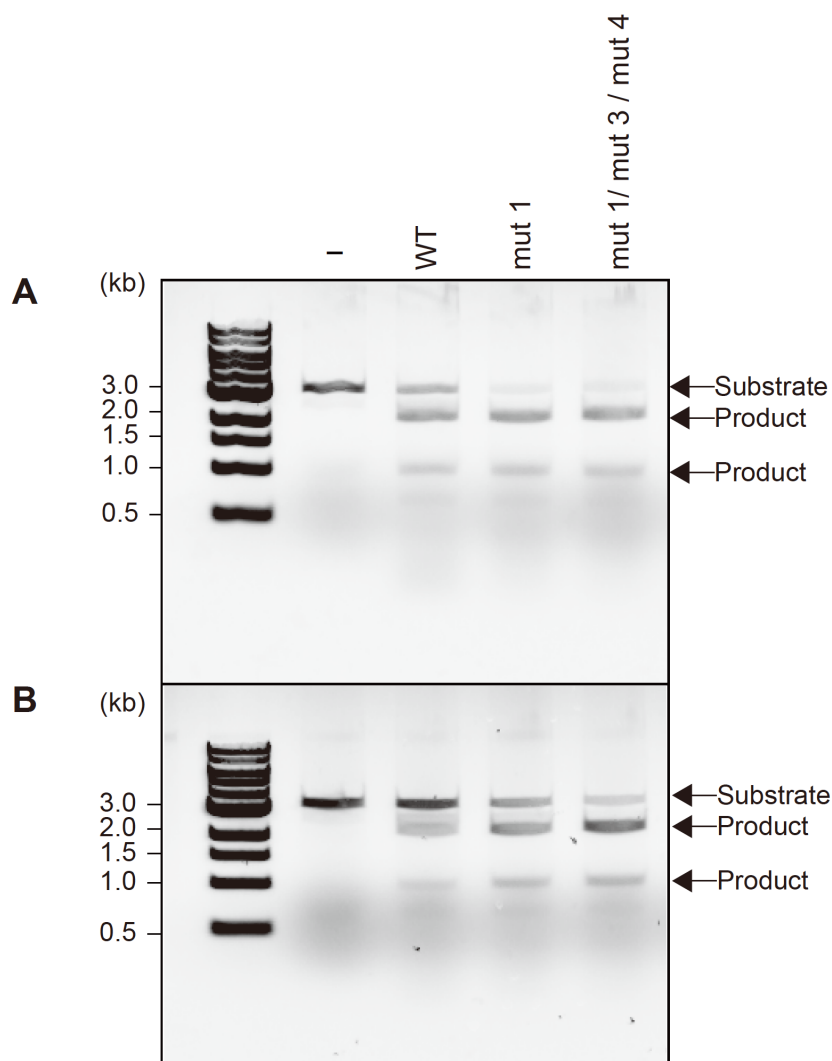


Figure 5-3 Comparison of the DNA cleavage activities of WT and mutants

The linearized plasmid targets with the the 5'-AGAAACCC-3' PAM (100 ng) were incubated at 37° C for 30 min with CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5) in the presence of their cognate sgRNAs, respectively. Two dilution series were prepared. (A) The reaction products were diluted to 1/4 with reaction buffer. (B) The reaction products were diluted to 1/8 with reaction buffer.

