농학석사학위논문

*Xanthomonas albilineans*에서 유래된 CRISPR-associated Protein Cas2의 결정 구조

Crystal Structure of CRISPR-associated Protein Cas2 from *Xanthomonas albilineans*

2016년 8월

서울대학교 대학원

농생명공학부 응용생명화학전공

정 유 진

A dissertation for the Degree of Master of Science

Crystal Structure of CRISPR-associated Protein Cas2 from *Xanthomonas albilineans*

*Xanthomonas albilineans*에서 유래된 CRISPR-associated Protein Cas2의 결정 구조

August 2016

U geene Jeong

Applied Life Chemistry Major

Department of Agricultural Biotechology

Seoul National University

*Xanthomonas albilineans*에서 유래된 CRISPR-associated Protein Cas2의 결정 구조

Crystal Structure of CRISPR-associated Protein Cas2 from *Xanthomonas albilineans*

지도교수 배 의 영

이 논문을 농학석사학위논문으로 제출함

2016년 7월

서울대학교 대학원

농생명공학부 응용생명화학전공

정 유 진

정 유 진의 석사학위논문으로 인준함

2016년 7월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Abstract

Crystal Structure of CRISPR-associated Protein Cas2 from *Xanthomonas albilineans*

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) system that provides RNA-based adaptive immunity to archaea and bacteria is a defense system against extrachromosomal genomic elements. And their variable CRISPR-associated (Cas) proteins are known as immune effectors which are composed of distinct subtypes. There are three steps in CRISPR-Cas system; (1) Acquisition (2) Processing of crisprRNA and (3) Interference. All CRISPR systems have Cas1 and Cas2 proteins. *cas1* and *cas2* genes are core *cas* genes in all CRISPR-Cas subtypes. The *cas2* gene from *Xanthomonas albilineans* GPE PC73 genomic DNA (Gene ID: 8702143) was amplified by polymerase chain reaction for generation of a C-terminal His tag and it was cloned into pET21a. The *Xacas2* gene was expressed in *Escherichia coli* BL21 (DE3). The XaCas2 protein was purified by two steps and I could get pure XaCas2 protein to make protein crystals. The protein crystals were made in a condition that contains 0.17 *M* ammonium acetate, 0.085 *M* sodium acetate pH 4.6, 25.5% (w/v) polyethylene glycol 4000 and 15% (v/v) glycerol. The conditions of protein crystals were optimized for making big and angulate protein crystals. The protein crystals of XaCas2 were made at two different temperatures, 20 ℃ and 4℃. The structures of XaCas2 were solved with molecular replacement. Both structures of XaCas2 consist of N-terminal ferredoxin fold and C-terminal segment which contains a short 3_{10}-helix and a fifth β-strand. And there was no difference between two structural features.

However, conformational differentiations were found in C-terminal end between A and B protomers within each crystal structure. The fifth β-strand crosses the four stranded β-sheets on the other protomer, making five stranded antiparallel β-sheets. The swapping of C-terminal segment stabilize the dimerization. The dimerization is supported by hydrogen bonds and salt bridges in β-sheets. Comparing with other Cas2 homologues, XaCas2 likely to have double strand DNases activity.

Key Words: Cas2, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), CRISPR associated (Cas) proteins, *Xanthomonas albilineans* GPE PC73, x-ray crystallography

Student Number: 2013-21186

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Asp	Aspartic acid
Bh	*Bacillus halodurans*
BME	β-mercaptoethanol
Cas	CRISPR-associated
crRNA	crisprRNA
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Dv	*Desulfovibrio vulgaris*
DTT	Dithiothreitol
E. coli	*Escherichia coli*
IMAC	Immobilized metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
MR	Molecular replacement
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
RMSD	Root mean square deviation
Sp	*Streptococcus pyogenes*
Ss	*Sulfolobus solfataricus*
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Xa	*Xanthomonas albilineans*

Introduction

Clustered Regularly Interspaced Short Palindromic Repeats (Bolotin et al. 2005) and their associated proteins are compose of defense mechanisms in archaeal and bacterial genomes. This defense system is a nucleic acid-based immune system adjusted by variable protein families which are composed of distinct subtypes.

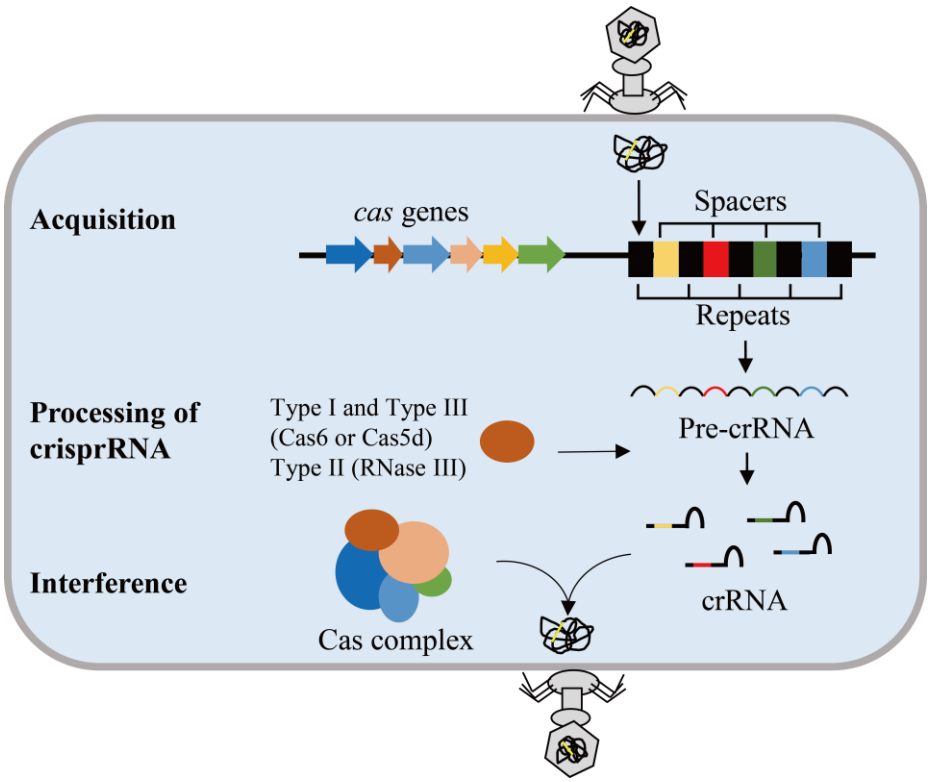
CRISPR array is a unique sequence storing the immunological memory in genomic DNA (Mojica et al. 2005). A CRISPR locus consist of 'spacer' and 'repeat' (Fig. 1). The 20 to 50 base pairs spacers are short DNA sequences originated from invading pathogens that intercalated in CRISPR DNA repeats. The regulatory leader sequence is located in front of CRISPR array (Makarova et al. 2011)

In 1987, CRISPR locus and their functions were firstly analyzed in silico (Ishino et al. 1987). The first description about CRISPR locus was published in 2000 (Mojica et al. 2000). They named the repetitive sequences as CRISPR locus. And the relation between *cas* genes and CRISPR repeats was found in 2002 (Jansen et al. 2002). In 2005, it is confirmed that the spacers correspond to sequences in phages and plasmids (Bolotin et al. 2005, Mojica et al. 2005). The researchers proposed that the CRISPR-Cas systems are against the invasions of phage infection and plasmid conjugation.

There are major three steps in CRISPR-Cas systems: (1) Acquisition, (2) Processing of crisprRNA (crRNA) and (3) Interference (Fig. 1). Three major types have same defense pathways but distinct differences exist in mechanism. The universal proteins, Cas1 and Cas2, are involved in adaptation step. The

Figure 1. Overview of CRISPR-Cas system

Overview of the CRISPR-Cas system with their associated Cas proteins. In first step, a short DNA sequence is stored at CRISPR array in genomic DNA. And a long pre-crRNA is processed into short mature-crRNAs by Cas6, Cas5d or RNase III. In the final step, crRNA and the remaining Cas proteins form a ribonucleoprotein complex and the complex destroys foreign nucleic acid.


Acquisition
cas genes
Spacers
Repeats
Processing of crisprRNA
Type I and Type III
(Cas6 or Cas5d)
Type II (RNase III)
Pre-crRNA
Interference
crRNA
Cas complex

protein complex of Cas1 and Cas2 has endonuclease activity and is essential for insertion of foreign nucleic acid into CRIPSR array (Makarova et al. 2011, Nuñez et al. 2014). The short DNA sequences from foreign nucleic acids are called spacer. In the second stage, pre-crRNA is processed into mature-crRNA. The distinct three types of CRISPR-Cas systems have different methods to processing. The crRNA is processed by Cas6 or Cas5 endoribonuclease in the type I and III (Carte et al. 2008) while RNase III endoribonuclease in the type II (Deltcheva et al. 2011). Either a Cas protein or Cas protein complex that forming the effector complex with mature-crRNA is used in interference step in the final step of CRISPR-Cas system. The mature-crRNA binds to Cas9 protein (a single multidomain protein) in type Ⅱ. In case of type I and type III, a mature-crRNA binds to multi subunit complex. (Cas3 and other proteins in type I and Cas10 and other proteins in type III) (Pattanayak et al. 2013, Semenova et al. 2011, Wiedenheft et al. 2011, Yang et al. 2013, Banga and Boyd 1992, Beumer et al. 2013). The crRNA helps the effector proteins to cleave target nucleic acid (Fig. 1).

CRISPR-Cas system is divided into three major types (Type Ⅰ, II, III) on the basis of the topology of the Cas1 phylogenetic tree (Makarova et al. 2011) and two major types (Type IV and V) were added in a recent study. There are two classes (class 1 and class 2) and five types (I, II, III, IV and V). 'Class 1' system contains ribonucleotide complexes while 'Class 2' include a single protein as effector protein. And the five types are divided into 16 subtypes including two new types (IV and V systems) and three new subtypes (II-C, III-C and III-D) by their proteins.

In subtype I-C, there are seven Cas proteins including Cas1, Cas2, Cas3, Cas4, Cas5d, Cas7 and Cas8 encoded in a single operon (Makarova et al. 2015b). The Cas1 and Cas2 are involved in the adaptation step. The Cas4 is suggested as a member of adaptation stage in recent study (Zhang et al. 2012). The Cas5d is a subunit of complex and processes a crRNA to short mature-crRNA (Koo et al. 2013). The Cas7 is used as backbone of effector protein and binds to mature-crRNA (Jackson et al. 2014). The Cas8 is a hallmark of subtype I-C, involved in PAM (protospacer adjacent motif) recognition (Cass et al. 2015). And the Cas3 has the DNA nuclease activity which is used in interference step (Makarova and Koonin 2015a).

