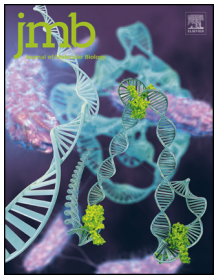

Crystal Structures of Csm2 and Csm3 in the Type III-A CRISPR–Cas Effector Complex

Daijiro Takeshita, Momoe Sato, Hideko Inanaga and Tomoyuki Numata

Biomedical Research Institute, *National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Higashi, Tsukuba-shi, Ibaraki 305-8566, Japan*

Correspondence to Tomoyuki Numata: *t-numata@aist.go.jp*
https://doi.org/10.1016/j.jmb.2019.01.009
Edited by Philip C Bevilacqua

Abstract

Clustered regularly interspaced short palindromic repeat (CRISPR) loci and CRISPR-associated (Cas) genes encode CRISPR RNAs (crRNA) and Cas proteins, respectively, which play important roles in the adaptive immunity system (CRISPR–Cas system) in prokaryotes. The crRNA and Cas proteins form ribonucleoprotein effector complexes to capture and degrade invading genetic materials with base complementarity to the crRNA guide sequences. The Csm complex, a type III-A effector complex, comprises five Cas proteins (Csm1–Csm5) and a crRNA, which co-transcriptionally degrades invading DNA and RNA. Here we report the crystal structures of the *Staphylococcus epidermidis* Csm2 (*Se*Csm2) and *Thermoplasma volcanium* Csm3 (*Tv*Csm3) at 2.4- and 2.7-Å resolutions, respectively. *Se*Csm2 adopts a monomeric globular fold by itself, in striking contrast to the previously reported *Thermotoga maritima* Csm2, which adopted an extended conformation and formed a dimeric structure. We propose that the globular monomeric form is the bona fide structure of Csm2. *Tv*Csm3 forms a filamentous structure in the crystals. The molecular arrangement of *Tv*Csm3 is similar to that of the stacked Cmr4 proteins in the Cmr complex, suggesting the functionally relevant architecture of the present Csm3 structure. We constructed model structures of the Csm complex, which revealed that Csm3 binds the crRNA and periodically deforms the crRNA-target duplex by a similar mechanism to that of Cmr4 in the Cmr complex. The model and mutational analysis suggest that the conserved lysine residue of Csm2 is important for target RNA binding, and Csm2 stabilizes the active structure of the Csm complex to facilitate the reaction.

Introduction

Clustered regularly interspaced short palindromic repeat (CRISPR) loci and CRISPR-associated (Cas) genes serve as a prokaryotic adaptive immune system that mediates RNA-guided interference to protect bacteria and archaea against invading genetic elements, such as phages and plasmids [1–4]. CRISPR loci contain arrays of short invariant repeats interspaced with variable sequences that are derived from previously encountered foreign genetic materials [5–7]. Prokaryotes incorporate these invader-derived sequences into CRISPR loci to protect themselves from subsequent genetic invasions. The primary transcripts of CRISPR loci are processed within each repeat to generate mature CRISPR RNAs (crRNAs) [8–10]. The crRNAs and Cas protein(s) form effector complexes that degrade invading nucleic acids in a base complementary manner, by using the crRNAs as guides [8,11,12].

CRISPR–Cas systems are divided into two main classes, based on the Cas protein composition of the effector complexes [13]. The class 1 effector complexes comprise multiple Cas proteins, while only a single Cas protein is required for the class 2 effector complexes. The class 1 CRISPR–Cas system is further divided into two different types, I and III. The type I effector complex, as represented by Cascade (CRISPR-associated complex for antiviral defense), is a large multisubunit ribonucleoprotein complex that targets double-stranded DNA [8]. Cascade unwinds the target double-stranded DNA to form an R-loop structure, which subsequently recruits the helicase/nuclease Cas3 to degrade the target DNA [8,14,15]. In

contrast, the type III effector complex, which is also a multisubunit ribonucleoprotein complex [11,16], exhibits dual DNA and RNA interference activities, in which the DNA degradation is mediated in a transcription-dependent manner [17–20]. The structures of the type I and III effector complexes have been characterized by crystallography and cryoelectron microscopy (cryoEM). These analyses, together with a bioinformatics study, revealed their structural similarity and divergent evolution from a common ancestral complex [21–38].

The type III CRISPR–Cas system is further classified into four subtypes, III-A, III-B, III-C, and III-D, based on the composition and organization of the type III CRISPR–Cas loci [13]. The Csm and Cmr complexes, which are well-characterized subtype III-A and III-B effector complexes, respectively, comprise five (Csm1–Csm5) and six (Cmr1–Cmr6) proteins in addition to the crRNA [11,16]. Csm1 and Cmr2 in the Csm and Cmr complexes, respectively, belong to a Cas10 protein family that is a hallmark subunit of the type III CRISPR–Cas system [13] and exhibits DNase activity [17–20]. The structural studies of the type III effector complexes revealed their filamentous architecture consisting of two helical chains [26–31]. One chain is formed from a Csm4/Cmr3 molecule and several Csm3/Cmr4 molecules, and the other is from a Csm1/Cmr2 molecule and several Csm2/Cmr5 molecules in the Csm/Cmr complexes, respectively. The crystal structure of the hybrid Cmr complex lacking Cmr1 (simply called the Cmr complex hereafter) elucidated the interactions between the Cmr proteins and the crRNA-target duplex [30]. In the Cmr complex, the crRNA-target duplex adopts an unwound ribbon-like structure due to the intercalations of the idiosyncratic loops (thumbs) of the stacked Cmr4 subunits. These intercalated Cmr4 thumbs periodically break the base-pairings between the crRNA and target, causing base-flipping from the duplex at 6-nt intervals. The flipped target nucleotides are placed near the conserved aspartate of Cmr4, which serves as a catalytic residue in the Cmr complex [30,39,40]. The structure explains the mechanism by which the Cmr complex cleaves the target at multiple sites with 6-nt intervals.

The cryoEM structures and biochemical experiments of the Csm complex suggested that the mechanisms of target RNA recognition and cleavage by this complex are analogous to those by the Cmr complex [26,29,41]. The crystal structures of all Csm protein subunits, except for Csm5, have been reported [42–45]. Partial structural models of the Csm complex were proposed, by superimposing the crystal structures of the Csm proteins onto the cryoEM electron density. However, the complex model should be further examined in greater detail, since the subunit compositions of Csm2, Csm3, and Csm4 in the Csm complex have remained unclear. In particular, the subunit composition of Csm2 and its overall structure in the functional state need to be re-assessed. In the crystal structure of *Thermotoga maritima* Csm2 (*Tm*Csm2), the protein reportedly formed a homodimer [45], which is apparently inconsistent with the results of cryoEM and biochemical experiments showing that the Csm complex contains three Csm2 molecules [29,41]. Venclovas [46] pointed out that the recombinant *Tm*Csm2 prepared by Gallo *et al.* [45] existed as both a monomer and dimer, and suggested the possibility that the monomer form of Csm2 is incorporated into the Csm complex in nature. Therefore, the functional state of the Csm2 protein has been controversial and remains to be elucidated. Csm3 is an important protein component that forms a helical backbone and binds the crRNA in the Csm complex [26,29,41]. However, the mechanisms of the self-association and crRNA binding by Csm3 have remained elusive, because the crystal structures of Csm3 determined to date only exhibited a monomeric form or heterodimeric form with Csm4 [42,43]. Therefore, these diverse findings have hampered our understanding of the structural and functional relationships of the Csm complex.

In this study, we solved the crystal structures of *Staphylococcus epidermidis* Csm2 (*Se*Csm2) in a monomeric form and *Thermoplasma volcanium* Csm3 (*Tv*Csm3) in a multimerized form at 2.4- and 2.7-Å resolutions, respectively. The *Se*Csm2 monomeric structure consistently explains its subunit composition in the Csm complex suggested previously [29,41]. The multimerized *Tv*Csm3 forms a helical filament almost identical to the backbone of the Csm complex [29]. Based on these structures, together with the other previously reported structures, we constructed a model of the Csm complex that delineates the interactions between the Csm proteins and the crRNA-target duplex.

Results and Discussion

Structure determination of *S. epidermidis* Csm2

In the crystal structure of *T. maritima* Csm2 (*Tm*Csm2), the protein reportedly formed a homodimer, in which the H3 helix from one protomer interacts with the H3 helix from the other protomer to form an interdigitated dimer with a symmetric structure [45]. However, this crystallographic result is inconsistent with the finding that the Csm complex consists of three Csm2 molecules, as suggested by biochemical and cryoEM structural studies [29,41], because if Csm2 formed a homodimer, then the Csm complex would contain an even number of Csm2 molecules. Therefore, a reanalysis of the Csm2 structure is essential to clarify the structural and functional relationships of the Csm complex.

To reveal the bona fide structure of Csm2, we sought to determine the crystal structures of Csm2 proteins

from other species. We expressed and purified the Csm2 proteins from *S. epidermidis* (*Se*Csm2) and *Methanocaldococcus jannaschii* (*Mj*Csm2). We also prepared *Tm*Csm2, to reevaluate its oligomeric state in solution. The size exclusion chromatography (SEC) analysis revealed that the molecular weights of *Se*Csm2, *Mj*Csm2, and *Tm*Csm2 are 10.3, 15.0, and 16.1 kDa, respectively (Fig. 1a and Supplementary Fig. 1). The molecular weights of *Mj*Csm2 and *Tm*Csm2 are similar to the theoretical molecular weight values of these proteins (17.8 and 18.8 kDa, respectively), while that of *Se*Csm2 is slightly smaller than the theoretical value (17.6 kDa). The minor difference between the calculated and theoretical molecular weight values of *Se*Csm2 might be attributable to its molecular properties under our assay conditions. Together, these results demonstrated that the Csm2 proteins from various species exist as monomers in solution. The previously purified *Tm*Csm2 by Gallo *et al*. [45] contained both monomeric and dimeric forms, in contrast to the present study where we did not detect the peak corresponding to the *Tm*Csm2 dimer (Fig. 1a). The difference between these two studies might be due to variations in the experimental conditions.

We solved the crystal structure of *Se*Csm2 by the single-wavelength anomalous diffraction (SAD) method, using selenomethionine-labeled crystals. The structure was refined at 2.4-Å resolution, with an R_{work}/R_{free} of 0.212/0.233 (Table 1). The current model contains two molecules in the asymmetric unit (molecules A and B) and includes residues 1–13 and 25–125 for both molecules. The structures of the two *Se*Csm2 molecules in the asymmetric unit are quite similar, with a r.m.s. deviation of 0.206 Å in the superimposition of the corresponding 114 Cα-atoms.

