농학석사학위논문

Anti-CRISPR 단백질, AcrIE4-F7 의 생화학적 연구

Biochemical Study of an Anti-CRISPR Protein, AcrIE4-F7

2023년 2월

서울대학교 대학원

농생명공학부 응용생명화학전공

이 규 진

A Dissertation for the Degree of Master of Science

Biochemical Study of an Anti-CRISPR Protein, AcrIE4-F7

February 2023

Gyujin Lee
Applied Life Chemistry Major
Department of Agricultural Biotechnology
Seoul National University

Anti-CRISPR 단백질, AcrIE4-F7 의 생화학적 연구

Biochemical Study of an Anti-CRISPR Protein, AcrIE4-F7

지도교수 배 의 영

이 논문을 농학석사학위논문으로 제출함
2023년 1월

서울대학교 대학원
농생명공학부 응용생명화학전공
이 규 진

이규진의 석사학위논문을 인준함
2023년 1월

위 원 장 송 영 훈 (인)

부 위 원 장 배 의 영 (인)

위 원 이 상 기 (인)

Abstract

To defend invasions of bacteriophages and plasmids, bacteria and archaea employ the clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) defense system. Type I-E and I-F CRISPR-Cas systems exploit CRISPR-associated complex for antiviral defense (Cascade) to recognize the invading DNA. In turn, bacteriophages have evolved anti-CRISPR (Acr) proteins to neutralize CRISPR-mediated immunity. AcrIE4-F7 encoded by *Pseudomonas citronellolis* is a fused protein of AcrIE4 and AcrIF7 which inactivates both type I-E and I-F CRISPR-Cas system. In this study, I reported the target Cas proteins of AcrIE4-F7 by analyzing the interaction between AcrIE4-F7 and Cas proteins which comprise the type I-E and I-F Cascade complex. The N- and C-terminal domains of AcrIE4-F7 inhibit target DNA binding by interacting with the PAM recognition site of Cas8e and Cas8f, respectively. Pairwise sequence alignment and mutation analyses demonstrated that conserved negative-charged residues in each Acr domain are essential for interaction with their Cas8 target. These results suggest PAM recognition sites are the main targets of AcrIE4-F7, which is a common inhibition mechanism against divergent CRISPR-Cas types.

Keywords: CRISPR-Cas, Anti-CRISPR, AcrE4-F7, Cascade complex, *Pseudomonas aeruginosa*, PAM recognition site.

***Student Number*: 2021-27675**

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Aca	Acr-associated
Acr	Anti-CRISPR
AcrIE4-F7^{NTD}	N-terminal domain of AcrIE4-F7
AcrIE4-F7^{CTD}	C-terminal domain of AcrIE4-F7
APBS	Adaptive Poisson-Boltzmann Solver
BME	β-mercaptoethanol
Cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defense
CBASS	Cyclic oligonucleotide-based anti-phage signaling system
CD	Circular dichroism
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
Cryo-EM	Cryogenic electron microscopy
DTT	1,4-Dithiothreitol
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IMAC	Immobilized metal ion chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
MBP	Maltose-binding protein
MOPS	3-(N-morpholino)propanesulfonic acid
NMR	Nuclear magnetic resonance
PAM	Protospace adjacent motif
PCR	Polymerase chain reaction
Pre-crRNA	Precursor crRNA
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco etch virus
Tris	Tris(hydroxymethyl)aminomethane

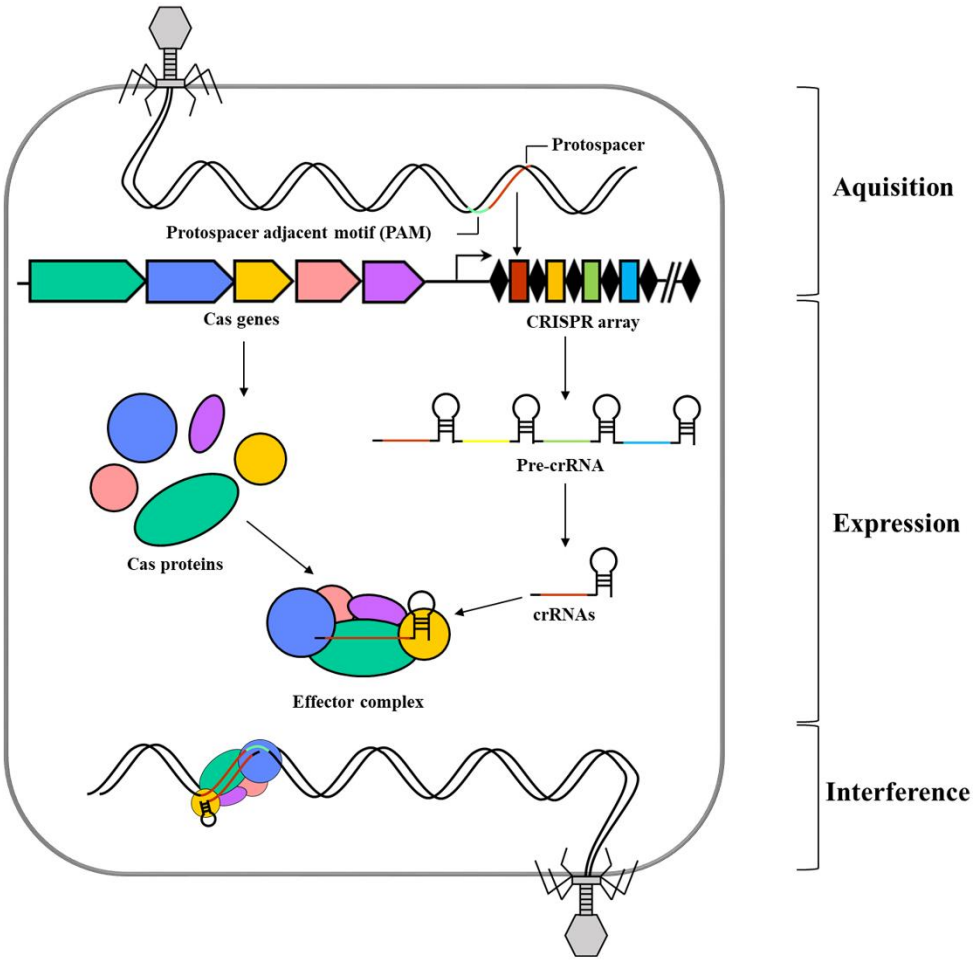
Introduction

CRISPR-Cas system

To defend against invading genetic elements, bacteria have evolved diverse defense immune systems such as restriction-modification, abortive infection, and cyclic oligonucleotide-based anti-phage signaling system (CBASS) (Bernheim and Sorek, 2020). Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) defense system provides bacteria and archaea with an adaptive immune system (Barrangou et al., 2007; Brouns et al., 2008; Wiedenheft et al., 2012). CRISPR-Cas system acquires genetic fragments from the foreign nucleic acids and gives sequence-specific protection against them (Marraffini, 2015). CRISPR-Cas system consists of Cas genes and a CRISPR array (Fig. 1). CRISPR array has 'spacer' sequences derived from bacteriophages or plasmid (protospacer) and 'repeat' sequences which are located on both sides of 'spacer' sequence. CRISPR-Cas system operates through three stages (Fig. 1): acquisition, expression, and interference (Marraffini, 2015). In the acquisition step, protospacer from foreign DNAs is integrated into the host CRISPR array. During the acquisition step, the protospacer adjacent motif (PAM) sequence determines the candidate of the spacer (Mojica et al., 2009; Datsenko et al., 2012; Fineran and Charpentier, 2012). In the expression step, Cas genes are transcribed and translated. CRISPR array is transcribed into a precursor CRISPR RNA (Pre-crRNA), which is processed into a crRNA that has a single 'spacer' sequence. The mature crRNA and Cas proteins form an effector complex. In the interference step, the effector complex recognizes the target DNA which has the complementary sequence with a spacer sequence,

Figure 1. Overview of the CRISPR-Cas system

CRISPR-Cas system is composed of Cas genes and a CRISPR array. The black diamond and the colored rectangles represent repeats and phage-derived spacers, respectively. CRISPR-Cas system works through three stages: acquisition, expression, and interference. During the acquisition step, genetic fragment from invading DNA is integrated into the 'spacer' region. In the expression step, Cas genes are transcribed and translated. The CRISPR array is transcribed and processed into a crRNA, which forms an effector complex with single or several Cas proteins. In the interference step, the effector complex recognizes the target DNA and degrades the target DNA. The PAM sequence is essential for the acquisition and interference step.


Protospacer
Protospacer adjacent motif (PAM)
Cas genes
CRISPR array
Pre-crRNA
Cas proteins
crRNAs
Effector complex
Aquisition
Expression
Interference

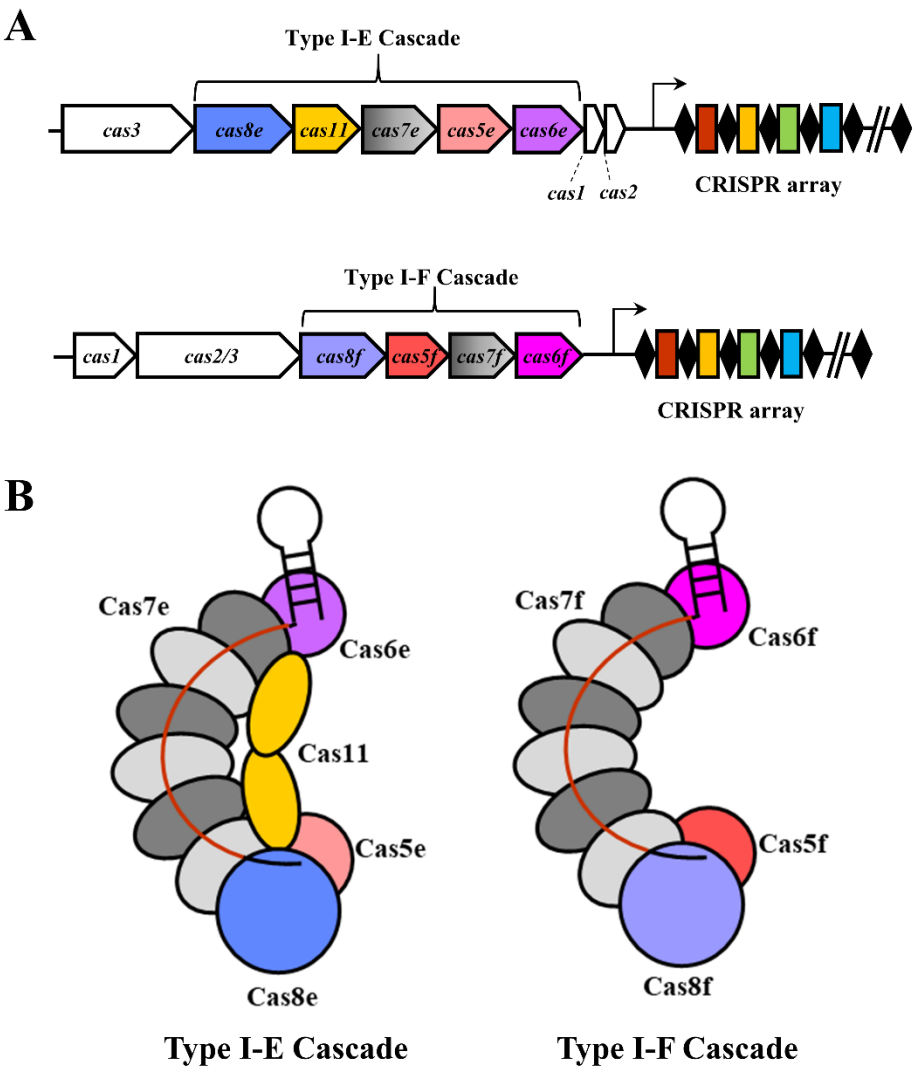
and the target DNA which has the complementary sequence with a spacer sequence, and cleaves the target DNA. During the target recognition, the effector complex recognizes the PAM sequence to distinguish the interference target (Mojica et al., 2009; Datsenko et al., 2012).

The updated classification of CRISPR-Cas systems includes two classes, six types, and 33 subtypes (Makarova et al., 2020). The divergent CRISPR-Cas systems were classified by effector complex (class), signature gene (type), and cas gene operon (subtype) (Makarova et al., 2020). While Class 1 (Type I, III, and IV) composes multi-subunit effector complexes, Class 2 (Type II, V, and VI) possesses a single multi-domain protein to recognize the target. Type I CRISPR-Cas system has Cas3 as its signature gene and it is categorized into 9 subtypes. They constitute a seahorse-formed CRISPR-associated complex for antiviral defense (Cascade) to defend against invading DNAs.

In the subtype I-E and I-F CRISPR-Cas system, eight and six Cas proteins exist in the Cas locus, respectively (Fig. 2A). Compared to the type I-E CRISPR-Cas system, the type I-F CRISPR-Cas system does not have the *cas11* gene and possesses the fusion gene of *cas2* and *cas3*, demonstrating a functional connection between the acquisition and interference step. Cas1-Cas2 complex and Cas1-Cas2/3 mediate the acquisition process. And Cas3 and Cas2/3 function as nucleases, so they are recruited to the Cascade to degrade the target DNA. While the type I-E Cascade is composed of Cas8e, Cas11, Cas7e, Cas5e, Cas6e, and crRNA with a stoichiometry of $Cas8e_1$:$Cas11_2$:$Cas7e_6$:$Cas5e_1$:$Cas6e_1$:$crRNA_1$, the type I-F Cascade consists of Cas8f, Cas7f, Cas5f, Cas6f, and crRNA with a stoichiometry of $Cas8f_1$:$Cas7f_6$:$Cas5f_1$:$Cas6f_1$:$crRNA_1$ (Xue and Sashital, 2019) (Fig. 2B). Cas6 cleaves Pre-crRNA to make mature crRNA, and interacts with the 3' stem of crRNA.

Figure 2. Type I-E and I-F CRISPR-Cas system

(A) Schematic representation of the type I-E (*top*) and I-F (*bottom*) CRISPR-Cas system. Type I-E and I-F CRISPR-Cas systems possess eight and six Cas genes, respectively. (B) The architecture of type I-E (*left*) and I-F (*right*) Cascade complex. While the type I-E Cascade complex exhibits a subunit stoichiometry of $Cas8e_1$:$Cas11_2$:$Cas7e_6$:$Cas5e_1$:$Cas6e_1$:$crRNA_1$, the type I-F Cascade complex shows a subunit stoichiometry of $Cas8f_1$:$Cas7f_6$:$Cas5f_1$:$Cas6f_1$:$crRNA_1$.


A
Type I-E Cascade
cas3
cas8e
cas11
cas7e
cas5e
cas6e
cas1
cas2
CRISPR array
Type I-F Cascade
cas1
cas2/3
cas8f
cas5f
cas7f
cas6f
CRISPR array
B
Cas7e
Cas6e
Cas11
Cas5e
Cas8e
Type I-E Cascade
Cas7f
Cas6f
Cas5f
Cas8f
Type I-F Cascade

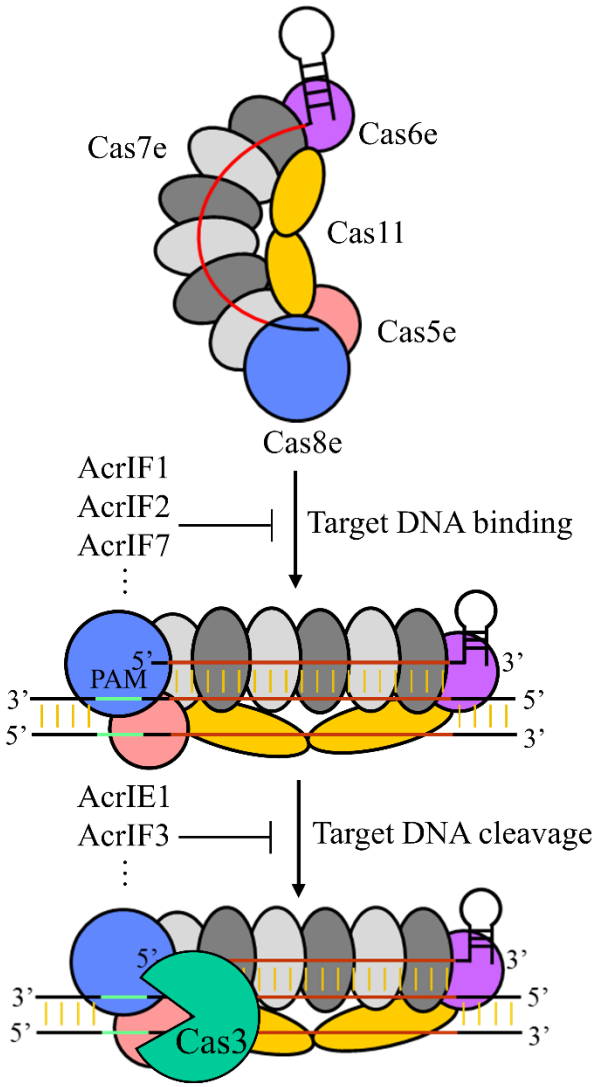
Cas5 recognizes the repeat-derived 5' handle of mature crRNA. Cas8 is responsible for PAM recognition and binds to Cas5 to form the tail of Cascade. Six copies of Cas7 interact with the spacer part of crRNA, making the Cascade backbone. Two copies of Cas11 associate with Cas7 subunits through electrostatic interaction and form the belly which stabilizes the crRNA and target DNA.

Anti-CRISPR proteins

In response to the CRISPR-Cas system, phages have evolved anti-CRISPR (Acr) proteins to suppress the host CRISPR-Cas proteins. Acr proteins were first discovered in phages infecting *Pseudomonas aeruginosa* UCBPP-PA14 which has type I-F CRISPR-Cas system, and the anti-CRISPR activities of these proteins were confirmed by phage plaque assay (Pawluk et al., 2014). Many Acr proteins have been found in the genomic DNA of phages or mobile genetic elements through bioinformatics analyses and functional assay (Bondy-Denomy et al., 2013; Pawluk et al., 2016; Marino et al., 2018; Pawluk et al., 2018; Pinilla-Redondo et al., 2020; Leon et al., 2021). So far, 24 Acr proteins were described to target the type I-F CRISPR-Cas system (i.e., AcrIF1-24), and 9 Acr proteins were found to inhibit the type I-E CRISPR-Cas system (i.e., AcrIE1-9). Type I-E and I-F anti-CRISPR proteins suppress the CRISPR-Cas system by inhibiting either target DNA binding or target DNA cleavage (Fig. 3). To prevent the Cascade from binding to target DNA, Acr proteins directly associate with Cascade. AcrIE2 was reported to bind to Cascade, but the identification of the target Cas protein was still unknown (Mejdani et al., 2021). The complex structures of Cascade and Acr proteins revealed that

Figure 3. Inhibitory mechanism of type I-E and I-F Acr proteins

Type I-E and I-F anti-CRISPR proteins suppress the CRISPR-Cas system by inhibiting either target DNA binding or target DNA cleavage. Most Acr proteins bind to the Cascade to interrupt the recruitment of target DNA or nuclease. Also, an Acr protein which has enzymatic activity modifies the key residue of the Cas protein.


Cas7e
Cas6e
Cas11
Cas5e
Cas8e
AcrIF1
AcrIF2
AcrIF7
Target DNA binding
5'
PAM
3'
3'
5'
5'
3'
AcrIE1
AcrIF3
Target DNA cleavage
5'
3'
3'
5'
5'
3'
Cas3

AcrIF2/6/7/10 target the PAM recognition site of Cas8f (Chowdhury et al., 2017; Guo et al., 2017; Zhang et al., 2020; Gabel et al., 2021) Also, AcrIF13 was found to have the same binding interface as AcrIF2 through competitive assay (Wang et al., 2022). In addition, AcrIF1/8/9/14/24 interact with Cas7f subunits which compose the Cascade backbone (Guo et al., 2017; Zhang et al., 2020; Gabel et al., 2021; Yang et al., 2022). Moreover, there is an Acr protein that has enzymatic activity. AcrIF11 functions as a mono-ADP-ribosyltransferase, which modifies N250 of Cas8f, a key residue for PAM sequence recognition of target DNA (Niu et al., 2020). The other inhibitory mechanism of Acr proteins is to prevent the degradation of target DNA. AcrIF3 and AcrIE1 directly associate with the Cas2/3 and Cas3 nuclease and prevents their interaction with the Cascade complex (Wang et al., 2016; Pawluk et al., 2017). AcrIF4 and AcrIF5 block the Cas2/3 recruitment by interacting with Cascade. AcrIF4 binds to the helical bundle of Cas8f and hinders the rotation of Cas8f induced by dsDNA binding (Gabel et al., 2021; Gao et al., 2022). AcrIF5 binds only to the Cas7f subunits of the Cascade backbone where the conformational change occurred by dsDNA binding (Xie et al., 2022).

