Biochemical and Biophysical Research Communications 441 (2013) 720–725

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity

Tae-Yang Kim [a,1], Minsang Shin [a,1], Ly Huynh Thi Yen [b], Jeong-Sun Kim [a,b,*]

[a] Department of Chemistry, Chonnam National University, Gwangju 500-757, Republic of Korea
[b] Interdisciplinary Graduate Program in Molecular Medicine, Gwangju 501-746, Republic of Korea

ARTICLE INFO

Article history:
Received 19 October 2013
Available online 5 November 2013

Keywords:
Cas1
CRISPR
Bacterial immune system
Crystal structure

ABSTRACT

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and CRISPR-associated (Cas) proteins are involved in bacterial acquired immunity against incoming hazardous genetic materials. Cas1 is ubiquitous in CRISPR-containing microorganisms and supposed to recognize and cleave a foreign nucleic acid, and integrate the cleaved fragment into host genome using a yet unidentified mechanism. However, all the reported Cas1s did not show the nucleolytic activity, which makes its role still obscure. The elucidated crystal structure of Cas1 from *Archaeoglobus fulgidus* (*Af*Cas1) shows a butterfly-like dimeric structure. The Asp out of three confirmed nucleolytic residues of Glu, His, and Asp in other Cas1s is replaced with Glu in *Af*Cas1. Further, insertion of five residues into one of two loops, which are close to the catalytic center of and disordered in other Cas1 structures, partially covers the active site of *Af*Cas1. Nonetheless, *in vitro* assays show that its nucleic acid-binding activity was not impaired against the tested single-stranded (ss) DNA, various forms of double-stranded (ds) DNA, or ssRNA with a hydrolyzing activity against ssRNA and dsDNA in a metal ion-dependent way. These results support the proposed Cas1's function at the early step of this bacterial immune system.

1. Introduction

The heritable bacterial immunity against the invading foreign nucleic acids is accomplished by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) and scores of accompanying CRISPR-associated (Cas) proteins [1–4]. The information of past visitors is recorded in the variable sequence (a spacer) between the two specific repeats within the CRISPR locus [1,4,5]. The CRISPR/Cas-based adaptive system includes at least three distinct steps: the recognition and cleavage of incoming exogenous genetic elements and the incorporation of the cleaved fragment into the spacer sequence of the host CRISPR locus, the transcription of the CRISPR locus by the host and processing of a long transcript into a number of small crispr (cr) RNAs, and the silencing of exogenous gene using the small crRNAs as a bait [6].

Among several Cas proteins, Cas1 and Cas2 are ubiquitous, since they are found in almost all the microorganisms that possess the CRISPR/Cas system. Therefore, their existence in a microorganism is a hallmark to identify whether or not an organism contains this kind of immune system. The function of Cas1 has been implicated from the elucidated structure and structure-based functional analysis of Cas1 from *Pseudomonas aeruginosa* strain 14 (*Pa*Cas1), which is a metal-dependent double-stranded (ds) DNA-specific endonuclease and produces dsDNA fragments of approximately 80 base pairs in length. Based on these data, Cas1 has been conceived to take part in the reactions of the early stage in an acquired bacterial immune system by playing a role in the recognition and cleavage of foreign nucleic acids, and possibly in the integration of their fragments into CRISPR loci [7]. Cas1 from *Escherichia coli* K12 (*Ec*Cas1) has also shown an endonucleolytic activity, but it has specificity against the single-stranded (ss) and branched dsDNAs in a metal-dependent manner. Particularly, the recognition and cleavage of branched dsDNAs may coincide with the proposed Cas1 function that integrates the cleaved exogenous products into the CRISPR locus. Further, *in vivo* experiments have exhibited that *Ec*Cas1 physically and genetically interacted with key components of DNA repair systems such as recB, recC, and ruvB, strongly implicating that Cas1 is additionally involved in DNA repair [8]. However, the SSO1450 product, which is a Cas1 from *Sulfolobus solfataricus* (*Ss*Cas1), did not cleave the added nucleic acids in the presence of Mg ions or Ca ions. Since its nuclease activity was tested only with metal ions that are commonly used for the detection of nuclease activity, there is still a possibility that *Ss*Cas1 may require another kind of a metal ion for its nuclease activity, or that it is, as suggested, a restriction endonuclease with a strong zpreference for specific sequences that have yet to be identified

* Corresponding author. Address: Department of Chemistry, College of Natural Science, Chonnam National University, 300 Yongbong-dong, Buk-gu, Gwangju 500-757, Republic of Korea. Fax: +82 62 530 3389.
E-mail address: jsunkim@chonnam.ac.kr (J.-S. Kim).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.bbrc.2013.10.122

[9]. Nonetheless, these contradictory results implicate that nucleolytic activity may not be a common function of all Cas1s and the role of Cas1 is still ambiguous.