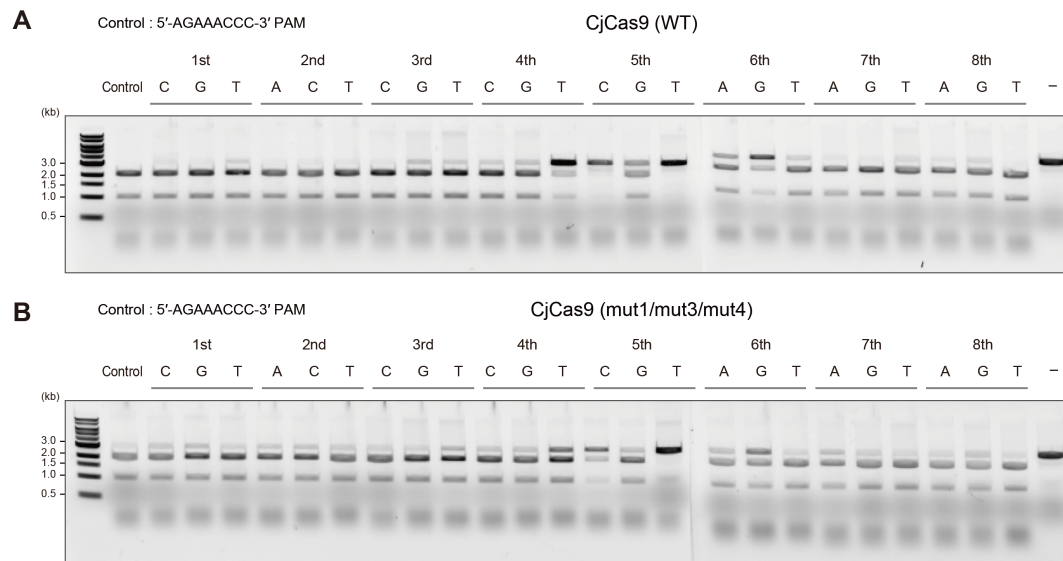


Figure 5-4 Comparison of the PAM preference between WT and the mut1/mut3/mut4 mutant

The linearized plasmid targets with the 24 PAM mutants were incubated at 37°C for 30 min with WT CjCas9-sgRNA complex (100 nM, molar ratio, 1:1.5) (A) or CjCas9 (mut1/mut3/mut4) -sgRNA complex (100 nM, molar ratio, 1:1.5) (B).

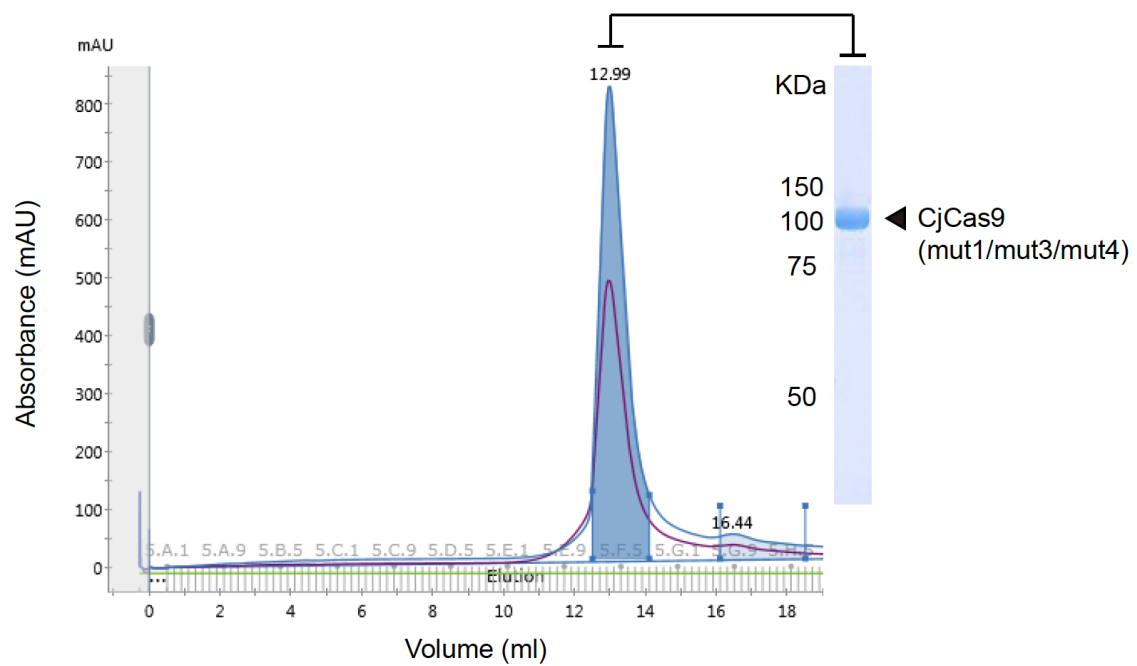


Figure 5-5 Purification of the CjCas9 mutant (mut1/mut3/mut4)

Chromatograms of the CjCas9 mutant (mut1/mut3/mut4) and SDS-PAGE analysis with coomassie brilliant blue staining.