Cas1 proteins are identified as metal dependent endonuclease (Nuñez et al. 2014, Wiedenheft et al. 2009). Although Cas2 proteins are universally known core protein, the function of Cas2 is not fully studied. In one study by Yosef et al., (2012), they give the direct evidence that the complex of Cas1 and Cas2 involves in adaptation step (Yosef et al.2012). And the crystal structure of complex of Cas1 and Cas2 from *E. coli* including dsDNA strand is described in 2014 (Nuñez et al. 2014). However, the mechanism of selection of foreign DNAs remains unknown.

Some crystal structures of Cas2 from *Bacillus halodurans* (BhCas2), *Desulfovibrio vulgaris* (DvCas2), *Sulfolobus solfataricus* (SsCas2), and *Streptococcus pyogenes* (SpCas2) are identified (Ka et al. 2014, Nam et al. 2012, Beloglazova et al. 2008). SsCas2 has ssRNase activity while BhCas2 and SpCas2 have dsDNase activity (Ka et al. 2014, Nam et al. 2012, Beloglazova et al. 2008). A pair of aspartate residues in N-terminal in SsCas2 is assumed as catalytic residues coordinating the catalytic divalent metal ion(s) (Beloglazova

et al. 2008).

Here, I report the structures of Cas2 originated from *Xanthomonas albilineans* GPE PC73 (XaCas2). X-ray crystallization is used for structural determination and XaCas2 has typical characteristics of Cas2 proteins. XaCas2 has N-terminal ferredoxin fold and C-terminal segment which contains a short 3_{10}-helix and a fifth β-strand. Comparing with Cas2 homologues, the features of XaCas2 are similar with BhCas2 and SpCas2, meaning XaCas2 might bind to dsDNA and cleaves it.

Materials and Methods

Cloning of *Xacas2* gene

cas2 gene from *Xanthomonas albilineans* GPE PC73 genomic DNA (Gene ID: 8702143) was synthesized in pBHA vector through gene synthesis service in Bioneer (Korea). *Xacas2* gene was amplified from pBHA vector using polymerase chain reaction (PCR) with primers which is including 5`- NdeI and 3`- XhoI restriction enzyme sites. PCR was performed with materials as shown below: synthesized *Xacas2* DNA, primers (forward primer: 5`-AGA TAT ACA TAT GAT GGT ATT GGT TTC GTA TGA CG-3`, reverse primer: 5`-AAT TCC TCG AG AAA AAT CAG TGG GCC TTT AAG ATC GAG G-3`), dNTPs and pfu polymerase (Bioneer). Steps of PCR condition started from initial denaturation (at 95℃ for 5 min) to 30 cycles of the PCR reaction (denaturation at 95℃ for 30 sec, annealing at 55℃ for 30 sec, and amplification at 72℃ for 60 sec). And resulting product was digested with NdeI and XhoI restriction enzymes. Digested product was cloned into pET21a (Novagen), an expression vector, which has a C-terminal $(His)_6$ tag via NdeI and XhoI restriction enzyme sites. The plasmid was transformed into *E. coli* BL21 (DE3 lysogen) (Enzynomics) to express XaCas2 proteins.

Protein expression and purification of XaCas2

The transformed *E. coli* BL21 (DE3) cells were grown in Luria-Bertani (LB) broth medium containing 0.1 mg ml^{-1} ampicillin at 37℃ until optical density reached 0.7 at A600. Protein expression was induced by adding a isopropyl-β-D-thiogalactopyranoside (IPTG) making a final concentration 0.5 mM. And

incubation was continued at 17℃ for 17 h with continuous shacking. Cells were harvested by centrifugation and resuspended in lysis buffer (20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5% (v/v) glycerol, 25 m*M* imidazole, 5 m*M* β-mercaptoethanol (BME), 0.25 m*M* pheny -lmethylsulfonyl fluoride (PMSF)) in the ice to achieve cell lysate.

The lysate was sonicated to reducing viscosity and centrifuged to separate insoluble components (30000 g, 55 min, 4℃). The supernatant was filtered with 0.45 μm filter and loaded onto a 5 mL HisTrap HP column (GE Healthcare, USA) pre-equilibrated with binding buffer (20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5% (v/v) glycerol, 25 m*M* imidazole, 5 m*M* BME). The column was washed with 20 column volumes (CV) of binding buffer and the bound proteins were eluted by applying a linear gradient of imidazole up to 500mM with elution buffer (20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5% (v/v) glycerol, 500 m*M* imidazole, 5 m*M* BME).

The pooled fractions of IMAC sample were analyzed by sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) and a HiLoad 16/60 Superdex 200 column (GE Healthcare, USA) is used as further purification step to confirm that XaCas2 forms a oligomer separating by sizes of proteins. The column was equilibrated with size exclusion chromatography buffer (20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5% (v/v) glycerol, 2 m*M* dithiothreitol (DTT)). All purification process were performed at 4℃.

Crystallization of XaCas2

The XaCas2 protein in buffer (20 m*M* HEPES pH 7.5, 200 m*M* NaCl, 5% (v/v) glycerol, 2 m*M* DTT) was concentrated to 6 mg ml $^{-1}$. Crystallization was tested

864 conditions for screening by mixing the protein solution with reservoir solution in a 1 : 1 ratio. One condition (0.17 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 25.5% (w/v) polyethylene glycol 4000 (PEG 4000), 15% (v/v) glycerol) made crystals and the crystals were confirmed as protein crystals with UVEX UV fluorescence imaging systems (Molecular Dimension). Crystals of XaCas2 were optimized by the hanging drop vapor diffusion method at 4℃ and 20℃. The crystal at 20℃ was made in 0.05 *M* acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 23.0% (w/v), PEG4000 and 10% (v/v) glycerol. The crystals at 4℃ were made in 0.10 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 23.0% (w/v) PEG 4000 and 10% (v/v) glycerol. The crystals grown at 4℃ were transferred into cryo-protection buffer (0.10 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 23.0% (w/v) PEG4000, 12.5% (v/v) glycerol) before flash frozen in liquid nitrogen after 4 days.

X-ray data collection

Diffraction data of the XaCas2 was collected at the wavelength of 0.9793 in the Pohang accelerator laboratory beamline 7A (Pohang, Korea) at 100 K. Each crystal was soaked in a cryo-protection solution that 2% (v/v) glycerol is added in reservoir buffer for data collection.

Model building and refinement

The data was processed with HKL2000 (Otwinowski and Minor 1997). The *B. halodurans* Cas2 structure (PDB code: 4ES2) was used as a starting model for molecular replacement (MR) phasing in PHASER (McCoy et al. 2007) (Table

1). The final structures of XaCas2 were completed using alternate cycles of manual fitting in COOT (Emsley and Cowtan 2004) and refined using PHNIX.REFINE (Afonine et al. 2012). The stereochemical quality of the final models was assessed using MolProbity (Chen et al. 2010).

Table 1. Classification of Cas2 homologues and sequence similarity to XaCas2

Subtype	PDB ID	Origins of *cas2* gene	Sequence identity of amino acid (%)
I-A	3EXC	*Sulfolobus solfataricus*	33.72
I-C	4ES2	*Bacillus halodurans*	62.50
I-C	3OQ2	*Desulfovibrio vulgaris*	62.50
I-C	4QR0	*Streptococcus pyogenes*	52.08

* XaCas2 is used as standard sequence to estimate sequence identity.

Results

Expression and purification of XaCas2

PCR analyses was performed to amplified the 300 bp-*Xacas2* gene including 5`- NdeI and 3`- XhoI restriction enzyme sites (Fig. 2). The resulting XaCas2-pET21a plasmid contains 96-amino acid of XaCas2 and C-terminal $(His)_6$ tag. DNA sequence of *Xacas2* gene was confirmed through a sequencing service provided by Bioneer (Korea) (Table 2). The XaCas2-$(His)_6$ protein (molecular weight: 11.92 kDa) was successfully induced with adding 0.5 mM IPTG. As expected, XaCas2 was separated into high purity at linear gradient of imidazole from immobilized metal affinity chromatography (IMAC) in peak 2 (Fig. 3A). Peak 1 contained that XaCas2 with other proteins and the XaCas2 in peak 2 was pooled and concentrated after SDS-PAGE analysis (Fig. 3B). Protein was aggregated during concentration due to the protein buffer conditions or other reasons and filtered with 0.45μm filter. XaCas2 forms a homodimer corresponding a peak for 20 kDa (Peak 2) and uncertain substances were eluted in peak 1 in size exclusion chromatography (Fig. 4A). The uncertain substances were assumed as nucleic acids because there are no stained protein in peak 1 shown in SDS-PAGE gel (Fig. 4B).

Crystallization and X-ray analysis of XaCas2

The screening kit from Molecular Dimension (USA) was used for screening the initial crystallization. One condition in MCSG1 (C4: 0.17 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 25.5% (w/v) PEG 4000, 15% (v/v) glycerol) made the protein crystals among 864 kinds of conditions (Fig. 5). The

Figure 2. PCR and colony PCR product of *Xacas2* gene loaded on agarose gel

(A) A 291 bp - *Xacas2* gene was amplified and was confirmed it's size using agarose gel electrophoresis. (B) Agarose gel electrophoresis of colony PCR product. Additional sequences (179 bp) in vector are added to 291 bp - *Xacas2* gene due to using the T7 primers.

