FEBS Letters

FEBS PRESS science publishing by scientists

Structure of Csm2 elucidates the relationship between small subunits of CRISPR-Cas effector complexes

Česlovas Venclovas

Institute of Biotechnology, Vilnius University, Lithuania

Correspondence
Č. Venclovas, Institute of Biotechnology, Graičiūno 8, Vilnius University, Vilnius LT–02241, Lithuania
Fax/Tel: +370 52602116
E-mail: ceslovas.venclovas@bti.vu.lt

(Received 23 January 2016, revised 3 March 2016, accepted 7 April 2016, available online 5 May 2016)

doi:10.1002/1873-3468.12179

Edited by Christian Griesinger

Type I and type III CRISPR-Cas effector complexes share similar architecture and have homologous key subunits. However, the relationship between the so-called small subunits of these complexes remains a contentious issue. Here, it is shown that the recently solved structure of *Thermotoga maritima* Csm2 represents a dimer with the extensive structure swapping between monomers. Unswapping the structure generates a compact globular monomer which shares similar structure and surface properties with Cmr5, the small subunit of a related Cmr complex. Detailed analysis of available structures of small subunits reveals that they all have a common fold suggesting their common origin.

Keywords: CRISPR-Cas; Csm2; protein fold; protein homology; small subunits; structural similarity

Approximately half of bacteria and the majority of archaea possess Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) arrays and CRISPR-associated (Cas) proteins [1,2]. The CRISPR array consists of conserved repeat sequences interrupted with so-called spacer sequences that originate from foreign DNA. The CRISPR locus transcription and subsequent processing produce short CRISPR RNA (crRNA) molecules, each carrying a single spacer. Together with Cas proteins crRNAs form large ribonucleoprotein effector complexes that detect and destroy virus or plasmid sequences if they have sequence complementary to the spacer region of crRNA. Thus, CRISPR-Cas systems function as adaptive immunity systems that protect their prokaryotic hosts from the invading foreign mobile genetic elements [3,4].

A recent classification of CRISPR-Cas systems divides them into two broad classes depending on the composition of the crRNA-effector complexes [5]. Class 1 encompasses multisubunit effector complexes, whereas class 2 effector complexes consist of a single protein. The two classes of CRISPR-Cas systems are further subdivided into five different types and multiple subtypes; however, except for a recently characterized type V system [6], only the major CRISPR-Cas types (I–III) have been studied in detail. Type II is a class 2 system that includes a single effector protein, Cas9. In contrast, types I and III are class 1 systems represented by multiprotein effector complexes. Best studied among the latter are Cascade (type I-E), Csm (III-A), and Cmr (III-B) complexes.

Computational and experimental studies revealed that the Cascade and type III CRISPR-Cas complexes are linked by evolution [5,7–12], but the relationship is intricate. Thus, all three complexes share a similar architecture of two intertwined helical protein filaments. The major filament, the backbone, in all three complexes is formed by Cas5 and Cas7 family proteins (Fig. 1). The minor filament or the 'belly' in Cascade is formed by two copies of the small subunit (Cse2) together with a four-helix bundle of the large subunit. A corresponding minor filament in the Cmr complex is composed of multiple copies of the small subunit (Cmr5) and the C-terminal domain of the large

Abbreviations
CRISPR, Clusters of regularly interspaced short palindromic repeats; Cas, CRISPR associated; crRNA, CRISPR RNA; PDB, protein data bank.

