농학박사학위논문

화농성연쇄상구균 유래 Type II-A 크리스퍼 관련 단백질들의 구조 및 상호작용 연구

Structures and Interactions of Type II-A CRISPR-associated Proteins from *Streptococcus pyogenes*

2019년 8월

서울대학교 대학원
농생명공학부 응용생명화학전공
가 동 현

A Dissertation for the Degree of Doctor of Philosophy

Structures and Interactions of Type II-A CRISPR-associated Proteins from *Streptococcus pyogenes*

August 2019

Donghyun Ka
Applied Life Chemistry Major
Department of Agricultural Biotechnology
Seoul National University

Abstract

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (*cas*) genes constitute a prokaryotic adaptive immune system against invading foreign genetic elements like phages and plasmids. CRISPR acts as genetic memory storing short unique DNA sequences that match up with nucleic acid sequences from the intruders that previously tried to attack the host cell. The CRISPR array is accompanied by a set of *cas* genes encoding the proteins for CRISPR-mediated immunity. Cas1 and Cas2, universally conserved Cas proteins, form a complex and enables 'memorizing intruders' by integrating short fragments from the foreign nucleic acids into the host CRISPR array. In addition, Cas9 and Csn2 are also essential for the immunization process in the type II-A CRISPR-Cas system. In this thesis, I reported the X-ray crystal structures of type II-A Cas1 and Cas2 from *Streptococcus pyogenes*, and characterized the interaction network between four type II-A Cas proteins. Csn2, type II-A specific Cas protein forms a larger complex with the Cas1–Cas2 complex and also interacts directly with Cas9, while the Cas9 does not interact with the Cas1–Cas2 complex. The stoichiometry and the binding affinity of the interaction network among the four type II-A Cas proteins were determined. These results provide experimental evidence to infer the role of Csn2 and how Cas9 and Cas1–Cas2 integrase module are connected in the type II-A CRISPR-Cas systems.

Keywords: CRISPR-Cas, protein complex, sgRNA, *Streptococcus pyogenes*, structural biology, x-ray crystallography

***Student number*:** 2012-23366

Table of contents

List of Tables

List of Figures

List of Abbreviations

BME	Beta-mercaptoethanol
Cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defense
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HEPN	Higher eukaryotes and prokaryotes nucleotide-binding domain
ITC	Isothermal titration calorimetry
MALDI-TOF	Matrix-assisted laser desorption ionization time-of-flight
MBP	Maltose-binding protein
NTP	Nucleoside triphosphate
PCR	Polymerase chain reaction
Pre-crRNA	Precursor CRISPR RNA
RNase	Ribonuclease
RMSD	Root mean square deviation
SAD	Single-wavelength anomalous diffraction
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SeMet	Selenomethionine
sgRNA	Single-guide RNA
SUMO	Small ubiquitin-like modifier
TEV	Tobacco etch virus
tracrRNA	Trans-activating crRNA
Tris	tris(hydroxymethyl)aminomethane

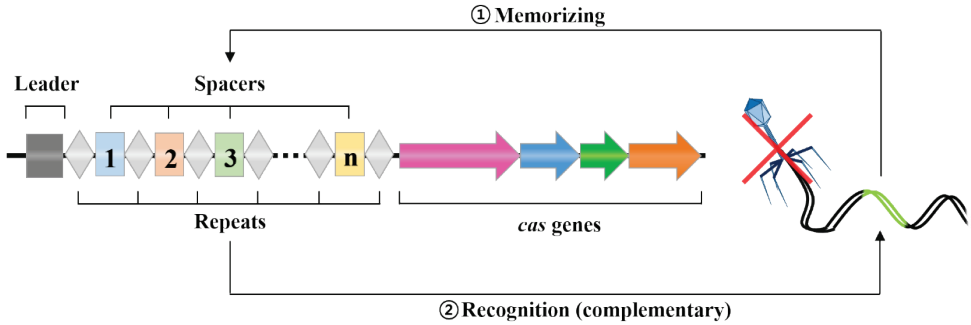
Introduction

CRISPR-Cas system

Against the threat of invading genetic elements like phages or plasmids, prokaryotes have evolved various countermeasures for disarming the predators, including restriction-modification, abortive infection, and surface modification (Rocha et al., 2001; Thomas and Nielsen, 2005; Van Houte et al., 2016). Most archaea and one-half of bacteria also have RNA-based adaptive immune system called 'CRISPR-Cas' which consists of clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (*cas*) genes (Fig. 1) (Jansen et al., 2002; Grissa et al., 2007). CRISPR-Cas system provides most bacteria and archaea with acquired resistance to the foreign nucleic acids by employing genetic memory (CRISPR) and the CRISPR-associated protein modules (Cas proteins) (Barrangou et al., 2007). CRISPR acts as an archive that has an array of 'repeat' sequences and saves 'spacer' sequences that match up with the sequences from plasmids or bacteriophages (protospacer), between two repeats. CRISPR locus contains AT-rich region called 'leader' that has promoter elements (Bolotin et al., 2005; Mojica et al., 2005; Pourcel et al., 2005; Lillestol et al., 2009). A set of *cas* genes encoding the proteins for CRISPR-mediated immunity (Cas proteins) is near the CRISPR array. The CRISPR-Cas immune system works through three stages: adaptation, expression, and interference. Firstly, the host forms a memory of previous infection by memorizing invading nucleic acid. This step is called adaptation or spacer acquisition. Universally conserved Cas1 and Cas2 proteins constitute a integrase complex for acquiring short DNA fragments from the invading nucleic acids and integrate the fragment into the host CRISPR loci, resulting a new repeat-spacer unit (Van Der Ploeg, 2009; Yosef et al., 2012; Nunez et al., 2014; Fagerlund et al., 2017; Wright et al., 2017; Xiao et al., 2017).

Figure 1. Features of CRISPR locus

CRISPR-Cas system is comprised of a CRISPR array and multiple CRISPR-associated (*cas*) genes. The CRISPR is composed of identical repeats (grey rhombus) separated by spacers (coloued boxes). The spacers originate from foreign mobile genetic elements like phages or plasmids. *cas* genes, which are associated with CRISPR loci, are usually present on one side of the array.


① Memorizing
Leader
Spacers
1
2
3
n
Repeats
cas genes
② Recognition (complementary)

The CRISPR array serves as a genetic bank and is the reason why the CRISPR-Cas system is adaptive immunity. Secondly, during the expression stage, CRISPR array is transcribed into a long precursor CRISPR RNA (crRNA). Subsequently, the CRISPR-derived transcript is processed into short crRNAs that have only one distinct spacer parts. The mature crRNA is loaded onto a single or several Cas protein(s) to form a ribonucleoprotein complex (Brouns et al., 2008; Carte et al., 2008; Deltcheva et al., 2011; Jore et al., 2011). Finally, in the interference stage, the invading nucleic acid that has complementary sequences to a crRNA get caught by the crRNA–Cas protein surveillance (or effector) complex (Hale et al., 2008; Marraffini and Sontheimer, 2008; Hale et al., 2009; Jinek et al., 2012; Westra et al., 2012; Rouillon et al., 2013).

Classification of the CRISPR-Cas systems

There is a wide divergence between CRISPR-Cas systems. These systems can be classified by the criteria of the characteristic of effector complex (Classes), the presence of signature *cas* genes (types) and the distinctive composition of *cas* gene operon (subtypes) (Fig. 2) (Makarova et al., 2011a; Makarova and Koonin, 2015; Makarova et al., 2015; Koonin et al., 2017). At the first set out, the primary classification reflected the evolutionary relationship between CRISPR loci and based on the phylogenetic tree of *cas1* gene which is universally conserved and thought to be a core gene (Haft et al., 2005; Makarova et al., 2006; Kunin et al., 2007). However, given the absence of the universal *cas* gene and the stepped-up variability of the genomic architecture of CRISPR-Cas systems, the 'polythetic approach' is adopted that determines 'signature *cas* gene' and then takes account of the divergence of *cas* gene architectures (i.e., evolutionary relationships between each CRISPR-Cas system). Through this two-step classification, each of CRISPR-Cas systems is assigned to their type (I–VI) and subtype. Furthermore, each of the six types belongs to one of the two classes that are divided by the different principles of *cas* gene

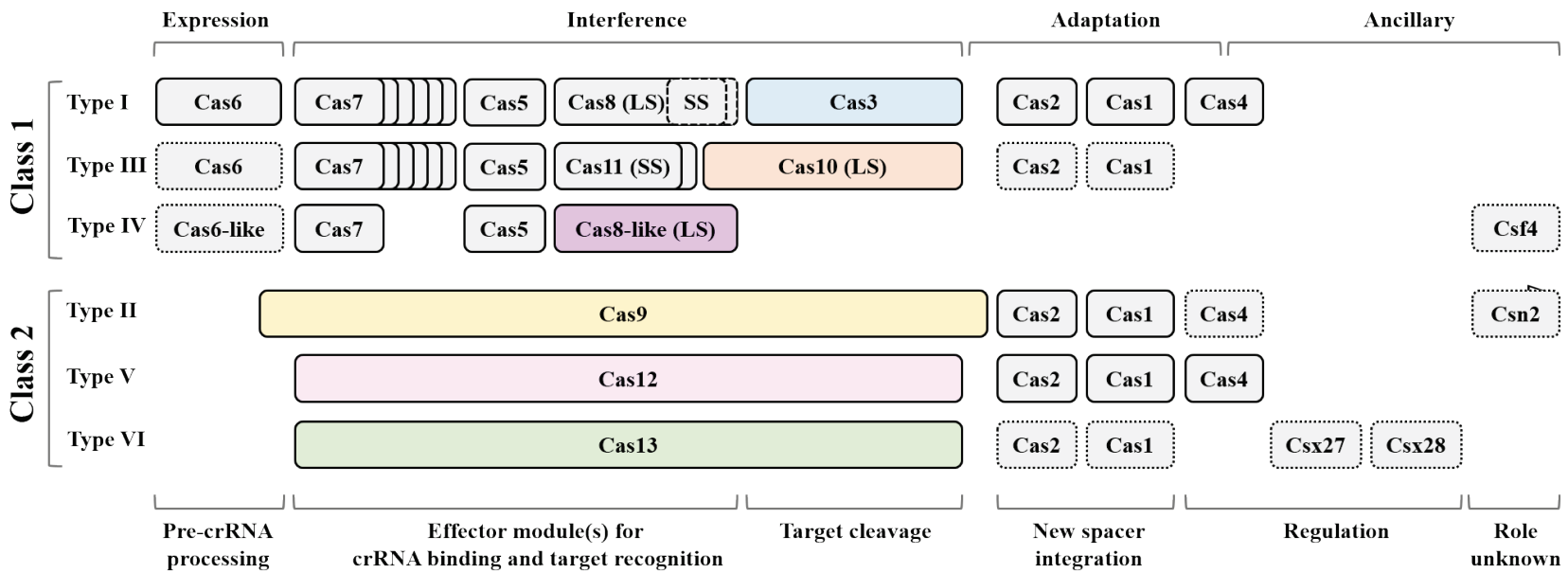
composition for effector complex; that is, whether the system possesses multi-subunit modules or a single, large, multi-domain Cas protein for effector complex. Class 1 system includes type I, III and IV CRISPR-Cas systems, and Class 2 system includes type II, V, and VI.

Type I and III CRISPR-Cas systems own *cas3* and *cas10* genes for their signature genes respectively, and both systems have structurally similar seahorse-formed CRISPR-associated complex for antiviral defense (Cascade) despite the low sequences similarity between each corresponding modules of the effector complex. Cas6 protein family functions as the specific RNase in precursor crRNA processing and binds the 3′-stem of matured crRNA. Conversely, Cas5 family binds the 5′-handle of the crRNA. The spacer part of the crRNA is bound with the helical backbone of Cascade constituted of several Cas7s which contains RNA recognition motif. The other two subunits (large and small subunit) consist the inner side of the backbone. The large subunit (Cas8 in type I and Cas10 in type III) interacts with Cas5, forming the tail of Cascade. And the small subunit, typically present in several copies but in some cases it does not exist at all, interacts with Cas7 backbone. Type IV system is regarded as a putative CRISPR-Cas system because of its recognizable features that are lacking Cas1 and Cas2 module, not in proximity to a CRISPR array (or having no detectable CRISPR arrays), encoding a minimal multi-subunit effector complex predictively composed of Cas5, Cas7 and unique large subunit, Csf1 (Cas8-like, signature gene).

Type II, type V and type VI systems have a single effector protein and also as the signature genes. On the whole, Cas9 (type II effector) and Cas12a/Cas12b (type V effectors) proteins share bi-lobed jaw-like shape and have RuvC-like nuclease domain. However, it does not show any other structural relationships apart from the RuvC-like domain. In the aspect of the nuclease activity, Cas9 has HNH nuclease family for cleaving target DNA strand. Cas12a and Cas12b also have their own unique nuclease domain corresponding to the HNH domain of Cas9. Type VI

Figure 2. Classification of the CRISPR-Cas systems

CRISPR-Cas systems have been classified into two classes, six types and many subtypes according to their mechanistic and machinery details. Although all CRISPR-Cas systems share the same mechanism of action as a whole, the class 1 and class 2 system adopts a distinctive effector complex in the interference stages. While class 1 systems employ multiple Cas proteins for an effector complex, in Class 2 systems, only one Cas protein is engaged. Signature genes of each type are shown in color.


Expression
Interference
Adaptation
Ancillary
Class 1
Type I
Cas6
Cas7
Cas5
Cas8 (LS)
SS
Cas3
Cas2
Cas1
Cas4
Type III
Cas6
Cas7
Cas5
Cas11 (SS)
Cas10 (LS)
Cas2
Cas1
Type IV
Cas6-like
Cas7
Cas5
Cas8-like (LS)
Csf4
Class 2
Type II
Cas9
Cas2
Cas1
Cas4
Csn2
Type V
Cas12
Cas2
Cas1
Cas4
Type VI
Cas13
Cas2
Cas1
Csx27
Csx28
Pre-crRNA processing
Effector module(s) for crRNA binding and target recognition
Target cleavage
New spacer integration
Regulation
Role unknown

effectors (Cas13a and Cas13b) show a distinctive feature of possessing two HEPN domains implying the dedicated RNA-targeting CRISPR-Cas systems. Nonetheless, the structural analysis of Cas13a revealed that general bi-lobed architecture of class 2 system is still standing.

Mechanism of the CRISPR-Cas system

Despite the differences between the CRISPR-Cas systems in terms of the combination of Cas proteins, so far it has been known that most of CRISPR-Cas systems relies on the same working principle and shares basic mode of action. (Fig. 3) During the adaptation stage, foreign nucleic acids (protospacers) are selected, processed, and integrated into a CRISPR array resulting in the incorporation of the new spacer at the first repeat next to leader sequence. A short conserved sequence (2-3 nt) named 'protospacer-adjacent motif' (PAM) on either side of the protospacer determines the candidate that will become a new spacer (Deveau et al., 2008; Mojica et al., 2009). Subsequently, the selected protospacer is processed into the prespacer, which has the correct size for integration by Cas1–Cas2 integrase complex. Much of knowledge about the detailed adaption mechanism has been gained from the studies in type I-E CRISPR-Cas system. In this system, Cas1 and Cas2 are the only Cas proteins required for the adaptation (Nunez et al., 2014). The Cas1–Cas2 integrase complex captures DNA fragment with 3′ overhang and the Cas1 subunit preferentially binds to the PAM-complementary sequence in the 3′ overhang. Also, the Cas1 cleaves the overhang to generate a 33-nt-length product. This Cas1-mediated processing of prespacer creates the nucleophilic 3′-OH, which can attack the leader-proximal repeat (Wang et al., 2015). Both ends of the first repeat (or leader-proximal repeat) are nicked by the Cas1 subunits and are ligated to 3′-strands of the prespacer. Finally, host DNA repair enzymes fill the gap after the integration. Several type I CRISPR-Cas systems (I-A, I-B, I-C, I-D, and I-U) have an additional nuclease protein, Cas4 which is associated with Cas1–Cas2 integrase for the

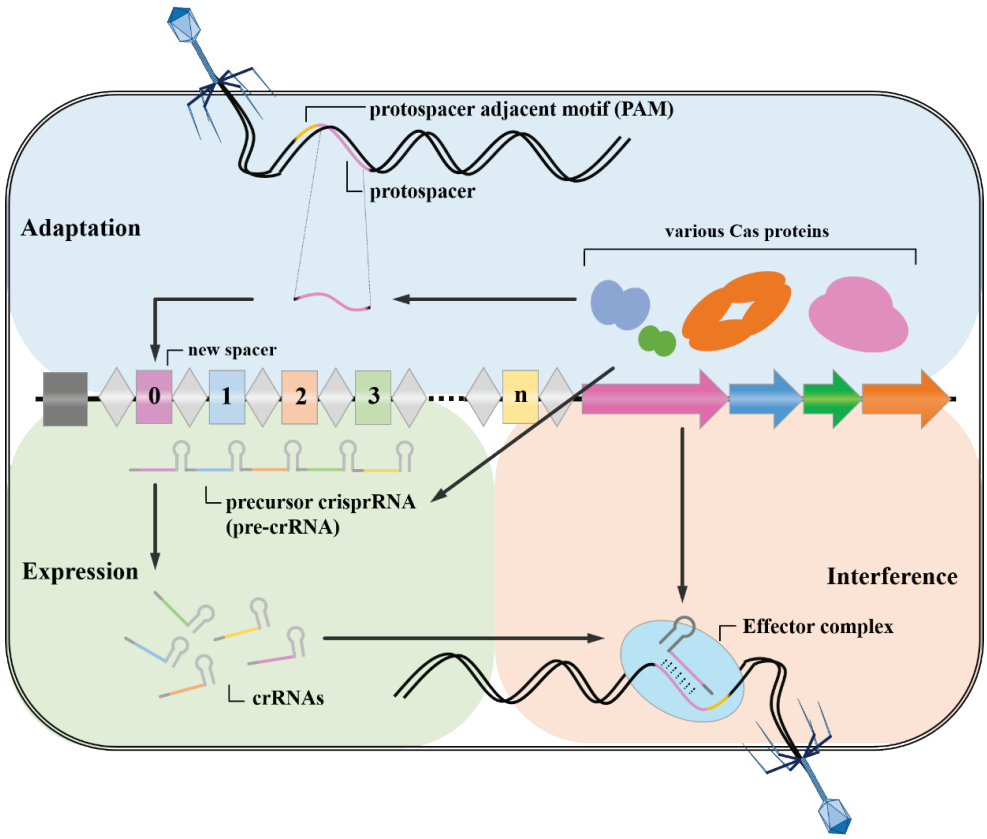
CRISPR adaptation (Makarova et al., 2018). In type I-C CRISPR-Cas system, Cas4 forms a complex with Cas1–Cas2 integrase complex and recognizes the PAM sequences within the DNA fragment. After that, the Cas4 cleaves the 3′ overhangs of DNA fragment resulting in the preparation of proper prespacer for integration step (Lee et al., 2018; Lee et al., 2019).

For the utilization of stored memory of intruders, an entire CRISPR array is transcribed at once using the leader sequence preceding the array as a start point. Subsequently, the long transcript (precursor-crRNA, pre-crRNA) is processed into matured crRNAs (or guide RNA), which consists of a spacer for target recognition and a repeat segment for binding to Cas protein(s). The goal of the expression stage is generating a guide RNA having only one spacer sequence. In the Class 1 systems, both type I and III systems employ Cas6 or Cas5 which process the repeat part of pre-crRNA and, in many cases remain bound to the crRNA acting as a scaffold protein for the effector complex (Carte et al., 2008; Garside et al., 2012; Nam et al., 2012b). However, Class 2 systems exploit an effector Cas protein and, sometimes, a host RNase III protein for crRNA maturation. Type II and type V-B systems require an additional RNA component, trans-activating crRNA (tracrRNA) which is capable of complementary binding to the repeat portion of crRNA. The crRNA-tracrRNA duplex binds to the effector protein (Deltcheva et al., 2011; Shmakov et al., 2015).

CRISPR-Cas systems exploit the complementary base pairing between the guide RNA and the target nucleic acid for recognizing and cleaving the invading genetic elements. However, the only dependence on the sequence-specific targeting causes an unintended immune response leading to the cleavage of CRISPR array. Hence, the CRISPR-Cas systems developed an additional authentication mechanism that involves the recognition of the protospacer adjacent motif (PAM) by the effector complex. This PAM-based target recognition prevents the self-destruction of CRISPR-array which does not have the PAM. Interaction with the PAM initiates unwinding of the target DNA and subsequent pairing of the guide RNA and the

Figure 3. Mechanism of CRISPR-Cas adaptive immunity

CRISPR-Cas adaptive immunity works through three stages: adaptation, expression, and interference. In the adaptation stage, short nucleic acid fragment from the invading foreign genetic element is inserted into CRISPR array. In the second stage, the CRISPR array is transcribed as long precursor CRISPR RNA, called pre-crRNA and then processed into mature crRNAs which contain a part of repeat and unique spacer sequence. This mature crRNAs can act as guide RNAs which recognize spacer-complementary sequences (protospacer) on the invading nucleic acids. In the last stage, the interference step, an effector complex formed by Cas protein(s) and the guide RNA captures and cleaves the target. In addition to the Interference step, Cas proteins also participate in the adaptation and expression stages.


protospacer adjacent motif (PAM)
protospacer
Adaptation
various Cas proteins
new spacer
0
1
2
3
n
precursor crisprRNA
(pre-crRNA)
Expression
Interference
Effector complex
crRNAs

protospacer. The PAM proximal region of the target nucleic acid, termed the seed sequence is crucial for stable binding of the effector complex to the target. Mutations in this region significantly impair the protospacer binding affinity of effector complex (Wiedenheft et al., 2011).

Type II-A CRISPR-Cas system

The characterization of the II-A CRISPR-cas system has taken the field of genome engineering to a new stage. At the heart of this enormous change is Cas9 protein. Cas9 has a nuclease activity which causes sequence-specific DNA double-strand breaks (DSBs) and can be easily reprogrammed to target desire sequences by designing a guide RNA that contain complementary sequence to the DNA of interest. This engineering application of Cas9 in the biotechnology is rooted in the fundamental researches on the mechanism of Type II-A CRISPR-Cas system. Type II-A CRISPR-Cas system contains four *cas* genes (Cas9, Cas1, Cas2, and Csn2), and trans-activating crRNA (tracrRNA) in addition to CRISPR array (Fig. 4). tracrRNA carries anti-repeat sequences which base pair with pre-crRNA and is transcribed in the vicinity of the *cas* operon. After formation of the ribonucleoprotein, the host factor RNase III processes a crRNA-tracrRNA duplex, for making a matured guide RNA. In genome editing application, each single strand RNA from tracrRNA and crRNA, which constitute the guide RNA can be artificially fused into the chimeric RNA called 'single-guide RNA' (sgRNA). The RNA-guided Cas9 endonuclease scans for PAM sequences (NGG for *S. pyogenes* Cas9). Recognizing PAM, the spacer part of the guide RNA is used to identify the target DNA as the protospacer. After going through the two-step interrogation, Cas9 cleaves the captured DNA making blunt-ended DSBs. The HNH and RuvC-like domains of Cas9 cleave the complementary and non-complementary strand of the dsDNA target, respectively.

Interestingly, the studies of the type II-A system from *S. pyogenes* and *S. thermophilus* showed that all Cas components (Cas9, Cas1, Cas2, and Csn2)

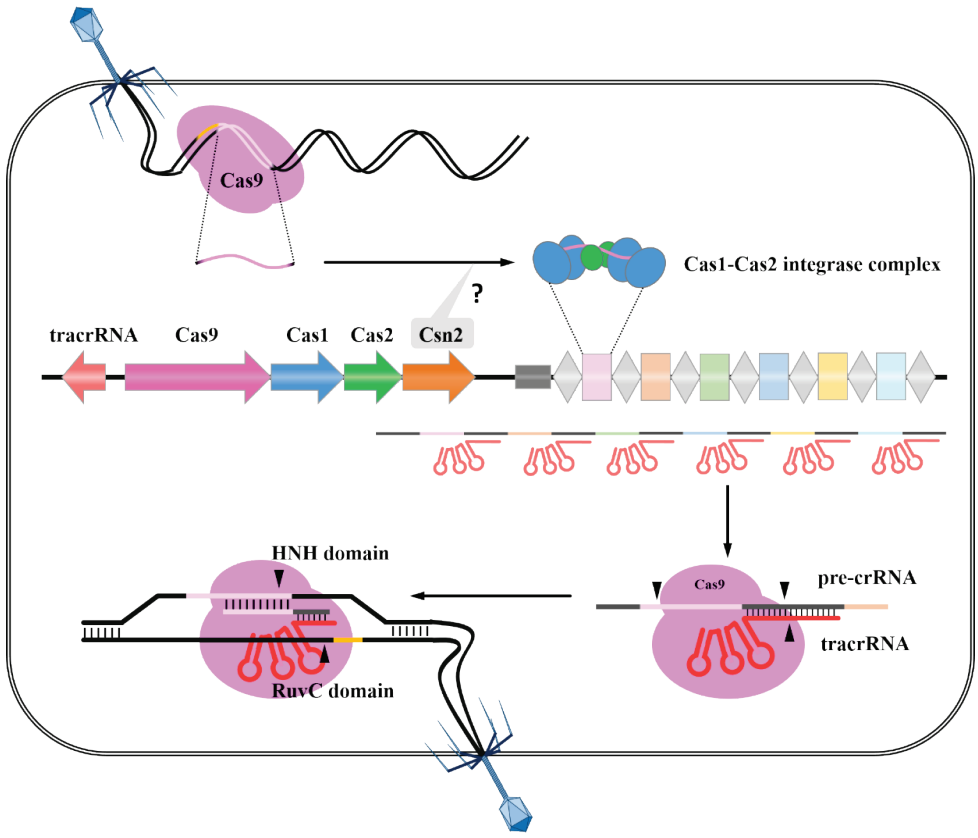
including tracrRNA, are essential for adaptation (Fig. 4). The free end of invader DNA, which is injected firstly into the cell is the preferred candidates for the new spacers (Modell et al., 2017). Cas9 recognizes the PAM sequence on the invader DNA and allows appropriate spacers to be integrated. However, the catalytic activity of Cas9 is dispensable for adaptation (Heler et al., 2015; Wei et al., 2015). Type II-A CRISPR-Cas system is one of the systems of which the spacer integration mechanism is relatively well characterized. At the leader-repeat boundary, repeat proximal four nucleotides in leader sequence are recognized by the Cas1–Cas2 complex and specify the first spacer integration site (Mcginn and Marraffini, 2016; Wright and Doudna, 2016; Xiao et al., 2017). The successful second integration depends on the size of prespacer and the recognition of the opposite end of a repeat (Wright and Doudna, 2016). Csn2 is a subtype-specific signature Cas protein and has been implicated in the adaptation stage of CRISPR-mediated immunity because it was required to obtain new spacers during *in vivo* adaptation (Barrangou et al., 2007; Heler et al., 2015). Csn2 can bind to the DNA end and slide along the dsDNA, which implies that the protein could have a role related to DNA fragments such as prespacer (Koo et al., 2012; Arslan et al., 2013). Given that II-A Cas1–Cas2 complex is capable of prespacer integration *in vitro* (Wright and Doudna, 2016; Xiao et al., 2017), Csn2 might be required for other steps in the adaptation stage such as prespacer selection or processing.