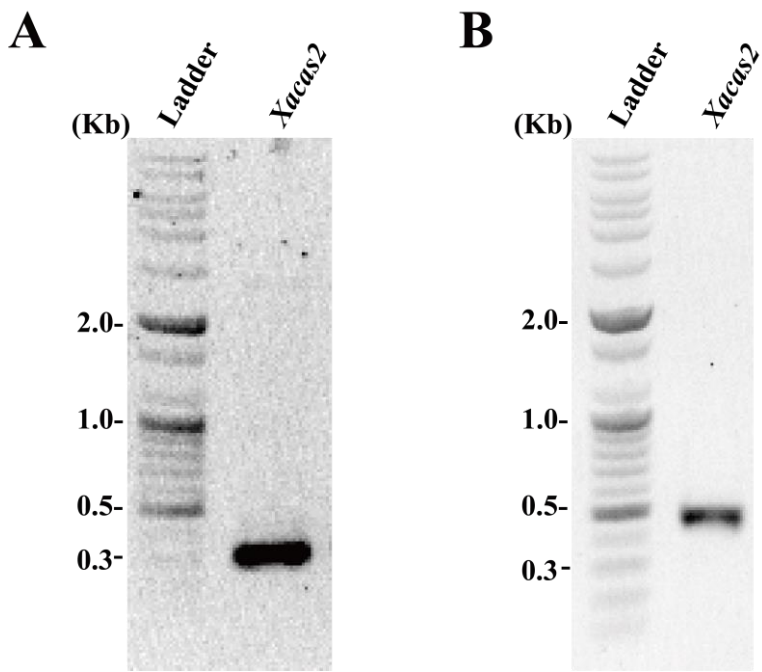

A
Ladder
Xacas2
(Kb)
2.0-
1.0-
0.5-
0.3-
B
Ladder
Xacas2
(Kb)
2.0-
1.0-
0.5-
0.3-

Table 2. Sequences of XaCas2

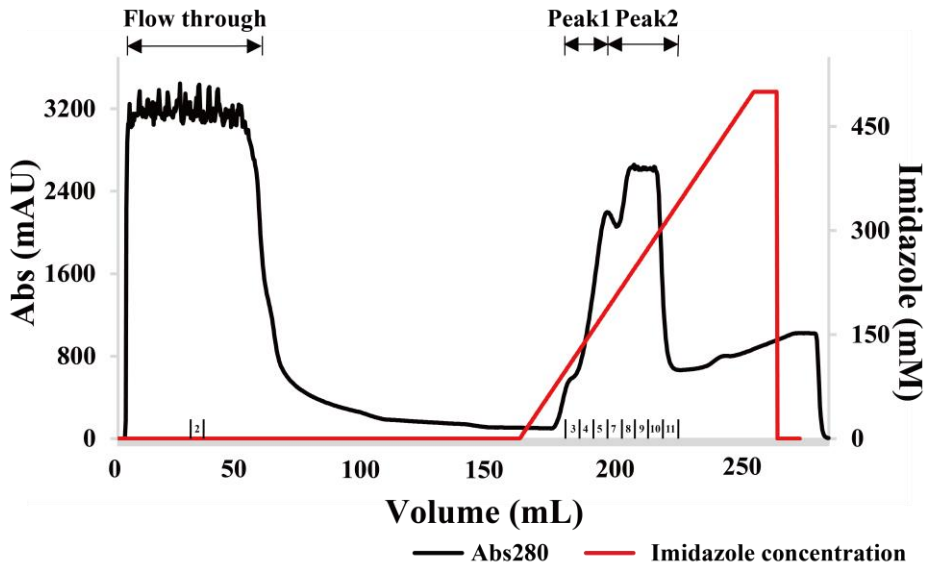
The sequences of XaCas2-(His)$_6$ protein	
Sequence of DNA	5'-*CAT<u>ATG</u>*ATGGTATTGGTTTCGTATGACGTGT CAAATCCAGTCCGGGTGGAGATAAACGTCT GCGTAAGGTAGCAAAAGCCTGCCGTGATTTAG GGCAACGAGTTCAATTTTCAGTCTTTGAGATCG AAGTCGACCCGGCGCAGTGGACAGCACTTAGA CAGCGGTTATGTGATCTGATAGATCCTGATATT GATTCTCTGCGCTTCTATCATTTAGGTGCTAAA TGGGAAGCTCGCGTGGAACACGTTGGCGCCAA ACCCAGCCTCGATCTTAAAGGCCCACTGATTTT T*CTCGAG***CACCACCACCACCACCAC**-3
Sequence of amino acid with (His)$_6$ tag	5'-*<u>M</u>*MVLVSYDVSTSSPGGDKRLRKVAKACRDLG QRVQFSVFEIEVDPAQWTALRQRLCDLIDPDIDSL RFYHLGAKWEARVEHVGAKPSLDLKGPLIF*LE***HH HHHH**-3'

*The start residue (Metionoine) is underlined.
* Additional sequences encoding 5`- NdeI and 3`- XhoI restriction enzyme sites are slanted.
* Sequences encoding (His)$_6$ tag are marked in bold

Figure 3. Elution profile of affinity chromatography with SDS-PAGE analysis of XaCas2

(A) The IMAC (immobilized metal affinity chromatography) purification of XaCas2 is shown. The blue line displays imidazole gradient and the black line is elution profile of proteins. (B) SDS-PAGE gel of IMAC step product; Lane 1: soluble proteins; supernatant after centrifugation, Lane 2: the pass through proteins, Lane 3-5 (peak1): eluted of XaCas2 with other proteins, Lane 6: molecular weight marker (kDa), Lane 7-11 (peak2): eluted protein of XaCas2.

A


Flow through
Peak1 Peak2
3200
2400
1600
800
0
Abs (mAU)
450
300
150
0
Imidazole (mM)
0
50
100
150
200
250
Volume (mL)
Abs280
Imidazole concentration

B

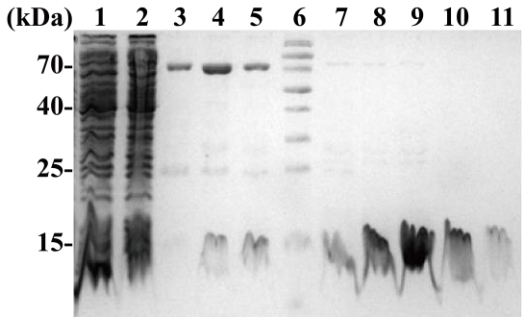
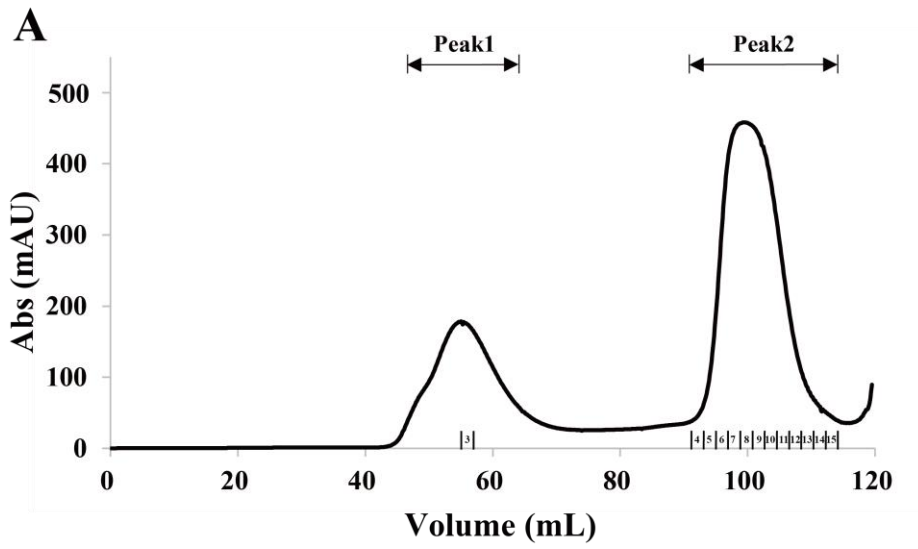

(kDa)
1
2
3
4
5
6
7
8
9
10
11
70-
40-
25-
15-

Figure 4. Elution profile of size exclusion chromatography with SDS-PAGE analysis of XaCas2

(A) The size exclusion chromatography purification step using HiLoad 16/60 Superdex 200 column of XaCas2 is shown. The XaCas2 was purified as dimer (20 kDa: 11kDa x 2). (B) SDS-PAGE gel of second step product; Lane 1: pooled proteins from IMAC purification step, Lane 2: molecular weight marker (kDa), Lane 3: eluted uncertain substance, Lane 4-15 (peak2): eluted protein of XaCas2.


A
Peak1
Peak2
Abs (mAU)
500
400
300
200
100
0
0
20
40
60
80
100
120
Volume (mL)
3
4
5
6
7
8
9
10
11
12
13
14
15

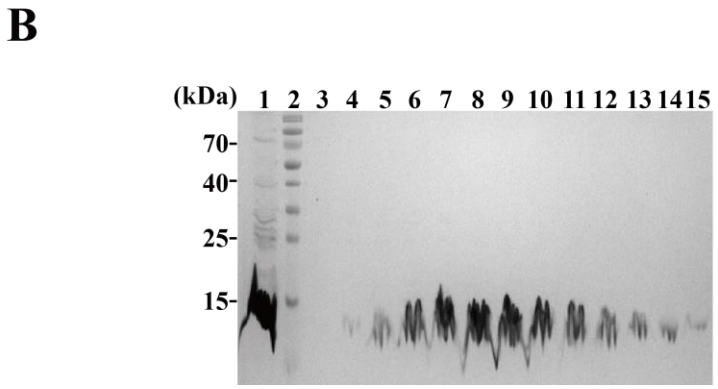
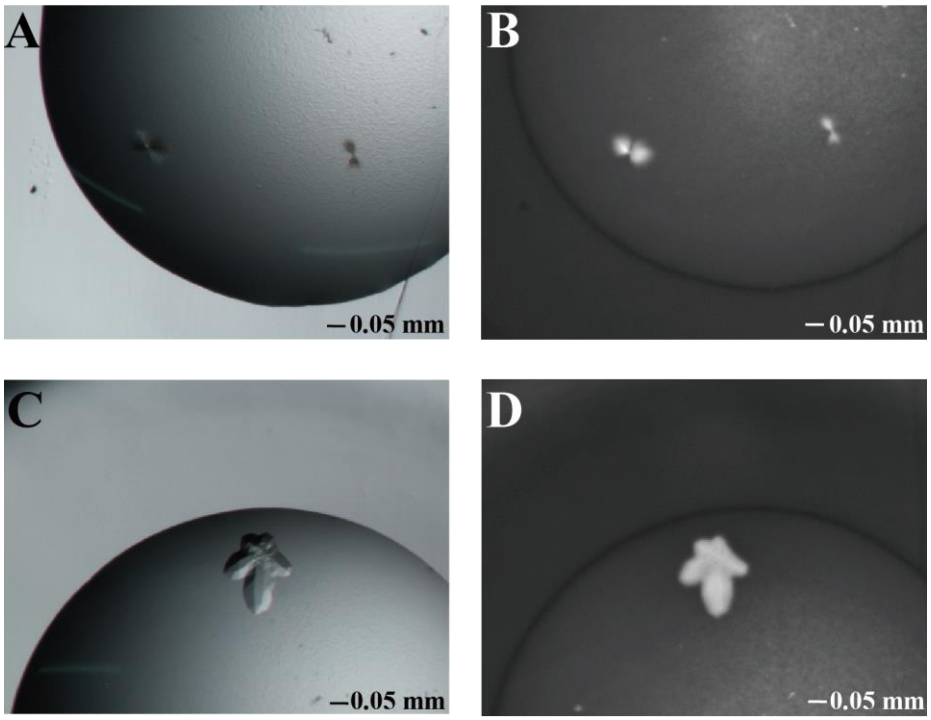

B
(kDa)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
70-
40-
25-
15-

Figure 5. Initial protein crystals of XaCas2

(A) Needle shape and (C) flattened pillar shape are found in MCSG1 (C4: 0.17 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 25.5% (w/v) PEG 4000, 15% (v/v) glycerol) using stereoscopic microscope. (B and D) The image of crystals from UV fluorescence imaging systems.