Anti-CRISPR protein, AcrIE4-F7

The *acrIE4-F7* gene was identified from a homolog search of *aca1* (Acr- associated gene) in the *Pseudomonas* genome (Marino et al., 2018). AcrIE4-F7 which is a fusion protein of AcrIE4 and AcrIF7 was isolated from *Pseudomonas citronellois* and exhibited the dual inhibition of type I-E and I-F CRISPR-Cas system (Marino et al., 2018). The N-terminal domain of AcrIE4-F7 (AcrIE4-F7^{NTD}, residues 1-52) and C-terminal domain of AcrIE4-F7 (AcrIE4-F7^{CTD}, residues 53-119) have high sequence identity with AcrIE4 (69%) and AcrIF7 (55%). The structure and inhibition

mechanism of AcrIF7 was revealed (Kim et al., 2020; Gabel et al., 2021), but that of AcrIE4 is still unknown. AcrIF7 targets the PAM recognition site of Cas8f and competes with target DNA for Cas8f binding.

In this study, I performed biochemical studies of AcrIE4-F7 by identifying its target Cas proteins and analyzing the binding interface of each domain. To figure out the target Cas protein of AcrIE4-F7^{NTD}, I conducted the analytical size exclusion chromatography between AcrIE4-F7 and Cas components which comprise the type I-E Cascade. Also, I tested which domain of AcrIE4-F7 is responsible for Cas8e binding. From the interaction test, I found the AcrIE4-F7^{NTD} bind to the Cas8e subunit. In the next step, I modeled the *P. aeruginosa* Cas8 based on the structure of *T. fusca* Cascade (PDB: 5U07). Three Lys residues are located near the putative PAM recognition site, and mutational analyses of Cas8e revealed that these residues are essential for Cas8e binding. Moreover, mutational analyses of AcrIE4-F7 demonstrate the negative-charge residues in the α2 helix are crucial for Cas8e binding. AcrIE4-F7^{CTD} binds to the Cas8f:Cas5f heterodimer in a similar manner to AcrIF7 which targets the PAM recognition site of Cas8f. According to the pairwise sequence alignment of AcrIF7 and AcrIE4-F7, the key residues for Cas8f binding are conserved in AcrIE4-F7^{CTD}. Also, AcrIF7 and AcrIE4-F7^{CTD} have very similar structures. These results indicate that AcrIE4-F7^{CTD} is a functional and structural homolog of AcrIF7. In summary, the biochemical analyses in this study revealed that AcrIE4-F7 targets the PAM recognition site of its Cas8 targets, emphasizing a common inhibition mechanism against divergent CRISPR-Cas types.

Materials and Methods

Cloning

The synthetic gene of AcrIE4-F7 (NCBI accession number WP_064584002.1) was amplified using polymerase chain reaction (PCR). The gene was cloned into pET28a containing N-terminal $(His)_6$-maltose binding protein (MBP) tag with a tobacco etch virus (TEV) protease cleavage site. Mutant AcrIE4-F7 genes were generated by PCR with mutagenic primers, and these were made by Dr. Sung-Hyun Hong at Seoul National University. The genetic fragments encoding the N-terminal and C-terminal domain of AcrIE4-F7 were amplified by PCR from its full-length gene and cloned into pET21a with C-terminal $(His)_6$ tag and pET28a with N-terminal $(His)_6$-MBP tag with a TEV protease cleavage site, respectively.

The genes of type I-E Cas protein (i.e. Cas5e, Cas6e, Cas7e, Cas8e, and Cas11) were amplified by PCR from *P. aeruginosa* PRD-10 and *Escherichia coli* DH5α genomic DNAs and cloned into pET28a with N-terminal $(His)_6$-MBP tag and TEV protease cleavage site (Hong et al., 2022). The constructs of type I-E Cas proteins from *P. aeruginosa* PRD-10 were made by Mr. Jasung Koo at Seoul National University. $(His)_6$-MBP tagged Cas8e construct plasmid (pET28a) was used as a template for site-directed mutagenesis. Mutant Cas8e genes were generated by PCR with mutagenic primers.

The synthetic gene of Cas8f and Cas5f genes from *Xanthomonas albilineans* GPE PC73 were cloned into pET28a with N-terminal $(His)_6$-MBP tag and a TEV cleavage site and pET21a without a tag, respectively (Hong et al., 2018).

Table 1. The information on genes used in this study

Source organism	Protein name	Genbank ID
Pseudomonas citronellolis	AcrIE4-F7	WP_064584002.1
Pseudomonas aeruginosa PRD-10	Cas5e	WP_004348746.1
	Cas6e	WP_004348749.1
	Cas7e	WP_004348745.1
	Cas8e	WP_004348741.1
	Cas11	WP_016561674.1
Escherichia coli K-12	Cas8e	NP_417240.1
Xanthomonas albilineans GPE PC73	Cas8f	WP_012917520.1
	Cas5f	WP_012917521.1

Table 2. The primers used for cloning

Primer		5'-3' Sequence
AcrIE4-F7 N-terminal $(His)_6$-MBP	Forward	ATTTCCAGGGCCATATGTCTACCCAGTATACCTATCAG
	Reverse	GGTGGTGGTGCTCGAGTTAAGCTTCGTGTTCAACCAG
AcrIE4-F7^{NTD} C-terminal $(His)_6$	Forward	AAGGAGATATACATATGTCTACCCAGTATACCTATCAG
	Reverse	GGTGGTGGTGCTCGAGTTTTTCTTCACCGAACGCTTC
AcrIE4-F7^{CTD} N-terminal $(His)_6$-MBP	Forward	ATTTCCAGGGCCATATGTCTCCGAAATTCAGCACC
	Reverse	GGTGGTGGTGCTCGAGTTAAGCTTCGTGTTCAACCAG
E. coli Cas8e N-terminal $(His)_6$-MBP	Forward	ATTTCCAGGGCCATATGAATTTGCTTATTGATAACTGGATC
	Reverse	GGTGGTGGTGCTCGAGTTAGCCATTTGATGGCCCTC
P. aeruginosa Cas8e K176E	Forward	**GAA**ACTGTTCGCGATTCGG
	Reverse	**TTC**GACCAACCCGCCCG
P. aeruginosa Cas8e K183E	Forward	GACTGTTCGCGATTCGGAC**GAA**GCCGGCGCACTGGCC
	Reverse	GGCCAGTGCGCCGGC**TTC**GTCCGAATCGCGAACAGTC
P. aeruginosa Cas8e K357E	Forward	GCCAGCGACCAGGCC**GAA**CTGTTGCGCTGGCGCTC
	Reverse	GAGCGCCAGCGCAACAG**TTC**GGCCTGGTCGCTGGC

*The bold type indicates the desired mutation.

Protein expression and purification

The AcrIE4-F7 construct was transformed into *E. coli* BL21(DE3) cell, which was grown in the LB medium at 37℃ to an optical density at 600 nm of 0.6. Protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 17℃ for 16 h. The cells were harvested by centrifugation at 5000 g for 5 min and resuspended in lysis buffer (20 mM 3-(N-morpholino)propanesulfonic acid (MOPS), pH 7.0, 300 mM NaCl, 5 mM β-mercaptoethanol (BME), 10% (w/v) glycerol, 30 mM imidazole, 0.3 mM phenylmethylsulfonyl fluoride, and 0.02% (w/v) Triton X-100). After sonication and centrifugation, the resulting supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer A (20 mM MOPS, pH 7.0, 300 mM NaCl, 5 mM BME, 10% (w/v) glycerol and 30 mM imidazole). After washing with the same buffer, the bound protein was eluted by applying a linear gradient of imidazole (up to 500 mM). Pooled fractions were treated with TEV protease to cleave the $(His)_6$-MBP, and dialyzed with dialysis buffer (20 mM MOPS, pH 7.0, 300 mM NaCl, 5 mM BME, 10% (w/v) glycerol). The $(His)_6$-MBP tag was separated using the 5 mL HisTrap HP column (GE Healthcare). AcrIE4-F7 was further purified by size exclusion chromatography (SEC) using HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with SEC buffer A (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0, 150 mM NaCl and 2 mM 1,4-dithiothreitol (DTT)).

The genetic fragments encoding N- and C-terminal domains of AcrIE4-F7 were transformed into *E. coli* BL21 (DE3) cells and expressed as described above for the full-length AcrIE4-F7. The proteins were loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer A. After washing the column

with the same buffer, the bound proteins were eluted by applying a linear gradient of imidazole (up to 500 mM). The $(His)_6$-MBP tag of the C-terminal domain was cleaved by TEV protease during dialysis and separated with the 5 mL HisTrap HP column (GE Healthcare). Finally, the proteins were further purified by SEC using a HiLoad 16/60 Superdex 75 column (GE Healthcare) pre-equilibrated with SEC buffer A. The wild-type and mutant constructs of type I-E Cas proteins were transformed into *E. coli* BL21(DE3) cells and expressed individually as described above for the expression of AcrIE4-F7. Because of the poor expression and low solubility of the proteins, the protein samples were purified without cleavage of the N-terminal tag. The Cas proteins were loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer B (20 mM HEPES, pH 7.0, 500 mM NaCl, 5 mM BME, 20% (w/v) glycerol and 30 mM imidazole). After washing the columns with the same buffer, the bound proteins were eluted by applying a linear gradient of imidazole (up to 500 mM). Finally, the proteins were further purified by SEC using HiLoad 16/60 Superdex 200 column (GE Healthcare) or HiLoad 26/60 Seperdex 200 column (GE Healthcare) pre-equilibrated with SEC buffer A and 10% (w/v) glycerol.

To produce Cas8f:Cas5f heterodimer, the Cas8f construct containing N-terminal $(His)_6$-MBP tag and Cas5f construct with no tag were co-transformed into *E. coli* BL21 (DE3) and co-expressed with 0.5 mM IPTG at 17°C at 16 h. The $(His)_6$-MBP tagged Cas8f:Cas5f heterodimer was loaded onto a 5 mL HisTrap HP column (GE Healthcare) pre-equilibrated with binding buffer C (20 mM tris(hydroxymethyl)aminomethane (Tris)-HCl, pH 7.5, 300 mM NaCl, 5 mM BME 10% (w/v) glycerol and 30 mM imidazole). After washing the column with the same buffer, the protein sample was eluted by applying a linear gradient of imidazole (up

to 500 mM). The N-terminal $(His)_6$-MBP tag was cleaved by TEV protease during dialysis and separated on a 5 mL HisTrap HP column (Healthcare). The Cas8f:Cas5f heterodimer was finally purified by SEC using HiLoad 16/60 superdex 200 column (GE Healthcare) pre-equilibrated with SEC buffer C (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM DTT and 5% (w/v) glycerol).

Analytical size exclusion chromatography (SEC)

Analytical SEC was performed using a Superdex 200 10/300 GL column (GE Healthcare). To determine the interaction between Acr and type I-E, I-F Cas proteins, the column was equilibrated with analytical SEC buffer A (20 mM HEPES, pH 7.0, 150 mM NaCl, 2 mM DTT, 5% (w/v) glycerol) and aSEC buffer B (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 2 mM DTT, 5% (w/v) glycerol). Proteins (20 µM each) were mixed and incubated for 1 hour at 4 ℃. After incubation, the samples were loaded onto the column at a flow rate of 0.5 ml/min. The eluted SEC fractions were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by Coomassie staining.

Circular dichroism (CD)

To check the secondary structure of mutant proteins, CD spectra were measured with protein samples (0.3 µM) in 500 µL buffer (10 mM sodium phosphate, pH 7.2) at 25℃ using a J-815 CD spectropolarimeter (Jasco).

Pairwise sequence alignment

To compare the sequence of Cas and Acr proteins, sequences were aligned by the Clustal Omega program (Sievers et al., 2011), and visualized by the Jalview program (Waterhouse et al., 2009).

RESULTS

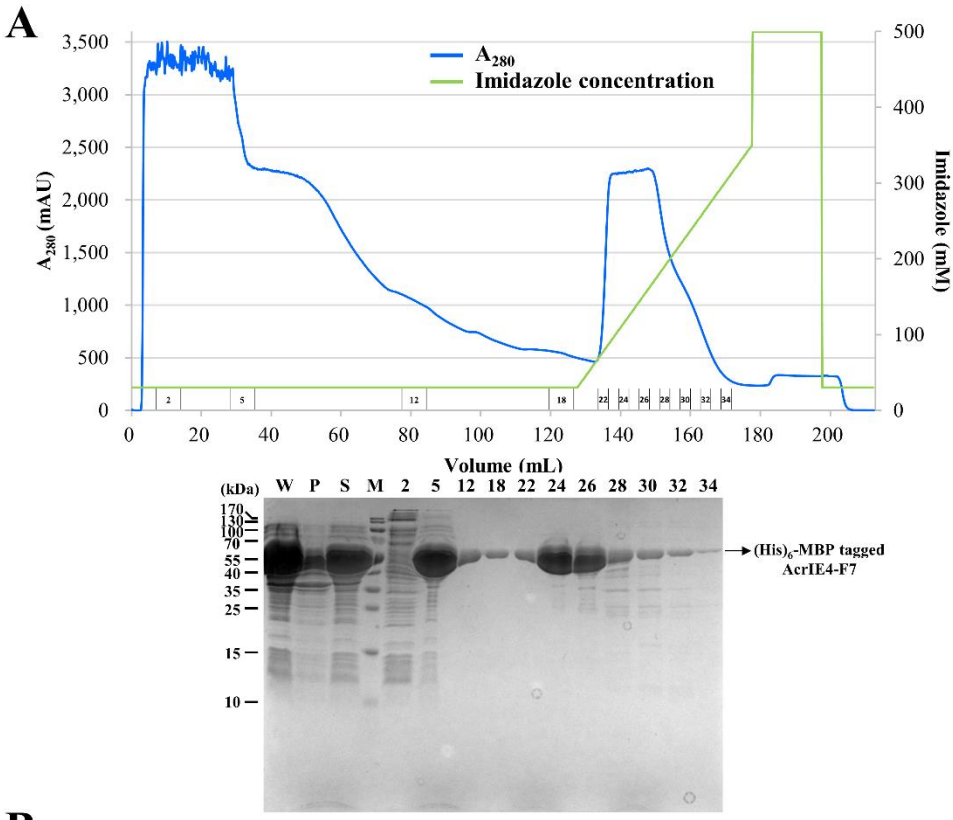
Purification of AcrIE4-F7

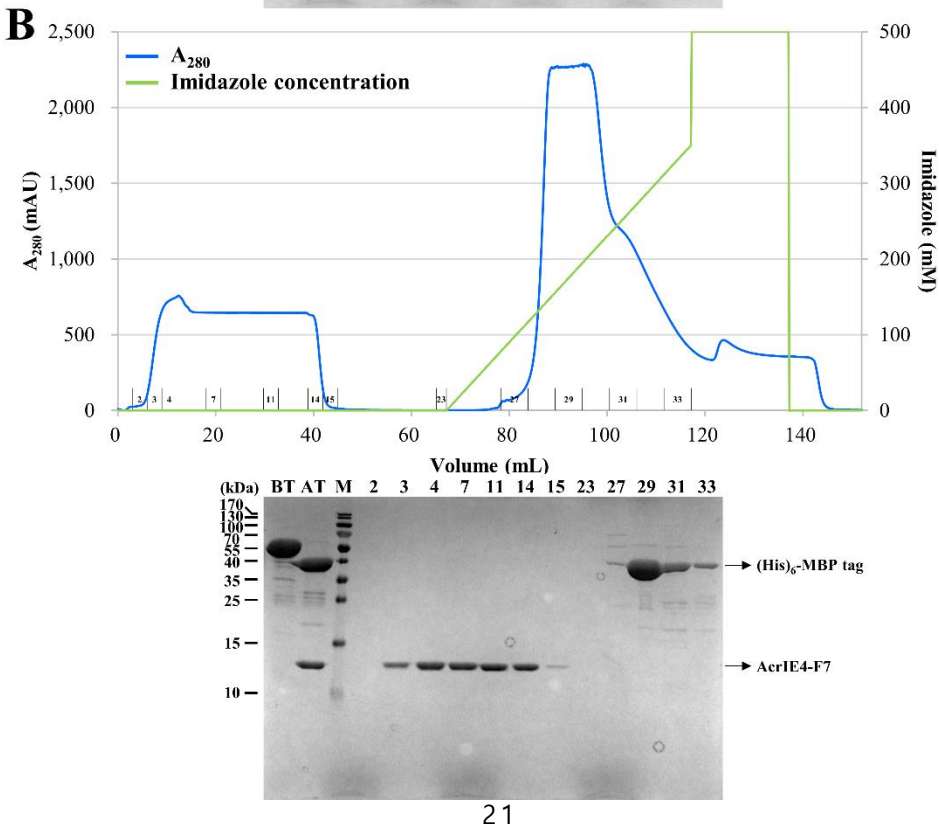
To purify AcrIE4-F7, AcrIE4-F7 with N-terminal $(His)_6$-MBP tag was expressed in *E. coli* BL21(DE3) cells. AcrIE4-F7 was purified in three-step including first immobilized metal ion chromatography (IMAC) – second IMAC – SEC. During the first IMAC, $(His)_6$-MBP tagged AcrIE4-F7 bound to the 5mL HisTrap HP column by coordination bond between $(His)_6$ tag and nickel ion in the column. As the concentration of imidazole increased, the bound protein was eluted between 62 µM and 325 µM of imidazole (Fig. 4A). After TEV protease was treated during the dialysis, the second IMAC was performed to separate the $(His)_6$-MBP tag. Since $(His)_6$-MBP tag was cleaved from AcrIE4-F7, AcrIE4-F7 didn't interact with the column and eluted at the early fraction (Fig. 4B). After concentration of pooled fractions, size exclusion chromatography was performed for high purity of AcrIE4-F7. AcrIE4-F7 (calculated MW: 14 kDa) was eluted at 81.8 mL (measured MW: 12.3 kDa) and it forms a monomer (Fig. 4C). I collected the number of fractions 19 to 26 from the SEC and the samples were concentrated.

To test which domain of AcrIE4-F7 interacts with Cas proteins, AcrIE4-F7^{NTD} with C-terminal $(His)_6$ tag and AcrIE4-F7^{CTD} with N-terminal $(His)_6$-MBP tag was expressed in *E. coli* BL21(DE3) cells and purified. AcrIE4-F7^{NTD} was purified in two-step including IMAC – SEC. During the IMAC, AcrIE4-F7^{NTD} was eluted between 59 µM and 250 µM of imidazole (Fig. 5A). In the SEC, AcrIE4-F7^{NTD} (calculated MW: 7 kDa) was eluted at 92.6 mL (measured MW: 6.6 kDa) and it forms a monomer (Fig. 5B). I collected the number of fractions 29 to 35 from the

Figure 4. Elution profiles of AcrIE4-F7

Elution profiles (*top*) and SDS-PAGE (*bottom*) analyses of AcrIE4-F7 from each purification step. The protein absorbance at 280 nm and concentration of imidazole are shown in *blue* and *green*, respectively. Eluted fractions were analyzed with 15% SDS-PAGE gel and visualized by Coomassie staining. (A) The first IMAC purification. The bound $(His)_6$-MBP tagged AcrIE4-F7 was eluted with a linear gradient of imidazole. (B) The second IMAC purification. After TEV proteolysis, $(His)_6$-MBP tag was removed. (C) The SEC purification. Lane designation on the SDS-PAGE gel: W, whole cell lysate; P, pellet; S, supernatant; M, protein size marker; BT, sample before TEV proteolysis; AT, sample after TEV proteolysis; I, injection sample (AcrIE4-F7: 14 kDa, $(His)_6$-MBP tag: 43kDa).