Crystal structure of *S. epidermidis* Csm2

*Se*Csm2 is a globular protein with six α-helices that exists as a monomer in the crystals (Fig. 1b), consistent with the SEC analysis (Fig. 1a). The *Se*Csm2 structure is quite different from the previously reported *Tm*Csm2 structure [45], which exhibited the extended structures of each protomer and formed a homodimer (Fig. 1c). The structural difference is exclusively attributable to the different arrangement of the helices/helix between these two Csm2 proteins. In the *Tm*Csm2 crystal structure, residues 61–103 form a long H3 helix, thereby resulting in the extended conformation of this protein [45] (Fig. 1c). In contrast, the corresponding region in *Se*Csm2 (residues 59–101) is divided into three helices, α3, α4, and α5, and the latter two helices are folded back to form a globular

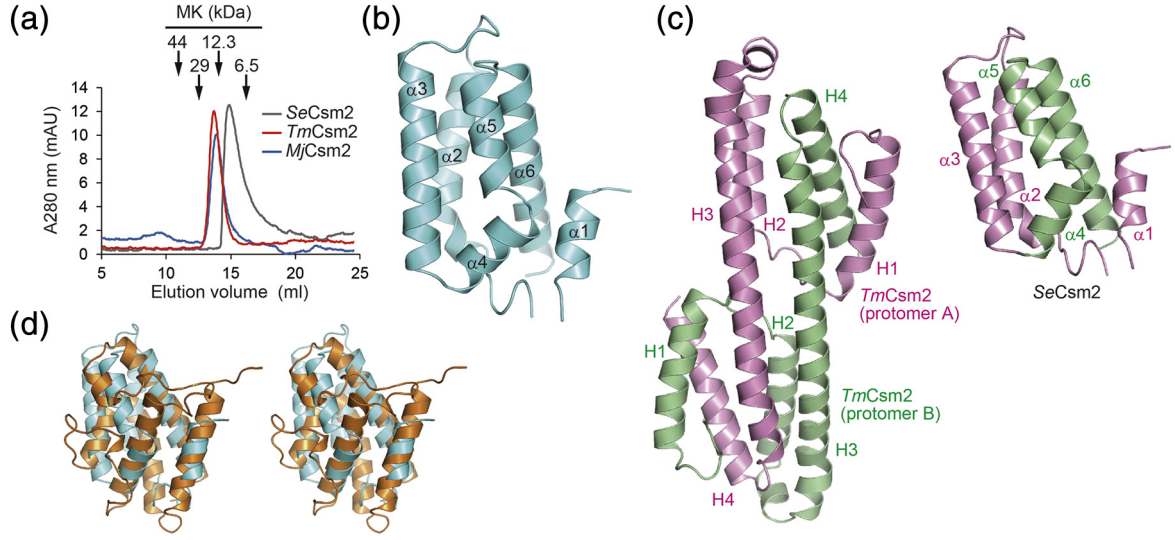

Fig. 1. Structure of *S. epidermidis* Csm2. (a) SEC of the purified Csm2 proteins from *S. epidermidis* (gray), *T. maritima* (red), and *M. jannaschii* (blue) on a Superdex 75 HR 10/300 column. For calibration, conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), cytochrome *C* (12.3 kDa), and aprotinin (6.5 kDa) were used as molecular weight markers to calculate the standard curve for the column. The elution volumes for these molecular weight markers are indicated above the chromatograms. (b) Overall structure of *Se*Csm2. The secondary structure elements are labeled. (c) Structural comparison of *Se*Csm2 (right) and the *Tm*Csm2 dimer [45] (PDB ID: 5AN6) (left). The A and B protomers of the *Tm*Csm2 dimer are pink and green, respectively. *Se*Csm2 is colored pink (residues 1–79) and green (residues 80–125), and these regions can be superimposed onto the N-terminal part of the A protomer and the C-terminal part of the B protomer of the *Tm*Csm2 dimer, respectively. The H3 helices in both protomers of *Tm*Csm2 are involved in the dimerization of this protein. The corresponding region of *Se*Csm2 is divided into three parts (helices α3, α4, and α5). (d) Stereoview of the superimposed structures of *Se*Csm2 (light blue) and *A. fulgidus* Cmr5 [30] (orange).

Table 1. Summary of data collection and refinement statistics

	*Se*Csm2	*Tv*Csm3
Data collection		
Space group	*C*2	$P4_12_12$
Cell dimensions		
a, *b*, *c* (Å)	98.5, 67.5, 69.6	78.3, 78.3, 182.7
Wavelength	0.97875	0.9788
Resolution (Å)	50–2.4 (2.44–2.40)	50–2.7 (2.75–2.70)
R_{merge}	0.080 (0.505)	0.080 (0.466)
$I/\sigma I$	21.4 (2.3)	60.7 (11.5)
Completeness (%)	100.0 (99.8)	99.7 (100.0)
Redundancy	6.6 (5.6)	13.9 (14.2)
Refinement		
Resolution (Å)	49.3–2.4	39.5–2.7
No. reflections	12,643	16,271
R_{work}[a]/R_{free}[b]	0.212/0.233	0.211/0.243
No. atoms	1,940	2,867
B-factors (Å^2)	61.7	49.7
R.m.s. deviations		
Bond lengths (Å)	0.002	0.002
Bond angles (°)	0.376	0.592
Ramachandran plot		
Most favored (%)	94.9	90.9
Additionally allowed (%)	5.1	8.8
Generously allowed (%)	0	0.3
Disallowed (%)	0	0

The values in parentheses are for the outermost shell.
[a] $R_{work} = \Sigma | F_o - F_c | / \Sigma F_o$ for reflections of working set.
[b] $R_{free} = \Sigma | F_o - F_c | / \Sigma F_o$ for reflections of test set (5.0% of total reflections).

structure (Figs. 1c and 2a). The α6 helix is located in a similar position to the C-terminal H4 helix of protomer B of *Tm*Csm2 (Fig. 1c). Gallo *et al.* [45] reported previously that *Tm*Csm2 exists as both a monomer and dimer in solution, as judged by SEC and LC-MS experiments. Therefore, the monomeric structure of *Tm*Csm2 seems to be quite similar to the present *Se*Csm2 structure. The reported dimeric structure of *Tm*Csm2 would not be a functionally relevant state, since the *Thermus thermophilus* Csm complex reportedly contained three Csm2 molecules [29], which would be incompetent if Csm2 existed as a dimer. Based on the sequence alignment and the secondary structure prediction of Csm2, Venclovas [46] suggested that the long H3 α-helix observed in the reported *Tm*Csm2 structure can be divided into two parts, and its C-terminal portion and the following H4 helix fold back into the N-terminal region of the protein. This possible structural rearrangement may form a globular monomeric structure of Csm2 that would be similar to the present *Se*Csm2 structure. Considering these findings together, we strongly propose that Csm2 alone forms a globular conformation, as observed in the present work.

The *Se*Csm2 structure shares structural similarity with Cmr5, which is the smallest subunit of the type III-B Cmr complex [30,47] (Fig. 1d), although these two protein families only have low-sequence similarity. The *Se*Csm2 structure was superimposed onto the *Archaeoglobus fulgidus* and *T. thermophilus* Cmr5 structures (PDB IDs: 3X1L and 2ZOP, respectively) with r.m.s. deviations of 2.81 Å (84 aligned Cα atoms) and 2.71 Å (88 aligned Cα atoms), respectively.

Structure determination of *T. volcanium* Csm3

Several Csm3 molecules associate around the crRNA to form the helical backbone of the Csm complex [26,29]. The crystal structures of *Methanopyrus kandleri* Csm3 (*Mk*Csm3) in the monomeric form [42] and *M. jannaschii* Csm3 (*Mj*Csm3) in the complex with Csm4 [43] have been determined. However, there is no detailed information about the multimeric Csm3 structure that can provide molecular insights into its function, in the context of the Csm complex. In an attempt to obtain the multimerized Csm3 structure, we tried to crystallize the Csm3 proteins from several bacterial and archaeal species (*M. jannaschii*, *Pyrococcus horikoshii*, *T. volcanium*, *S. epidermidis*, and *T. maritima*). The *T. volcanium* Csm3 (*Tv*Csm3) crystals were successfully obtained, and the crystal structure was solved by the molecular replacement method, using the *Mj*Csm3 structure as the search model. The *Tv*Csm3 structure was refined to an R_{work}/R_{free} of 0.211/0.243 at 2.7-Å resolution (Table 1). There are two molecules in the asymmetric unit (molecules A and B), and the model contains residues 5–26, 37–91, 102–132, and 144–216 for molecule A and residues 5–26, 37–91, 102–132, and 145–216 for molecule B. The structures of the two *Tv*Csm3 molecules in the asymmetric unit can be superimposed with an r.m.s. deviation of 0.673 Å for the corresponding 179 Cα-atoms. The *Tv*Csm3 structure resembles the Cmr4 structure reported previously [30,39], consistent with the fact that the Csm3 and Cmr4 proteins belong to the Cas7 family [13].

Crystal structure of *T. volcanium* Csm3

Two *Tv*Csm3 molecules interact with each other in the asymmetric unit (Fig. 3a), in which the interface of the proteins buries a solvent-accessible surface area of 1,017 Å^2, corresponding to approximately 10% of the total solvent-accessible surface area of the protein. The residues (Glu128, Lys130, Arg149, Asp186, and Ser187) from one-molecule hydrogen bond with the residues (Tyr57, Ser61 Arg108, Asp109, Lys210, and Tyr215) of the other molecule (Fig. 3b). Many of these residues are well-conserved in the Csm3 protein family (Fig. 2b). Furthermore, many hydrophobic residues from each protomer strengthen the interactions at the interface.