A
—0.05 mm
B
—0.05 mm
C
—0.05 mm
D
—0.05 mm

protein crystals of XaCas2 optimized by hanging drop vapor diffusion method at 20℃ and 4℃ (Fig. 6A and C). The crystals were fully grown in four days and over the size of all crystals were over 0.2 mm. Both crystals were confirmed as protein crystals through UV fluorescence imaging systems (Fig. 6B and D). The X-ray data of crystals was collected by the beamline 7A at a wavelength of 0.9793 in Pohang accelerator laboratory (Pohang, Korea). Data collection and refinement statistics of 20℃ and 4℃ are shown in Table 3 and Table 4 respectively. The space group of both conditions are $P6_3$.

Crystal structures of XaCas2

The Cas2 homologues share common structural features. Both crystal structures of 20℃ and 4℃ have homo-dimeric assembly which contains N-terminal *βαββαβ* ferredoxin fold and C-terminal residues (Fig. 7A and B). Some residues in C-terminal segments (residues in 87-96 of both protomers) in both structures are not shown in final structure due to deficiency of electron density. And two serine residues (Ser^{12} and Ser^{13}) in 20℃ also are not shown as the same reason. The root mean square deviation (RMSD) of C_α is 0.2 between 20℃ and 4℃. The N-terminal *βαββαβ* ferredoxin fold in each protomer is composed of a four-stranded antiparallel *β*-sheet (*β*1–4) and two *α*-helices (*α*1, *α*2), and C-terminal segment includes short 3_{10} helix and fifth *β*-strand (*β*5) (Fig. 7C). The fifth *β*-strand (*β*5) in C-terminal segments of one promoter has interaction with fourth *β*- strand (*β*4) in four-stranded antiparallel *β*-sheet of another protomer, leading two five-stranded antiparallel *β* -sheets in a single XaCas2 (Fig. 7A and B).

When A and B protomers were aligned based on structure within XaCas2, they had similar overall structure and the RMSD of of C_α is 0.1. But there was

Figure 6. Images of crystals at 20℃ and 4℃ used in data collection

(A) The optimized XaCas2 at 20℃. This crystal has flatted hexagon shape. The reservoir buffer contains 0.05 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 23% (w/v) PEG 4000 and 10.0% (v/v) glycerol. (B) The image of crystal from UV fluorescence imaging systems of 20℃ is shown. (C) The optimized XaCas2 at 4℃. The crystals have flatted pillar shape. 0.10 *M* ammonium acetate, 0.085 *M* sodium acetate: HCl pH 4.6, 22.5% (w/v) PEG 4000 and 10.0% (v/v) glycerol are belongs in reservoir buffer. (D) The image of crystals from UV fluorescence imaging systems of 4℃ is shown.

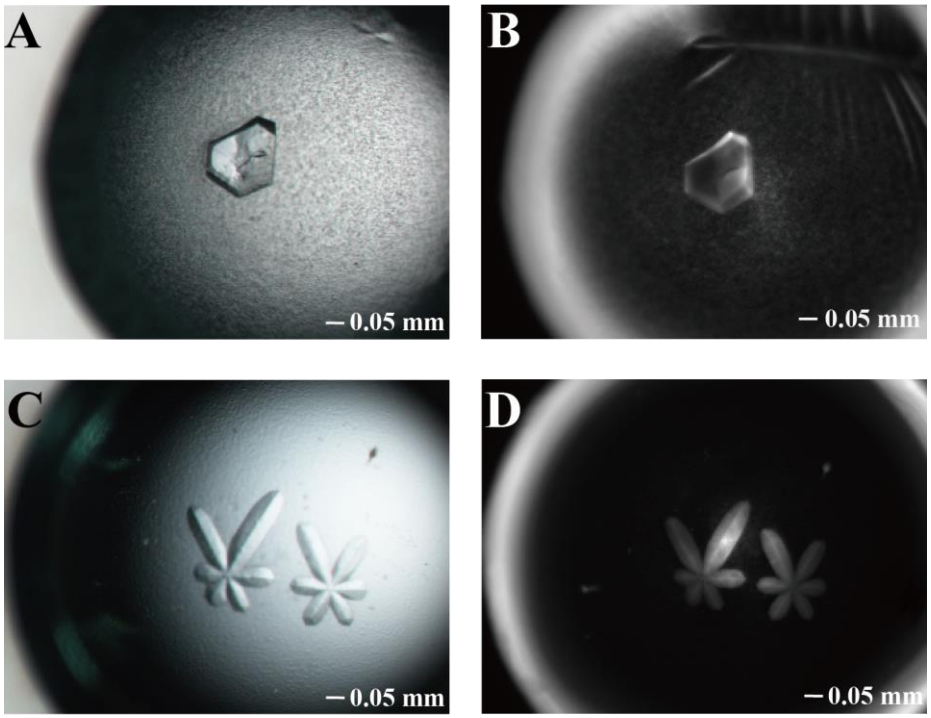

A
— 0.05 mm
B
— 0.05 mm
C
— 0.05 mm
D
— 0.05 mm

Table 3. Data collection and refinement statistics of crystal structure of XaCas2 at 20℃

Data collection	**20℃**
Space group	$P6_3$
Unit cell parameters (Å)	$a = b = 90.39$, $c = 50.83$
Wavelength (Å)	0.9793
Data collection statistics	
Resolution range (Å)	50.00–1.65 (1.71–1.65)
Number of reflections	345997 (28581)
Completeness (%)	100.0 (99.7)
R_{merge}[a]	0.071 (0.809)
Redundancy	12.1 (11.5)
Mean I/σ	33.1 (3.1)
Refinement statistics	
Resolution range (Å)	29.59–1.65
R_{cryst}[b]/R_{free}[c] (%)	16.7/20.1
RMSD bonds (Å)	0.006
RMSD angles (deg)	1.006
Average B-factor (Å^2)	40.11
Number of water molecules	179
Ramachandran favored (%)	98.8
Ramachandran allowed (%)	1.2

Values in parentheses are for the highest-resolution shell.

[a] $R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, 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.

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

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

Table 4. Data collection and refinement statistics of crystal structure of XaCas2 at 4℃

Data collection	**4℃**
Space group	$P6_3$
Unit cell parameters (Å)	$a = b = 90.63$, $c = 50.06$
Wavelength (Å)	0.9793
Data collection statistics	
Resolution range (Å)	50.00–1.75 (1.81–1.75)
Number of reflections	292642 (23845)
Completeness (%)	99.9 (100)
R_{merge}[a]	0.067 (0.700)
Redundancy	12.3 (12.3)
Mean I/σ	35.8 (3.7)
Refinement statistics	
Resolution range (Å)	29.66–1.75
R_{cryst}[b]/R_{free}[c] (%)	16.7/19.9
RMSD bonds (Å)	0.006
RMSD angles (deg)	1.029
Average B-factor (Å^2)	37.70
Number of water molecules	161
Ramachandran favored (%)	97.0
Ramachandran allowed (%)	3.0

Values in parentheses are for the highest-resolution shell.

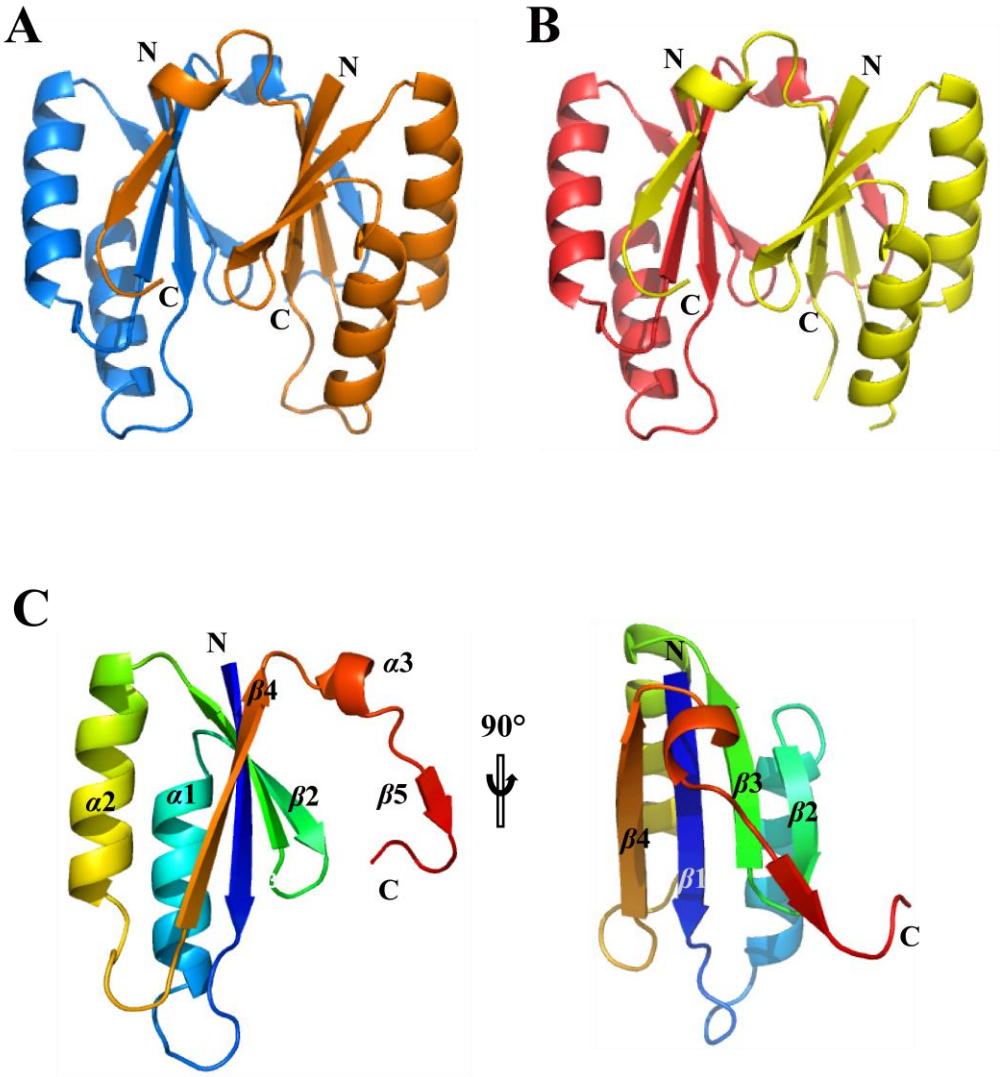
[a] $R_{merge} = \Sigma_h \Sigma | I_i(h) - \langle I(h) \rangle | / \Sigma_h \Sigma_i I_i(h)$, where $I_i(h)$ is the intensity of an individual measurement of the reflection and $\langle I(h) \rangle$ is the mean intensity of the reflection.