Type II-A CRISPR adaptation

For successful adaptive immunity, the CRISPR-Cas system is thought to be prudent in the choice of protospcaer. For example, the integration of prespacer DNA from mobile genetic elements outweighs that from the host DNA, which causes inadvertent autoimmunity (Yosef et al., 2012; Heler et al., 2015; Levy et al., 2015). In terms of efficiency, several type I CRISPR-Cas systems adopt a crRNA-directed adaptation called ‘primed adaptation’ for keeping the defense up-to-dates (Li et al.,

Figure 4. Type II-A CRISPR-Cas system

II-A CRISPR-Cas system possesses simple *cas* operon in terms of the number of the genes. Cas9, the signature protein for type II systems is a multi-domain protein which forms an effector complex with crRNA:tracrRNA duplex and also has a fucntion of target DNA cleavage. Although Cas1-Cas2 complex is responsible for the spacer integration process, it has been confirmed that all four proteins (Cas9, Cas1, Cas2, and Csn2) are required for adaptation *in viv*o. For Cas9, the presence of DNA cleavage activity appears not to effect on adaptation.


Cas9
Cas1-Cas2 integrase complex
?
tracrRNA
Cas9
Cas1
Cas2
Csn2
HNH domain
RuvC domain
Cas9
pre-crRNA
tracrRNA

2014; Richter et al., 2014; Levy et al., 2015). This enhanced adaptation requires additional Cascade and Cas3 nuclease and intensifies spacer uptake compared to the standard adaptation (called naïve adaptation). The crRNAs target the memorized protospacers and trigger additional spacer acquisitions from previously encountered foreign genetic elements. The primed adaptation is thought to be a strategy against phage escape mutants (Datsenko et al., 2012). Although priming has been observed only in type I systems, recent studies revealed that the II-A CRISPR-Cas system needs all Cas proteins including the Cas9 interference machinery for adaptation (Heler et al., 2015; Wei et al., 2015). Also, when *S. pyogenes* type II-A Cas operon was expressed in *Escherichia coli*, four Cas proteins were co-purified (Heler et al., 2015), suggesting the formation of a multi-protein Cas complex in type II-A CRISPR adaptation. In previous studies, type II-A CRISPR-Cas system seems to have different adaptation strategy from other systems. Therefore, identifying the adaptation mechanism will provide an understanding of the diversity of adaptation strategies among the CRISPR-Cas systems. Here, I reveal the detailed interaction network between the four *S. pyogenes* II-A Cas proteins and report the structural features of *S. pyogenes* II-A Cas1 and Cas2. Csn2 associates with the Cas1–Cas2 complex to form a higher-order complex via the interaction between Csn2 and Cas1 (Ka et al., 2016). The stoichiometry and topology of II-A Cas1–Cas2–Csn2 complex (hereafter referred to as the "type II-A CRISPR adaptation module") were determined by size exclusion chromatography (SEC) and isothermal titration calorimetry (ITC) analyses. Furthermore, I demonstrate that Cas9 interacts with Csn2, but not with Cas1–Cas2 complex to constitute the adaptation module (Ka et al., 2018). The binding stoichiometry and affinity between Cas9 and Csn2 were also characterized using SEC and surface plasmon resonance (SPR) assays. These results provide experimental evidence of how Cas9 interference module and Cas1–Cas2 integrase complex are connected in the adaptation stage and a basis for understanding the adaptation mechanism of the type II-A CRISPR-Cas system.

Materials and Methods

Cloning of type II-A cas genes

The *S. pyogenes cas1* gene was cloned into a pHMGWA vector that has an N-terminal $(His)_6$-Maltose binding protein (MBP) tag and a tobacco etch virus (TEV) protease cleavage site, for crystallization (Busso et al., 2005). The *cas1* gene was also cloned into the pET28 vector with a C-terminal $(His)_6$ affinity tag for the interaction test. The *S. pyogenes cas2* gene was cloned into pET28 vector which contains an N-terminal $(His)_6$-MBP tag and a TEV protease cleavage site. The *S. pyogenes csn2* gene was cloned into a pHMGWA vector using the same method with *S. pyogenes cas1* gene. The *S. pyogenes cas9* gene was cloned into pET28 vectors with a C-terminal $(His)_6$ affinity tag. For *S. pyogenes* Cas1–Cas2 complex co-expression. The genes of Cas1 and Cas2 were cloned into pET21 vector with no affinity tag, and pET28 vector with an N-terminal $(His)_6$-MBP tag, respectively.

Single amino acid mutants of the SpCas1 (K14A, K18A, and K25A), SpCsn2 (E23, E199A) and Cas2 (R92stop) were generated by site-directed mutagenesis using mismatched PCR primers. The deletion mutation of the *S. pyogenes* Cas2 (residues 92–113 and residues 6–91) was prepared using In-Fusion HD Cloning Kit (Takara, Japan), which recognizes and fuses both 15-bp-overlapping ends of the linearized vectors prepared by PCR procedure. The mutations were also confirmed by DNA sequencing. The genes and primer sequences used in this study are shown in Table 1 and 2, respectively.

Purification of *S. pyogenes* II-A Cas proteins

E. coli BL21 (DE3 lysogen) cells containing these all constructs were cultured in LB medium at 37ºC until the optical density at 600 nm reached 0.7. The expression

Table 1. Information of the genes used in this study

Name	Organism	Locus tag	Gene ID
Cas9	Streptococcus pyogenes M1 GAS (strain: SF370, serotype: M1, Taxonomy ID: 160490)	SPy_1046	901176
Cas1		SPy_1047	901177
Cas2		SPy_1048	901178
Csn2		SPy_1049	901179

Table 2. Sequences of primers used in this study (Continued on next page)

Primer		5′-3′ Sequence	Description
Cas9-C-terminal $(His)_6$	Forward[a]	AGATATACCATGGACAAAAAA TATAGCATTGG	Cloned into pET28a vector (Merck) using NcoI and XhoI restriction sites
	Reverse	AATTCCTCGAGATCGCCACCC AGCTGGC	
Cas1 C-terminal $(His)_6$	Forward	AGATATACCATGGCTGGTTGG CGTACT	
	Reverse	AATTCCTCGAGTATCCTAAATT CAGGAACTCCTTTC	
Cas1 untagged	Forward	AGATATACATATGGCTGGTTG GCGTAC	Cloned into pET21a vector using NdeI and XhoI restriction sites
	Reverse	AATTCCTCGAGTCATATCCTAA ATTCAGGAACTCC	
Cas1 K14A	Forward[b]	GTTGTGGTAAATACCCACTCG **GCA** TTATCCTATAAGAATAATCATC	$(His)_6$-MBP-TEV-Cas1 construct plasmid (pHMGWA) was used as a template for Site directed mutagenesis
	Reverse	GATGATTATTCTTATAGGATA A**TGC**CGAGTGGGTATTTACCA CAAC	
Cas1 K18A	Forward	CCCACTCGAAATTATCCTAT**G CG**A ATAATCATCTGATTTTTAAGG ATG	
	Reverse	CATCCTTAAAAATCAGATGAT TATT**CGC**ATAGGATAATTTCG AGTGGG	
Cas1 K25A	Forward	GAATAATCATCTGATTTTT**GC G**GA TGCCTATAAAACGGAGCTGAT CC	
	Reverse	GGATCAGCTCCGTTTTATAGG CATCC**GCA**AAAATCAGATGAT TATTC	
Cas2 N-terminal $(His)_6$-MBP-TEV protease site	Forward	GATATACATATGAGTTATAGA TATATGAGAATGATACTTATG	Cloned into modified pET28a vector containing sequences for the $(His)_6$-tagged maltose-binding protein. (His-MBP) and TEV protease recognition site using NdeI and XhoI restriction sites
	Reverse	AATTCCTCGAGTTAAGATTCAT CAAAAGCCTCC	

Cas2 (residues 1-91)	Forward	CGAATGATTTATTTACATGGTG AA**TAATAA**AATTGTATTGCAA ACTCCGATG	$(His)_6$-MBP-TEV-Cas2 construct plasmid was used as a template for Site directed mutagenesis
	Reverse	CATCGGAGTTTGCAATACAAT T**TTATTA**TTCACCATGTAAAT AAATCATTCG	
Cas2 (residues 92-113) N-terminal $(His)_6$-MBP-TEVprotease site	Forward[c]	*CAGAAATAATTGTAT*TGCAAACT CCG	$(His)_6$-MBP-TEV-Cas2 construct plasmid was used as a template for PCR.
	Reverse	*ATACAATTATTTCTG*CCCTGGAA ATACAGGTTTTC	
Cas2 (for crystallization, residues 6-91)	Forward	*AGCGGTAGCGGCAGC*ATGAGA ATGATACTTATGTTTGATATGC	$(His)_6$-MBP-TEV-Cas2 (N-terminal region) plasmid was used as a template for PCR
	Reverse	*GCTGCCGCTACCGCT*GCCCTGG AAATACAGGTTTTC	
Csn2 E23A	Forward	CCATGTTAAGAAGTGCTC**GCG** GATGTCTGTGTATTTTC	
	Reverse	GAAAATACACAGACATC**CGC**G AGCACTTCTTAACATGG	
Csn2 E199A	Forward	CTCTTTTTAGAACCACGT**GCG** CTATATGATTTTCCGCAG	
	Reverse	CTGCGGAAAATCATATAG**CGC** ACGTGGTTCTAAAAAGAG	

[a]The underlined nucleotides indicate the restriction recognition sites
[b]The bold type indicates the a desired mutation
[c]The italic font indicates the 15 bp overlap region for a cloning reaction

of the proteins was induced by the addition of 0.3 mM isopropyl-β-$_D$-thiogalactopyranoside, followed by incubation at 17ºC for 16 h. The cells were harvested by centrifugation with 500 *X g* and 5 min and resuspended in lysis buffer (300 mM NaCl, 10% (w/v) glycerol, 5 mM β-mercaptoethanol (BME), 20 mM Tris-HCl pH 7.5). After sonication and centrifugation, the supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) that was pre-equilibrated with binding buffer (300 mM NaCl, 10% (w/v) glycerol, 5 mM BME, 30 mM imidazole, 20 mM Tris-HCl pH 7.5). After washing the column with the 10 column volumes of the binding buffer, so the absorbance reached a flat baseline, the bound protein was eluted by applying a linear gradient of imidazole (up to 450 mM) with 10 column volumes.

Cas1 was dialyzed against dialysis buffer (175 mM NaCl, 10% (w/v) glycerol, 5 mM BME, 20 mM Tris-HCl pH 7.5) and, for the $(His)_6$-MBP tagged construct, the tag was cleaved by TEV protease during the dialysis. Cas1 was further purified by using a 5-mL HiTrap Heparin HP column (GE Healthcare) eluting with a linear gradient of 175 mM – 1 M NaCl. The $(His)_6$-MBP tag of Csn2 was also cleaved by TEV protease and separated on a 5-mL HisTrap HP column (GE Healthcare); however, that of Cas2 was maintained due to its low solubility below 500 mM NaCl concentration. Cas9 was dialyzed against dialysis buffer (175 mM NaCl, 10% (w/v) glycerol, 5 mM BME, 20 mM Tris-HCl pH 7.5) further purified on a 5-mL SP-Sepharose Hitrap column (GE Healthcare), eluting with a linear gradient of 175 mM – 1 M NaCl. After the purification process applied to each protein, all individual proteins were purified finally using a HiLoad 16/60 Superdex200 column (GE Healthcare) equilibrated with size-exclusion chromatography buffer (200 mM NaCl, 5% (w/v) glycerol, 2 mM dithiothreitol (DTT), 20 mM Tris-HCl pH 8.0). The mutant proteins of individual Cas proteins were expressed and purified as described for wild type proteins, except for the Cas2 (residues 92–113) used for crystallization. In that case, the protein was purified as described above for the Csn2.

For the Cas1–Cas2 complex, two Cas proteins for the complex were co-expressed in *E. coli* BL21 (DE3) cells containing both constructs. Protein expression and sample preparation for purification were performed as described above for the other individual Cas proteins. Clarified lysate was loaded onto a 5 mL HisTrap HP column (GE Healthcare). The column was washed with the binding buffer and the bound Cas1–Cas2 complex was eluted by applying a linear gradient of imidazole (up to 450 mM). The Cas1–Cas2 complex was dialyzed overnight against the dialysis buffer (500 mM NaCl, 10% (w/v) glycerol, 5 mM BME, 20 mM Tris-HCl pH 7.5). The Cas1–Cas2 complex was loaded onto a 5-mL HiTrap Heparin HP column (GE Healthcare), eluting with a linear gradient of 500 mM – 1 M NaCl. Lastly, The Cas1–Cas2 complex was further purified on size exclusion chromatography using a HiLoad 16/60 Superdex200 column (GE Healthcare) which was equilibrated with the same SEC buffer as described above for the individual Cas proteins.

For the Cas9–single-guide RNA (sgRNA) complex, firstly Cas9 in the size-exclusion chromatography buffer was incubated with sgRNA at the molar ratio 1:1.3 on ice for 1 hour. When mixing the two samples, the sgRNA was slowly injected while stirring the solution containing Cas9. After that, the mixed sample was filtered at 0.20 μm, followed by being applied onto a HiLoad 16/60 Superdex200 column (GE Healthcare).

Crystallization and structure determination of the *S. pyogenes* II-A Cas1 and Cas2

To determine the X-ray crystal structure of *S. pyogenes* type II-A Cas1, selenomethionine-containing protein was expressed in *E. coli* BL21 (DE3) cells grown in M9 medium supplemented with 0.25 mM selenomethionine (Mark et al., 2001). The selenomethionine-substituted protein was purified as described above for native Cas1 protein. The selenomethionyl Cas1 protein crystals were grown at 20ºC by the hanging-drop method from 4.2 mg/mL protein solution in buffer (300 mM

NaCl, 10% (w/v) glycerol, 2 mM DTT, 20 mM Tris-HCl pH 8.0) mixed with an equal amount of the reservoir (0.6 M sodium dihydrogen phosphate/0.6 M potassium dihydrogen phosphate, 27% (w/v) glycerol and 130 mM HEPES pH 7.0). The crystals were flash-frozen in liquid nitrogen without additional cryoprotectant.

To determine the X-ray crystal structure of *S. pyogenes* type II-A Cas2, five amino acids in the N-terminal end were replaced with Ser-Gly-Ser-Gly-Ser residues for the solubility issue, and 22 amino acids in the C-terminal tail (residues 92–113) was removed for improving the crystallization. The detailed method for the construct preparation, expression, and purification as described above. The Crystals were grown at 20°C by the sitting-drop vapor diffusion method from 14 mg/mL Cas2 protein solution in buffer (300 mM NaCl, 5% (w/v) glycerol, 5 mM BME, 20 mM HEPES pH 7.0) mixed with an equal amount of reservoir solution (22.0% (w/v) PEG 4000, 325 mM ammonium sulfate, 0.1 M sodium acetate pH 4.6). To determine the crystal structure using single-wavelength anomalous diffraction method, the selenomethionyl Cas2 was also expressed in *E. coli* BL21 (DE3) cells grown in M9 medium supplemented with 0.25 mM selenomethionine, as described previously (Mark et al., 2001). The selenomethionine-substituted protein was purified in the same manner as the native protein, and its crystals were grown under similar but not same condition (26.5% (w/v) PEG 4000, 225 mM ammonium sulfate, 0.1 M sodium acetate pH 3.6). The native and selenomethionyl crystals were both flash-frozen in liquid nitrogen without additional cryoprotective reagents.

Diffraction data were collected at the beamline 7A of the Pohang Accelerator Laboratory at 100 K. Diffraction images were processed with the HKL2000 program (Otwinowski and Minor, 1997). Determinations of selenium positions, density modification, and initial modeling of the selenomethionyl protein were done using PHENIX (Adams et al., 2002; Adams et al., 2010). In the case of Cas2, the initial model of the selenomethionyl Cas2 structure was used for phasing of the native structure in PHASER (Mccoy et al., 2007). The structure of the selenomethionyl

Cas1 and native Cas2 was completed using alternate cycles of manual fitting in COOT (Emsley and Cowtan, 2004) and refinement in PHENIX (Adams et al., 2010). The stereochemical quality of the final model was assessed using MolProbity web server (Chen et al., 2010). The detailed statistics are summarized in Table 3.

Nuclease activity assay of *S. pyogenes* II-A Cas1

Linearized pUC19 plasmids (100 ng) or PCR products containing the *S. pyogenes* type II CRISPR sequence (200 ng) were incubated with the Cas1 proteins (20 μM) in reaction buffer (100 mM KCl, 10% (w/v) glycerol, 1 mM DTT, 25 mM HEPES pH 7.5) in a final volume of 5 μl at 20°C for 2 hours. Each reaction was supplemented with no metal, EDTA or various divalent ions (2 mM Mg^{2+}, Ca^{2+}, Mn^{2+}, Ni^{2+}, and Co^{2+}). The activity assay for Cas1 in the presence of one or all of Cas9, Cas2, and Csn2, DNA substrates were incubated with Cas1 (20 μM) in the presence of one or all, SpCas9 (10 μM), N-terminal $(His)_6$-MBP-tagged Cas2 (10 μM) and Csn2 (20 μM).

The reaction mixture was quenched with proteinase K treatment at 20°C for 1 hour by adding 1 ul of proteinase K stock (125 μg/ml) for protein digestion. After that, to remove residual peptide fragments from the reaction mixture containing the nucleic acid substrate, phenol-chloroform extraction procedure was performed. Briefly, 300 μl of DDW and the equal volume of phenol : chloroform : isoamyl alcohol (25:24:1) premix (P2069, Sigma Aldrich) was added. The mixture was thoroughly mixed using a vortex mixer and centrifuged to take 270 μl of supernatant (aqueous phase) which contains nucleic acids substrate. The supernatant was concentrated by ethanol precipitation. The final reaction products from the linearized plasmids and the PCR products were analyzed on 1% (w/v) and 2.5% (w/v) agarose gels and visualized using ethidium bromide, respectively. In the case of activity assay against other types of nucleic acid (short dsDNA, ssDNA, ssRNA), Each nucleic acid substrate analyzed in a 12% (w/v) polyacrylamide gel and visualized by

Table 3. Data collection and refinement statistics[a] for *S. pyogenes* II-A Cas1 and Cas2

	Cas1	Cas2	
	SeMet ^{peak}	SeMet ^{peak}	Native
Space group	$P3_121$	$P2_12_12_1$	$P2_12_12_1$
Unit cell parameters (Å)	a = b = 95.2, c = 210.8	a=51.1, b=76.4, c=145.2	a=51.1, b=75.6, c=146.4
Wavelength (Å)	0.9794	0.9793	0.9793
Data collection statistics			
Resolution range (Å)	50.00–2.25 (2.33–2.25)	50.00–2.00 (2.07–2.00)	50.00–1.76 (1.82–1.76)
Number of reflections	53138 (5224)	38716 (3813)	56417 (5499)
Completeness (%)	99.6 (99.5)	99.9 (100.0)	99.7 (98.6)
R_{merge}[b]	0.078 (0.657)	0.099 (0.644)	0.071 (0.452)
Redundancy	10.9 (11.1)	7.2 (7.3)	7.2 (6.5)
Mean I/σ	21.2 (3.9)	12.0 (3.1)	22.5 (3.5)
Phasing statistics			
f', f'' used in phasing	-8.0, 6.1	-8.0, 4.3	
Figure of merit	0.367	0.414	
Refinement statistics			
Resolution range (Å)	35.14–2.25		47.96–1.76
R_{cryst}[c]/R_{free}[d] (%)	18.2/21.1		19.7/23.3
RMSD bonds (Å)	0.019		0.007
RMSD angles (deg)	1.76		1.00
Average B-factor (Å^2)	70.8		42.0
Number of water molecules	249		293
Ramachandran favored (%)	96.3		99.2
Ramachandran allowed (%)	3.5		0.6

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

[b]$R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

[c]$R_{cryst} = \Sigma_h | |F_{obs}| - |F_{calc}| | / \Sigma_h |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

[d]R_{free} was calculated as R_{cryst} using ~ 5% of the randomly selected unique reflections that were omitted from structure refinement.

SYBR Gold staining.

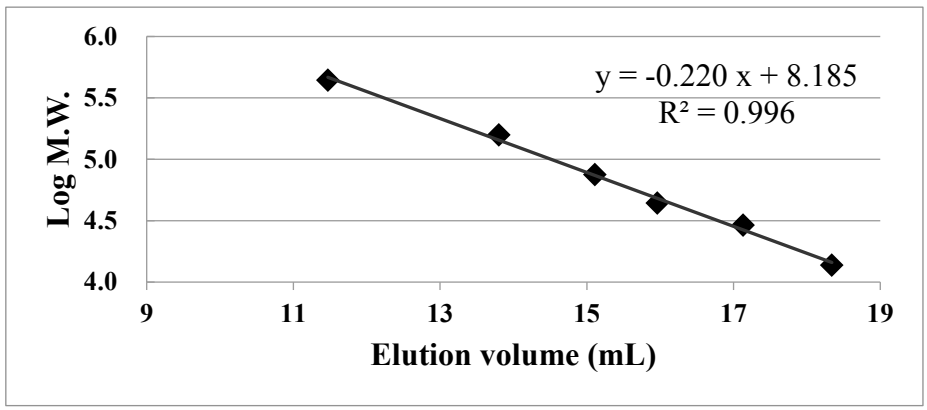
Analytical size exclusion chromatography (SEC)

Analytical SEC was performed on Superdex 200 10/300 GL columns (GE Healthcare). The columns were equilibrated with buffer (150 mM NaCl, 2 mM DTT, 20 mM Tris-HCl pH 8.0). For complex formation test, the 700 µl of samples with a mixture of two (or more) Cas proteins were incubated together in the buffer (200 mM NaCl, 2 mM DTT, 20 mM Tris-HCl pH 8.0) at 4°C for 1 hour. 450 µl of the incubated sample was loaded onto the columns at a flow rate of 0.5 mL/min. The experiments with different salt concentrations were performed by applying different concentrations of NaCl in the column. Control experiments for single Cas protein were performed for making reference curves in the same way as the mixture samples. Experiments with varying salt concentrations were performed by applying different amounts of NaCl to the column. A standard curve for the analytical SEC column is provided in Table 4. Elution fractions were analyzed by 12% (w/v) sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by Coomassie blue staining.

Isothermal titration calorimetry (ITC) experiment

The two protein sample pairs used in the experiment were prepared by dialysis with the same buffer (150 mM NaCl, 10% (w/v) glycerol, 20 mM Tris pH 8.0) for over 12 hours in order to avoid unintended heat effects due to the composition difference between the respective buffers in which those two samples are dissolved. The ITC experiments were performed using a MicroCal iTC200 system (GE Healthcare) at 25°C except for the $(His)_6$-MBP-tagged Cas1–Csn2 interaction which carried out at 20°C due to the instability of reaction. 200 µl of protein sample in the cell being stirred at 800 rpm were titrated with 20 consecutive 2-µL injections. Each injection

Table 4. Standard curve for the analytical SEC column used in this study



Standard	Molecular weight (M.W.; kDa)	Elution volume (mL)
Ferritin	440	11.5
Aldolase	158	13.8
Conalbumin	75	15.1
Ovalbumin	44	16.0
Carbonic anhydrase	29	17.1
Ribonuclease A	13.7	18.3

was carried out at 180 seconds spacing and over 4 seconds duration. Origin software (OriginLab) was used for processing and analysis of the ITC titration data with a non-linear curve fitting assuming one set of the binding site and multiple identical independent binding sites. The unique sets of fitting parameters (*N*; the number of binding sites, *K*; binding constant in M^{-1} and ΔH in cal/mole) are calculated from the fitting analysis.

***In vitro* transcription and purification of the sgRNA**

The sgRNA was in vitro transcribed using T7 RNA polymerase (P266L mutant) and linearized plasmid template that carries a T7 promoter sequence in front of the template sequence. The 103-nucleotide template sequence for sgRNA in vitro transcription (5′-G***AAATTAGGTGCGCTTGGC***GTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCTTGGATCC-3′) harbors 18 bases as a guide sequence corresponding to the 3′ end side of the spacer4 from the *S. pyogenes* CRISPR1 (bold italic), one G at the 5′ end for efficient transcription and BamHI restriction site for plasmid linearization (underlined). The DNA template for the in vitro transcription was prepared by using the *AccuPrep*® Nano-Plus Plasmid Extraction Maxiprep Kit (Bioneer, Korea). The P266L-mutated T7 RNA polymerase gene inserted in the pAR1219 vector, which contains N-terminal $(His)_6$ tag (Davanloo et al., 1984; Guillerez et al., 2005). The protein expression and sample preparation for purification were performed as described above for the other proteins. The supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) that was pre-equilibrated with binding buffer (300 mM NaCl, 5% (w/v) glycerol, 5 mM BME, 30 mM imidazole, 20 mM Sodium phosphate pH 7.4). The bound protein was washed with the binding buffer and eluted by applying a linear gradient of imidazole (up to 450 mM) The protein was further purified using a HiLoad 16/60 Superdex200 column (GE Healthcare) equilibrated with size-exclusion chromatography buffer

(200 mM NaCl, 2 mM dithiothreitol (DTT), 2 mM Ethylenediaminetetraacetic acid (EDTA), 20 mM sodium phosphate pH 7.4). The purified T7 RNA polymerase was dialyzed against storage buffer (100 mM NaCl, 50% (w/v) glycerol, 1 mM DTT, 1 mM EDTA, 20 mM sodium phosphate pH 7.4).

As a detailed experiment procedure, the sgRNA was transcribed by mixing 4 unit/μl T7 RNA polymerase, 60 nM linear template DNA and 2 mM of each NTPs (ATP, GTP, CTP and UTP) in the reaction buffer (25 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 2 mM spermidine, 40 mM Tris-HCl pH 8.25) at 37°C for 6 hours. A stock solution of 1 M Tris-HCl for 10X reaction buffer was strictly adjusted to pH 8.25, not 8.0, which is recommended in general in vitro transcription protocol, because the reaction is very pH-sensitive. After removing the pyrophosphate pellet by centrifugation (15,000 $\times g$, 1 min), the nucleic acid precipitate was collected by ethanol precipitation followed by centrifugation (15,000 $\times g$, 30 min at 4°C) and air-dried. The pellet was resuspended in the RNA SEC buffer (150 mM NaCl, 10 mM sodium phosphate pH 6.6) and the dissolved RNA transcript was purified by size exclusion chromatography on a HiLoad 16/60 Superdex200 column (GE Healthcare) equilibrated from the same RNA SEC buffer for being separated with the linearized plasmid templates and the remaining nucleotides.