Intriguingly, the two interacting *Tv*Csm3 molecules in the asymmetric unit contact the adjacent molecules on both sides in the crystal lattice (Fig. 3c), where the same interactions observed between the two *Tv*Csm3 molecules in the asymmetric unit stabilize the contacts. Therefore, many of the residues involved in the interactions with the symmetry-related molecules are

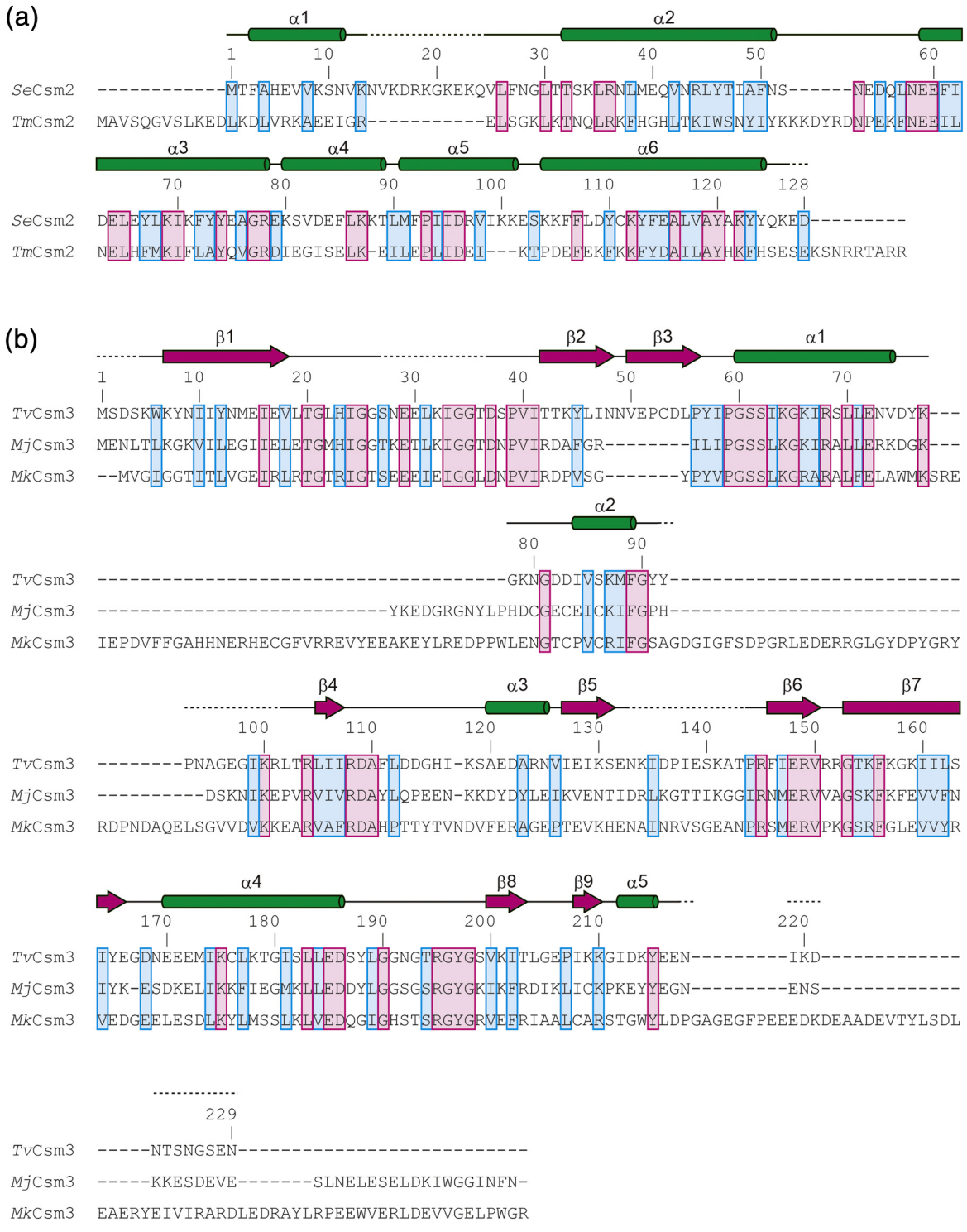

Fig. 2. Structure-based sequence alignments of Csm2 (a) and Csm3 (b). The sequence alignments were performed with ClustalW, with manual editing based on the aligned structures. The proteins from *S. epidermidis*, *T. maritima*, *T. volcanium*, *M. jannaschii*, and *M. kandleri* are abbreviated as *Se*, *Tm*, *Tv*, *Mj*, and *Mk*, respectively. The amino acid residue numberings and secondary structure elements of *Se*Csm2 and *Tv*Csm3 are indicated above the alignments. Conserved residues and conservative replacements are colored pink and cyan, respectively. The α-helices and β-sheets are colored green and magenta, respectively. The disordered regions are shown as dotted lines.

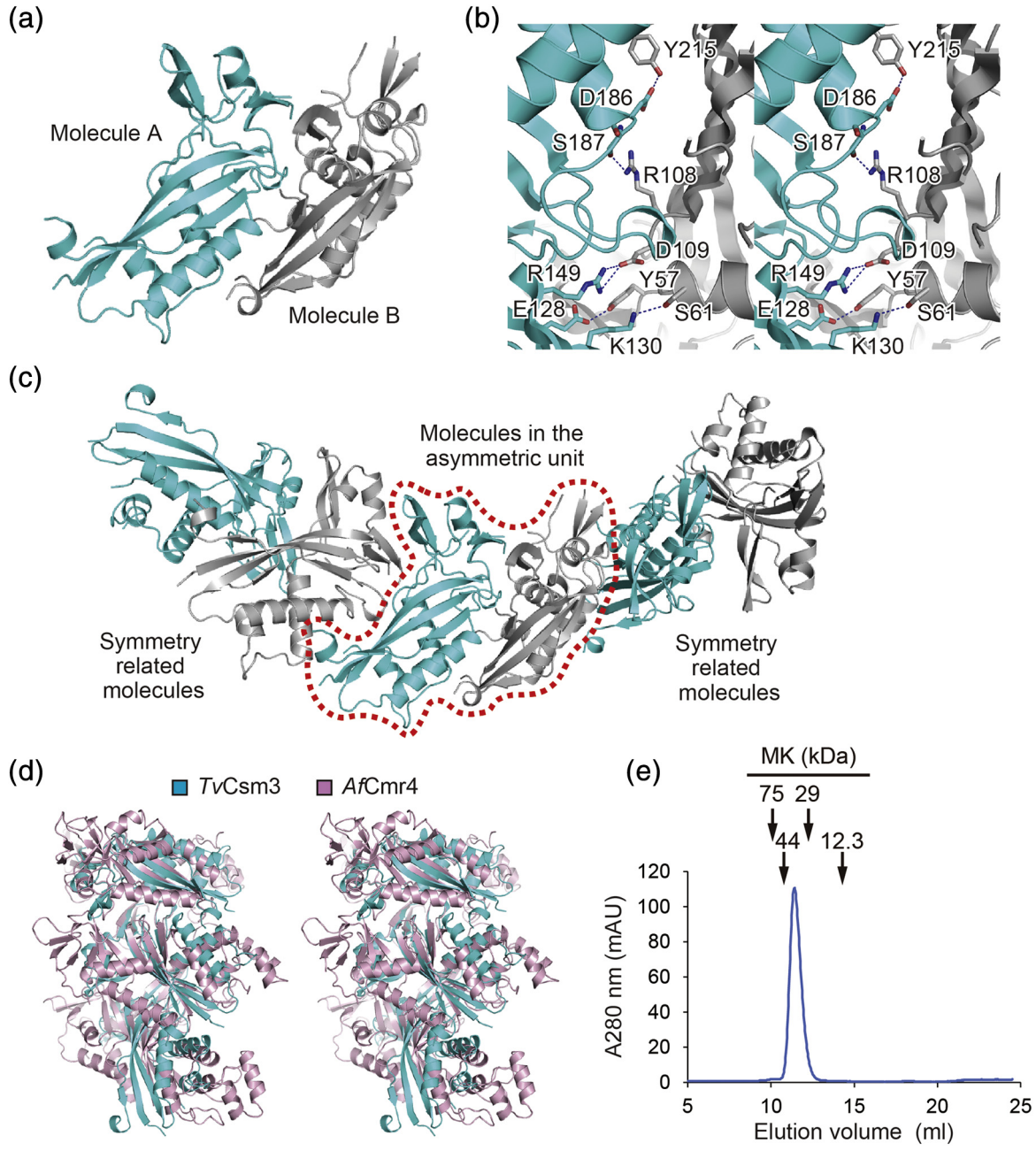

Fig. 3. Structure of the *T. volcanium* Csm3. (a) Crystal structure of the *Tv*Csm3 proteins in the asymmetric unit. Molecules A and B are colored cyan and gray, respectively. (b) Stereoview of the interactions between two *Tv*Csm3 molecules in the asymmetric unit, with the same color codes as in panel a. Hydrogen bonds are depicted by dashed lines. (c) Molecular arrangement of *Tv*Csm3 in the crystal lattice. Two molecules in the crystallographic asymmetric unit are surrounded by a dashed line. The two symmetry-related molecules on each side are shown with the same color codes. (d) Stereoview of the superimposition of the stacked *Tv*Csm3 molecules in the crystal lattice onto the stacked Cmr4 molecules in the Cmr complex [30] (PDB ID: 3X1L). *Tv*Csm3 and Cmr4 are colored cyan and pink, respectively. (e) Size exclusion chromatogram of the purified *Tv*Csm3, analyzed with a Superdex 75 HR 10/300 column. The elution volumes for the molecular weight markers, conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), and cytochrome *C* (12.3 kDa), are indicated above the chromatogram. The SEC analysis revealed the calculated molecular weight of *Tv*Csm3 as 39.9 kDa, which is slightly larger than the theoretical molecular weight value of this protein (25.5 kDa). The difference may be due to the asymmetric molecular shape of this protein.

also well conserved among the Csm3 family members. The buried surface area between the symmetry-related molecules is comparable (1,021 Å^2) to that between the two *Tv*Csm3 molecules in the asymmetric unit. Thus, the *Tv*Csm3 molecules are stacked in a regular manner, giving rise to the multimerization of *Tv*Csm3 in the crystals (Fig. 3c).

The multimerized *Tv*Csm3 molecules exhibit a helical arrangement, which is similar to the stacked structure of the three Cmr4 molecules in the Cmr complex [30] (Fig. 3d). In contrast, the SEC analysis revealed that *Tv*Csm3 exists as a monomer in solution (Fig. 3e). The inconsistency between the *Tv*Csm3 states in the crystal and in solution might be due in part to the differences in the experimental conditions. It is also likely that the crystal packing interaction forces *Tv*Csm3 to adopt this stable molecular arrangement, which may also be found in nature. Intriguingly, a cryoEM study demonstrated that the recombinant Cmr4, together with Cmr5, formed a long filamentous structure with an architecture resembling that of the helical backbone of the Cmr complex [27]. A similar helical filament was also reported for the *Sulfolobus solfataricus* Cas7 protein from the type I-A system, despite the fact that the recombinant Cas7 protein expressed in *Escherichia coli* was monomeric in solution [48]. These findings suggest that the Cas7 family proteins have an intrinsic property to polymerize by themselves. The multimerized *Tv*Csm3 structure in the crystals seems to represent a functionally relevant state of this protein, because the residues involved in the molecular contacts are well conserved in the Csm3 family (Figs. 2b and 3b) and the molecular arrangement is similar to that of Cmr4 in the Cmr complex (Fig. 3d). While the Csm3 and Cmr4 families exhibit similar molecular arrangements in their polymerized structures (Fig. 3d), the residues responsible for their multimerizations are not conserved in these proteins, showing that they diverged from a common ancestor during evolution.