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

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

Figure 7. Crystal structures of XaCas2

Molecular replacement (MR) phasing in PHASER is used to determining the structures. The crystal structures of 4℃ (A) and 20℃ (B) are shown. Deletion of Ser^{12} and Ser^{13} in crystal stricter of 20℃ is due to the lack of electron density. They have butterfly shapes with homo-dimeric position. (C) The A protomer in 4℃ are shown as monomer. XaCas2 has N-terminal ferredoxin fold (β1-α1-β2-β3-α2-β4) and C-terminal residues (short 3_{10} helix and β5).


A
N
N
C
C
B
N
N
C
C
C
N
α3
β4
α2
α1
β2
β5
C
90°
N
β3
β2
β4
β1
C

conformational differentiation of C-terminal segment although the sequences of both A and B protomers were same (Fig. 8).

The Asp^{8} residues are conserved and the distance of two conserved residues is 11.03 Å in 20℃ and 11.04 Å in 4℃ (Fig. 9). It is difficult to say that the structures of XaCas2 are active form because the distance of Asp^{8}, putative active site (Nam et al. 2012), is too far to interaction with metal ion(s). Consequentially, the two XaCas2 structures are inactive conformational states and conformational changes induced by accession of substrates or metal ion(s) would occur.

Dimer interface of XaCas2

As mentioned earlier, the fifth β-strand (β5) of one protomer across the four-stranded antiparallel β-sheet makes β-sandwich at the dimer interface. And the dimerization is stabilized by formation of β-sandwich (Fig. 10). According to PDBePISA analysis, the total buried dimerization area is ~1356 $Å^2$ in 20℃ and ~ 1353 $Å^2$ in 4℃ (Krissinel and Henrick 2007). There are 35 hydrogen bonds and 13 salt bridges between the two protomers. The significant hydrogen bonds and salt bridges contain following residues (< 3 Å contained in β-sheets): Ser^{6} (β1)- Gln^{35} (between β2 and β3) (2.66 Å), Tyr^{69} (β4)- Glu^{40} (β4) (2.66 Å), Arg^{67} (β4) - Glu^{40} (β4) (2.70 Å), Asp^{8} (β1) - Phe^{36} (between β2 and β3) (2.83 Å), Ser^{65} (β4) - Gln^{35} (between β2 and β3) (2.99 Å).

Structural comparison with other Cas2 homologues

Four Cas2 homologues were selected to comparison (BhCas2, DvCas2, SpCas2, SsCas2) as amino acid sequence identity. All Cas2 proteins belong to subtype

Figure 8. Comparison of A and B protomers in XaCas2

The A and B protomers were aligned within Cas2 proteins and conformational differentiation is exist in after fourth β-strand. Each protomer is coloured in orange (protomer A) and blue (protomer B). The C-terminal segments are highlighted in square.

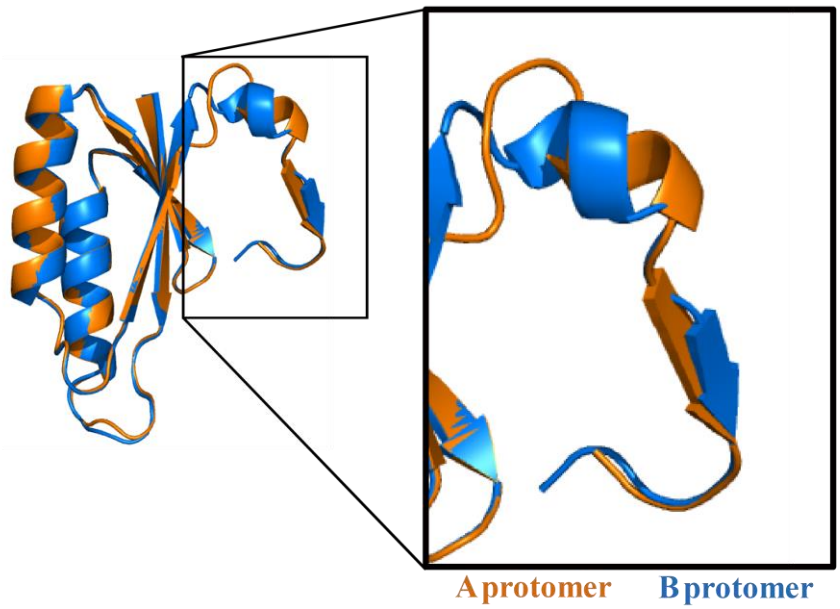

A protomer
B protomer

Figure 9. Conserved Asp and its distance in XaCas2

The dimer formation in (A) 20℃ and (B) 4℃ are shown. The distances between the two Asp^{8} residues are over 11 Å in each crystal structure.

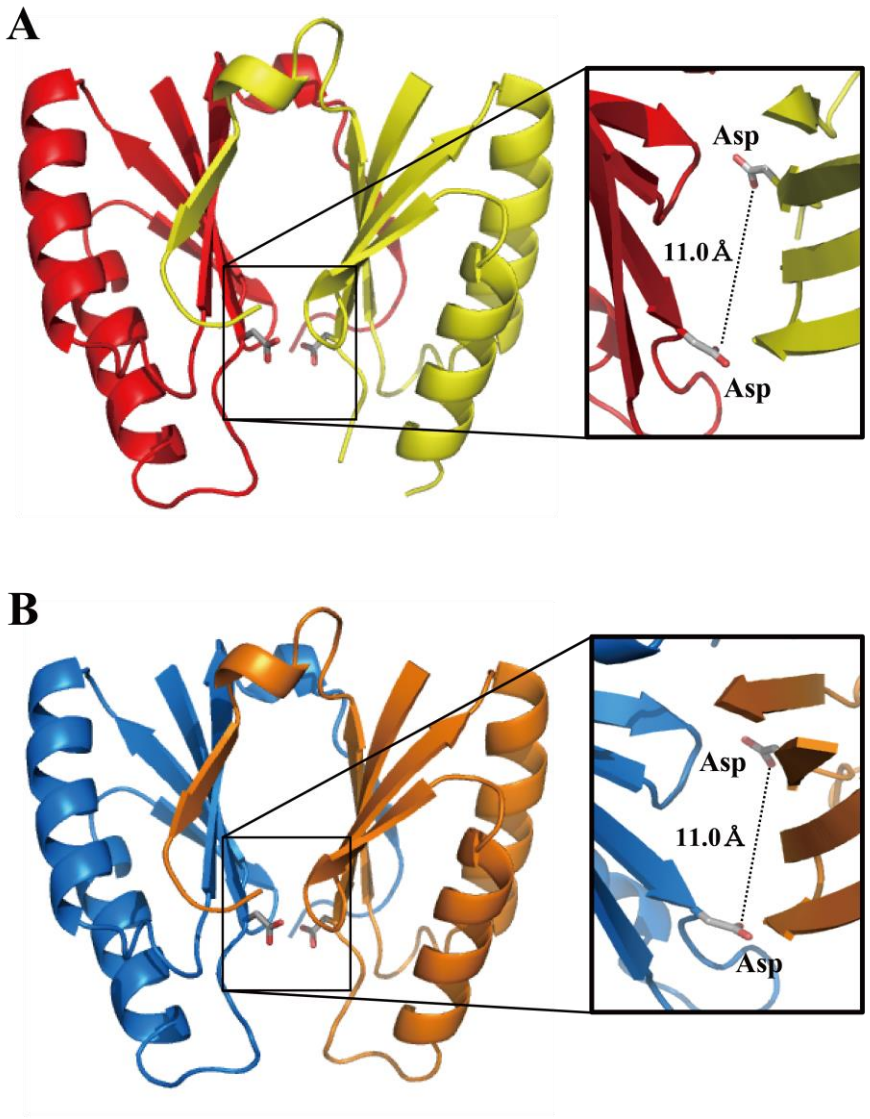
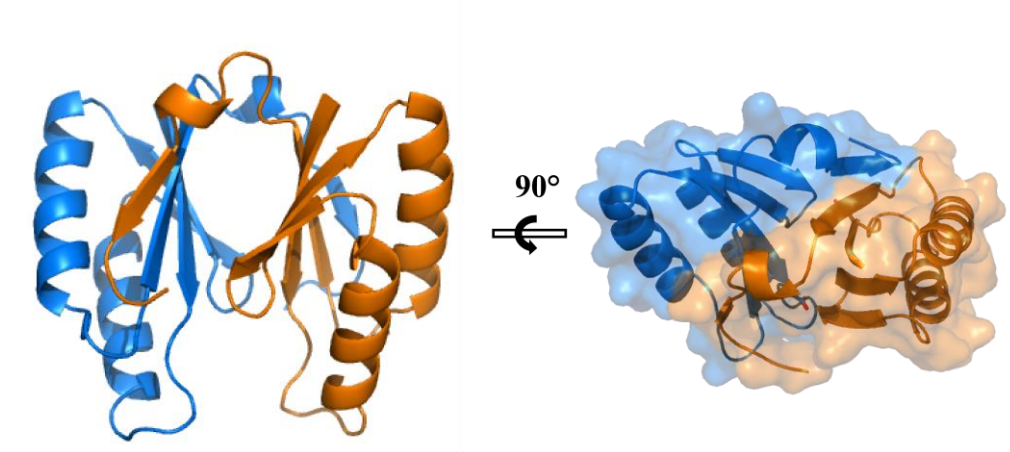

A
Asp
11.0 Å
Asp
B
Asp
11.0 Å
Asp

Figure 10. Dimerization of XaCas2

Total buried dimerization area is ~1353 Å^2 (23% of the total surface) in crystal structure of 4℃. The dimerization is stabilized by swapping of C-terminal segments.