Surface plasmon resonance (SPR) experiment

The SPR binding assays were carried out on a series S sensor chip CM5 with a Biacore T200 (GE Healthcare). The ligand immobilization procedure was performed using the amine coupling method at a flow rate of 5 μl/min. The carboxyl groups of dextran matrix covalently attached to the gold surface of the CM5 chip was activated with a mixture of 0.1 M N-hydroxysuccinimide and 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride at a ratio of 1:1 for 400 s. Subsequently, 60 μg/mL of the Cas9 dissolved in 10 mM sodium acetate (pH 5.0) or equivalent concentration of sgRNA-loaded *S. pyogenes* Cas9 in 10 mM sodium

acetate (pH 4.0) was injected for 30 seconds or 400 seconds, respectively. In this procedure, the ligand was passed over the surface, and the ester groups reacted with primary amine groups of ligand molecules linking covalently. The remaining activated carboxyl groups on the sensor chip surface were deactivated with 1 M ethanolamine (pH 8.5) for 400 s. After that, a short injection (24 seconds) of 10 mM NaOH was used to remove loosely bound ligands and confirmed the stabilized baseline. The multi-cycle analysis was performed at a flow rate of 30 µl/min. Each concentration of Csn2 in running buffer (150 mM NaCl, 10 mM Tris-HCl pH 8.0) was injected for 600 s, followed by dissociation for 1200 s in a separate analysis cycle. The sensor chip surface was regenerated with 10 mM NaOH between cycles. Data were fit with the bivalent analyte model. The equilibrium dissociation constant (K_d) was determined based on kinetic rate constants calculated using Biacore T200 evaluation software (GE Healthcare).

Results

Structural analysis of *S. pyogenes* II-A Cas1

For the structural basis of Type II Cas1, the crystal structure of *S. pyogenes* II-A Cas1 was solved at a resolution of 2.25 Å, using SAD phasing method (Table 3) (protein data bank; PDB ID code 4ZKJ). The *S. pyogenes* II-A Cas1 monomer possesses an N-terminal β-sandwich-like domain (residues 1–81) and a C-terminal α-helical domain (residues 88–289) that are connected by a short flexible linker (residues 82–87) (Fig. 5A). Topology and domain folds of *S. pyogenes* II-A Cas1 monomer are also similar to those of other type I Cas1 proteins (PDB ID: 4N06 for I-A Cas1 from *Archaeoglobus fulgidus* (Kim et al., 2013), 3NKD for I-E Cas1 from *Escherichia coli* (Babu et al., 2011), and 3GOD I-F Cas1 from *Pseudomonas aeruginosa* (Wiedenheft et al., 2009)). The N-terminal domain is comprised of two β-sheets (sheet 1; β1, β5, β7, β8, and sheet 2; β2–β4, β6) that are stacked almost in parallel and two helices (η1, α1) located next to the sandwich-like part. The C-terminal domain includes nine α-helices of various lengths (α2–α10), three 3_{10} helices (η2–η4) and two continuous short β-strands (β9 and β10).

Each domain of *S. pyogenes* II-A Cas1 is superposed well with the corresponding domain of the previously solved type I Cas1 structures. The monomer structure of *S. pyogenes* II-A Cas1 shows a two-domain architecture similar to that seen in other type I Cas1 structures despite low sequence identities (below 20%) (Fig. 5B and Fig. 6). The structural similarity of individual domains between *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 was accessed because *A. fulgidus* I-A Cas1 shows the highest sequence similarity among type I Cas1 structures. The root mean square deviation (RMSD) values of alpha carbon positions for the each domain between the two Cas1 structures were calculated to be 2.0 Å for 83 Cα atoms

Figure 5. Monomer structure of *S. pyogenes* II-A Cas1 and structural comparison with *A. fulgidus* I-A Cas1

(A) The monomer structure of *S. pyogenes* II-A Cas1 is colored in rainbow format. (N-terminus = blue, C-terminus = red) (B) Structural alignment of for the N- and C-terminal domains between *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 monomers. (B) Structural alignment of *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 monomers based on their N-terminal domains. *S. pyogenes* II-A Cas1 is colored in blue, and *A. fulgidus* I-A Cas1 is shown in pink.

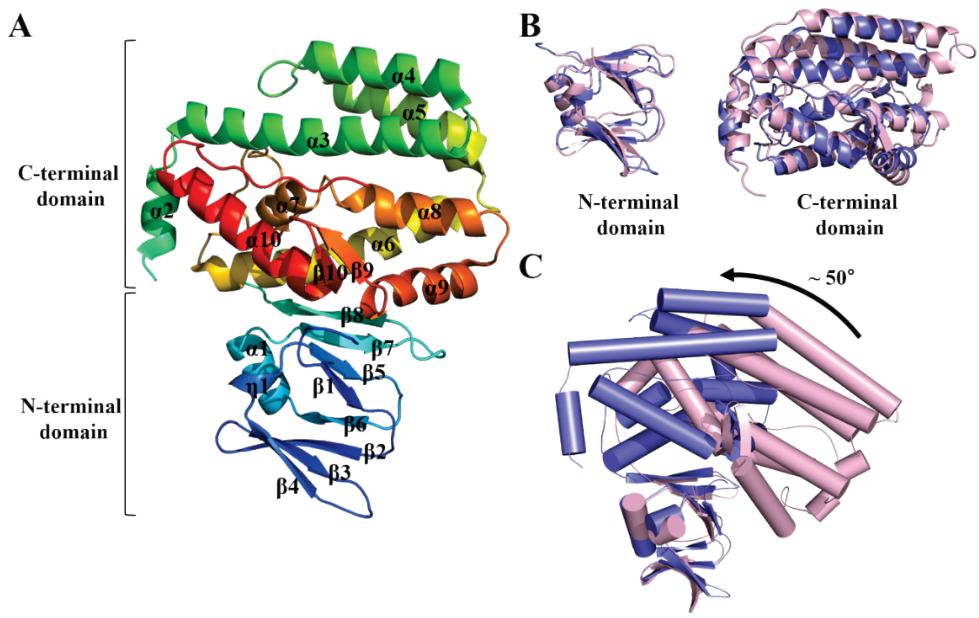
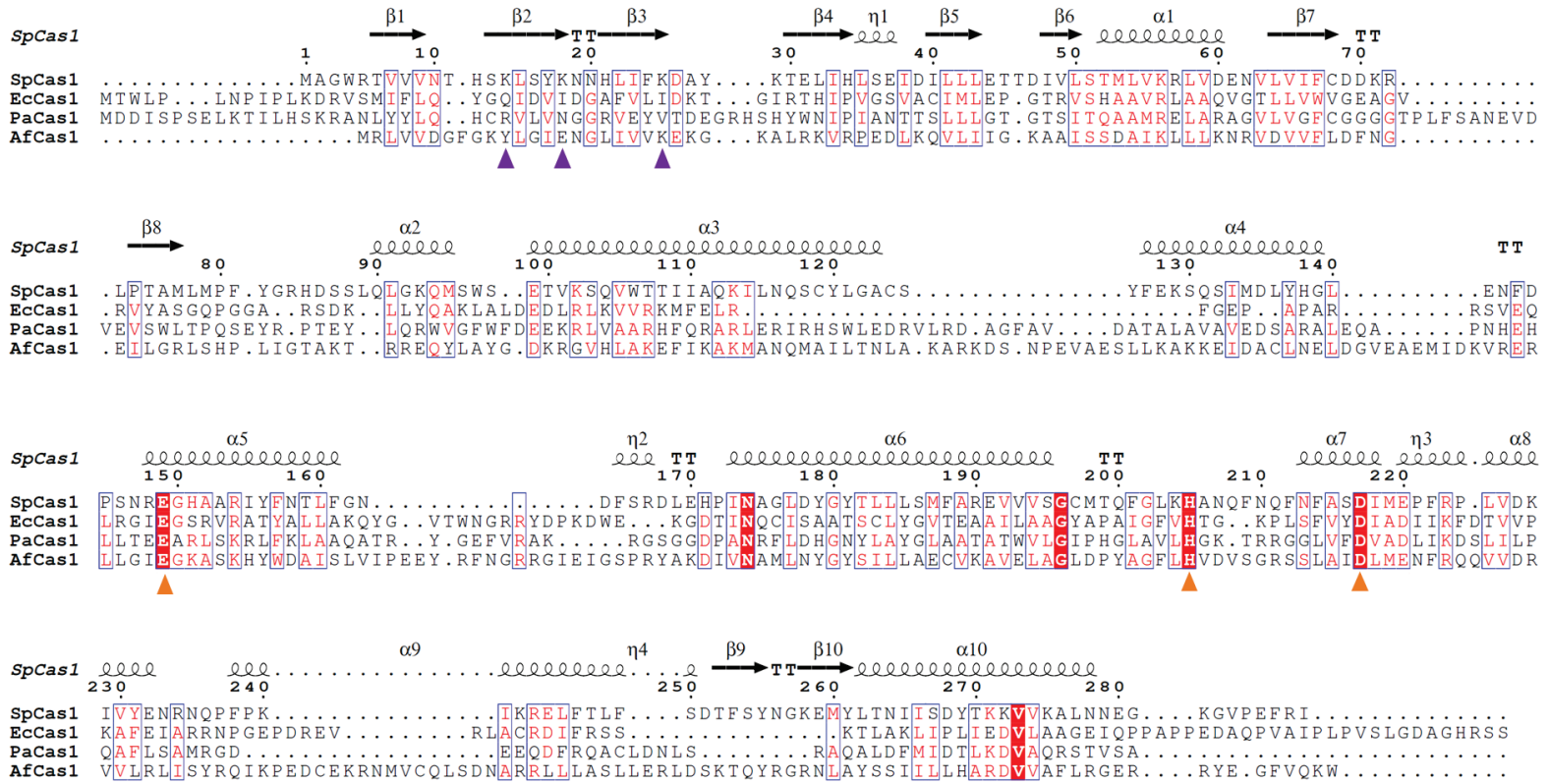

A
C-terminal domain
N-terminal domain
α4
α5
α3
α2
α7
α8
α10
α6
β10
β9
α9
β8
α1
β7
β5
β1
β6
β2
β3
β4
B
N-terminal domain
C-terminal domain
C
~ 50°

Figure 6. Structure-based sequence alignment of *S. pyogenes* II-A Cas1 and type I Cas1 proteins

Secondary structure annotations are indicated based on *S. pyogenes* II-A Cas1. Three conserved residues coordinating a metal ion for catalysis are marked with red triangles. The residues which are predicted to participate in interactions with *S. pyogenes* Csn2 also indicated by purple triangles. (Sp; *S. pyogenes, Ec;E. coli, Pa; P. aeruginosa, Af; A. fulgidus)*


SpCas1
β1 β2 TT β3 β4 η1 β5 β6 α1 β7 TT
SpCas1 ..............MAGWRTVVVNT.HSKLSYKNNHLIFKDAY....KTELIHLSEIDILLLETTDIVLSTMLVKRLVDENVLVIFCDDKR.....
EcCas1 MTWLP...LNPIPLKDRVSMIFLQ..YGQIDVIDGAFVLIDKT...GIRTHIPVGSVACIMLEP.GTRVSHAAVRLAAQVGTLLVWVGEAGV.....
PaCas1 MDDISPSELKTILHSKRANLYYLQ..HCRVLVNGGRVEYVTDEGRHSHYWNIPIANTTSLLLGT.GTSITQAAMRELARAGVLVGFCGGGGTPLFSANEVD
AfCas1 ..................MRLVVDGFGKYLGIENGLIVVKEKG...KALRKVRPEDLKQVLIIG.KAAISSDAIKLLLKNRVDVVFLDFNG.....
β8 α2 α3 α4 TT
SpCas1 .LPTAMLMPF.YGRHDSSLQLGKQMSWS..ETVKSQVWTTIIAQKILNQSCYLGACS..............YFEKSQSIMDLYHGL.....ENFD
EcCas1 .RVYASGQPGGA..RSDK..LLYQAKLALDEDLRLKVVRKMFELR.............................FGEP..APAR.....RSVEQ
PaCas1 VEVSWLTPQSEYR.PTEY..LQRWVGFWFDEEKRLVAARHFQRARLERIRHSWLEDRVLRD.AGFAV....DATALAVAVEDSARALEQA.....PNHEH
AfCas1 .EILGRLSHP.LIGTAKT..RREQYLAYG.DKRGVHLAKEFIKAKMANQMAILTNLA.KARKDS.NPEVAESLLKAKKEIDACLNELDGVEAEMIDKVRER
α5 η2 TT α6 TT α7 η3 α8
SpCas1 PSNREGHAARIYFNTLFGN.....DFSRDLEHPINAGLDYGYTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDIMEPFRP.LVDK
EcCas1 LRGIEGSRVRATYALLAKQYG..VTWNGRRYDPKDWE...KGDTINQCISAATSCLYGVTEAAILAAGYAPAIGFVHTG..KPLSFVYDIADIIKFDTVVP
PaCas1 LLTEEARLSKRLFKLAAQATR..Y.GEFVRAK.....RGSGGDPANRFLDHGNYLAYGLAATATWVLGIPHGLAVLHGK.TRRGGLVFDVADLIKDSLILP
AfCas1 LLGIEGKASKHYWDAISLVIPEEY.RFNGRRGIEIGSPRYAKDIVNAMLNYGYSILLAECVKAVELAGLDPYAGFLHVDVSGRSSLAIDLMENFRQQVVDR
α9 η4 β9 TT β10 α10
SpCas1 ....
SpCas1 IVYENRNQPFPK.....IKRELFTLF....SDTFSYNGKEMYLTNIISDYTKKVVKALNNEG....KGVPEFRI.....
EcCas1 KAFEIARRNPGEPDREV.....RLACRDIFRSS.....KTLAKLIPLIEDVLAAGEIQPPAPPEDAQPVAIPLPVSLGDAGHRSS
PaCas1 QAFLSAMRGD.....EEQDFRQACLDNLS.....RAQALDFMIDTLKDVAQRSTVSA.....
AfCas1 VVLRLISYRQIKPEDCEKRNMVCQLSDNARRLLLASLLERLDSKTQYRGRNLAYSSILLHARDVVAFLRGER....RYE.GFVQKW.....

(N-terminal domain) and 2.3 Å for 200 Cα atoms (N-terminal domain). However, *S. pyogenes* II-A Cas1 monomer exhibits a different relative orientation of the two domains, which is distinct from those seen in other type I Cas1 structures (Fig. 5C). To manifest this difference, the two Cas1 monomers were aligned using only the N-terminal domains, and remarkable deviation in the location of their C-terminal regions was observed. When the N-terminal domain is set to a fixed point, the C-terminal portion rotates about 50 degrees. With this change, the RMSD value of alpha carbon atomic positions between intact *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 monomers is higher (4.9 Å) than those calculated separately for the N- and C-terminal domains.

This unique conformation of *S. pyogenes* II-A Cas1 monomer also provides a more extensive interface for dimer formation than the homologous structures in type I systems (Fig. 7A and B). The dimeric state of type I Cas1 proteins is mainly formed by the interaction of two N-terminal domains especially involving three secondary structural elements (β6, α1, β8) that are forming hydrophobic packing. In addition to the interactions between the N-terminal domains, another dimer interface is newly formed by the two C-terminal regions in the Crystal structure of *S. pyogenes* II-A Cas1. The loop of 16 amino acids in length (residues 196–211) connecting α6 and α7 protrudes from the C-terminal domain and contacts extensively with residues on the concave section formed by four helices (α3 and α5–α8) of the opposite C-terminal domain from the other monomer. This structural feature has not been previously reported in other type I Cas1 proteins. As a result, there are 19 hydrogen bond interactions and 2707 $Å^2$ of surface area between the two C-terminal domains, while the *A. fulgidus* I-A Cas1 dimer has only 4 hydrogen bonds and bury 869 $Å^2$ of surface area between two C-terminal domains.

The overall shape of *S. pyogenes* II-A Cas1 dimer is significantly different from those of the type I Cas1 dimers (Fig. 7C). The silhouette of type I Cas1 dimers has been previously described as a “butterfly shape” in which the N- and C-terminal

domains corresponded to the hindwing and forewing, respectively (Wiedenheft et al., 2009). In *S. pyogenes* II-A Cas1, while the N-terminal portion maintains similar appearance with other type I Cas1 proteins, its C-terminal domains slide into the center and also make an extensive contact surface as described previously (Fig. 7B). The stacked wingspan portion of *S. pyogenes* II-A Cas1 is only ~65 Å width whereas *A. fulgidus* I-A Cas1 has a wingspan of ~102 Å. This change makes that the *S. pyogenes* II-A Cas1 has a more globular shape and compact dimeric form compared to the type I Cas1 homologs.

Disruption of putative metal binding sites in *S. pyogenes* II-A Cas1

Type I Cas1 proteins have been reported to cleave various nucleic acid substrates in sequence-independent and metal-dependent manners (Wiedenheft et al., 2009; Babu et al., 2011; Kim et al., 2013). In many cases, these Cas1 proteins show divalent metal ion-dependent nuclease activities against dsDNAs (Wiedenheft et al., 2009; Babu et al., 2011; Kim et al., 2013). Three conserved residues (two carboxylates; aspartate and glutamate and one histidine) in the concave of the C-terminal domain are responsible for metal ion binding and have been identified to be critical for the enzymatic activity in the previous research (Fig. 7B and Fig. 8). In the X-ray crystal structure of *P. aeruginosa* I-F Cas1, those three residues (E190, H254, and D268) coordinate a catalytic manganese ion (Wiedenheft et al., 2009). In the other type I Cas1 structures, despite the absence of the metal ion, the positions of these three residues are structurally conserved (Babu et al., 2011; Kim et al., 2013). In previous studies, mutations of these three residues or addition of EDTA in the dsDNA cleavage assays inhibited the catalytic activity of Cas1 (Wiedenheft et al., 2009; Babu et al., 2011; Kim et al., 2013).

In type II Cas1 proteins including the *S. pyogenes* II-A Cas1, the three residues (E149, H205, and E220 in *S. pyogenes* II-A Cas1) are also conserved in their sequences, but they are not found within proximity in the structure of *S. pyogenes*

Figure 7. Comparison of *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 dimer
(A) Dimeric structure of *S. pyogenes* II-A Cas1. Chain A and B are shown in blue and cyan, respectively. (B) Structural alignment of *S. pyogenes* II-A Cas1 and *A. fulgidus* I-A Cas1 dimers based on their N-terminal domains. *A. fulgidus* is shown in pink and orange, representing chain A and B, respectively. (C) Side-by-side comparison of the two Cas1 dimers. (Left; *S. pyogenes* II-A Cas1, right; *A. fulgidus* I-A Cas1)

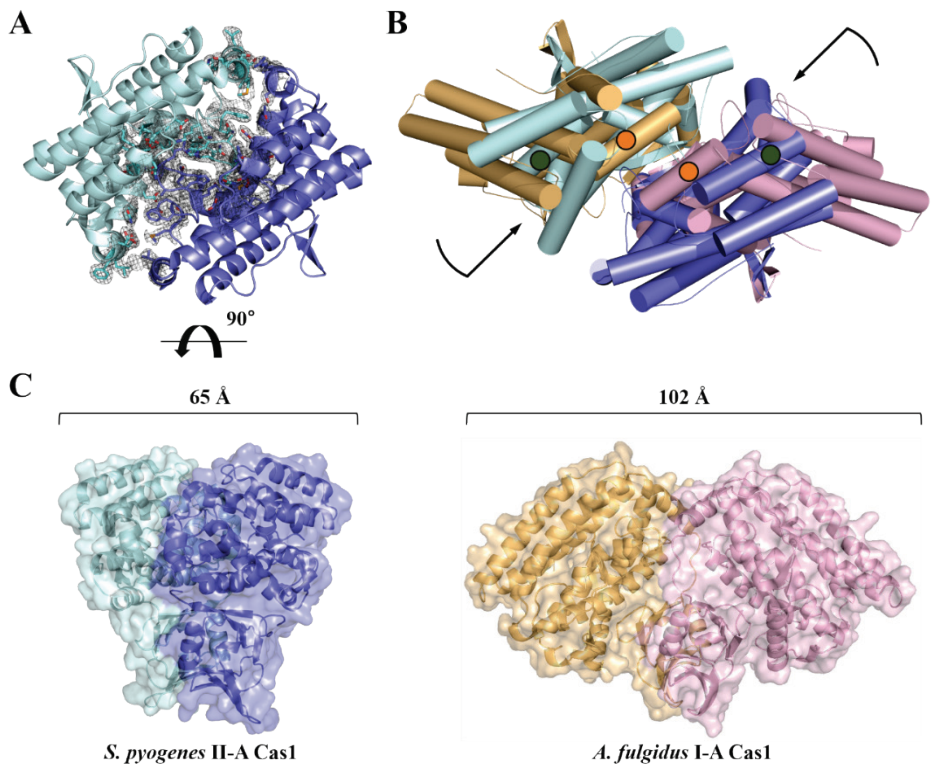
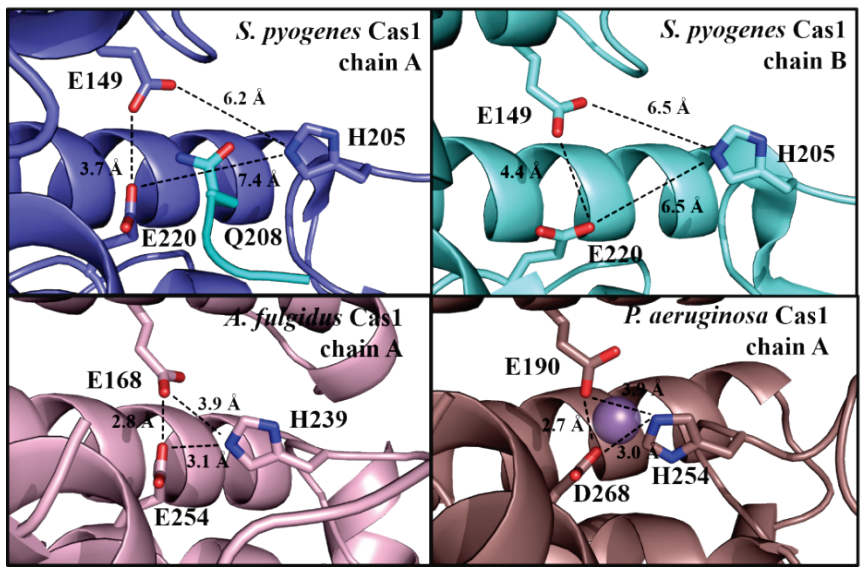

A
B
90°
C
65 Å
102 Å
S. pyogenes II-A Cas1
A. fulgidus I-A Cas1

Figure 8. Disruption of the potential metal-binding site in *S. pyogenes* II-A Cas1

Close-up views of three conserved catalytic residues in the potential metal-binding site. A purple sphere represents the bound manganese ion. (*S. pyogenes* II-A Cas1, up-left for chain A and up-right for chain B; *A. fulgidus* I-A Cas1, down-left; *P. aeruginosa* I-F Cas1, down-right)


S. pyogenes Cas1
chain A
E149
6.2 Å
H205
3.7 Å
7.4 Å
E220
Q208
S. pyogenes Cas1
chain B
E149
6.5 Å
H205
4.4 Å
6.5 Å
E220
A. fulgidus Cas1
chain A
E168
3.9 Å
2.8 Å
H239
3.1 Å
E254
P. aeruginosa Cas1
chain A
E190
2.7 Å
3.0 Å
H254
D268

II-A Cas1 (Fig. 6). The three residues are distant from each other. Especially the histidine residue is twice as long as those of other4 type I Cas1 (Fig. 8). Interestingly, the side chain of E208 from another protomer (i.e., chain B) is found in the center space between the three conserved residues of chain A, occupying the putative metal-binding site in the *S. pyogenes* II-A Cas1 monomer. This structural feature suggests that the extended interactions between two C-terminal domains of *S. pyogenes* II-A Cas1 lead to the disruption of the positioning of potential catalytic residues for metal-binding. No significant amount of metal ion was co-purified with the *S. pyogenes* II-A Cas1, as measured by inductively coupled plasma mass spectrometry (Table 5). The different conformational state of *S. pyogenes* II-A Cas1 dimer also results in the burial of these three residues. The two glutamate residues (E149 and E220) of *S. pyogenes* II-A Cas1 are buried on the surface created by the dimerization of the Cas1, while the corredponding glutamate residues in the type I Cas1 structures are exposed to the outside (Table 6). The distance between the two potential metal-binding sites of the *S. pyogenes* II-A Cas1 dimer is ~15 Å whereas that of *A. fulgidus* Cas1 dimer is ~43 Å. This structural rearrangement of the C-terminal domain portion may result in limited accessibility of the putative catalytic sites for its substrates such as dsDNAs. From this structural characteristics, it is likely that the *S. pyogenes* II-A Cas1 dimer may malfunction in binding and/or cleavage of dsDNA unless another conformational change occurs, because the potential metal binding sites appear to be disrupted.

With this structural information, the nuclease activity assay was performed to confirm whether *S. pyogenes* II-A Cas1 exhibits the catalytic function for various nucleic acid substrates (Fig. 9 and Fig. 10). Firstly, the dsDNA nuclease activity of *S. pyogenes* II-A Cas1 against linearized pUC19 plasmids was examined with various divalent ions; however, no cleavage was detected. Its nuclease function for another dsDNA substrate containing *S. pyogenes* type II-A CRISPR locus was also

Table 5. Concentration (ppb) of metals in the purified *S. pyogenes* II-A Cas1 sample measured by inductively coupled plasma mass spectrometry

Metal	Concentration (ppb)
Mg	3.94
Mn	Not detected
Co	Not detected
Ni	24.62
Cu	7.38
Zn	9.42

Table 6. Surface area analysis of the potential metal binding sites in the Cas1 structures

	S.pyogenes II-A Cas1		*A. fulgidus* I-A Cas1		*P. aeruginosa* I-F Cas1	
	E149	E220	E168	E254	E190	D268
Buried surface area by dimerization (Å^2)	30.6	17.2	0.0	0.0	0.0	0.0
Accessible surface area in dimer (Å^2)	3.7	1.3	8.5	3.4	21.7	15.8

Average values calculated for multiple independent chains in asymmetric unit are reported.

tested, but no DNase activity was detected in this assay. With other types of nucleic acid substrates such as ssDNAs, dsRNAs, and ssRNAs, *S. pyogenes* II-A Cas1 did not display nuclease activity under the conditions supplemented with various divalent metal ions. Although it cannot be excluded that more specific experimental conditions may be required for the catalytic function of *S. pyogenes* II-A Cas1, these results suggest at least that *S. pyogenes* II-A Cas1 may not have a nuclease activity against the tested nucleic acid substrates.

Structural analysis of *S. pyogenes* II-A Cas2

Among the four *S. pyogenes* type II-A Cas proteins, the crystal structures of Csn2 and Cas9 have been reported previously (Koo et al., 2012; Anders et al., 2014; Jinek et al., 2014; Nishimasu et al., 2014; Jiang et al., 2015). After determining the crystal structure of *S. pyogenes* II-A Cas1, the structure of *S. pyogenes* II-A Cas2 was also determined to fill the knowledge gap on the structure of individual *S. pyogenes* II-A Cas proteins. At first, an N-terminal $(His)_6$-MBP tag was required for *S. pyogenes* II-A Cas2 because there was the precipitation problem after removing the tag from the Cas2 protein. However, it was not able to crystallize the $(His)_6$-MBP-tagged form of *S. pyogenes* II-A Cas2. Thus, another construct was needed that can increase the solubility of the *S. pyogenes* II-A Cas2. In the sequence alignment with other II-A Cas2 proteins, the *S. pyogenes* II-A Cas2 have additional five residues at N-terminus (Met-Ser-Tyr-Arg-Tyr). Assuming that these residues are not related to the structure of the Cas2 protein, first five N-terminal residues (Met-Ser-Tyr-Arg-Tyr) were replaced with a flexible segment (Ser-Gly-Ser-Gly-Ser) which could act as a linker because some Cas2 proteins require a linker to remove the tag. Finally, it was able to gain the soluble form of *S. pyogenes* II-A Cas2 without the $(His)_6$-MBP-tag. The crystal structure of *S. pyogenes* II-A Cas2 was solved using the construct from which the C-terminal tail (residues 92–113) was additionally eliminated. In the case of the *S. pyogenes* II-A Cas2 containing the C-terminal tail portion, protein crystals

Figure 9. Various nucleic acid cleavage assays with *S. pyogenes* II-A Cas1

(A) Liearized pUC19 (left) or *the S. pyogenes* type II CRISPR loci DNA fragments were incubated with *S. pyogenes* II-A Cas1. (B) Cleavage assay of *S. pyogenes* II-A Cas1 against various types of nucleic acid substrates. Sequences of nucleic acid substrates used in the assay are 5′-TTGTACAGACGTTCCCACATTCTTGT-3′ for ssDNA, 5′-GUUCAGCGUGUCCGGC GAGT-3′ for dsRNA and 5′-AGAGCAGUGGCUGGUUGAGAUUUAAAGAGC-3′ for ssRNA. Each reaction was supplemented with no metal, EDTA or various divalent ions. DNase I was used as a positive control which cleaves the dsDNA substrates.

