농 학 석 사 학 위 논 문

Structural and functional study of *Streptococcus pyogenes* Csn2, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated protein

*Streptococcus pyogenes*의

CRISPR-associated protein Csn2의

구조결정 및 기능 연구

2012년 8월

서울대학교 대학원

농생명공학부 응용생명화학전공

구 윤

A Thesis for the Degree of MASTER OF SCIENCE

Structural and functional study of *Streptococcus pyogenes* Csn2, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated protein

Advisor: Euiyoung Bae

A dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

to the Faculty of the school of Agricultural Biotechnology
at
SEOUL NATIONAL UNIVERSITY
Seoul, Korea

by
Yoon Koo

August 2012

농 학 석 사 학 위 논 문

Structural and functional study of *Streptococcus pyogenes* Csn2, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated protein

*Streptococcus pyogenes*의 CRISPR-associated protein Csn2의 구조결정 및 기능 연구

지도교수 배 의 영

이 논문을 농학석사 학위논문으로 제출함

2012년 8월

서울대학교 대학원

농생명공학부 응용생명화학전공

구 윤

구윤의 농학석사 학위논문을 인준함

2012년 8월

위 원 장 이 상 기 (인)

부위원장 배 의 영 (인)

위 원 김 수 언 (인)

ABSTRACT

Structural and functional study of *Streptococcus pyogenes* Csn2, CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-associated protein

Yoon Koo

Major in Applied Life Chemistry

School of Agricultural Biotechnology

The Graduate School

Seoul National University

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) system functions as an adaptive immune system that protects prokaryotic cells from invading genomic nucleic acids, such as bacteriophages and conjugative plasmids.

Csn2 is a Nmeni subtype-specific Cas protein, recently classified as type Ⅱ–A, required for new spacer acquisition in the adaptation process of this immune system. In this study, the crystal structure of *Streptococcus pyogenes* Csn2 revealed

a ring-like quaternary structure including calcium binding sites. Also, the Csn2 has a non-specific double-stranded (ds-) DNA-binding activity.

In summary, the results support the suggestions of Csn2 functioning at the CRISPR adaptation stage.

Key words: CRISPR-Cas system, *Streptococcus pyogenes* SF370, Csn2, x-ray crystallography, DNA binding

***Student Number*: 2010-23441**

Contents

List of Figures

List of Tables

List of Abbreviations

BME	β-mercaptoethanol
Cas	CRISPR-associated protein
CRISPR	Clustered regularly interspaced short palindromic repeats
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium bromide
ICP-AES	Inductively coupled plasma atomic emission spectroscopy
ICP-MS	Inductively coupled plasma mass spectrometry
IMAC	Immobilized-metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
MAD	Multi-wavelength anomalous dispersion
MBP	Malotose binding protein
PDB	Protein data bank
PMSF	Phenylmethanesulfonyl fluoride
RMSD	Root-mean-square deviation
SDS-PAGE	Sodium Dodecyl Sulfate - Poly Acrylamide Gel
SeMet	Selenomethionine
S. pyogenes	*Streptococcus pyogenes*
TEV	Tobacco etch virus

1. INTRODUCTION

Many prokaryotes have developed several defense mechanisms to resist invasions by foreign nucleic acids. The recently discovered CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) system confers inheritable and adaptive immunity against invading genetic elements, such as phages and plasmids in archaea and bacteria (1-8). CRISPR loci are found in more than 90 % of sequenced archaeal genomes and nearly 46 % of sequenced bacterial genomes, including industrially or clinically important pathogens (9-12).

A CRISPR locus consists of short, often partially palindromic, 20 to 50 base-pairs-long DNA repeat sequences, interspersed by similar length of short variable spacer sequences, some of them are identical to phages and plasmids sequences, which represent the acquired memory of immunity from the invader nucleic acids (13-16), and in general, is accompanied by linked sets of CRISPR-associated (*cas*) genes that encode Cas proteins (1). Many of Cas proteins were analyzed, and they contain motifs and/or domains that are characteristic of nucleic acid binding and processing (12).

Although the details of molecular mechanisms remain to be determined, the CRISPR-Cas system mediates the immune response via a three-stage process; adaptation, expression and interference (2, 17, 18). During the adaptation stage,

short foreign nucleic acids are integrated into the CRISPR loci as variable spacers. In the second stage, expression, the long primary transcript of CRISPR (pre-crRNA) is generated and matured into short crRNAs (17). The next, interference process, the crRNAs recognize and destruct re-invading foreign genetic elements. In all three processes, the Cas proteins are involved (17).

According to the recently updated CRISPR-Cas systems classification, three different types (TypeⅠ, Ⅱ and Ⅲ) exist, which differ in the Cas protein composition and mechanisms of immune response (17). These all three types of immune systems contain two universal genes; *cas1*, encoding a DNAse (19, 20), and *cas2*, encoding an endoribonuclease (21). Csn2 is a Cas protein classified in typeⅡ-A (a Nmeni subtype in the old classification) comprising four Cas proteins (17). These four Cas proteins include two core Cas proteins (Cas1 and Cas2), and two subtype-specific Cas proteins (Csn1 and Csn2). Previous studies on this subtype CRISPR-Cas systems in *Streptococcus thermophilus* and *Streptococcus pyogenes* indicated that Csn1, as known as Cas9 (17), is the only Cas protein required for the expression and interference processes, also, suggesting that Csn2 protein is involved in adaptation process by knocking-out the corresponding cas genes (22, 23).

Here, the *Streptococcus pyogenes* Csn2 protein characterized as a double-stranded DNA binding protein and its x-ray crystal structure was determined at 2.2 Å resolution (24). The revealed structure allows for a more detailed structural analysis of *S. pyogenes* Csn2 protein. The nature of quaternary structure from *S. pyogenes* Csn2 explains its double-stranded DNA-binding function (24). Furthermore, calcium was bound in Csn2 but it seems not to be essential for the tetramerization of the protein since Csn2 shows that a stable tetramer from oligomeric state analysis (24). Before entering the main, I would like to make clear that Du-kyo Jung and Euiyoung Bae also have contributed for the study.

2. MATERIALS AND METHODS

2.1 Protein expression and purification

The plasmid constructed *S. pyogenes* csn2 gene cloned pHMGWA vector containing a $(His)_6$-malotose binding protein (MBP) tag and a tobacco etch virus (TEV) protease cleavage site (25) was used to transform *Escherichia coli* BL21 (DE3) cells. The transformed cells were grown in LB medium at a temperature of 37 ℃ until the OD_{600} reached 0.6 and then protein expression was induced by adding a final concentration of 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Following induction, the cells were grown at 17 ℃ for 18 hours and harvested by centrifugation.

The harvested cells were resuspended in lysis buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 20 % (w/v) glycerol, 5 mM β-mercaptoethanol (BME), 0.1 % (v/v) Triton X-100, 10 mM imidazole, 0.25 mM phenylmethanesulfonyl fluoride) and lysed by sonication. The lysate was clarified by centrifugation and the supernatant filtered through 0.45 μm filter. The filtered supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare, USA) pre-equilibrated with running buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 20 % (w/v) glycerol, 5 mM BME, 20 mM imidazole). After the column was washed with 10 column volumes (CV) of running buffer, the bound protein was eluted by a linear gradient of imidazole up to 500 mM.

To remove the N-terminal $(His)_6$-MBP tag, the fractions containing the bound protein was pooled with TEV protease and pour in 10 K membrane to dialyze against TEV proteolysis buffer (20 mM sodium phosphate pH 7.4, 500 mM NaCl, 20 % (w/v) glycerol, 5 mM BME). The untagged-Csn2 protein and the $(His)_6$-MBP tag were separated on another HisTrap HP column. Fractions containing Csn2 protein were pooled and further purified on a HiLoad 16/60 Superdex200 column (GE Healthcare, USA) equilibrated with sizing column buffer (20 mM HEPES pH 7.5, 500 mM KCl, 2 mM DTT, 5 % (w/v) glycerol).

2.2 Expression and purification of selenomethionyl protein

For crystal structure determination, selenomethionine-labelled *S. pyogenes* Csn2 protein was expressed in *E. coli* BL21 (DE3) cells and purified.