90°

I-C CRISPR-Cas systems (Makarova et al. 2011) and sequence identity is higher than 50% except SsCas2. SsCas2 belongs to subtype I-A CRISPR-Cas systems and sequence identity is lower than 30% (Table 1). The crystal structure of 4℃ was used to comparison with other Cas2 homologues as a representative of XaCas2.

The A and B protomers were well aligned but they have different C-terminal segments although each protomer has same sequence in one dimer (Fig. 11A). Contrastively, other Cas2 homologues such as BhCas2, DvCas2, SpCas2 and SsCas2 have highly similar conformation states (Fig. 11B-E). Also all A protomers in Cas2 homologues have similar conformation states except C-terminal end and $\beta 1 - \alpha 1$ loop (Fig. 12A and B). Differences of C - terminal end are responsible for conformational changes. As I mentioned earlier, C-terminal end of one promoter has interactions with four stranded antiparallel β-sheet of another protomer, making five-stranded antiparallel β-sheets. The variations of C-terminal of Cas2 proteins may lead to change of structural features. These differences provide a basis for the hypothesis that conformational changes induced by substrates are occur.

XaCas2 is similar with SpCas2 in overall structure (Fig. 13A). The RMSD of Cα between XaCas2 and SpCas2 is 1.2 Å. The SpCas2 has N-terminal $\beta\alpha\beta\beta\alpha\beta$ ferredoxin, C-terminal segment and Aps^8 pairs. And distance of two Aps^8 residues in SpCas2 is 11.4 Å. SpCas2 is also considered as inactive conformational state (Ka et al. 2014). BhCas2 has same conformational state with XaCas2 and there were no specific structural changes (RMSD of Cα is 1.3 Å). The distance of conserved two Aps^8 residues is 10.6 Å. However, there is

Figure 11. Structural comparison within Cas2 proteins

The A and B protomers were aligned within Cas2 proteins and conformational differentiation is exist in after fourth β-strand in (A) XaCas2 (A: blue, B: gray). But other Cas2 homologous ((B) SpCas2 (A: green, B: gray), (C) BhCas2 (A: yellow, B: gray), (D) DvCas2 (A: purple, B: gray) and (E) SsCas2 (A: Red, B: gray) have same conformational state in C-terminal segment.

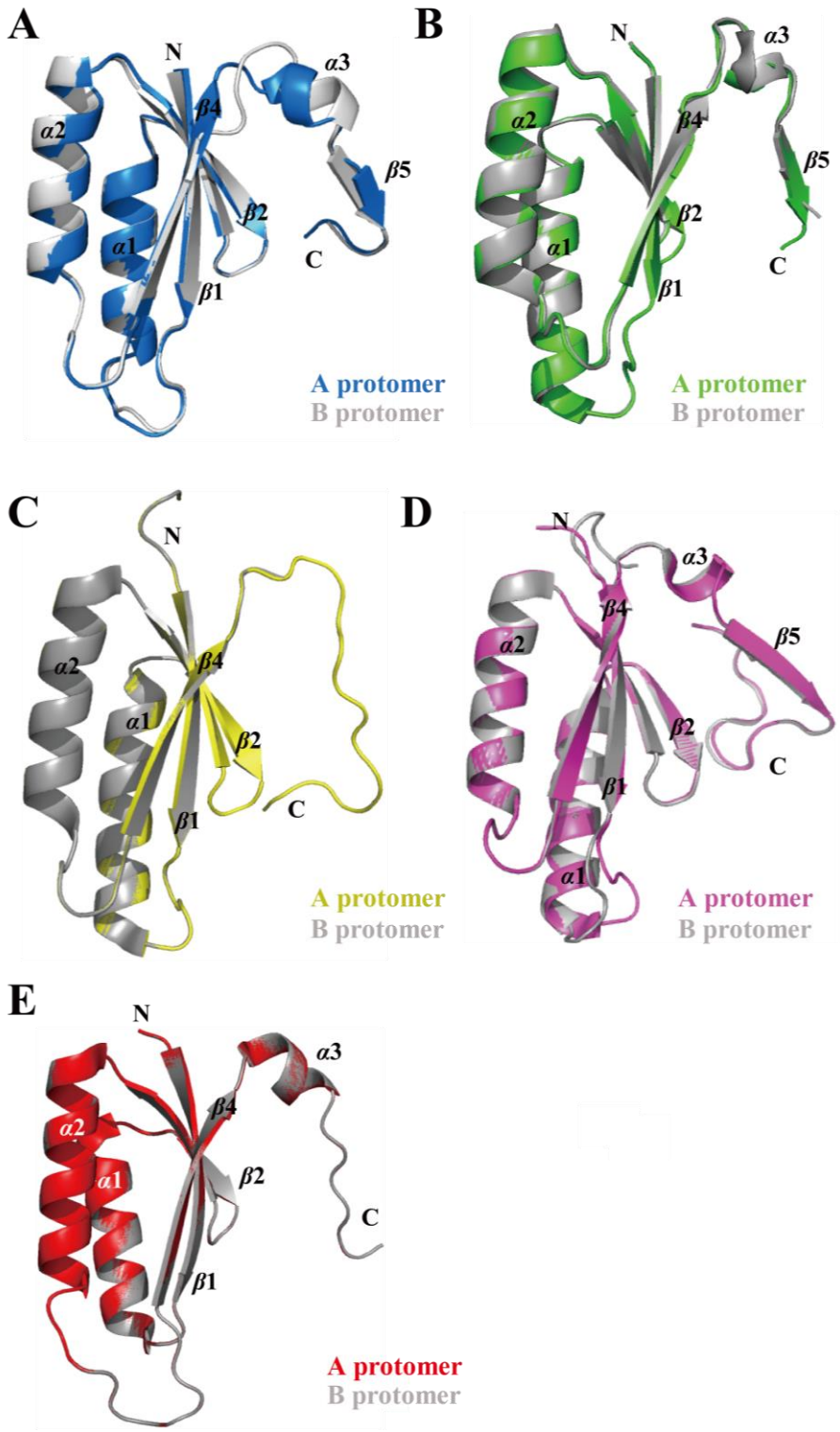
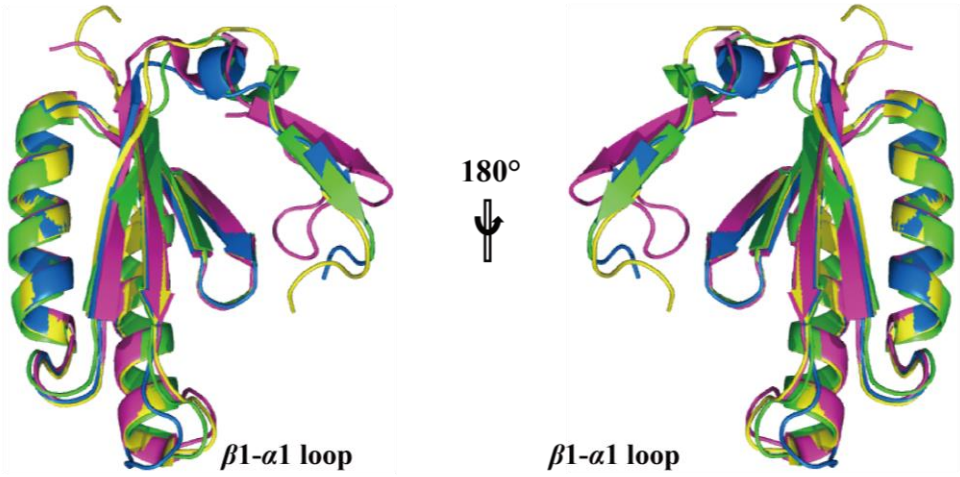

A
N
α3
β4
α2
β5
β2
α1
C
β1
A protomer
B protomer
B
N
α3
α2
β4
β5
β2
α1
C
β1
A protomer
B protomer
C
N
α2
β4
α1
β2
C
β1
A protomer
B protomer
D
N
α3
β4
α2
β5
β2
C
β1
α1
A protomer
B protomer
E
N
α3
β4
α2
β2
α1
C
β1
A protomer
B protomer

Figure 12. Sequence alignment and structural comparison of Cas2 homologues with A protomers

(A) Superposition of SpCas2 (green), BhCas2 (yellow) and DvCas2 (purple) with XaCas2 (blue) is shown. The A protomer is used to comparison. All A protomers in Cas have similar structure except C terminal end and $\beta 1 - \alpha 1$ loop. (B) Sequence alignment of Cas2 homologues. The secondary structures are indicated based on BhCas2 using ESPript3.0 program.