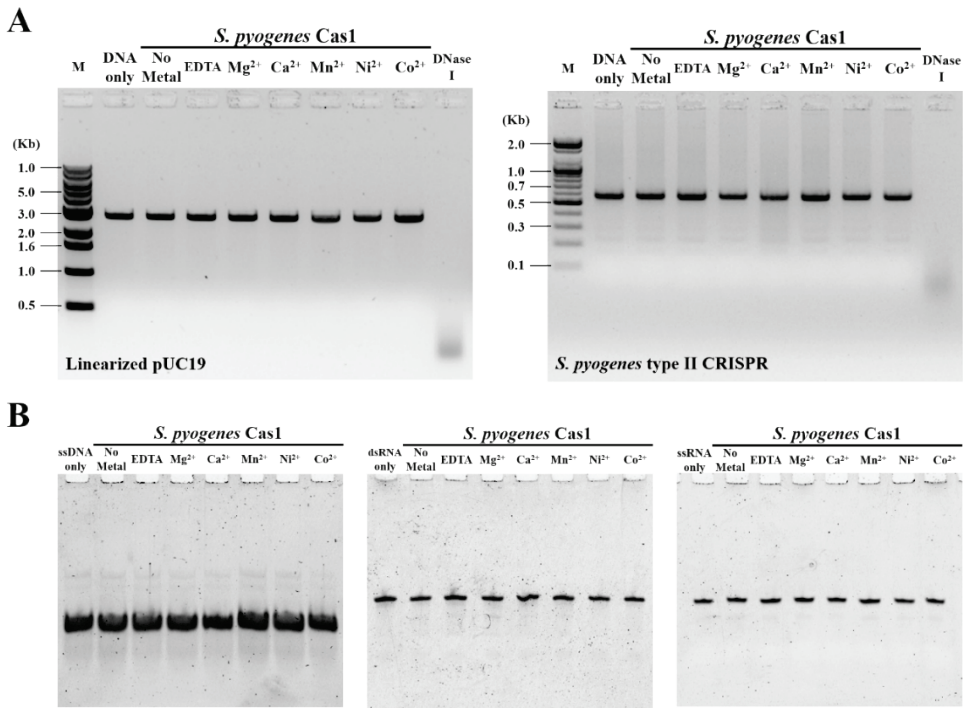
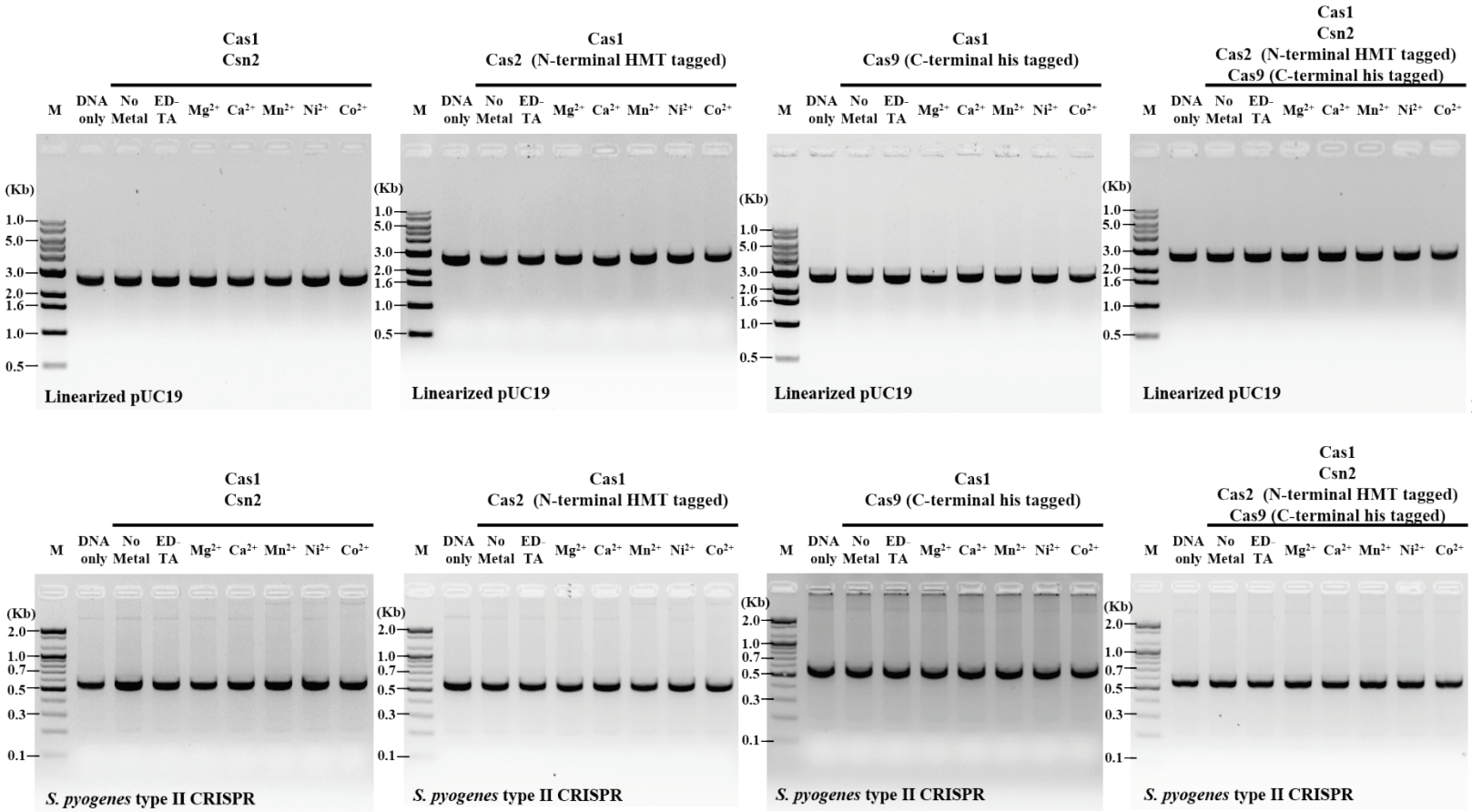

A
S. pyogenes Cas1
M
DNA only
No Metal
EDTA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+
DNase I
(Kb)
1.0
5.0
3.0
2.0
1.6
1.0
0.5
Linearized pUC19
S. pyogenes Cas1
M
DNA only
No Metal
EDTA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+
DNase I
(Kb)
2.0
1.0
0.7
0.5
0.3
0.1
S. pyogenes type II CRISPR
B
S. pyogenes Cas1
ssDNA only
No Metal
EDTA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+
S. pyogenes Cas1
dsRNA only
No Metal
EDTA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+
S. pyogenes Cas1
ssRNA only
No Metal
EDTA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+

Figure 10. dsDNA nuclease activity assay for *S. pyogenes* II-A Cas1 in the presence of one or all of II-A Cas2, Csn2, and Cas9

Linearized pUC19 plasmids or the *S. pyogenes* type II CRISPR sequence fragments were incubated with *S. pyogenes* II-A Cas1 in the presence of one or all of N-terminal $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2, Csn2 and C-terminal $(His)_6$-tagged Cas9.


Cas1
Csn2
Cas1
Cas2 (N-terminal HMT tagged)
Cas1
Cas9 (C-terminal his tagged)
Cas1
Csn2
Cas2 (N-terminal HMT tagged)
Cas9 (C-terminal his tagged)
M
DNA only
No Metal
ED-TA
Mg2+
Ca2+
Mn2+
Ni2+
Co2+
(Kb)
1.0
5.0
3.0
2.0
1.6
1.0
0.5
Linearized pUC19
2.0
1.0
0.7
0.5
0.3
0.1
S. pyogenes type II CRISPR

were also obtained but failed to resolve the X-ray structure. The C-terminal tail was often unresolved in the crystal structures of other sole Cas2 homologs due to insufficient electron density (Nam et al., 2012a; Ka et al., 2014; Ka et al., 2017). Data collection, phasing, and refinement statistics are summarized in Table 4.

Six *S. pyogenes* II-A Cas2 molecules are forming three dimers with non-crystallographic pseudo-two-fold symmetry in asymmetric unit. The three Cas2 dimers structures are very similar, where the RMSD values of Cα atomic positions among them range from 0.7 Å to 1.3 Å. Therefore, the single Cas2 dimer molecule comprised of chain A and B was decided to represent *S. pyogenes* II-A Cas2 dimer structure because it includes the most amino acid residues (residues 6–91). (Fig. 11) The crystal structure of *S. pyogenes* II-A Cas2 reveals the topology and fold that are identical to those of the other homologous structures, including the Cas2 from the *Enterococcus faecalis* type II-A system (Nam et al., 2012a; Ka et al., 2014; Ka et al., 2017; Xiao et al., 2017) (Fig.11A and B). The protomer of *S. pyogenes* II-A Cas2 contains a canonical ferredoxin fold comprised of a four-stranded antiparallel β-sheet (β1–4) and two α-helices (α1, α2) at N-terminus and an additional C-terminal segment including a short α-helix (α3) and a β-strand (β5). The C-terminal β5 strand in one protomer aligns with the β4 strand from the other protomer, extending the antiparallel β-sheet of the N-terminal ferredoxin fold. Through this β5 strand swapping, the two Cas2 molecules form one dimer creating a dimer interface between two antiparallel β-sheets from each protomer. The dimerization of *S. pyogenes* II-A Cas2 forms 22 hydrogen bonds and buries 1,355 $Å^2$ of the surface area.

However, there is a notable conformational difference in the relative protomer orientation at the dimer interface (Fig. 11C). In the *S. pyogenes* II-A Cas2 structure, two protomers orient approximately parallel and are facing each other, while the *E. faecalis* II-A Cas2 dimerizes at a tilted angle, creating V-shape space (Xiao et al., 2017). When those two II-A Cas2 structures are aligned based on only one of their

protomers, the significant structural deviation is noted in the positioning of the other protomer. (Fig. 11D). In this alignment, there is an approximately 43° rigid-body rotation between the two unaligned protomers, and it seems to rotate around α3 helix as a putative hinge part. As the result of this rotation, the two conserved D13 residues at the tip of the β1 strands are further apart in the *E. faecalis* II-A Cas2 dimer (9.6 Å) than in the *S. pyogenes* II-A Cas2 structure (3.2 Å). However, it is unclear which the conformation of the Cas2 dimer structure is biologically relevant, and both may be the possible forms of II-A Cas2 dimer (see discussion).

Characterization of *S. pyogenes* II-A Cas1–Cas2 complex formation

In *E. coli* type I-E and *E. faecalis* type II-A systems, Cas1 and Cas2 proteins form a heterohexameric integrase complex, in which a single Cas2 dimer serves as a bridge between two Cas1 dimers at both ends (Nunez et al., 2014; Xiao et al., 2017). Disturbance of the interaction between the Cas1 and Cas2 proteins obstructed spacer acquisition *in vivo* (Nunez et al., 2014), indicating that the Cas1–Cas2 complex formation is crucial for the adaptation stage of CRISPR-Cas system. In this study, several biochemical analyses were performed to characterize the interaction between Cas1 and Cas2 from *S. pyogenes* type II-A CRISPR-Cas system. First, with individually purified *S. pyogenes* II-A Cas1 and Cas2, the complex formation was confirmed by analytical SEC (Fig. 12A). All results were obtained using the N-terminal $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2. The full-length *S. pyogenes* type II-A Cas2 without solubility tag was not suitable for the binding assays requiring low salt concentration condition (150 mM NaCl) and was soluble at only a high salt concentration above 500 mM NaCl. In the analytical SEC, the pre-incubated *S. pyogenes* type II-A Cas1 and $(His)_6$-MBP tagged *S. pyogenes* type II-A Cas2 migrated together with one peak and showed shorter retention time than those of the individual proteins, implying that the two Cas proteins form a complex. Then the quantitative analysis was perform using ITC for obtaining their thermodynamic

Figure 11. X-ray crystal structure of *S. pyogenes* II-A Cas2 and structural comparison with *E. faecalis* II-A Cas2

(A) The monomer structure is shown in rainbow color from the N-terminus (blue) to the C-terminus (red). Secondary structure elements are also indicated. (B) The sequence alignment of *S. pyogenes* (Sp) and *E. faecalis* (Ef) II-A Cas2 based on the two protomer structures. Secondary structure elements of the two Cas2 homologs are indicated (Top, *S. pyogenes* II-A Cas2; bottom, *E. faecalis* II-A Cas2). (C) Side-by side comparison of *S. pyogenes* and *E. faecalis* II-A Cas2 dimeric structures. (D) Structural alignment of the two type II-A Cas2 dimers based on one of the two protomers. The rigid-body rotation between the two unaligned protomers is observed.

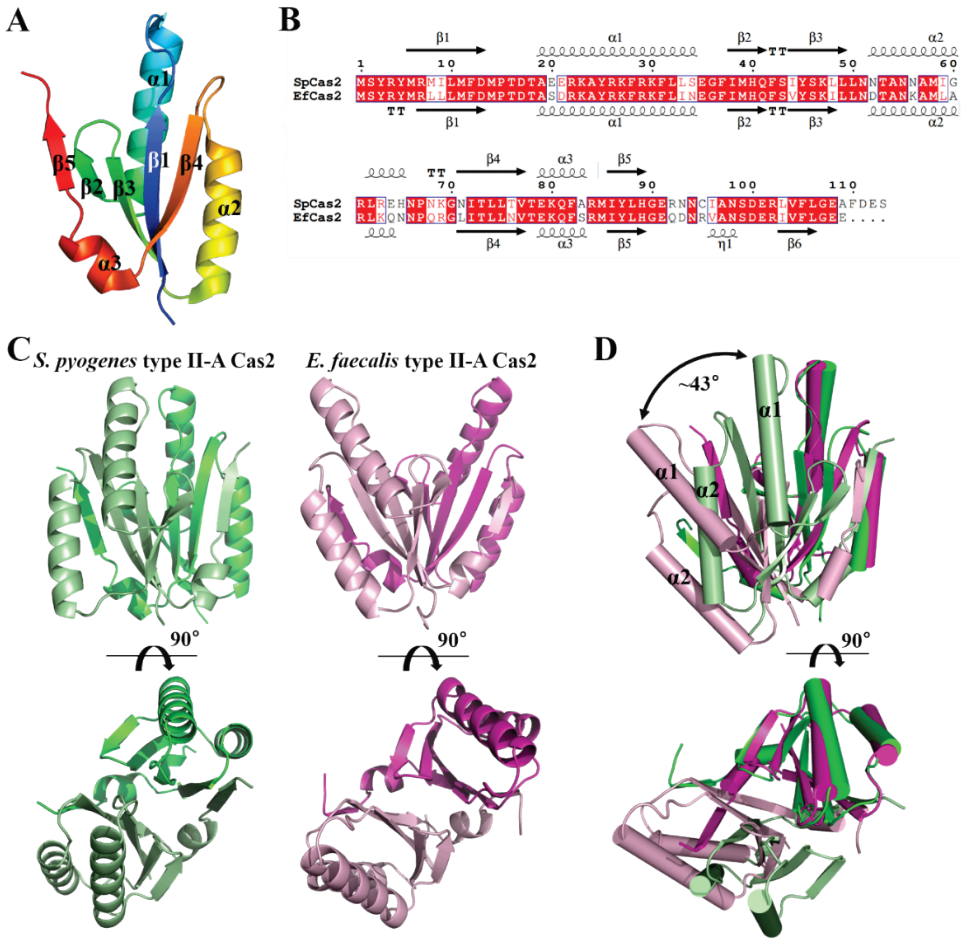
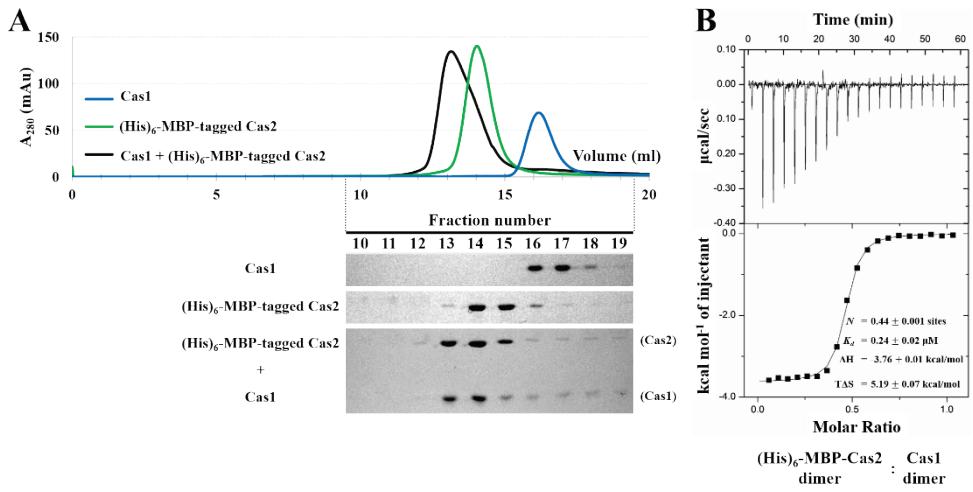

A
α1
β5
β2
β3
β1
β4
α2
α3
B
β1
α1
β2
β3
α2
SpCas2
EfCas2
β4
α3
β5
η1
β6
C
S. pyogenes type II-A Cas2
E. faecalis type II-A Cas2
90°
90°
D
~43°
α1
α1
α2
α2
90°

Figure 12. Interaction between *S. pyogenes* II-A Cas1 and Cas2

(A) Interaction between individually purified *S. pyogenes* II-A Cas1 (20 μM) and $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 (20 μM) determined by analytical SEC. Elution fractions were also analyzed by SDS-PAGE. (B) ITC analysis of the binding between $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 and the *S. pyogenes* II-A Cas1. 400 μM of the *S. pyogenes* II-A Cas2 was added consecutively to the chamber containing 80 μM of *S. pyogenes* II-A Cas1. $K_d(= 1/K_a)$ and ΔH value were determined from ITC profile fitting to Origin 7 software. The $T\Delta S$ value was determined using the equations $\Delta G = -RT \ln K_a$ and $T\Delta S = \Delta H - \Delta G$.


A
150
100
50
0
A_{280} (mAu)
Cas1
(His)6-MBP-tagged Cas2
Cas1 + (His)6-MBP-tagged Cas2
Volume (ml)
0
5
10
15
20
Fraction number
10 11 12 13 14 15 16 17 18 19
Cas1
(His)6-MBP-tagged Cas2
(His)6-MBP-tagged Cas2
+
Cas1
(Cas2)
(Cas1)
B
Time (min)
0 10 20 30 40 50 60
0.10
0.00
-0.10
-0.20
-0.30
-0.40
µcal/sec
0.0
-2.0
-4.0
kcal mol-1 of injectant
N = 0.44 ± 0.001 sites
Ka = 0.24 ± 0.02 µM
ΔH = -3.76 ± 0.01 kcal/mol
TΔS = 5.19 ± 0.07 kcal/mol
0.0
0.5
1.0
Molar Ratio
(His)6-MBP-Cas2 dimer : Cas1 dimer

properties, including the reaction stoichiometry during the binding event between *S. pyogenes* II-A Cas1 and Cas2 (Fig. 12B). The $(His)_6$-MBP-tagged *S. pyogenes* type II-A Cas2 was gradually titrated into the cell containing the *S. pyogenes* type II-A Cas1. The submicromolar binding affinity (K_d = 0.24 μM) for the complex formation was derived using the multiple identical independent binding sites model fitting. The reaction stoichiometry (N) of $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 to *S. pyogenes* II-A Cas1 was calculated to be 0.44 at the center of the binding isotherm. This binding stoichiometry is also consistent with the I-E Cas1–Cas2 heterohexameric complex formation in which two Cas1 dimers and one Cas2 dimer participated.

In the *E. coli* and *E. faecalis* Cas1–Cas2 complexes, the C-terminal tail of the Cas2 mediates the interaction between Cas1 and Cas2 (Nunez et al., 2014; Xiao et al., 2017). To determine whether, and to what extent, the C-terminal tail of *S. pyogenes* II-A Cas2 contributes to the interaction with *S. pyogenes* II-A Cas1, I analyzed the binding of the truncated Cas2 variants to the *S. pyogenes* II-A Cas1. In the analytical SEC, the C-terminally truncated variant (residues 1–91) of the $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 did not interact with the Cas1, whereas the $(His)_6$-MBP-tagged C-terminal tail of *S. pyogenes* II-A Cas2 co-eluted with the Cas1 (Fig. 13A and B). In the ITC experiment using a peptide corresponding to the C-terminal tail of *S. pyogenes* II-A Cas2, the stoichiometry (N) and dissociation constant (K_d) were determined to be almost identical to those obtained with the full-length *S. pyogenes* II-A Cas2 (Fig. 13C). These results suggest that the C-terminal tail region of *S. pyogenes* II-A Cas2 is primarily responsible for the interaction with *S. pyogenes* II-A Cas1 in the formation of Cas1–Cas2 complex.

Complex formation of *S. pyogenes* II-A Cas1 with a type II-A specific Cas protein, Csn2

Csn2 is a type II-A specific Cas protein involved in the spacer acquisition process

during the type II-A CRISPR-mediated immunity. Also, Csn2 possessed the affinity to dsDNA (Nam et al., 2011; Koo et al., 2012; Arslan et al., 2013). Despite the specific role of Csn2 in adaptation is enigmatic, Csn2 is essential for adaptation *in vivo*. (Barrangou et al., 2007; Heler et al., 2015; Wei et al., 2015). Since Csn2 was not required for prespacer DNA integration, it has been assumed that Csn2 is involved in a different step of adaptation mechanism, such as the generation of prespacers (Wright and Doudna, 2016). Recently, it has been reported that the four Cas proteins of the type II-A CRISPR-Cas system can interact with each other. (Heler et al., 2015). Thus, I tried to determine the detailed interaction network among the type II-A Cas proteins that are functionally related. When C-terminal $(His)_6$-tagged *S. pyogenes* type II-A Cas1 was expressed with untagged *S. pyogenes* Csn2 in *E. coli* BL21 (DE3) cells, the Csn2 was co-purified following nickel-affinity chromatography, while the untagged Csn2 showed no affinity to the nickel column (Figure 14A and table 7). Subsequently, Size-exclusion chromatography method was used to demonstrate the direct interaction between *S. pyogenes* II-A Cas1 and Csn2. The pre-incubated Cas1 and Csn2 were co-eluted at a shorter retention time than those of the two individual proteins alone, suggesting the formation of *S. pyogenes* II-A Cas1–Csn2 complex (Fig. 14B).

Additional interaction tests were performed with increasing amounts of the *S. pyogenes* II-A Cas1 to assess the stoichiometry of the complex formation (Fig. 15A). As the result, the peak corresponding to the complex formation showed more height and decreased retention time until the molar ratio between the *S. pyogenes* II-A Cas1 and Csn2 reached ~ 1:1. Considering the oligomeric state of the two proteins, this result suggests that two Cas1 dimers can bind to one Csn2 tetramer. The ITC experiment was also performed for a detailed quantitative investigation of the interaction between the *S. pyogenes* II-A Cas1 and Csn2 (Fig. 15B). In the ITC experiment, *S. pyogenes* II-A Cas1 needed an N-terminal $(His)_6$-MBP-tag for stability, otherwise a slight precipitation occurred during ITC assay. The solubility

Figure 13. Interaction tests between *S. pyogenes* II-A Cas1 and truncated *S. pyogenes* II-A Cas2 variants

(A and B) *S. pyogenes* II-A Cas1 (40 μM) did not interact with the C-terminal-truncated form (residues 1–91) of *S. pyogenes* II-A Cas2 (A), but formed a stable complex with the C-terminal tail (residues 92–113) of *S. pyogenes* II-A Cas2 (B). Both two *S. pyogenes* II-A Cas2 variants (20 μM) were purified with an N-terminal $(His)_6$-MBP-tag for their solubility. (C) ITC analysis for the binding of the C-terminal tail part of *S. pyogenes* II-A Cas2 to *S. pyogenes* II-A Cas1. The 400 μM C-terminal tail of *S. pyogenes* II-A Cas2 sample was injected into 80 μM of *S. pyogenes* II-A Cas1. The N and K_d values are also determined experimentally. The peptide corresponding to the C-terminal tail of *S. pyogenes* II-A Cas2 (residues 92–113) was commercially synthesized (BIONICS, Korea). ITC derived thermodynamic parameters for the binding are calculated as mentioned in Fig 12B.

