Biodesign

pISSN 2288-6982 | eISSN 2288-7105

CRYSTALLIZATION P 37-41

Purification, crystallization and X-ray crystallographic analysis of Csm1/Csm4 sub-complex in CRISPR/Cas Type III-A system

In-Young Baek[1,2], Kwang-Hyun Park[2], Woo-chan Ahn[2], Yan An[2], In-kyu Hwang[1]* and Euijeon Woo[2,3]*

[1]Immunology and Immunopharmacology Laboratory, College of Pharmacy, Chungnam National University, Daejeon 34134, 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, University of Science and Technology, Daejeon 34141, Korea
*Correspondence: ejwoo@kribb.re.kr

The CRISPR/Cas system is an adaptive prokaryotic immune response that defends against exogenous genetic elements. The CRISPR/Cas Type III-A system targets RNA and DNA that is coupled with transcription. The effector complex of the Type III-A CRISPR-Cas system has five Csm components, Csm1~Csm5. The Csm1/Csm4 sub-complex is placed in the crRNA 5' ends of the effector RNP complex and likely involved in the discriminative function for self vs non-self DNA. Here, we prepared Csm1/Csm4 proteins from *Thermococcus onnurineus* NA1 and crystallized the sub-complex. The Csm1/Csm4 was crystallized using hanging-drop vapor diffusion from a reservoir solution containg 200 mM sodium chloride, 100 mM sodium acetate trihydrate, pH 4.6, 26% (+/-)-2-methyl-2,4-pentanediol. The diffraction data to 2.8 Å resolution show that the crystal belongs to the space group P6522 with unit cell parameters of a=154.91Å, b=154.91Å, c=182.25Å, $\alpha=\beta=90°$ and $\gamma=120°$.

INTRODUCTION

The clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) systems are the prokaryotic immune response by crRNA-guided nuclease effector complexes (Jinek et al., 2012; Kunin et al., 2007). From the three distinct mechanisms of the CRISPR/Cas immune system, CRISPR loci are composed of repeats and spacer sequences that are derived from DNA fragments of invaders (Bolotin et al., 2005; Terns and Terns, 2011). Transcripts from CRISPR loci are processed to yield short small interfering CRISPR RNAs (crRNAs) that constitutes the crRNA-effector complex and guides the complex to detect target DNA or RNA (Brouns et al., 2008). Based on the distinct set of *cas genes*, CRISPR/Cas systems are classified into 2 classes, six major types (types I–VI) and at least 16 subtypes (Mohanraju et al., 2016). Class 1 includes type I/III/IV systems composed of multi-subunit crRNA-effector complex such as Cascade, Csm and Cmr complex. In comparison, class 2 includes type II/V/VI systems that function as a single protein such as Cas9, Cpf1 and C2C2 (Fagerlund et al., 2015; Makarova et al., 2015; Zetsche et al., 2015). The type III system can be classified into 4 subtypes including the main type III-A Csm and type III-B Cmr complexes (Makarova et al., 2015; Zhang et al., 2016). The Csm complex in type III-A system consists of five subunits (Csm1~5) and cleaves target RNAs or ssDNAs with the complementary guide crRNA (Hale et al., 2014; Staals et al., 2014; Zhang et al., 2012). The Csm1 subunit of the Csm complex has an HD domain that functions as DNA nuclease activity, and a palm domain associated with oligoadenylates synthesis (Cocozaki et al., 2012; Makarova et al., 2011; Zhu and Ye, 2012). In the Csm1/Csm4 sub-complex, the palm domain is located between Csm1 and Csm4 and synthesizes oligoadenylates from ATP with a cofactor Mg^{2+}.

In the electron microscopy analysis of Csm structure, the 5'-terminal region of the crRNA contacts Csm1/Csm4 while the crRNA 3'-terminal region interacts with Csm5, which reportedly constitutes the capping of the complex (Park et al., 2017; Staals et al., 2014; Walker et al., 2017). To date, two structures of Csm components have been reported in the PDB databank, the Csm1 from *Thermococcus onnurinues* (Jung et al., 2015) and the Csm3/Csm4 sub-complex from *Methanocaldococcus jannaschii* (Numata et al., 2015). The crystal structure of Csm1/Csm4 sub-complex was not reported yet. In this study we performed purification, crystallization and preliminary X-ray crystallographic study of *Thermococcus onnurineus* Csm1 and Csm4 in Type III-A CRISPR-Cas system.