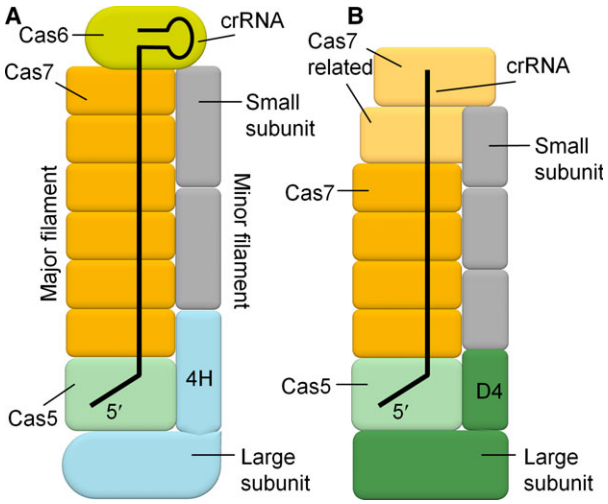

Fig. 1. Schematic representation of generic Cascade (A) and Cmr/Csm (B) complexes. Cas7 and Cas5 subunits common to all complexes are shown correspondingly in light green and gold. The C-terminal 4-helix bundle of the large subunit (Cse1) of Cascade is denoted as '4H'. The C-terminal domain within the unrelated large subunit (Cmr2/Csm1) of the Cmr/Csm complex is denoted as 'D4'. 'Cas7 related' subunits in Cmr correspond to Cmr6 and Cmr1, whereas in Csm they correspond to the extra Cas7 subunit and Csm5.

subunit, Cmr2 [11,13,14]. Although for the Csm complex only low-resolution EM structures are available [15,16], it is believed that the small (Csm2) and large (Csm1) subunits play similar architectural role in the formation of the analogous filament. At the same time, Cascade and type III complexes have unrelated large subunits at the base and those capping the intertwined-filament structure at the head.

One of the open questions in our understanding of the origin and evolution of multisubunit CRISPR-Cas effector complexes is the relationship between their small subunits. Based on the α-helical nature of small subunits and their ubiquitous presence in CRISPR-Cas systems, it was hypothesized that these subunits might be related and therefore tentatively could be unified into a single group (Cas11) [7]. However, sequence-based analyses failed to establish homology between different families of small subunits. In such cases experimentally determined structures are usually more informative, because protein structures tend to be more strongly conserved than sequences. However, even crystal structures were unable to give a straightforward answer.

Until recently, only structures for Cse2 (type I-E), Csa5 (type I-A), and Cmr5 (type III-B) subunits from different organisms were solved. Cse2 is a small two-domain all α-helical protein [17,18], Csa5 also has two domains but only the α-helical domain (hereafter D1) is conserved within the family [19], whereas Cmr5 is a single-domain α-helical protein [20,21]. The first intriguing observation was made by noting the structural similarity between Cmr5 and the C-terminal (D4) domain of Cmr2 suggesting that they may interact within the Cmr complex [22], which was subsequently confirmed by the crystal structure of the Cmr complex [11]. Reeks *et al.* [19] found that domain D1 of the Csa5 subunit is related to the C-terminal (D2) domain of Cse2. They also noted similarity between Cmr5 and the N-terminal (D1) domain of Cse2. These observations led to the conclusion that there are two structural families, one represented by domain D1 of Cse2 and Cmr5 and the other one represented by domains D2 of Cse2 and D1 of Csa5 and that the two families are unrelated to each other [19]. Most recently the missing structure of Csm2, the small subunit of the Csm complex, has been solved [23]. It revealed a dimer of α-helical monomers. Based on the structural analysis the authors concluded that Csm2 has a different fold than either Cse2, Cmr5, or Csa5 and thus might be a functional analog rather than structural homolog of these other small subunits [23].

In this study, the structural relationship between small subunits was revisited. The analysis revealed that the Csm2 structure represents a dimer with the extensive swapping of structural elements of individual protein chains. The expected monomeric structure of Csm2 was found to exhibit close structural and functional resemblance to Cmr5. Furthermore, the exhaustive structural comparison of the available structures of small subunits revealed that they all share the same overall fold.