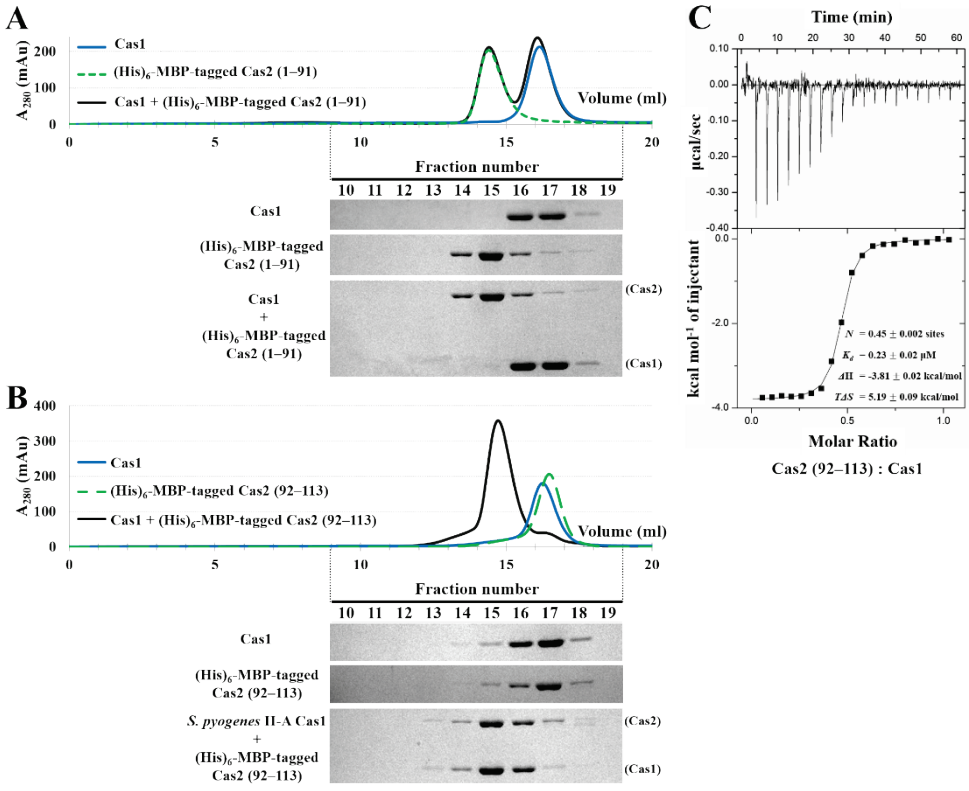

A
Cas1
(His)6-MBP-tagged Cas2 (1–91)
Cas1 + (His)6-MBP-tagged Cas2 (1–91)
A280 (mAu)
Volume (ml)
Fraction number
Cas1
(His)6-MBP-tagged Cas2 (1–91)
Cas1 + (His)6-MBP-tagged Cas2 (1–91)
(Cas2)
(Cas1)
B
Cas1
(His)6-MBP-tagged Cas2 (92–113)
Cas1 + (His)6-MBP-tagged Cas2 (92–113)
S. pyogenes II-A Cas1 + (His)6-MBP-tagged Cas2 (92–113)
C
Time (min)
µcal/sec
kcal mol-1 of injectant
N = 0.45 ± 0.002 sites
Ka = 0.23 ± 0.02 µM
ΔH = -3.81 ± 0.02 kcal/mol
TΔS = 5.19 ± 0.09 kcal/mol
Molar Ratio
Cas2 (92–113) : Cas1

Figure 14. Direct interaction between *S. pyogenes* II-A Cas1 and type II-A specific Cas protein, Csn2

(A) Co-purification of *S. pyogenes* II-A Cas1 and Csn2. Fractions for total soluble expressed proteins (T), flow-through (F), and 450 mM imidazole-elution (E) were analyzed by SDS-PAGE. A control experiment expressing the untagged *S. pyogenes* II-A Cas1 alone are also shown. Separately purified C-terminal $(His)_6$-tagged *S. pyogenes* II-A Cas1 and untagged Csn2 were analyzed together on the right side for comparison. (B) The *S. pyogenes* II-A Cas1 and Csn2 form a complex in size-exclusion chromatography. Separately purified *S. pyogenes* II-A Cas1 (20 μM) and Csn2 (40 μM) were used for the experiment.

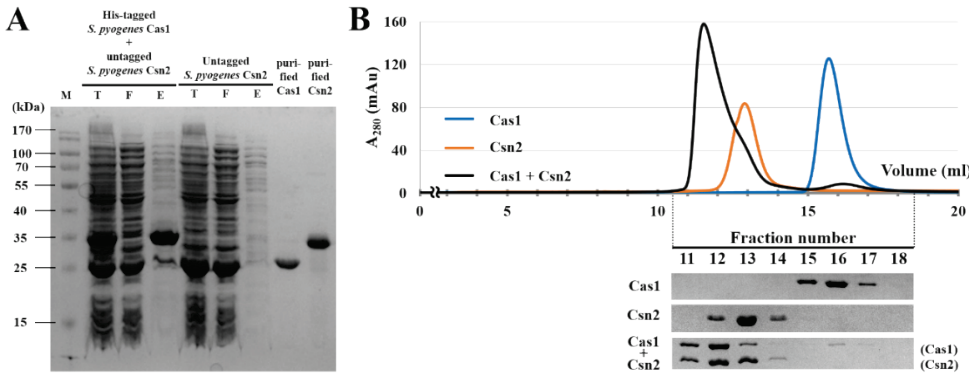

A
His-tagged
S. pyogenes Cas1
+
untagged
S. pyogenes Csn2
Untagged
S. pyogenes Csn2
puri-fied Cas1
puri-fied Csn2
M
T
F
E
T
F
E
(kDa)
170
100
70
55
40
35
25
15
B
160
120
80
40
0
A280 (mAu)
Cas1
Csn2
Cas1 + Csn2
Volume (ml)
0
5
10
15
20
Fraction number
11 12 13 14 15 16 17 18
Cas1
Csn2
Cas1
+
Csn2
(Cas1)
(Csn2)

Table 7. Mass spectrometry analysis of protein bands from the *S .pyogenes* II-A Cas1 and Csn2 co-purification experiment. (Related to Fig. 14A)

Protein	% Coverage	Total peptides	Unique peptides	MS/MS counts	Score	Intensity (cps)
S. pyogenes II-A Cas1	88.6	28	2	121	323.31	1.85×10^{11}
S. pyogenes II-A Csn2	91.8	14	2	119	293.04	1.87×10^{11}

Two major bands from Lane 4 (E, $(His)_6$-tagged Cas1 + untagged Csn2) of Figure 16A SDS-PAGE result were analyzed by using LC-MS/MS after trypsin digestion.

tag of *S. pyogenes* II-A Cas1 did not interfere the *S. pyogenes* II-A Cas1–Csn2 complex formation (Fig. 15C). The dissociation constant (K_d) of the *S. pyogenes* II-A Cas1–Csn2 interaction was determined to be ~ 5.0 μM about 20 times weaker than the value identified between the *S. pyogenes* II-A Cas1 and Cas2. The binding ratio of the Cas1 dimer to the Csn2 tetramer was calculated to be ~ 1.9:1 suggesting that two Cas1 dimers can interact with one Csn2 tetramer.

Identification of the binding interface of *S. pyogenes* II-A Cas1–Csn2 complex

Further experiments were performed to designate the potential binding interface between the two II-A Cas proteins. *S. pyogenes* II-A Cas1 has two parts which are divided into the N-terminal and C-terminal domain (Fig 5A and B). To determine which domain of *S. pyogenes* II-A Cas1 is involved in the interaction, each domain of the Cas1 was prepared and tested the interaction with the *S. pyogenes* Csn2 (Fig. 16A and B). In the analytical size-exclusion chromatography, the N-terminal domain of *S. pyogenes* II-A Cas1 eluted earlier with *S. pyogenes* Csn2. The N-terminal $(His)_6$-MBP-tag was added for the solubility because the sole N-terminal domain part was not soluble in the buffer condition used for the protein purification. The C-terminal domain of the *S. pyogenes* II-A Cas1 did not interact with *S. pyogenes* Csn2 and showed the unchanged retention time comparing those of the C-terminal domain of the Cas1 alone. This results suggest that only the N-terminal domain of *S. pyogenes* II-A Cas1 is responsible for the interaction with *S. pyogenes* Csn2. Interestingly, as the concentration of NaCl increased, the interaction between *S. pyogenes* II-A Cas1 and Csn2 was disrupted. At a concentration of 500 mM NaCl, the two proteins were utterly separated into two peaks corresponding to Cas1 and Csn2, respectively (Fig. 16C). The disruption of the complex formation in high salt environment implies that ionic interaction play a crucial role in the formation of *S. pyogenes* II-A Cas1–Csn2 complex.

Since *S. pyogenes* Csn2 is a stable tetramer, it was hard to select a specific

Figure 15. Binding stoichiometry analysis of the *S. pyogenes* II-A Cas1–Csn2 complex formation

(A) Binding stoichiometry analysis using by size-exclusion chromatography analysis. Increasing amounts of the *S. pyogenes* II-A Cas1 were incubated with Csn2 and analyzed by size-exclusion chromatography. (B) ITC trace for *S. pyogenes* II-A Cas1 binding to Csn2. N-terminal $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas1 was added consecutively to the chamber containing Csn2 because untagged the Cas1 precipitated under the experimental conditions used for ITC analyses. ITC derived thermodynamic parameters for the binding are calculated as mentioned in Fig 12B. (C) The *S. pyogenes* II-A Cas1 (20 μM) with Csn2 (40 μM) was confirmed in analytical size-exclusion chromatography before trying ITC analysis.

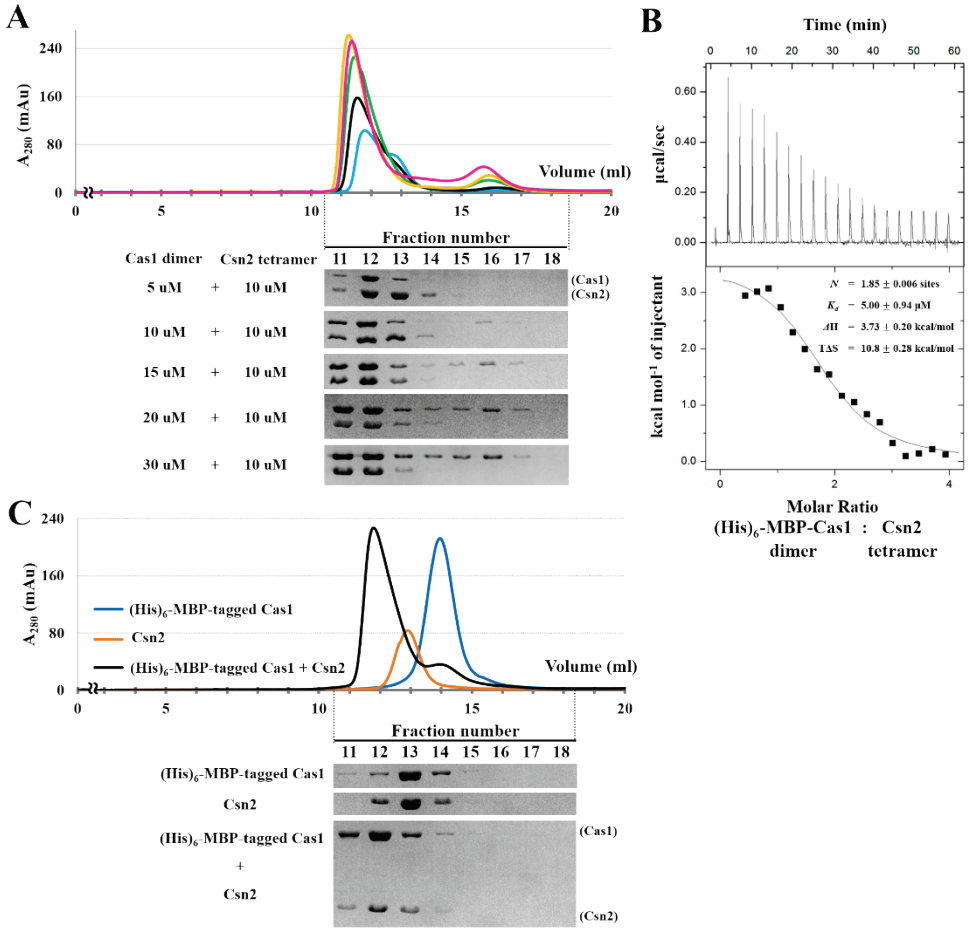
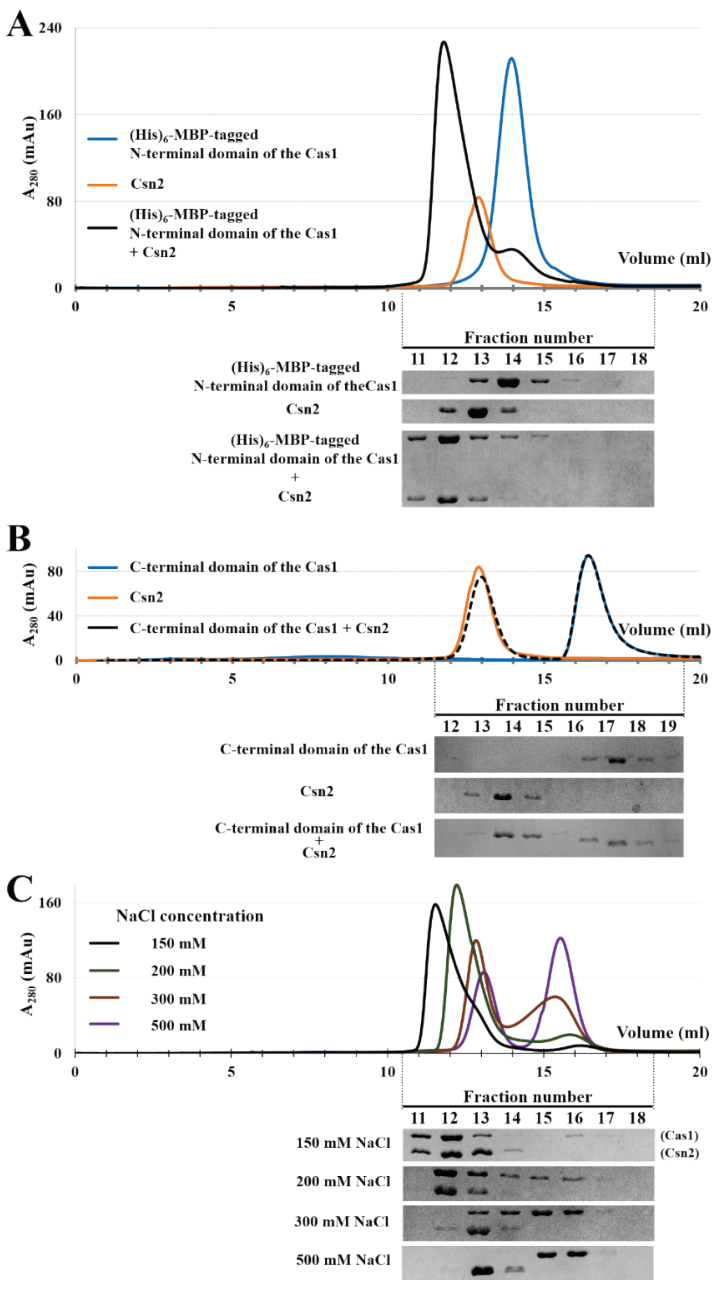

A
240
160
80
0
A_{280} (mAu)
Volume (ml)
0
5
10
15
20
Fraction number
Cas1 dimer
Csn2 tetramer
11 12 13 14 15 16 17 18
5 uM + 10 uM
10 uM + 10 uM
15 uM + 10 uM
20 uM + 10 uM
30 uM + 10 uM
(Cas1)
(Csn2)
B
Time (min)
0 10 20 30 40 50 60
µcal/sec
0.60
0.40
0.20
0.00
kcal mol-1 of injectant
3.0
2.0
1.0
0.0
N = 1.85 ± 0.006 sites
K_d = 5.00 ± 0.94 µM
ΔH = -3.73 ± 0.20 kcal/mol
TΔS = 10.8 ± 0.28 kcal/mol
0
2
4
Molar Ratio
(His)6-MBP-Cas1 : Csn2
dimer tetramer
C
240
160
80
0
A_{280} (mAu)
(His)6-MBP-tagged Cas1
Csn2
(His)6-MBP-tagged Cas1 + Csn2
Volume (ml)
0
5
10
15
20
Fraction number
11 12 13 14 15 16 17 18
(His)6-MBP-tagged Cas1
Csn2
(His)6-MBP-tagged Cas1
+
Csn2
(Cas1)
(Csn2)

Figure 16. Characterization of the interaction between *S. pyogenes* II-A Cas1 and Csn2

(A) Interaction test between N-terminal of *S. pyogenes* II-A Cas1 (20 μM) and Csn2 (40 μM). Since the untagged N-terminal domain of the Cas1 was not soluble under the analysis condition, the $(His)_6$-MBP-tagged form of the *S. pyogenes* II-A Cas1 was used (B) Interaction test between C-terminal of *S. pyogenes* II-A Cas1 (20 μM) and Csn2 (40 μM). Dotted chromatograms were applied to show the overlapping clearly. (C) Effect of salt concentration on the *S. pyogenes* II-A Cas1–Csn2 complex formation. The behavior of the complexes in the different salt concentration buffers was monitored by analytical chromatography.


A
240
160
80
0
A_{280} (mAu)
(His)$_6$-MBP-tagged N-terminal domain of the Cas1
Csn2
(His)$_6$-MBP-tagged N-terminal domain of the Cas1 + Csn2
Volume (ml)
0
5
10
15
20
Fraction number
11 12 13 14 15 16 17 18
(His)$_6$-MBP-tagged N-terminal domain of theCas1
Csn2
(His)$_6$-MBP-tagged N-terminal domain of the Cas1 + Csn2
B
80
40
0
A_{280} (mAu)
C-terminal domain of the Cas1
Csn2
C-terminal domain of the Cas1 + Csn2
Volume (ml)
0
5
10
15
20
Fraction number
12 13 14 15 16 17 18 19
C-terminal domain of the Cas1
Csn2
C-terminal domain of the Cas1 + Csn2
C
160
80
0
A_{280} (mAu)
NaCl concentration
150 mM
200 mM
300 mM
500 mM
Volume (ml)
0
5
10
15
20
Fraction number
11 12 13 14 15 16 17 18
150 mM NaCl
200 mM NaCl
300 mM NaCl
500 mM NaCl
(Cas1)
(Csn2)

domain for the interaction test. Nevertheless, the symmetric structure of *S. pyogenes* Csn2 was suitable for the docking simulation to deduce the binding interface between the N-terminal domain of *S. pyogenes* II-A Cas1 and the Csn2 tetramer. To predict the overall structure and the detailed potential binding interface of the *S. pyogenes* II-A Cas1–Csn2 complex, *in silico* docking simulation was performed using available *S. pyogenes* II-A Cas1 and Csn2 structures. In the calculated *S. pyogenes* II-A Cas1–Csn2 complex model, the bottom face of the dimeric N-terminal domain part of the Cas1 participates in the interface with the α/β domains of the Csn2 (Fig. 17A). Among the interface residues, I paid attention to the three lysine residues (K14, K18, and K25) of *S. pyogenes* II-A Cas1, and two glutamate residues (E23 and E199) of *S. pyogenes* Csn2. Since the size-exclusion chromatography analysis showed that the complex formation is dependent on ionic interaction, these charged residues could be responsible for the complex formation (Fig. 17B). One of the three lysine residues of *S. pyogenes* II-A Cas1 was mutated to alanine for the experimental confirmation of *in silico* docking simulation. Three *S. pyogenes* II-A Cas1 mutants (K14A, K18A, and K25A) were prepared and their interactions with the *S. pyogenes* Csn2 was examined in size-exclusion chromatography (Fig. 17C and D). Of the three mutants, the K14A mutant did not interact with *S. pyogenes* Csn2 at all and showed a simple sum of the chromatograms of *S. pyogenes* II-A Cas1 and Csn2 alone. It implies that the K14 residue is critical for the formation of the *S. pyogenes* II-A Cas1–Csn2 complex. Also, the K25A mutant displayed a significantly different chromatogram pattern compared to wild-type *S. pyogenes* II-A Cas1. The complex of *S. pyogenes* II-A Cas1 K25A mutant and Csn2 was eluted considerably later than the wild type complex. The peak corresponding to the unbound *S. pyogenes* II-A Cas1 was also shown. K18A mutant of *S. pyogenes* II-A Cas1 displayed a chromatogram similar to that of wild type *S. pyogenes* II-A Cas1 but had a slightly longer retention time for the complex. This suggests that the K18 residue is the least essential residue for the complex formation among the three lysine residues in *S.*

pyogenes II-A Cas1. In the interface of the predicted complex structure model, both K14 and K25 of *S. pyogenes* II-A Cas1 form the core of the binding interface, while K18 residue is found at the edge of the binding interface (Fig. 17B). It is also noteworthy that K18 is less conserved than K14 and K25 in type II Cas1 homologs (Fig. 18). K14 is strictly conserved, and K25 is either conserved or mutated to arginine in most of type II Cas1 proteins. Interestingly, these three lysine residues are not conserved in the type I Cas1 homologs (Fig. 6). Moreover, *S. pyogenes* Csn2 did not interact with type I-F Cas1 from *Xanthomonas campestris*, implying that Csn2 interacts with II-A Cas1 specifically (Fig. 19).

In the interface of the Cas1–Csn2 docking model, two negatively charged residues (E23 and E199) from the α/β domain of *S. pyogenes* Csn2 are located near the three lysine residues of *S. pyogenes* II-A Cas1. To further validate the *in silico* docking prediction, I generated two *S. pyogenes* Csn2 mutants (E23A and E199A) and tested their interaction with the wild-type *S. pyogenes* II-A Cas1 (Fig. 17B). The E23A mutation of *S. pyogenes* Csn2 reduced the affinity of Csn2 for *S. pyogenes* II-A Cas1, whereas the E199A mutant of the Csn2 showed the same chromatogram as the wild-type Csn2. These results indicate that the E23 residue is involved in the interaction with *S. pyogenes* II-A Cas1 for complex formation. E23 residue is well conserved in other Csn2 homologs, but the E199 is not (Fig. 20). Taken together, I concluded that the computational prediction for the binding interface is valid, and the charged residues at the *S. pyogenes* II-A Cas1–Csn2 binding interface participate in the electrostatic interactions. The conservation of the charged residues suggests that the formation of II-A Cas1–Csn2 complex is a common feature of the type II-A CRISPR-Cas systems.

Three-component complex formation in *S. pyogenes* Type II-A CRISPR-Cas system

S. pyogenes II-A Cas1, Cas2, and Csn2 are required for type II-A CRISPR adaptation

Figure 17. *In silico* docking model of *S. pyogenes* II-A Cas1–Csn2 complex

(A) In the overall view of the model, The *S. pyogenes* II-A Cas1 dimer is colored as in blue and cyan for each protomer and *S. pyogenes* Csn2 is shown in orange and bright orange for chain A and B in the asymmetric unit. Tetrameric arrangement of the Csn2 was obtained by generating symmetry mate from crystal lattice information. (B) Enlarged view of the putative *S. pyogenes* II-A Cas1–Csn2 complex binding interface (the rectangle indicated by a dotted line) in the *in silico* docking model (right). The candidate residues suspected to be involved in electrostatic interactions between the Cas1 and Csn2 are annotated. (C) The analytical SEC using three *S. pyogenes* II-A Cas1 mutants (20 μM) and Wild type (WT) *S. pyogenes* Csn2 (40 μM). (D) The same analytical SEC but using *S. pyogenes* Csn2 mutants and Wild type (WT) II-A Cas1.

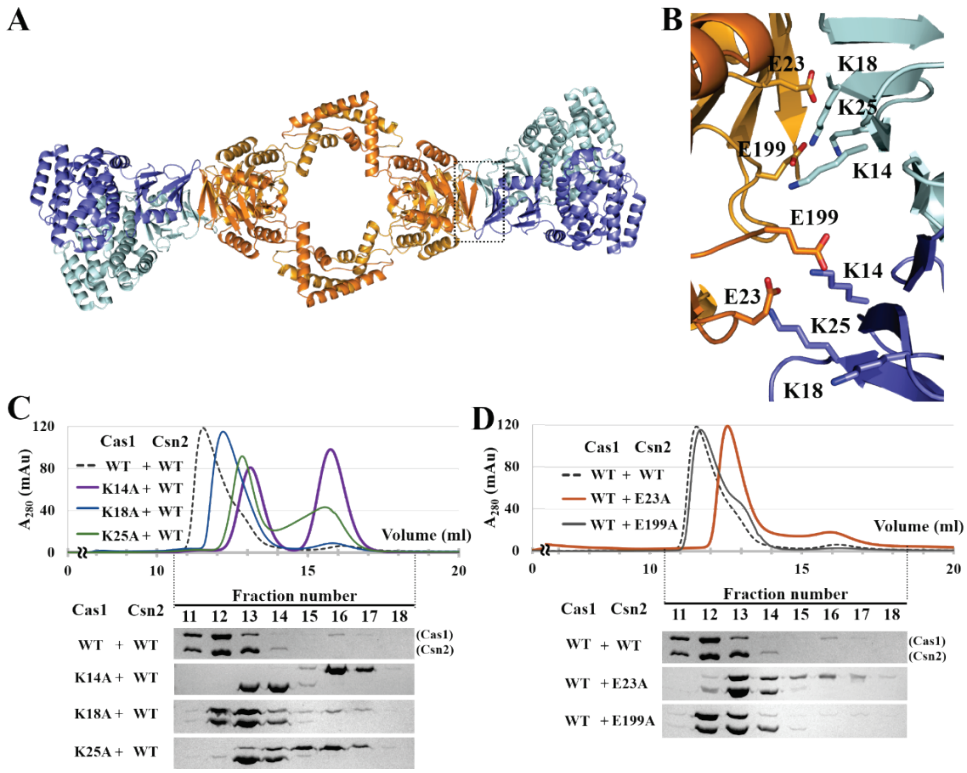

A
B
E23
K18
K25
E199
K14
E199
K14
E23
K25
K18
C
Cas1 Csn2
WT + WT
K14A + WT
K18A + WT
K25A + WT
A280 (mAu)
Volume (ml)
Fraction number
11 12 13 14 15 16 17 18
(Cas1)
(Csn2)
D
Cas1 Csn2
WT + WT
WT + E23A
WT + E199A
A280 (mAu)
Volume (ml)
Fraction number
11 12 13 14 15 16 17 18
(Cas1)
(Csn2)

Figure 18. Sequence alignment of Cas1 homologues from type II-A CRISPR-Cas systems

The blue triangles indicate the residues predicted to participate in electrostatic interactions for the Cas1–Csn2 complex formation.

