Biodesign

pISSN 2288-6982 | eISSN 2288-7105

CRYSTALLIZATION P 62-65

Purification, crystallization and X-ray crystallographic analysis of Csm5 in Type III-A Crispr-Cas system

Yan An[1,2], Kwang-Hyun Park[2], Minho Lee[2], Woo-chan Ahn[2], In-Young Baek[2], Tae-Jip Kim[1] and Euijeon Woo[2,3,*]

[1]Division of Animal, Horticultural and Food Sciences, Graduate School of Chungbuk National University, Cheongju 28644, Korea, [2]Disease Target Structure Research Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon 34141, Korea, [3]Department of Structural Biology, KRIBB school of Bioscience, Korea University of Science and Technology, Daejeon 34141, Korea.
*Correspondence: ejwoo@kribb.re.kr

The CRISPR-Cas system is a microbial adaptive and heritable immune system, of which mechanism relies on the effector ribonucleoprotein (RNP) complex to degrade foreign genetic elements. The effector surveillance complex of the Type III-A CRISPR-Cas system has five Csm components, Csm1~Csm5. The Csm5 protein is placed in the crRNA 3' ends in the effector RNP complex, leading to speculations that crRNA maturation may be catalyzed by the Csm5 subunit. However, the crystal structure and the detailed function of Csm5 still remain elusive. In this study, the Csm5 from *Thermococcus onnurineus* NA1 was purified and crystallized by the sitting drop method in the condition of 20% w/v PEG 8000 and 100 mM CHES/sodium hydroxide pH 9.5 at 291K. The diffraction data were collected to a resolution of 3.5 Å. The crystal belonged to space group $P4_32_12$, with unit cell parameters *a*=81.0 Å, *b*=81.0 Å, *c*=169.1 Å. One protomer was presented in the asymmetric unit with a corresponding V_M of 2.10 $Å^3$ Da^{-1} and solvent content of 50%.

INTRODUCTION

Clustered regularly interspaced short palindromic repeat (CRISPR) and the CRISPR-associated system (CAS) is a defense mechanism present in bacteria and archaea that employs crRNA-guided effector complexes (Jinek, Chylinski et al., 2012, Kunin, Sorek et al., 2007). CRISPR loci is arranged by a repeat motif and short DNA fragments, known as spacer, derived from past invaders genetic elements that provide the specificity for genetic memory (Deveau, Garneau et al., 2010, Szczepankowska, 2012). *Cas* genes encode the proteins responsible for mediating the CRISPR response to foreign nucleic acids (Garneau, Dupuis et al., 2010, Horvath & Barrangou, 2010). Transcripts from CRISPR loci are processed to generate a short crRNAs that assemble with CRISPR-associated (Cas) proteins into a ribonucleoprotein (RNP) complex. Based on *cas* gene content and distinct mechanism, CRISPR-Cas systems are classified into five major types (Types I~VI) (Cao, Gao et al., 2016, Makarova, Wolf et al., 2015, Walker, Chou-Zheng et al., 2017). Type I/III/IV systems are mediated by the multiprotein complex; Cascade or Csm/Cmr complex, whereas Type II/V/VI systems are mediated by a single RNP complex; Cas9 or Cpf1 or C2C2 (Fagerlund, Staals et al., 2015, Makarova et al., 2015, Zetsche, Gootenberg et al., 2015). The Type III system, commonly found in a wide variety of bacteria and archaea, can be further classified into four subtypes, Type III A-D (Han, Li et al., 2017, Hatoum-Aslan, Maniv et al., 2014). The Type III-A system, appearing more commonly in archaea, harbors five *csm* genes *csm1-csm5* and the effector Csm complex cleaves target RNAs or single-strand DNAs in the region of crRNA complementarity (Liu, Iavarone et al., 2017, Park, An et al., 2017) .