Cas1 structures from different organisms have been deposited in the Protein Data Bank (PDB) and publicly released, including Cas1 from *Pyrococcus horikoshii* (*Ph*Cas1; PDB ID 3pv9); Cas1 from *E. coli* K12 (*Ec*Cas1; PDB ID 3nkd) [8]; Cas1 from *Thermotoga maritima* (*Tm*Cas1; PDB ID 3lfx; Midwest Center for Structural Genomics); Cas1 from *P. aeruginosa* UCBPP-PA14 (*Pa*Cas1; PDB ID 3god) [7]; and Cas1 from *Aquifex aeolicus* (*Aa*Cas1; PDB ID 2yzs; RIKEN Structural Genomics). They show a homo-dimeric structure, whose monomer is composed of two distinct domains. The catalytic residues responsible for the hydrolysis of a nucleic acid are located on the charged surface within the C-terminal α-helical domain. However, two loops at the nucleolytic cores are disordered in most of the revealed Cas1 structures, which may play a critical role in the substrate binding, as well as the nucleolysis [7,8].

In this study, the crystal structure of Cas1 from *Archaeoglobus fulgidus* (*Af*Cas1) was determined at a resolution of 2.40 Å, and its *in vitro* function was investigated.

2. Materials and methods

2.1. Cloning, expression, and purification of AfCas1

There are two genes annotated as Cas1 in *A. fulgidus*, for example, AF1878 of 345 residues and AF2435 of 322 residues. The gene coding for *Af*Cas1 (AF1878, Met1-Trp345) was amplified from *A. fulgidus* chromosomal DNA by the polymerase chain reaction (PCR). The PCR product was cloned into pET21a (Invitrogen) derivative, which expresses 25 extra amino acids at the N-terminus including 6 continuous His residues and the Tobacco etch virus (TEV) cleavage sequence. The expression construct was transformed into *E. coli* B834(DE3) strain, which was grown in Luria–Bertani medium containing 100 μg/ml ampicillin at 310 K. After induction by the addition of 0.5 mM IPTG, the culture medium was maintained for a further 8 h at 310 K. Cells were harvested by centrifugation at 5000*g* and 277 K. The cell pellet was re-suspended in buffer A (20 mM Tris–HCl at pH 7.5 and 500 mM NaCl) and then disrupted by ultrasonication. Cell debris was removed by centrifugation at 11,000*g* for 1 h. The fusion protein was purified using a 5-ml HisTrap HP chelating column (GE healthcare, Uppsala, Sweden). The column was washed with buffer A and the bound protein was eluted with a linear gradient of 0–500 mM imidazole in buffer A. After treatment with recombinant TEV protease to cleave the 6 His residue tag and removal of salts by dialysis, the protein, which contains 5 additional amino acids (GGGGG) at the N-terminus and has a theoretical isoelectric point (pI) of 9.45, was purified using a 5 ml HiTrap SP HP cation-exchange column (GE healthcare, Uppsala, Sweden). The column was washed with buffer B (20 mM Tris–HCl at pH 7.5) and the bound protein was then eluted with a linear gradient of 0–1000 mM NaCl in buffer B. Seleno-L-methionine (SeMet) substituted protein was prepared by a procedure similar to that for the native protein. In both cases, the purified protein was >95% pure as judged by Coomassie Blue-stained SDS–PAGE (not shown).

2.2. Crystallization, data collection, and structure determination

For crystallization, the purified protein was concentrated to 11.5 mg ml^{-1} in a buffer of 20 mM Tris–HCl at pH 7.5 and 300 mM NaCl. Its concentration was determined using an extinction coefficient at 280 nm of 0.781 mg ml^{-1} cm^{-1} calculated from the amino-acid sequence. Crystallization of *Af*Cas1 was attempted at 295 K by the sitting-drop vapor-diffusion method. The initial crystallizing condition of the native *Af*Cas1 was obtained from Sparse Matrix Screening [10]. Native crystals suitable for diffraction experiments were obtained within 3 days using a hanging-drop vapor-diffusion method at 295 K by mixing 1 μl of each protein solution and reservoir solutions and equilibrating against 200 μl of reservoir solution, consisting of 30% (w/v) Polyethyleneglycol Monomethylether 550, 0.1 M Bicine at pH 9.0, and 0.1 M Sodium Chloride. The Se-Met substituted protein crystals were obtained under the same crystallizing condition as that of the native crystals. For data collection, 10% (w/v) glycerol was added to the crystallizing precipitant as a cryo-protectant, and the crystals were immediately placed in a 100 K nitrogen-gas stream. Diffraction data for high resolution and those for single-wavelength anomalous dispersion (SAD) with a SeMet crystal were collected at the BL8.3.1 of the Advanced Light Source (ALS) in the United States of America. The indexing, integration, and scaling of the reflections were conducted using a MOSFLM [11] and the ELVES software suite [12]. Thirteen Se sites out of the expected 14 in the asymmetric unit were identified at a resolution of 2.80 Å using the program Phenix [13] combined with SOLVE [14]. The electron density was further improved by density modification using the program Phenix [13] combined with RESOLVE [15], resulting in automatic modeling of approximately 70% of all modeled residues. Further model building was performed manually using the programs WinCoot [16]. When approximately 80% of the residues were built, the initial phase was given up. Subsequently, refinement was performed with PHENIX [13] including TLS restraint using the diffraction data to a final resolution of 2.40 Å. During the refinement, additional amino acids and water molecules were placed, leading to the present final model. The statistics for the collected data and refinement are summarized in Table 1. The quality of the model was analyzed with WinCoot and MolProbity [16,17]. Figures for the ribbon diagram and stick model were prepared using the PyMol Molecular graphics program (Delano Scientific).