A
A_{280}
Imidazole concentration
A_{280} (mAU)
Imidazole (mM)
Volume (mL)
(kDa) W P S M 2 5 12 18 22 24 26 28 30 32 34
(His)$_6$-MBP tagged AcrIE4-F7


B
A_{280}
Imidazole concentration
A_{280} (mAU)
Imidazole (mM)
Volume (mL)
(kDa) BT AT M 2 3 4 7 11 14 15 23 27 29 31 33
(His)$_6$-MBP tag
AcrIE4-F7

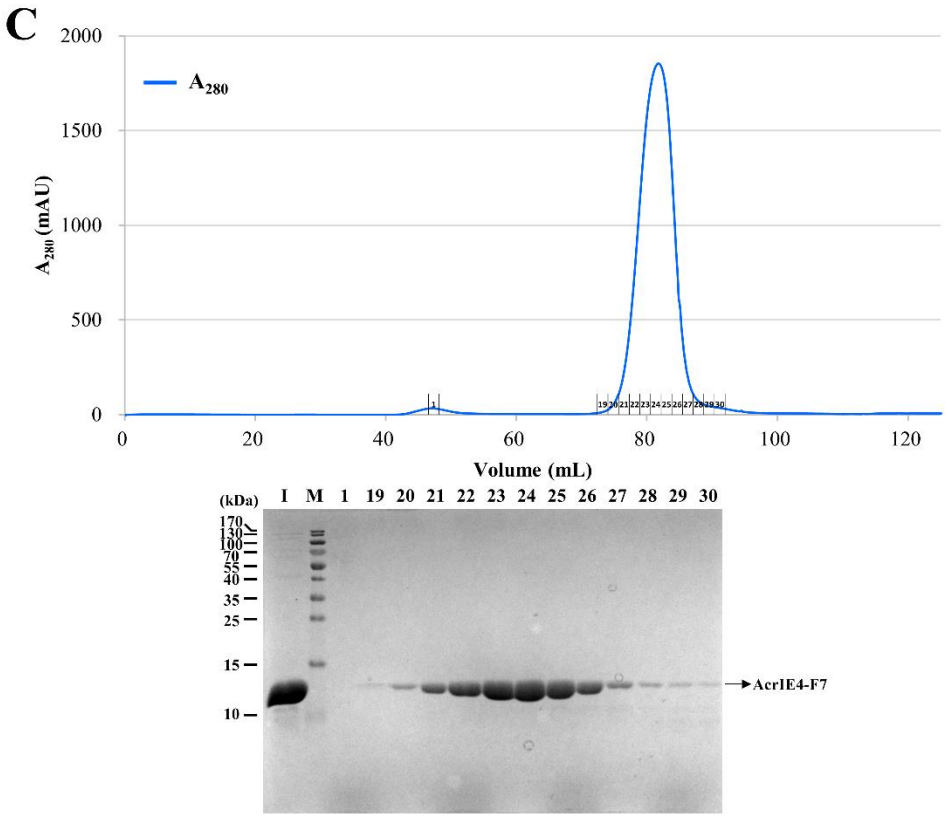
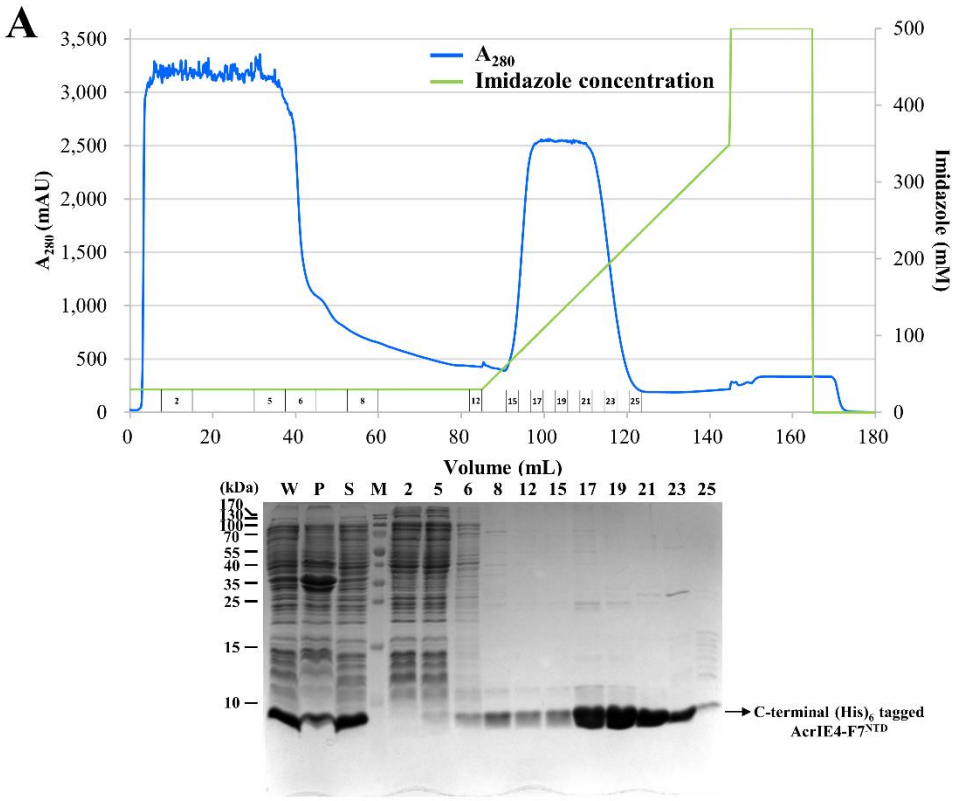
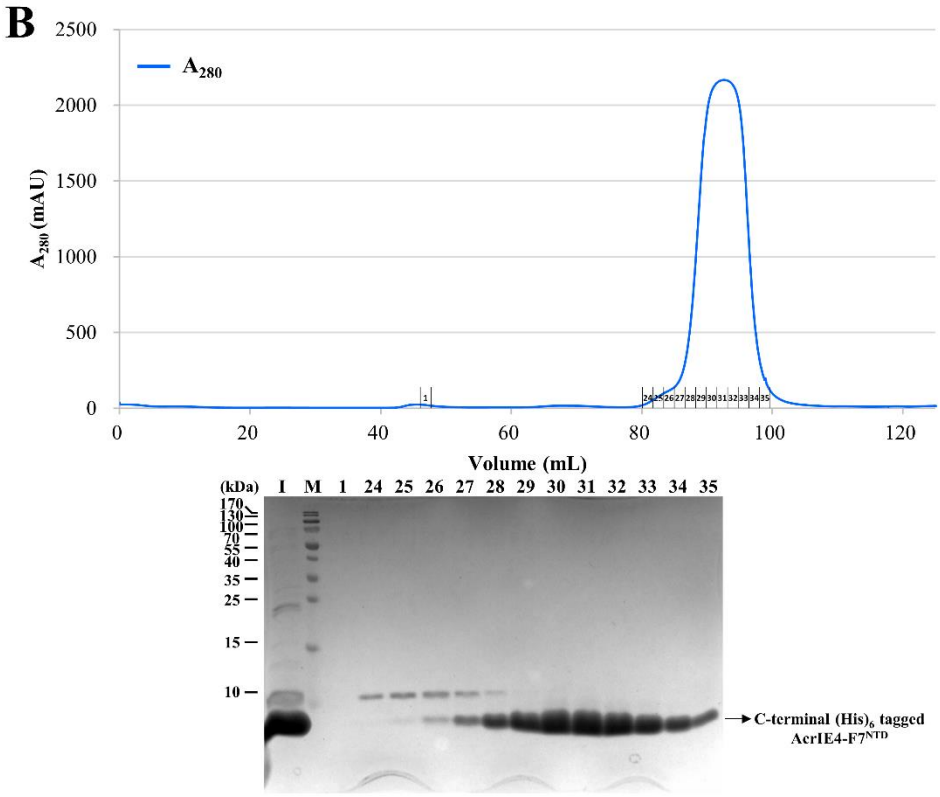

C
2000
1500
1000
500
0
A280
A280 (mAU)
0
20
40
60
80
100
120
Volume (mL)
(kDa)
I M 1 19 20 21 22 23 24 25 26 27 28 29 30
170
130
100
70
55
40
35
25
15
10
AcrIE4-F7

Figure 5. Elution profiles of AcrIE4-F7^{NTD}

Elution profiles (*top*) and SDS-PAGE (*bottom*) analyses of AcrIE4-F7^{NTD} from each purification step. The AcrIE4-F7 with C-terminal $(His)_6$ tag was purified by two-step purification including IMAC (A) – SEC (B). The protein absorbance at 280 nm and concentration of imidazole are shown in *blue* and *green*, respectively. Eluted fractions were analyzed with 15% SDS-PAGE gel and visualized by Coomassie staining. Lane designation on the SDS-PAGE gel: W, whole cell lysate; P, pellet; S, supernatant; M, protein size marker; I, injection sample (AcrIE4-F7^{NTD}: 7 kDa).


A
A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)
(kDa) W P S M 2 5 6 8 12 15 17 19 21 23 25
C-terminal (His)6 tagged AcrIE4-F7NTD


B
A280
A280 (mAU)
Volume (mL)
(kDa) I M 1 24 25 26 27 28 29 30 31 32 33 34 35
C-terminal (His)6 tagged AcrIE4-F7NTD

SEC. AcrIE4-F7^{CTD} was purified by the same method with full-length AcrIE4-F7. During the first IMAC, AcrIE4-F7^{CTD} was eluted between 69 μM and 375 μM of imidazole (Fig. 6A). After TEV protease proteolysis, $(His)_6$-MBP tag was separated in the second IMAC (Fig. 6B). In the SEC, AcrIE4-F7^{CTD} (calculated MW: 8 kDa) was eluted at 88.0 mL (measured MW: 8.6 kDa) and it forms a monomer (Fig. 6C). I collected the number of fractions 24 to 33 from the SEC and the samples were concentrated.

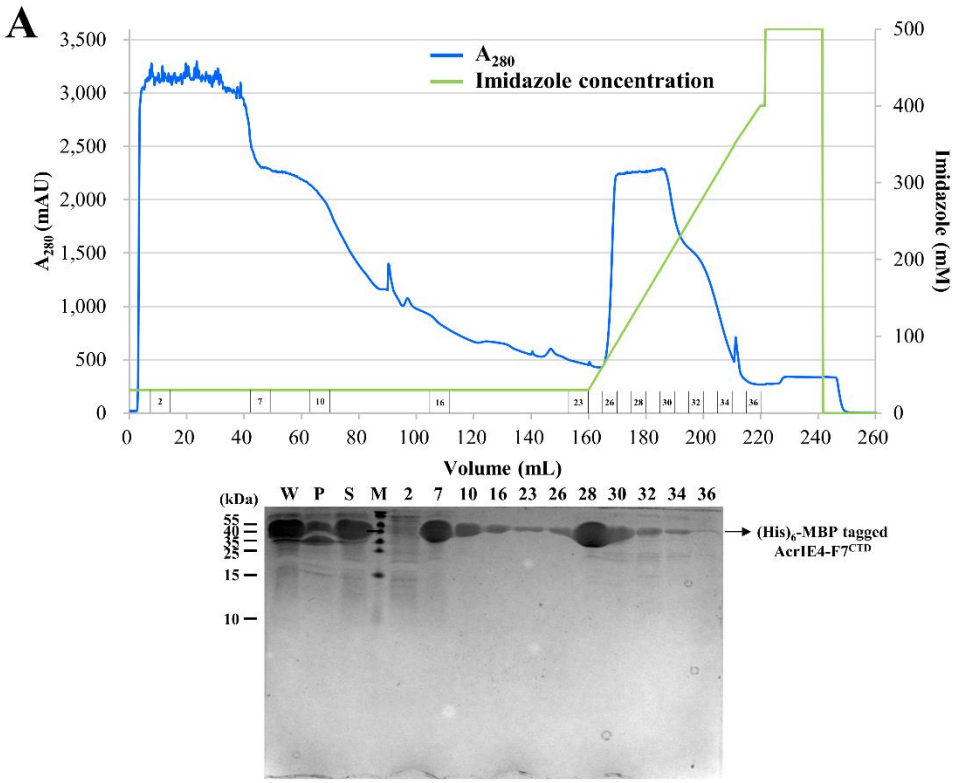
Purification of type I-E Cas proteins

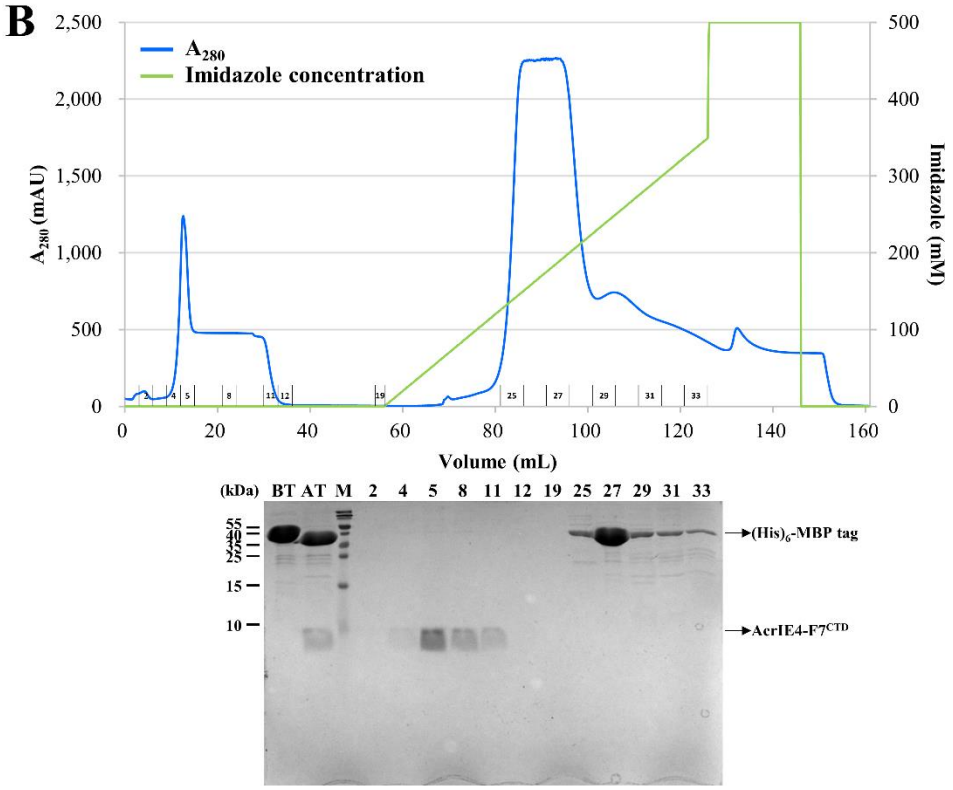
To purify type I-E Cas proteins of *P. aeruginosa* (Cas5e, Cas6e, Cas7e, Cas8e, and Cas11) which comprise Cascade complex, these proteins with N-terminal $(His)_6$-MBP tag were expressed in *E. coli* BL21(DE3) cells. Type I-E Cas proteins were purified in two-step including IMAC – SEC. Due to the poor expression and low solubility of individual type I-E Cas proteins, it has been difficult to identify the targets of type I-E Acr (Mejdani et al., 2021). I could express and purify these proteins by adding an N-terminal soluble $(His)_6$-MBP tag and not cleaving them (Fig. 7). During the IMAC, each Cas protein was eluted as the concentration of imidazole increased. All of these proteins were purified in high yields with a purity of more than 90%. The eluted fractions of Cas5e (25-32), Cas6e (23-31), Cas7e (22-29), Cas8e (19-26), and Cas11 (23-31) from the SEC were collected and the samples were concentrated.

To test whether AcrIE4-F7 could target the type I-E CRISPR-Cas system of *E. coli*, *E. coli* Cas8e with N-terminal $(His)_6$-MBP tag was expressed in *E. coli* BL21(DE3) cells. *E. coli* Cas8e was purified in the same method as *P. aeruginosa* Cas8e. During the IMAC, $(His)_6$-MBP tagged *E. coli* Cas8e was eluted between 58

Figure 6. Elution profiles of AcrIE4-F7^{CTD}

Elution profiles (*top*) and SDS-PAGE (*bottom*) analyses of AcrIE4-F7^{CTD} from each purification step. AcrIE4-F7^{CTD} with N-terminal $(His)_6$-MBP tag was purified by the same method of full-length AcrIE4-F7. (A) The first IMAC purification. (B) The second IMAC purification. (C) The SEC purification. The protein absorbance at 280 nm and concentration of imidazole are shown in *blue* and *green*, respectively. Eluted fractions were analyzed with 18% SDS-PAGE gel and visualized by Coomassie staining. Lane designation on the SDS-PAGE gel: W, whole cell lysate; P, pellet; S, supernatant; M, protein size marker; BT, sample before TEV proteolysis; AT, sample after TEV proteolysis; I, injection sample (AcrIE4-F7^{CTD}: 7 kDa, $(His)_6$-MBP tag: 43kDa).


A
A_{280}
Imidazole concentration
A_{280} (mAU)
Imidazole (mM)
Volume (mL)
(kDa)
W P S M 2 7 10 16 23 26 28 30 32 34 36
55
40
35
25
15
10
(His)$_6$-MBP tagged AcrIE4-F7^{CTD}


B
A_{280}
Imidazole concentration
A_{280} (mAU)
Imidazole (mM)
Volume (mL)
(kDa)
BT AT M 2 4 5 8 11 12 19 25 27 29 31 33
55
40
35
25
15
10
(His)$_6$-MBP tag
AcrIE4-F7^{CTD}

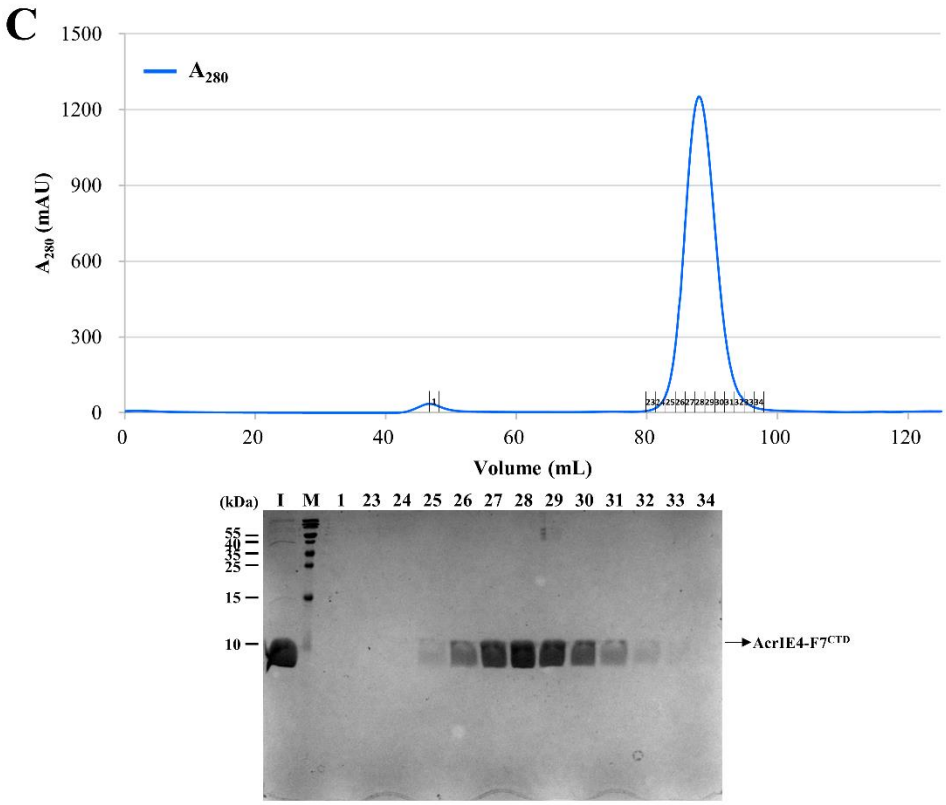
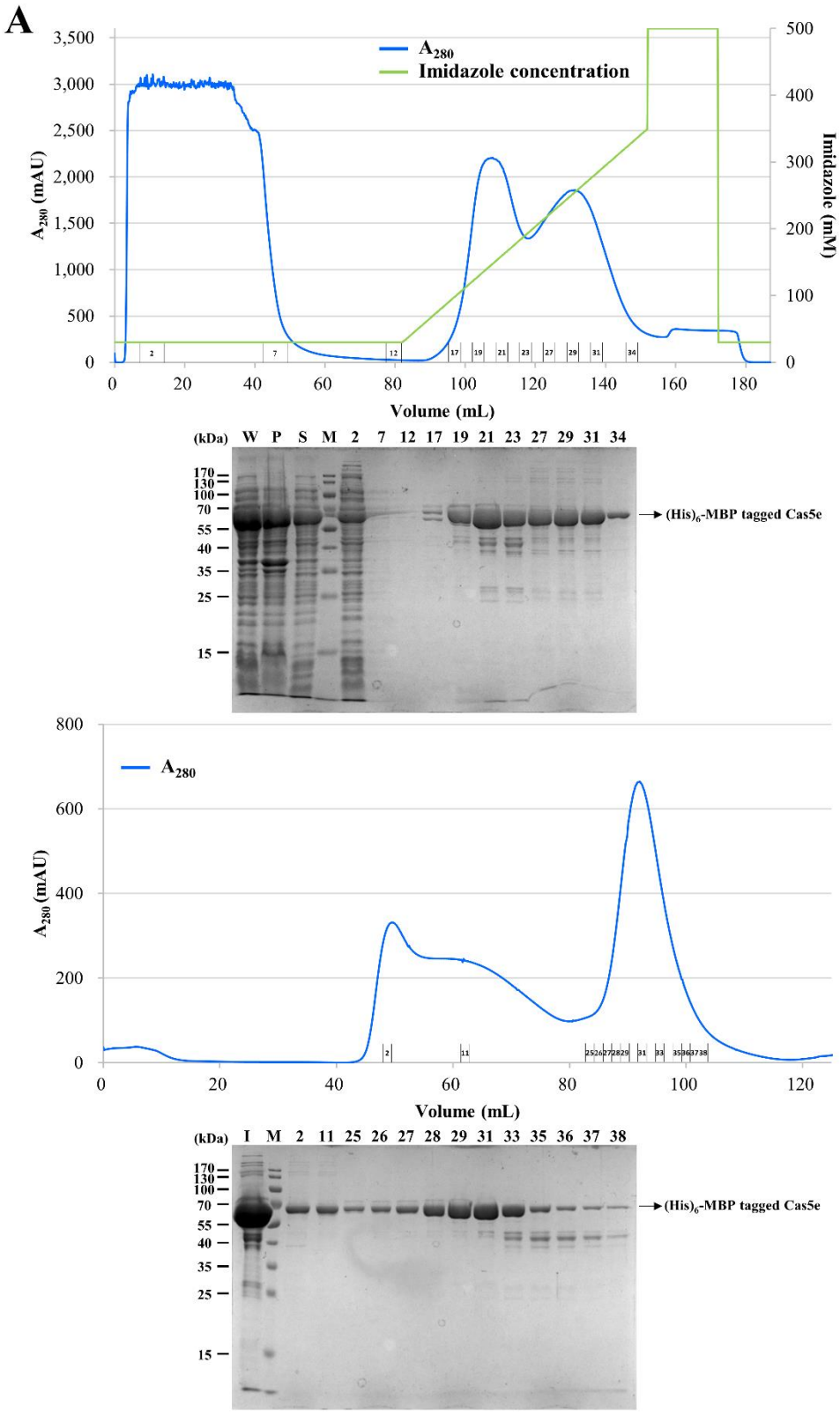
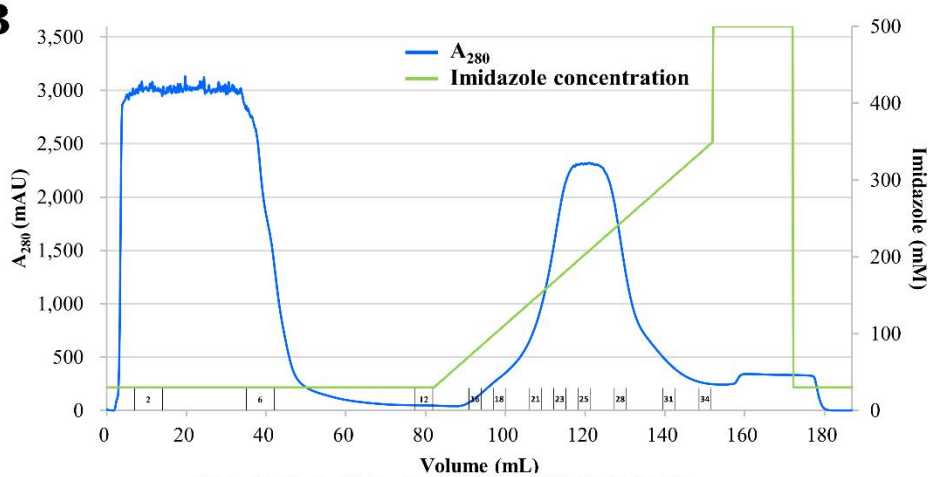