Materials and methods

Protein structures analyzed in this work were obtained from PDB (http://www.pdb.org). Sequence searches for close homologs were performed with BLAST [24] against the nonredundant sequence database obtained from NCBI (http://www.ncbi.nlm.nih.gov). Multiple sequence alignments were constructed with MAFFT v7.221 [25]. Remote homology detection was performed using HMM profile–profile searches with HHPRED [26]. Predicted secondary structures were obtained using PSIPRED [27], while the actual secondary structures were derived from crystal structures with DSSP [28]. Protein structure comparison and superposition was done using DALILITE v2.4.1 [29]. Protein structure and surface visualization and analysis as well as construction of models of protein complexes were performed using UCSF CHIMERA [30]. Electrostatic properties of protein surfaces were computed with the APBS (Adaptive Poisson–

Boltzmann Solver) tool [31]. Figures representing protein sequence alignments and three-dimensional structures were prepared with ESPRIPT [32] and UCSF CHIMERA, respectively.

Results and Discussion

Csm2 structure represents a swapped dimer

A recently solved *Thermotoga maritima* Csm2 crystal structure (PDB id: 5AN6) is a dimer with an extensive interface between the two chains. Each chain of the dimer is composed of four α-helices and bears no resemblance to the structures of other small subunits [23]. However, it should be emphasized that in solution Csm2 exists in both dimeric and monomeric forms [23]. The solved structure accounts for the dimeric form, but raises a question as to the structure of a monomer. An extensive hydrophobic surface buried at the dimer interface would be exposed to the solution in the isolated individual chain and the N-terminal α-helix would become detached from the second and the third α-helices (Fig. 2B). This observation argues strongly against the monomeric structure being the same as that of an individual chain of the dimer. What then could be the structure of a monomer? The analysis of homologous sequences from other *Thermotoga* species and the secondary structure prediction provide an immediate clue. Thus, close Csm2 homologs from *Thermotoga naphthophila* (82% identical to *T. maritima* Csm2) and *Thermotoga profunda* (43% sequence identity) feature respectively an insertion and a deletion within the long H3 α-helix (Fig. 2A) suggesting that in these homologs the corresponding α-helix is disrupted. Moreover, secondary structure prediction for the *T. maritima* Csm2 confidently assigns two α-helices in place of the single H3 α-helix (Fig. 2A). Notably, the linker region connecting the two predicted helices coincides with the insertion/deletion in the Csm2 homologs. Taken together these observations suggest that the H3 α-helix is predisposed to form two helices. The break in H3 would allow the C-terminal region of a chain to fold back onto its own N-terminal region (Fig. 2B). This structural rearrangement would produce a globular monomeric structure consisting of five α-helices and corresponding to half of the dimer (Fig. 2C). In other words, the solved Csm2 structure apparently represents a structure-swapped dimer which can be unswapped by introducing a break into the H3 α-helix. Extensive structural rearrangements required for the transition between monomeric and dimeric forms explain the observed relative stability of monomeric and dimeric populations of the *T. maritima* Csm2 in solution [23].