2.3. Nucleic acid binding and cleavage assays

For assay on agarose gel, the purified wild-type and mutant proteins were incubated with nucleic acids for 10 min at 298 K in a buffer solution containing 20 mM Tris–HCl at pH 7.5 and 300 mM NaCl. The total volume of the mixture was 10 μl, and the concentration of nucleic acids was 2 μM. The mixtures were loaded onto a 1% (w/v) agarose gel, and electrophoresis was conducted for 40 min at 100 V and 298 K in Tris/Borate/EDTA buffer (89 mM Tris, 89 mM Tris borate, and 2 mM EDTA). The sequences of the ssDNA and ssRNA used were 5′-TTTCTTACACCTCTGT TGCGGTCCACTTGTGTCCACACC-3′ and 5′-AUUGAAAGCAGGAGG-GACCGGAAACACACGGUUGAAGGG-3′, respectively. The dsDNAs used were the amplified *Af*Cas1 gene and plasmid. For assay on a native polyacrylamide gel, ssRNA and dsDNA were end-labeled with [γ-^{32}P] ATP using T4 polynucleotide kinase (Roche). The labeled ssRNA and dsDNA (0.1 μM) were incubated with wild-type and mutant proteins (0.1 μM) for 30 min at 298 K in a buffer solution containing 20 mM Tris–HCl at pH 7.5, 10 mM Magnesium Acetate, 300 mM KCl, 100 μg/ml BSA, and 100 μg/ml Heparin. 10 μl of the mixture was loaded onto a 12% (w/v) non-denaturing polyacrylamide gel (40:1), and electrophoresis was conducted for 60 min at 70 V and 298 K in Tris/Borate/EDTA buffer.

3. Results and discussion

3.1. Overall structure

The crystal structure of *Af*Cas1 was elucidated using SAD method and refined to a resolution of 2.40 Å. In the refined structure, there are three regions that have been poorly traced: seven residues of Glu196-Tyr202 and five of Asp241-Arg245 around the

Table 1
Data collection and structure refinement statistics.

Data collection	Native	Se-peak
Synchrotron	8.3.1, ALS	
Wavelength (Å)	0.98968	0.97956
Space group	$P2_1$	
Cell parameters	a = 70.8 Å, b = 41.91 Å, c = 125.21 Å α = 90°, β = 98.02°, γ = 90°	
Resolution (Å)	65.0–2.4 (2.47–2.4)	100.0–2.8 (2.91–2.80)
Completeness (%)	99.4 (100.0)	95.3 (84.4)
R_{merge}[a]	17.9 (76.4)	12.9 (35.7)
Reflections (total/unique)	80,121/28,962	129,027/18,177
Multiplicity	2.8	7.0
Temperature (K)	100	100
$I/Sigma$[b] (I)	5.1 (1.5)	9.4 (3.8)
FOM[c], solve/resolve (100–2.8 Å)		0.27/0.71
Refinement		
R_{factor}[d]/R_{free}[e]	0.20 (0.25)/0.24 (0.30)	
No. of atoms (protein/water)	5491/355	
Ramachandran plot		
Favored region (%)	98.3	
Allowed region (%)	1.7	
Outliers (%)	0	
Poor rotamers (%)	4.9	
Average B factors (Å^2)		
Protein	29.3	
Solvent	26.7	
R.m.s.d.s		
Bonds (Å)	0.003	
Angles (°)	0.75	

Values in parentheses are for the highest-resolution shell.
FOM, figure of merit; rmsds, root-mean-square-deviations from ideal values [18].
[a] $R_{merge} = \sum_{hkl}[\sum_i(|I_{hkl,i} - < I_{hkl} >|)]/\sum_{hkl,i} < I_{hkl} >$, where $I_{hkl,i}$ is the intensity of an individual measurement of the reflection with Miller indices hkl and $<I_{hkl}>$ is the mean intensity of that reflection.
[b] I/Sigma means <I>/<Sigma>.
[c] Figureof merit $= |\sum P(\alpha)e^{i\alpha}/\sum P(\alpha)|$, where $P(\alpha)$ is the phase probability distribution and α is the phase.
[d] $R_{factor} = \sum_{hkl}||F_{obs}| - |F_{calc}||/\sum_{hkl}|F_{obs}|$; where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively, for the reflections hkl included in the refinement.
[e] R_{free} is the same as R_{factor} but calculated over a randomly selected fraction (6%) of the reflection data not included in the refinement.

catalytic site, and two residues of Lys344-Trp345 at the C-terminus (red-dotted boxes in Fig. 1A). Two protein molecules are present in the asymmetric unit (ASU) with a root-mean-square deviations (rmsds) value of 0.73 Å among the aligned Cα atoms. Analyses of the modeled residues by MolProbity [17] reveal that all refined residues are located in the valid region of the Ramachandran plot (Table 1).