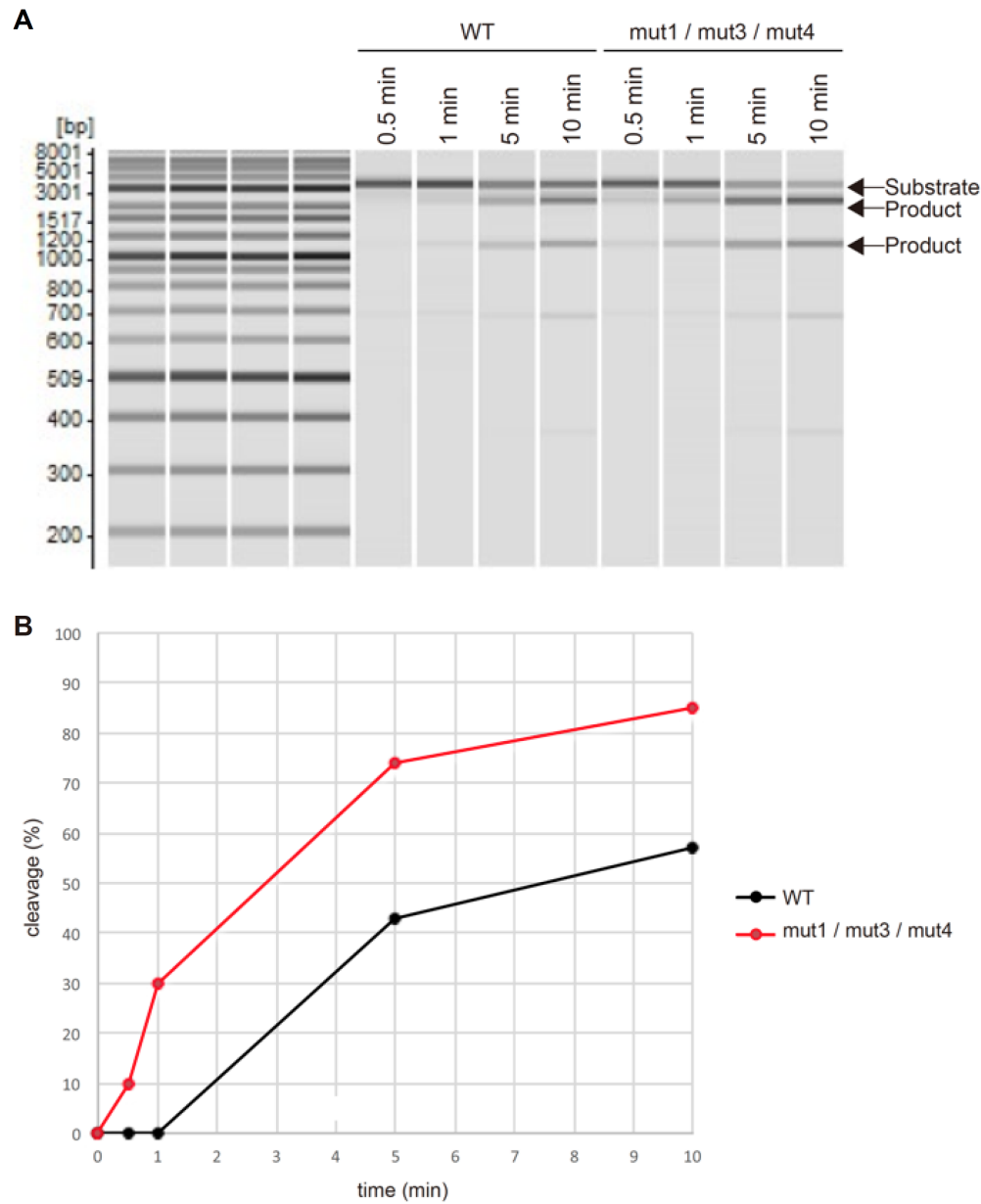


Figure 5-6 Comparison of the cleavage activities between WT and mutant (mut1/mut3/mut4)

The linearized plasmid targets with the the 5'-AGAAACCC-3' PAM were incubated at 37°C for 10 min with CjCas9 (100 nM) in the presence of their cognate sgRNAs, respectively. The samples were collected at each time (0.5, 1, 5, and 10 min). The concentration of cleavage products were measured by MultiNA. The cleavage product bands are showed in (A). The graph for cleavage efficiency is shown in (B).