C
1500
1200
900
600
300
0
A280
A280 (mAU)
0
20
40
60
80
100
120
Volume (mL)
(kDa)
I
M
1
23
24
25
26
27
28
29
30
31
32
33
34
55
40
35
25
15
10
AcrIE4-F7CTD

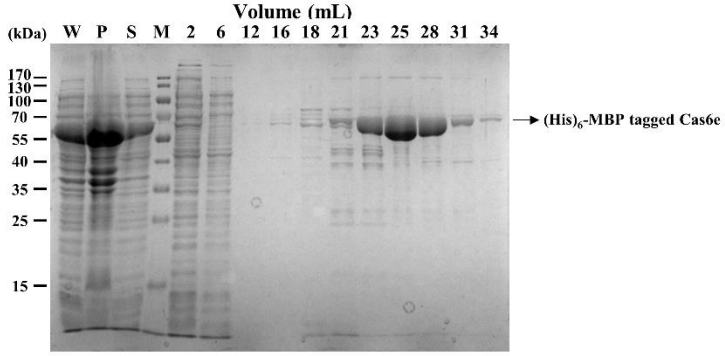
Figure 7. Elution profiles of *P. aeruginosa* type I-E Cas proteins

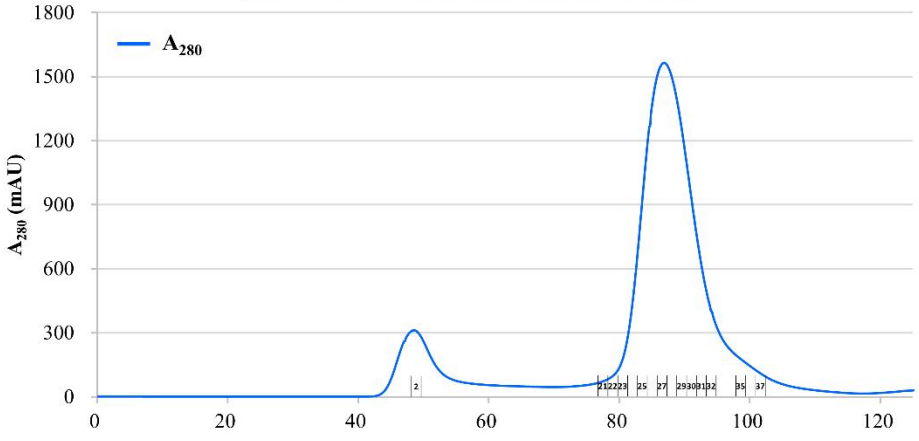
Elution profiles and SDS-PAGE analyses of *P. aeruginosa* type I-E Cas proteins that comprise the type I-E Cascade complex. Cas5e (A), Cas6e (B), Cas7e (C), Cas8e (D), and Cas11 (E) were purified by the two-step purification including IMAC – SEC. Type I-E Cas proteins were purified with $(His)_6$-MBP tag for their solubility. The protein absorbance at 280 nm and concentration of imidazole are shown in *blue* and *green*, respectively. Eluted fractions were analyzed with 10.5% or 12% SDS-PAGE gel and visualized by Coomassie staining. Lane designation on the SDS-PAGE gel: W, whole cell lysate; P, pellet; S, supernatant; M, protein size marker; I, injection sample ($(His)_6$-MBP tagged Cas5e: 67 kDa, $(His)_6$-MBP tagged Cas6e: 65 kDa, $(His)_6$-MBP tagged Cas7e: 84 kDa $(His)_6$-MBP tagged Cas8e: 100 kDa, $(His)_6$-MBP tagged Cas11: 62 kDa).

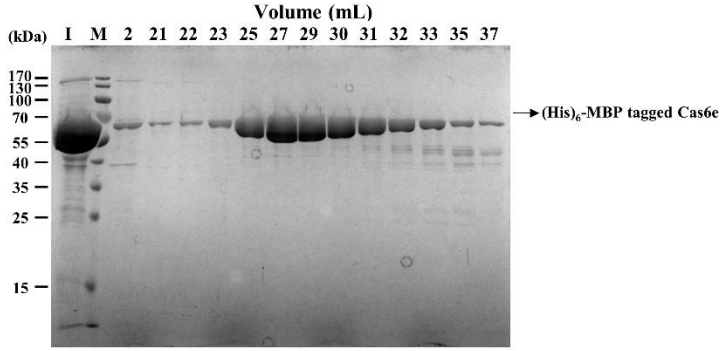

A
A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)
(kDa) W P S M 2 7 12 17 19 21 23 27 29 31 34
(His)6-MBP tagged Cas5e
A280
A280 (mAU)
Volume (mL)
(kDa) I M 2 11 25 26 27 28 29 31 33 35 36 37 38
(His)6-MBP tagged Cas5e

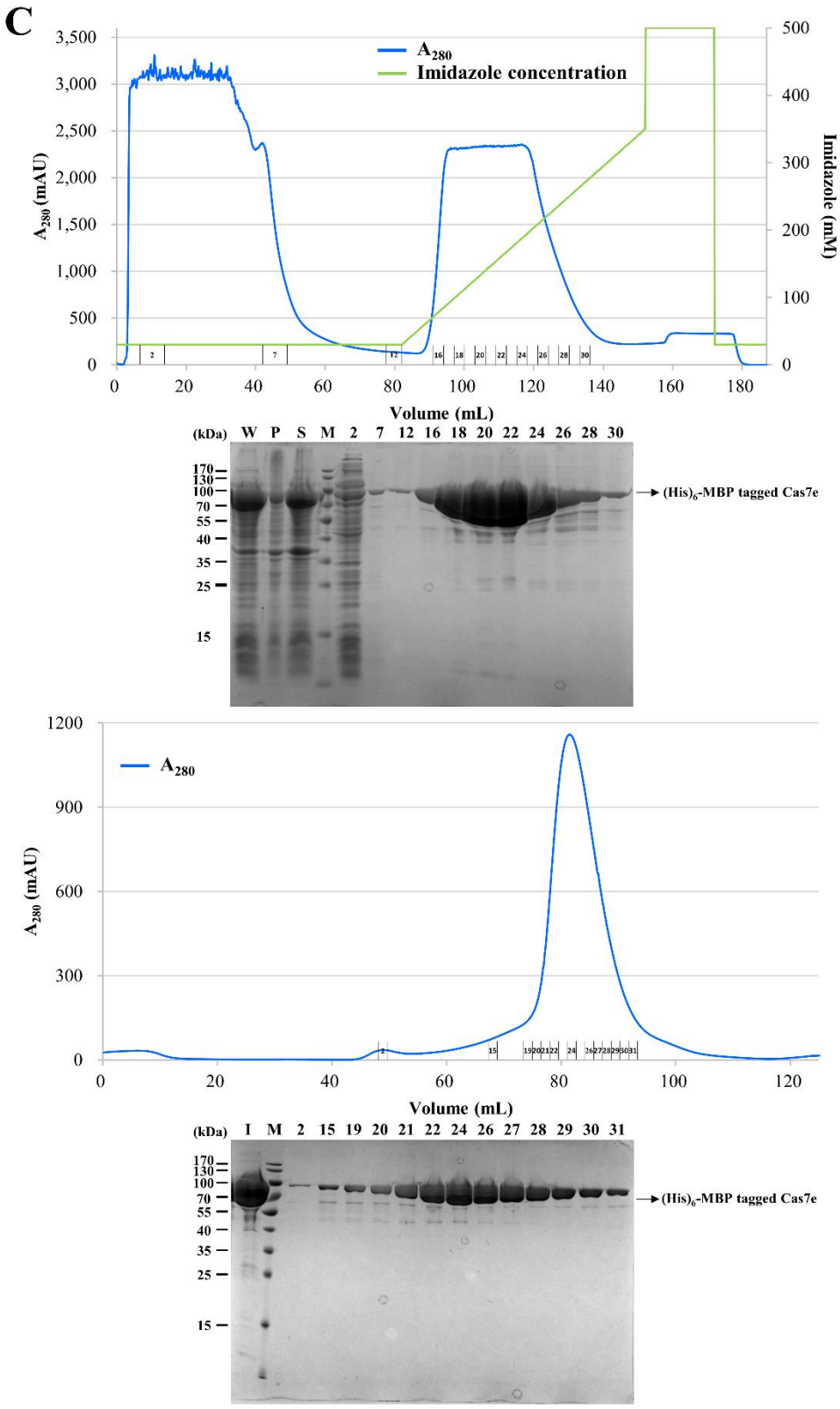
B


A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)

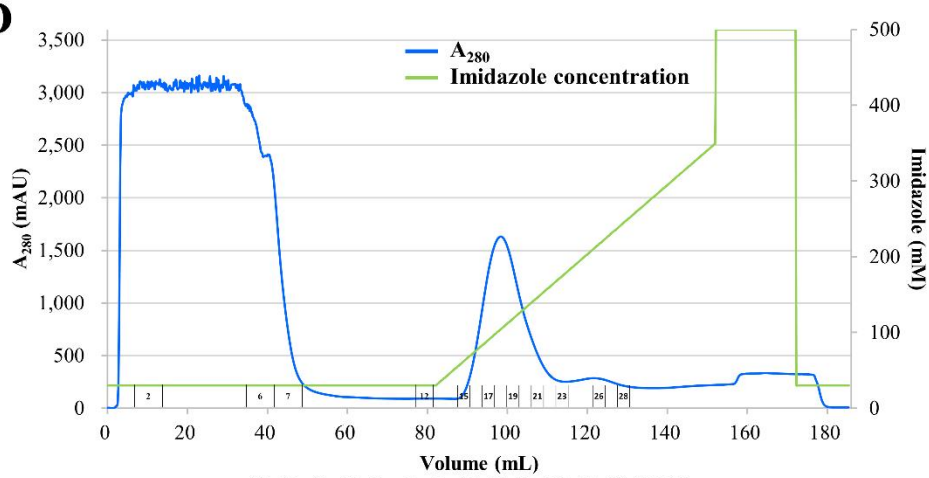

(kDa) W P S M 2 6 12 16 18 21 23 25 28 31 34
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas6e

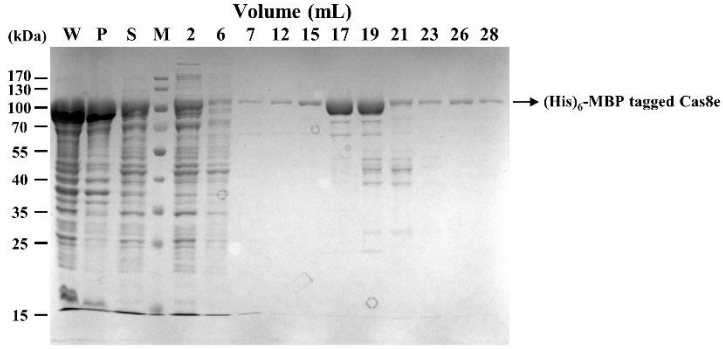

A280
A280 (mAU)
Volume (mL)

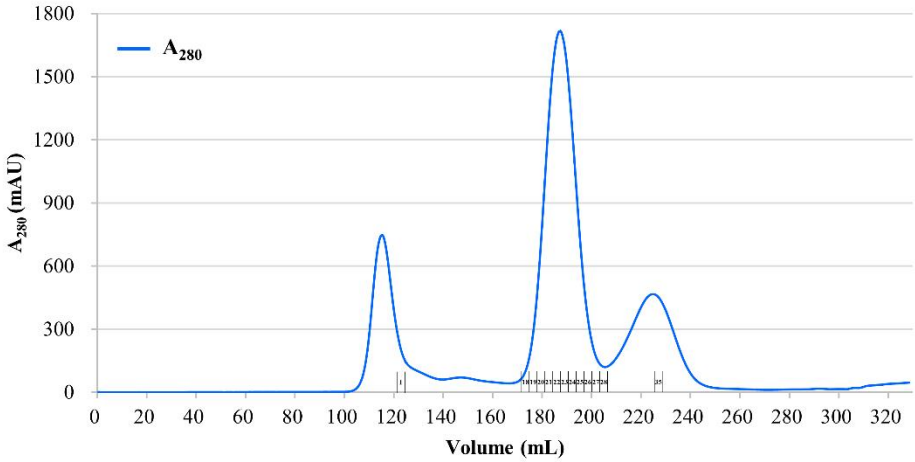

(kDa) I M 2 21 22 23 25 27 29 30 31 32 33 35 37
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas6e

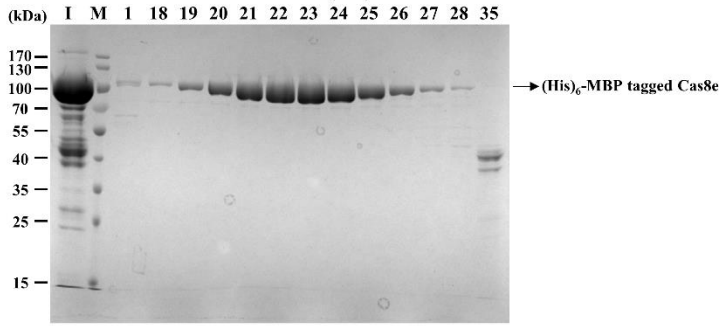

C
A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)
(kDa) W P S M 2 7 12 16 18 20 22 24 26 28 30
(His)6-MBP tagged Cas7e
A280
A280 (mAU)
Volume (mL)
(kDa) I M 2 15 19 20 21 22 24 26 27 28 29 30 31
(His)6-MBP tagged Cas7e

D

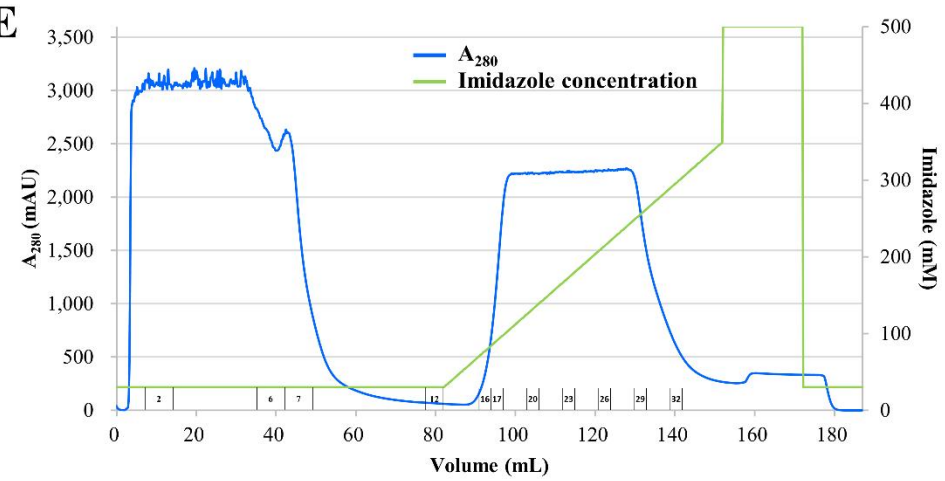

A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)

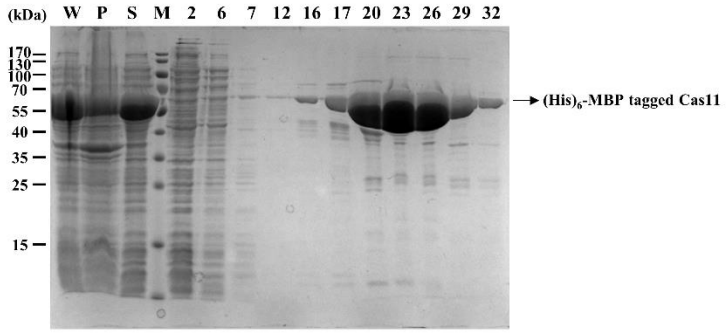

(kDa) W P S M 2 6 7 12 15 17 19 21 23 26 28
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas8e

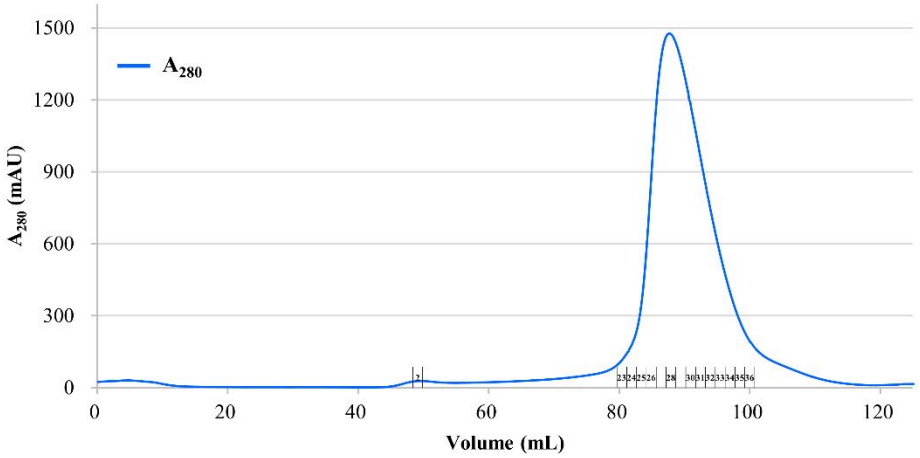

A280
A280 (mAU)
Volume (mL)

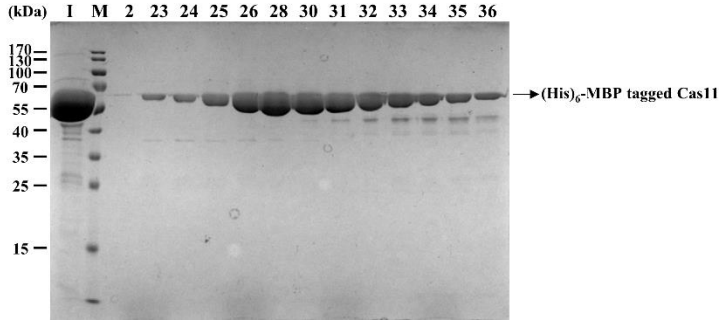

(kDa) I M 1 18 19 20 21 22 23 24 25 26 27 28 35
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas8e

E


A280
Imidazole concentration
A280 (mAU)
Imidazole (mM)
Volume (mL)
0
500
1,000
1,500
2,000
2,500
3,000
3,500
100
200
300
400
20
40
60
80
100
120
140
160
180
2
6
7
12
16
17
20
23
26
29
32


(kDa)
W P S M 2 6 7 12 16 17 20 23 26 29 32
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas11


A280
A280 (mAU)
Volume (mL)
0
300
600
900
1200
1500
20
40
60
80
100
120


(kDa)
I M 2 23 24 25 26 28 30 31 32 33 34 35 36
170
130
100
70
55
40
35
25
15
(His)6-MBP tagged Cas11

μM and 345 μM of imidazole (Fig. 8A). In the SEC, $(His)_6$-MBP tagged *E. coli* Cas8e (calculated MW: 99 kDa) was eluted at 75.4 mL (measured MW: 98.2 kDa) and it forms a monomer (Fig. 8B). I collected the number of fractions 17 to 23 from the SEC and the samples were concentrated.