The monomeric *Af*Cas1 forms two globular domains (Fig. 1B): an N-terminal domain (NTD, Met1-His77) of a β-sandwich with a short α-helix and a C-terminal domain (CTD, Ala83-Cys343) of several α-helices with two pairs of short β-strands. Two domains are connected by a loop (Pro78-Thr82) between the β8-strand of the NTD and the α2-helix of the CTD (β8–α2 loop). Like other related structures, a catalytic site is formed within the α-helical CTD. Two *Af*Cas1 molecules in the ASU further assemble to form a butterfly-like dimeric structure (Fig. 1B).

3.2. Comparison of AfCas1's NTD with those of other Cas1s and its interaction for forming a dimeric structure

The NTD comprises one α-helix (α1) and eight short β-strands (β1–β8). The eight β-strands constitute two β-sheets that are formed by four β-strands (β2, β3, β4, and β6) and by the other four (β1, β5, β7, and β8), respectively. The α1-helix is inserted between the β6- and the β7-strands of the second β-sheet (Fig. 2A). A number of non-polar residues emanating from these two β-sheets are put together to form a sandwich-like NTD with openings at both sandwich edges.

In spite of sequence variations (Fig. 1A), the NTDs is well-superposed on all the reported Cas1 structures. *Pa*Cas1 and *Ec*Cas1 contain approximately ten extra residues at the N-terminus compared with other Cas1s, which form an α-helix in *Pa*Cas1 and no secondary structure in *Ec*Cas1 [7,8]. *Pa*Cas1 has another insertion of ten residues between the β7- and the β8-strands, which form a short β-sheet in one protomer, but is disordered in the other protomer [7].

A dimeric *Af*Cas1 structure is formed mainly via the interaction of two NTDs. Three secondary structural elements are involved in the dimerization of NTD, where both polar and non-polar interactions are simultaneously observed. The β6- and the β8-strands from both protomers interact hydrophobically to form two curved β-sheets of eight β-strands (Fig. 2A). The inserted α1-helix further contributes to the dimerization of two NTDs, where the N-terminal α-helical cap of one protomer interacts hydrophilically with main-chain atoms of the β5–β6 loop, and hydrophobically with the β8-strand of the other protomer (Fig. 2B). The extra ten residues at the N-terminus of *Pa*Cas1 and *Ec*Cas1 and another inserted region between the β7- and the β8-strands in *Pa*Cas1 are located outside from the β-sandwich core. Overall, the NTD structure of *Af*Cas1 and its interaction in the dimeric structure are well conserved.

3.3. Comparison of AfCas1's CTD with those of other Cas1s

Except for several invariant or highly conserved residues that constitute a catalytic site, the CTDs of Cas1s also show sequence diversity (Fig. 1A). Approximately C-terminal 260 residues of *Af*Cas1 comprise eight α-helices. Compared with *Ec*Cas1 and *Pa*Cas1, *Af*Cas1 has some insertions of four short β-strands between the α7- and the α8-helices (β9 and β10) and the α8- and the α9-helices (β11 and β12) (Fig. 1). Nonetheless, *Af*Cas1's CTD is superposed well with those of other Cas1s with small rmsd values of less than 3.0 Å among the compared 230 Cα atoms.

A closer look at the aligned sequences, however, reveals several regions, resulting in the minute structural differences. First, the extreme C-terminus shows little sequence conservation among Cas1s (Fig. 1A), which are exposed to solvent. Second, the α4-helices especially vary in terms of their lengths (Fig. 1A). Third, another sequence-variable region between the α7- and the α8-helices forms a pair of β-strands in *Af*Cas1, *Tm*Cas1, and *Ph*Cas1. However, approximately fifteen residues at the corresponding region are lacking in *Pa*Cas1 and *Ec*Cas1 (Fig. 1A). Overall, these sequence variations appear not to have a significant effect on the enzymatic activity of Cas1 since they are located far apart from the putative nucleolytic core. However, it is worthwhile to mention the inserted region (^{194}GIEIG^{198}) between the α5- and the α6-helices (α5–α6 loop) in *Af*Cas1, which is unique in only *Af*Cas1 (Fig. 1A). The α5–α6 and the α6–α7 loops around the nucleolytic site have been poorly defined in other Cas1s without a bound substrate [7,8]. In *Af*Cas1, main-chain atoms of the α5–α6 loop have a weak calculated electron density (Fig. 3A), while the α6–α7 loop is well-traced. Therefore, this flexible α5–α6 loop is likely to undergo disorder-to-order changes upon substrate-binding.