Cells from overnight culture grown in LB medium were isolated by centrifugation at 5000 rpm for 2min and resuspended in M9 minimal medium with the same volume of removed LB medium. Inoculate 2 % of the resuspended cells to large volume of the same M9 medium. Cells were grown at 37 °C and when the OD_{600} reached 0.6, 50x amino acid mixture (containing Lys, Thr, Phe, Leu, Ile, Val) was supplemented into the culture. After a 30-min incubation, a final concentration of 0.25 mM selenomethionine (SeMet) and 0.5 mM IPTG were added, and then cells were grown at 17 °C for 12 hours (26).

Purification of selenomethionyl *S. pyogenes* Csn2 was identical with those for

native Csn2 protein.

2.3 Mass spectrometry

MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometry was used to confirm the selenomethionine-labeling of *S. pyogenes* Csn2 protein.

The protein solution sample was prepared 3 mg/mL concentration in a 500 μL of the protein buffer containing 100 mM HEPES pH 7.0, 150 mM KCl, 4 mM DTT. The native Csn2 at the same condition was used as a negative control of selenomethionine-labeling.

MALDI-TOF mass spectrometry was performed at the National Center for Inter-University Research Facilities (NCIRF).

2.4 Crystallization

Crystals of selenomethionyl *S. pyogenes* Csn2 were grown at 20 ℃ by the hanging drop vapor diffusion method. By mixing 15 mg/mL protein solution in buffer (100 mM HEPES pH 7.0, 150 mM KCl, 4 mM DTT) with reservoir solution (2.3 M sodium acetate pH 7.0) in a 1:1 ratio. Suitable crystals for x-ray diffraction were cryo-protected by soaking in the reservoir solution supplemented with 20 % (v/v) ethylene glycol, and flash-frozen in liquid nitrogen.

The native *S. pyogenes* Csn2 crystals were obtained from reservoir solution of

2.8 M sodium acetate pH 7.0, and treated using the same method.

2.5 Structure determination and refinement

Multiwavelength anomalous diffraction (MAD) data for the selenomethionyl *S. pyogenes* Csn2 were collected to 2.2 Å resolution at the Photon Factory (KEK, Japan) on the AR-NW12A beamline. Data sets were indexed, integrated, and scaled using HKL2000 (27). The selenium positions of selenomethionyl *S. pyogenes* Csn2 were determined using SOLVE. Subsequent density modification was performed using RESOLVE and initial electron density map was produced (28,29,30). The structure was completed by alternate cycles of manual fitting in COOT (31), refinement in REFMAC5 (32) and PHENIX suite (33) with default geometry restraints. Also, Translation Libration Screw (TLS) refinement of four groups corresponding to individual domains in the asymmetric unit was used.

Diffraction data for the native *S. pyogenes* Csn2 were collected to 2.9 Å resolution at the Pohang Accelerator Laboratory (Pohang, Korea) on beamline 6C. Data were processed with iMOSFLM (34). The structure of selenomethionyl *S. pyogenes* Csn2 was used as a starting model for molecular replacement (MR) phasing in PHASER (35). Structure building and refinement were carried out with the programs COOT (31) and REFMAC5 (32), respectively.

The final model quality was assessed using MOLPROBITY (36). Sequence alignment at Figure 2.A was performed using ClustalW (37) and ESPRIPT (38).

Hydrogen bonds were identified with PISA (39). Figures were prepared using PyMol (www.pymol.org).

2.6 Metal analysis

To reveal the identity and amount of *S. pyogenes* Csn2 bound metals, both native (3.2 mg/mL) and selenomethionyl (3.3 mg/mL) Csn2 proteins in buffer (100 mM HEPES pH 7.0, 150 mM KCl, 4 mM DTT) were analyzed and as a control, the buffer alone was also examined.

Inductively coupled plasma atomic emission spectroscopy (ICP-AES) (Ultima 2C, Jobin Yvon, France) was used for determining the concentration of calcium (Ca) in *S. pyogenes* Csn2 samples. By inductively coupled plasma mass spectrometry (ICP-MS) (Elan 6100, Perkin Elmer, USA), the concentration of magnesium (Mg), manganese (Mn), cobalt (Co) and Nickel (Ni) were measured. ICP-AES and ICP-MS were performed at the Korea Basic Science Institute.

2.7 Oligomeric state analysis

Analytical size-exclusion chromatography was performed to determine the oligomeric size of *S. pyogenes* Csn2 using a Superdex 200 10/300 GL column (GE Healthcare, USA) at 4 ℃. The column was equilibrated with the sizing column buffer containing 20mM HEPES pH 7.5, 500mM KCl, 5% (w/v) glycerol, 2mM DTT or 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 2 mM DTT and at 0.4 mL/min of

flow rate, 0.5 mL of 1.0 mg/mL *S. pyogenes* Csn2 was loaded on.

To analyze the calcium dependency of Csn2, following experiments were repeated using the sizing column buffer supplemented with 20 mM of $CaCl_2$, ethylenediamine tetraacetic acid (EDTA), or ethylene glycol tetraacetic acid (EGTA).

2.8 Electrophoretic mobility shift assay (EMSA)

Binding reactions between *S. pyogenes* Csn2 and ds-DNAs were carried out in binding buffer containing 20 mM HEPES pH 7.5, 50 mM NaCl without and with 20mM $CaCl_2$. Purified Csn2 proteins were incubated with 90 bp ds-DNA fragments and also with longer ds-DNAs at room temperature for 20 min.

Two 90 bp ds-DNAs were synthesized by using the synthesis service from Bioneer and two longer ds-DNAs were PCR amplified. One of 90 bp ds-DNAs was *S. pyogenes* CRISPR ds-DNA fragment and the other, control 1, was the promoter of *Early Responsive to Dehydration Stress* Ⅰ (*ERD*Ⅰ) ds-DNA fragment from *Arabidopsis thaliana*. For CRISPR (578 bp), the plasmid named SpCRISPR1 (*S. pyogenes* CRSIPR in pUC19) was used as a template with primers, Sp CRISPR Nmeni [F] (5'-AGATATA-GGTATAATACTCTTAATAAATGC-3') and Sp CRISPR Nmeni [R] (5'-AATTC-TTCTCAAGTTATCATCGGCAA-3'). As a result, this sequence contains 566 bp of *S. pyogenes* CRSIPR and 12 bp of extra sequence. AKmeso-ionpairs(R19K/D116R/Q198E/Q202E) was used as a template for control

2, which contains 654 bp of AKmeso-ionpairs(R19K/D116R/Q198E/Q202E) gene with extra 34 bp from pET11 vector sequence at the upper side of the gene. The control sequence has no sequence similarity with CRISPR but similar length.

All of ds-DNAs were prepared in distilled water and the Csn2 was prepared in protein buffer containing 100 mM HEPES pH 7.0, 150 mM KCl, 4 mM DTT. Then 1 μM of ds-DNAs were mixed with the same volume of 0, 1 or 10 μM of *S. pyogenes* Csn2 tetramer. Total reaction volume was 10 μL (4 μL of protein solution, 4 μL of ds-DNA solution and 2 μL of 5 x binding buffer) and all of the mixture was loaded on each lane with 10 x loading dye.

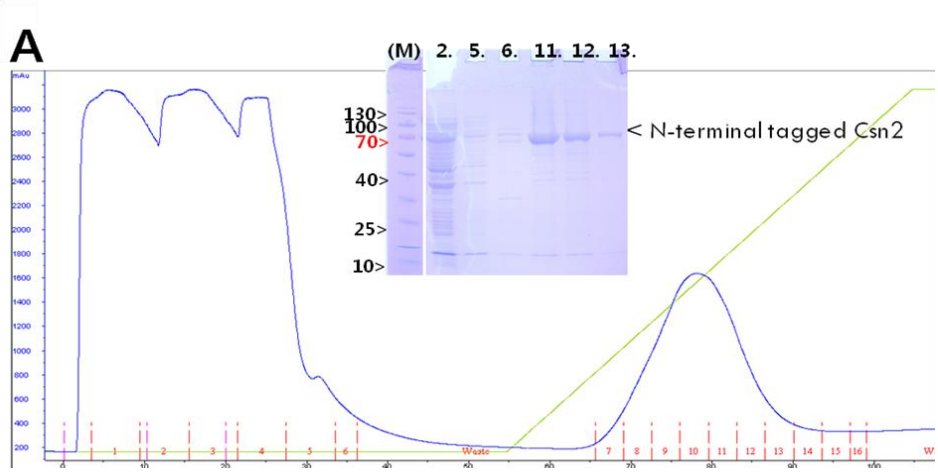
The complexes were separated on a 10 % native Tris-glycine polyacrylamide gel. 5 x Tris-glycine buffer (this buffer was made with the same composition of SDS-PAGE but free-SDS) was added in the 10 % polyacrylamide gel for one tenth, and the gel has been pre-run before separation with 0.5 x Tris-glycine buffer at 50 V for 30 to 40 min. The gel was visualized after 10 min of ethidium bromide (EtBr) staining and 10 min of washing on a rocker.