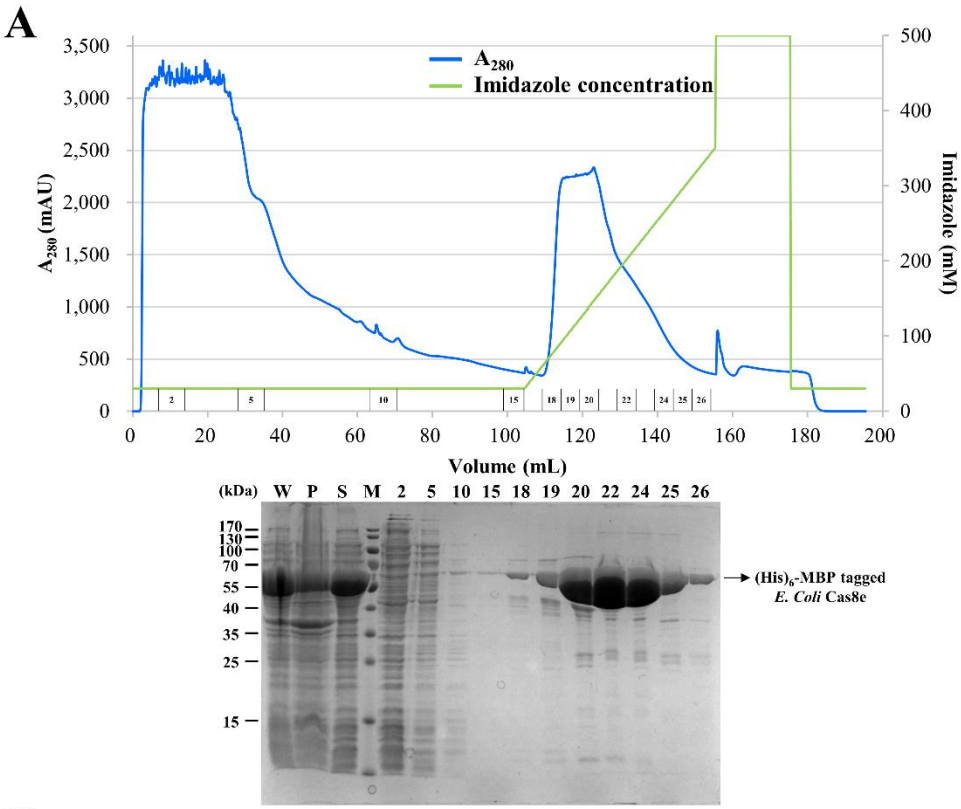
The N-terminal domain of AcrIE4-F7 binds to Cas8e

AcrIE4-F7 is a fusion protein of AcrIE4 and AcrIF7 which can inactivate Type I-E an I-F CRISPR-Cas system. The N-terminal domain of AcrIE4-F7 reveals a high sequence identity (69%) with AcrIE4. From the results of the phage plaque assay, it has been reported that AcrIE4 has a strong anti-CRISPR activity against the type I-E CRISPR-Cas system (Pawluk et al., 2014). However, the inhibitory mechanism of AcrIE4 is still unclear. Thus, I tried to figure out the target Cas proteins by analytical SEC to reveal the function of AcrIE4. Type I-E Cascade components are difficult to purify due to poor expression and solubility (Mejdani et al., 2021), but I could successfully express and purify those Cas proteins with the N-terminal $(His)_6$-MBP tag. The analytical SEC experiment was conducted between AcrIE4-F7 and N-terminal $(His)_6$-MBP tagged Cas proteins constituting the Cascade complex.

In the analytical SEC analyses, AcrIE4-F7 and $(His)_6$-MBP tagged Cas8e co-eluted at a smaller elution volume than those of each protein alone. This result demonstrates that AcrIE4-F7 interacts with Cas8e (Fig. 9). In an ITC experiment, the binding stoichiometry (N) and dissociation constant (K_D) of AcrIE4-F7 and $(His)_6$-MBP tagged Cas8e was determined as 1.0 ± 0.0 and 200 ± 28 nM (Hong et al., 2022), suggesting AcrIE4-F7 tightly binds to the Cas8e subunit with a submicromolar binding affinity. I also performed analytical SEC analyses between AcrIE4-F7 and other Cascade component (Cas5e, Cas6e, Cas7e, and Cas11), but

Figure 8. Elution profiles of *E. coli* Cas8e

Elution profiles and SDS-PAGE analyses of *E. coli* Cas8e. *E. coli* Cas8e was purified in the same method as *P. aeruginosa* Cas8e. (A) The IMAC purification. (B) The SEC purification. The protein absorbance at 280 nm and concentration of imidazole are shown in *blue* and *green*, respectively. Eluted fractions were analyzed with 12% SDS-PAGE gel and visualized by Coomassie staining. Lane designation on the SDS-PAGE gel: W, whole cell lysate; P, pellet; S, supernatant; M, protein size marker; I, injection sample (*E. coli* $(His)_6$-MBP tagged Cas8e: 99 kDa).


A
A_{280}
Imidazole concentration
A_{280} (mAU)
Imidazole (mM)
Volume (mL)
(kDa) W P S M 2 5 10 15 18 19 20 22 24 25 26
170
130
100
70
55
40
35
25
15
(His)$_6$-MBP tagged
E. Coli Cas8e

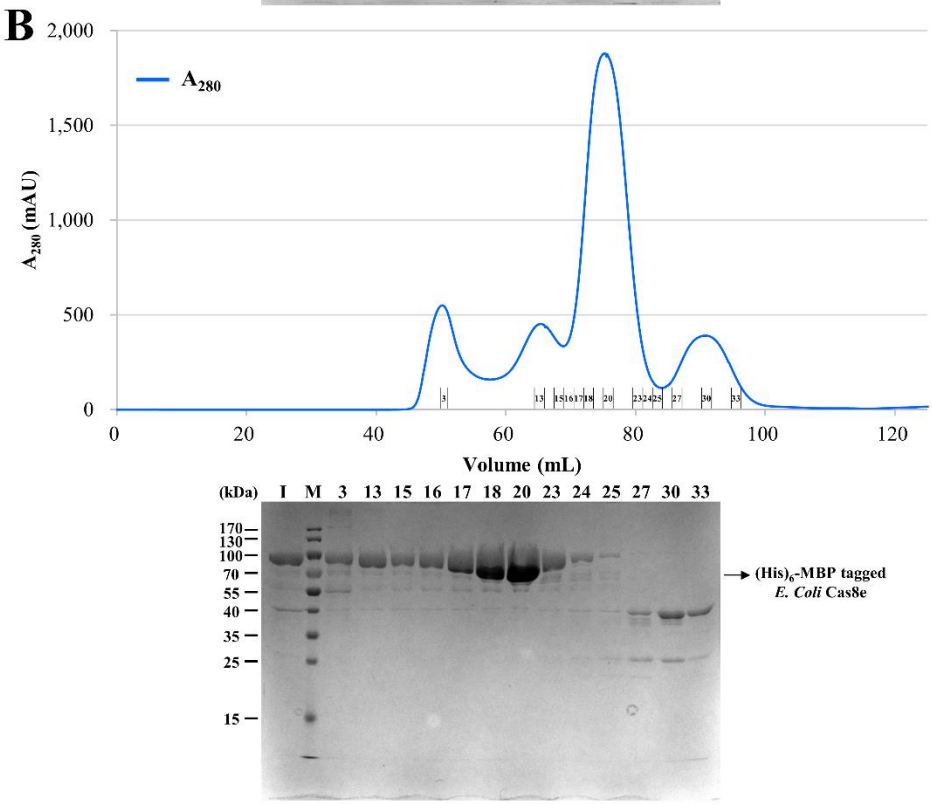
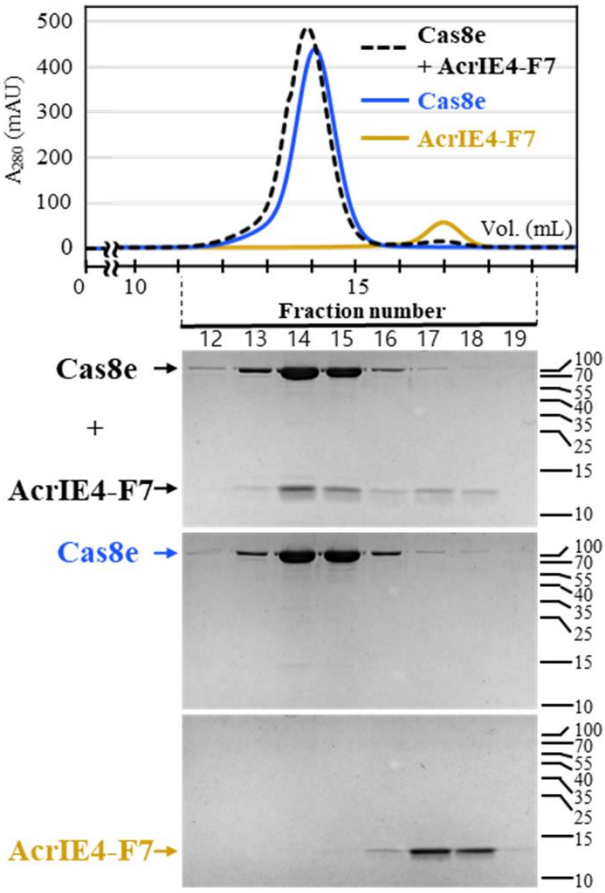

B
A_{280}
A_{280} (mAU)
Volume (mL)
(kDa) I M 3 13 15 16 17 18 20 23 24 25 27 30 33
170
130
100
70
55
40
35
25
15
(His)$_6$-MBP tagged
E. Coli Cas8e

Figure 9. AcrIE4-F7 binds to Cas8e.

Interaction between AcrIE4-F7 and Cas8e was determined by analytical SEC analyses. AcrIE4-F7 (20 μM) formed a stable complex with $(His)_6$-MBP tagged Cas8e (20 μM). Eluted fractions were analyzed by SDS-PAGE.


500
400
300
200
100
0
A_{280} (mAU)
Cas8e
+ AcrIE4-F7
Cas8e
AcrIE4-F7
Vol. (mL)
0
10
15
Fraction number
12 13 14 15 16 17 18 19
Cas8e
+
AcrIE4-F7
Cas8e
AcrIE4-F7
100
70
55
40
35
25
15
10

AcrIE4-F7 didn't interact with other Cas proteins (Fig. 10). In the next step, I analyzed the binding of N- and C-terminal domains of AcrIE4-F7 to Cas8e to determine which domain contributes to the Cas8e binding. In the analytical SEC, only AcrIE4-F7^{NTD} interacted with Cas8e, but not with AcrIE4-F7^{CTD} (Fig. 11). These results suggest that the N-terminal domain of AcrIE4-F7 is responsible for the interaction with Cas8e.

AcrIE4-F7^{NTD} targets the PAM recognition site of Cas8e.

From the analytical SEC and ITC experiments, I found that AcrIE4-F7 binds to the Cas8e subunit. In previous studies, some type I-F Acr proteins target the PAM recognition site of the Cas8f subunit (Jia and Patel, 2021). These proteins have low theoretical pI values and they compete with target DNA for the PAM interaction site of Cas8f (Yang et al., 2021). The positive-charged residues near the PAM recognition site of Cas8f are crucial to interact with the negative-charged residues of Acr (Kim et al., 2020). Since AcrIE4-F7^{NTD} has a low theoretical pI value (~4.2), I suspected that it may target the PAM recognition site of Cas8e. To prove this hypothesis, I tried to perform charge-reversal mutation for some residues near the PAM recognition site of Cas8e, but the structure of *P. aeruginosa* Cas8e was not identified. It was difficult to experimentally determine the structure of *P. aeruginosa* Cas8e because of the low expression and solubility. Therefore, I sought to model the *P. aeruginosa* Cas8e based on the Cas8e homolog.

The structure of *Thermobifida fusca* type I-E Cascade has been reported by cryogenic electron microscopy (cryo-EM) (Xiao et al., 2017). The sequence identity of *T. fusca* Cas8e and *P. aeruginosa* Cas8e is 24%, and these proteins recognize the

Figure 10. Analytical SEC analyses between AcrIE4-F7 and other type I-E Cas proteins (i.e, Cas5e, Cas6e, Cas7e, or Cas11).

Analytical SEC analyses were performed to examine whether AcrIE4-F7 binds to Cas5e (A), Cas6e (B), Cas7e (C), or Cas11 (D). AcrIE4-F7 did not interact with any of the other type I-E Cascade components. Eluted fractions were analyzed by SDS-PAGE.

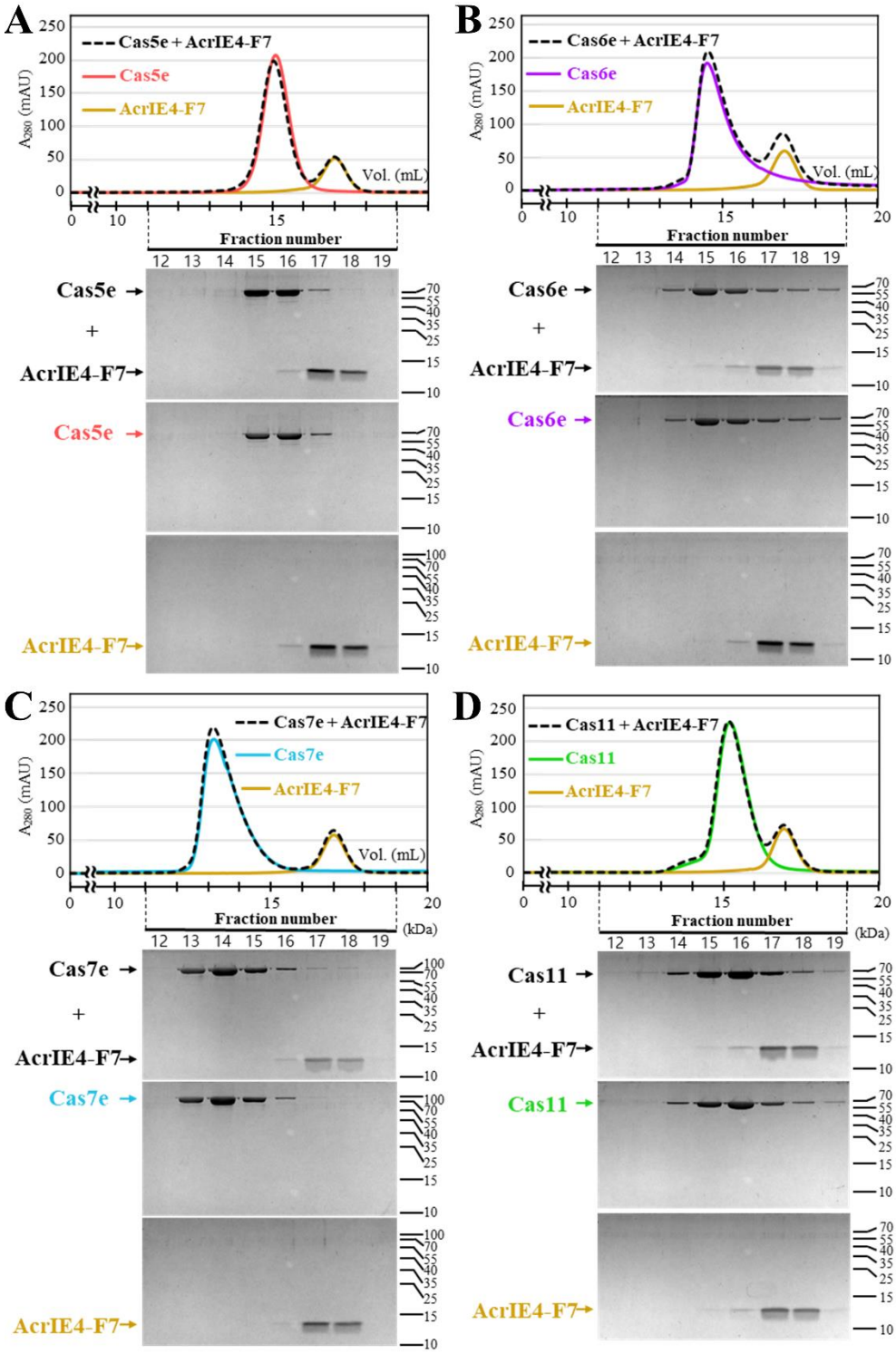
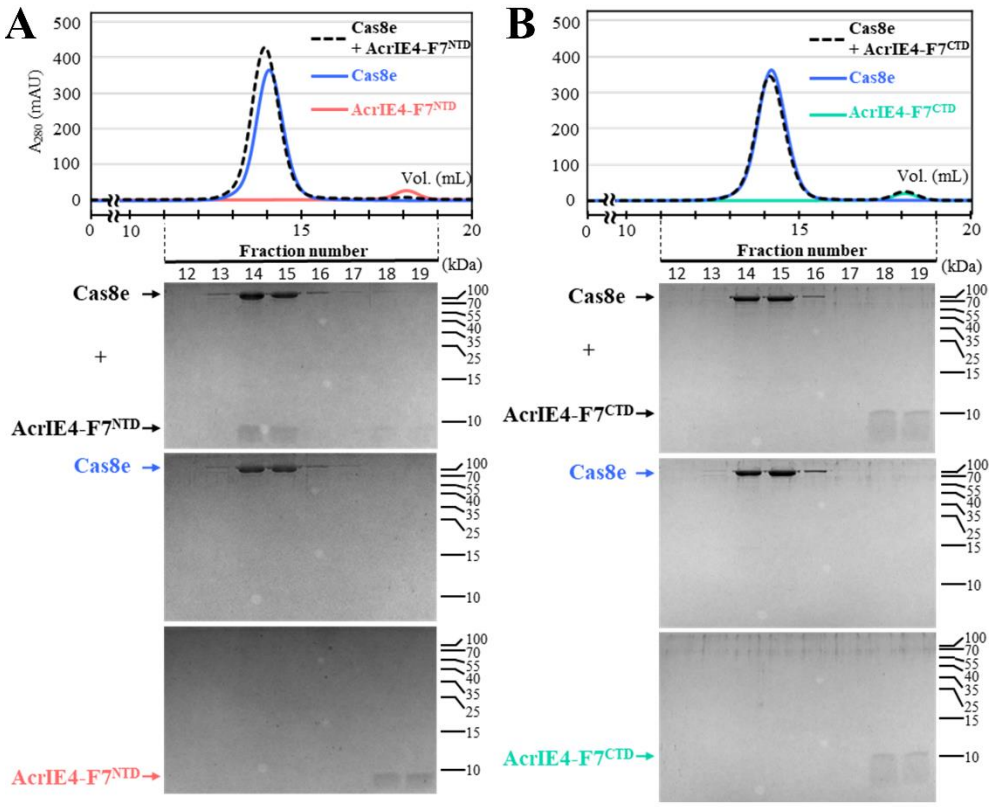

A
Cas5e + AcrIE4-F7
Cas5e
AcrIE4-F7
Vol. (mL)
Fraction number
Cas5e
+
AcrIE4-F7
B
Cas6e + AcrIE4-F7
Cas6e
AcrIE4-F7
Vol. (mL)
Fraction number
Cas6e
+
AcrIE4-F7
C
Cas7e + AcrIE4-F7
Cas7e
AcrIE4-F7
Vol. (mL)
Fraction number
(kDa)
Cas7e
+
AcrIE4-F7
D
Cas11 + AcrIE4-F7
Cas11
AcrIE4-F7
Vol. (mL)
Fraction number
(kDa)
Cas11
+
AcrIE4-F7

Figure 11. The N-terminal domain of AcrIE4-F7 is responsible for its interaction with Cas8e.

Analytical SEC analyses were performed to examine which domain of AcrIE4-F7 binds to Cas8e. Cas8e did not interact with the C-terminal domain of AcrIE4-F7 (*right*) but formed a stable complex with the N-terminal domain of AcrIE4-F7 (*left*). Eluted fractions were analyzed by SDS-PAGE.


A
B
Cas8e
+ AcrIE4-F7NTD
Cas8e
AcrIE4-F7NTD
Cas8e
+ AcrIE4-F7CTD
Cas8e
AcrIE4-F7CTD
A280 (mAU)
Vol. (mL)
Fraction number
(kDa)
Cas8e
+
AcrIE4-F7NTD
AcrIE4-F7CTD

same 5´-AAG PAM sequence, indicating that they share a conserved binding interface for PAM interaction (Pawluk et al., 2014; Xiao et al., 2017). I modeled the *P. aeruginosa* Cas8e structure (Fig. 12A) based on the *T. fusca* Cas8e structure (PDB: 5U07) using the Phyre[2] program (Kelley et al., 2015). Gly-rich loop, L1-loop, and Gln-wedge of *T. fusca* Cas8e are mediated for PAM recognition (Xiao et al., 2017). Since I guessed that AcrIE4-F7 may bind to Cas8e through charge interaction, I was looking for the positively charged residues near the Gly-rich loop, L1-loop, and Gln-wedge of *T. fusca* Cas8e. *T. fusca* Cas8e has Arg208 and Arg386 on the Gly-rich loop and the Gln-wedge, respectively. From the structural alignment of *T. fusca* and *P. aeruginosa* Cas8e, I identified three Lys residues on *P. aeruginosa* Cas8e are located within or adjacent to the Gly-rich loop (Lys176 and Lys183) or Gln-wedge (Lys357) (Fig. 12B). I made charge-reversed *P. aeruginosa* Cas8e mutants (K176E, K183E, K357E). After expressing and purifying each mutant, the CD spectra of Cas8e mutants were measured to confirm the change in secondary structure that might be induced by the mutation. Three mutants showed similar CD spectra to that of WT Cas8e, suggesting that charge reversal mutations in the putative PAM recognition site did not cause Cas8e misfolding (Fig. 13). In the analytical SEC analyses between AcrIE4-F7 and Cas8e mutants, AcrIE4-F7 did not co-eluted with any of Cas8e mutants, demonstrating that K176, K183, and K357 are essential for binding with AcrIE4-F7 (Fig. 14). These results reveal that AcrIE4-F7^{NTD} targets the PAM recognition site of Cas8e to suppress type I-E CRISPR-Cas system.