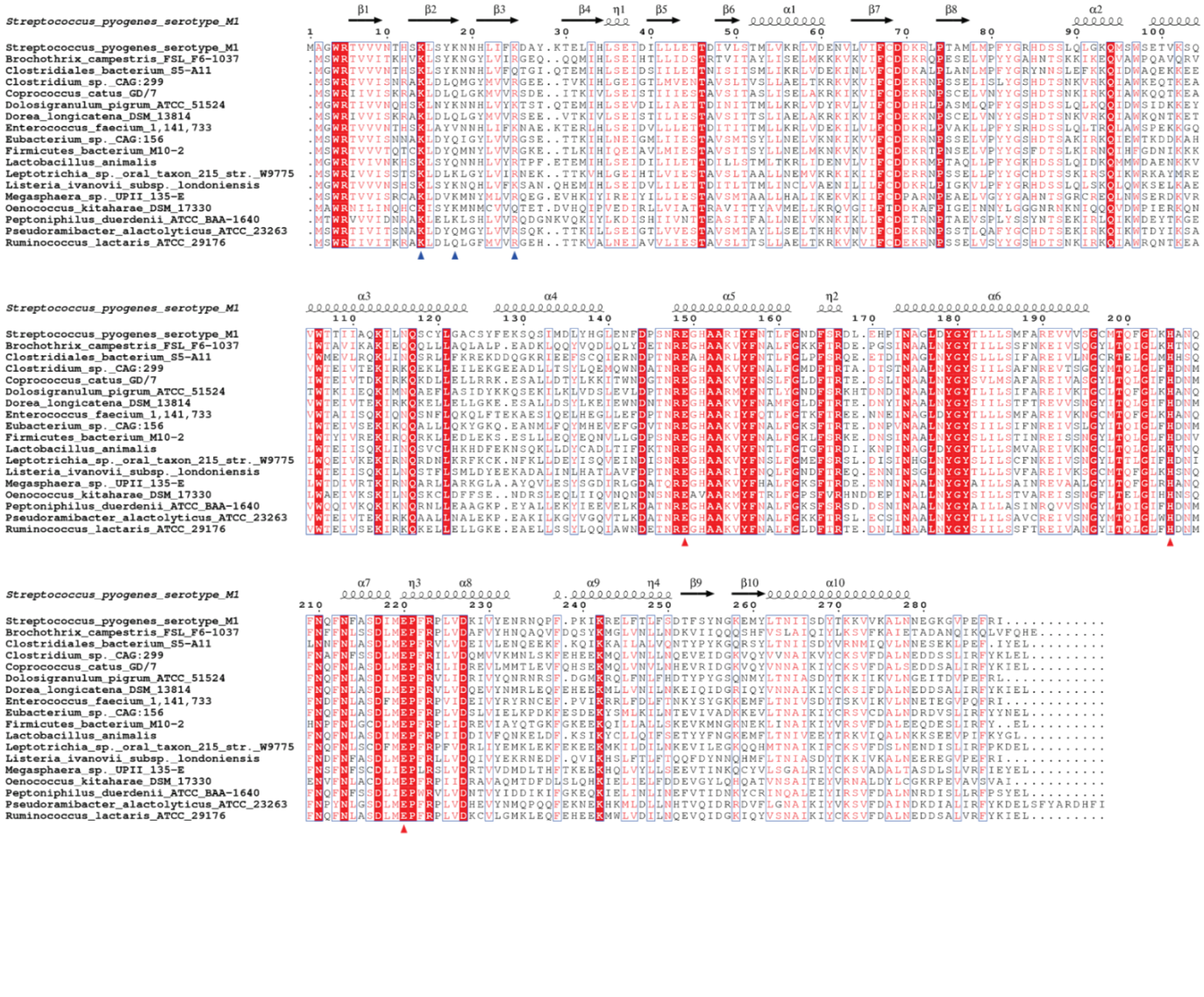

Streptococcus_pyogenes_serotype_M1
Streptococcus_pyogenes_serotype_M1
Brochothrix_campestris_FSL_F6-1037
Clostridiales_bacterium_S5-A11
Clostridium_sp._CAG:299
Coprococcus_catus_GD/7
Dolosigranulum_pigrum_ATCC_51524
Dorea_longicatena_DSM_13814
Enterococcus_faecium_1,141,733
Eubacterium_sp._CAG:156
Firmicutes_bacterium_M10-2
Lactobacillus_animalis
Leptotrichia_sp._oral_taxon_215_str._W9775
Listeria_ivanovii_subsp._londoniensis
Megasphaera_sp._UPII_135-E
Oenococcus_kitaharae_DSM_17330
Peptoniphilus_duerdenii_ATCC_BAA-1640
Pseudoramibacter_alactolyticus_ATCC_23263
Ruminococcus_lactaris_ATCC_29176

Figure 19. Interaction test between *S. pyogenes* Csn2 and *X. campestris* I-F Cas1

Dotted chromatograms were applied to show the overlapping clearly.

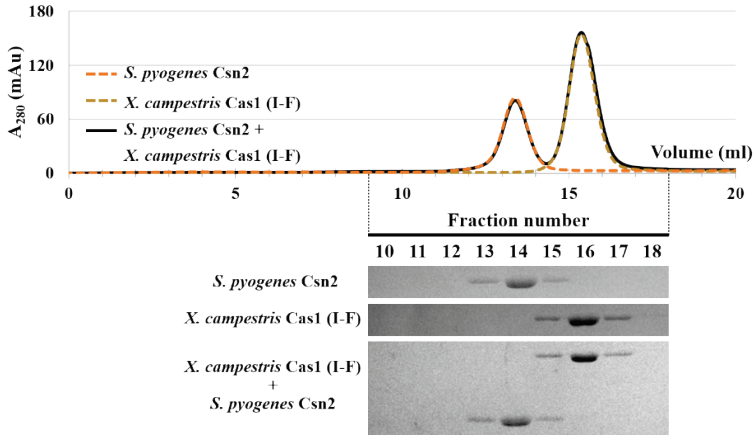
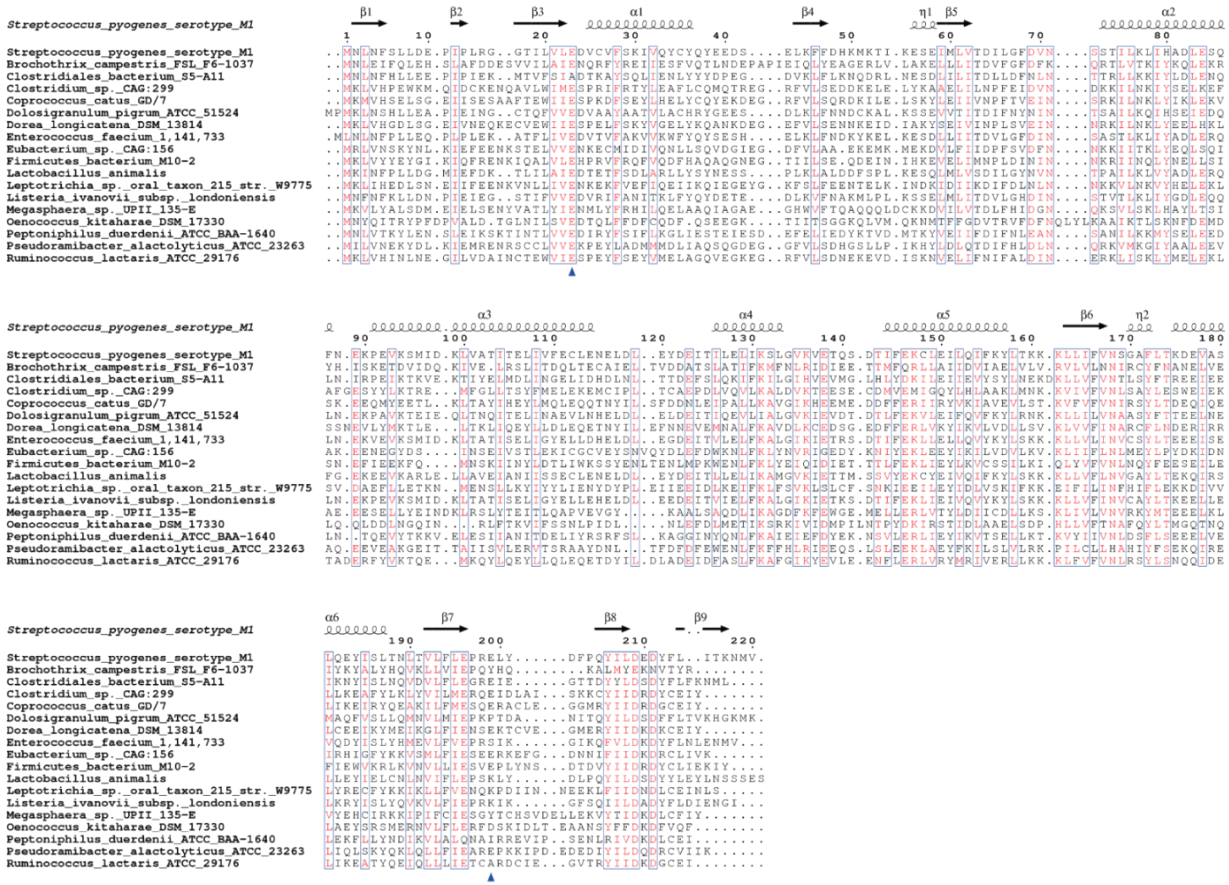

180
120
60
0
A_{280} (mAu)
S. pyogenes Csn2
X. campestris Cas1 (I-F)
S. pyogenes Csn2 +
X. campestris Cas1 (I-F)
Volume (ml)
0
5
10
15
20
Fraction number
10 11 12 13 14 15 16 17 18
S. pyogenes Csn2
X. campestris Cas1 (I-F)
X. campestris Cas1 (I-F)
+
S. pyogenes Csn2

Figure 20. Sequence alignment of Csn2 homologs from type II-A CRISPR-Cas systems

Residues predicted to participate in electrostatic interactions for the *S. pyogenes* II-A Cas1-Csn2 complex formation are indicated by blue triangles.


Streptococcus_pyogenes_serotype_M1
Streptococcus_pyogenes_serotype_M1
Brochothrix_campestris_FSL_F6-1037
Clostridiales_bacterium_S5-A11
Clostridium_sp._CAG:299
Coprococcus_catus_GD/7
Dolosigranulum_pigrum_ATCC_51524
Dorea_longicatena_DSM_13814
Enterococcus_faecium_1,141,733
Eubacterium_sp._CAG:156
Firmicutes_bacterium_M10-2
Lactobacillus_animalis
Leptotrichia_sp._oral_taxon_215_str._W9775
Listeria_ivanovii_subsp._londoniensis
Megasphaera_sp._UPII_135-E
Oenococcus_kitaharae_DSM_17330
Peptoniphilus_duerdenii_ATCC_BAA-1640
Pseudoramibacter_alactolyticus_ATCC_23263
Ruminococcus_lactaris_ATCC_29176

(Barrangou et al., 2007; Heler et al., 2015; Wei et al., 2015). Since the *S. pyogenes* II-A Cas1 has interactions with both *S. pyogenes* II-A Cas2 and Csn2 (Fig. 12 and Fig. 14), I have investigated whether these three II-A Cas proteins form a single three-component Cas complex. To confirm the formation of the putative type II-A CRISPR adaptation module, The analytical SEC was performed with the *S. pyogenes* II-A Cas1–Cas2 complex and Csn2. For preparing the II-A Cas1–Cas2 complex, Cas1 and $(His)_6$-MBP-tagged Cas2 were co-expressed in *E. coli* cells. Both proteins were purified together by affinity chromatography (Fig. 21A). Also, they co-migrated as one peak in size-exclusion chromatography. The elution volume of the co-expressed *S. pyogenes* II-A Cas1–Cas2 complex was identical to that of the reconstituted complex for which the separately purified Cas1 and the $(His)_6$-MBP-Cas2 are mixed at a 2:1 ratio (Fig. 21B). From the result, it was confirmed that *S. pyogenes* II-A Cas1 and Cas2 form a stoichiometric complex by the co-expression method. I have tried to remove the $(His)_6$-MBP tag from the *S. pyogenes* II-A Cas2; unfortunately, the untagged Cas2 even precipitated with the bound Cas1 at low salt concentrations below 300 mM NaCl. Therefore, the *S. pyogenes* II-A Cas1–$(His)_6$-MBP-Cas2 complex was used for the binding assays requiring low salt concentration (150 mM NaCl).

In the analytical SEC, *S. pyogenes* Cas1–$(His)_6$-MBP-Cas2 complex and *S. pyogenes* Csn2 migrated together with a smaller retention volume than the Cas1–$(His)_6$-MBP-Cas2 complex or the Csn2 tetramer alone, displaying that *S. pyogenes* II-A Cas1, Cas2, and Csn2 form a larger three-component protein complex (Fig. 22A). In the ITC experiment, the binding ratio of *S. pyogenes* Cas1–$(His)_6$-MBP-Cas2 heterohexamer complex to *S. pyogenes* Csn2 tetramer was calculated to be ~2:1, implying that two Cas1–Cas2 complexes interact with a single Csn2 tetramer (Fig. 22B). The binding constant (K_d) between *S. pyogenes* Cas1–$(His)_6$-MBP-Cas2 complex and *S. pyogenes* Csn2 was determined to be 0.52 μM whereas that between *S. pyogenes* Cas1 and Csn2 was measured as ~5.0 μM. This result shows that *S.*

pyogenes Csn2 bind more strongly to *S. pyogenes* II-A Cas1–Cas2 complex than *S. pyogenes* II-A Cas1 alone. This 10-fold difference in affinity for *S. pyogenes* Csn2 might be attributed from the direct interaction between $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 and *S. pyogenes* Csn2. However, the *S. pyogenes* II-A Cas2 did not appear to have intrinsic binding affinity to *S. pyogenes* Csn2. The two proteins did not interact with each other in SEC analysis (Fig. 22C).

Interaction between *S. pyogenes* Cas9 and Csn2

Cas9 from the type II-A CRISPR-Cas system is known to be responsible for both crRNA biogenesis and target nucleic acid degradation (Barrangou et al., 2007; Deltcheva et al., 2011; Jinek et al., 2012). Recently, several studies reported the evidence that Cas9 is involved in the adaptation of CRISPR immunity (Heler et al., 2015; Wei et al., 2015). *S. pyogenes* Cas9 designates protospacer candidates by recognizing the 5′-NGG-3′ PAM sequence during the adaptation stage, and associates with the other adaptation machinery (Cas1, Cas2, and Csn2) (Heler et al., 2015). To understand how *S. pyogenes* Cas9 is connected with the *S. pyogenes* II-A CRISPR-Cas adaptation module, the independent *in vitro* interaction tests were conducted using *S. pyogenes* Cas9 and the other II-A Cas proteins. In the analytical SEC, *S. pyogenes* Cas9 interacted with *S. pyogenes* Csn2 to form a complex. However, the Cas9 did not show the interaction with *S. pyogenes* II-A Cas1, $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2, or the Cas1–$(His)_6$-MBP-Cas2 complex (Fig. 23). The formation of the *S. pyogenes* Cas9–Csn2 complex was disturbed by increasing the salt concentration of the buffer, suggesting that electrostatic interactions play an essential role in the interaction between the two proteins (Fig. 24A). The reaction ratio between the two proteins was calculated from the SEC analyses with increasing amounts of *S. pyogenes* Cas9 against a fixed amount of *S. pyogenes* Csn2. Until the molar ratio of the *S. pyogenes* Cas9 to Csn2 reached approximately 2:1, the elution volume of the complex peak decreases and the peak

Figure 21. Complex formation of *S. pyogenes* II-A Cas1 and Cas2

(A) N-terminal $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 was expressed together with untagged *S. pyogenes* II-A Cas1 in *E. coli* BL21 (DE3) cells, and purified using dextrin sepharose-matrixed affinity chromatography (MBPTrap HP column, GE healthcare). Fractions for total soluble expressed proteins (T), flow-through (F) and 10 mM D-(+)-maltose-elution (E) were analyzed by SDS-PAGE. A control experiment expressing the untagged *S. pyogenes* II-A Cas1 alone are also shown on the left. Separately purified N-terminal $(His)_6$-MBP-tagged *S. pyogenes* II-A Cas2 and C-terminal $(His)_6$-tagged *S. pyogenes* II-A Cas1 were also analyzed together for comparison. (B) Elution volume comparison of the co-expressed and reconstituted *S. pyogenes* II-A Cas1–Cas2 complexes. The reconstituted complex sample was prepared by the stoichiometric mixing of individually purified Cas1 (40 μM) and $(His)_6$-MBP-Cas2 (20 μM).

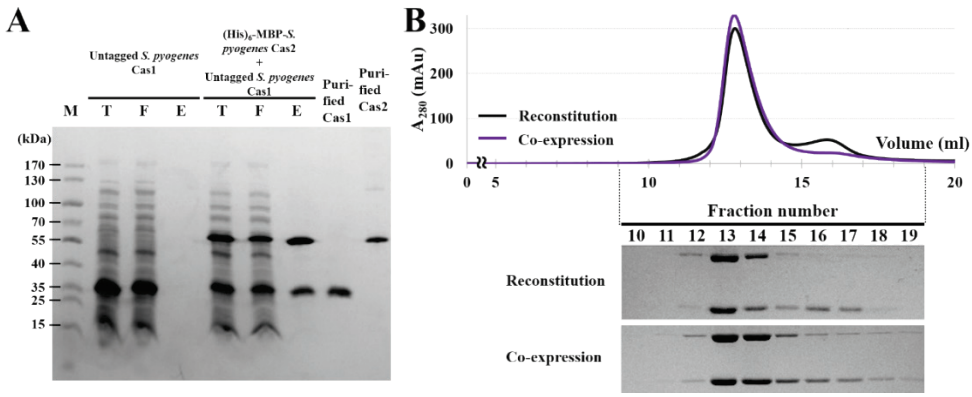
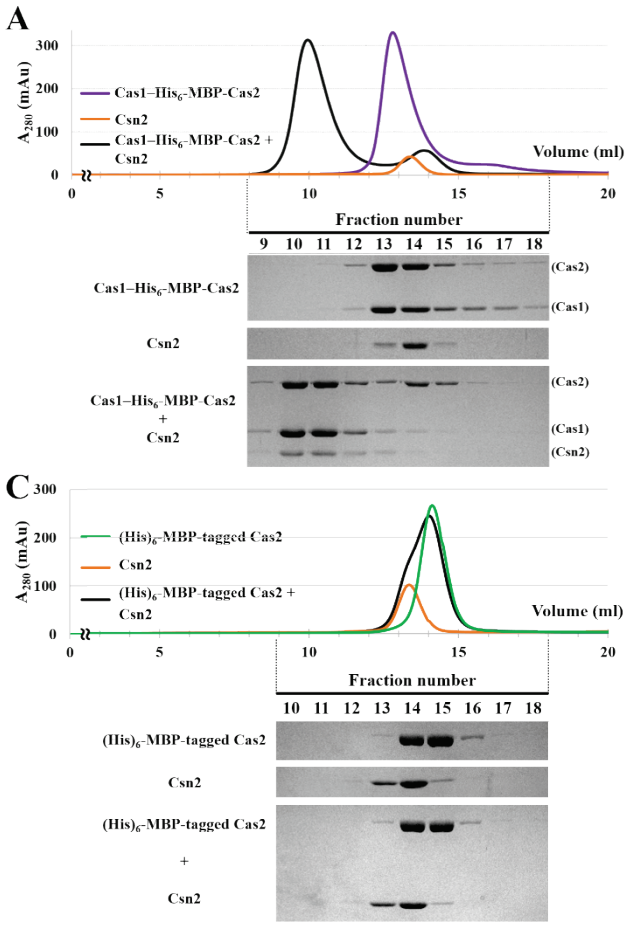

A
Untagged S. pyogenes Cas1
(His)6-MBP-S. pyogenes Cas2 + Untagged S. pyogenes Cas1
Purified Cas1
Purified Cas2
M T F E T F E
(kDa)
170
130
100
70
55
40
35
25
15
B
A280 (mAu)
300
200
100
0
Reconstitution
Co-expression
0 5 10 15 20
Volume (ml)
Fraction number
10 11 12 13 14 15 16 17 18 19
Reconstitution
Co-expression

Figure 22. Formation of type II-A CRISPR adaptation module consisting of *S. pyogenes* II-A Cas1–Cas2 complex and Csn2

(A) The analytical SEC assessed the interaction between the *S. pyogenes* II-A Cas1–$(His)_6$-MBP-Cas2 complex (10 μM) and Csn2 (40 μM). (B) ITC assay for the interaction between *S. pyogenes* II-A Cas1–$(His)_6$-MBP-Cas2 complex and Csn2. 50 μM of *S. pyogenes* Csn2 was titrated with 250 μM of Cas1–$(His)_6$-MBP-Cas2 complex. The experimentally determined N and K_d values are indicated. ITC derived thermodynamic parameters for the binding are calculated as mentioned in Fig 12B. (C) There was no interaction between the *S. pyogenes* Csn2 and II-A Cas2. 20 μM of $(His)_6$-MBP-tagged II-A Cas2 and 40 μM of Csn2 were incubated together or separately and analyzed by analytical SEC.



Cas1–$(His)_6$-MBP-Cas2 : Csn2
heterohexamer tetramer

Figure 23. Interaction analysis between *S. pyogenes* Cas9 and the other type II-A Cas proteins

(A) *S. pyogenes* Cas9 (10 μM) and Csn2 (40 μM) were incubated together and analyzed by analytical SEC. (B, C, and D) 10 μM of *S. pyogenes* Cas9 was incubated with (B) Cas1 (20 μM) or (C) $(His)_6$-MBP-Cas2 (20 μM) or (D) Cas1–$(His)_6$-MBP-Cas2 complex (10 μM), respectively. Subsequently, the mixture was analyzed by analytical SEC. Dotted chromatograms were applied to clearly show the overlapping.

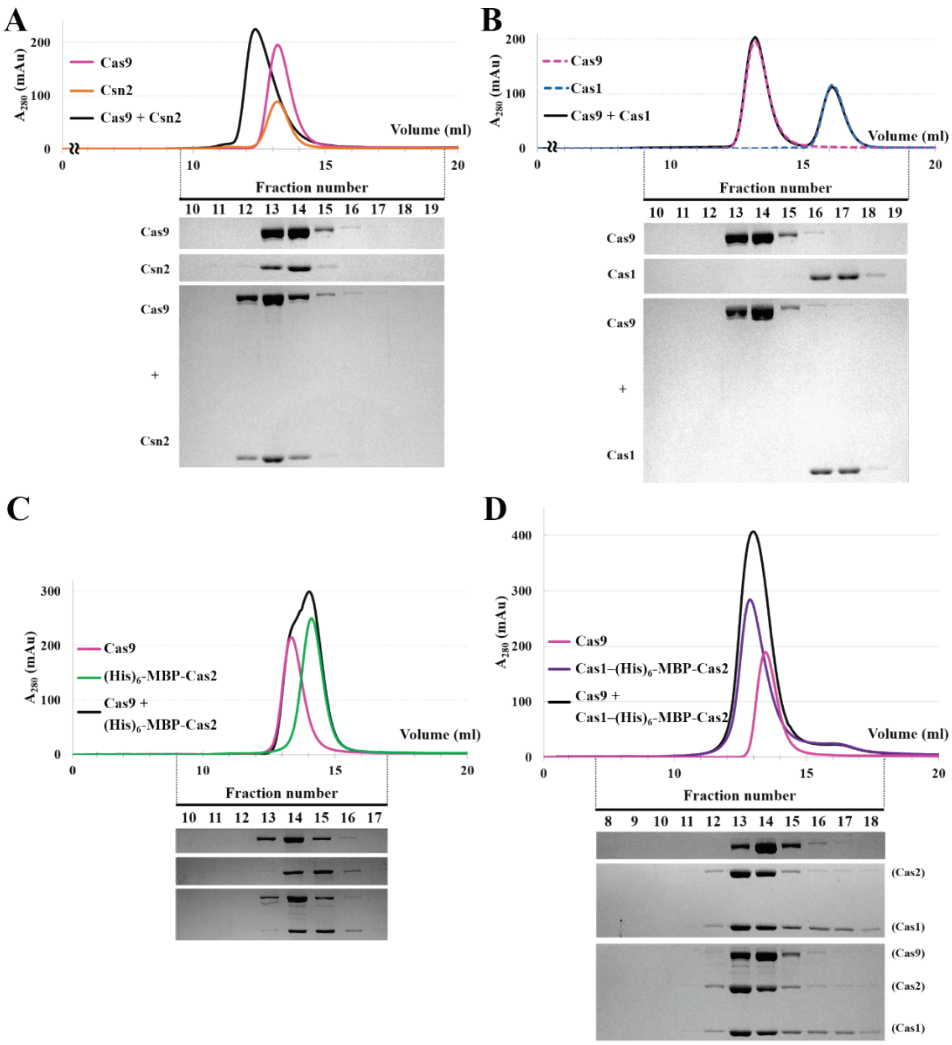
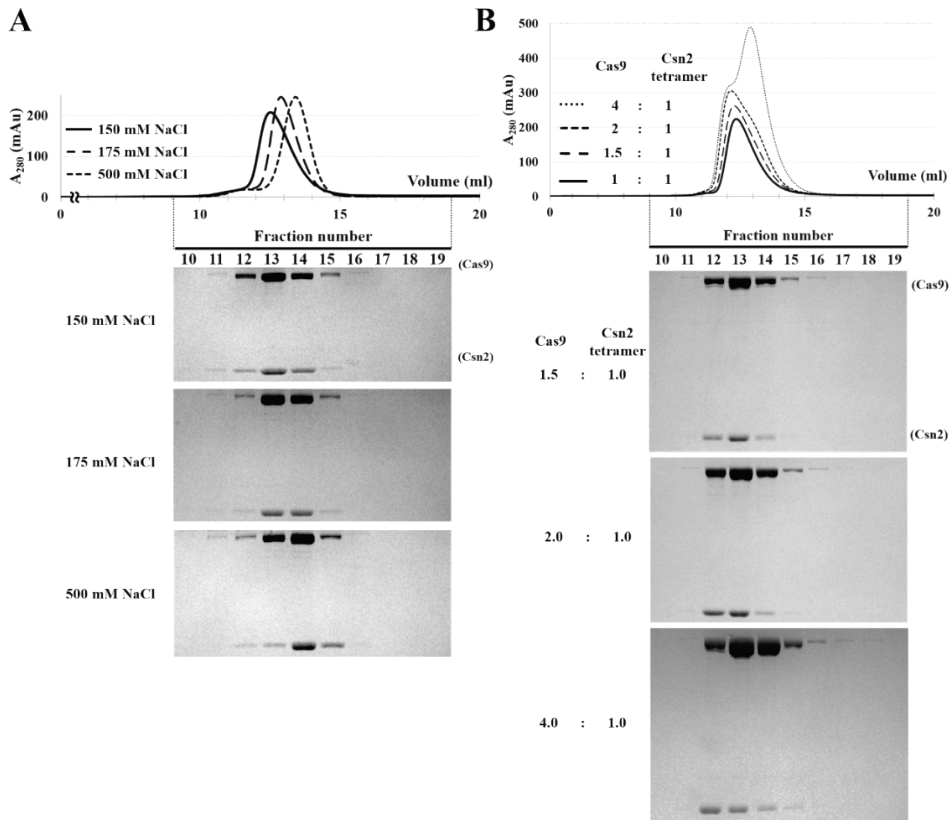

A
Cas9
Csn2
Cas9 + Csn2
Volume (ml)
Fraction number
Cas9
Csn2
Cas9
+
Csn2
B
Cas9
Cas1
Cas9 + Cas1
Volume (ml)
Fraction number
Cas9
Cas1
Cas9
+
Cas1
C
Cas9
(His)6-MBP-Cas2
Cas9 +
(His)6-MBP-Cas2
Volume (ml)
Fraction number
D
Cas9
Cas1–(His)6-MBP-Cas2
Cas9 +
Cas1–(His)6-MBP-Cas2
Volume (ml)
Fraction number
(Cas2)
(Cas1)
(Cas9)
(Cas2)
(Cas1)

Figure 24. Characterization of interaction between *S. pyogenes* Cas9 and Csn2

(A) Effect of salt concentration on the interaction between Cas9 and Csn2. Analytical size-exclusion chromatography was performed using increasing concentrations of NaCl. (B) Determination of the binding stoichiometry between *S. pyogenes* Cas9 and Csn2 by analytical SEC. Csn2 (40 μM) was pre-incubated with increasing amounts of Cas9 (10, 15, 20 and 40 μM), and the interaction was tested by analytical SEC. The chromatogram for the interaction between Cas9 (10 μM) and Csn2 (40 μM) was reproduced from Fig. 23A for comparison.


A
B
150 mM NaCl
175 mM NaCl
500 mM NaCl
Volume (ml)
Fraction number
(Cas9)
(Csn2)
Cas9
Csn2 tetramer
4 : 1
2 : 1
1.5 : 1
1 : 1
1.5 : 1.0
2.0 : 1.0
4.0 : 1.0

height increased (Fig. 24B). This result suggests that a maximum of two Cas9 proteins can interact with one Csn2 tetramer. The SPR binding assays were performed to determine the binding affinity between *S. pyogenes* Cas9 and Csn2. The ITC data could not be obtained due to the precipitation of the two proteins during titration. In the SPR analysis, the Cas9 was immobilized on the surface of sencor chip, and the Csn2 in the running buffer was injected repeatedly at different concentrations. The calculated dissociation constant value (K_d) of the *S. pyogenes* Cas9–Csn2 interaction was calculated to be ~1.4 μM (Fig. 25A). The *S. pyogenes* Csn2 can interact with the *S. pyogenes* Cas9 in the apo state, which lacks guide RNA. However, the previous *in vivo* adaptation studies used the active form of *S. pyogenes* Cas9 with guide RNA which consists of crRNA and tracrRNA (Heler et al., 2015). In order to fill this gap and clarify the interaction between *S. pyogenes* Csn2 and Cas9 more clearly, it was necessary to confirm the interaction of Csn2 with the Cas9 containing the RNA components. The sgRNA was synthesized by *in vitro* transcription methods (See the method section). Individually purified sgRNA and *S. pyogenes* Cas9 were combined to create a ribonucleoprotein complex. The sgRNA-loaded *S. pyogenes* Cas9 exhibited a higher binding affinity (K_d = 0.11 μM) for *S. pyogenes* Csn2 than the apo-Cas9 in the SPR analysis (Fig. 25B).