A structural comparison of three Csm3 molecules from different species revealed that *Mk*Csm3 has an additional domain adjacent to the RNA recognition motif (RRM) fold [42], which is absent in both the *Tv*Csm3 and *Mj*Csm3 structures [43] (Supplementary Fig. 2a–c). This additional domain in *Mk*Csm3 does not clash with the interface responsible for the multimerization of the Csm3 protein, as clarified by the present *Tv*Csm3 crystal structure (Supplementary Fig. 2d–f). These findings suggest some conservation in the mode of self-association among the Csm3 family members and support the idea that the multimerized *Tv*Csm3 structure in the crystals represents a functionally relevant state of this protein. The structural comparison between *Tv*Csm3 and the *M. jannaschii* Csm3–Csm4 complex [43] revealed that the concave surface involved in the multimerization of *Tv*Csm3 is equivalent to the interface between *Mj*Csm3 and *Mj*Csm4 (Supplementary Fig. 3). Therefore, Csm3 uses the same region for both Csm4 binding and its multimerization. This resembles the mechanism for Cmr complex formation, where Cmr4 and Cmr3 (orthologs of Csm3 and Csm4, respectively) also adopt a similar strategy [30].

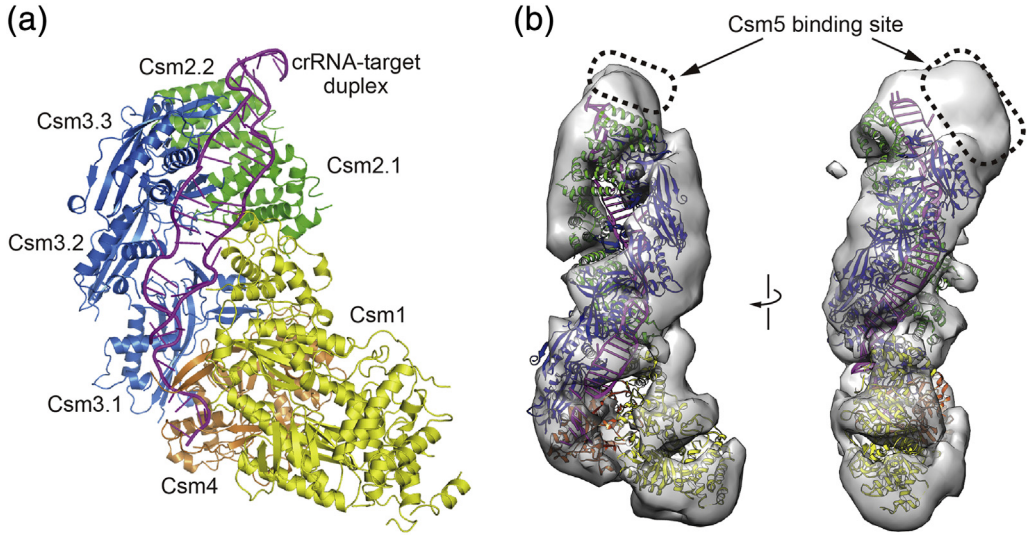

Fig. 4. Model of the Csm complex. (a) Model of the Csm complex with the protein stoichiometry of $Csm1_1 2_2 3_3 4_1 5_1$. Csm1, Csm2, Csm3, Csm4, and the crRNA-target duplex are colored yellow, green, blue, orange, and magenta, respectively. The model was constructed by superimposing each Csm protein onto the Cmr complex [30] (PDB ID: 3X1L). (b) Fitting of the Csm complex model (the protein stoichiometry of $Csm1_1 2_5 3_6 4_1 5_1$) into the EM reconstruction (EMD-2420) of the *T. thermophilus* Csm complex [29]. The $Csm1_1 2_5 3_6 4_1 5_1$ model was constructed based on the $Csm1_1 2_2 3_3 4_1 5_1$ model and manually docked into the EM reconstruction by using the UCSF Chimera program [50]. The coloring scheme is the same as in panel a.

Model of the Csm complex

The Csm complex comprises five Csm proteins (Csm1–5) and the crRNA [49]. Although the crystal structure of Csm5 is not presently available, it is possible to construct a partial model of the Csm complex by superimposing the Csm protein structures onto the crystal structure of the Cmr complex [30]. The multimerized *Tv*Csm3 structure can be superimposed well on the three stacked Cmr4 molecules in the Cmr complex [30]. Two *Se*Csm2 molecules are snugly accommodated in the positions that are occupied by two Cmr5 molecules in the Cmr complex [30]. Furthermore, we used the *M. jannaschii* Csm4 [43] and *Thermococcus onnurineus* Csm1 [44] structures, giving rise to the partial model of the Csm complex with the protein stoichiometry of one Csm1 molecule, two Csm2 molecules, three Csm3 molecules, one Csm4 molecule, and one Csm5 molecule ($Csm1_1 2_2 3_3 4_1 5_1$) (Fig. 4a). Like the Cmr complex [30], by regularly adding more of each subunit of Csm2 and Csm3 to the above-mentioned Csm complex model, a larger Csm complex model with the protein stoichiometry of $Csm1_1 2_5 3_6 4_1 5_1$ can be constructed, which fit well to the cryoEM map of the *T. thermophilus* Csm complex [29] (Fig. 4b).

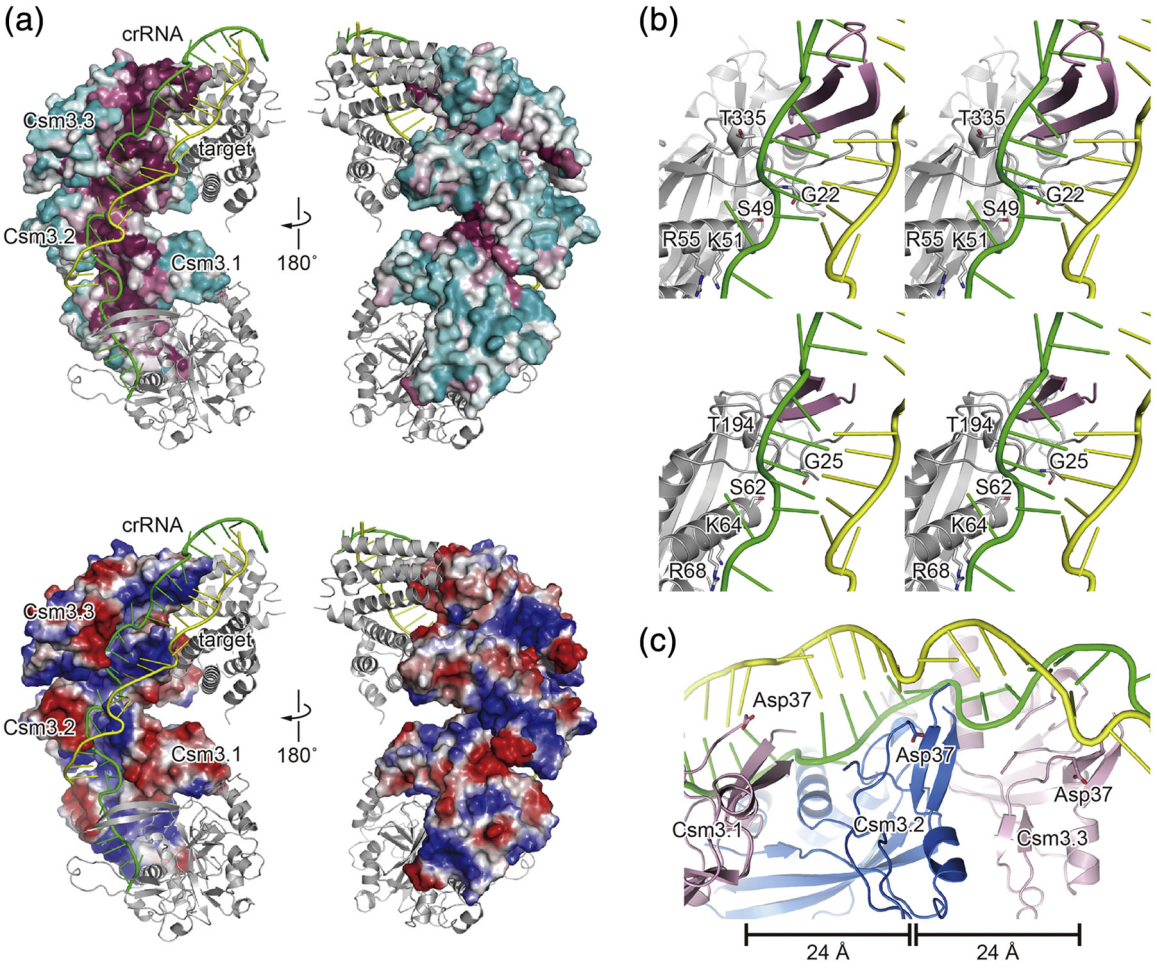

Fig. 5. Model of the Csm3 molecules bound to the crRNA-target duplex. (a) Interactions between the modeled Csm3 molecules and the crRNA-target duplex. The modeled atomic coordinates are the same as in Fig. 4a. Three Csm3 molecules (Csm3.1–3.3) are represented by surface models. The residues are color-coded (magenta: conserved → cyan: variable) according to the sequence conservation among the Csm3 family proteins, calculated by the ConSurf server [53] (top). The electrostatic potential (blue and red show electropositive and electronegative, respectively) of the stacked *Tv*Csm3 molecules (bottom). (Left) Front view of the model. (Right) Back view of the model, rotated by 180° from the left panel. The crRNA and the target strand are green and yellow, respectively. Two Csm2 molecules and Csm4 are shown in light gray ribbon models. Csm1 was omitted for clarity. (b) Stereoviews of Cmr4 (top) and Csm3 (bottom) interacting with the crRNA-target duplex in the Cmr complex crystal structure [30] (PDB ID: 3X1L) and in the Csm complex model, respectively. The thumbs of each complex are colored pink. The Cmr4 residues responsible for crRNA binding and their corresponding residues in Csm3 are depicted by stick models. (c) Catalytic aspartates (Asp37 residues) in the Csm complex model appear periodically at a distance of 24 Å.

The model of the Csm complex exhibits a filamentous architecture similar to that of the Cmr complex. The helical filament comprises two chains, the major and minor chains formed by Csm3 and Csm2, respectively. Like the Cmr complex [30], these two chains contact Csm4 (ortholog of Cmr3) and the D4 domain of Csm1 (ortholog of Cmr2), respectively (Fig. 4a). As proposed by Venclovas [46], the Csm2 structure resembles the D4 domain of Csm1, and these two proteins are arranged in a similar manner to those orthologs in the Cmr complex [30]. The molecular arrangement of Csm3 and Csm4 is analogous to that of Cmr4 and Cmr3 [30], as stated above. Therefore, the Csm complex forms the helical filament in the same manner as that of the Cmr complex.