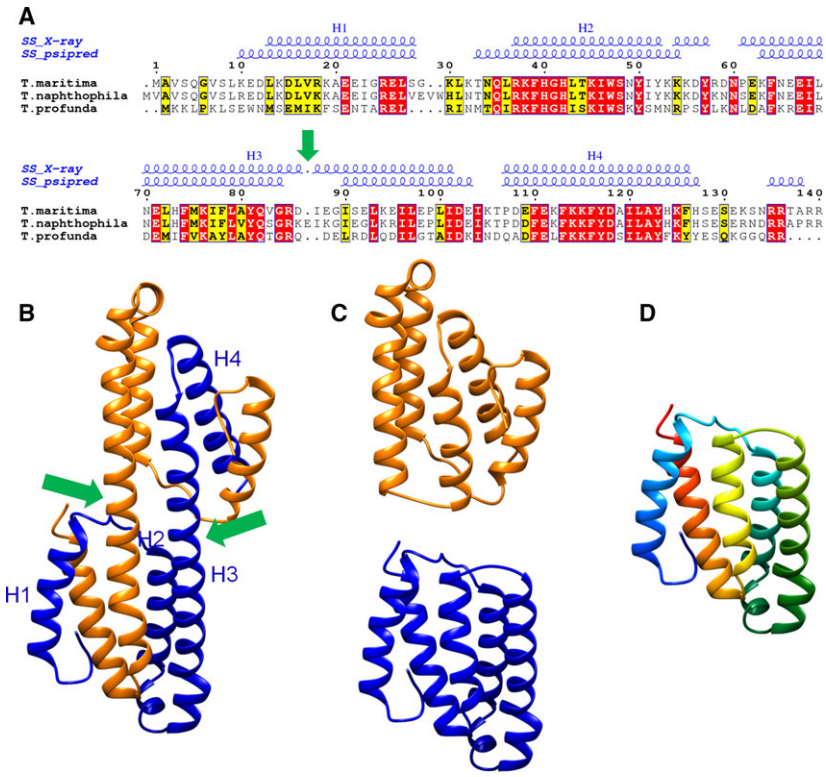

Fig. 2. Structure of *T. maritima* Csm2 represents a structure-swapped dimer. (A) Sequence alignment of Csm2 proteins from three *Thermotoga* species. Identical and similar residues are colored in red and yellow, respectively. Both predicted (*SS_psipred*) and actual (*SS_X-ray*) secondary structures are shown above the alignment. Green arrow indicates an expected break in the H3 helix of *T. maritima* Csm2. (B) Crystal structure of *T. maritima* Csm2 with two chains colored in blue and orange, respectively. Green arrows indicate breaks in H3 helices of corresponding chains for unswapping the dimeric structure. (C) Structures of monomers resulting from unswapping the dimeric structure. (D) Unswapped structure of a monomer with coloring indicating chain progression from N-terminus (blue) to C-terminus (red).

Unswapped Csm2 monomer is a structural homolog of Cmr5 with the expected similar binding properties

As the original Csm2 structure was found to lack similarity to other small subunits [23] an obvious question is what happens if the structure is unswapped. Structural comparison of the unswapped structure of Csm2 monomer with the Cmr5 subunit (PDB id: 2ZOP) of a related Cmr complex revealed that they do share similar structures (Fig. 3). Csm2 and Cmr5 could be superimposed with 2.9Å RMSD over 87 residue pairs. Moreover, the connectivity of α-helices in both structures is the same indicating that they have the same fold. Although structurally similar, superimposed Csm2 and Cmr5 display only negligible sequence similarity (~ 13% identical residues) providing an explanation of why the structural homology could not be identified previously. Structural similarity suggests that Csm2 and Cmr5 may perform identical or similar functions in the corresponding Csm and Cmr complexes. To date there is no high-resolution structure of a Csm complex. On the other hand, a crystal structure of Cmr bound to crRNA and single-stranded DNA has been recently reported [11]. Although the crystallized complex lacks one of the subunits (Cmr1) capping an end of the Cmr complex, it is active in cleaving the single-stranded RNA target complementary to the guide crRNA. This indicates that the Cmr crystal structure is functionally relevant [11]. The Cmr complex is composed of two helical protein filaments that form a groove for the binding of crRNA-target duplex. Two Cmr5 subunits together with the C-terminal domain (D4) of a large subunit, Cmr2, form the minor filament, which mostly contributes to the binding of target strand of the duplex (Figs 4A and S1). Consistent with this function, the Cmr5 subunits display an elevated positive electrostatic potential on the target-binding surface. To test whether Csm2 could perform a similar function, two Csm2 subunits were modeled into the filament in place of Cmr5 by simply overlaying them onto the Cmr5 subunits and then removing the latter (Fig. 4B). Although simplistic, the model offers two important observations. First, Csm2 subunits fit into the modeled filament without major steric clashes. Second, the Csm2 surface in the vicinity of the target strand also shows an increased positive electrostatic potential, which is even more strongly pronounced than in Cmr5. Taken together, these data establish that Csm2 and Cmr5 are homologous proteins performing the same or very similar function. It should be emphasized that the uncovered structural-functional similarity between Csm2 and Cmr5 is relevant for the unswapped structure of Csm2 monomer. The functional importance of the monomeric form of Csm2 is also indirectly supported by the composition of other Csm complexes. For example, the *Staphylococcus thermophilus* Csm complex containing a mature crRNA includes three Csm2 subunits [33], the estimated stoichiometries of *Thermus thermophilus* and *Sulfolobus solfataricus* Csm complexes also include three copies of Csm2 homologs [15,16]. Whether the Csm2 dimer has any role in the context of the Csm complex is not clear and additional experiments are needed to answer this question.