Since *S. pyogenes* Csn2 possesses a binding affinity for both *S. pyogenes* II-A Cas1–Cas2 complex and *S. pyogenes* Cas9, A further analytical SEC assay was conducted to confirm whether *S. pyogenes* Csn2 binds to the two molecules simultaneously forming a four-component Cas complex (Fig. 26). Separately purified Cas9, Csn2, and Cas1–$(His)_6$-MBP-Cas2 complex were incubated together and then injected into the analytical SEC column. The analytical SEC result showed that the four Cas components do not form a stable complex. The retention volume of the putative four-component Cas complex (Fractions 10 and 11) is similar to that of the three-component complex constituted of *S. pyogenes* II-A Cas1, $(His)_6$-MBP-Cas2, and Csn2. In addition, the unbound form of the *S. pyogenes* Cas9 was observed

(Fractions 13 and 14).

Figure 25. SPR analysis of the *S. pyogenes* Cas9–Csn2 interaction

(A) Sensorgrams show the binding of increasing concentrations (0.10, 0.20, 0.39, 0.78, 1.56, 3.13, and 6.25 μM) of free *S. pyogenes* Csn2 to immobilized *S. pyogenes* Cas9. The association rate constant k_a is calculated to be 4.04 x 10^2 $M^{-1}s^{-1}$, and the dissociation rate constant k_d is calculated to be 5.69 x 10^{-4} s^{-1}. (B) Sensorgrams show the binding of increasing concentrations (3.13, 6.25, 12.5, 25, 50, 100 and 200 nM) of free *S. pyogenes* Csn2 tetramer to immobilized *S. pyogenes* Cas9-sgRNA complex. The association rate constant k_a is calculated to be 1.25 x 10^4 $M^{-1}s^{-1}$, and the dissociation rate constant k_d is calculated to be 1.40 x 10^{-3} s^{-1}. The calculated K_d values for the two sets of sensorgrams are also shown.

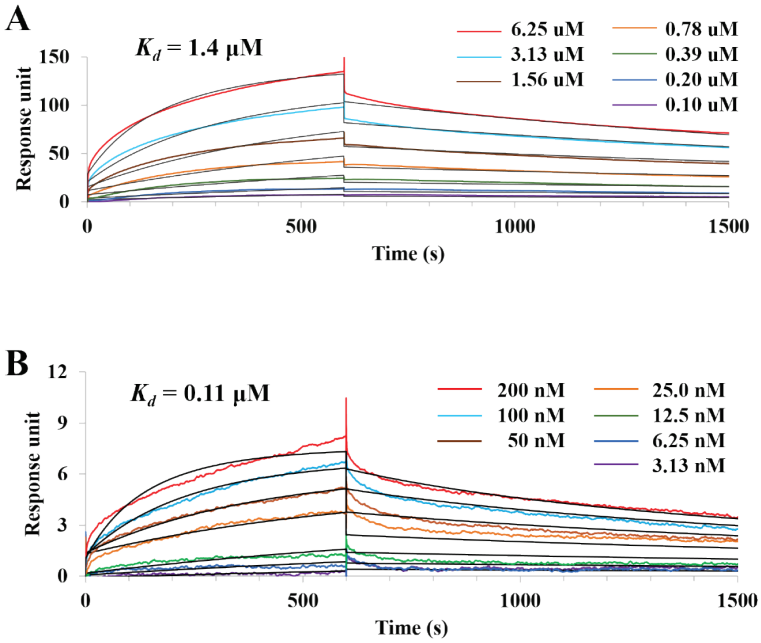
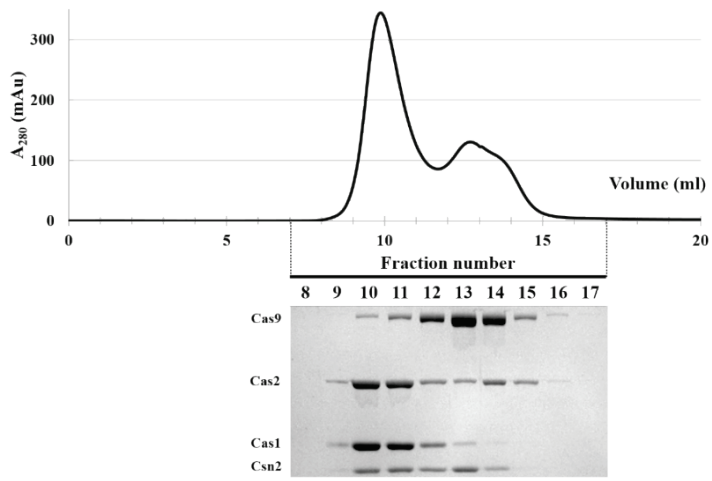

A
K_d = 1.4 µM
6.25 uM
3.13 uM
1.56 uM
0.78 uM
0.39 uM
0.20 uM
0.10 uM
Response unit
150
100
50
0
0
500
1000
1500
Time (s)
B
K_d = 0.11 µM
200 nM
100 nM
50 nM
25.0 nM
12.5 nM
6.25 nM
3.13 nM
Response unit
12
9
6
3
0
0
500
1000
1500
Time (s)

Figure 26. Test of four-component II-A Cas complex formation in analytical SEC

Separately purified four *S. pyogenes* type II Cas proteins (10 μM of Cas9, 10 μM of Cas1–$(His)_6$-MBP-Cas2 complex and 40 μM of Csn2) were incubated together and injected into analytical size-exclusion chromatography column.


300
200
100
0
A_{280} (mAu)
Volume (ml)
0
5
10
15
20
Fraction number
8 9 10 11 12 13 14 15 16 17
Cas9
Cas2
Cas1
Csn2

Discussion

Structural feature of *S. pyogenes* II-A Cas1 and Cas2

Cas1 is a universally conserved Cas protein which is found in most CRISPR-Cas systems (Makarova et al., 2011b). Although the structural and functional features of the several Cas1 proteins from type I CRISPR-Cas systems have been described (Wiedenheft et al., 2009; Babu et al., 2011; Kim et al., 2013), homologs from the other CRISPR-Cas types have not been characterized until recently. I determined the crystal structures of the *S. pyogenes* II-A Cas1 and Cas2 proteins and expand on the structural knowledge of II-A Cas proteins. Despite the similarity of the domain folds, the crystal structure of *S. pyogenes* II-A Cas1 revealed a unique structural feature distinct from the type I Cas1 homologs. In the type I Cas1 structures, three conserved residues (Glu149, His205, and Glu220) in the C-terminal domain are responsible for DNA endonuclease activity and are in proximity to coordinate catalytic metal ions. However, in the *S. pyogenes* II-A Cas1 structure, extensive dimerization interaction between the two C-terminal domain causes the structural rearrangement of the three conserved residues and the burial of the potential metal-binding sites. Consistent with the structural features, the nuclease activity assays showed that *S. pyogenes* II-A Cas1 did not cleave any of the tested nucleic acid substrates.

Despite the negative results, I still cannot exclude the possibility that *S. pyogenes* II-A Cas1 may require more specific experimental conditions and/or nucleic acid substrates for its function as a nuclease. It is conceivable that the *S. pyogenes* II-A Cas1 structure represents an inactive conformation, and the binding of other factors such as Cas2 or Csn2 which form a complex with the Cas1 may induce conformational switching to a catalytically active state. Recently, several II-A Cas1–Cas2 complex structures have been determined. It is revealed that the Cas1

dimer in the type II-A Cas1–Cas2 complex consists of two protomers having a different conformational state of the C-terminal region (Xiao et al., 2017) (Fig. 27A). In particular, one of the two protomers has a conformational state similar to that seen in type I Cas1 homologs, implying that a protomer of the *S. pyogenes* II-A Cas1 dimer can be switched to the active state (Fig. 27B). To confirm this possibility, another set of nuclease activity tests was performed in the presence of *S. pyogenes* II-A Cas2, but no cleavage was detected (Fig. 10). Even in this case, the $(His)_6$-MBP tag, which was essential for solubilizing the Cas2 under the experimental condition which I used, might interfere with proper conformational change and/or binding of DNA substrate. In the type I-C Cas2 proteins, the N-terminal SUMO-tag can interfere with the hinge motion of the Cas2 dimer and inhibits dsDNase activity (Nam et al., 2012a). There is a possibility that the non-free state of *S. pyogenes* II-A Cas2 due to the $(His)_6$-MBP-tag may interfere with the active state of Cas1–Cas2 complex.

S. pyogenes II-A Cas2 revealed a dimer arrangement of two monomers parallel to each other. This arrangement differs from *E. faecalis* type II-A Cas2, in which two protomers are arranged at an angle each other to form a V-shape (Xiao et al., 2017) (Fig. 11C). Since the experimental setting is different from each other, it is difficult to determine which conformation is biologically relevant. Rather, these II-A Cas2 structures are likely to exhibit different conformational states indicative of the possibility of the hinge motion (Fig. 28A). While the *S. pyogenes* II-A Cas2 was crystallized singly, *E. faecalis* II-A Cas2 was crystallized as a constituent of the prespacer-bound II-A Cas1–Cas2 complex. The prespacer DNA spans the V-shaped concave of the *E. faecalis* Cas2. (Xiao et al., 2017). Recently, a *Streptococcus thermophilus* II-A Cas2 dimer in the II-A Cas1–Cas2–Csn2 complex structure has been reported to have another conformational state supporting the hypothesis (Wilkinson et al., 2019). Such a conformational switching in the II-A Cas2 dimer appears to be possible through the putative hinge region around α3 helix (Fig. 11A).

Figure 27. Asymmetric conformation of II-A Cas1 dimers in Cas1–Cas2 complex

(A) Side by side comparison of the three II-A Cas1 dimer structures. (B) Structural alignment of *S. pyogenes, S. themophilus,* and *E. faecalis* II-A Cas1 dimers based on their N-terminal domains.

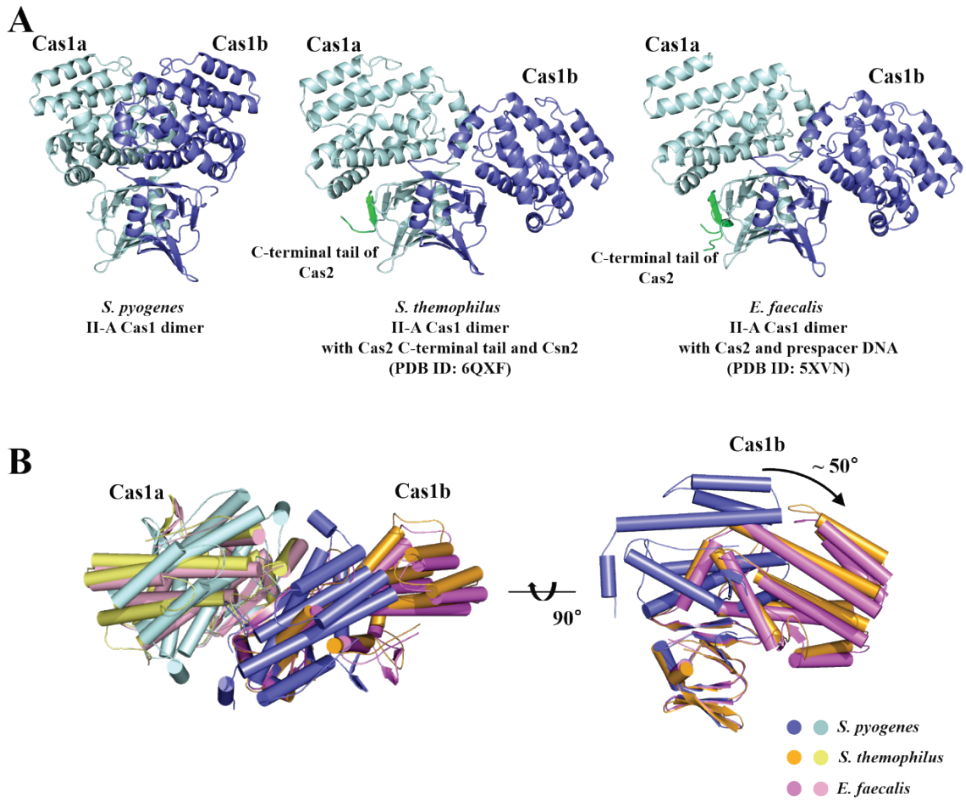
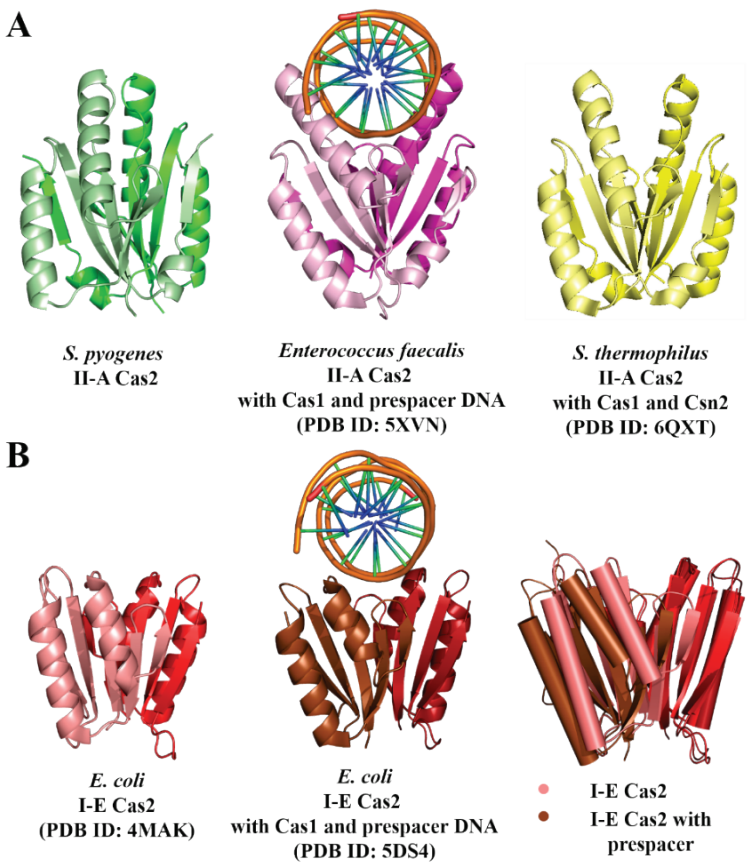

A
Cas1a
Cas1b
Cas1a
Cas1b
C-terminal tail of Cas2
Cas1a
Cas1b
C-terminal tail of Cas2
S. pyogenes
II-A Cas1 dimer
S. themophilus
II-A Cas1 dimer
with Cas2 C-terminal tail and Csn2
(PDB ID: 6QXF)
E. faecalis
II-A Cas1 dimer
with Cas2 and prespacer DNA
(PDB ID: 5XVN)
B
Cas1a
Cas1b
90°
Cas1b
~ 50°
S. pyogenes
S. themophilus
E. faecalis

Figure 28. Structural comparison between various II-A Cas2 homologs

(A) The three different conformational states of the II-A Cas2 homologs. The three II-A Cas2 homologs exhibit high sequence similarity to each other (over 80% sequence identity). (B) I-E Cas2 structures without hinge part. The conformational changes are rarely observed between the sole dimer state (left) and the state with prespacer (Cas1 does not shown for clarity) (center). When two structures overlap based on one protomer, the other protomers appear to move in parallel.


A
S. pyogenes
II-A Cas2
Enterococcus faecalis
II-A Cas2
with Cas1 and prespacer DNA
(PDB ID: 5XVN)
S. thermophilus
II-A Cas2
with Cas1 and Csn2
(PDB ID: 6QXT)
B
E. coli
I-E Cas2
(PDB ID: 4MAK)
E. coli
I-E Cas2
with Cas1 and prespacer DNA
(PDB ID: 5DS4)
I-E Cas2
I-E Cas2 with prespacer

The possibility of bending motion was previously reported in I-C Cas2 proteins which share the exchange of the C-terminal segments for dimerization (Nam et al., 2012a; Ka et al., 2014; Ka et al., 2017). In the case of I-E Cas2 without the exchange of the secondary structure elements, there is no bending motion between the sole and prespacer DNA-bound structures. Rather, it is confirmed that the two protomers move relatively parallel (Nunez et al., 2015) (Fig. 28B).

The C-terminal tail of the *S. pyogenes* II-A Cas2 is critical for its interaction with the *S. pyogenes* II-A Cas1. Interestingly, the binding stoichiometry (*N*) of the C-terminal tail of the Cas2 to the Cas1 was determined to be ~0.5 in the ITC analysis, indicating that only one C-terminal tail peptide of the Cas2 could bind to a single Cas1 dimer (Fig. 13C). Considering the relatively small size of the C-terminal tail (22 amino acids), it is difficult to imagine that the other binding site on the opposite side of the Cas1 dimer will be prevented by the steric hindrance caused by the first bound C-terminal tail of the Cas2. Instead, these observation suggests that the binding of the Cas2 to the Cas1 protomer causes a conformational change of the second binding interface in the unbound Cas1 protomer. It could decrease the affinity of the second binding site for the Cas2, similar to negative cooperativity. A significant conformational difference was observed between the Cas2-bound and unbound Cas1 protomers in the *E. faecalis* and *S. themophilus* II-A Cas1–Cas2 structures (Xiao et al., 2017; Wilkinson et al., 2019). In the binding interface between the *E. faecalis* II-A Cas1 and Cas2, the C-terminal tail of the Cas2 interacts with both N- and C- terminal domain of the Cas1, forming the hydrophobic core and causing conformational rearrangement of the Cas1 monomer (Fig. 29A). Moreover, it seems that only one of the two protomers can have the conformational state. This is because if the two protomers of the dimer maintain the same state, the steric hindrance is expected to occur between the two monomers (Fig. 29B).

The formation of type II-A CRISPR-Cas adaptation module

II-A CRISPR-Cas system requires all four Cas proteins, including Cas9 nuclease for adaptation. Also, when expressed in *E.coli*, the four *S. pyogenes* II-A Cas proteins were co-purified, implying that they would interact with each other. I confirmed that the presence of type II-A CRISPR adaptation module and their topology by confirming the interactions between the Cas proteins in detail. *S. pyogenes* Csn2 associates with the *S. pyogenes* type II-A Cas1–Cas2 complex via the interaction between the Csn2 and the Cas1. The binding ratio of *S. pyogenes* Csn2 tetramer to *S. pyogenes* II-A Cas1 dimer (or II-A Cas1–Cas2 heterohexamer) was determined to be 1:2 suggesting that the *S. pyogenes* Csn2 tetramer possesses two binding sites for the interaction with *S. pyogenes* II-A Cas1. The electrostatic interaction plays a crucial role in the complex formation. From the *in silico* docking simulation, I have specified several charged residues which can participate in electrostatic interactions in the putative binding interface.

Recently, three cryo-EM structures of *S. thermophilus* II-A Cas1–Cas2–Csn2 complexes containing dsDNA with different assemblies have been reported (Wilkinson et al., 2019) (Fig. 30A and B). One form is a monomer complex (or monomer unit) which consists of two Csn2 tetramers and two Cas1–Cas2 heterohexamer complexes with ~ 30 bp dsDNA. Another form is comprises of two independent monomer units connected with additional two Cas1–Cas2 heterohexamer complexes having a different conformational arrangement. The other is a filament form comprising continuous monomer units with long DNA duplex. The *S. thermophilus* II-A Cas1–Cas2–Csn2 monomer unit shows a Cas1–Csn2 binding interface different from the *S. pyogenes* II-A Cas1–Csn2 docking model (Fig. 30C). *S. thermophilus* II-A Cas1 retains the binding surface, whereas *S. thermophilus* Csn2 has a binding surface different from those seen in the docking model. *S. thermophilus* Csn2 involve D209 and D211 residues in the electrostatic interaction with the Cas1. Moreover, this two residues are strictly conserved in other

Figure 29. Structural rearrangement of the C-terminal domain of II-A Cas1 triggered by the interaction with II-A Cas2

(A) In both *S. themophilus* and *E. faecalis* Cas1-Cas2 complexes, the α9 helix in the C-terminal domain of the II-A Cas1 participates in the formation of a hydrophobic core. Also, this is accompanied by a conformational change of the C-terminal domain. (B) Assuming that both C-terminal domains in the II-A Cas dimer, interact with Cas2, each C-terminal domain could crash with each other creating steric hindrance.

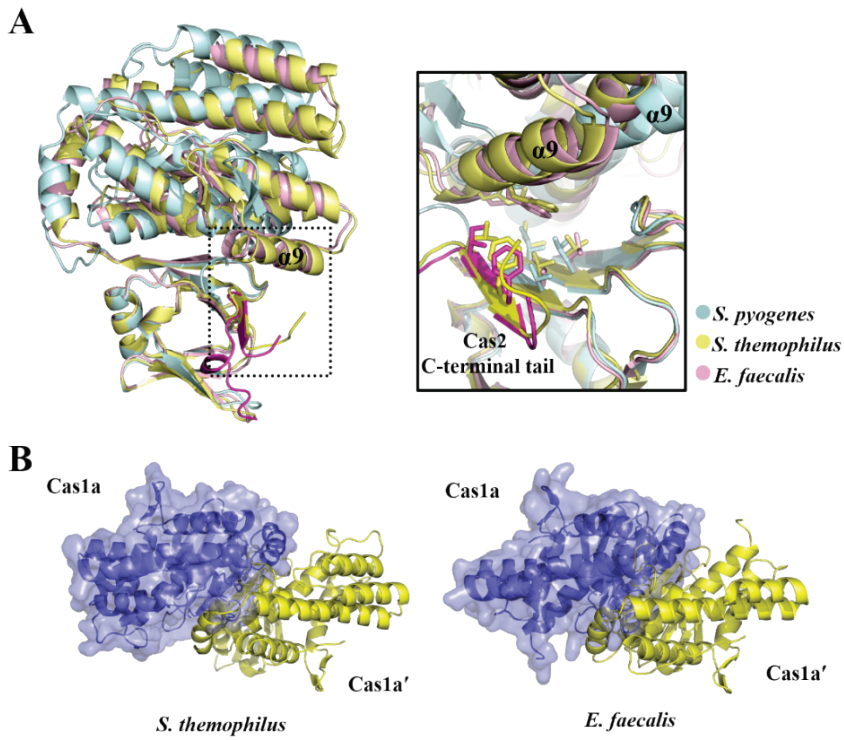
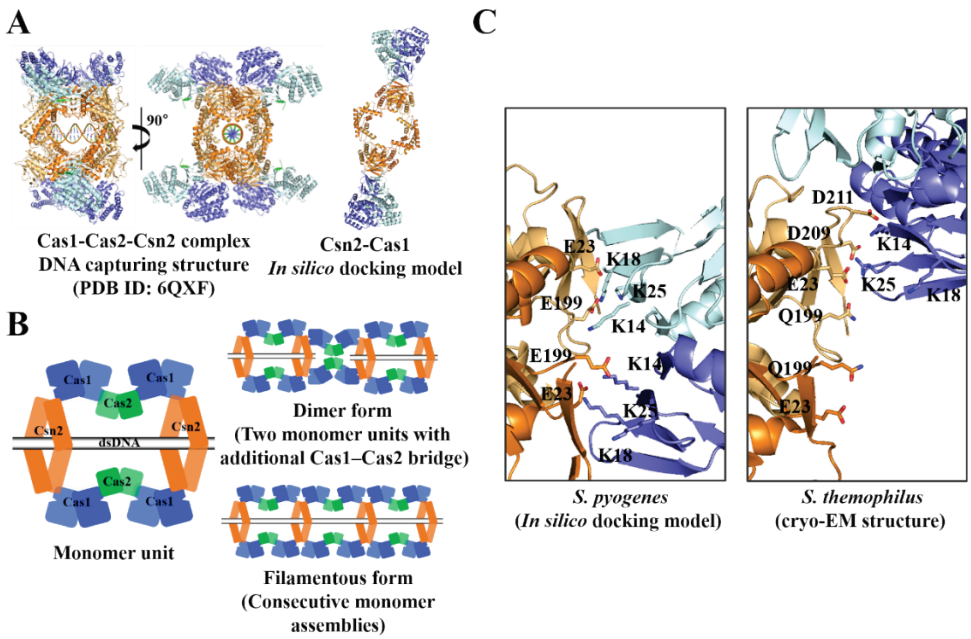

A
α9
α9
α9
Cas2
C-terminal tail
S. pyogenes
S. themophilus
E. faecalis
B
Cas1a
Cas1a′
S. themophilus
Cas1a
Cas1a′
E. faecalis

Figure 30. Differences of the Cas1–Csn2 interaction surface between the *in silico* docking model and *S. thermophilus* Cas1–Cas2–Csn2 complex monomer unit

(A) The cryo-EM structure of the *S. thermophilus* Cas1–Cas2–Csn2 complex (left) and the *in silico* docking model of the *S. pyogenes* II-A Cas1–Csn2 complex (right). The Cas1 dimer is colored as in blue and cyan for each protomer and Csn2 is shown in orange and bright orange for chain A and B in the asymmetric unit. The C-terminal tail of Cas2 is shown in green (only in Cas1-Cas2-Csn2 complex). (B) Schematic representation of the three different assemblies of the *S. thermophilus* Cas1–Cas2–Csn2 complexes with dsDNA. (C) The binding interface between the Cas1 and Csn2 of the *in silico* docking model and cryo-EM structure (monomer unit).


A
90°
Cas1-Cas2-Csn2 complex
DNA capturing structure
(PDB ID: 6QXF)
Csn2-Cas1
In silico docking model
B
Cas1
Cas2
Csn2
dsDNA
Monomer unit
Dimer form
(Two monomer units with
additional Cas1–Cas2 bridge)
Filamentous form
(Consecutive monomer
assemblies)
C
E23
K18
E199
K25
K14
E199
K14
E23
K25
K18
S. pyogenes
(*In silico* docking model)
D211
D209
K14
E23
K25
K18
Q199
Q199
E23
S. themophilus
(cryo-EM structure)

Csn2 homologs, including *S. pyogenes* Csn2 (Fig. 20). The E23 residue is located at the periphery of the binding interface.