RESULTS AND DISCUSSION

The histidine tagged Csm1/Csm4 was cloned and simultaneously expressed in *E.coli* and purified by affinity chromatography. In preliminary analysis, we found that the Csm1/Csm4 sub-complex was produced in a soluble form by *E. coli* co-expression techniques, while each individual subunit was insoluble. The *csm1* gene encoded a recombinant protein of 377 amino acid residues with a calculated molecular weight of 85.4 kDa and

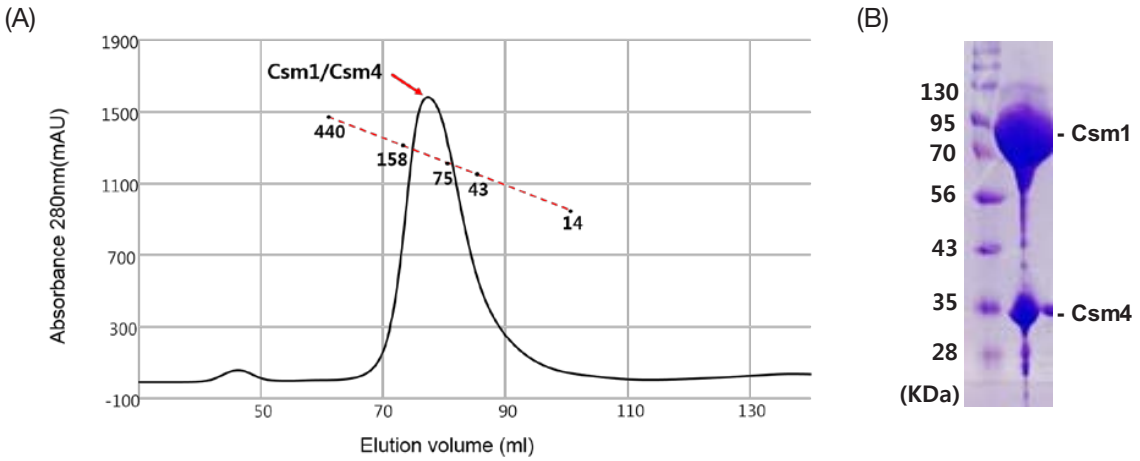

FIGURE 1 | Size-exclusion chromatography of Csm1/Csm4 sub-complex. (A) The major peak of Csm1/Csm4 represents the deduced molecular weight of 121 kDa. The size markers are ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa) and ribonuclease A (13.7 kDa). (B) SDS-PAGE analysis of the purified Csm1/Csm4 sub-complex.

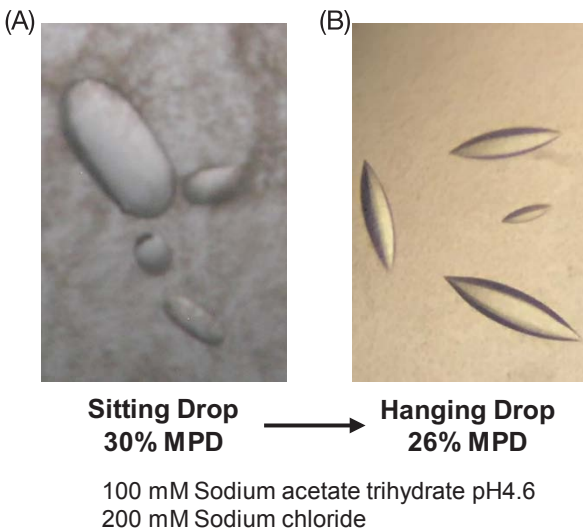

FIGURE 2 | Improvement of Csm1/Csm4 crystal. (A) Initial crystal of Csm1/Csm4 was obtained in oval shape by sitting drop method after 1 week. (B) Sharp-edged crystals were obtained by crystallization optimization with a hanging drop method.