In the Electron Microscopy structure analysis, the Csm complex exhibits a 'twisted bread stick' shape in that the Csm3 proteins are helically arranged to form a groove wrapping around the crRNA (Park et al., 2017). The 5'-terminal region of the crRNA contacts Csm1 while crRNA 3'-terminal region interact with Csm5, which reportedly constitutes the cap of the complex (Staals, Zhu et al., 2014, Walker et al., 2017). Despite numerous biochemical characterization and functional studies, the complete crystal structure of Type III-A Csm complex still remains elusive (Ichikawa, Cooper et al., 2017). To date, two structures of Csm components have been reported in the PDB databank, the Csm1 from *Thermococcus onnurinues* and the Csm3 from *Methanopyrus kandleri* (Hrle, Su et al., 2013, Jung, An et al., 2015), but the crystal structure of Csm5 was not reported yet. In this study we report the purification, crystallization and preliminary X-ray crystallographic study of Csm5 in Type III-A CRISPR-Cas system.

RESULTS AND DISCUSSION

The gene encoding *Csm5* (Ton_0897) organized in Csm complex was amplified by PCR from *T. onnurineus* genomic DNA (Jung, Park et al., 2016) and cloned into the pProEx-HTa vector to enable expression of the recombinant protein with an N-terminal six-histidine tag. The *Csm5* gene encoded a recombinant

protein of 397 amino acid residues with molecular masses of 45.2 kDa. The recombinant plasmid was well expressed in the *E. coli* BL21 (DE3). Host cells harboring the recombinant pProEx-HTa vector showed high expression after induction by 1 mM IPTG. Thermostable enzymes derived from thermophiles generally maintained their solubility even at high temperatures, so that those expressed in *E. coli* could effectively be separated by heat treatment. The crude lysate was centrifuged and the soluble supernatant was then heat treated at 75°C for 10 mins to precipitate the proteins from the host cell lysate. As a result, the thermostable Csm5 protein remained in the solution while most of other proteins precipitated after the heat treatment. After the affinity chromatography (Ni-NTA) column purification, the N-terminal six-histidine tag was easily cleaved by the TEV protease suggesting neither mis-folding nor structural hindrance. During the purification procedure at room temperature, the Csm5 protein easily resulted in aggregation in a buffer without glycerol, suggesting a sensitive feature of the protein to temperature. Addition of 5 % glycerol to the sample buffer helped to maintain the solubility of the Csm5 protein, which was removed at the last step of dialysis before crystallization setup. The protein appeared a single band on SDS-PAGE, with estimated purity over 95% (Figure 1). We obtained the Csm5 crystal in 20% w/v PEG 8000, 100mM CHES/sodium hydroxide pH 9.5 (Figure 2) after 2 weeks of crystallization setup. Before the X-ray diffraction, the crystals were transferred to the cryo-protection buffer containing 50% glycerol (v/v) to the mother liquor of the reservoir solution. The crystal X-ray diffraction data were collected to a resolution of 3.4 Å on beamline 7A at the Pohang Accelerator Laboratory (PAL, Pohang, Korea), using a Quantum 270 CCD detector (ADSC, USA). The Csm5 crystals belonged to $P4_32_12$ group with unit cell parameters a=81.0 Å, b=81.0 Å, c=169.1 Å, $\alpha=\beta=\gamma$=90°.

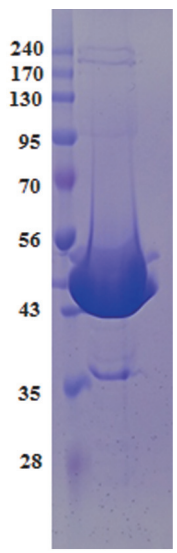

FIGURE 1 | The protein band of Csm5 on 12% SDS-PAGE gel. The Csm5 shows a single major band on 12% SDS-PAGE gel.