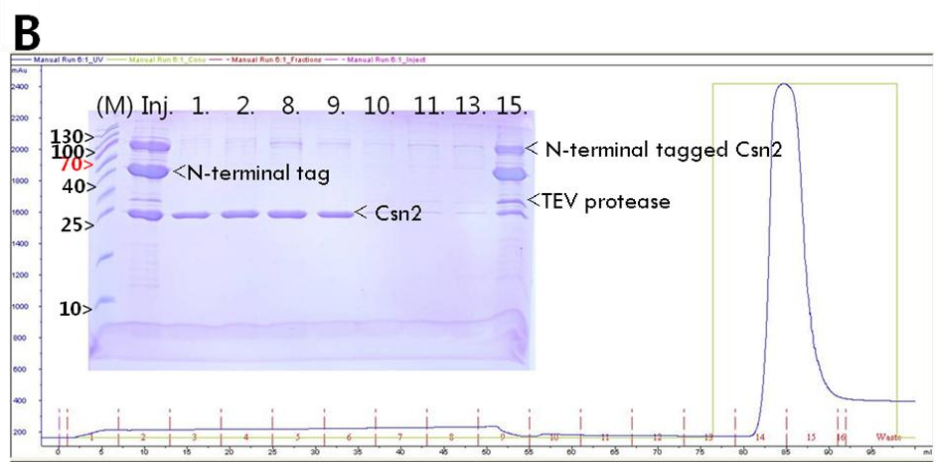
3. RESULTS

3.1 Expression, purification and crystallization of Csn2

Figuer 1 shows the elution profiles of chromatography with SDS-PAGE analysis from each purification steps. The $(His)_6$-MBP tagged-Csn2 protein (68.7 kDa) was successfully induced by IPTG. High purity of fusion protein was obtained at a linear gradient of imidazole from first IMAC (Figure 1A). After TEV protease proteolysis, the untagged-Csn2 protein (25.9 kDa), the $(His)_6$-MBP tag and TEV protease were separated on another IMAC. However, since the cleaved-tag and the protease have a $(His)_6$-tag but Csn2 has been untagged, the Csn2 protein was eluted in washing step (Figure 1B). Following size-exclusion chromatography confirmed that the Csn2 protein forms an oligomer (Figures 1C and 6).

The screen kits from Hampton research was used for the initial crystallization. In many of them have found crystals, and among them, the reservoir number 24 of Classics Ⅱ Suite, containing 2.8M sodium acetate pH 7.0, was selected to crystal optimization. The crystals of selenomethionyl and native *S. pyogenes* Csn2 were obtained each from 2.8 M sodium acetate pH 7.0 and 2.3 M sodium acetate pH 7.0 (Figure 2). By hanging-drop buffer diffusion method, these Csn2 crystals were fully grown in a day or two at 20 ℃.


A
(M) 2. 5. 6. 11. 12. 13.
130>
100>
70>
40>
25>
10>
< N-terminal tagged Csn2


B
(M) Inj. 1. 2. 8. 9. 10. 11. 13. 15.
130>
100>
70>
40>
25>
10>
<N-terminal tag
< Csn2
< N-terminal tagged Csn2
<TEV protease

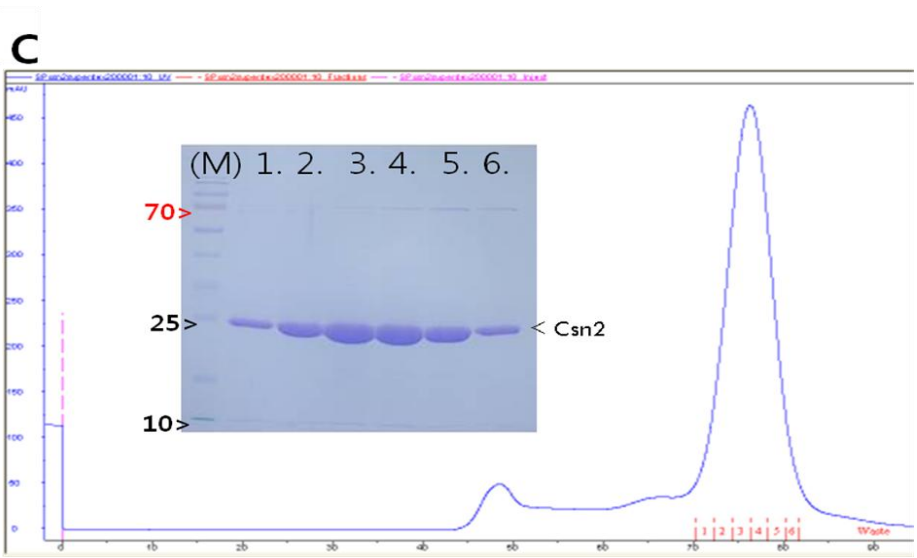

C
(M) 1. 2. 3. 4. 5. 6.
70>
25>
10>
< Csn2

Figure 1. Elution profiles with SDS-PAGE analysis of Csn2 protein from each purification step. (A) First IMAC purification using a 5 mL HisTrap HP column. The green line displays imidazole gradient. (B) After TEV proteolysis, second purification with the same IMAC column of the first purification. (C) Final size-exlusion chromatography purification using a HiLoad 16/60 Superdex200 column. The purified Csn2 protein was observed to form a tetrameric protein (100 kDa: 25.9 kDa x4). Using SDS-PAGE were 10% and the numbers of each lanes are identical to those of fractions.

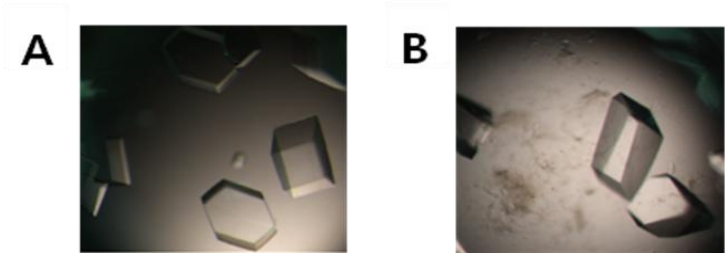


Figure 2. Crystals of the (A) selenomethionyl and (B) native *S. pyogenes* Csn2 protein.

3.2 Overall structure of *S. pyogenes* Csn2

The structure of *S. pyogenes* Csn2 was solved at a resolution of 2.2 Å using multi-wavelength anomalous diffraction (MAD) with a selenomethionine derivatized protein crystal. Data collection and structure refinement statistics are summarized in Table 1. In the asymmetric unit, the two molecules of Csn2, three calcium ions, two ethylene glycol molecules and 204 water molecules were contained. Due to insufficient electron density, several residues (protomer A: residues 40, 41, 48-50, 210-213, protomer B: residues 40, 41, 48-50, 140, 220) were not modeled into the final structure.

The two Csn2 in the asymmetric unit further assemble with a symmetry related Csn2 dimer. This results in diamond-shaped ring structure with a hole in the center with a diameter of ~30 Å. Besides, electrostatic analysis revealed the positively charged inner surface of *S. pyogenes* Csn2 ring structure (Figure 3) and the charge distribution is highly conserved among the Csn2 protein family (Figure 4). These structural features are interesting because the surface of inner ring may mediate interactions with the ds-DNA.

Each Csn2 protomer structure contains a globular α/β domain and an extended α-helical domain extruded from the α/β domains (Figure 5). The globular α/β domain (residues 1-62 and 144-219) composed of three flanking α–helices (α1,

α5, α6), a six-stranded β–sheet core(β3, β5-β9) and a three-stranded anti-parallel β-sheet (β1, β2, β4). The α-helical domain (residues 73-133) consists of three α-helices (α2-α4). These two domains are connected by two flexible hinge regions (residues 63-72 and 134-143).