3.4. AfCas1 has a metal ion-dependent nucleolytic activity against ssRNA and dsDNA

Even though there is non-definitive substrate specificity in the catalytically active Cas1s of *Pa*Cas1 and *Ec*Cas1, they degraded a

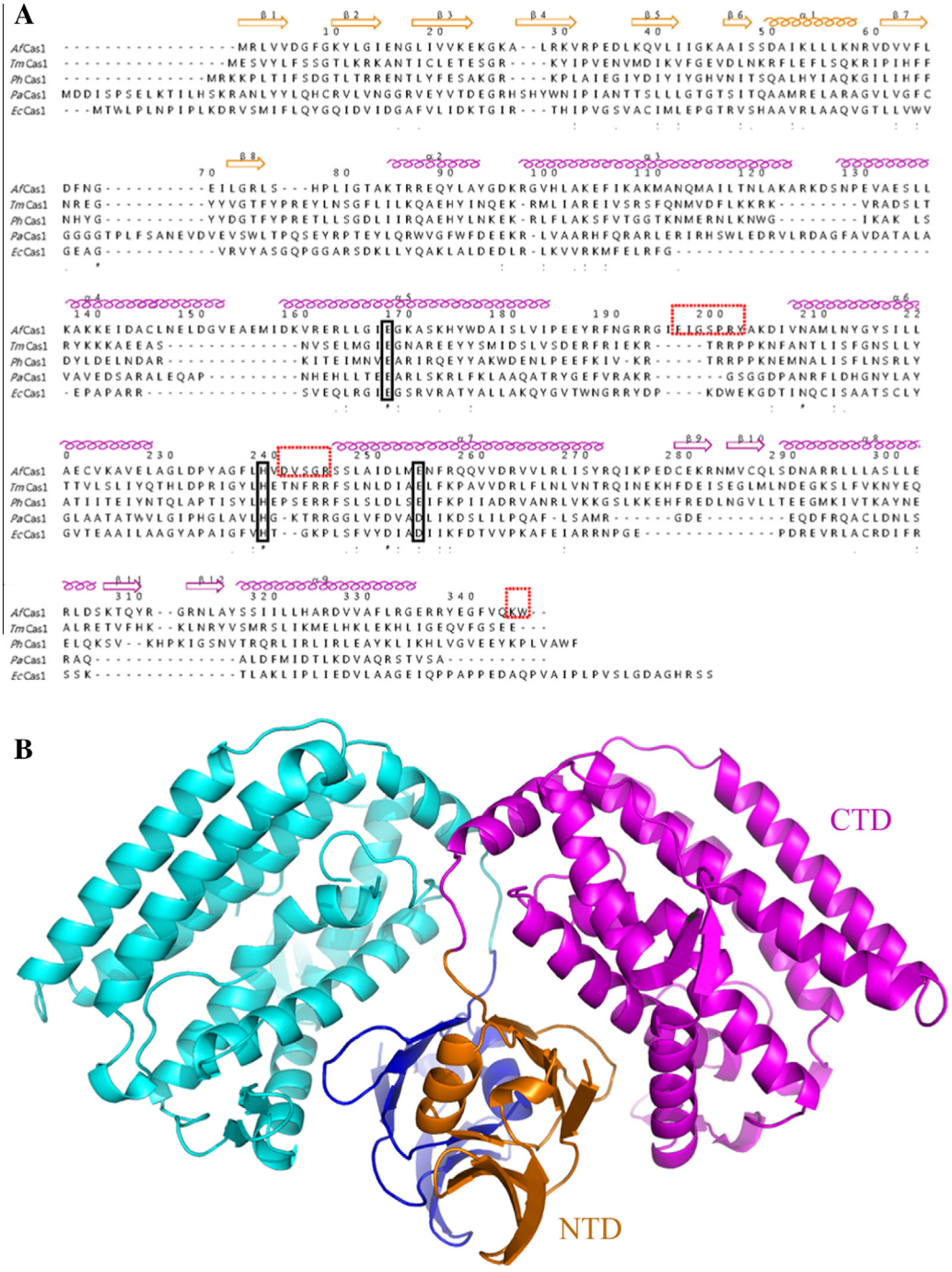

Fig. 1. Overall structure of *Af*Cas1. (A) Structure-based sequence alignment of Cas1. The amino acid sequences are shown for Cas1 from *Thermotoga maritima* (*Tm*Cas1, PDB ID 3lfx), Cas1 from *Pseudomonas aeruginosa* UCBPP-PA14 (*Pa*Cas1, PDB ID 3god), Cas1 from *Pyrococcus horikoshi* (*Ph*Cas1, PDB ID 3pv9), Cas1 from *Escherichia coli* str. K-12 (*Ec*Cas1, PDB ID 3nkd), and Cas1 from *Archeaoglobus fulgidus* (*Af*Cas1, PDB ID 4n06). The coils and arrows above the aligned sequences represent α-helices and β-strands observed in *Af*Cas1, respectively. The secondary structures of NTD are displayed in orange, whereas those of CTD are shown in magenta. The numbering scheme follows the amino acid sequence of *Af*Cas1. The residues that form a nucleolytic core are enclosed by black boxes. The regions that have weak electron density in the present structure are indicated in red-dotted boxes. Identical residues are marked by "*", while conserved residues are indicated by ":" and ".", respectively. The sequence alignment was prepared with ClustalW2 software from the European Bioinformatics Institute based on primary amino acid sequences [19]. (B) Ribbon diagram of the dimeric *Af*Cas1. In the dimeric structure, one protomer is gold and magenta, and the other is blue and cyan. The figures, except for Fig. 1(A), were prepared by PyMol Molecular Graphics software from Delano Scientific. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

substrate in a divalent metal-dependent manner [7–8]. Three polar residues are structurally conserved at the active site within the α-helical CTD, such as with Glu190, His254, and Asp268 in *Pa*Cas1 and Glu141, His208, and Asp221 in *Ec*Cas1, respectively. Substitutions of these residues into the respective Ala eliminated their nucleolytic activity. In this regard, the conservation of these three residues within the α-helical CTDs has been considered as a strong evidence for the classification of Cas1 as a nuclease.