Small subunits and C-terminal domains of large Cmr/Csm subunits feature the same overall fold

The uncovered similarity between Csm2 and Cmr5 prompted to take a fresh look at the relationship between all known structures of small subunits. To this end, the unswapped monomeric Csm2 structure, three Cmr5 structures, N- and C-terminal domains of four different Cse2 structures, Csa5 domain D1, and C-terminal domains of Cmr2 and Csm1 all were compared against each other using DaliLite. Results of structural comparisons are provided as Dali *Z*-scores in Table 1 and more details for each comparison including RMSD values and the number of aligned residues are available in Table S1. Dali *Z*-score above 2 usually suggests that

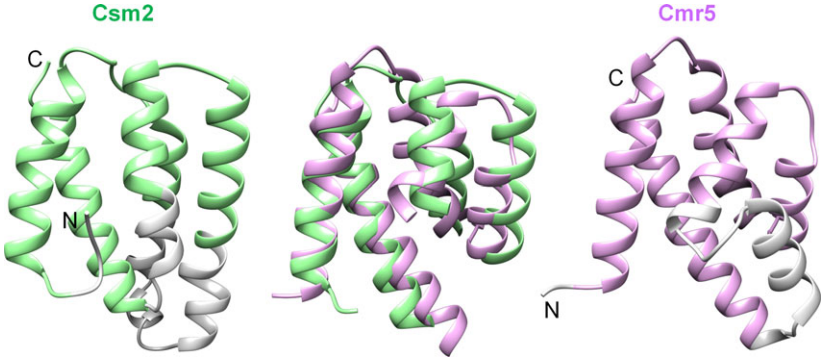

Fig. 3. Structural similarity between Csm2 and Cmr5. Corresponding structural regions of Csm2 (left) and Cmr5 (right) are colored in green and magenta, respectively. Middle panel shows structural superposition of these regions.

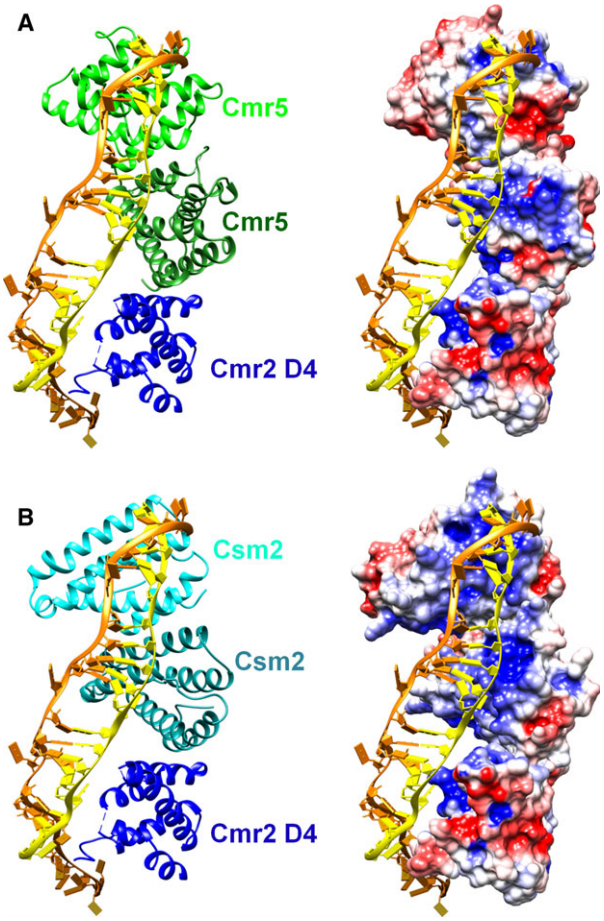

Fig. 4. Comparison of protein filaments involving Cmr5 and Csm2, respectively. (A) Crystal structure of the Cmr complex. For clarity only the Cmr2 D4 domain, two Cmr5 subunits, crRNA (orange) and the target oligonucleotide (yellow) are shown. Proteins are represented as a ribbon diagram (left) and as molecular surfaces (right) colored according to electrostatic potential (positive, blue; negative, red). (B) The same structure in which Cmr5 subunits were replaced with Csm2.