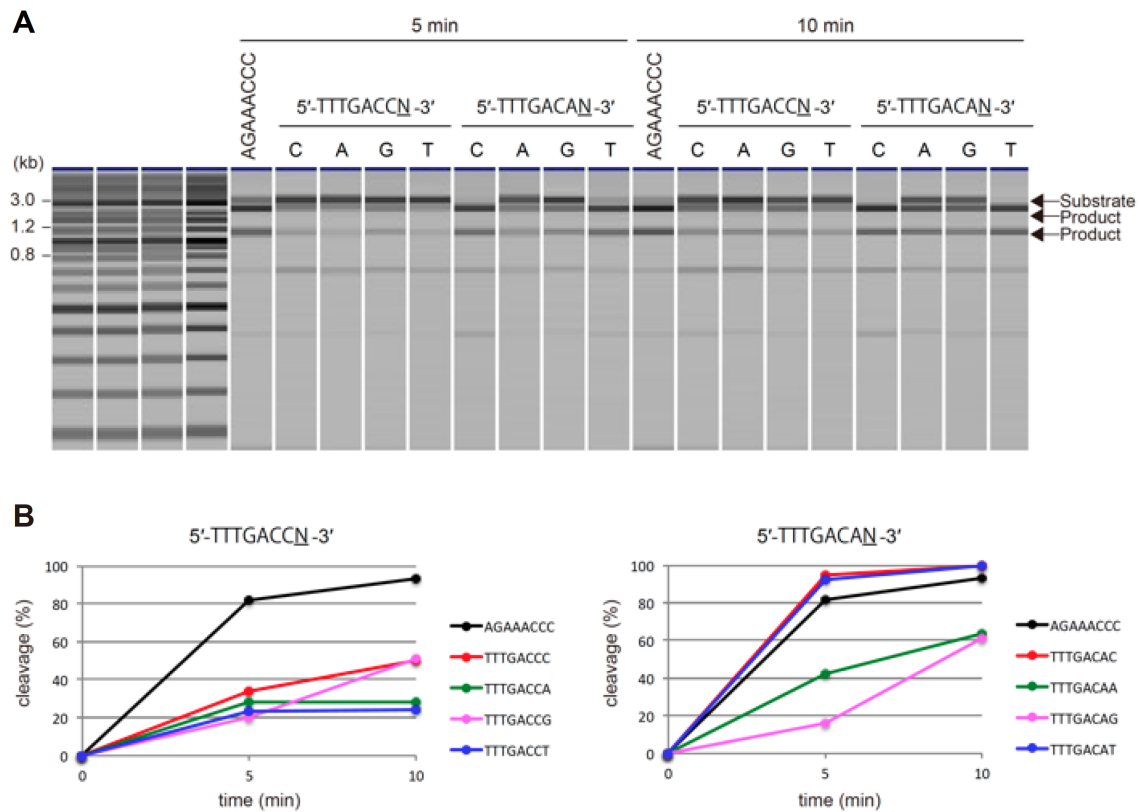


Figure 5-7 Comparison of the cleavage activities for the different PAM sequences

The linearized plasmid targets with the either the 5'-TTGACCN-3' or 5'-TTTGACAN-3' PAM were incubated at 37°C for 10 min with CjCas9 (100 nM) in the presence of their cognate sgRNAs, respectively, and then analyzed by MultiNA. The cleavage product bands are showed in (A). The graph for cleavage efficiency is shown in (B).

Chapter 6 General Discussion

6.1 Summary of this study

The author has determined the crystal structures of *Campylobacter jejuni* Cas9 (CjCas9), one of the smallest Cas9 orthologs, in complex with an sgRNA and its target DNA. The structures provided insights into a minimal Cas9 scaffold, and revealed the remarkable mechanistic diversity of the CRISPR-Cas9 systems. The CjCas9 guide RNA contains an unexpected triple-helix structure, which is distinct from known RNA triple helices, thereby expanding the natural repertoire of RNA triple helices. Furthermore, unlike the other Cas9 orthologs, CjCas9 contacts the nucleotide sequences in both the target and non-target DNA strands, and recognizes the 5'-NNNVR YM-3' as the protospacer adjacent motif. Collectively, these findings improve our mechanistic understanding of the CRISPR-Cas9 systems, and may facilitate Cas9 engineering.

Since small CjCas9 has advantages such as high efficiency of introduction into viral vectors, it is expected to be applied to genome editing. However, CjCas9 has weak DNA cleavage activity in eukaryotic cells because cleavage efficiency was lower than that of SpCas9.

From this background, the author conducted a search for activity enhancing mutants and succeeded in creating mutants with twice as much activity intensity as WT.

However, its activity is still lower than that of SpCas9. It is necessary for searching further activity enhancing mutants to obtain versatility.