The α/β domain participates in dimerization between two protomers in the same asymmetric unit by forming 9 hydrogen bonds and burying 2116 $Å^2$ of total solvent accessible surface area. But the extended α-helical domains of the two Csn2 protomers are involved in inter-subunit contacts with the α-helical domains of the symmetry-related dimer, not in internal contacts within the asymmetric unit. The interaction of the two α-helical domains forms 23 hydrogen bonds and buries 4317 $Å^2$ of total solvent accessible surface area. Buried area calculation and molecular contact analysis were carried out using the CCP4i suite (40). Hydrogen bonds were identified with PISA (39). After all, the dimerization of Csn2 dimers results in a tetramer. Also, the results of analytical size-exclusion chromatography support tetrameric structure of *S. pyogenes* Csn2 whose monomeric molecular weight is 25.9 kDa (Figure 6).

Table 1. Data collection and refinement statistics.

	SeMet Csn2			Native Csn2
Space group	$C222_1$			$C222_1$
Unit cell parameters (Å)	a=59.5, b=163.4, c=150.0			a=59.1, b=162.9, c=149.0
	Se-peak	**Se-edge**	**Se-remote**	
Wavelength (Å)	0.97910	0.97927	0.96404	1.00000
Data collection statistics				
Resolution range (Å)	50.00-2.20 (2.28-2.20)	50.00-2.20 (2.28-2.20)	50.00-2.20 (2.28-2.20)	49.68–2.90 (3.06 –2.90)
Number of reflections	36996 (3608)	37023 (3608)	37021 (3609)	16061 (2286)
Completeness (%)	98.7 (97.6)	98.7 (97.6)	98.7 (97.6)	98.4 (97.8)
R_{merge} [a]	6.8 (65.4)	6.5 (66.6)	6.0 (68.4)	10.9 (53.5)
Redundancy	13.5 (13.2)	13.5 (13.1)	13.5 (13.2)	6.6 (6.1)
Mean I/σ	31.1 (5.1)	32.8 (4.9)	36.9 (4.8)	11.7 (2.8)
Refinement statistics				
Resolution range (Å)	29.17–2.20 (2.26–2.20)			49.68–2.90
Number of reflections, total/test	35414 / 1777			15232 / 813
R_{cryst} [b]/R_{free} [c] (%)	21.5 / 25.8			21.5 / 26.7
RMSD bonds (Å)	0.008			0.017
RMSD angles (deg)	1.087			1.714
Average B factor (Å ^{2})	50.37			65.14
Number of water molecules	204			28
Ramachandran favored (%)	97.8			92.3
Ramachandran allowed (%)	2.2			7.7

Values in parentheses are for the highest-resolution shell.

[a] $R_{merge} = \Sigma_h \Sigma_i |I_i(h) - \langle I(h) \rangle| / \Sigma_h \Sigma_i I_i(h)$, ,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 5.0% of the randomly selected unique reflections that were omitted from structure refinement.

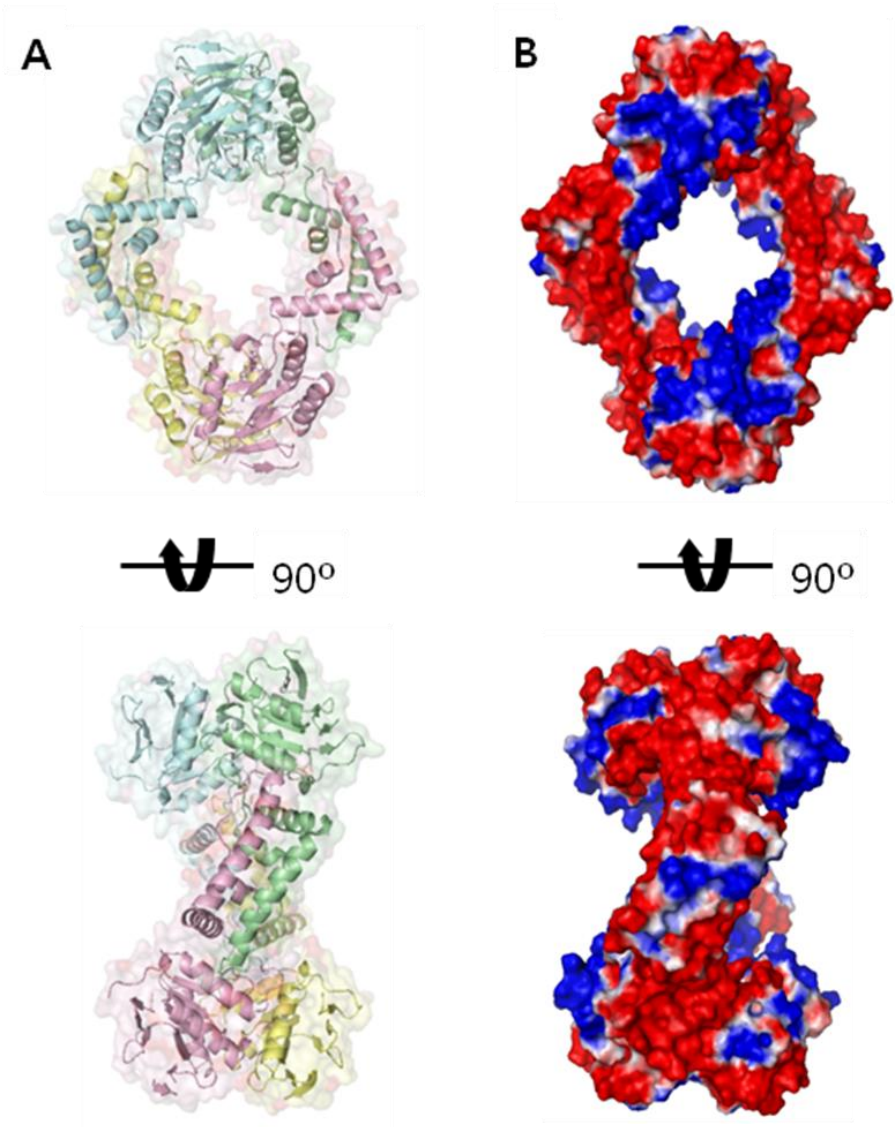


Figure 3. Overall structure of *S. pyogenes* Csn2 tetrameric ring structure. (A) The ring-like structure of tetrameric Csn2. The four protomers are colored pink, yellow, cyan and green, respectively. (B) The same views to reveal the electrostatic potential on the surface of Csn2 tetramer (red=-25 kT, blue=+25 kT). A clear division of positively charged inner and negatively charged outer portions of the tetrameric ring is observed.