During peer review of this manuscript, the high-resolution cryoEM structures of the Csm complexes from *Streptococcus thermophilus* and *T. onnurineus* were reported [51,52]. The current Csm complex model resembles the cryoEM structures, despite the differences of the protein stoichiometry of the complex. This finding shows that the structure and function of the Csm complex are well conserved among different species.

Implications for crRNA-target binding and target degradation

Based on the model of the Csm complex, the conserved residues among the Csm3 family members are localized on the possible crRNA binding face (Fig. 5a). Positively charged residues are clustered on this surface (Fig. 5a). The shape of the crRNA bound to the Cmr complex [30] fit well to the molecular surface of the multimerized *Tv*Csm3 structure (Fig. 5a), suggesting that these two type III effector complexes bind the crRNA-target duplex in a similar manner. In contrast, the opposite surface of the *Tv*Csm3 molecules is not conserved (Fig. 5a), consistent with the idea that this region faces the solvent. In the Cmr complex, the long loop of Cmr4, the so-called "thumb," plays an important role in deforming the crRNA-target duplex for periodically directing the scissile bonds toward the active sites [30]. The corresponding region of *Tv*Csm3 (residues 127–150) is partially folded to form the thumb-like structure (residues 133–143 are disordered) that intercalates the duplex in the model (Fig. 5b). In contrast, the corresponding regions in *Mk*Csm3 and *Mj*Csm3 were completely disordered in the crystals [42,43]. These findings suggest that the thumb is structured through subunit multimerization and is further stabilized by the interactions with the crRNA-target duplex.

A structural comparison between *Tv*Csm3 and *A. fulgidus* Cmr4 (*Af*Cmr4) [30] revealed that several residues (Gly22, Ser49, Lys51, Arg55, and Thr335) in *Af*Cmr4 responsible for the crRNA binding in the Cmr complex appear in the same places in the *Tv*Csm3 structure (Gly25, Ser62, Lys64, Arg68, and Thr194, respectively) (Fig. 5b). These residues are conserved among the Csm3 family proteins (Fig. 2b). These findings, together with the mutational analysis demonstrating the importance of Lys51 and Arg55 of *Af*Cmr4 in the target cleavage reaction by the Cmr complex [30], strongly suggest that these conserved residues among the Csm3 proteins bind the crRNA in the same manner as those of Cmr4 reported in the Cmr complex structure. These Csm3 residues are indeed located in close proximity to the phosphate groups of the crRNA in the recently published cryoEM structures [51,52].

A conserved Asp residue in Csm3 (Asp37 in *Tv*Csm3) serves as the catalyst in the target RNA cleavage reaction [17,41]. This aspartate is also conserved in Cmr4 and may serve as the acid catalyst in the Cmr complex [30,39,40]. In the Csm complex model, the Asp37 residues of *Tv*Csm3 are separated from each other by approximately 24 Å (Fig. 5c). A

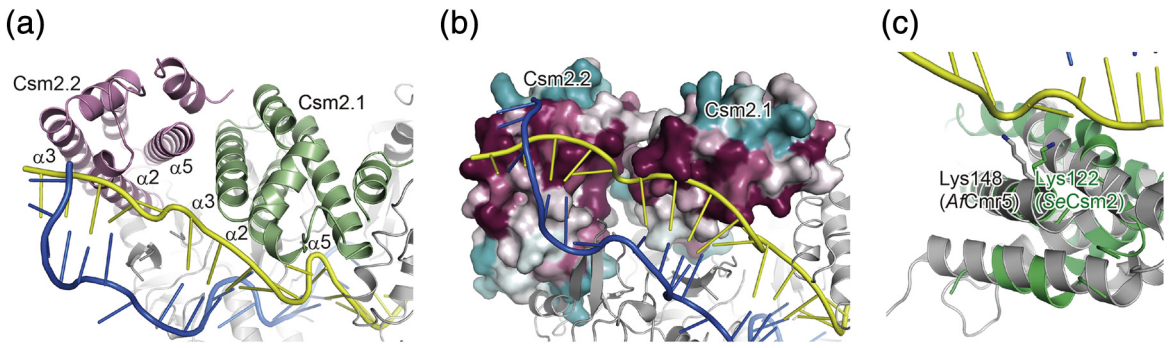

Fig. 6. The Csm2 molecules in the Csm complex model. (a) Model of the interactions between the Csm2 proteins and the crRNA-target duplex. The modeled atomic coordinates are the same as in Fig. 4a. The two Csm2 molecules (Csm2.1 and Csm2.2), the crRNA, and the target strand are colored light green, pink, blue, and yellow, respectively. Helices (α2, α3, and α5), which contact the target strand, are labeled in each molecule. (b) The conserved surface of Csm2 forms the target contact site. Two Csm2 molecules are represented by surface models and are color-coded according to the sequence conservation, as in Fig. 5a. (c) Superimposition of *Se*Csm2 onto *Af*Cmr5 in the Cmr complex [30] (PDB ID: 3X1L) shows that these two proteins have a lysine residue at the same position, close to the target strand.

comparable distance (23–24 Å) is observed in the Cmr complex, where the Asp31 residues in *Af*Cmr4 are close to the scissile bonds of the target strand [30]. Furthermore, the Asp37 residues in *Tv*Csm3 exist in quite similar places to those of the corresponding aspartates in *Af*Cmr4. These findings definitely support the biochemical experiments suggesting that the Csm complex periodically cleaves the target RNA at multiple sites with 6-nt intervals [29,41], in the same manner as the Cmr complex.

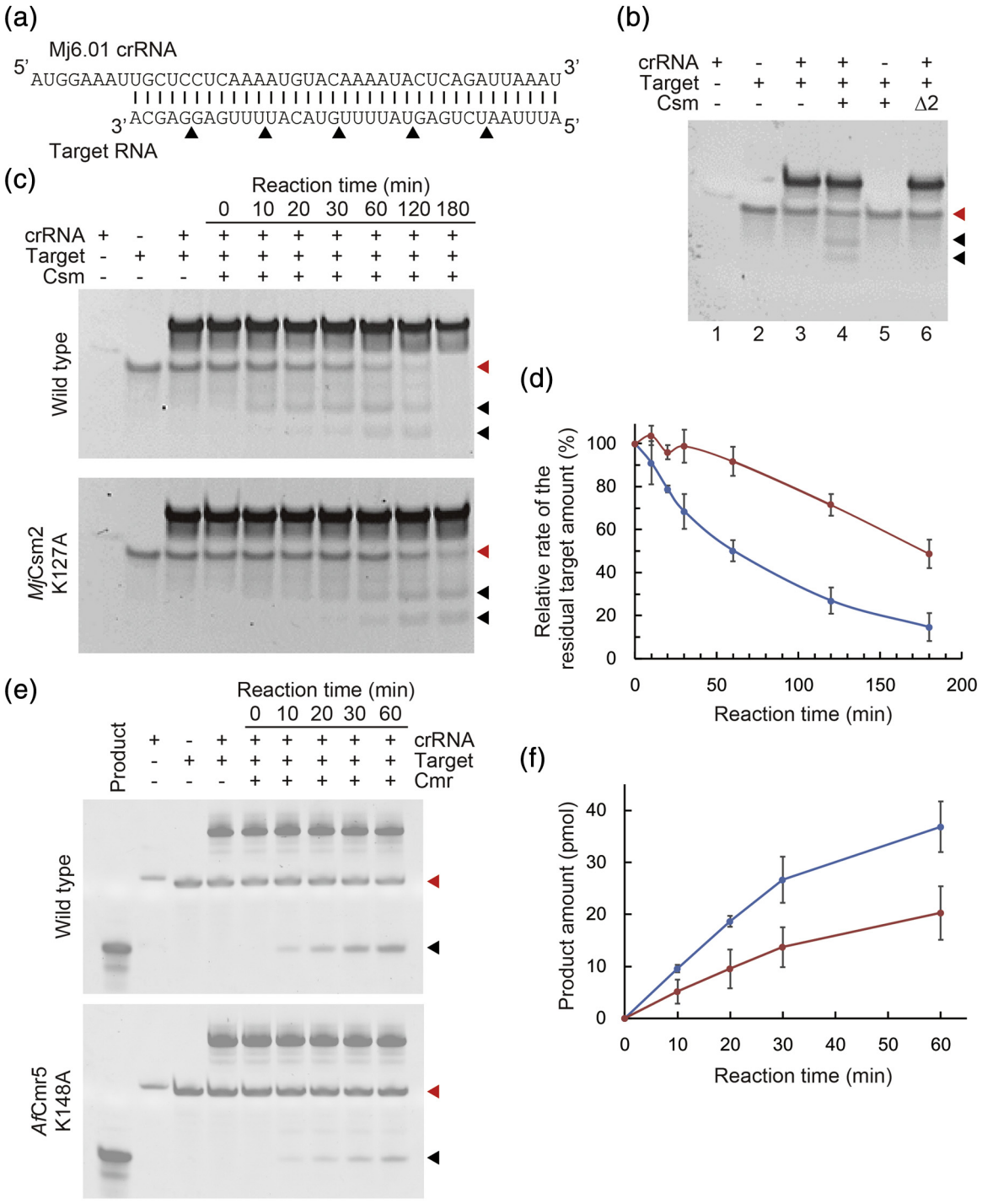

Fig. 7 (*legend on next page*)

Possible role of the small subunit of the type III effector complex

In the Csm complex model, the Csm2 molecules fit snugly with the target nucleic acid (Fig. 6a). Three helices (α2, α3, and α5) of *Se*Csm2 may form a target binding groove in the Csm complex (Fig. 6a). Intriguingly, the conserved residues in the Csm2 family are clustered on this proposed target binding surface (Fig. 6b). In particular, several basic and hydrophilic residues (Arg36, Tyr74, Arg78, and Lys122), which are conserved among the Csm2 family members (Fig. 2a), are located near the target strand. These residues are in close contact with the target RNA in the recent cryoEM structures of the Csm complex [51,52]. The structural alignment between *Se*Csm2 and *Af*Cmr5 [30] revealed that *Af*Cmr5 has the corresponding lysine residue (Lys148) at the same position as Lys122 of *Se*Csm2 (Fig. 6c). In the Cmr complex, Lys148 of *Af*Cmr5 is located near the scissile bonds of the target strand [30] (Fig. 6c). This lysine residue is well conserved in the Csm2 family, but not in the Cmr5 family.