The catalytic site of *Af*Cas1 can be found by superposition onto the *Pa*Cas1 and *Ec*Cas1 structures in combination with a multiple sequence alignment (Figs. 1A and 3A). Three polar residues of Glu168, His239, and Glu254 are located within a short distance

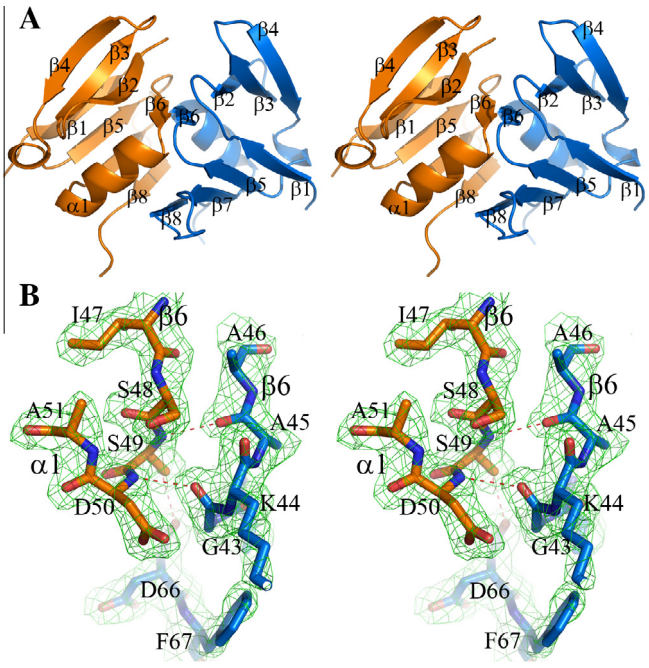

Fig. 2. Dimerization of *Af*Cas1 through the interaction between the two NTDs. (A) Stereo representation of two sandwich-like NTDs. The respective NTDs are shown as a ribbon diagram and differentiated by blue and gold for each protomer. (B) Stereo representation of the polar interaction between the two NTDs. $2F_o$-F_c density (marine) is contoured at 1.0 σ. The residues at the contacting edge between the α1-helix and the β5–β6 region are represented as stick models. N and O atoms are colored blue and red, respectively. Hydrogen bonds are displayed as red-dotted lines between the atoms. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

at the structurally equivalent positions to the catalytic residues of *Ec*Cas1 and *Pa*Cas1 (Fig. 3A). The last Asp among the three residues in other active Cas1s is replaced with Glu residue in *Af*Cas1, which is conserved in *Tm*Cas1 and *Ph*Cas1 (Fig. 1A). In the overall CTD structures, the residues on the α5–α7 secondary elements form the putative nucleolytic site of *Af*Cas1 (Fig. 3A). Even though two loops of α5–α6 and α6–α7 are partially disordered (Fig. 1A), the catalytic His239 residue on the α6–α7 loop (Figs. 1A and 3A) has a well-defined electron density in *Af*Cas1. Five residues inserted in the α5–α6 loop of *Af*Cas1 form an elongated structure that partially covers the putative nucleolytic core of *Af*Cas1 (Fig. 3A), suggesting that it might not be easily accessible from solvent.

The mapping of charge potentials on the surface of Cas1 dimeric structures displays interesting polarized features, showing a positively charged surface in the vicinity of the nucleolytic core and a negatively charged surface far from the core (Fig. 3B). Another positively charged surface is located on the other side of the active site with the CTD of one Cas1 protomer, but faces the same side as the active site of the other Cas1 protomer. These overall charge distributions on the surface may accommodate easy and focused access of a nucleic acid to the nucleolytic core of Cas1.

In spite of several minute structural differences around the putative active site, the partial conservation of common nucleolytic residues of Cas1s in *Af*Cas1 suggest that *Af*Cas1 may function as a nuclease. The nucleolytic activity of the recombinant *Af*Cas1 was tested against various nucleic acids, such as ssDNA, ssRNA, or dsDNA, with different divalent metal ions. *Af*Cas1 was bound to all the tested nucleic acids regardless of metal ions (Fig. 4A). However, *Af*Cas1 cleaved the ssRNA and dsDNA independent of their sequences, but dependent on metal ions in our tested conditions (Fig. 4B). To confirm the involvement of three conserved residues in its nucleolytic activity, they were mutated into Ala residue. All the three respective mutations dramatically decreased its nuclease activity against ssRNA (Fig. 4B), indicating that they are responsible for the nucleolytic activity of *Af*Cas1 like in other Cas1 proteins. Therefore, the replacement of Asp residue with the Glu in the third catalytic residue and the presence of five residues inserted in the α5–α6 loop of *Af*Cas1 at the catalytic site do not