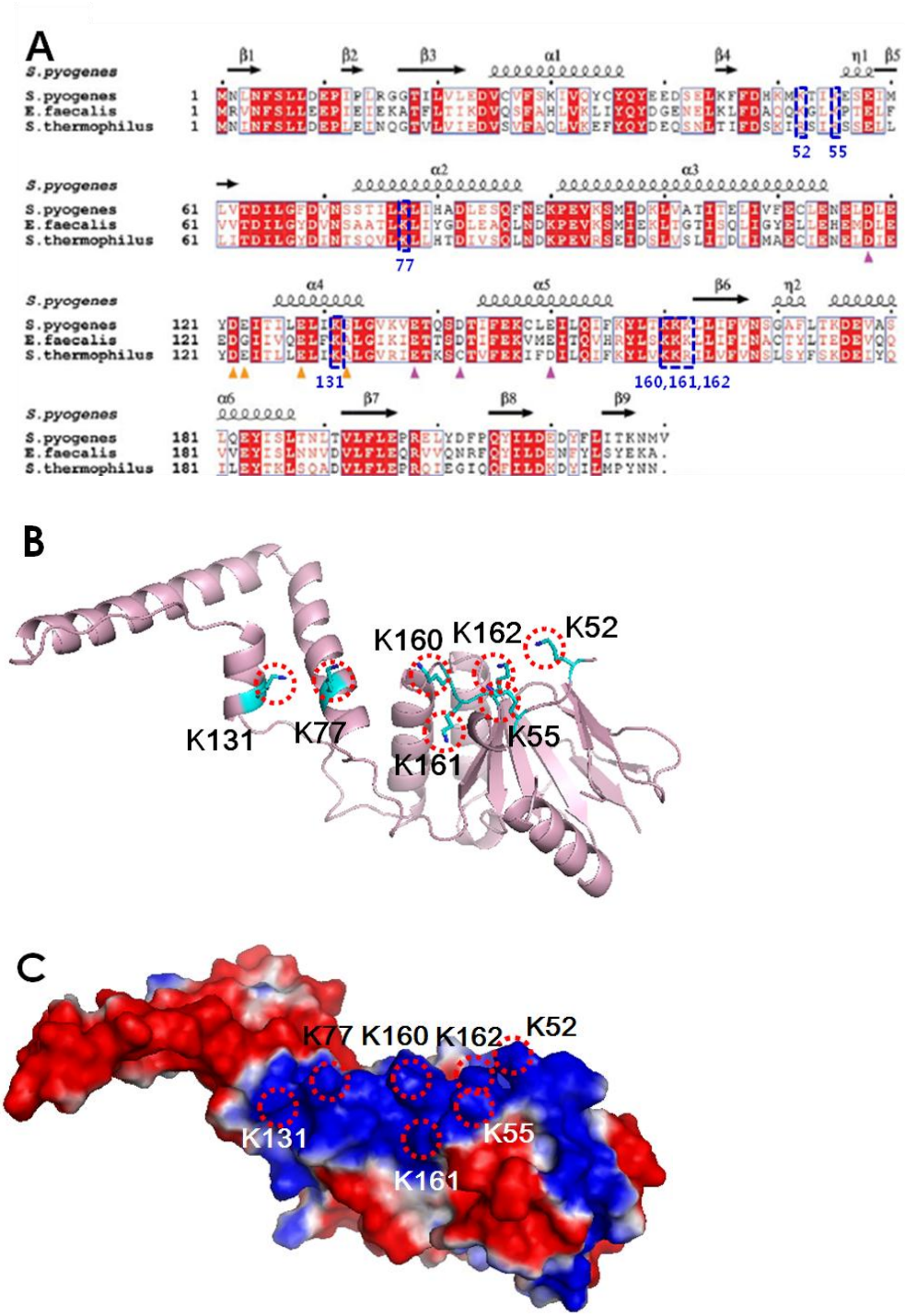


Figure 4. Sequence alignment of Csn2 homologues. (A) Sequence alignment of Csn2 homologues from *S. pyogenes*, *E. faecalis* and *S. thermophius* was performed

using ClustalW (37) and ESPRIPT (38). Based on the *S. pyogenes* Csn2 structure, secondary structures are marked. The orange and purple triangles are indicating the calcium coordinating residues within the CA1 and CA2 sites (Adapted from (24)). (B) The conserved positive charges from the lysine residues. (C) Electrostatic potential of Csn2 monomer (red=-25 kT, blue=+25 kT).

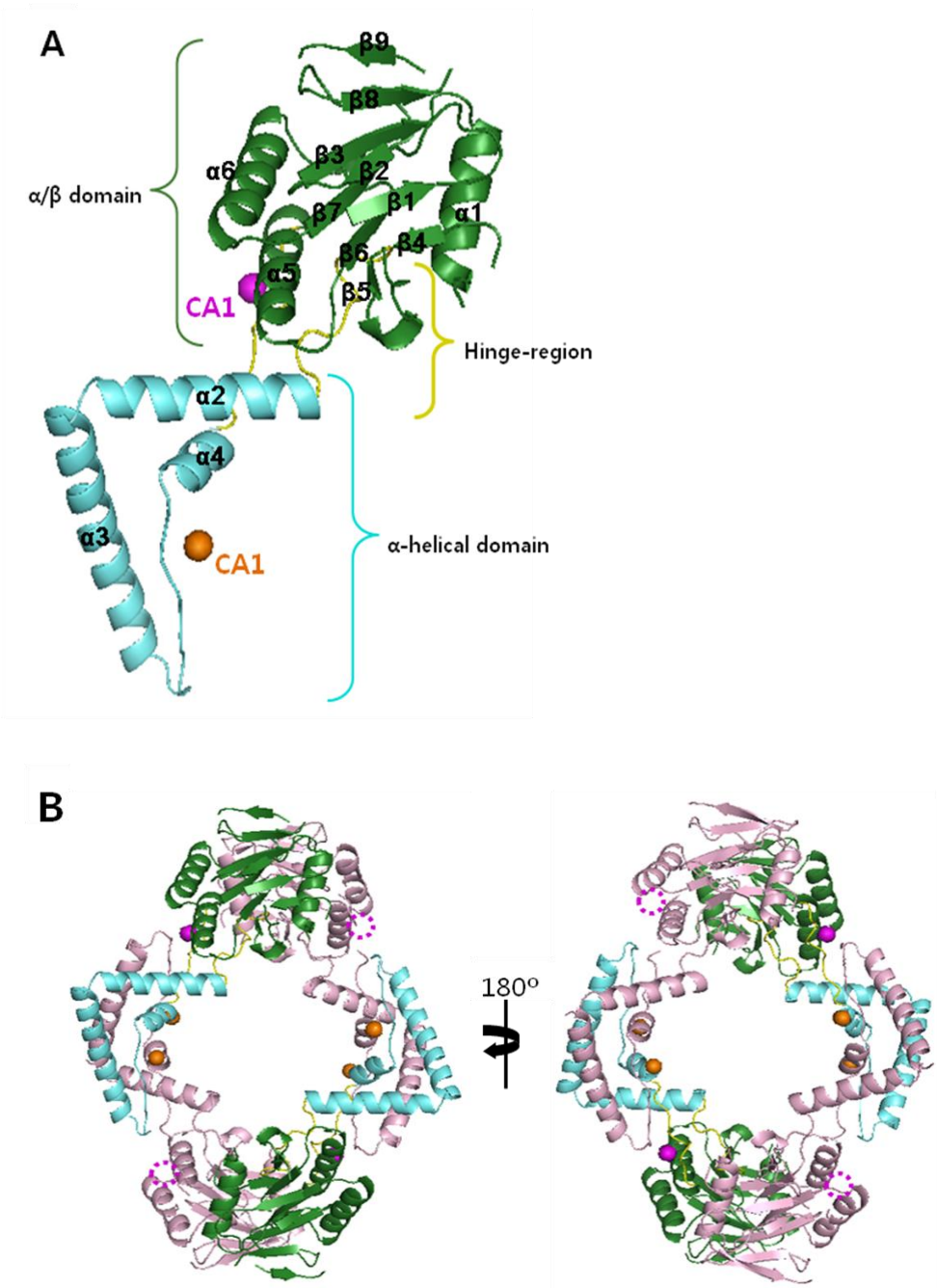


Figure 5. The monomeric structure of the *S. pyogenes* Csn2 protein. (A) The α/β domain, α-helical domain and hinge regions are colored green, cyan and yellow.

Bound calcium ions in the CA1 and CA2 sites are represented as orange and purple spheres, respectively. Secondary structures are indicated, also. (B) Tetrameric arrangement of Csn2 is viewed along the two-fold symmetry axis. Monomer A is colored as in Figure 3A, and monomer B is colored pink. Unfound CA2 sites were indicated with dotted line circles.