Considering the high sequence identities between the Cas proteins (60% for Csn2 and Cas9, and 80% Cas1 and Cas2), the limited experimental settings for the docking simulation might have led to the difference of the Cas1–Csn2 binding mode between the *in silico* docking model and cryo-EM structures. In the docking simulation, only one Csn2 tetramer and one Cas1 dimer were used to find the putative binding mode. In contrast, the *S. thermophilus* II-A Cas1–Cas2–Csn2 monomer unit exhibits a more systematic binding mode in which two Csn2 tetramers and two Cas1–Cas2 heterohexamers associate with each other through four identical binding interfaces. Despite the difference in the binding mode, the interaction surface of the *S. pyogenes* type II-A Cas1 is still valid. The mutational analysis of *S. pyogenes* type II-A Cas1 is also consistent with cryo-EM structure of the *S. thermophilus* II-A Cas1–Cas2–Csn2 monomer unit. The K14 and R25 residues of the *S. thermophilus* II-A Cas1 participate in electrostatic interactions with E209 and E211 residues from the *S. thermophilus* Csn2 while K18 residue is far from the interaction surface. Although not as much as the K18 residue of the Cas1, the E23 residue of the Csn2 is also off the center of the ionic interaction. This may explain why the E23A mutation of *S. pyogenes* Csn2 does not show the drastic inhibition of the interaction with the *S. pyogenes* II-A Cas1 (Fig. 17D).

The binding stoichiometry between the *S. pyogenes* II-A Cas1 dimer and the *S. pyogenes* Csn2 tetramer was calculated to be 2:1 implying that maximum two Cas1 dimers can interact with one Csn2 tetramer. From the standpoint of the Csn2, simple stoichiometry is consistent with the *S. thermophilus* II-A Cas1–Cas2–Csn2 monomer unit (Fig. 30B). In this case, the monomer unit can be interpreted as two Cas1–Csn2 complexes ($Cas1_4$–$Csn2_4$) connected with two Cas2 dimers. However, when the monomer unit is interpreted as two Cas1–Cas2 heterohexamer complexes ($Cas1_4$–$Cas2_4$) connected by two Csn2 tetramers, there is inconsistency with the ITC assay

result using *S. pyogenes* II-A Cas1–Cas2 complex and Csn2. It is possible that the binding ratio calculated from the ITC assay reflects the filamentous form of the II-A Cas1–Cas2–Csn2 complex (Fig. 30B). In fact, The *S. thermophilus* Csn2 tetramer has a maximum four independent binding interfaces for the Cas1 dimer because the protomer of the Csn2 can interact with one Cas1 dimer. The filamentous form revealed that the Cas1 dimer occupies all binding sites of the Csn2 resulting in interaction with four different Cas1–Cas2 heterohexamer complexes. Since the filamentous form is a continuous assembly of the monomer units, the binding ratio of the Cas1–Cas2 complex to the Csn2 is maintained at 2: 1. However, as noted by the authors, a small number of filamentous forms were observed. Therefore, it is unlikely that the filamentous form is a predominant species over the monomer unit in the ITC analysis. Rather, the binding ratio of *S. pyogenes* II-A Cas1–Cas2 with *S. pyogenes* Csn2 remained the simple stoichiometry because of different experimental setting. In the ITC experiment, I used a *S. pyogenes* II-A Cas1–Cas2 complex having a relatively large size tag that could interfere with stable binding. Also, I did not consider dsDNA fragment, another component of the *S. thermophilus* II-A Cas1–Cas2–Csn2 complex structures.

The role of Csn2 in type II-A CRISPR adaptation mechanism

The involvement of Csn2 in the adaptation process in type II-A CRISPR-Cas system has been demonstrated (Barrangou et al., 2007; Heler et al., 2015). Also, the free dsDNA end-binding and sliding property of Csn2 imply that the involvement of Csn2 with prespacers (Nam et al., 2011; Koo et al., 2012; Arslan et al., 2013). However, its precise role therein remains unclear. *In vivo* deletion of the *csn2* gene inhibits insertion of a new spacer into the CRISPR array upon bacteriophage challenge (Heler et al., 2015; Wei et al., 2015). The II-A Cas1–Cas2 complex catalyzes full-site spacer integration when provided with proper substrates, indicating the dispensability of Csn2 in prespacer integration *in vitro* (Wright and

Doudna, 2016; Xiao et al., 2017). Based on these two observations, it has been suggested that Csn2 may be involved in different step(s) of the type II-A CRISPR adaptation process, such as the generation of prespacers (Wright and Doudna, 2016). In the present study, I found that *S. pyogenes* Csn2 interacts with *S. pyogenes* Cas9 directly and also forms an adaptation module with the *S. pyogenes* II-A Cas1–Cas2 complex, respectively. These results suggest that the Csn2 may serve as a partner for the Cas9 which participates in the adaptation process and/or also can mediate the Cas9 and the II-A Cas1–Cas2 complex. In the former case, Csn2 can be loaded onto the free end of the viral DNA injected into the host cell and then recruit the Cas9 which is responsible for PAM recognition. In type II-A CRISPR-Cas systems, Cas9 plays pivotal roles in crRNA biogenesis and interference stages and also possesses PAM recognition property. (Barrangou et al., 2007; Deltcheva et al., 2011; Jinek et al., 2012). Cas9 is required in the adaptation of CRISPR immunity specifying functional PAM sequences (Heler et al., 2015; Wei et al., 2015; Heler et al., 2017). So, Csn2 can designate the foreign DNA as an object of the survey for protospacer candidate. It was reported that new spacers are obtained immediately after infection from the free end of the viral DNA, which is first injected into the host cell. Also, the targeting of an early-injected region of phage DNA provides better immunity than that of late-injected sequences (Modell et al., 2017).

The latter case highlights the role of Csn2 in serving as a scaffold protein and coordinating the functions of Cas9 and Cas1–Cas2 complex within *S. pyogenes* type II-A CRISPR adaptation. Interestingly, the analytical SEC result suggests that the four *S. pyogenes* II-A Cas components cannot from a stable four-component complex simultaneously. If the binding interfaces for *S. pyogenes* II-A Cas1–$(His)_6$-MBP-Cas2 complex and *S. pyogenes* Cas9 partially overlap each other, the binding of the Cas1–$(His)_6$-MBP-Cas2 complex to the Csn2 may hinder the interaction between the Cas9 and Csn2. The dissociation constant (K_d) between *S. pyogenes* Csn2 and *S. pyogenes* Cas1–$(His)_6$-MBP-Cas2 complex is smaller than that between

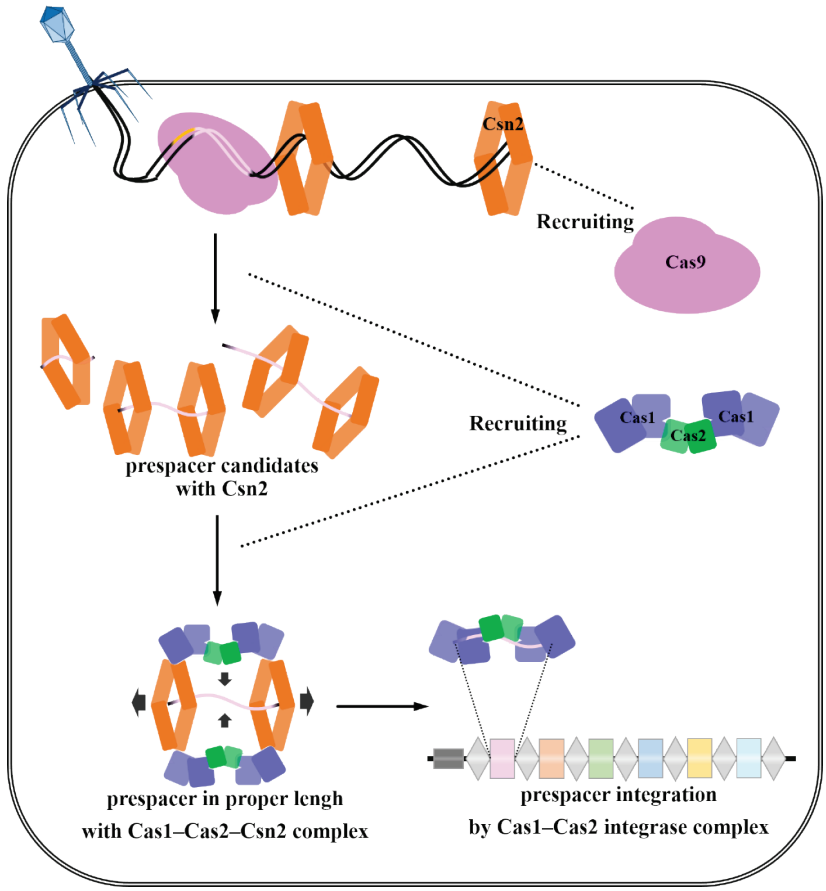
S. pyogenes Csn2 and *S. pyogenes* Cas9 (Figs. 26 and 33) even when the Cas9 forms a complex with sgRNA. It is also conceivable that the $(His)_6$-MBP tag, which was required for solubilizing the *S. pyogenes* Cas2 can interfere with the Cas9–Csn2 interaction. Csn2 may keep the dsDNA substrates after the recognition of foreign DNA by Cas9 (Fig. 31). In this case, Csn2 retains a functional spacer candidate (prespacer) and acts as a mid-bridge to recruit II-A Cas1–Cas2 integrase complex. This possibility is consistent with the result in which *S. pyogenes* II-A Cas1–Cas2 complex does not interact with *S. pyogenes* Cas9 but forms a complex with *S. pyogenes* Csn2. It is noteworthy that the II-A Cas1–Cas2–Csn2 with dsDNA complex and the guide RNA-loaded Cas9 are in stable but there is no complexes consisting of the two in the cryo-EM analysis of the co-expressed *S. thermophilus* II-A CRISPR-Cas proteins (Wilkinson et al., 2019).

Another Cas protein, Cas4, is known as a participant in the adaptation stage in addition to Cas1 and Cas2 and is conserved in several types of CRISPR-Cas system. Recently, it has been reported that Cas4 forms a complex with Cas1 and shows Cas1–Cas2-dependent nuclease activity trimming the single strand overhangs of prespacer DNA to correct its size and also defining PAM-compatible spacer in at least three type I CRISPR-Cas systems (I-A, I-C and, I-D) (Kieper et al., 2018; Lee et al., 2018; Rollie et al., 2018; Shiimori et al., 2018; Almendros et al., 2019). This property suggests that the Cas4 protein ensures spacer fidelity, and so blocks the integration of premature prespacers. However, Csn2 and Cas4 do not appear to be functionally homologous to each other because the enzymatic activity of Csn2 has not been reported. Instead, the free dsDNA end-loading property of Csn2 indicates that Csn2 could be associate with prespacer DNA in a manner similar to Ku protein. The heterodimeric Ku protein forms a ring-like structure which can bind at the DNA ends similar to the Csn2. (Dynan and Yoo, 1998; Walker et al., 2001; Arslan et al., 2013). The Ku-heterodimer binds to the double-strand breakpoint and recruits other factors involved in non-homologous end joining (NHEJ) (Feldmann et al., 2000). It is

noteworthy that Csn2 can inhibit NHEJ repair by blocking the access of Ku protein to DNA when the host cell can express both NHEJ system and Csn2 (Bernheim et al., 2017). The previous studies of Cas4 and Ku proteins suggest that Csn2 preserves the prespacer DNA and recruits the Cas1–Cas2 integrase complex (Fig. 31). Supporting this assumption, the dsDNA-bound Cas1–Cas2–Csn2 complex structure from *S. thermophilus* reveals that Csn2 prevents the Cas1–Cas2 integrase complex from accessing the prespacer DNA directly but also primes the complex to capturing the prespacer DNA (Fig. 30A and B).

Figure 31. The possible roles of Csn2 in type II-A CRISPR-Cas adaptation

The interaction between Cas9 and Csn2 might provide two profits in adaption. Immediate binding of Csn2 to the viral DNA end can help Cas9 in searching proper prespacers. Alternatively, Csn2 may be waiting in advance to quickly deliver the prespacer selected by Cas9 to the next step. Because Csn2 blocks the direct interaction between the prespacer and Cas1–Cas2 integrases shown in the *S. thermophiles* Cas1–Cas2–Csn2 DNA capturing complex, the dissociation of the capturing complex should occur in an unknown mechanism. One possibility is that the process might depend on the length of the prespacer.


Csn2
Recruiting
Cas9
Recruiting
Cas1
Cas2
Cas1
prespacer candidates
with Csn2
prespacer in proper lengh
with Cas1–Cas2–Csn2 complex
prespacer integration
by Cas1–Cas2 integrase complex

References

Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr* 66(Pt 2), 213-221.

Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, Mccoy AJ, Moriarty NW et al. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr D Biol Crystallogr* 58(Pt 11), 1948-1954.

Almendros C, Nobrega FL, Mckenzie RE and Brouns SJJ (2019) Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation, *Nucleic Acids Res* 47(10), 5223-5230.

Anders C, Niewoehner O, Duerst A and Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease, *Nature* 513(7519), 569-573.

Arslan Z, Wurm R, Brener O, Ellinger P, Nagel-Steger L, Oesterhelt F et al. (2013) Double-strand DNA end-binding and sliding of the toroidal CRISPR-associated protein Csn2, *Nucleic Acids Res* 41(12), 6347-6359.

Babu M, Beloglazova N, Flick R, Graham C, Skarina T, Nocek B et al. (2011) A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair, *Mol Microbiol* 79(2), 484-502.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes, *Science* 315(5819), 1709-1712.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H et al. (2000) The Protein Data Bank, *Nucleic Acids Res* 28(1), 235-242.

Bernheim A, Calvo-Villamanan A, Basier C, Cui L, Rocha EPC, Touchon M et al. (2017) Inhibition of NHEJ repair by type II-A CRISPR-Cas systems in bacteria, *Nat Commun* 8(1), 2094.

Bolotin A, Quinquis B, Sorokin A and Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, *Microbiology* 151(Pt 8), 2551-2561.

Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes, *Science* 321(5891), 960-964.

Busso D, Delagoutte-Busso B and Moras D (2005) Construction of a set Gateway-based destination vectors for high-throughput cloning and expression screening in *Escherichia coli*, *Anal Biochem* 343(2), 313-321.

Carte J, Wang R, Li H, Terns RM and Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes, *Genes Dev* 22(24), 3489-3496.

Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr* 66(Pt 1), 12-21.

Datsenko KA, Pougach K, Tikhonov A, Wanner BL, Severinov K and Semenova E (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial

immunity system, *Nat Commun* 3, 945.

Davanloo P, Rosenberg AH, Dunn JJ and Studier FW (1984) Cloning and expression of the gene for bacteriophage T7 RNA polymerase, *Proc Natl Acad Sci U S A* 81(7), 2035-2039.

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III, *Nature* 471(7340), 602-607.

Dynan WS and Yoo S (1998) Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids, *Nucleic Acids Res* 26(7), 1551-1559.

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

Fagerlund RD, Wilkinson ME, Klykov O, Barendregt A, Pearce FG, Kieper SN et al. (2017) Spacer capture and integration by a type I-F Cas1-Cas2-3 CRISPR adaptation complex, *Proc Natl Acad Sci U S A* 114(26), E5122-E5128.

Feldmann E, Schmiemann V, Goedecke W, Reichenberger S and Pfeiffer P (2000) DNA double-strand break repair in cell-free extracts from Ku80-deficient cells: implications for Ku serving as an alignment factor in non-homologous DNA end joining, *Nucleic Acids Res* 28(13), 2585-2596.

Garside EL, Schellenberg MJ, Gesner EM, Bonanno JB, Sauder JM, Burley SK et al. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases, *RNA* 18(11), 2020-2028.

Grissa I, Vergnaud G and Pourcel C (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats, *BMC Bioinformatics* 8, 172.

Guillerez J, Lopez PJ, Proux F, Launay H and Dreyfus M (2005) A mutation in T7 RNA polymerase that facilitates promoter clearance, *Proc Natl Acad Sci U S A* 102(17), 5958-5963.

Haft DH, Selengut J, Mongodin EF and Nelson KE (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes, *PLoS Comput Biol* 1(6), e60.

Hale C, Kleppe K, Terns RM and Terns MP (2008) Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*, *RNA* 14(12), 2572-2579.

Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L et al. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex, *Cell* 139(5), 945-956.

Heler R, Samai P, Modell JW, Weiner C, Goldberg GW, Bikard D et al. (2015) Cas9 specifies functional viral targets during CRISPR-Cas adaptation, *Nature* 519(7542), 199-202.

Heler R, Wright AV, Vucelja M, Bikard D, Doudna JA and Marraffini LA (2017) Mutations in Cas9 Enhance the Rate of Acquisition of Viral Spacer Sequences during the CRISPR-Cas Immune Response, *Mol Cell* 65(1), 168-175.

Jansen R, Embden JD, Gaastra W and Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes, *Mol Microbiol* 43(6), 1565-1575.

Jiang F, Zhou K, Ma L, Gressel S and Doudna JA (2015) STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition, *Science* 348(6242), 1477-1481.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA and Charpentier E (2012) A

programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, *Science* 337(6096), 816-821.

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E et al. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation, *Science* 343(6176), 1247997.

Jore MM, Lundgren M, Van Duijn E, Bultema JB, Westra ER, Waghmare SP et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by Cascade, *Nat Struct Mol Biol* 18(5), 529-536.

Ka D, Hong S, Jeong U, Jeong M, Suh N, Suh JY et al. (2017) Structural and dynamic insights into the role of conformational switching in the nuclease activity of the *Xanthomonas albilineans* Cas2 in CRISPR-mediated adaptive immunity, *Struct Dyn* 4(5), 054701.

Ka D, Jang DM, Han BW and Bae E (2018) Molecular organization of the type II-A CRISPR adaptation module and its interaction with Cas9 via Csn2, *Nucleic Acids Res* 46(18), 9805-9815.

Ka D, Kim D, Baek G and Bae E (2014) Structural and functional characterization of *Streptococcus pyogenes* Cas2 protein under different pH conditions, *Biochem Biophys Res Commun* 451(1), 152-157.

Ka D, Lee H, Jung YD, Kim K, Seok C, Suh N et al. (2016) Crystal Structure of *Streptococcus pyogenes* Cas1 and Its Interaction with Csn2 in the Type II CRISPR-Cas System, *Structure* 24(1), 70-79.

Kieper SN, Almendros C, Behler J, Mckenzie RE, Nobrega FL, Haagsma AC et al. (2018) Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation, *Cell Rep* 22(13), 3377-3384.

Kim TY, Shin M, Huynh Thi Yen L and Kim JS (2013) Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity, *Biochem Biophys Res Commun* 441(4), 720-725.

Koo Y, Jung DK and Bae E (2012) Crystal structure of *Streptococcus pyogenes* Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure, *PLoS One* 7(3), e33401.

Koonin EV, Makarova KS and Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems, *Curr Opin Microbiol* 37, 67-78.

Kunin V, Sorek R and Hugenholtz P (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats, *Genome Biol* 8(4), R61.

Lee H, Dhingra Y and Sashital DG (2019) The Cas4-Cas1-Cas2 complex mediates precise prespacer processing during CRISPR adaptation, *Elife* 8.

Lee H, Zhou Y, Taylor DW and Sashital DG (2018) Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays, *Mol Cell* 70(1), 48-59 e45.

Levy A, Goren MG, Yosef I, Auster O, Manor M, Amitai G et al. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA, *Nature* 520(7548), 505-510.

Li M, Wang R and Xiang H (2014) Haloarcula hispanica CRISPR authenticates PAM of a target sequence to prime discriminative adaptation, *Nucleic Acids Res* 42(11), 7226-7235.

Lillestol RK, Shah SA, Brugger K, Redder P, Phan H, Christiansen J et al. (2009) CRISPR

families of the crenarchaeal genus Sulfolobus: bidirectional transcription and dynamic properties, *Mol Microbiol* 72(1), 259-272.

Makarova KS, Aravind L, Wolf YI and Koonin EV (2011a) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems, *Biol Direct* 6, 38.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI and Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action, *Biol Direct* 1, 7.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P et al. (2011b) Evolution and classification of the CRISPR-Cas systems, *Nat Rev Microbiol* 9(6), 467-477.

Makarova KS and Koonin EV (2015) Annotation and Classification of CRISPR-Cas Systems, *Methods Mol Biol* 1311, 47-75.

Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ et al. (2015) An updated evolutionary classification of CRISPR-Cas systems, *Nat Rev Microbiol* 13(11), 722-736.

Makarova KS, Wolf YI and Koonin EV (2018) Classification and Nomenclature of CRISPR-Cas Systems: Where from Here?, *CRISPR J* 1, 325-336.

Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG and James MN (2001) Crystallographic evidence for substrate-assisted catalysis in a bacterial beta-hexosaminidase, *J Biol Chem* 276(13), 10330-10337.

Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA, *Science* 322(5909), 1843-1845.

Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software, *J Appl Crystallogr* 40(Pt 4), 658-674.

Mcginn J and Marraffini LA (2016) CRISPR-Cas Systems Optimize Their Immune Response by Specifying the Site of Spacer Integration, *Mol Cell* 64(3), 616-623.

Modell JW, Jiang W and Marraffini LA (2017) CRISPR-Cas systems exploit viral DNA injection to establish and maintain adaptive immunity, *Nature* 544(7648), 101-104.

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J and Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, *J Mol Evol* 60(2), 174-182.

Nam KH, Ding F, Haitjema C, Huang Q, Delisa MP and Ke A (2012a) Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein, *J Biol Chem* 287(43), 35943-35952.

Nam KH, Haitjema C, Liu X, Ding F, Wang H, Delisa MP et al. (2012b) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system, *Structure* 20(9), 1574-1584.

Nam KH, Kurinov I and Ke A (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca2+-dependent double-stranded DNA binding activity, *J Biol Chem* 286(35), 30759-30768.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N et al. (2014) Crystal

structure of Cas9 in complex with guide RNA and target DNA, *Cell* 156(5), 935-949.

Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW and Doudna JA (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity, *Nat Struct Mol Biol* 21(6), 528-534.

Nunez JK, Lee AS, Engelman A and Doudna JA (2015) Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity, *Nature* 519(7542), 193-198.

Otwinowski Z and Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode, *Method Enzymol* 276, 307-326.

Pourcel C, Salvignol G and Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, *Microbiology* 151(Pt 3), 653-663.

Richter C, Dy RL, Mckenzie RE, Watson BN, Taylor C, Chang JT et al. (2014) Priming in the Type I-F CRISPR-Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer, *Nucleic Acids Res* 42(13), 8516-8526.

Rocha EP, Danchin A and Viari A (2001) Evolutionary role of restriction/modification systems as revealed by comparative genome analysis, *Genome Res* 11(6), 946-958.

Rollie C, Graham S, Rouillon C and White MF (2018) Prespacer processing and specific integration in a Type I-A CRISPR system, *Nucleic Acids Res* 46(3), 1007-1020.

Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G et al. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade, *Mol Cell* 52(1), 124-134.

Shiimori M, Garrett SC, Graveley BR and Terns MP (2018) Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci, *Mol Cell* 70(5), 814-824 e816.

Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E et al. (2015) Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems, *Mol Cell* 60(3), 385-397.

Thomas CM and Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria, *Nat Rev Microbiol* 3(9), 711-721.

Van Der Ploeg JR (2009) Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages, *Microbiology* 155(Pt 6), 1966-1976.

Van Houte S, Buckling A and Westra ER (2016) Evolutionary Ecology of Prokaryotic Immune Mechanisms, *Microbiol Mol Biol Rev* 80(3), 745-763.

Walker JR, Corpina RA and Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair, *Nature* 412(6847), 607-614.

Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M et al. (2015) Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems, *Cell* 163(4), 840-853.

Wei Y, Terns RM and Terns MP (2015) Cas9 function and host genome sampling in Type II-A CRISPR-Cas adaptation, *Genes Dev* 29(4), 356-361.

Westra ER, Van Erp PB, Kunne T, Wong SP, Staals RH, Seegers CL et al. (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively

supercoiled invader DNA by Cascade and Cas3, *Mol Cell* 46(5), 595-605.

Wiedenheft B, Van Duijn E, Bultema JB, Waghmare SP, Zhou K, Barendregt A et al. (2011) RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions, *Proc Natl Acad Sci U S A* 108(25), 10092-10097.

Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W and Doudna JA (2009) Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense, *Structure* 17(6), 904-912.

Wilkinson M, Drabavicius G, Silanskas A, Gasiunas G, Siksnys V and Wigley DB (2019) Structure of the DNA-Bound Spacer Capture Complex of a Type II CRISPR-Cas System, *Mol Cell*.

Wright AV and Doudna JA (2016) Protecting genome integrity during CRISPR immune adaptation, *Nat Struct Mol Biol* 23(10), 876-883.

Wright AV, Liu JJ, Knott GJ, Doxzen KW, Nogales E and Doudna JA (2017) Structures of the CRISPR genome integration complex, *Science* 357(6356), 1113-1118.

Xiao Y, Ng S, Nam KH and Ke A (2017) How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration, *Nature* 550(7674), 137-141.

Yosef I, Goren MG and Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli, *Nucleic Acids Res* 40(12), 5569-5576.

Accession numbers

The Atomic coordinates and structure factors for the SpCas1 and SpCas2 crystal structure have been deposited with the Protein Data Bank (Berman et al., 2000) under accession number 4ZKJ and 5ZYF, respectively.

Abstract in Korean

국문 초록

서울대학교 대학원
농생명공학부 응용생명화학전공
가동현

CRISPR (Clustered regularly interspaced shrot palindromic repeats, 주기적으로 반복되는 짧은 길이의 회문 구조를 가진 염기서열 집합체) 및 CRISPR 연관 유전자 (CRISPR-associated; *cas* gene)는 파아지 및 플라스미드와 같은 외래 유전 물질들의 침입에 대항하는 원핵생물의 적응 면역체계를 구성한다. CRISPR 배열은 이전에 숙주 세포를 공격하려고 시도하였던 침입자의 핵산 서열과 일치하는 짧은 고유 DNA 서열을 저장하는 유전적 기억 장치 역할을 한다. CRISPR 배열은 일련의 cas 유전자를 동반하며, 이 유전자들은 CRISPR-Cas 체계의 동작에 중요한 역할을 하는 단백질을 암호화 하고 있다. 여러 CIRPSR-Cas 체계 보편적으로 보존 된 2 개의 Cas 단백질인 Cas1 과 Cas2 단백질은 복합체를 형성하여 외래 핵산 단편을 숙주 CRISPR 배열로 통합시킨다. 유형 II-A CRISPR-Cas 시스템에서는 Cas1 및 Cas2 이외에도 다른 Cas 구성 요소인 Cas9 및 Csn2의 존재가 면역화 과정에 필수적이다. 이 논문에서는 *Streptococcus pyogenes* 의 type II-A Cas1 과 Cas2 단백질의 X 선 결정 구조를 보고하고 두 단백질 간의 상호 작용 특성을 규명하였다. 타입 II-A 특이 Cas 단백질로 알려진 Csn2 는 Cas1-Cas2 복합체와 함께 더 큰 복합체를 형성함을 확인되었다. Csn2 는 또한

Cas9와 직접 상호작용하는 반면, Cas9 은 Cas1-Cas2 복합체 또는 복합체의 개별 구성체와는 상호작용을 하지 않았다. 마지막으로 4 가지 II-A Cas 단백질 사이의 상호 작용들의 화학량론 및 결합 친화도를 측정 하였다. 이러한 결과는 지금까지 기술되지 않은 Csn2 의 역할과 Cas9 및 Cas1-Cas2 인테그라아제 모듈이 II-A CRISPR-Cas 시스템에서 어떻게 연결되는지를 보여주는 실험적 증거를 제공한다.

주요어: 크리스퍼, 단백질 복합체, sgRNA, *Streptococcus pyogenes*, 구조 생물학, X 선 결정학

학번: 2012-23366