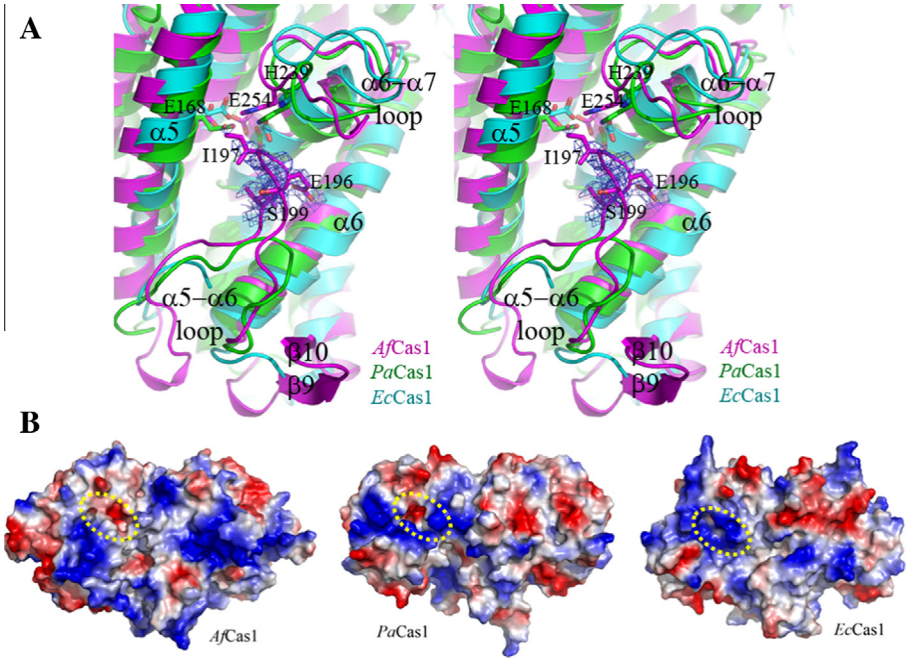

Fig. 3. Nucleolytic core of *Af*Cas1. (A) Comparison of the nucleolytic cores of three Cas1s. The residues that constitute a nucleolytic core are displayed with stick models and differentiated by colors. Only the residues of *Af*Cas1 are labeled for clarity. The five residues inserted in the α5-α6 loop of *Af*Cas1 are displayed as stick models with $2F_o$-F_c density (blue) contoured at 0.5 σ. (B) Surface potential map of Cas1s. The positively and negatively charged surfaces are colored in blue and red, respectively. The nucleolytic cores of Cas1s are indicated by yellow-dotted circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