To test the importance of this residue in the target RNA cleavage activity, we prepared the wild-type and mutant Csm complexes for the reaction. At first, we reconstituted the *Se*Csm complex, but the target RNA cleavage activity by the wild-type *Se*Csm complex was very low in our assay conditions. We then prepared the *M. jannaschii* Csm (*Mj*Csm) complex, which exhibited robust target RNA degrading activity (Fig. 7a and b). We could not observe the particle formation of the Csm complex by the SEC analysis, possibly due in part to the low affinities of the Csm proteins, which did not form a stable complex in our experimental conditions. However, when *Mj*Csm2 was removed from the reconstitution, the RNA cleavage activity disappeared (lane 6 of Fig. 7b). This result strongly suggests that the Csm complex was properly reconstituted in the presence of all of the protein components and the crRNA. We next sought to explore the role of the lysine residue in the target RNA cleavage reaction. Lys127 in *Mj*Csm2, corresponding to the Lys122 residue of *Se*Csm2, was mutated to alanine, and the mutant Csm2-containing *Mj*Csm complex was reconstituted. The *Mj*Csm2 K127A mutation-containing *Mj*Csm complex exhibited reduced target RNA cleavage activity as compared to the wild-type *Mj*Csm complex (Fig. 7c and d), showing that this conserved lysine residue plays an important role, possibly by binding the target strand during the reaction. This finding also supports the accuracy of the present Csm complex model. We also prepared the K148A mutant of *Af*Cmr5 and reconstituted the mutant Cmr complex. The mutant Cmr complex exhibited decreased RNA degrading activity, as compared to the wild-type complex (Fig. 7e and f). These mutagenesis experiments also suggested that the lysine residues in Csm2 and Cmr5 are not directly involved in catalysis during the RNA hydrolysis reaction. In the Cmr complex structure, Cmr5 and Cmr4 associate with the crRNA-target duplex [30]. Cmr5 is located near the target strand but only interacts with the target sparsely, in striking contrast to the extensive interactions between Cmr4 and the crRNA-target duplex [30]. The Csm complex model suggested a similar environment around the crRNA-target duplex. This is also supported by the recently published cryoEM structures of the Csm complex [51,52]. Furthermore, neither Csm2 nor Cmr5 has other polar or hydrophilic conserved residues in the vicinity of the scissile bonds. Together, these findings suggest that the small subunit proteins (Csm2 and Cmr5) of the type III effector complexes do not participate in the catalysis during the RNA cleavage reaction. Instead, these proteins may be predominantly involved in the formation of the helical filament of the effector complex, to stabilize the active structure of the complex that facilitates the RNA cleavage reaction.

The helical filament in the Csm complex is reportedly capped by Csm5 [29,51,52], by a capping mechanism analogous to that by Cmr6 in the Cmr complex [30]. In contrast to the Csm complex that comprises five Csm proteins, the Cmr complex has an extra protein, Cmr1, in addition to their five counterparts [11]. A recent study of the *Sulfolobus islandicus* Cmr complex suggested that Cmr1 enhances the RNA and DNA interference activities of the Cmr complex, by boosting the binding ability to both the crRNA and target RNA [54]. Csm5

Fig. 7. Target RNA cleavage activity. (a) Nucleotide sequences of *Mj*6.01 crRNA and the target RNA used in the RNA cleavage experiment of the reconstituted *Mj*Csm complex. The complementary region between both strands is depicted by lines. The possible cleavage sites on the target RNA strand are indicated by triangles. (b) RNA cleavage activity of the reconstituted *Mj*Csm complex. The RNA cleavage reaction was performed for 1 h at 70 °C. Δ2 in lane 6 denotes the reaction by the particle reconstituted under the *Mj*Csm2-deficient conditions. The target RNA and cleavage products are indicated by red and black triangles, respectively. (c) Time courses of the RNA cleavage experiments with the wild-type (top) and *Mj*Csm2 K127A mutant-containing (bottom) *Mj*Csm complexes. (d) RNA degrading activities of the wild-type (blue) and mutant (red) *Mj*Csm complexes. The relative amounts of the residual target RNA were calculated from panel c, where the band intensities of the target (indicated by red triangles) were normalized to that of the reaction time of 0 min. The error bars indicate the standard deviation determined from three independent measurements. (e) Time courses of the RNA cleavage experiment by the wild-type (top) and Cmr5 K148A mutant-containing (bottom) Cmr complexes. The nucleotide sequences of the *Pf*7.01 crRNA and its target used in this assay are in Ref. [30]. The synthetic RNA corresponding to the reaction product (20 mer) (indicated by the black triangles) was loaded in the first lanes to calculate the product amount of each reaction. (f) RNA degrading activities of the wild-type (blue) and mutant (red) Cmr complexes. The product amounts of each reaction were calculated from panel e. The error bars indicate the standard deviation determined from two independent experiments.

might perform multiple roles in the CRISPR interference process that are accomplished by two protein components (Cmr1 and Cmr6) in the Cmr complex. Alternatively, the detailed action mechanisms for target recognition and cleavage might have diverged between these two type III effector complexes. Biochemical study of Csm5 is needed to understand the role of Csm5 and to clarify the differences in the interference mechanisms between these type III effector complexes. In the present study, we solved the structures of the functional states of the Csm2 and Csm3 proteins, and successfully constructed a model of the Csm complex, which will contribute to the elucidation of the RNA-guided interference mechanism of the type III effector complexes.

Materials and Methods

Protein expression and purification

The synthetic genes encoding *S. epidermidis* Csm2 and *T. volcanium* Csm3 were purchased from Takara Bio, and were cloned into the *Nde*I and *Bam*HI sites of the pET28b vector (Novagen). The recombinant proteins were produced in *E. coli* strain BL21-CodonPlus (DE3)-RIL (Agilent). The harvested *E. coli* cells were lysed by sonication in lysis buffer, containing 20 mM Tris–HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole, 1 mM PMSF, 1 mM benzamidine, and 7 mM 2-mercaptoethanol. After sonication, the cell debris was removed by centrifugation at 8000 rpm for 15 min. For *Se*Csm2, the solution was further centrifuged at 11,000 rpm for 15 min, and the subsequent supernatant was used for purification. For *Tv*Csm3, the solution was subjected to heat treatment at 65 °C for 20 min, and the heat-denatured proteins were removed by centrifugation at 8000 rpm for 15 min. Both recombinant proteins have the His-tag sequence derived from the expression plasmid at the N-terminus, and therefore, the clear lysate was applied to a Ni-NTA column (Qiagen), equilibrated with 20 mM Tris–HCl (pH 8.0), 300 mM NaCl, 10 mM imidazole, and 7 mM 2-mercaptoethanol. Both proteins were eluted with the same buffer containing 250 mM imidazole. The eluate of *Se*Csm2 was dialyzed against 20 mM Tris–HCl (pH 7.0), 400 mM NaCl, and 7 mM 2-mercaptoethanol. The supernatant of the dialysate was loaded onto a Resource Q column (GE Healthcare), previously equilibrated with the same buffer, to separate *Se*Csm2 from the nucleic acids. The flow-through fraction containing *Se*Csm2 was collected and dialyzed against 20 mM Tris–HCl (pH 7.0), 200 mM NaCl, and 7 mM 2-mercaptoethanol. The dialysate was loaded onto a Resource S column (GE Healthcare), previously equilibrated with the same buffer. *Se*Csm2 was eluted with a linear gradient of 0.2–1.0 M NaCl in the same buffer. The protein-rich fractions were pooled and dialyzed against 20 mM Tris–HCl (pH 8.0), 200 mM NaCl, and 1 mM DTT.

After purification on a Ni-NTA column, the *Tv*Csm3 solution was dialyzed against 20 mM Tris–HCl (pH 7.0), 200 mM NaCl, and 7 mM 2-mercaptoethanol. After the concentration of NaCl in the protein solution was reduced to 100 mM by mixing with an equal volume of 20 mM Tris–HCl (pH 7.0) and 7 mM 2-mercaptoethanol, the protein solution was purified on a HiTrap Heparin column (GE Healthcare), previously equilibrated with 20 mM Tris–HCl (pH 7.0), 100 mM NaCl, and 7 mM 2-mercaptoethanol. A linear gradient from 0.1 to 1.0 M NaCl was developed in the same buffer. The protein-rich fractions were pooled and dialyzed against 20 mM Tris–HCl (pH 8.5), 100 mM NaCl, and 7 mM 2-mercaptoethanol. *Tv*Csm3 was then purified on a Resource Q column, previously equilibrated with the same buffer. The protein was eluted with a linear gradient of 0.1–1.0 M NaCl in the same buffer. The fractions rich in *Tv*Csm3 were pooled and dialyzed against 10 mM Tris–HCl (pH 8.0), 200 mM NaCl, and 1 mM DTT.

The purified *Se*Csm2 and *Tv*Csm3 proteins were concentrated to 0.485 and 13.7 mg/ml and were flash-frozen in liquid nitrogen before storage at −80 °C. The protein concentrations of *Se*Csm2 and *Tv*Csm3 were determined using the molar extinction coefficients of 13,410 and 21,890 M^{-1} cm^{-1} at 280 nm, respectively. The selenomethionine (SeMet)-labeled *Se*Csm2 was produced in the methionine auxotroph *E. coli* strain B834-CodonPlus (DE3)-RIL and was purified in the same manner as the native protein.

Size exclusion chromatography

The molecular weights of the purified proteins were analyzed by SEC on a Superdex 75 HR 10/300 column, which was equilibrated with 10 mM Tris–HCl (pH 8.0), 200 mM NaCl, and 1 mM DTT. Conalbumin (75 kDa), ovalbumin (44 kDa), carbonic anhydrase (29 kDa), cytochrome *C* (12.3 kDa), and aprotinin (6.5 kDa) were used as molecular weight marker proteins to calculate the standard curve for the column.

Crystallization and data collection

*Se*Csm2 was crystallized by the sitting-drop vapor diffusion method at 20 °C, under conditions containing 180 mM Li_2SO_4, 90 mM Hepes-Na (pH 7.5), 22.5% (w/v) PEG3350, and 10 mM sarcosine. Sitting drops were prepared by mixing 1 µl of the reservoir solution with 1 µl of the protein solution (0.485 mg/ml), and were equilibrated against 100 µl of reservoir solution. The *Se*Csm2 crystals grew within 3 days. The SeMet-labeled *Se*Csm2 crystals were also obtained under the same conditions as those used for the native crystals. The *Tv*Csm3 crystals were obtained by the

sitting-drop vapor diffusion method at 20 °C, with a reservoir composed of 180 mM di-ammonium tartrate, 18% (w/v) PEG3350, and 3% (w/v) PEG1500. The *Tv*Csm3 crystals grew within 10 days.

For data collection, the *Se*Csm2 and *Tv*Csm3 crystals were transferred to cryoprotectant solutions supplemented with 20% (w/v) PEG400 and 20% (w/v) ethylene glycol, respectively, in each reservoir solution. The crystals were mounted in a nylon loop, and then flash-cooled in a nitrogen stream at 95 K. X-ray diffraction data were collected at the beam-line BL-17A of the Photon Factory (Ibaraki, Japan), using Pilatus3 S6M and ADSC Q315r detectors, respectively. Diffraction data were integrated and scaled with the program HKL2000 [55]. The processing statistics are summarized in Table 1.