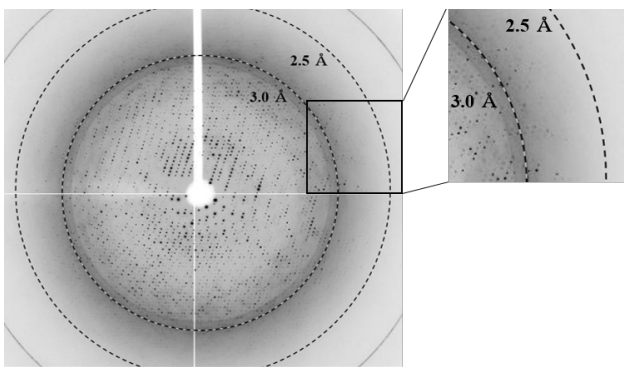

FIGURE 3 | Diffraction image. A 1.0° oscillation image of the Csm1/Csm4 crystal.

csm4 gene encoded a recombinant protein of 289 amino acid residues with a calculated molecular weight of 31.8 kDa. Host cells harboring the recombinant pRSFDuet vector showed high expression after induction by 0.5 mM IPTG at 18°C. Because Csm1/Csm4 protein was derived from thermophiles, heat treatment was initially applied after centrifugation and it showed good solubility at high temperatures. Purification was performed at room temperature using Ni-NTA affinity chromatography and size exclusion chromatography. During the purification, Csm1/Csm4 was easily degraded and aggregated at low salt. Addition of 1 mM PMSF and 1 M NaCl to the lysis buffer helped to avoid aggregation and maintained the solubility. From the size-exclusion analyses, Csm1 and Csm4 subunits formed a stable complex as a major single peak of molecular weight of 121 kDa (Figure 1A). The major peak showed two protein bands on SDS-PAGE, with estimated purity over 95% (Figure 1B). Crystals were initially obtained at 18°C after 1 week in a condition of 200 mM sodium chloride, 100 mM sodium acetate trihydrate, pH 4.6, 30% (+/-)-2-methyl-2,4-pentanediol (MPD) in the presence of ATP and EDTA. Because Csm1 is known to produce oligoadenylate from ATP molecules, we setup crystallization with ATP in order to understand binding and cleavage pattern of the ATP substrate. After refinement of the initial crystallization condition, sharp-edged crystals were obtained in reservoir solutions containing 26% MPD as a precipitant (Figure 2). For data collection, crystals were transferred to cryo-protection solution consisting of the mother liquor, 1 mM ATP and 25% glycerol. The diffraction data were collected to a resolution of 2.8 Å on beamline 7A at the Pohang Accelerator Laboratory (PAL, Pohang, Korea), using a Quantum 270 CCD detector (ADSC, USA). The Csm1/Csm4 sub-