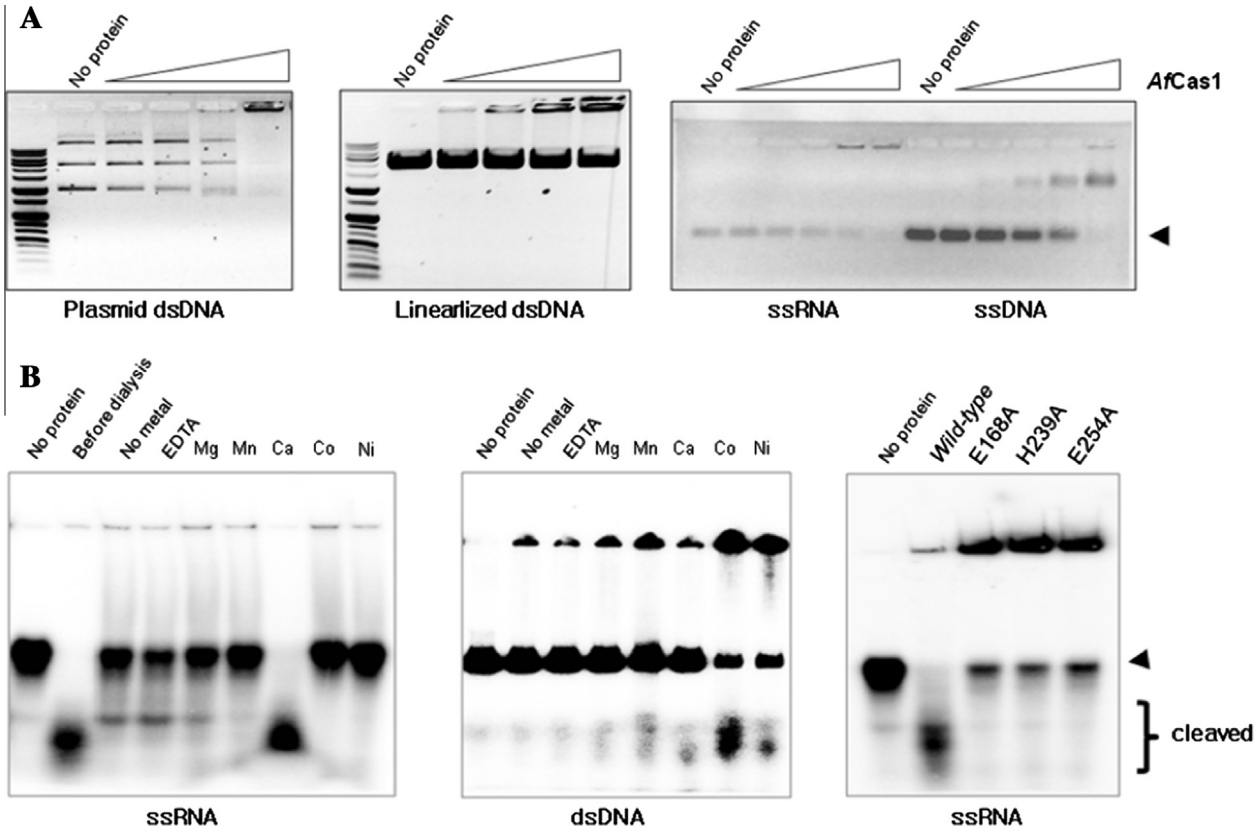

Fig. 4. (A) Nucleolytic activity of *Af*Cas1. Nucleic acid-binding activity of *Af*Cas1. The mobility shifts of plasmid and linearized dsDNAs (left), and ssRNA and ssDNA (right) were carried out on 1% (w/v) agarose gel and visualized by Ethidium bromide staining. The triangle indicates the added nucleic acids. (B) Nucleolytic activity of *Af*Cas1. The mixtures of radioactively-labeled ssRNA and dsDNA and the proteins were analyzed on 12% (w/v) polyacrylamide gel with various divalent metal ions. The triangle indicates the added nucleic acids.

interfere the dsDNA-specific nuclease activity of *Af*Cas1. The presented data here confirm that the proposed Cas1's activity in the early stage of bacterial innate immunity is applied to *Af*Cas1, although the role of its ssRNA-specific nucleolytic activity is unclear still yet.

Data deposition

The coordinates and structure factors of *Af*Cas1 were deposited in the Protein Data Bank under the accession number 4n06.

Acknowledgments

This study was financially supported by Chonnam National University, 2011. The authors wish to thank the staff scientists at ALS for their assistance with data collection.

References

[1] A. Bolotin, B. Quinquis, A. Sorokin, S.D. Ehrlich, Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin, Microbiology 151 (2005) 2551–2561.
[2] F.J. Mojica, C. Diez-Villasenor, J. Garcia-Martinez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, J. Mol. Evol. 60 (2005) 174–182.
[3] R. Sorek, V. Kunin, P. Hugenholtz, CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea, Nat. Rev. Microbiol. 6 (2008) 181–186.
[4] J. van der Oost, M.M. Jore, E.R. Westra, M. Lundgren, S.J. Brouns, CRISPR-based adaptive and heritable immunity in prokaryotes, Trends Biochem. Sci. 34 (2009) 401–407.
[5] S.J. Brouns, M.M. Jore, M. Lundgren, E.R. Westra, R.J. Slijkhuis, A.P. Snijders, M.J. Dickman, K.S. Makarova, E.V. Koonin, J. van der Oost, Small CRISPR RNAs guide antiviral defense in prokaryotes, Science 321 (2008) 960–964.
[6] H. Deveau, J.E. Garneau, S. Moineau, CRISPR/Cas system and its role in phage-bacteria interactions, Annu. Rev. Microbiol. 64 (2010) 475–493.
[7] B. Wiedenheft, K. Zhou, M. Jinek, S.M. Coyle, W. Ma, J.A. Doudna, Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated antiviral defense, Structure 17 (2009) 904–912.
[8] G. Brown, A. Binkowski, S. Phanse, A. Joachimiak, E.V. Koonin, A. Savchenko, A. Emili, J. Greenblatt, A.M. Edwards, A.F. Yakunin, A dual function of the CRISPR–Cas system in bacterial antivirus immunity and DNA repair, Mol. Microbiol. 79 (2011) 484–502.
[9] D. Han, K. Lehmann, G. Krauss, SSO1450 – A CAS1 protein from *Sulfolobus solfataricus* P2 with high affinity for RNA and DNA, FEBS Lett. 583 (2009) 1928–1932.
[10] J. Jancarik, S.-H. Kim, Sparse matrix sampling: a screening method for crystallization of proteins, J. Appl. Crystallogr. 24 (1991) 409–411.
[11] A.G. Leslie, Joint CCP4 ESF-EAMCB, Newslett. Protein Crystallogr. 26 (1992) 1–3.
[12] J. Holton, T. Alber, Automated protein crystal structure determination using elves, Proc. Natl. Acad. Sci. USA 101 (2004) 1537–1542.
[13] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. D66 (2010) 213–221.
[14] T.C. Terwilliger, J. Berendzen, Automated MAD and MIR structure solution, Acta Crystallogr. D55 (1999) 849–861.
[15] T.C. Terwilliger, Automated main-chain model building by template matching and iterative fragment extension, Acta Crystallogr. D59 (2003) 38–44.
[16] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D60 (2004) 2126–2132.
[17] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, L.W. Murray, W.B. Arendall 3rd, J. Snoeyink, J.S. Richardson, D.C. Richardson, MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes, Nucleic Acids Res. 32 (2007) W615–W619.
[18] R.A. Engh, R. Huber, Accurate bond and angle parameters for X-ray protein structure refinement, Acta Crystallogr. A 47 (1991) 392–400.
[19] M.A. Larkin, G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, Clustal W and Clustal X version 2.0, Bioinformatics 23 (2007) 2947–2948.