Structure determination and refinement

We determined the *Se*Csm2 structure by the SAD method, using the data from the SeMet-labeled crystals, with the AutoSol program in the Phenix suite [56]. The SAD experimental electron density map at 2.4-Å resolution had sufficient quality for tracing two *Se*Csm2 molecules. The initial model was manually built with the program Coot [57] using the SeMet locations as guides. The model was then improved by iterative cycles of refinement with the program PHENIX and manual rebuilding with Coot. The *Se*Csm2 structure was finally refined to an R_{work} of 21.2% and an R_{free} of 23.3% at 2.4-Å resolution. The final model contains two *Se*Csm2 molecules (residues 1–13 and 25–125 for both molecules). A Ramachandran analysis by PROCHECK [58] showed 94.9%, 5.1%, 0%, and 0% of the protein residues in the most favored, additionally allowed, generously allowed, and disallowed regions, respectively.

The structure of *Tv*Csm3 was solved by the molecular replacement method with the program Molrep [59] using the *M. jannaschii* Csm3 structure [43] (PDB ID: 4QTS) as the search model. The *Tv*Csm3 structure was refined to an R_{work} of 21.1% and an R_{free} of 24.3% at 2.7-Å resolution, with the program PHENIX. The final model contains two *Tv*Csm3 molecules (residues 5–26, 37–91, 102–132, and 144–216 for molecule A and residues 5–26, 37–91, 102–132, and 145–216 for molecule B). The stereochemistry of the models was verified by PROCHECK (90.9%, 8.8%, 0.3%, and 0% in the most favored, additionally allowed, generously allowed, and disallowed regions, respectively). Refinement statistics are summarized in Table 1. Molecular graphics were illustrated with PyMol (http://www.pymol.org/).

Modeling of the Csm complex

To build the Csm complex model, we first superimposed one copy of *To*Csm1 [44], two copies of *Se*Csm2, three copies of *Tv*Csm3, and one copy of *Mj*Csm4 [43] onto the individual Cmr2, Cmr5, Cmr4, and Cmr3 proteins in the Cmr complex [30], to generate the Csm complex model with the protein stoichiometry of $Csm1_1 2_2 3_3 4_1 5_1$. To distinguish each protomer of the Csm2 and Csm3 proteins, the Csm2 molecule adjacent to Csm1 in the model is named Csm2.1, and the other Csm2 molecule is Csm2.2 (Fig. 4a). The Csm3 molecule contacting Csm4 is named Csm3.1, and its adjacent molecule and the other subunit are Csm3.2 and Csm3.3, respectively (Fig. 4a). Although the Csm complex model was not subjected to energy minimization, the subunits fit the model without any major steric clashes, and only minor clashes of several side chains occurred between Csm1 and Csm4, Csm3 and Csm4, and Csm2 and Csm3. These minor clashes would be attributable to the use of Csm proteins from different species for constructing the model. The crRNA-target duplex in the Cmr complex [30] fit well to the $Csm1_1 2_2 3_3 4_1 5_1$ model without any adjustment.

Although the cryoEM reconstruction and mass spectrometric experiment suggested that the stoichiometry of the *T. thermophilus* Csm complex was $Csm1_1 2_3 3_6 4_2 5_1$ [29], we estimated it as $Csm1_1 2_5 3_6 4_1 5_1$, based on the structural analogies between the Csm and Cmr proteins and the assembly pattern of the Cmr proteins in the Cmr complex. To build the $Csm1_1 2_5 3_6 4_1 5_1$ model, the $Csm2_2 3_3$ subcomplex was isolated from the $Csm1_1 2_2 3_3 4_1 5_1$ model. Csm2.1 and Csm3.2 of the $Csm2_2 3_3$ subcomplex were superimposed onto Csm2.2 and Csm3.3 in the $Csm1_1 2_2 3_3 4_1 5_1$ model, respectively, to generate new Csm2.3 and Csm3.4 molecules, resulting in the $Csm1_1 2_3 3_4 4_1 5_1$ model. Similarly, the $Csm2_3 3_4$ subcomplex was isolated from the $Csm1_1 2_3 3_4 4_1 5_1$ model. Csm2.1 and Csm3.2 of the $Csm2_3 3_4$ subcomplex were then superimposed onto Csm2.3 and Csm3.4 in the $Csm1_1 2_3 3_4 4_1 5_1$ model, respectively, which created new Csm2.4, Csm2.5, Csm3.5, and Csm3.6 molecules to give rise to the $Csm1_1 2_5 3_6 4_1 5_1$ model. The duplex between the crRNA guide and target was extended to fit the $Csm1_1 2_5 3_6 4_1 5_1$ model in a regular manner. The $Csm1_1 2_5 3_6 4_1 5_1$ model was fit to the cryoEM map of the *T. thermophilus* Csm complex [29] by using the UCSF Chimera program [50].

Target RNA cleavage activity

M. jannaschii Csm proteins were prepared by the method previously reported [43]. The single-stranded RNA (43-mer) from the *M. jannaschii* CRISPR locus (*Mj*6.01-crRNA: AUGGAAAUUGCUCCUCAAAAU GUACAAAAUACUCAGAUUAAAU) and its target sequence (AUUUAAUCUGAGUAUUUUGUA CAUUUUGAGGAGCA) that is complementary to the guide of *Mj*6.01-crRNA were purchased from Sigma. The purified *Mj*Csm1–5 proteins and *Mj*6.01-crRNA were mixed in a molar ratio of 1:2:3:1:1:1, respectively, and incubated at 70 °C for 30 min to form the *Mj*Csm

complex. Although we did not observe the formation of the stable complex by the SEC analysis, we estimated the composition of the reconstituted *Mj*Csm complex to be 1:2:3:1:1:1 for *Mj*Csm1, *Mj*Csm2, *Mj*Csm3, *Mj*Csm4, *Mj*Csm5, and *Mj*6.01-crRNA, respectively. The RNA-cleavage reaction was performed at 70 °C in a 50-μl reaction mixture containing 400 nM *Mj*Csm complex and 3.4 μM target RNA in the reaction buffer [20 mM Hepes-Na (pH 7.0), 250 mM KCl, 2 mM $MgCl_2$, 10 mM DTT, and 1 U ribonuclease inhibitor]. The RNA cleavage reaction was terminated by phenol–chloroform–isoamyl alcohol extraction, and the RNAs were precipitated with ethanol and then dissolved in 10 μl of loading dye (3.5 M urea, 0.05% bromophenol blue, and 0.05% xylene cyanol). The mixture was incubated at 95 °C for 10 min and then on ice for 2 min, and was fractionated on a 15%–20% denaturing polyacrylamide gel containing 7 M urea. After electrophoresis, the gels were stained with SYBR Green II. The target RNA cleavage activity of the wild-type and mutant Cmr5-containing Cmr complex was analyzed as previously described [30].

Accession codes

The atomic coordinates and the structure factors of *Se*Csm2 and *Tv*Csm3 have been deposited in the Protein Data Bank under the accession codes 6AE1 and 6AE2, respectively.

Acknowledgments

We thank the beam-line staff at BL-17A of the Photon Factory (Ibaraki, Japan) for technical assistance during data collection. This work was supported by a Grant-in-Aid for Scientific Research (B) from the Japan Society for the Promotion of Science and grants from the Kurata Memorial Hitachi Science and Technology Foundation and the Mochida Memorial Foundation for Medical and Pharmaceutical Research to T.N.

Author Contributions: T.N. designed the research; D.T., M.S., H.I., and T.N. performed the research; D.T. and T.N. analyzed the data; and T.N. wrote the manuscript.

Conflict of Interest: The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2019.01.009.

Received 4 August 2018;
Received in revised form 22 December 2018;
Accepted 5 January 2019
Available online 11 January 2019

Keywords:
CRISPR;
Cas;
crRNA;
Csm complex;
RNA-based defense system

Abbreviations used:
CRISPR, clustered, regularly interspaced, short palindromic repeat; Cas, CRISPR associated; crRNA, CRISPR RNA; Cascade, CRISPR-associated complex for antiviral defense; nt, nucleotide; cryoEM, cryoelectron microscopy; SEC, size exclusion chromatography; SAD, single-wavelength anomalous diffraction.

References

[1] R. Sorek, C.M. Lawrence, B. Wiedenheft, CRISPR-mediated adaptive immune systems in bacteria and archaea, Annu. Rev. Biochem. 82 (2013) 237–266.
[2] R.M. Terns, M.P. Terns, CRISPR-based technologies: prokaryotic defense weapons repurposed, Trends Genet. 30 (2014) 111–118.
[3] J. van der Oost, E.R. Westra, R.N. Jackson, B. Wiedenheft, Unravelling the structural and mechanistic basis of CRISPR–Cas systems, Nat. Rev. Microbiol. 12 (2014) 479–492.
[4] B. Wiedenheft, S.H. Sternberg, J.A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea, Nature 482 (2012) 331–338.
[5] 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.
[6] F.J. Mojica, C. Díez-Villaseñor, J. García-Martínez, E. Soria, Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements, J. Mol. Evol. 60 (2005) 174–182.
[7] C. Pourcel, G. Salvignol, G. Vergnaud, CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies, Microbiology 151 (2005) 653–663.
[8] 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.
[9] J. Carte, R. Wang, H. Li, R.M. Terns, M.P. Terns, Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes, Genes Dev. 22 (2008) 3489–3496.
[10] C. Hale, K. Kleppe, R.M. Terns, M.P. Terns, Prokaryotic silencing (psi)RNAs in *Pyrococcus furiosus*, RNA 14 (2008) 2572–2579.
[11] C.R. Hale, P. Zhao, S. Olson, M.O. Duff, B.R. Graveley, L. Wells, R.M. Terns, M.P. Terns, RNA-guided RNA cleavage by a CRISPR RNA–Cas protein complex, Cell 139 (2009) 945–956.
[12] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, E. Charpentier, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity, Science 337 (2012) 816–821.

[13] K.S. Makarova, Y.I. Wolf, O.S. Alkhnbashi, F. Costa, S.A. Shah, S.J. Saunders, R. Barrangou, S.J. Brouns, E. Charpentier, D.H. Haft, P. Horvath, S. Moineau, F.J. Mojica, R.M. Terns, M.P. Terns, M.F. White, A.F. Yakunin, R.A. Garrett, J. van der Oost, R. Backofen, E.V. Koonin, An updated evolutionary classification of CRISPR–Cas systems, Nat. Rev. Microbiol. 13 (2015) 722–736.