One promotor was present in the asymmetric unit with a corresponding V_M of 2.10 $Å^3$ Da^{-1} and solvent content of 50%. The data collection statistics were summarized in Table 3. We have attempted to solve the structure by Molecular Replacement method using the structure of Cmr1, a close homologues in Type III-B Crispr-Cas system. However due to the low sequence similarity, it was not solved by the molecular replacement method. In an attempt to obtain the phase of the crystal by single anomalous dispersion (SAD) method, currently the selenomethionine substituted protein of Csm5 is being purified for crystallization setup.

METHODS

Construction of csm5

The gene coding of insert Csm5 (Ton_0897) protein was amplified from *Thermococcus onnurineus* NA1 genomic DNA by PCR. The primer sequence was described in Table 1. PCR products were inserted into the pProEx-HTa vector (BamHI and XhoI restriction enzyme site) using EZ-Fusion™ Cloning core kit (Enzynomics). N-terminus of the protein contains the histidine tag and TEV cleavage site to enable purification. The Csm5 macromolecular information was given in Table 1.

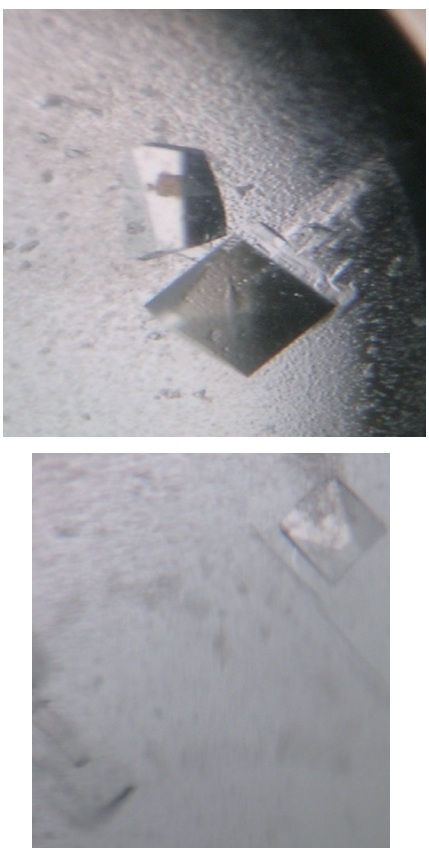

FIGURE 2 | Crystals of Csm5. The crystals were obtained in two week using a reservoir solution containing 20 % PEG 8000 (w/v), 100 mM CHES/sodium hydroxide pH 9.5 at 291 K.

TABLE 1 | Csm5 production information

Source organism	*Thermococcus onnurineus* NA1
DNA source	Chromosomal DNA
Forward primer	TTTCAGGGCGCCATGGATCCGATGACCGAGCGAACA
Reverse primer	GGTACCGCATGCCTCGAGTCATTCCAGCACCAC
Cloning vector	pProEx-HTa
Expression host	*E. coli* BL21 (DE3)
Complete amino acid sequence of the construct produced	MTERTLKVLSPLHIGTGNELTPVDIYPRENIIHVLDTERLVNDLMNLGVELNEILALLKN PPGDAYIWKGYIEEFHLDPSDYSIYTLKIHGKIGRKSMQIKEFIKLNGRPYIPGSSLKGA IRTAVLYKALKECGDARAVMRVVSKVNGDVARDIGRSEDVLDYYMSFLSRARIDRKRADD LLEAIVFGMEPDRRSKIRYEPKRDPMKALIVRDSKPVGRKHLAVYHVEVIGNPQPIPIWV EAIEPGAATDVEIHVDTEALRLNADYFNGLLWECLKERGEPGEVFEDFLWEAVDEFYTAV MKYETIEVQKFGRYTSQVRSFYASLEDHSGHVLRLGWGSGWLAMTIGLLLVEKGYKWENV RKKLGLGKKPGGSGFSREFPKTRRLADGMPMGWVVLE