two structures might have the same fold, and the higher the *Z*-score value, the more significant structural similarity is expected to be. However, this *Z*-score threshold has been derived empirically and may vary for individual protein structural families. In addition, *Z*-score strongly depends on the size of the structure, with large structures producing higher *Z*-scores than the small ones. The Dali *Z*-scores (Table 1) immediately reveal highly similar groups of structures that comprise the Cmr5 family and individual domains of the Cse2 family. Most pairs within these groups have *Z*-scores > 5 and some even have *Z*-scores > 10. Somewhat lower scoring pairs include previously noted similarities between Csa5 D1 and Cse2 D2 domains [19], Cmr5 and Cse2 D1 domains [19], Cmr5 and Cmr2 D4 [22] as well as D4 domains of homologous large Cmr/Csm subunits (Cmr2 and Csm1, respectively) [34]. As expected from the established structural homology between Csm2 and Cmr5, Csm2 is also similar to the D4 domains of Cmr2/Csm1. Unexpectedly, however, DaliLite detected significant similarity (*Z*-scores > 3) of Csm2 with domain D2, but not with domain D1 of Cse2. Since the homologous relationship of Csm2 and Cmr5 is evident, this suggests that Cse2 D1 and D2 domains might also have similar folds but this similarity is perhaps too subtle for automatic detection. The protein fold similarity is commonly defined as the same major secondary structure elements arranged similarly in the three-dimensional space and connected in the same order [35]. Let us consider in the light of this definition the structures/domains of small subunits and Cmr2/Csm1 D4 domains. Apparently, all these structures have similarly arranged α-helices which are connected in the same order (Fig. 5). Admittedly, there is a fairly large variability of the corresponding helices in both size and orientation—some are even broken into two, some are shifted relative to the others. Nevertheless, all the structures can be described by the common α-helical fold consisting of five helices, arranged as two intercalating helix pairs (α1–α2 and α4–α5) each shaped as letter 'V' and connected together by the α3-helix (Fig. 5).

Why has the existence of a common fold until now escaped detection? This can be explained by at least two factors—the small domain size and the plasticity of the fold. The impact of the small size can be illustrated with relatively poor Dali *Z*-scores obtained by comparing Cse2 D1 domains with each other (Table 1). The structural plasticity or variability can be exemplified with C-terminal (D4) domains of Cmr2 and Csm1 subunits, both being members of the Cas10 family. The homology between the corresponding D4 domains can be detected with high confidence even from sequence data, for example, by using profile–profile comparison with HHsearch. However, despite obvious homology their structures feature a number of differences, in particular α4-helix, which is nearly completely replaced by the coil region in Csm1 D4 (Fig. 5). Thus, relating structures of small subunits appears by no means a trivial task and the Csm2 structure has contributed in a major way by helping to reveal previously unrecognized similarity between the two groups of structures represented by the individual domains of Cse2.

A common fold does not necessarily imply a homologous relationship between proteins. However, in this case it seems unlikely that small subunits of Cascade/Cmr/Csm and D4 domains of Cmr2/Csm1 had originated independently. With the established relationship between Csm2 and Cmr5, the homology between Csm

Table 1. Dali *Z*-scores of pairwise comparison of structures.[a]