A


180°
β1-α1 loop
β1-α1 loop
XaCas2 SpCas2 BhCas2 DvCas2

B

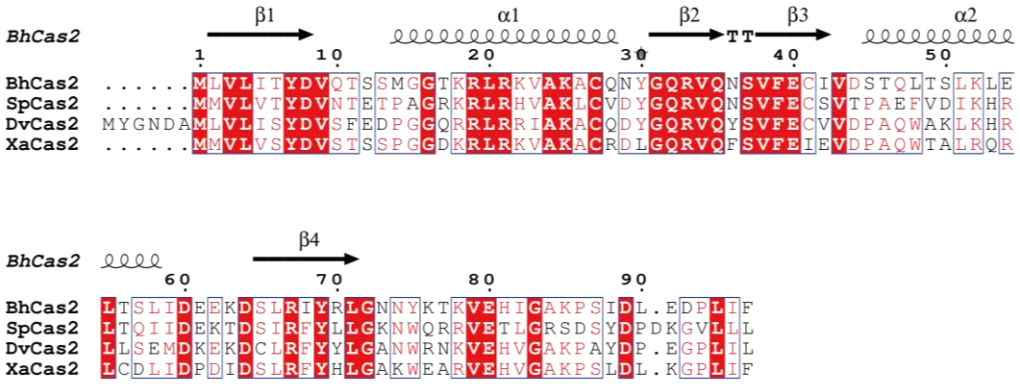
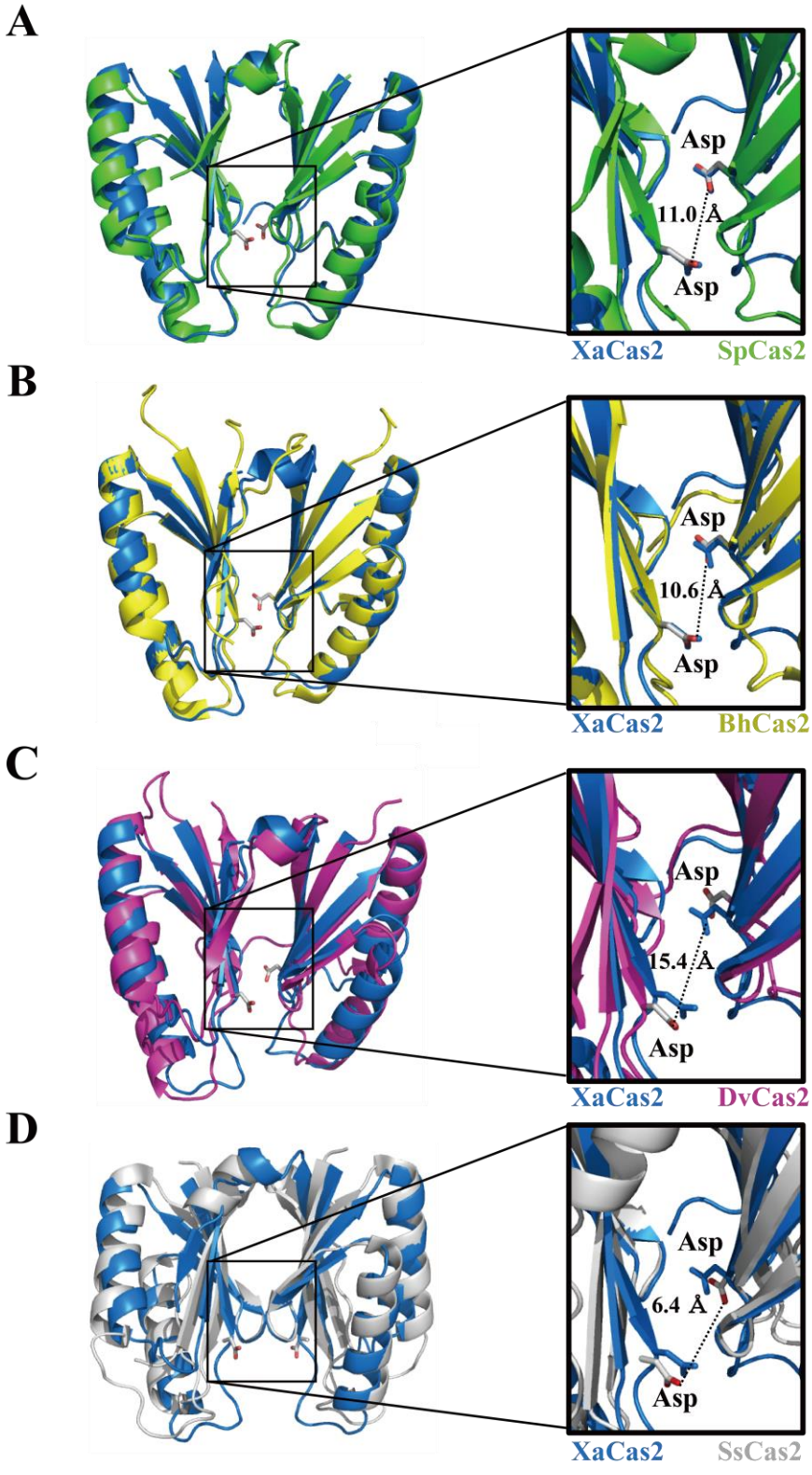

BhCas2
β1
α1
β2
β3
α2
1 10 20 30 40 50
BhCas2 ......MLVLITYDVQTSSMGGTKRLRKVAKACQNYGQRVQNSVFECIVDSTQLTSLKLE
SpCas2 ......MMVLVTYDVNTETPAGRKRLRHVAKLCVDYGQRVQNSVFECSVTPAEFVDIKHR
DvCas2 MYGNDAMLVLISYDVSFEDPGGQRRLRRIAKACQDYGQRVQYSVFECVVDPAQWAKLKHR
XaCas2 ......MMVLVSYDVSTSSPGGDKRLRKVAKACRDLGQRVQFSVFEIEVDPAQWTALRQR
BhCas2
β4
60 70 80 90
BhCas2 LTSLIDEEKDSLRIYRLGNNYKTKVEHIGAKPSIDL.EDPLIF
SpCas2 LTQIIDEKTDSIRFYLLGKNWQRRVETLGRSDSYDPDKGVLLL
DvCas2 LLSEMDKEKDCLRFYYLGANWRNKVEHVGAKPAYDP.EGPLIL
XaCas2 LCDLIDPDIDSLRFYHLGAKWEARVEHVGAKPSLDL.KGPLIF

Figure 13. Structural comparison of Cas2 homologues

Superposition of (A) SpCas2 (green), (B) BhCas2 (yellow), (C) DvCas2 (purple) and (D) SsCas2 (gray) with XaCas2 (blue). The proteins were well aligned except loop and C-terminal end. These differences might be caused by compositional changes as their origins. The distance between the two Asp^{8} residues are shown in square boxes; 11.4 Å in SpCas1, 10.6 Å in BhCa2 and 15.4 Å in DvCas1.


A
Asp
11.0 Å
Asp
XaCas2
SpCas2
B
Asp
10.6 Å
Asp
XaCas2
BhCas2
C
Asp
15.4 Å
Asp
XaCas2
DvCas2
D
Asp
6.4 Å
Asp
XaCas2
SsCas2

slight difference between β4-strand and β5-strand in XaCas2. XaCas2 has a short 3_{10} helix after β4-strand but the helix did not exist in BhCas2 (Fig. 13B). And it is proposed that the BhCas2 has more flexible C-terminal segments.

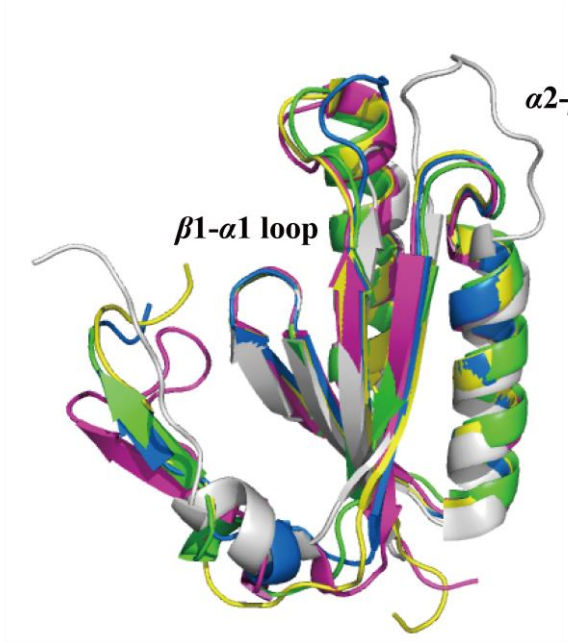
Distance of the Asp^{14} residues in DvCas2 is 15.4 Å and it is the longest distance of all Cas2 proteins (Fig. 13C). The RMSD value of Cα atoms is 1.4 Å. The five-stranded antiparallel β –sheets of DvCas2 is more twisted than XaCas2. This contortion makes the distance of Asp pairs further than other Cas2 homologues proteins.

The structural difference is notable (RMSD of Cα is 2.39) comparing with SsCas2. And distance of two conserved Aps^{10} residues is 6.4 Å (Fig. 13D). These variation of distances also make possibility of substrate induced conformational changes. The large conformational differences were found in the α2 - β4 and β1 – α1 loop. The length β1 – α1 loop is longer and α2 - β4 loop is shorter than SsCas2 (Fig. 14). As previous study, SsCas2 has ssRNA-specific endoribonucleases and the α2 - β4 loop may has responsibility for recognition of ssRNA substrates (Beloglazova et al. 2008).

All three A protomers in Cas2 proteins including BhCas2, SpCas2 and XaCas2 have similar structure excluding SsCas2 (Fig. 14). And Nam et al.,(2012) suggested that these distinctions would make the differences of width in binding site (Nam et al. 2012). The long α2 - β4 loop SsCas2 could make deep and narrow substrate binding groove and makes it possible to bind ssRNA substrates. BhCas2 has longer and more flexible β1 – α1 loop that makes a wider binding groove for dsDNA binding. And BhCas2 and SpCas2 actually

Figure 14. Structural comparison of loops in Cas2 homologues

Each Cas2 protein are coloured in green (SpCas2) yellow (BhCas2) purple (DvCas2) gray (SsCas2) and blue (XaCas2). The length of α2 - β4 and β1 – α1 loop are different. SsCas2 has longer α2 - β4 loop while β1 – α1 loop is shorter than other Cas2 homologues. Subtype specific substrate binding would be influenced by these differences of loops.



XaCas2 SpCas2 BhCas2 DvCas2 SsCas2

Although conformational states of XaCas2 are different in $\beta 1 - \alpha 1$ loop with Bhcas2 and Spcas2, lengths are as same as other proteins. The features of loops and overall structure in XaCas2 are highly similar with BhCas2, meaning XaCas2 might bind and cleaves the dsDNAs.

Discussion

Cas2 proteins are considered as universal and important proteins of CRISPR-Cas systems. The structure of Cas2 proteins are well known in various subtypes but the studies for the functions are still insufficient. Cas2 proteins have DNA or RNA nuclease activity but these activities are not essential for acquisition step (Nam et al. 2012, Beloglazova et al. 2008, Ka et al. 2014). Cas2 may also have another function on other steps in CRISPR-Cas system or metabolic activity. Cas1 has been known as metal dependent endonuclease protein. (Nuñez et al. 2014, Wiedenheft et al. 2009) And the nuclease activities are different from among subtype (Ka et al. 2014, Beloglazova et al. 2008). And a crystal structure that complex of Cas1 and Cas2 protein with dsDNA strand substrate from *E. coli* is determined in 2014 (Nuñez et al. 2014). However, it could not explain how the complex recognizes the substrate and cut foreign nucleic acid into small nucleic acid segments. Crystal structure of complex is not sufficient to applying in subtype I-C due to the *E.coli* has subtype I-E.