[14] E.R. Westra, P.B. van Erp, T. Künne, S.P. Wong, R.H. Staals, C.L. Seegers, S. Bollen, M.M. Jore, E. Semenova, K. Severinov, W.M. de Vos, R.T. Dame, R. de Vries, S.J. Brouns, J. van der Oost, CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3, Mol. Cell 46 (2012) 595–605.

[15] T. Sinkunas, G. Gasiunas, S.P. Waghmare, M.J. Dickman, R. Barrangou, P. Horvath, V. Siksnys, *In vitro* reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*, EMBO J. 32 (2013) 385–394.

[16] L.A. Marraffini, E.J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA, Science 322 (2008) 1843–1845.

[17] P. Samai, N. Pyenson, W. Jiang, G.W. Goldberg, A. Hatoum-Aslan, L.A. Marraffini, Co-transcriptional DNA and RNA cleavage during type III CRISPR–Cas immunity, Cell 161 (2015) 1164–1174.

[18] M. Kazlauskiene, G. Tamulaitis, G. Kostiuk, Č. Venclovas, V. Siksnys, Spatiotemporal control of type III-A CRISPR–Cas immunity: coupling DNA degradation with the target RNA recognition, Mol. Cell 62 (2016) 295–306.

[19] J.R. Elmore, N.F. Sheppard, N. Ramia, T. Deighan, H. Li, R.M. Terns, M.P. Terns, Bipartite recognition of target RNAs activates DNA cleavage by the type III-B CRISPR–Cas system, Genes Dev. 30 (2016) 447–459.

[20] M.A. Estrella, F.T. Kuo, S. Bailey, RNA-activated DNA cleavage by the type III-B CRISPR–Cas effector complex, Genes Dev. 30 (2016) 460–470.

[21] M.M. Jore, M. Lundgren, E. van Duijn, J.B. Bultema, E.R. Westra, S.P. Waghmare, B. Wiedenheft, Ü. Pul, R. Wurm, R. Wagner, M.R. Beijer, A. Barendregt, K. Zhou, A.P. Snijders, M.J. Dickman, J.A. Doudna, E.J. Boekema, A.J. Heck, J. van der Oost, S.J. Brouns, Structural basis for CRISPR RNA-guided DNA recognition by Cascade, Nat. Struct. Mol. Biol. 18 (2011) 529–536.

[22] B. Wiedenheft, G.C. Lander, K. Zhou, M.M. Jore, S.J. Brouns, J. van der Oost, J.A. Doudna, E. Nogales, Structures of the RNA-guided surveillance complex from a bacterial immune system, Nature 477 (2011) 486–489.

[23] R.N. Jackson, S.M. Golden, P.B. van Erp, J. Carter, E.R. Westra, S.J. Brouns, J. van der Oost, T.C. Terwilliger, R.J. Read, B. Wiedenheft, Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*, Science 345 (2014) 1473–1479.

[24] S. Mulepati, A. Héroux, S. Bailey, Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target, Science 345 (2014) 1479–1484.

[25] H. Zhao, G. Sheng, J. Wang, M. Wang, G. Bunkoczi, W. Gong, Z. Wei, Y. Wang, Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*, Nature 515 (2014) 147–150.

[26] C. Rouillon, M. Zhou, J. Zhang, A. Politis, V. Beilsten-Edmands, G. Cannone, S. Graham, C.V. Robinson, L. Spagnolo, M.F. White, Structure of the CRISPR interference complex CSM reveals key similarities with Cascade, Mol. Cell 52 (2013) 124–134.

[27] M. Spilman, A. Cocozaki, C. Hale, Y. Shao, N. Ramia, R.M. Terns, M.P. Terns, H. Li, S. Stagg, Structure of an RNA silencing complex of the CRISPR–Cas immune system, Mol. Cell 52 (2013) 146–152.

[28] R.H. Staals, Y. Agari, S. Maki-Yonekura, Y. Zhu, D.W. Taylor, E. van Duijn, A. Barendregt, M. Vlot, J.J. Koehorst, K. Sakamoto, A. Masuda, N. Dohmae, P.J. Schaap, J.A. Doudna, A.J.R. Heck, K. Yonekura, J. van der Oost, A. Shinkai, Structure and activity of the RNA-targeting Type III-B CRISPR–Cas complex of *Thermus thermophilus*, Mol. Cell 52 (2013) 135–145.

[29] R.H. Staals, Y. Zhu, D.W. Taylor, J.E. Kornfeld, K. Sharma, A. Barendregt, J.J. Koehorst, M. Vlot, N. Neupane, K. Varossieau, K. Sakamoto, T. Suzuki, N. Dohmae, S. Yokoyama, P.J. Schaap, H. Urlaub, A.J. Heck, E. Nogales, J.A. Doudna, A. Shinkai, J. van der Oost, RNA targeting by the type III-A CRISPR–Cas Csm complex of *Thermus thermophilus*, Mol. Cell 56 (2014) 518–530.

[30] T. Osawa, H. Inanaga, C. Sato, T. Numata, Crystal structure of the CRISPR–Cas RNA silencing Cmr complex bound to a target analog, Mol. Cell 58 (2015) 418–430.

[31] D.W. Taylor, Y. Zhu, R.H. Staals, J.E. Kornfeld, A. Shinkai, J. van der Oost, E. Nogales, J.A. Doudna, Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning, Science 348 (2015) 581–585.

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

[33] R.P. Hayes, Y. Xiao, F. Ding, P.B. van Erp, K. Rajashankar, S. Bailey, B. Wiedenheft, A. Ke, Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*, Nature 530 (2016) 499–503.

[34] M.L. Hochstrasser, D.W. Taylor, J.E. Kornfeld, E. Nogales, J.A. Doudna, DNA targeting by a minimal CRISPR RNA-guided Cascade, Mol. Cell 63 (2016) 840–851.

[35] S. Chowdhury, J. Carter, M.F. Rollins, S.M. Golden, R.N. Jackson, C. Hoffmann, L. Nosaka, J. Bondy-Denomy, K.L. Maxwell, A.R. Davidson, E.R. Fischer, G.C. Lander, B. Wiedenheft, Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex, Cell 169 (2017) 47–57.

[36] Y. Xiao, M. Luo, R.P. Hayes, J. Kim, S. Ng, F. Ding, M. Liao, A. Ke, Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR–Cas system, Cell 170 (2017) 48–60.

[37] T.W. Guo, A. Bartesaghi, H. Yang, V. Falconieri, P. Rao, A. Merk, E.T. Eng, A.M. Raczkowski, T. Fox, L.A. Earl, D.J. Patel, S. Subramaniam, Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR–Cas surveillance complex, Cell 171 (2017) 414–426.

[38] Y. Xiao, M. Luo, A.E. Dolan, M. Liao, A. Ke, Structure basis for RNA-guided DNA degradation by Cascade and Cas3, Science 361 (2018), eaat0839.

[39] C. Benda, J. Ebert, R.A. Scheltema, H.B. Schiller, M. Baumgärtner, F. Bonneau, M. Mann, E. Conti, Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4, Mol. Cell 56 (2014) 43–54.

[40] X. Zhu, K. Ye, Cmr4 is the slicer in the RNA-targeting Cmr CRISPR complex, Nucleic Acids Res. 43 (2015) 1257–1267.

[41] G. Tamulaitis, M. Kazlauskiene, E. Manakova, Č. Venclovas, A.O. Nwokeoji, M.J. Dickman, P. Horvath, V. Siksnys, Programmable RNA shredding by the type III-A CRISPR–

Cas system of *Streptococcus thermophilus*, Mol. Cell 56 (2014) 506–517.
[42] A. Hrle, A.A. Su, J. Ebert, C. Benda, L. Randau, E. Conti, Structure and RNA-binding properties of the type III-A CRISPR-associated protein Csm3, RNA Biol. 10 (2013) 1670–1678.
[43] T. Numata, H. Inanaga, C. Sato, T. Osawa, Crystal structure of the Csm3–Csm4 subcomplex in the type III-A CRISPR–Cas interference complex, J. Mol. Biol. 427 (2015) 259–273.
[44] T.Y. Jung, Y. An, K.H. Park, M.H. Lee, B.H. Oh, E. Woo, Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity, Structure 23 (2015) 782–790.
[45] G. Gallo, G. Augusto, G. Rangel, A. Zelanis, M.A. Mori, C.B. Campos, M. Würtele, Structural basis for dimer formation of the CRISPR-associated protein Csm2 of *Thermotoga maritima*, FEBS J. 283 (2016) 694–703.
[46] Č. Venclovas, Structure of Csm2 elucidates the relationship between small subunits of CRISPR–Cas effector complexes, FEBS Lett. 590 (2016) 1521–1529.
[47] K. Sakamoto, Y. Agari, K. Agari, S. Yokoyama, S. Kuramitsu, A. Shinkai, X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from *Thermus thermophilus* HB8, Proteins 75 (2009) 528–532.
[48] N.G. Lintner, M. Kerou, S.K. Brumfield, S. Graham, H. Liu, J.H. Naismith, M. Sdano, N. Peng, Q. She, V. Copié, M.J. Young, M.F. White, C.M. Lawrence, Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE), J. Biol. Chem. 286 (2011) 21643–21656.
[49] A. Hatoum-Aslan, P. Samai, I. Maniv, W. Jiang, L.A. Marraffini, A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs, J. Biol. Chem. 288 (2013) 27888–27897.
[50] T.D. Goddard, C.C. Huang, T.E. Ferrin, Visualizing density maps with UCSF Chimera, J. Struct. Biol. 157 (2007) 281–287.
[51] L. You, J. Ma, J. Wang, D. Artamonova, M. Wang, L. Liu, H. Xiang, K. Severinov, X. Zhang, Y. Wang, Structure studies of the CRISPR-Csm complex reveal mechanism of co-transcriptional interference, Cell 176 (2019) 239–253.
[52] N. Jia, C.Y. Mo, C. Wang, E.T. Eng, L.A. Marraffini, D.J. Patel, Type III-A CRISPR-Cas Csm complexes: Assembly, periodic RNA cleavage, DNase activity regulation, and autoimmunity, Mol. Cell 73 (2019) 264–277.
[53] H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko, N. Ben-Tal, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules, Nucleic Acids Res. 44 (2016) W344–W350.
[54] Y. Li, Y. Zhang, J. Lin, S. Pan, W. Han, N. Peng, Y.X. Liang, Q. She, Cmr1 enables efficient RNA and DNA interference of a III-B CRISPR–Cas system by binding to target RNA and crRNA, Nucleic Acids Res. 45 (2017) 11305–11314.
[55] Z. Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode, Methods Enzymol. 276 (1997) 307–326.
[56] 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. D 66 (2010) 213–221.
[57] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D 60 (2004) 2126–2132.
[58] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[59] A. Vagin, A. Teplyakov, Molecular replacement with MOLREP, Acta Crystallogr. D66 (2010) 22–25.