Organism	Protein domain[b]	PDB chain domain[c]	5AN6u	2ZOPA	2OEBA	4GKFA	3W2WA D4	4UW2B D4	3ZC4A D1	2ZCAA D1	3WA8A D1	4H79A D1	4QYZB D1	2ZCAA D2	3WA8A D2	4H79A D2	4QYZB D2
T. maritima	Csm2	5AN6u	26.9	5.4	4.9	5	3.3	4.3	2.9	1.6	1.2	1.6	1.2	3	4	3.8	3.1
T. thermophilus	Cmr5	2ZOPA	5.4	27.2	14.6	14.4	4.4	5.1	2.3	5	3.7	5.3	3.8	1.6	2.1	2.6	2.2
A. fulgidus	Cmr5	2OEBA	4.9	14.6	33.5	18.3	4.3	4.3	1.4	3.6	3.2	3.9	2	2.1	1.8	1.9	1.6
P. furiosus	Cmr5	4GKFA	5	14.4	18.3	32.9	4.3	3.6	1.7	3.8	3.2	4.2	2.5	1.3	2.3	1.6	1.4
P. furiosus	Cmr2_D4	3W2WA D4	2.8	4.4	4.3	4.3	24.5	4.1	4.1	2.3	2.4	1.1	2.1	3.1	3	2.7	3.3
T. onnurineus	Csm1_D4	4UW2B D4	4.3	5.1	4.3	3.6	4.1	26.6	2.5	2	1	2.2	1.3	2.2	2.8	1.7	2.6
S. solfataricus	Csa5_D1	3ZC4A D1	2.9	2.3	1.4	1.7	4.1	2.5	22.7	1	1.3	2.3	1.3	6.4	6.7	6.8	5.9
T. thermophilus	Cse2_D1	2ZCAA D1	1.2	5	3.6	3.8	2.3	2	1	19.5	9.4	7	5.6	1.4	1.6	1.7	1.2
M. ruber	Cse2_D1	3WA8A D1	1.2	3.7	3.2	3.2	2.4	1	1	9.4	18.8	5.5	5.5	1.2	1.7	0.6	1.3
T. fusca	Cse2_D1	4H79A D1	1.6	5.3	3.9	4.2	1.3	2.2	2.3	7	5.5	20.6	3.9	1	2	0.6	0.7
E. coli	Cse2_D1	4QYZB D1	1.2	3.8	2	2.5	2.1	1.3	1.9	5.6	5.5	3.9	18.5	1.3	0.7	1.3	0.9
T. thermophilus	Cse2_D2	2ZCAA D2	3	1.6	2.1	1.3	3.1	2.2	6.4	1.4	1.2	1	1.3	20.1	13	11.9	9.8
M. ruber	Cse2_D2	3WA8A D2	4	2.1	1.8	2.3	3	2.8	6.7	1.6	1.7	2	0.5	13	23.5	11.8	10.4
T. fusca	Cse2_D2	4H79A D2	3.8	2.6	1.9	1.6	2.7	1.7	6.8	1.7	1.3	0.7	0.5	11.9	11.8	21.2	9.6
E. coli	Cse2_D2	4QYZB D2	3.1	2.2	1.6	1.4	3.3	2.6	5.9	1.2	1.3	0.8	0.9	9.8	10.4	9.6	19.7

A. fulgidus, Archaeoglobus fulgidus; M. ruber, Meiothermus ruber; T. fusca, Thermobifida fusca.

[a] Dali *Z*-scores in general are not symmetric, therefore, values for comparisons in both directions are provided; darker shading corresponds to more significant *Z*-scores.

[b] Protein name and domain: D1, N-terminal domain; D2 and D4, C-terminal domains; Csa5_D1, conserved α-helical domain.

[c] PDB code, chain id and domain number if applicable; 5AN6u indicates the unswapped Csm2 structure.