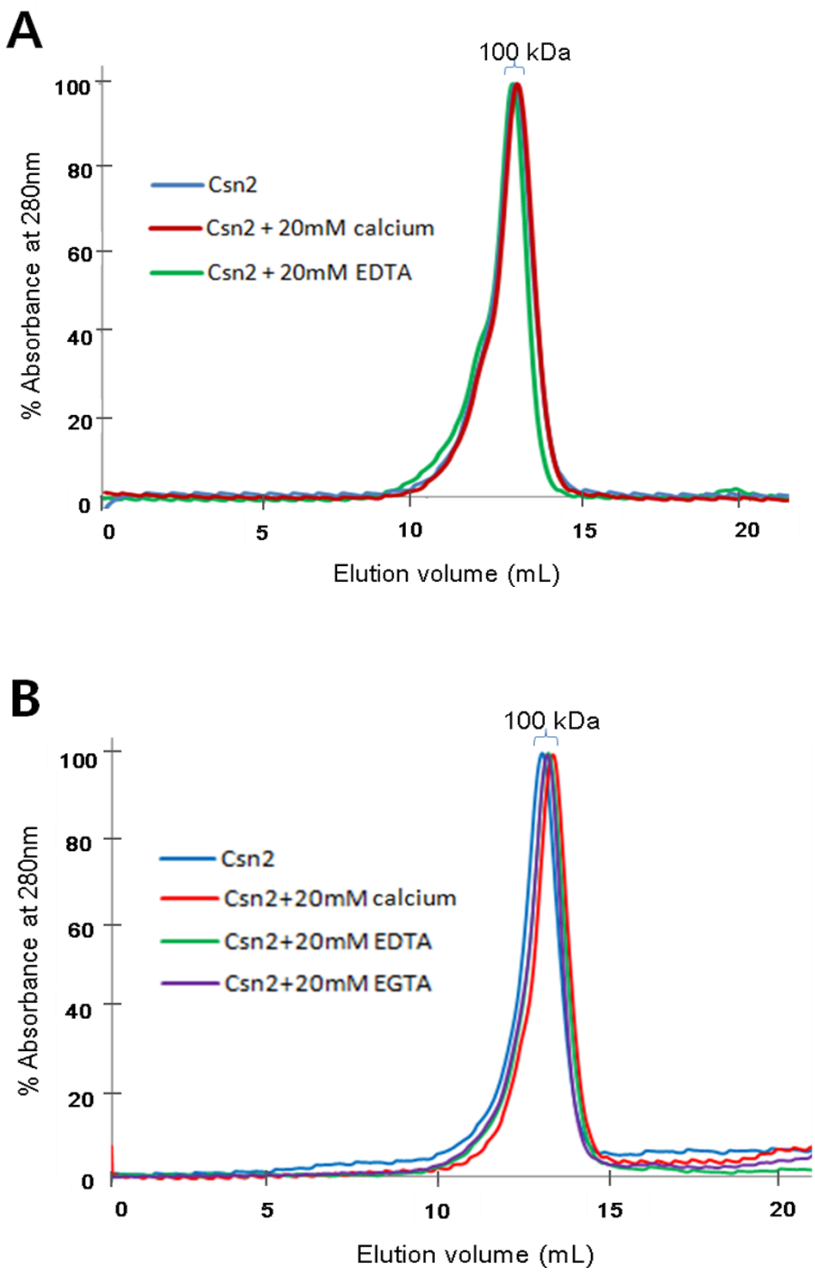


Figure 6. Analytical size-exclusion chromatography of *S. pyogenes* Csn2 showing the results with the sizing column buffer containing (A) 20mM HEPES pH 7.5, 500mM KCl, 5% (w/v) glycerol, 2mM DTT and (B) 10 mM Tris-HCl pH 8.0, 200

mM NaCl, 2 mM DTT. The blue line of upper figure covered behind the red and green line. In both of buffer condition, no difference in formation of Csn2 oligomer whether the presence of calcium ions or complete removal of the ions using EDTA or EGTA.

3.3 Calcium binding sites in Csn2 protomer

Within the asymmetric unit of Csn2 structure, three potential metal binding sites were found. To identify the bound metals, the concentrations of Ca, Mg, Mn, Co, and Ni in the protein solution samples were determined using inductively coupled plasma mass spectrometry (ICP-MS) and inductively coupled plasma atomic emission spectroscopy (ICP-AES). Among five metals, only calcium was detected in significant amounts in the samples (Table 2).

There are two different calcium-binding sites in Csn2 protomer, CA1 and CA2 (Figure 5A). However, not four calcium ions but only three were identified in the asymmetric unit of *S. pyogenes* Csn2 and in the tetrameric structure, six calcium ions were found, four in CA1 sites and two in CA2 sites (Figure 5B).

The CA1 sites are placed at the middle of the interface of interacting α-helical domains, and on those sites, the calcium ions are coordinated by four residues (Asp122, Glu123, and Glu128 from protomer A and Ser132 from protomer B) and one water molecule (Figure 7A). On the CA2 sites, there is only one calcium ion in the structure of asymmetric unit which is located adjacent to one of two hinge regions in protomer A. The calcium of CA2 site is coordinated by four amino acid residues (Glu138, Asp142, and Glu150 from protomer A and Asp118 from protomer B of the symmetry-related dimer) and two water molecules (Figure 7B).

Besides, the results of analytical size-exclusion chromatography show that *S. pyogenes* Csn2 acts as a stable tetramer in solution irrelevant to the presence of

calcium ions (Figure 6).

3.3 *S. pyogenes* Csn2 binds to ds-DNA

An electrophoretic mobility shift assay (EMSA) was performed to analyze the ds-DNA binding activity of Csn2. In the presence of *S. pyogenes* Csn2 protein, the migrations of all short or longer sequences of ds-DNAs in the native gel were slower and the more Csn2 were added in, the greater shifting was observed (Figures 8 and 9).

Reaction in the extra calcium added binding buffer doesn't show any differences from the buffer without calcium, since Csn2 has already included calcium ions as seen in its crystal structure (Figure 8). One of two 90 base-pair ds-DNAs was a section of *S. pyogenes* CRISPR that included the first repeat and spacer sequences and the other was a section of the promoter site of *Early Responsive to Dehydration Stress Ⅰ* (ERDⅠ) gene from *Arabidopsis thaliana* (Figure 8C). EMSA was also performed with the longer sequences of ds-DNAs and they all showed up shifted (Figure 9).

To summarize the results from the assay, *S. pyogenes* Csn2 binds to ds-DNA in a non-specific fashion.

Table 2. Concentration (ppb) of metals in *S. pyogenes* Csn2.

	Native Csn2	**SeMet Csn2**	**Buffer**
Ca	$5.105x10^3$	$8.579x10^3$	$<0.3x10^3$
Ni	20.71	5.63	<5
Mn	11.9	8.0	10.3
Mg	<5	<5	<5
Co	<5	<5	<5

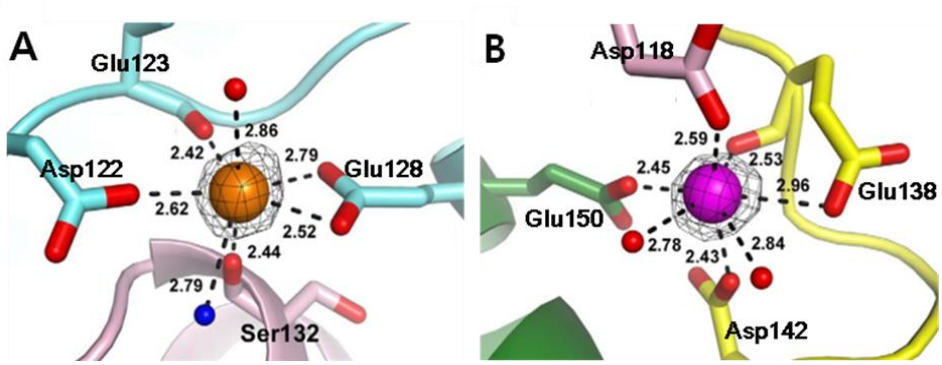


Figure 7. Calcium binding sites in *S. pyogenes* Csn2 structure. (A) CA1 site and (B) CA2 site are colored as in Figure 3A. Coordinating oxygen atoms are shown in red and the missing water molecule in the CA1 site, represented as a blue sphere, is modeled based on a comparison with the other CA1 site. The distances are also indicated. The composite omit electron density contoured at the 10 σ.