6.2 Future prospects

The structure of CjCas9 reported by us was a mutant deleting the HNH domain for obtaining high-quality crystals. Although the HNH domain is a highly conserved and the structure has high similarity through organism species, the cooperation with other domains was lost due to the HNH domain deletion. Therefore, the author thinks that it is important to obtain the complete CjCas9 structure in order to gain the accurate information of the interaction with each domain.

The smallest CjCas9 has attracted attention as a genome editing tool *in vivo*, but due to its low activity, it has not been practically available as a tool in general. the author conducted a search for activity enhancing mutants of CjCas9 and succeeded in creating mutants with double activity intensity. However, this activity enhancement mutant has been found to be still less active than SpCas9 or SaCas9.

As a result of comparing the time when the Cas9 proteins cleavage the target DNA completely among SpCas9, SaCas9 and CjCas9, it showed that SpCas9 has 3.0-4.0 times and SaCas9 has 1.5-2.0 times higher activities than CjCas9 triple mutant

(mut1/mut3/mut4), respectively (data not shown). Therefore, it is necessary to search for further activity enhancement mutants. The author continued searching for activity enhancing mutants and found a few new mutants with almost the same activity intensity as mut1 (data not shown). It is expected further activity improvement when this new mutants is multiple with the triple mutant (mut1/mut3/mut4).

In addition, it was reported that genome editing using CjCas9 was successful by examining the sequence of the guide RNA and PAM (Kim *et al.*,2017). As mentioned above, the author examined the 7th and 8th PAM sequences, and certainly it was confirmed that the efficiency was better when the 7th and 8th PAM sequences were AC *in vitro* (Figure 5-7). Furthermore, it is necessary to study the length of the guide sequence in sgRNA. According to previous studies (Kim *et al.*,2017), it was reported that the genome editing efficiency *in vivo* using CjCas9 is the best when the length of the guide sequence is 22 nt. Currently, in addition to the 20 nt guide sequence, the author is conducting experiments to see the cleavage efficiency in both *in vivo* and *in vitro* using sgRNAs with 22 nt and 24 nt guide sequences.

Furthermore, the author has focused on the pseudoknot structure containing triple helix which is characteristic of CjCas9 sgRNA. The RNA structure of triple helix is more compact than other RNA structures such as stem loop. The author think that small Cas9

has evolved to recognize RNA with minimal surface area. As mentioned above, when the author mutated a triple helix site included a variety of base interactions or a flip out base recognized by CjCas9, the activity of CjCas9 decreases or disappears (Figure4-5 D). From this study, these elements are essential for CjCas9 for the accurate recognition of its sgRNA and CjCas9-sgRNA complex formation. Therefore, to attempt changing the recognition of the pseudoknot site may lead to an improvement the cleavage activity of CjCas9.

In addition to the above experiments, the author will search genomic editing experiments, transcription repression experiments *in vivo* and crystal structure analysis. By feeding back these information, the author will search for further activity enhancing mutants. In the future, the author hopes that CjCas9 gains the versatility as an application tool that surpasses SpCas9.

References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 213–221.
- Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* *513*, 569–573.
- Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. *Nat. Biotechnol.* *34*, 933–941.
- Brown, J.A., Bulkley, D., Wang, J., Valenstein, M.L., Yario, T.A., Steitz, T.A., and Steitz, J.A. (2014). Structural insights into the stabilization of MALAT1 non-coding RNA by a bipartite triple helix. *Nat. Struct. Mol. Biol.* *21*, 633–640.
- 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.
- Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* *42*, 6091–6105.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* *339*, 819–823.
- Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* *62*, 1002–1011.
- 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., Labonté 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., L  crivain, 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.
- Wataru Fujii, Arisa Ikeda, Koji Sugiura, and Kunihiro Naito (2017). Efficient Generation of Genome-Modified Mice Using *Campylobacter jejuni*-Derived CRISPR/Cas International Journal of Molecular Sciences. *18*, 2286, 1–8.
- Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magada  n, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* *468*, 67–71.
- Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* *109*, E2579–E2586.
- Gilbert, S.D., Rambo, R.P., Van Tyne, D., and Batey, R.T. (2008). Structure of the SAM-II riboswitch bound to *S*-adenosylmethionine. *Nat. Struct. Mol. Biol.* *15*, 177–182.
- Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F., et al. (2016). Structure and engineering of *Francisella novicida* Cas9. *Cell* *164*, 950–961.
- Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell* *157*, 1262–1278.
- Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* *348*, 1477–1481.
- Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* *351*, 867–871.

- Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. *eLife* 2, e00471.
- Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343, 1247997.
- Karvelis, T., Gasiunas, G., Young, J., Bigelyte, G., Silanskas, A., Cigan, M., and Siksnys, V. (2015). Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* 16, 253.
- Kim, E., Koo, T., Park, S. W. et al.: *In vivo* genome editing with a small Cas9 orthologue derived from *Campylobacter jejuni*. *Nat. Commun.*, 8, 14500 (2017)
- Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* 62, 137–147.
- Leontis, N.B., Stombaugh, J., and Westhof, E. (2002). The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res.* 30, 3497–3531.
- Ma, E., Harrington, L.B., O’Connell, M.R., Zhou, K., and Doudna, J.A. (2015). Single-stranded DNA cleavage by divergent CRISPR-Cas9 enzymes. *Mol. Cell* 60, 398–407.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015).
- An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61.

- Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353, aad5147.
- Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155, 733–740.
- Nishimasu, H., and Nureki, O. (2016). Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr. Opin. Struct. Biol.* 43, 68–78.
- Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156, 935–949.
- Nishimasu, H., Cong, L., Yan, W.X., Ran, F.A., Zetsche, B., Li, Y., Kurabayashi, A., Ishitani, R., Zhang, F., and Nureki, O. (2015). Crystal structure of *Staphylococcus aureus* Cas9. *Cell* 162, 1113–1126.
- Ondov, B.D., Bergman, N.H., and Phillippy, A.M. (2011). Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* 12, 385.
- 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.
- 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.
- Theimer, C.A., Blois, C.A., and Feigon, J. (2005). Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol. Cell* 17, 671–682.
- Waterman, D.G., Winter, G., Parkhurst, J.M., Fuentes-Montero, L., Hattne, J., Brewster, A., Sauter, N.K., Evans, G., and Rosenstrom, P. (2013). The DIALS framework for integration software. *CCP4 Newsletter* 49, 16–19.
- Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and applications of CRISPR systems: harnessing nature’s toolbox for genome engineering. *Cell* 164, 29–44.

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.

Zhang, Y., Rajan, R., Seifert, H.S., Mondragón, A., and Sontheimer, E.J. (2015). DNase H activity of *Neisseria meningitidis* Cas9. *Mol. Cell* 60, 242–255.

Original Paper

Mari Yamada, Yuto Watanabe, Jonathan S. Gootenberg, Hisato Hirano, F. Ann Ran, Takanori Nakane, Ryuichiro Ishitani, Feng Zhang, Hiroshi Nishimasu, Osamu Nureki (2017).

Crystal Structure of the Minimal Cas9 from *Campylobacter jejuni* Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Molecular Cell*. 65, 1109–1121.

Conference presentations and awards

- International conferences

1. RNA 2016, Kyoto, Japan, June 28th~July 2nd, 2016.
2. CRISPR 2017, Big Sky Montana, USA, June 8th~11th, 2017.

- Domestic conferences

1. Biochemistry and Molecular Biology 2015, Kobe, Japan, December 1st~4th, 2015.
2. The 89th Meeting of the Japanese Biochemical Society, Sendai, Japan, September 25th~27th, 2016.
3. The Annual Meeting 2016 of the Crystallographic Society of Japan, Mito, Japan, November 17th~18th, 2016.

- Awards

1. Best poster prize award of RNA 2016
2. Young best oral presentation award of the 89th Meeting of the Japanese Biochemical Society
3. Best poster prize award of the Annual Meeting 2016 of the Crystallographic Society of Japan

Acknowledgement

The author thanks Professor Osamu Nureki and Dr. Hiroshi Nishimasu for the supervisions during the master's and Ph.D. courses. I am so glad to have teachers like you. Thank you for helping me bloom from the bottom of my heart.

I also thank Dr. Ryuichiro Ishitani, Dr. Tomohiro Nishizawa, Dr. Takanori Nakane, Yuto Watanabe, Hisato Hirano and my colleagues for technical advice and support on experiments, Ms. Rieko Yamazaki for secretarial assistance, the beamline scientists at PXI at the Swiss Light Source and BL32XU and BL41XU at SPring-8 for assistance with data collection, Feng Zhang, Jonathan S. Gootenberg and F. Ann Ran for the collaboration with the discovery of PAM analysis.

Finally, I appreciate my family who have supported me at any time from the bottom of my heart. Because of you, I am me.