TABLE 1 | Macromolecule-production information

Source organism	*Thermococcus onnurineus* NA1
DNA source	Genomic DNA
Forward primer*	Csm1 5'-TCATCACCACAGCCAGGATCCATGGAAATCGATGAA -3'. Csm4 5'-TAAGAAGGAGATATACATATGATGCCGAAGTTCATT -3'
Reverse primer*	Csm1 5'-GCATTATGCGGCCGCAAGCTTTCACCTCCTAACCGC -3' Csm4 5'-TTTACCAGACTCGAGGGTACCTCATTCCAGACCCTC -3'
Cloning vector	pRSFDuet-1
Expression vector	pRSFDuet-1
Expression host	*Escherichia coli* strain BL21(DE3)
Complete amino-acid sequence of the construct produced	Csm1 MEIDELTALGGLLHDIGKPVQRAGLYSGDHSTQGARFLRDLAENTGRAEYELLSLFSEFHHKGHMKNDE LMIRRIKELSPERFGLTMEDVLNALWIVYEADNLASGEREEGQPQASRPLYSVFNPGKAYPWAELDFEK ELPVPGDVFSIRSQDYRELVKRLWEELSKAKLRSDRLLPVLEKYLTFVSSVTSEGNIISLYDHMRMTSA IALAMLRAGCTAEDVRSGRCRKEKRFLLIEGDFSGIQDFIYRVSGKGTLKYLRARSAYLELIGWDVVLE ILSRLGLTRANVVFNAGGHFMIIAQNTPDAVKELEEIRAKAVEWLYREFESDLYLAIEWEPVSGREFGR EGGKNLFAEARKRLKHKLTVRKLKRFGEIKGLFEHGHTERLAECPVCGRELPEGKLEPSASDPETKVCP TCNRLVSLGGNLPKLLGFGRTAKNDAGVLVEGPFSGFVPYLQGGRPVGEQILVKNTLNPGEIPESAQFV PYFVADYFKKDPKGGVATFEELSMASTGTRRLGVMKGDVDRLGEFFSSMDSPSKLATASRFMDYFFKGY IGAIIEGKFGYIIGDVPSLRDWPEEPDIVVVYAGGDDFFIVGAWDQIFELAFRVRRAFNAYTGGKLTLS VGLGYFDERTPIYRMADVVSERLDTAKDEGRNRVFVVGRSRPLDGKHKLSYEWNHYEELWRTYAPRIYA GNGRLKGKLESKKGLLWKLLEIRELYVRDPNDVRWAYLTAYLLGRHGLSDLFPELVGIDTKAVERKEPQ PVYWVDGVLKIVLMAVRR Csm4 MPKFIAVKLIPKGPFRDIPRADTLFGAIGNAISAIHGQSAVEELVDAFVGGARISSAFPYSGDTYYLPK PLSVEPALEGILTGLDEEERYTTAKRLRKAKYLDLKNFELALRLRPFTIPEEIPYARVDVPRVVLDRVT QDSSIYFWEEIRFREKSGVYFLYSGPREVFDGYIAPAMRFLGDTGIGGKSTWGAGLFEVEFHEMKIDAP GSEYSVTLSNALPTKTPVLWRLLRKGGWSFGRRKPRMTFIAEGSIVKNDPGGMERLELGLSHEVYVYGLTFPLGVELPEGLE

*Restriction enzyme sites are underlined.

TABLE 2 | Crystallization condition

Method	Vapour diffusion in hanging drop
Plate type	24-Well Hanging Drop plate
Temperature (°C)	18
Protein concentration (mg ml^{-1})	16
Buffer composition of protein solution	50 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM β-mercaptoethanol, 5% Glycerol
Composition of reservoir solution	200 mM sodium chloride, 100 mM sodium acetate trihydrate, pH 4.6, 26% (+/-)-2-methyl-2,4-pentanediol
Volume and ratio of drop	1.5 ul protein with 1.5 ul reservoir solution
Volume of reservoir (ul)	1000

complex crystals belonged to the space group $p6_522$ with unit cell parameters of a=154.91Å, b=154.91Å, c=182.25Å, α=β=90° and γ=120°. The crystal of Csm1/Csm4 diffracted to 2.8 Å resolution (Figure 3). Detailed information of data collection and statistics were summarized in Table 3. Phasing of the data is in progress by molecular replacement and the preliminary structure refinement is currently under way using PHENIX.

METHODS

Protein expression and purification

The gene encoding *csm1 (Ton_0893)* and *csm4 (Ton_0896)* were amplified by PCR from T. onnurineus NA1 genomic DNA and were inserted into a single pRSFDuet vector (BamHI/HindIII and NdeI/KpnI sites) (Table 1). Csm1 was modified to contain a histidine tag and a TEV cleavage site at N-terminal for efficient purification. The protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18°C. Harvested cells were lysed in the 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol and 1 mM phenylmethylsulfonyl fluoride. After cell lysis, heat treatment at 65°C for 5min was applied for purification. Following centrifugation, the soluble supernatant was purified by affinity chromatography using His-Trap HP column (GE Healthcare). The protein was eluted by 250 mM imidazole in addition to the lysis buffer. Eluted fractions contained Csm1/Csm4 and subsequently applied to size exclusion chromatography in 50 mM Tris–HCl, pH 8.0, 500 mM NaCl, 5% glycerol and 5 mM β-mercaptoethanol buffer.