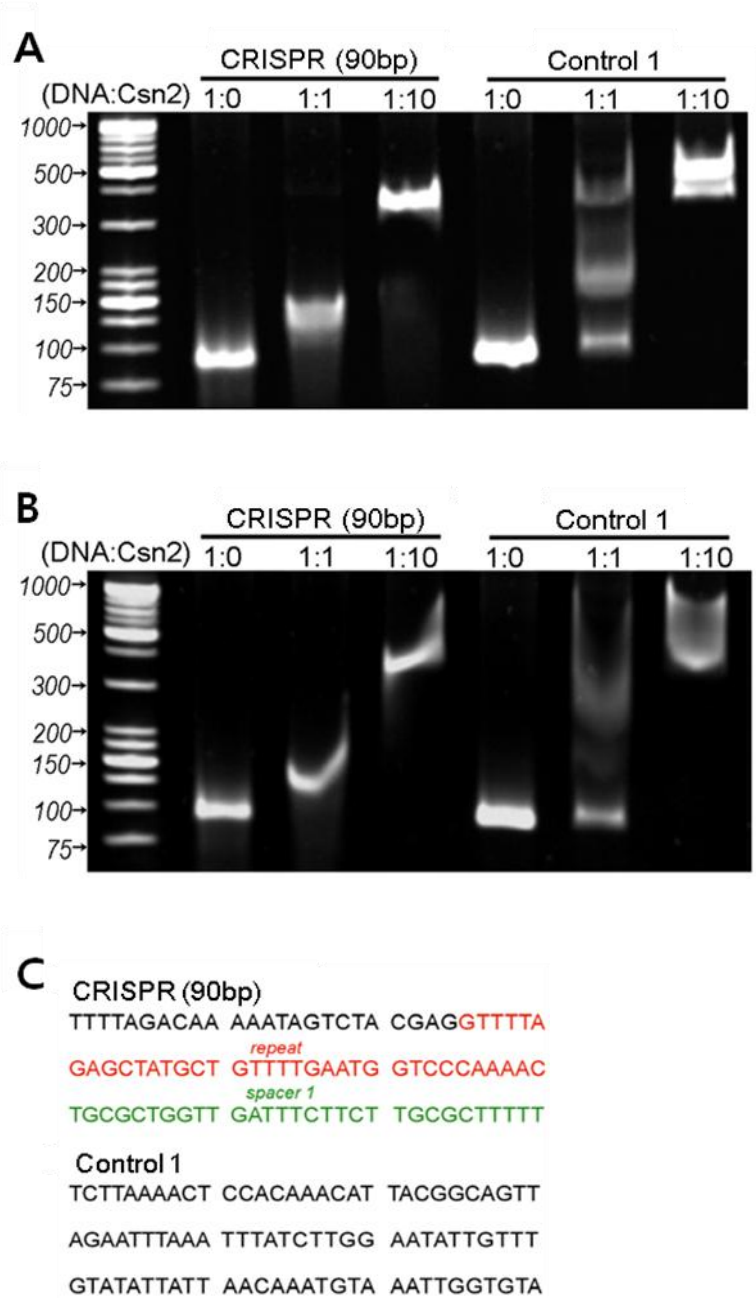


Figure 8. Double-stranded DNA binding of *S. pyogenes* Csn2. (A) EMSA was performed with 90 bp ds-DNA and increasing concentrations of *S. pyogenes* Csn2 in binding buffer containing 20 mM HEPES pH 7.5, 50 mM NaCl. The molar ratio

of DNA to Csn2 tetramer is indicated for each lane. (B) Performed with the same method of figure (A), except the binding buffer supplemented with 20mM $CaCl_2$ was used. (C) Sequences of ds-DNA fragments from *S. pyogenes* CRISPR (CRISPR) and promoter site of *A. thaliana* ERD Ⅰ (Control) used for the electrophoretic mobility shift assay (EMSA). In (CRISPR), the repeat and the first spacer of the CRISPR are colored red and green, respectively.

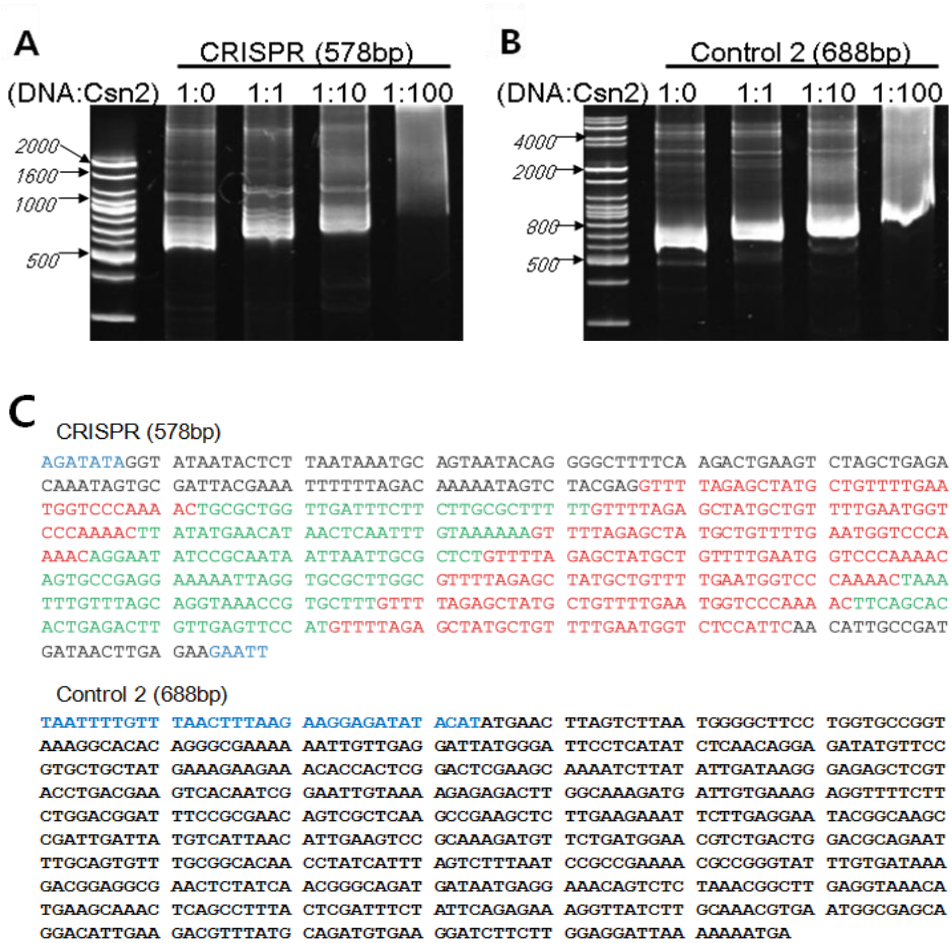


Figure 9. DNA binding assay of Csn2 with longer ds-DNAs. EMSA was performed with the same method of Figure 8A, but only with (A) the longer ds-DNA of CRISPR and (B) similar length of control ds-DNA, control 2. As the concentration of Csn2 tetramer increases, both of ds-DNA bands have shifted. (C) Sequences of ds-DNAs used in the figure 9. In CRISPR (578 bp), repeats are colored red, spacers are colored green and in both of ds-DNA sequences, the extra added base-pairs from primers are colored blue.

4. DISCUSSION

The Csn2 is regarded as an essential protein of the CRISPR-Cas system, because it provides a unique opportunity to study the adaptation process in the immune system. The functional information of this protein comes from knockout studies (41). Deletion of *csn2* gene in *Streptococcus thermophilus* led to the assumption that Csn2 is required for spacer integration (41).

The crystal structure of *S. pyogenes* Csn2 reveals that the protein functions at the quaternary structure level. Recently, the crystal structure of *Enterococcus faecalis* Csn2 was solved at a resolution of 2.7 Å (42). Like *S. pyogenes* Csn2, the structure revealed a ring-like structure with a hole in the middle (Figure 3). Considering the results from the ring-like quaternary structure, the diameter of the hole and the conserved positively charged lysine residues on the inner surface (Figure 4), coincide quite well with the non-specific ds-DNA binding function of Csn2 (Figures 8 and 9). However, the binding affinity of Csn2 appeared to be non-specific, as the shift occurred among the short and long CRISPRs and two different control DNAs, which have no relationship to the CRISPR-Cas system and have no similar sequences with CRISPR.

The results of metal analysis using inductively coupled plasma mass spectrometry (ICP-MS) indicate that the bound metals in *S. pyogenes* Csn2 were calcium (Table 2). The *E. faecalis* Csn2 structure (42) revealed calcium ions at the

equivalent sites of *S. pyogenes* (24), and the coordinating side-chains are also conserved (Figure 4A). From these observations, the metals at the sites were assigned as calcium ions.

The subunits of tetramer Csn2 from *S. pyogenes* display heterogeneity in calcium binding (Figure 5B). This may indicate that the difference of calcium ion affinity between CA1 and CA2.

Although in the study of *E. faecalis* Csn2, the authors found that the introduction of calcium ions is essential for its crystallization, and behavior as a tetramer (42), Csn2 from *S. pyogenes* included calcium ions in the crystal structure despite any addition of calcium ions through the whole experimental procedures from purification to crystallization. This suggests that those bound calcium ions were incorporated during protein expression.

More importantly, the results of the size-exclusion chromatography reveal that *S. pyogenes* Csn2 acted as a tetramer not only without the addition of calcium but also in the presence of metal-chelating agents such as EGTA and EDTA (Figure 6). This implies that for *S. pyogenes* Csn2, calcium binding in the structure may not be essential for the tetramerization but strengthen the stabilization of the interactions between protomers. In contrast, for *E. faecalis* Csn2, calcium is the key element to stabilize the tetrameric ring-like structure (42).