This work is the first step of researching acquisition step in Subtype I-C. I identified structures of Cas2 from *Xanthomonas albilineans*. *X. albilineans* has Subtype I-C and XaCas2 is a butterfly shaped protein with homo-dimeric assembly like other Cas2 homologues. The crystal structures of XaCas2 in different two temperature are same. It is determined that a temperature is not effect on conformational states. Both two crystal structures have conserved Asp^{8} at end of first β-strand, N-terminal ferredoxin fold and C-terminal segments. The dimerization is stabilized by swapping of C-terminal segments.

There was conformational differentiation of C-terminal segment between A and B protomers in one dimer structure. Also, no metal ion(s) were observed near neither of the conserved Asp^{8} residues and their distance is far to interaction with metal ion(s). Synthetically, structures of XaCas2 are regarded as inactive form. The conformational change caused by substrates or metal ion(s) would be appear temporarily due to the energetically instability in the absence of substrate.

According to previous studies, metal dependent nuclease activity of BhCas2 and SpCas2 effects on dsDNA and both Cas2 proteins were belongs to subtype I-C CRISPR-Cas systems (Ka et al. 2014, Nam et al. 2012). Metal dependent nuclease activity of BhCas2 and SpCas2 was detected over pH 7 (Nam et al. 2012, Ka et al. 2014). However, the conditions of BhCas2 and SpCas2 where crystallized were not over pH than 7. And all distances of Asp pairs are too far to interact with metal ion(s) in each Cas2 proteins. Although the crystal structures were confirmed as inactive forms, the Cas2 proteins have metal dependent nuclease activity. XaCas2 would has metal dependent nuclease activity like other Cas2 proteins despite crystal structures of XaCas2 were inactive form.

Contrastively, *S. solfataricus* has subtype I-A CRISPR-Cas systems and SsCas2 has metal dependent nuclease activity against RNA (Beloglazova et al. 2008). It is suggesting that the choice of substrate might be subtype specific. The lengths of loops above the active site pocket (α2 - β4 loop and β1 – α1 loop) in XaCas2 are differ from loops of SsCas2. The α2 - β4 loop in SsCas2, containing the important residues for ssRNA recognition (Beloglazova et al. 2008, Callaghan et al. 2005), is definitely longer than in XaCas2. Contrastively,

α2 - β4 loop in XaCas2 is longer and more flexible than loop in. If the length of loops were responsible to choosing substrate, it is important to know what residues would be needed. On the basis of structural analysis, XaCas2 more resembles in BhCas2 and SpCas2 than SsCas2. Consequentially, XaCas2 likely function as nuclease against dsDNA like BhCas2 and SpCas2. The overall structures of Cas2 homologues are similar with each other although the origin of Cas2 proteins are different. And the conserved aspartate residues in end of each β1strand in all Cas2 proteins are considered as key residues of activity.

To identify the supposition, nuclease activity assays of XaCas2 against DNA or RNA is needed. And the structures containing metal ions with substrates should be determined. As the nuclease activity is effected by pH and metal ions, the structures of XaCas2 should defined on various pH. Making mutation of on loops by adding or deletion of amino acid will solve the hypothesis.

Kwon AR et al., (Kwon et al. 2012) also support the supposition that the conserved Asp pairs are important to activity. They found that a VapD protein, a virulence-associated protein, is similar to Cas2 proteins. VapD has similar structure with Cas2 proteins such as features of dimerization and N-terminal ferredoxin fold. And it also has endonuclease activity to RNA. They found the conserved residues including Asp^{7} in β1-strand on each monomer by structural based sequence alignment with other VapD homologues. They speculate that Asp^{7} pairs are related to catalytic activity and confirmed it by making mutants. Although VapD is different from Cas2, VapD has similar crystal structures, conformational states and putative catalytic conserved Asp residues. This research assist that XaCas2 has nuclease activity and Asp^{8} pairs are important.

Determination of crystal structures with specific metal ion(s) in active site are necessary to find accurate active site and confirm the conformational change occurred by metal ions.

Determining the structures of Cas2 along or complex with Cas1 helps us to understanding mechanism of acquisition step. We can predict putative active site and binding site by analyzing structure and figure out how acquisition step works. It is still uncertain that Aps residues are active site of metal dependent nuclease activity and conformational changes are motivated by substrates. To confirm this hypothesis, metal dependent nuclease activity test and determining structures of Cas2 with metal ions or substrate are required. Making mutants that mutate putative active site is also help understanding how Cas2 works.

As additional study, solving structures of Cas2 with metal ions or substrate by NMR methods help to understanding conformational states. We can observe conformational changes of proteins by NMR methods because proteins were exist in solvent. A real time movement Cas2 protein will observed using PAL-XFEL in Pohang accelerator laboratory (Pohang, Korea). It is more powerful X-ray beam than previous X-ray beam that can catch movement of atoms in substance. The ways how cleaves substrates and make complex with Cas1 is confirmed as using these methods.

References

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M et al. (2012) Towards automated crystallographic structure refinement with phenix. refine. *Acta Crystallogr D* 68, 352-367.

Banga SS and Boyd JB (1992) Oligonucleotide-directed site-specific mutagenesis in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 89, 1735-1739.

Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M et al. (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem* 283, 20361-20371.

Beumer KJ, Trautman JK, Mukherjee K and Carroll D (2013) Donor DNA utilization during gene targeting with zinc-finger nucleases. *Genes Res* 3, 657-664.

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

Callaghan AJ, Marcaida MJ, Stead JA, McDowall KJ, Scott WG and Luisi BF (2005) Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature* 437, 1187-1191.

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. *Gene Dev* 22, 3489-3496.

Cass SD, Haas KA, Stoll B, Alkhnbashi OS, Sharma K, Urlaub H et al. (2015) The role of Cas8 in type I CRISPR interference. *Bioscience Rep* 35, e00197.

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

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

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

IshinoY, Shinagawa H, Makino K, Amemura M and Nakata A (1987) Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli* and identification of the gene product. *J bacterriol* 169, 5429-5433.

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

Jackson R, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ et al. (2014) Crystal structure of the CRISPR RNA–guided surveillance complex from *Escherichia coli. Science* 345, 1473-1479.

Koo Y, Ka D, Kim EJ, Suh N and Bae E (2013) Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *J Mol Biol* 425, 3799-3810.

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

Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol biol* 372, 774-797.

Kwon AR, Kim JH, Park SJ, Lee KY, Minn YH, Im H et al. (2012) Structural and

biochemical characterization of HP0315 from *Helicobacter pylori* as a VapD protein with an endoribonuclease activity. *Nucleic Acids Res* gkr1305.

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

Makarova KS and Koonin EV (2015a) Annotation and classification of CRISPR-Cas systems. *CRISPR: Methods and Protocols*. Humana Press, USA.

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

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

Mojica FJ, Díez-Villaseñor C, Soria E and Juez G (2000) Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol Microbiol* 36, 244-246.

Mojica FJ, M García-Martínez J and Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60, 174-182.

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

Nuñez 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 Biol* 21, 528-534.

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

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA and Liu DR (2013) High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 31, 839-843.

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, 653-663.

Samai P, Smith P and Shuman S (2010) Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr F* 66, 1552-1556.

Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B et al. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc Natl Acad Sci USA* 108, 10098-10103.

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, 904-912.

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 USA* 108, 10092-10097.

Yang H, Wang H, Shivalila CS, Cheng AW, Shi L and Jaenisch R (2013) One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* 154, 1370-1379.

Yosef I, Goren MG and Qimron U (2012) Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic acids Res* gks216.

Zhang J, Kasciukovic T and White MF (2012) The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS One* 7 e47232.

Abstract in Korean

*Xanthomonas albilineans*에서 유래된 CRISPR-associated Protein Cas2의 결정 구조

서울대학교 대학원

농생명공학부 응용생명화학전공

정유진

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) 시스템은 외부로부터 유입된 파지 (phage) 또는 플라스미드 (plasmid) 등에 대항하는 고세균과 세균에서 발견되는 방어기작이다. Cas (CRISPR associated) 단백질들은 면역 반응 기관으로서 RNA를 기반으로 작용된다. 이는 후천적으로 얻은 면역력의 세대 계승이 가능하다. 모든 CRISPR 시스템은 Cas1과 Cas2를 가진다. Cas2는 보편적으로 보존되어 있는 단백질이며 새로운 스페이서 (spacers)의 유입에 관여한다. 단백질 결정화를 통해 몇 몇의 Cas2 단백질의 구조가 규명되었다 그 예로서, *Bacillus halodurans* (BhCas2) *Desulfovibrio vulgaris* (DvCas2), 그리고 *Streptococcus pyogenes* (SpCas2) 이 존재한다. 모든 Cas2 단백질은 N 말단에 ferredoxin fold (*βαββαβ*)와 C 말단에 짧은 3_{10} helix와 다섯 번째 *β*-strand를 가진다. 본 논문에서 이용된 *Xanthomonas albilineans*는는 Subtype Ⅰ-C의 CRISPR를 가지고 있으며 Cas1, Cas2, Cas3, Cas4, Cas5d, Cas7, Cas8의 Cas 단백질로 이루어져 있다. 본 논문에서는 *Xanthomonas albilineans*에서 유래된

Cas2 (XaCas2) 단백질의 발현, 분리 및 정제, 결정화 한 후 X-ray 결정 구보 분석법을 이용하여 단백질의 3차 구조를 규명 하였다. 정제된 단백질로 결정화를 시도 했을 때 20℃와 4℃에서 결정을 얻을 수 있었고 그 구조적인 특성은 전형적인 Cas2의 특징을 가지고 있었으며, 온도가 다르다 하더라도 그 모양이 서로 같았다. 20℃에서 자란 결정은 1.65 Å으로, 4℃에서 자란 결정은 1.75Å으로 데이터를 모을 수 있었다. XaCas2는 수소결합 (hydrogen bonds)과 염다리 (salt bridges) 로 다이머 (dimer)를 유지하고 있으며 다른 동족체 (homologues)와 시퀀스 (sequence)와 3차원적 구조를 비교했을 때 Cas2 단백질의 구조는 기질 (substrate)이 접근할 때 변화 할 것으로 예측된다.

주요어: Cas2, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), CRISPR associated (Cas) proteins, *Xanthomonas albilineans* GPE PC73, x-ray crystallography

학 번: 2013-21186