Protein crystallization

Prior to crystallization, purified protein was concentrated to 16 mg/ml and concentration was determined by the nanodrop method. Preliminary crystallization screening for Csm1/Csm4 was conducted by sitting drop (96-well sitting-drop high throughput crystallography plate) method at 18°C. The drop contained 0.2 µl protein and 0.2 µl reservoir solution. The primary crystals were obtained in condition of 200 mM sodium chloride, 100 mM sodium acetate trihydrate, pH 4.6, 30% (+/-)-2-methyl-2,4-pentanediol in presence of 1 mM ATP and 1 mM EDTA at 18°C after 1 week of crystallization. The best quality crystal appeared in 200 mM sodium chloride, 100 mM sodium acetate trihydrate, pH 4.6, 26%

TABLE 3 | Data collection and statistics

	Csm1/Csm4 sub-complex
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	$P6_522$
a, b, c (Å)	154.9, 154.9, 182.2
α, β, γ (°)	90, 90, 120
Resolutaion range (Å)	50.0 - 2.8
Total No. of reflections	951237
Completeness (%)	99.6 (100)
Redundancy	29.4 (30.6)
<I/σ(I)>	46.06 (5.09)
R_{merge} (%)	9.0 (96.9)

(+/-)-2-methyl-2,4-pentanediol from hanging drop method (Table 2).

Data collection and processing

Before data collection, Csm1/Csm4 crystals were transferred to cryo-protectant solution consisting of the mother liquor, 1 mM ATP and 25% glycerol. The diffraction data was collected at 2.8 Å on beamline 7A at the Pohang Accelerator Laboratory (PAL, Pohang, Korea). A range of 360° was covered with a rotation angle per image of 1° and 1 s exposure. Frame 50 to 300 data sets were collected, indexed, merged, and scaled by HKL 2000 software (Otwinowski and Minor, 1997). The data collection and processed statistics are summarized in Table 3.

CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

ACKNOWLEDGEMENTS

We thank the staff members at Pohang Accelerator Laboratory beamline 7A for data collection. This research was partly supported by the Marine Biotechnology Program of the Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries (MOF) (No. 20170488) and partly by the National Research Fund (NRF-2018R1A2A2A05021648) and KRIBB Research Initiative.

REFERENCES

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.

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.

Original Submission: Apr 10, 2018
Revised Version Received: Jun 4, 2018
Accepted: Jun 5, 2018

Cocozaki, A.I., Ramia, N.F., Shao, Y., Hale, C.R., Terns, R.M., Terns, M.P., and Li, H. (2012). Structure of the Cmr2 subunit of the CRISPR-Cas RNA silencing complex. *Structure* **20**, 545-553.

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

Hale, C.R., Cocozaki, A., Li, H., Terns, R.M., and Terns, M.P. (2014). Target RNA capture and cleavage by the Cmr type III-B CRISPR-Cas effector complex. *Genes Dev* **28**, 2432-2443.

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.

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

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* **6**, 38.

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.

Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**, aad5147.

Numata, T., Inanaga, H., Sato, C., and Osawa, T. (2015). Crystal structure of the Csm3-Csm4 subcomplex in the type III-A CRISPR-Cas interference complex. *J Mol Biol* **427**, 259-273.

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.

Terns, M.P., and Terns, R.M. (2011). CRISPR-based adaptive immune systems. *Curr Opin Microbiol* **14**, 321-327.

Walker, F.C., Chou-Zheng, L., Dunkle, J.A., and 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., and Zhang, F. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771.

Zhang, J., Graham, S., Tello, A., Liu, H., and White, M.F. (2016). Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res* **44**, 1789-1799.

Zhang, J., Rouillon, C., Kerou, M., Reeks, J., Brugger, K., Graham, S., Reimann, J., Cannone, G., Liu, H., Albers, S.V., Naismith, J.H., Spagnolo, L., and White, M.F. (2012). Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol Cell* **45**, 303-313.

Zhu, X., and Ye, K. (2012). Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett* **586**, 939-945.