TABLE 2 | Crystallization

Method	Sitting drop vapor diffusion
Plate type	96-Well Sitting-Drop High-Throughput Crystallography Plate
Temperature (K)	291
Protein concentration (mg ml^{-1})	15
Buffer composition of protein solution	50 mM Tirs pH 7.5, 150 mM NaCl
Composition of reservoir solution	20 % w/v PEG 8000, 100 mM CHES/sodium hydroxide pH 9.5
Volume and ratio of drop	2µl; 1:1 ratio of protein and reservoir solution
Volume of reservoir (µl)	80

TABLE 3 | Data collection and processing statistics

Diffraction source	Beamline 7A, PAL
Wavelength (Å)	1.000
Temperature (K)	100
Detector	Quantum 270 CCD
Crystal-to-detector distance (mm)	350
Rotation range per image (°)	1
Total rotation range (°)	360
Exposure time per image (s)	1
Space group	P43212
a, *b*, *c* (Å)	81.0 , 81.0, 169.1
α, β, γ (°)	90, 90, 90
Resolution range (Å)	50.0-3.5
Total No. of reflections	13405
Completeness (%)	99.0 (100)
R_{sym} or R_{merge} (%)	14.0 (73.0)
I/δ (I)	62.09 (11.5)
Redundancy	55.2 (58.8)

Purification

The Plasmid DNA was used for transformation into E. coli BL21-RIL cells. Cells were cultured in 9L M9 media. When the cell culture reached OD value of 0.6 it was induced with 1mM IPTG and incubated at 18°C for 15 hours. The cells were collected by the centrifuge at 4,000 rpm for 30 mins. The collected cells were lysed by sonication (amplitude-60%, Pulse-1sec on/ 1sec off, total time was 10 mins). The lysis buffer contained 50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol, 1 mM Phenylmethylsulfonyl fluoride. The crude lysate was centrifuged and the soluble supernatant was then heat treated at 75°C for 10 mins to precipitate the proteins originated from the host cell lysate, before centrifugation at 15,000 rpm for 30 mins at room temperature. Only the soluble fraction of the heat treated supernatant was loaded onto affinity chromatography (Ni-NTA) column. The lysis buffer was used as the loading buffer and the elution buffer contained 250 mM Imdazole in addition to the lysis buffer. After the first purification, the N-terminal six-histidine tag was cleaved by incubation with the TEV protease at room temperature overnight. The tag cleavage efficiency of TEV protease was ~90% and most of the target protein was cleaved. After cutting the histidine tag, the protein was heated at 75°C for about 3mins. In second purification, size-exclusion chromatography (SEC, Hiload 16/200 Superdex 200pg GE Healthcare) was performed in 50 mM Tris HCl pH 7.5, 500 mM NaCl, 5 mM β-mercaptoethanol, 5 % glycerol, 1 mM Phenylmethylsulfonyl fluoride buffer.

Crystallization

The purified protein was used for crystallization set up. The final concentration of Csm5 protein was 15 mg/ml. The crystallization set up

used 980 conditions used sitting drop plate (96-Well Sitting-Drop High-Throughput Crystallography Plate) by Art Robbing instruments at 18°C. The drop contained 0.2 μl protein and 0.2 μl reservoir solution. The crystals were observed in a buffer containing 20 % PEG 8000 (w/v), 100 mM CHES/sodium hydroxide pH 9.5 after 2 weeks of crystallization. For the improvement and optimization of the crystal, the Csm5 protein was setup in sitting drop plate with 2 μl of protein sample and 2 μl reservoir solution (Table 2).

Data collection and processing

For data collection of the crystal in cryogenic condition, all crystals were transferred to 10 μl of cryo-protection buffer contained 5 μl 50% glycerol and 5 μl reservoir solution. X-ray datasets were collected at 2.5 Å on beamline 7A at the Pohang accelerator Laboratory (PAL, Pohang, Korea). Complete diffraction datasets were subsequently processed, merged, and scaled with HKL 2000 (Otwinowski & Minor, 1997, Suk-Youl Park, Sung-Chul Ha et al., 2017) (Table 3).