In the previous study, AcrIE4 potently inhibit the type I-E CRISPR-Cas system of *P. aeruginosa*, but not that of *E. coli* (Pawluk et al., 2014). To determine the interaction between AcrIE4-F7 and *E. coli* Cas8e, I performed analytical SEC as the

Figure 12. A model structure of *P. aeruginosa* Cas8e

(A) The structural comparison of the *T. fusca* Cas8e (PDB: 5U07) and *P. aeruginosa* Cas8e model. (B) *P. aeruginosa* Cas8e was modeled on the cryo-EM structure of the *T. fusca* Cascade complex (PDB: 5U07) using the Phyre[2] program. K176, K183, and K357 in the putative PAM recognition site of Cas8e are represented in *orange*, *green*, and *cyan*, respectively.

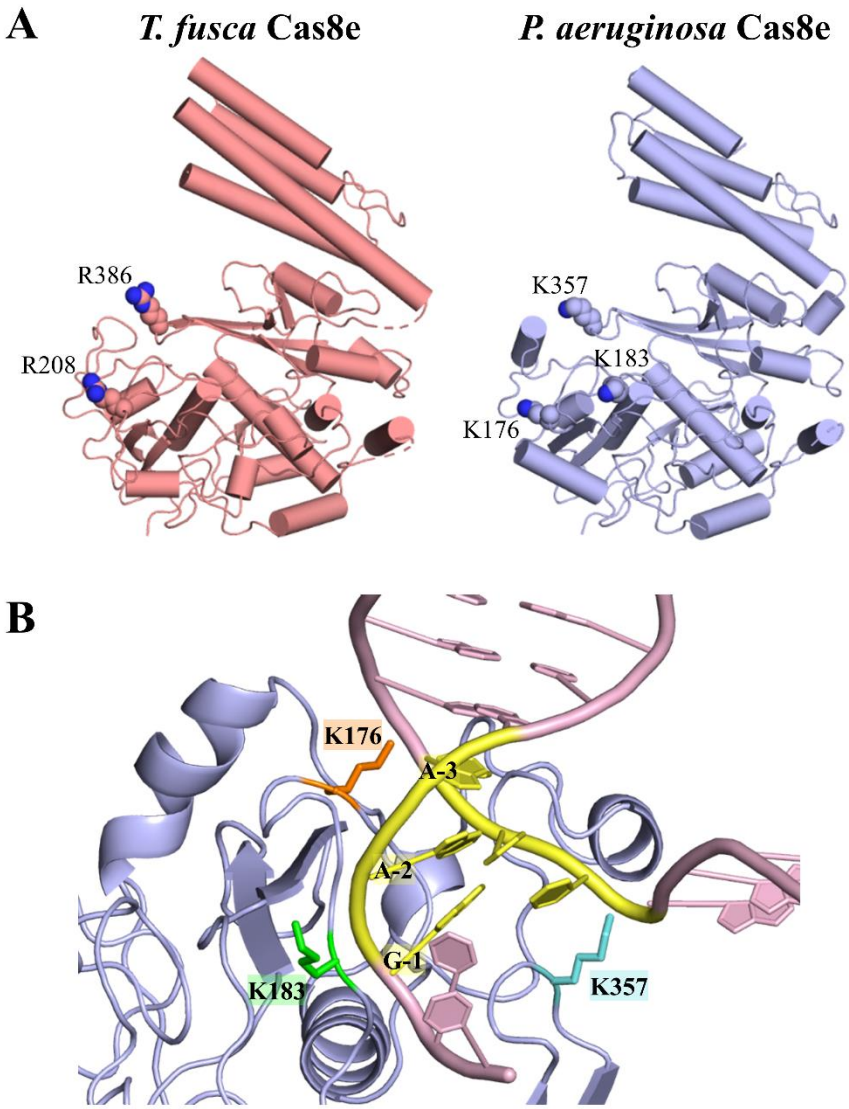

A
T. fusca Cas8e
P. aeruginosa Cas8e
R386
R208
K357
K183
K176
B
K176
A-3
A-2
G-1
K183
K357

Figure 13. CD spectra of WT Cas8e and its mutants

Cas8e mutants (K176E, K183E, and K357E) displayed similar CD similar spectra to that of WT Cas8e, indicating that charge reversal mutations in the putative PAM recognition site did not cause misfolding of Cas8e.

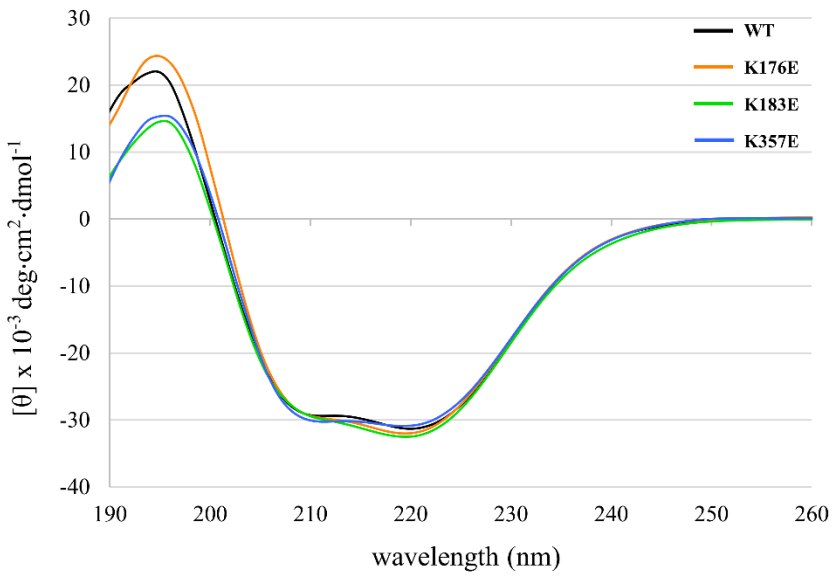
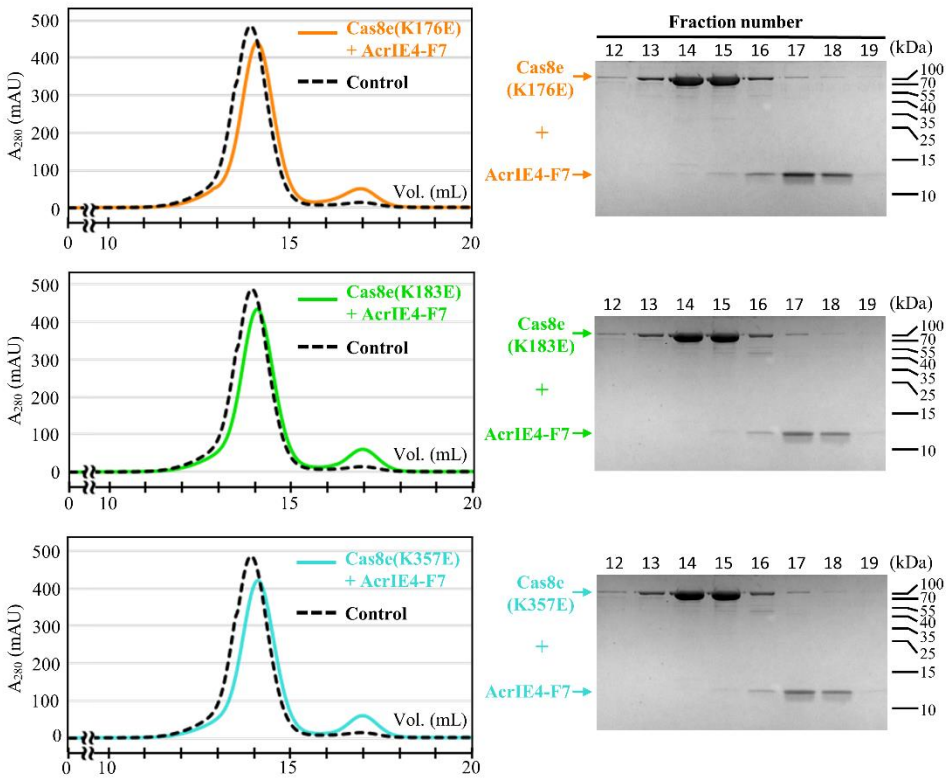

WT
K176E
K183E
K357E
30
20
10
0
-10
-20
-30
-40
190
200
210
220
230
240
250
260
[θ] x 10-3 deg·cm2·dmol-1
wavelength (nm)

Figure 14. AcrIE4-F7 targets the putative PAM recognition site of Cas8e.

Analytical SEC analyses between AcrIE4-F7 and Cas8e mutants. AcrIE4-F7 did not co-elute with Cas8e mutants (K176E, K183E, and K357E), demonstrating that the interaction with AcrIE4-F7 was hindered by the charge reversal mutations in the putative PAM recognition site of Cas8e. The dashed lines indicate the SEC chromatography for the binding to WT Cas8e.


Cas8e(K176E)
+ AcrIE4-F7
Control
Vol. (mL)
A280 (mAU)
Fraction number
12 13 14 15 16 17 18 19 (kDa)
Cas8e
(K176E)
+
AcrIE4-F7
Cas8e(K183E)
+ AcrIE4-F7
Control
Vol. (mL)
A280 (mAU)
Cas8e
(K183E)
+
AcrIE4-F7
Cas8e(K357E)
+ AcrIE4-F7
Control
Vol. (mL)
A280 (mAU)
Cas8e
(K357E)
+
AcrIE4-F7
100
70
55
40
35
25
15
10

same analytical SEC method between AcrIE4-F7 and *P. aeruginosa* Cas8e. In the analytical SEC analyses, AcrIE4-F7 didn't interact with *E. coli* Cas8e (Fig. 15), which is consistent with the previous report. Sequence identity scores from pairwise sequence alignment vary from 7% to 34%, suggesting that individual Cas components of *P. aeruginosa* and *E. coli* are distantly related (Fig. 16). Cas8e revealed the lowest sequence identity score (7%) and *P. aeruginosa* Cas8e and *E. coli* Cas8e recognize 5´-AAG and 5´-ATG PAM sequence, respectively. Therefore, the divergent PAM interaction surfaces of *P. aeruginosa* and *E. coli* Cas8e subunits may cause the specific anti-CRISPR activity of AcrIE4.

Identification of AcrIE4-F7^{NTD} binding interface for Cas8e binding.

From collaboration with Dr. Sung-Hyun Hong and Prof. Jeong-Yong Suh at Seoul National University, the NMR structure of AcrIE4-F7 was determined (PDB: 7VZM) (Hong et al., 2022). To identify the binding interface of AcrIE4-F7^{NTD} to Cas8e, mutation analyses of AcrIE4-F7^{NTD} were performed. Since AcrIE4-F7 interacted with positively-charged residues in Cas8e, negative-charged residues of AcrIE4-F7^{NTD} are important for Cas8e binding. Mutation sites in the AcrIE4-F7^{NTD} were selected based on the surface charge distribution of AcrIE4-F7^{NTD} (Fig. 17). The surface charge distribution of AcrIE4-F7^{NTD} was calculated using Adaptive Poisson-Boltzmann Solver (APBS) Plugin in PyMOL, and *red* and *blue* colors represent the negatively and positively charged surfaces, respectively. In addition, the Tyr20 was chosen for mutagenesis because it was directed outward rather than toward a hydrophobic core and might contribute to DNA mimic. Hence, six AcrIE4-F7 mutants (E12K/D13K, E19K/D22K, Y20A, D30K/E31K, E38K/D39K, and E46K) were generated. The interactions between Cas8e and AcrIE4-F7 mutants were

Figure 15. AcrIE4-F7 does not interact with *E. coli* Cas8e.

Analytical SEC analyses were performed to examine whether AcrIE4-F7 binds to *E. coli* Cas8e. AcrIE4-F7 did not co-eluted with *E. coli* Cas8e. Eluted fractions were analyzed by SDS-PAGE.

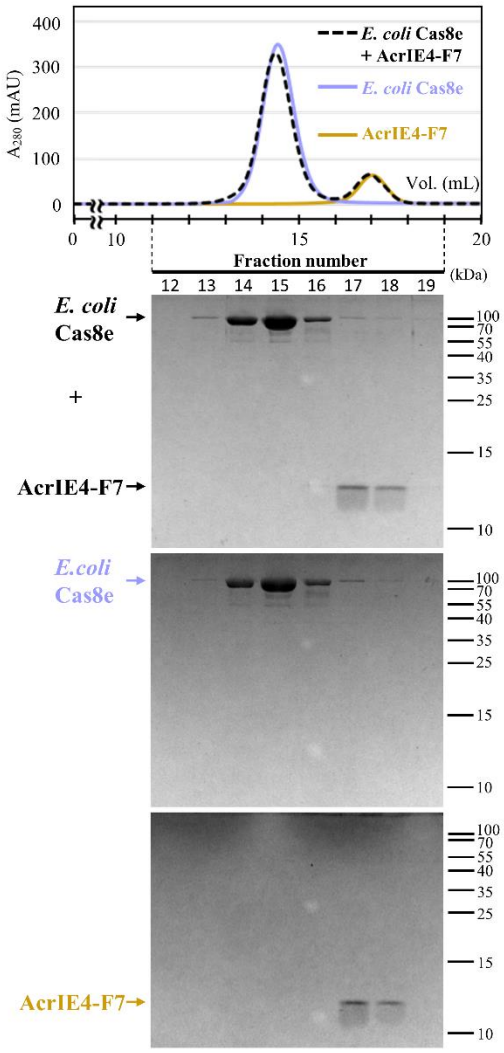

400
300
200
100
0
A_{280} (mAU)
E. coli Cas8e
+ AcrIE4-F7
E. coli Cas8e
AcrIE4-F7
Vol. (mL)
0
10
15
20
Fraction number
12 13 14 15 16 17 18 19
(kDa)
E. coli
Cas8e
+
AcrIE4-F7
E.coli
Cas8e
AcrIE4-F7
100
70
55
40
35
25
15
10

Figure 16. Sequence identity of *P. aeruginosa* and *E. coli* type I-E Cascade components.

The sequence identities of *P. aeruginosa* and *E. coli* type I-E Cascade components are calculated from the pairwise sequence alignment. Type I-E CRISPR-Cas system of *P. aeruginosa* and *E. coli* are distantly related. The sequence identity scores between them vary from 7% to 34%.

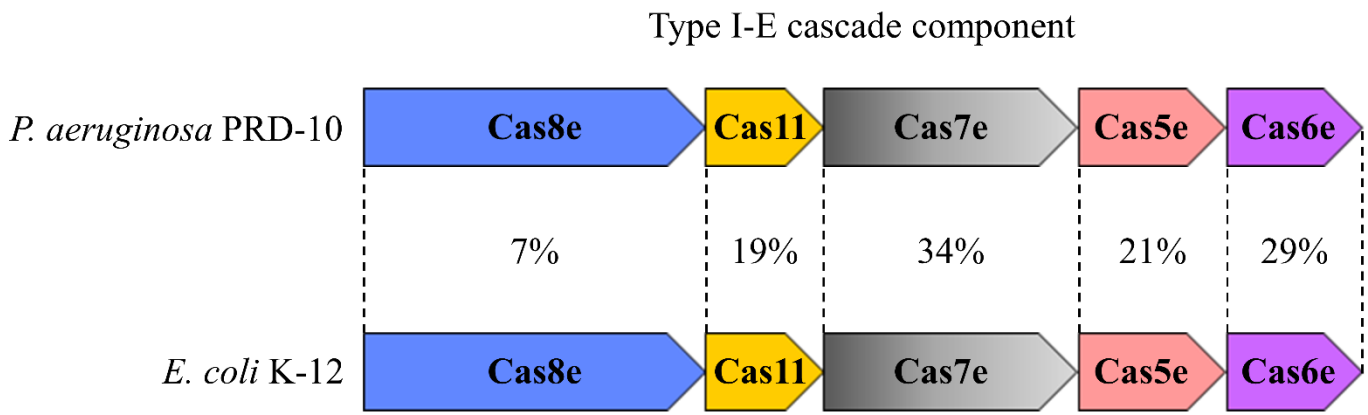
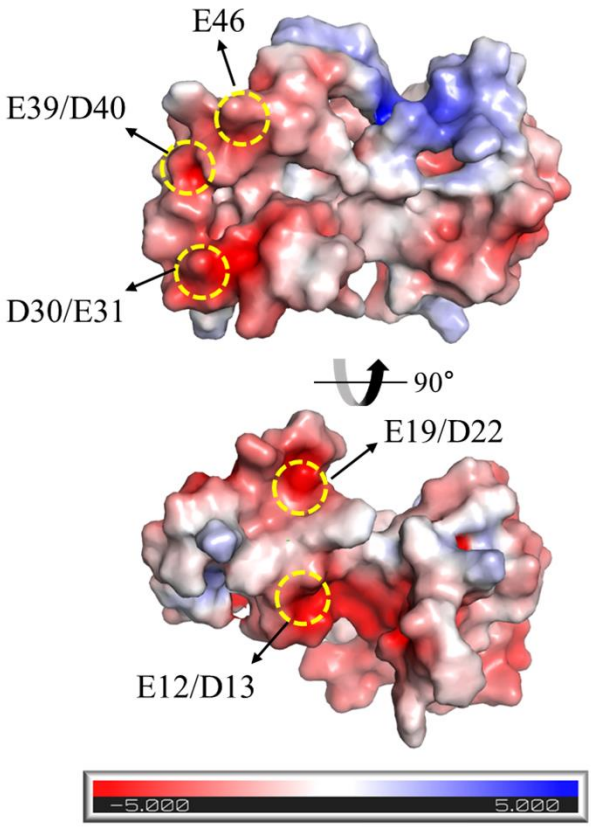

Type I-E cascade component
P. aeruginosa PRD-10
Cas8e
Cas11
Cas7e
Cas5e
Cas6e
7%
19%
34%
21%
29%
E. coli K-12
Cas8e
Cas11
Cas7e
Cas5e
Cas6e

Figure 17. The surface charge distribution of AcrIE4-F7^{NTD}

The NMR structure of AcrIE4-F7 (PDB: 7VZM) was identified through collaboration with Dr. Sung-Hyun Hong and Prof. Jeong-Yong Suh at Seoul National University. The surface charge distribution of AcrIE4-F7^{NTD} was calculated by the Adaptive Poisson-Boltzmann Solver (APBS) Plugin in PyMOL, and *red* and *blue* colors indicate the negatively and positively charged surfaces, respectively. Negatively charged residues for mutagenesis are shown with dotted *yellow* circles.


E46
E39/D40
D30/E31
90°
E19/D22
E12/D13
-5.000
5.000

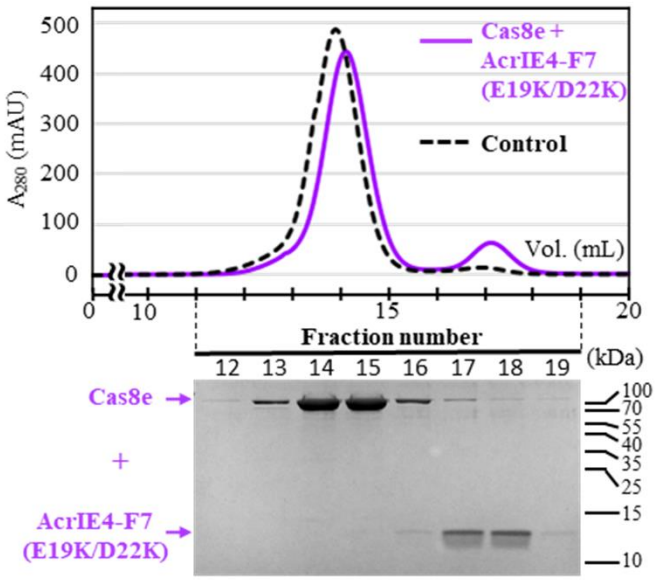
determined using SEC and ITC experiments. In the SEC and ITC experiment, Cas8e did not interact with the E19K/D22K mutant of AcrIE4-F7, demonstrating mutation in the α2 helix (E19K/D22K) disrupted the binding to Cas8e (Fig. 18) (Hong et al., 2022). In the next step, two single mutants (E19K and D22K) were generated to investigate which residue contributes more to the interaction. In the ITC experiment, these mutants did not interact with Cas8e, indicating that both residues were essential for the Cas8e binding (Hong et al., 2022). Mutation in the α3 helix (D30K/E31K) had a remarkable effect on Cas8e binding, resulting in a 24-fold reduction in binding affinity (Hong et al., 2022). However, mutations in the α1 helix (E12K/D13K) and α4 helix (E38K/E39K and E46K) had little or no effect on binding affinity. While the binding affinity of E12K/D13K and E46K mutants was reduced by ~2-fold, that of the E38K/D39K mutant did not reduce at all (Hong et al., 2022). Lastly, the Y20A mutant showed a ~18-fold reduction in Cas8e binding (Hong et al., 2022). These results suggest that the α2 helix of AcrIE4-F7 works as the main binding interface for Cas8e. ITC data and the production of AcrIE4-F7 mutants were provided by Mr. Changkon Park, Dr. Sung-Hyun Hong, and Prof. Jeong-Yong Suh at Seoul National University.