Collectively, the observations from these studies seem to be a supporting clue for Csn2 participates in the adaptation process. Although at present, the precise role of Csn2 within the CRISPR-Cas system is not yet clear. To find out the exact function

of Csn2, further studies are required, for example, co-crystallization of Csn2 protein with ds-DNA to find the exact binding mode.

In addition, clarifying the role of the CRISPR-Cas system could lead to another development of novel antibiotics against *S. pyogenes*, the human pathogen. For instance, a Csn2 inhibitor reduces the resistance to phage infections of pathogen *S. pyogenes*, consequently, reduces the pathogen's vitality.

5. Accession numbers

The coordinates and structure factors have been deposited in the Protein Data Bank (PDB) (43) with the accession code 3TOC (selenomethionyl *S. pyogenes* Csn2) and 3V7F (native *S. pyogenes* Csn2) (24).

6. REFERENCES

1. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology 151: 2551–2561.
2. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327: 167–170.
3. Karginov, F.V., and Hannon, G.J. (2010) The CRISPR system: small RNAguided defense in bacteria and archaea. Mol. Cell 37: 7–19.
4. Marraffini, L.A., and Sontheimer, E.J. (2010a) CRISPR interference: RNAdirected adaptive immunity in bacteria and archaea. Nat. Rev. Genet. 11: 181–190.
5. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. 60: 174–182.
6. Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR–a widespread system that provides acquired resistance against phages in bacteria and archaea. Nat.Rev. Microbiol. 6: 181–186.
7. van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ (2009) CRISPRbased adaptive and heritable immunity in prokaryotes. Trends Biochem Sci 34:401–407.

8. Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. Cell 136: 615–628.
9. Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. 1: e60.
10. Grissa, I., Vergnaud, G., and Pourcel, C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics 8: 172.
11. Jansen, R., Embden, J.D., Gaastra, W., and Schouls, L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. 43: 1565–1575.
12. Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol Direct 1: 7.
13. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315: 1709–1712.
14. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science

321: 960–964.

15. Marraffini, L.A., and Sontheimer, E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322: 1843–1845.
16. 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.
17. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, et al. (2011) Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9: 467–477.
18. Marraffini LA, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11: 181–190.
19. Wiedenheft B, Zhou K, Jinek M, Coyle SM, Ma W, Doudna JA(2009) Structural basis for DNase activity of a conserved protein implicated in CRISPRmediated genome defense. Structure 17(6): 904-912.
20. Marraffini LA, Sontheimer EJ (2009) Invasive DNA, chopped and in the CRISPR. Structure 17(6): 786-788.
21. Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, Kochinyan S, Wang S, Chruszcz M, Minor W, et al (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. J Biol Chem

283(29): 20361-20371.

22. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, et al. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471: 602–607.
23. Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, et al. (2011) The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic Acids Res 39: 9275–9282.
24. Koo Y, Jung DK, Bae E (2012) Crystal structure of Streptococcus pyogenes Csn2 reveals calcium-dependent conformational changes in its tertiary and quaternary structure. PLoS One 7(3): e33401.
25. Busso D, Delagoutte-Busso B, Moras D (2005) Construction of a set Gatewaybased destination vectors for high-throughput cloning and expression screening in Escherichia coli. Anal Biochem 343: 313–321.
26. Mark BL, Vocadlo DJ, Knapp S, Triggs-Raine BL, Withers SG, et al. (2001) Crystallographic evidence for substrate-assisted catalysis in a bacterial betahexosaminidase. J Biol Chem 276: 10330–10337.
27. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol 276: 307–326.
28. Terwilliger TC, Berendzen J (1999) Automated MAD and MIR structure solution. Acta Crystallogr D Biol Crystallogr 55: 849–861.
29. Terwilliger TC (2000) Maximum-likelihood density modification. Acta Crystallogr D Biol Crystallogr 56: 965–972.

30. Terwilliger TC (2003) Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr D Biol Crystallogr 59: 38–44.
31. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60: 2126–2132.
32. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53: 240–255.
33. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66: 213–221.
34. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr D Biol Crystallogr 67: 271–281.
35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) Phaser crystallographic software. J Appl Crystallogr 40: 658–674.
36. Chen VB, Arendall WB, 3rd, Headd JJ, Keedy DA, Immormino RM, et al. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66: 12–21.
37. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic

Acids Res 22: 4673–4680.

38. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15: 305–308.

39. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372: 774–797.

40. Potterton E, Briggs P, Turkenburg M, Dodson E (2003) A graphical user interface to the CCP4 program suite. Acta Crystallogr D Biol Crystallogr 59: 1131–1137.

41. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819): 1709-1712

42. Nam KH, Kurinov I, Ke A (2011) Crystal structure of clustered regularly interspaced short palindromic repeats (CRISPR)-associated Csn2 protein revealed Ca^{2+}-dependent double-stranded DNA binding activity. J Biol Chem 286: 30759–30768.

43. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242.

Abstract in Korean

CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated proteins) system은 archaea와 bacteria의 genome에 널리 분포하고 있는 미생물 면역 시스템으로 숙주의 체내에 침입한 bacteriophages 또는 plasmids의 유전물질들을 파괴하여 면역성을 나타내며 후천적으로 얻은 면역력의 세대 계승이 가능하다.

이 면역 시스템의 작용단계는 크게 adaptive stage, expression stage 그리고 interference stage로 구분한다. Adaptive process는 침입 유전물질의 일부 fragment를 자신의 CRISPR내부의 spacer로 도입해 들여오는 단계이며, expression process에서는 CRISPR가 전사된 후 crRNA라는 외래유전물질을 공격하기 직전의 최종형태로 다듬어지는 과정까지의 단계이다. 이 crRNA와 Cas protein들이 침입유전물질을 파괴하는 최종단계가 interference process이다.

Csn2는 Nmeni subtype Cas 단백질에 속하며, 최근 발표된 새로운 CRIPSR-Cas system 분류법에 따르면 TypeⅡ-A system에 포함된다. 현재까지의 연구 결과에 의하면 Csn2는 CRISPR-Cas system의 작용단계 중 adaptive process에 관여할 것으로 보여지고 있다.

본 연구에서는 *Streptococcus pyogenes* Csn2 단백질을 발현, 분리 및 정제, 결정화한 후 x-ray 결정 구조 분석법을 통하여 단백질의 3차

구조를 규명하였다. 이 구조적 실험 자료와 기존의 연구 결과들을 바탕으로 Csn2가 이중 DNA와 결합 특성이 있음을 확인하였다.

주요어: CRISPR-Cas system, *Streptococcus pyogenes* SF370, Csn2, x-ray crystallography, DNA binding

Student Number : 2010-23441

Acknowledgments

2년 전 아무것도 모르던, 부족한 저를 좋은 결과와 함께 무사히 졸업할 수 있게 성심껏 지도해주신 배의영 교수님께 진심으로 감사 드립니다. 그리고 조언과 도움을 구하면 언제나 기꺼이 답해 주시던 이상기 교수님, 구조생물학 연구실의 선배님들께도 깊이 감사 드립니다. 생물리화학 연구실에서 함께 하며 저에게 많은 실험들을 가르쳐 주었던 두교 선배 그리고 졸업 전, EMSA 실험방법에 대한 조언과 control DNA를 제공해주었던 분자생물학 연구실의 이상준 씨에게도 고마움을 전하고 싶습니다. 또한 2년간 함께 생활한 동기 소진이와 곧 새롭게 대학원 생활을 시작할 동현이, 앞으로 하는 일 모두 잘 되기를 바랍니다.

제 인생에서 가장 소중한 가족들에게 고맙고 사랑한다고 전하고 싶습니다. 오랜 세월, 딸 뒷바라지에 고생하신 아버지, 어머니께 죄송스러운 마음이 앞서지만 한결같은 사랑으로 키워주신 은혜에 조금이라도 보답하기 위해 더욱 부지런히 그리고 열정적으로 살고자 합니다. 언제나 철없는 동생을 귀여워해주는 하나뿐인 오빠의 새로운 시작을 진심으로 축하하고 미진 언니와 함께 오래오래 행복하기를 바랍니다.

끝으로, 연진, 나라 그리고 진효 모두에게 자랑스러운 친구가 되겠습니다. 지난 대학원 생활 동안 함께하였던 모든 인연들에 항상 감사하는 마음, 잊지 않겠습니다. 감사합니다.