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

We thank Dr. Kang Sung Gyun in KIOST for providing the genomic DNA and the staff members at Pohang accelerator Laboratory beamline 7A for data collection. This work was supported by National Research Fund (2015R1A2A2A03006970) and KRIBB Research Initiative.

Original Submission: May 24, 2017
Revised Version Received: Jun 5, 2017
Accepted: Jun 5, 2017

REFERENCES

Cao, L., Gao, C.H., Zhu, J., Zhao, L., Wu, Q., Li, M., and Sun, B. (2016). Identification and functional study of type III-A CRISPR-Cas systems in clinical isolates of Staphylococcus aureus. *Int J Med Microbiol* **306**, 686-696.

Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* **64**, 475-493.

Fagerlund, R.D., Staals, R.H., and Fineran, P.C. (2015). The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol* **16**, 251.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71.

Han, W., Li, Y., Deng, L., Feng, M., Peng, W., Hallstrom, S., Zhang, J., Peng, N., Liang, Y.X., White, M.F., and She, Q. (2017). A type III-B CRISPR-Cas effector complex mediating massive target DNA destruction. *Nucleic Acids Res* **45**, 1983-1993.

Hatoum-Aslan, A., Maniv, I., Samai, P., and Marraffini, L.A. (2014). Genetic characterization of antiplasmid immunity through a type III-A CRISPR-Cas system. *J Bacteriol* **196**, 310-317.

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170.

Hrle, A., Su, A.A., Ebert, J., Benda, C., Randau, L., and Conti, E. (2013). Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3. *RNA Biol* **10**, 1670-1678.

Ichikawa, H.T., Cooper, J.C., Lo, L., Potter, J., Terns, R.M., and Terns, M.P. (2017). Programmable type III-A CRISPR-Cas DNA targeting modules. *PLoS One* **12**, e0176221.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821.

Jung, T.Y., An, Y., Park, K.H., Lee, M.H., Oh, B.H., and Woo, E. (2015). Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure* **23**, 782-790.

Jung, T.Y., Park, K.H., An, Y., Schulga, A., Deyev, S., Jung, J.H., and Woo, E.J. (2016). Structural features of Cas2 from Thermococcus onnurineus in CRISPR-cas system type IV. *Protein Sci* **25**, 1890-1897.

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

Liu, T.Y., Iavarone, A.T., and Doudna, J.A. (2017). Correction: RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PLoS One* **12**, e0175612.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J., Terns, R.M., Terns, M.P., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736.

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

Park, K.H., An, Y., Jung, T.Y., Baek, I.Y., Noh, H., Ahn, W.C., Hebert, H., Song, J.J., Kim, J.H., Oh, B.H., and Woo, E.J. (2017). RNA activation-independent DNA targeting of the Type III CRISPR-Cas system by a Csm complex. *EMBO Rep* **18**, 826-840.

Staals, R.H., Zhu, Y., Taylor, D.W., Kornfeld, J.E., Sharma, K., Barendregt, A., Koehorst, J.J., Vlot, M., Neupane, N., Varossieau, K., Sakamoto, K., Suzuki, T., Dohmae, N., Yokoyama, S., Schaap, P.J., et al. (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of Thermus thermophilus. *Mol Cell* **56**, 518-530.

Park, S.Y., Ha, S.C., and Kim, Y.G. (2017). The Protein Crystallography Beamlines at the Pohang Light Source II. *Biodesign* **5**, 30-34.

Szczepankowska, A. (2012). Role of CRISPR/cas system in the development of bacteriophage resistance. *Adv Virus Res* **82**, 289-338.

Walker, F.C., Chou-Zheng, L., Dunkle, J.A., Hatoum-Aslan, A. (2017). Molecular determinants for CRISPR RNA maturation in the Cas10-Csm complex and roles for non-Cas nucleases. *Nucleic Acids Res* **45**, 2112-2123.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., Koonin, E.V., Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771.