AcrIE4-F7^{CTD} binds to Cas8f:Cas5f heterodimer

The C-terminal domain of AcrIE4-F7 shows a high sequence similarity (55%) with AcrIF7. AcrIF7 has a strong anti-CRISPR activity against type I-F CRISPR-Cas system (Pawluk et al., 2016) and its structure and inhibition mechanism were reported (Kim et al., 2020; Gabel et al., 2021). In the previous study, AcrIF7 tightly binds to the Cas8f:Cas5f heterodimer which comprises the PAM recognition tail of the type I-F Cascade complex (Kim et al., 2020). To test whether AcrIE4-F7 interacts

Figure 18. Analytical SEC analyses between Cas8e and the E19K/D22K mutant of AcrIE4-F7.

Analytical SEC analyses were performed to test whether Cas8e binds to the E19K/D22K mutant of AcrIE4-F7. The mutant did not interact with Cas8e, indicating the E19K/D22K mutations disrupted Cas8e binding. The dashed lines represent the SEC chromatography for the binding to WT Cas8e. Eluted fractions were analyzed by SDS-PAGE.


500
400
300
200
100
0
A_{280} (mAU)
Cas8e +
AcrIE4-F7
(E19K/D22K)
Control
Vol. (mL)
0
10
15
20
Fraction number
12 13 14 15 16 17 18 19
(kDa)
Cas8e
+
AcrIE4-F7
(E19K/D22K)
100
70
55
40
35
25
15
10

with Cas8f:Cas5f heterodimer in a similar manner to AcrIF7, I performed analytical SEC analyses between AcrIE4-F7 and Cas8f:Cas5f heterodimer. In the analytical SEC analyses, AcrIE4-F7 and Cas8f:Cas5f heterodimer co-eluted at a smaller elution volume than those of each protein alone. This result highlights that AcrIE4-F7 interacts with Cas8f:Cas5f heterodimer (Fig. 19). In an ITC experiment, the dissociation constant (K_D) of AcrIE4-F7 and $(His)_6$-MBP tagged Cas8e was determined as ~26 nM (Hong et al., 2022), which is equivalent to the K_D of ~46 nM between AcrIF7 and Cas8f:Cas5f. In the next step, I analyzed the binding of the N- and C-terminal domains of AcrIE4-F7 to Cas8f:Cas5f to determine which domain contributes to the binding to Cas8f:Cas5f. In the analytical SEC, AcrIE4-F7^{CTD} solely interacted with Cas8e, but not with AcrIE4-F7^{NTD} (Fig. 20). These results suggest that the C-terminal domain of AcrIE4-F7 is responsible for the interaction with Cas8f:Cas5f heterodimer.

The binding interface between AcrIE4-F7^{CTD} and Cas8f

The structure of *P. aeruginosa* type I-F Cascade:AcrIF7 (PDB: 7JZX) was determined by cryo-EM (Gabel et al., 2021). AcrIF7 mimics the surface potential of the PAM sequence to compete with target DNA for Cas8f binding (Kim et al., 2020). According to the pairwise sequence alignment of AcrIE4-F7^{CTD}, the key residues for Cas8f binding (Asp44, Asp49, and Glu50 in the AcrIF7) were conserved in AcrIE4-F7^{CTD} (Asp65, Asp80, and Glu86) (Fig. 21). Also, I compared the structure of AcrIE4-F7^{CTD} with the AcrIF7 whose structure was previously reported by NMR (PDB: 6M3N) and cryo-EM (PDB: 7JZX). From the structural alignment of AcrIF7 and AcrIE4-F7^{CTD}, each protein is superposed well except β1- β2 loop region

Figure 19. AcrIE4-F7 binds to the Cas8f:Cas5f heterodimer.

Interaction between AcrIE4-F7 and Cas8f:Cas5f was determined by analytical SEC analyses. AcrIE4-F7 (20 μM) formed a stable complex with Cas8f:Cas5f heterodimer (20 μM). Eluted fractions were analyzed by SDS-PAGE.

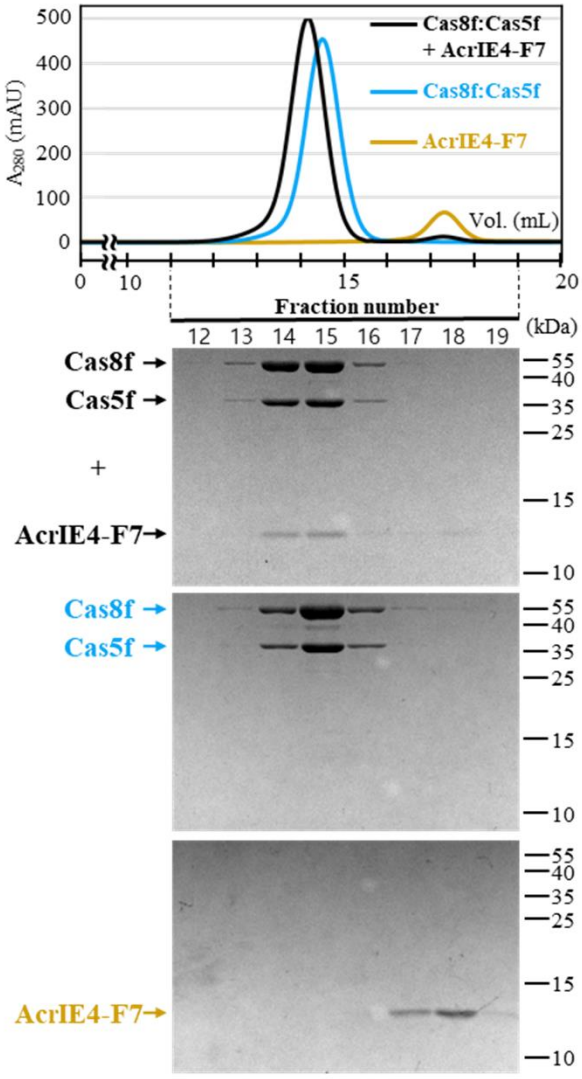

Cas8f:Cas5f
+ AcrIE4-F7
Cas8f:Cas5f
AcrIE4-F7
Vol. (mL)
A_{280} (mAU)
500
400
300
200
100
0
0
10
15
20
Fraction number
12 13 14 15 16 17 18 19
(kDa)
Cas8f
Cas5f
+
AcrIE4-F7
55
40
35
25
15
10
Cas8f
Cas5f
55
40
35
25
15
10
AcrIE4-F7
55
40
35
25
15
10

Figure 20. Cas8f:Cas5f heterodimer interacts with the C-terminal domain of AcrIE4-F7, not with its C-terminal domain.

Analytical SEC analyses were performed to examine which domain of AcrIE4-F7 binds to Cas8f:Cas5f heterodimer. Cas8f:Cas5f did not interact with the N-terminal domain of AcrIE4-F7 (*left*) but formed a stable complex with the C-terminal domain of AcrIE4-F7 (*right*). Eluted fractions were analyzed by SDS-PAGE.

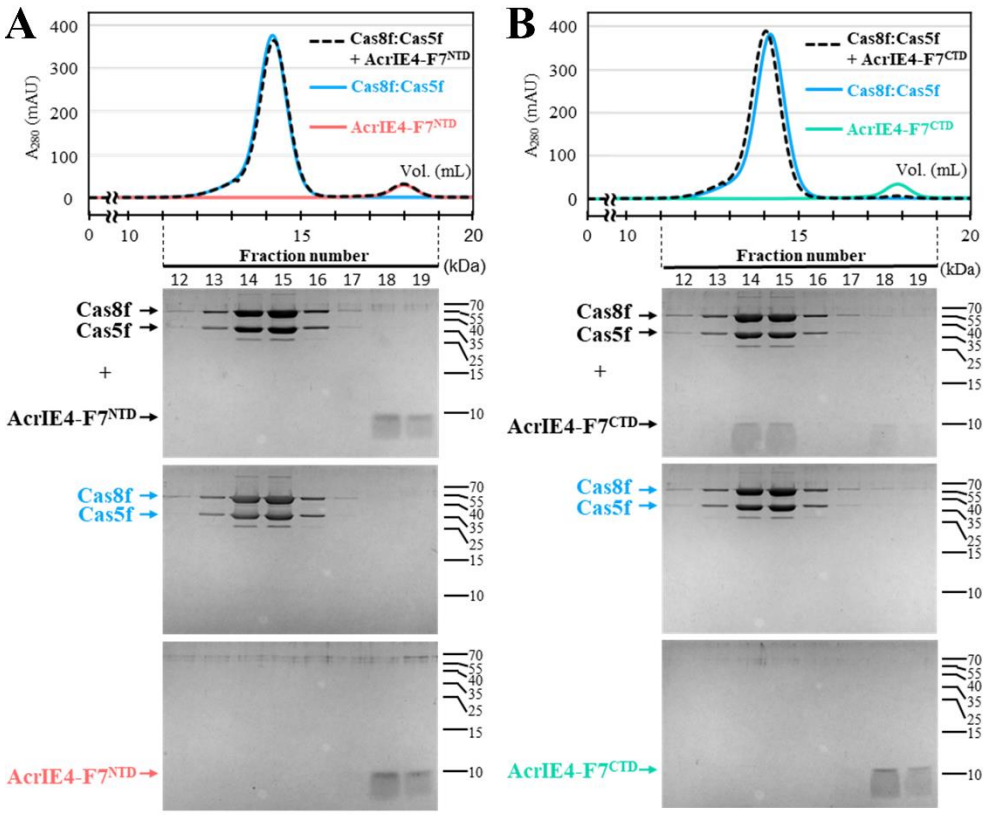

A
B
Cas8f:Cas5f
+ AcrIE4-F7NTD
Cas8f:Cas5f
AcrIE4-F7NTD
Cas8f:Cas5f
+ AcrIE4-F7CTD
Cas8f:Cas5f
AcrIE4-F7CTD
A280 (mAU)
Vol. (mL)
Fraction number
(kDa)
Cas8f
Cas5f
+
AcrIE4-F7NTD
AcrIE4-F7CTD

Figure 21. Pairwise sequence alignment of AcrIF7 and AcrIE4-F7^{CTD}

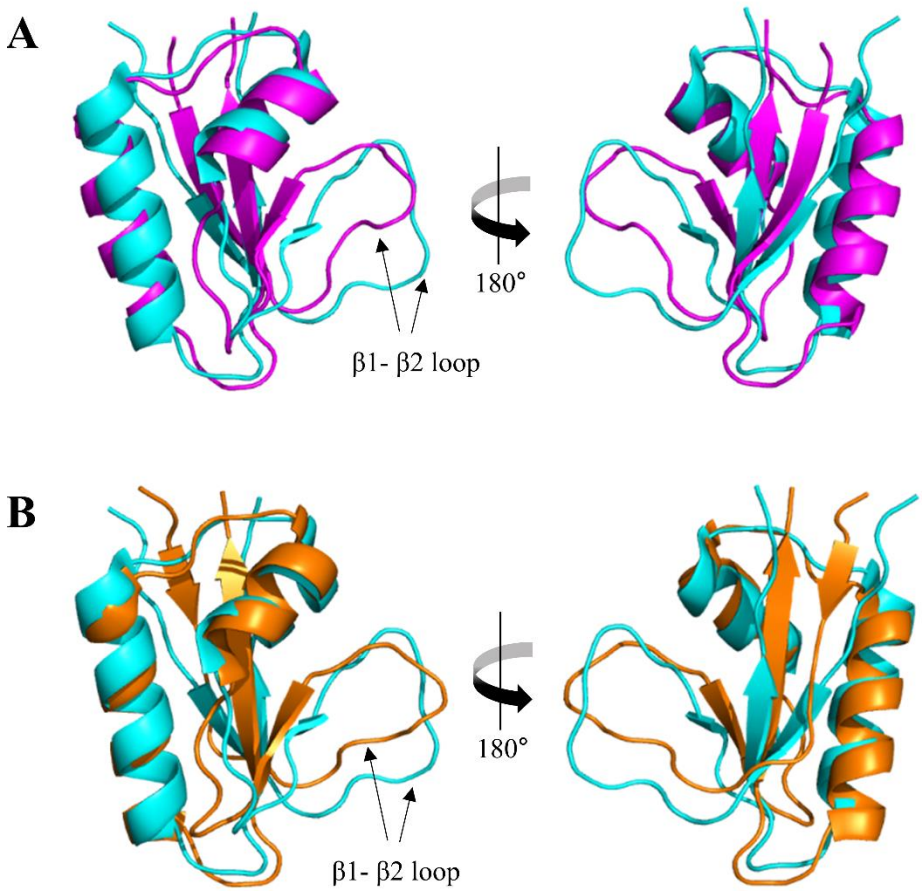
Sequences of AcrIF7 and AcrIE4-F7^{CTD} were aligned using the Clustal Omega program. The key residues of AcrIF7 for Cas8f binding were conserved in AcrIE4-F7^{CTD} and shown in red boxes.

```
                                  60        70        80        90        100       110
AcrIE4-F7/53-119   53 SPKFSTKVTTKPDFDGFQFYIEAGRDFDGDAYTEAYGVAVPTNIAARIQAQAAELNAGEWLLVEHEA 119
AcrIF7/1-67         1 MTTFTSIVTTNPDFGGFEFYVEAGQQFDDSAYEEAYGVSVPSAVVEEMNAKAAQLKDGEWLNVSHEA  67
```

(Fig. 22). The root mean square deviation (RMSD) values of alpha carbon between AcrIE4-F7^{CTD} and AcrIF7 were calculated at 1.4 Å (with cryo-EM structure) and 1.7 Å (with NMR structure), respectively. These results demonstrate AcrIE4-F7^{CTD}, which is likely a structural and functional homolog of AcrIF7, has a similar binding interface as AcrIF7 for Cas8f binding.

Figure 22. Structural alignment of AcrIF7 and AcrIE4-F7^{CTD}

The NMR structure of AcrIE4-F7 (PDB: 7VZM) was determined through collaboration with Dr. Sung-Hyun Hong and Prof. Jeong-Yong Suh at Seoul National University. (A) Structural alignment of AcrIE4-F7^{CTD} (*cyan*) and cryo-EM structure of AcrIF7 complexed with the type I-F Cascade complex (PDB: 7JZX; *magenta*). (B) Structural alignment of AcrIE4-F7^{CTD} and AcrIF7 (PDB: 6M3N; *orange*). The RMSD values of alpha carbon between AcrIE4-F7^{CTD} and AcrIF7 were calculated at 1.4 Å (cryo-EM structure) and 1.7 Å (NMR structure), respectively.


A
180°
β1- β2 loop
B
180°
β1- β2 loop

Discussion

In this study, biochemical experiments were conducted to understand the inhibition mechanism of AcrIE4-F7. Through experiments by the analytical SEC and ITC, I revealed the AcrIE4-F7 targets the PAM recognition site of Cas8 in the type I-E and I-F CRISPR-Cas system. During the acquisition and interference step, recognition of the PAM sequence from the target DNA is a crucial process of distinguishing the target DNA. To escape the host CRISPR-Cas system, it is frequently found that phages mutated the PAM sequences or near the PAM site (Sun et al., 2013), indicating mutations of this region provide phages a strong selective advantage. According to the studies of type I-F anti-CRISPR proteins, the cryo-EM structures of type I-F Cascade:Acr complex revealed that AcrIF2, AcrIF6, AcrIF7, and AcrIF10, which have no structural and sequence similarity between them, target the PAM recognition site by interacting with positively charged DNA vise of Cas8f (Guo et al., 2017; Zhang et al., 2020; Gabel et al., 2021). However, in the type I-E CRISPR-Cas system, only two Acr proteins (AcrIE1 and AcrIE2) have reported inhibitory mechanisms. From the *in-vivo* experiments, it was found that AcrIE1 and AcrIE2 target Cas3 and Cascade, respectively (Pawluk et al., 2017; Mejdani et al., 2021). Thus, the inhibition mechanism of AcrIE4-F7^{NTD} was first reported in the type I-E CRISPR-Cas system, highlighting that targeting the PAM recognition site is a common inhibition mechanism.

In a previous phage plaque assay result, the anti-CRISPR activity of AcrIE4-F7 was estimated using *P. aeruginosa* SMC4386 (type I-E) and *P. aeruginosa* PA14 (type I-F) (Marino et al., 2018). However, *P. aeruginosa* PRD-10 Cas proteins and

X. albilineans Cas8f:Cas5f were used for the binding analyses with AcrIE4-F7 instead of the *P. aeruginosa* homologs. I wanted to acquire *P. aeruginosa* SMC4386 cell or genomic DNA but I couldn't. So, I performed the binding analyses of AcrIE4-F7^{NTD} with *P. aeruginosa* PRD-10 which have a high sequence identity with *P. aeruginosa* SMC4386. Sequence identities of type I-E Cascade components from *P. aeruginosa* PRD-10 and *P. aeruginosa* SMC4386 are more than 90% and the average sequence identity is 98% (Fig. 23). For the binding analyses of AcrIE4-F7^{CTD}, I used *X. albilineans* Cas8f:Cas5f heterodimer. Since the inhibition mechanism of AcrIF7 was characterized using *X. albilineans* Cas8f:Cas5f heterodimer (Kim et al., 2020), I wanted to compare binding affinity between AcrIE4-F7 and Cas8f:Cas5f heterodimer with that of AcrIF7 and Cas8f:Cas5f heterodimer. *X. albilineans* Cas8f:Cas5f heterodimer has been reported to bind with type I-F Acr proteins such as AcrIF2 and AcrIF7 (Hong et al., 2018; Kim et al., 2020). The sequence identity of *P. aeruginosa* and *X. albilineans* Cas8f is 38.9% and AcrIF7-interacting residues are completely conserved. Arg24, Lys28, Arg58, Lys71, and Arg78 of *P. aeruginosa* Cas8f which are AcrIF7-interacting residues through charged interaction (Gabel et al., 2021) are conserved in Arg25, Lys29, Arg62, Lys75, and Arg82 of *X. albilineans* Cas8f (Fig. 24). Asn111 and Asn250 of *P. aeruginosa* Cas8f which were reported to recognize the PAM sequence (Gabel et al., 2021) are conserved in Asn115 and Asn249 of *X. albilineans* Cas8f (Fig. 24).

Previously, the structures of the type I-E Cascade complex in *T. fusca* and *E. coli* were determined (Hayes et al., 2016; Xiao et al., 2017). To recognize the PAM sequence of target DNA, the type I-E Cascade of *E. coli* has a Lys-finger, Gly-loop, and Gln-wedge (Hayes et al., 2016), while the type I-E Cascade of *T. fusca* harbors a Gly-rich loop, L1-loop, and Gln-wedge (Xiao et al., 2017). Arg208 of *T. fusca*

Figure 23. Sequence identity of *P. aeruginosa* SMC4386 and *P. aeruginosa* PRD-10 type I-E Cascade components.

Type I-E Cascade components of *P. aeruginosa* PRD-10 share a high sequence identity with that of *P. aeruginosa* SMC4386. All of the type I-E Cascade components have more than 90% of sequence identity and the average sequence identity is 98%, indicating they compose almost the same type I-E Cascade.

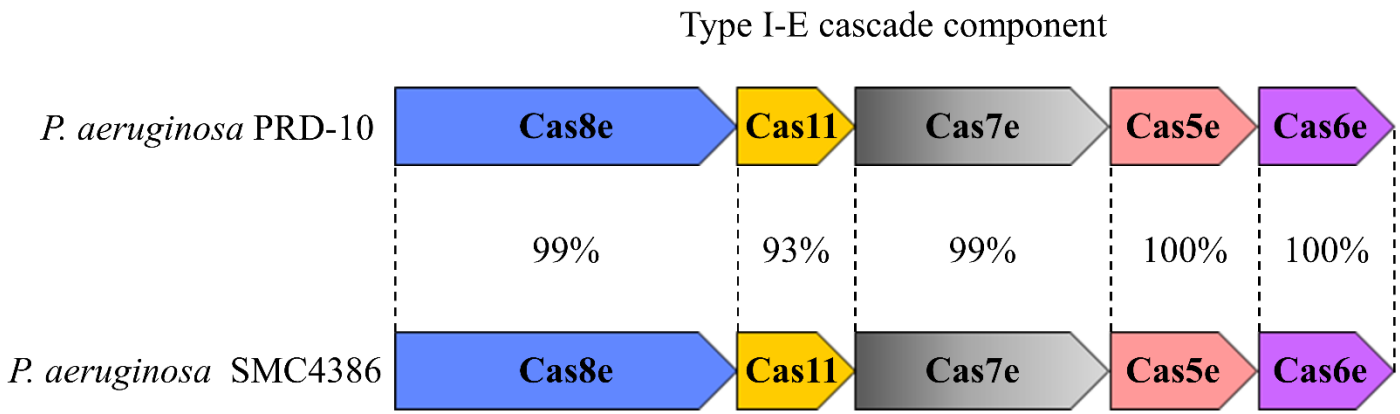

Type I-E cascade component
P. aeruginosa PRD-10
Cas8e
Cas11
Cas7e
Cas5e
Cas6e
99%
93%
99%
100%
100%
P. aeruginosa SMC4386
Cas8e
Cas11
Cas7e
Cas5e
Cas6e

Figure 24. Pairwise sequence alignment of *X. albilineans* and *P. aeruginosa* Cas8e

Sequence identity score of *X. albilineans* and *P. aeruginosa* Cas8e is 38.9%. The key residues of *P. aeruginosa* Cas8e for AcrIF7 binding were conserved in *X. albilineans* Cas8e. The residues for charge interaction and PAM recognition are shown in *red* and *orange* boxes, respectively.

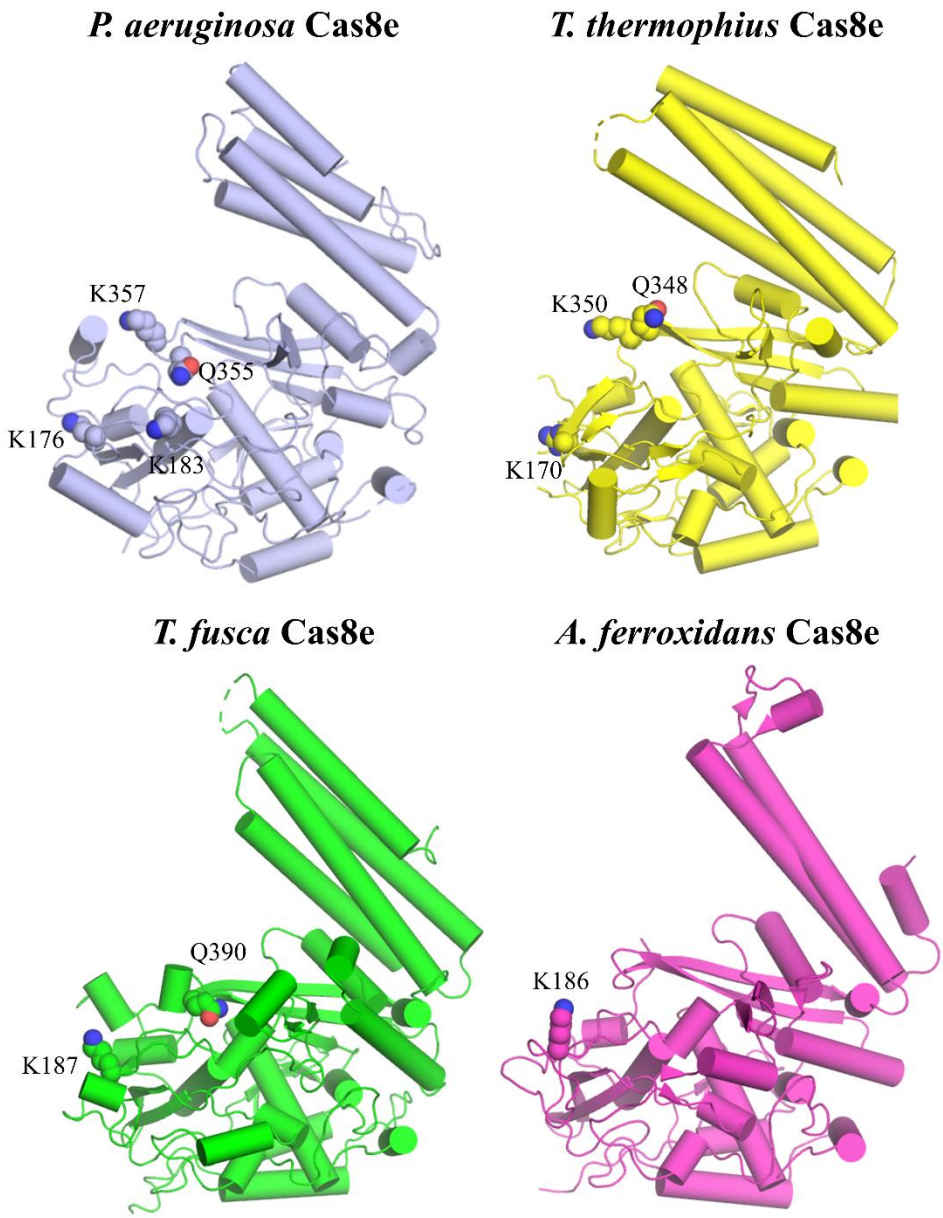
```
X.albilineans/1-455     1 MTEDPQHLTRSERFRSAIAAFIDARREAKLKGNEGDSDT-----D  40
P.aeruginosa/1-434      1 -MTSPLPTPTWQELRQFIESFIQERLQGKLDKLQPDEDDKRQTLL  44

X.albilineans/1-455    41 SKYDYATWLSDAARRVSQIQAVTHVLKATHPDARGSSLHVAPTRL  85
P.aeruginosa/1-434     45 ATHRREAWLADAARRVGQLQLVTHTLKPIHPDARGSNLHSLPQAP  89

X.albilineans/1-455    86 QAHTEVGTHVLGTDYAEDVVGNAAALDVFKLLKLEVDGRRLLDWM 130
P.aeruginosa/1-434     90 GQPGLAGSHELGDRLVSDVVGNAAALDVFKFLSLQYQGKNLLNWL 134

X.albilineans/1-455   131 QDDDADLRVALHDNAKVASDWMGAFCNLVRHDALPSSHQAAKQVY 175
P.aeruginosa/1-434    135 TEDSAEALQALSDNAEQAREWRQAFIGITTVKGAPASHSLAKQLY 179

X.albilineans/1-455   176 WLVGDEPREDTHYHLLQPLFSSSLAHAVHADIQDARFGERNKQAR 220
P.aeruginosa/1-434    180 FPL-----PGSGYHLLAPLFPTSLVHHVHALLREARFGDAAKAAR 219

X.albilineans/1-455   221 QAYRDKHPFDGTYHDYRNLVARKLGGTKPQNISQLNSERGGINYL 265
P.aeruginosa/1-434    220 EARSRQESWPHGFSEYPNLAIQKFGGTKPQNISQLNNERRGENWL 264

X.albilineans/1-455   266 LASLPPRWTQEERPRTLLKLDSALDRFAYFDEVPTLIKKLADFLA 310
P.aeruginosa/1-434    265 LPSLPPNWQRQNVNAPMRHSSVFEHDFGRTPEVSRLTRTLQRFLA 309

X.albilineans/1-455   311 TDPPKNDATQKKREGMENDLINKLALFAAETAVRFEPGWSRNPDC 355
P.aeruginosa/1-434    310 KTVHNNLAIRQRRAQLVAQICDE-ALQYAARLRELEPGWSATPGC 353

X.albilineans/1-455   356 VLPLSQQLWLDAERVELPIRIDPEHAEWEQQDRNFVDTYHLGDWQ 400
P.aeruginosa/1-434    354 QLHDAEQLWLDPLRA--------------QTDETFLQRRLRGDWP 384

X.albilineans/1-455   401 DEIAGQFATWLNDRLRAAGITGLGDDQYRHWAKQAIVDAAWPVPM 445
P.aeruginosa/1-434    385 AEVGNRFANWLNRAVSSDS-QILGSPEAAQWSQELSKELTMFKEI 428

X.albilineans/1-455   446 RRRAPAGGVT                                     455
P.aeruginosa/1-434    429 LEDERD----                                     434
```

Cascade which is located in a Gly-rich loop plays the equivalent role of Lys finger in *E. coli* Cascade (Xiao et al., 2017). Since AcrIE4 does not inhibit the *E. coli* type I-E CRISPR-Cas system (Pawluk et al., 2014), I modeled *P. aeruginosa* Cas8e based on the structure *T. fusac* Cascade complex (PDB: 5U07) to identify the binding site in Cas8e. And I found three Lys residues near the Gly-rich loop (K176 and K183) and Gln-wedge (K357). These residues are also located in the same position as Cas8e homologs in *Thermus thermophilus* (PDB:4F3E), *Acidimicrobium ferroxidans* (PDB: 4H3T), *T. fusc*a (PDB:3WVO). *T. fusca* has two CRISPR-loci in its genome, and each Cas8e structure was determined in apo form (PDB: 3WVO) and Cascade structure (PDB: 5U07), respectively. Since the sequence identity scores among Cas8e homologs are low at about 20~30%, it was so difficult to find the PAM-interacting residues using multiple sequence alignment. Structural alignment of these Cas8e homologs reveals that positively charged residues near the Gly-rich loop are located in these Cas8e homologs. Lys187 of *T. fusca*, Arg170 of *T. thermophilus*, and Lys186 of *A. ferroxidans* are detected near the K176 and K183 residue of *P. aeruginosa* Cas8e model (Fig. 25). For the positively charged residues near the Gln-wedge, *T. fusca* and *T. thermophilus* Cas8e have the Gln-wedge residue, but only *T. thermophilus* Cas8e has the positively charged residue (Lys350) around Gln-wedge (Fig. 25).

To figure out the target Cas protein of AcrIE4-F7^{NTD}, I purified the Cas proteins with N-terminal $(His)_6$-MBP tag. I could find the target Cas subunit with submicromolar binding affinity, but this method has some limitations: (i) If the interaction between Acr and Cas protein occurs near the N-terminus, the interaction can be sterically hindered by the large N-terminal MBP tag whose molecular weight

Figure 25. The structural comparison of Cas8e homolog

Except for the structure of *T. fusca* Cascade:crRNA complex (PDB: 5U07), Cas8e homolog structures have been revealed in *T. thermophiles* (PDB: 4F3E), *A. ferroxidans* (PDB: 4H3T), and *T. fusca* (PDB: 3WVO). These Cas8e homologs also have positive-charged residues near the Gly-rich loop (Lys187 of *T. fusca*, Arg170 of *T. thermophilus*, and Lys186 of *A. ferroxidans). T. fusca* and *T. thermophilus* Cas8e have the Gln-wedge residue, but *T. thermophilus* Cas8e only has the positively charged residue (Lys350) around Gln-wedge.


P. aeruginosa Cas8e
K357
Q355
K176
K183
T. thermophius Cas8e
Q348
K350
K170
T. fusca Cas8e
Q390
K187
A. ferroxidans Cas8e
K186

is 43 kDa; (ii) Due to the MBP tag at the N-terminus, it is not possible to form a Cascade complex, and I can only check the interaction between Acr and one Cas protein. Thus, it is hard to identify the target Cas protein if the Acr has multiple binding interfaces with several Cas proteins. In fact, in the Cascade structure with AcrIF4, AcrIF6, and AcrIF10 interacting with the PAM recognition site, they contact other subunits near Cas8f (Guo et al., 2017; Zhang et al., 2020; Gabel et al., 2021). Also, it would be tough if Acr binds to the Cascade complex after the conformation change. AcrIF5 only binds to the target DNA-bound form of Cascade, indicating that AcrIF5 has a binding affinity on the Cascade where the conformation change occurs (Xie et al., 2022); (iii) the individual Cas proteins may not fold correctly without interacting with Cascade component.

Like the AcrIE4-F7 which has dual anti-CRISPR activity, some Acr proteins occasionally suppress the divergent CRISPR-Cas system. AcrIF6, AcrIF18.1, AcrIF18.2, and AcrIF22 previously inhibit both type I-E and I-F CRISPR-Cas systems (Pawluk et al., 2016; Pinilla-Redondo et al., 2020). The structure and type I-F inhibitory mechanism of AcrIF6 were discovered but its dual inhibition remains unknown, and those of AcrIF18.1, AcrIF18.2, and AcrIF22 are still unclear. Unlike these Acr proteins, AcrIE4-F7 is unique in that functionally independent Acr proteins are fused and individual domains are responsible for the inhibition of each CRISPR-Cas type. It may be useful to link different Acr proteins. It has been reported that Acr-phages cooperate to suppress the CRISPR-Cas system (Borges et al., 2018; Landsberger et al., 2018). To overcome the host CRISPR-Cas system, tethered Acr proteins which have different inhibitory mechanisms could work more effectively. Moreover, the combination of Acr proteins could provide the potential to act as an improved inhibitor in the CRISPR application fields.

References

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.

Bernheim A and Sorek R. (2020). The pan-immune system of bacteria: antiviral defence as a community resource. *Nat Rev Microbiol, 18*(2), 113-119.

Bondy-Denomy J, Pawluk A, Maxwell KL and Davidson AR. (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature, 493*(7432), 429-432.

Borges AL, Zhang JY, Rollins MF, Osuna BA, Wiedenheft B and Bondy-Denomy J. (2018). Bacteriophage Cooperation Suppresses CRISPR-Cas3 and Cas9 Immunity. *Cell, 174*(4), 917-925 e910.

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.

Chowdhury S, Carter J, Rollins MF, Golden SM, Jackson RN, Hoffmann C, et al. (2017). Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell, 169*(1), 47-57 e11.

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.

Fineran PC and Charpentier E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology, 434*(2), 202-209.

Gabel C, Li Z, Zhang H and Chang L. (2021). Structural basis for inhibition of the type I-F CRISPR-Cas surveillance complex by AcrIF4, AcrIF7 and AcrIF14. *Nucleic Acids Res, 49*(1), 584-594.

Gao Z, Zhang L, Ge Z, Wang H, Yue Y, Jiang Z, et al. (2022). Anti-CRISPR protein AcrIF4 inhibits the type I-F CRISPR-Cas surveillance complex by blocking nuclease recruitment and DNA cleavage. *J Biol Chem, 298*(11), 102575.

Guo TW, Bartesaghi A, Yang H, Falconieri V, Rao P, Merk A, et al. (2017). Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell, 171*(2), 414-426 e412.

Hayes RP, Xiao Y, Ding F, van Erp PB, Rajashankar K, Bailey S, et al. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature, 530*(7591), 499-503.

Hong S, Ka D, Yoon SJ, Suh N, Jeong M, Suh JY, et al. (2018). CRISPR RNA and anti-CRISPR protein binding to the Xanthomonas albilineans Csy1-Csy2 heterodimer in the type I-F CRISPR-Cas system. *J Biol Chem, 293*(8), 2744-2754.

Hong SH, Lee G, Park C, Koo J, Kim EH, Bae E, et al. (2022). The structure of AcrIE4-F7 reveals a common strategy for dual CRISPR inhibition by targeting PAM recognition sites. *Nucleic Acids Res, 50*(4), 2363-2376.

Jia N and Patel DJ. (2021). Structure-based functional mechanisms and biotechnology applications of anti-CRISPR proteins. *Nat Rev Mol Cell Biol, 22*(8), 563-579.

Kelley LA, Mezulis S, Yates CM, Wass MN and Sternberg MJ. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc, 10*(6), 845-858.

Kim I, Koo J, An SY, Hong S, Ka D, Kim EH, et al. (2020). Structural and mechanistic insights into the CRISPR inhibition of AcrIF7. *Nucleic Acids Res, 48*(17), 9959-9968.

Landsberger M, Gandon S, Meaden S, Rollie C, Chevallereau A, Chabas H, et al. (2018). Anti-CRISPR Phages Cooperate to Overcome CRISPR-Cas Immunity. *Cell, 174*(4), 908-916 e912.

Leon LM, Park AE, Borges AL, Zhang JY and Bondy-Denomy J. (2021). Mobile element warfare via CRISPR and anti-CRISPR in Pseudomonas aeruginosa. *Nucleic Acids Res, 49*(4), 2114-2125.

Makarova KS, Wolf YI, Iranzo J, Shmakov SA, Alkhnbashi OS, Brouns SJJ, et al. (2020). Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol, 18*(2), 67-83.

Marino ND, Zhang JY, Borges AL, Sousa AA, Leon LM, Rauch BJ, et al. (2018). Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science, 362*(6411), 240-242.

Marraffini LA. (2015). CRISPR-Cas immunity in prokaryotes. *Nature, 526*(7571), 55-61.

Mejdani M, Pawluk A, Maxwell KL and Davidson AR. (2021). Anti-CRISPR AcrIE2 Binds the Type I-E CRISPR-Cas Complex But Does Not Block DNA Binding. *J Mol Biol, 433*(3), 166759.

Mojica FJM, Diez-Villasenor C, Garcia-Martinez J and Almendros C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading), 155*(Pt 3), 733-740.

Niu Y, Yang L, Gao T, Dong C, Zhang B, Yin P, et al. (2020). A Type I-F Anti-CRISPR Protein Inhibits the CRISPR-Cas Surveillance Complex by ADP-Ribosylation. *Mol Cell, 80*(3), 512-524 e515.

Pawluk A, Bondy-Denomy J, Cheung VH, Maxwell KL and Davidson AR. (2014). A new group of phage anti-CRISPR genes inhibits the type I-E CRISPR-Cas system of Pseudomonas aeruginosa. *mBio, 5*(2), e00896.

Pawluk A, Davidson AR and Maxwell KL. (2018). Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol, 16*(1), 12-17.

Pawluk A, Shah M, Mejdani M, Calmettes C, Moraes TF, Davidson AR, et al. (2017). Disabling a Type I-E CRISPR-Cas Nuclease with a Bacteriophage-Encoded Anti-CRISPR Protein. *mBio, 8*(6).

Pawluk A, Staals RH, Taylor C, Watson BN, Saha S, Fineran PC, et al. (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol, 1*(8), 16085.

Pinilla-Redondo R, Shehreen S, Marino ND, Fagerlund RD, Brown CM, Sorensen SJ, et al. (2020). Discovery of multiple anti-CRISPRs highlights anti-defense gene clustering in mobile genetic elements. *Nat Commun, 11*(1), 5652.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol, 7*, 539.

Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C and Banfield JF. (2013). Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol, 15*(2), 463-470.

Wang H, Gao T, Zhou Y, Ren J, Guo J, Zeng J, et al. (2022). Mechanistic insights into the inhibition of the CRISPR-Cas surveillance complex by anti-CRISPR protein AcrIF13. *J Biol Chem, 298*(3), 101636.

Wang X, Yao D, Xu JG, Li AR, Xu J, Fu P, et al. (2016). Structural basis of Cas3 inhibition by the bacteriophage protein AcrF3. *Nat Struct Mol Biol, 23*(9), 868-870.

Waterhouse AM, Procter JB, Martin DM, Clamp M and Barton GJ. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics, 25*(9), 1189-1191.

Wiedenheft B, Sternberg SH and Doudna JA. (2012). RNA-guided genetic silencing

systems in bacteria and archaea. *Nature, 482*(7385), 331-338.
Xiao Y, Luo M, Hayes RP, Kim J, Ng S, Ding F, et al. (2017). Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell, 170*(1), 48-60 e11.
Xie Y, Zhang L, Gao Z, Yin P, Wang H, Li H, et al. (2022). AcrIF5 specifically targets DNA-bound CRISPR-Cas surveillance complex for inhibition. *Nat Chem Biol, 18*(6), 670-677.
Xue C and Sashital DG. (2019). Mechanisms of Type I-E and I-F CRISPR-Cas Systems in Enterobacteriaceae. *EcoSal Plus, 8*(2).
Yang L, Zhang L, Yin P, Ding H, Xiao Y, Zeng J, et al. (2022). Insights into the inhibition of type I-F CRISPR-Cas system by a multifunctional anti-CRISPR protein AcrIF24. *Nat Commun, 13*(1), 1931.
Yang L, Zhang Y, Yin P and Feng Y. (2021). Structural insights into the inactivation of the type I-F CRISPR-Cas system by anti-CRISPR proteins. *RNA Biol, 18*(sup2), 562-573.
Zhang K, Wang S, Li S, Zhu Y, Pintilie GD, Mou TC, et al. (2020). Inhibition mechanisms of AcrF9, AcrF8, and AcrF6 against type I-F CRISPR-Cas complex revealed by cryo-EM. *Proc Natl Acad Sci U S A, 117*(13), 7176-7182.

Abstract in Korean

Anti-CRISPR 단백질, AcrIE4-F7 의 생화학적 연구

서울대학교 대학원
농생명공학부 응용생명화학전공
이규진

박테리오파지와 플라스미드의 침입을 방어하기 위해서, 박테리아와 고세균은 CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats–CRISPR associated proteins) system 을 사용한다. Type I-E 와 I-F CRISPR-Cas system 은 침입한 DNA 를 인지하기 위해서 Cascade (CRISPR-associated complex for antiviral defense) 를 활용한다. 이에 대항하여, 박테리오파지는 CRISPR-관련 면역체계를 무력화하기 위해서 Acr (anti-CRISPR) 단백질을 진화시켰다. *Pseudomonas aeruginosa* 유래 AcrIE4-F7 은 AcrIE4 와 AcrIF7 의 fusion protein 으로 type I-E 와 I-F CRISPR-Cas system 을 모두 무력화시킨다. 본 연구에서는 type I-E 와 I-F Cascade complex 를 구성하는 Cas 단백질들과 AcrIE4-F7 사이의 상호작용을 분석하여 AcrIE4-F7 의 target Cas 단백질들을 보고했다. AcrIE4-F7 의 N-말단과 C-말단 도메인 각각이 Cas8e 와 Cas8f 의 PAM 인식 부위와 상호작용하여 target DNA 의 결합을 억제한다. Pairwise sequence alignment 와

mutation analyses 는 각각의 Acr 도메인에 존재하는 음전하 잔기들이 target Cas8 과의 상호 작용에 필수적이라는 것을 말해준다. 이러한 결과들은 PAM 인식 부위가 AcrIE4-F7 의 주요 target 이라는 것을 제안하고 이는 다양한 CRISPR-Cas type 들을 억제하는 일반적인 메커니즘이다.

주요어: CRISPR-Cas, Anti-CRISPR, AcrE4-F7, Cascade complex, *Pseudomonas aeruginosa*, PAM 인식 부위

학번: 2021-27675