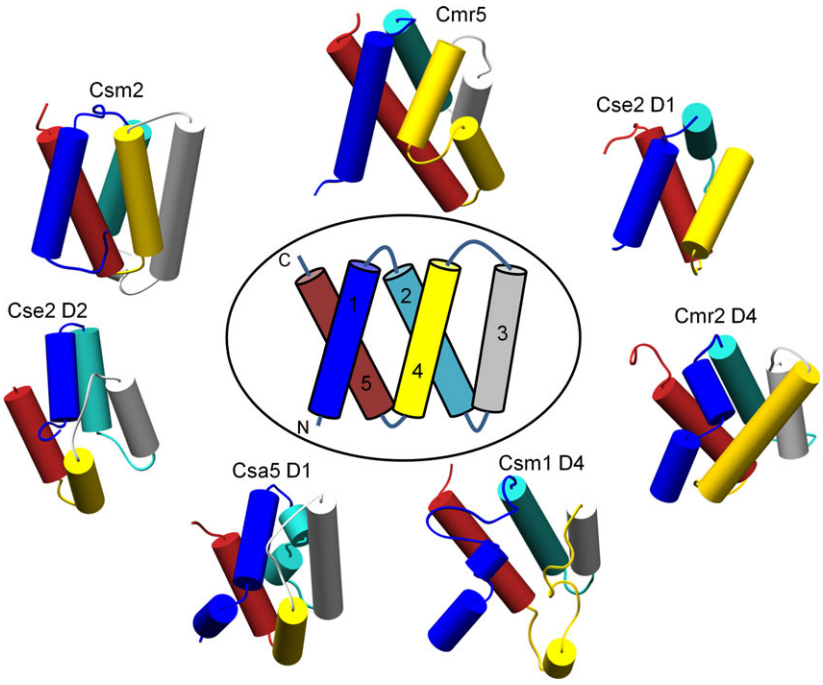

Fig. 5. Comparison of experimental structures of small subunits and D4 domains of large Cmr/Csm subunits. Structures include the unswapped structure of *T. maritima* Csm2 (PDB: 5AN6), *T. thermophilus* Cmr5 (PDB: 2ZOP), *T. thermophilus* Cse2 domains D1 and D2 (PDB: 2ZCA), domain D1 of *S. solfataricus* Csa5 (PDB: 3ZC4), and D4 domains of *Pyrococcus furiosus* Cmr2 (PDB: 3W2W) and *Thermococcus onnurineus* Csm1 (PDB: 4UW2). Structures are shown in a similar orientation and corresponding α-helices are colored in the same color. The encircled schematic diagram represents an idealized consensus fold.

and Cmr complexes can now be extended to all of their subunits. Furthermore, Cascade and Csm/Cmr complexes have also been suggested to have common roots [7]. These complexes share similar overall architecture and their corresponding major filaments include homologous subunits. Since small subunits Cse2, Cmr5, and Csm2 in their respective effector complexes play an equivalent architectural role by forming the minor protein filament, it is reasonable to assume that Cse2 might be homologous to Csm2 and Cmr5. It is also interesting to note that of all small subunits the Csm2 structure appears to be closest to the common fold (Fig. 5). This observation is in line with the proposed evolution of CRISPR-Cas complexes from an ancestral Csm-like complex [7]. Moreover, the present study could supply more details to the proposed evolutionary scenario. In particular, the observed higher similarity of Csm2 to Cse2 D2 and Cmr5 to Cse2 D1 is consistent with Cse2 originating from the fusion of Csm2-like and Cmr5-like subunits. In conclusion, both general and specific insights derived from the structural analysis in the present study are expected to facilitate structural, functional, and evolutionary studies of known and newly discovered CRISPR-Cas effector complexes.

Acknowledgements

The author thanks Virgis Siksnys for comments and suggestions and Ana Venclovienė for proof-reading the manuscript.

Author contributions

ČV designed the study, performed research and wrote the manuscript.

References

1 Grissa I, Vergnaud G and Pourcel C (2007) CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35**, W52–W57.
2 Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF *et al.* (2011) Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**, 467–477.
3 Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA and Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712.
4 Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845.
5 Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJ, Charpentier E, Haft DH *et al.* (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722–736.
6 Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A *et al.* (2015) Cpf1 Is a single

RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771.
7 Makarova KS, Aravind L, Wolf YI and Koonin EV (2011) Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* **6**, 38.
8 Jackson RN, Golden SM, van Erp PB, Carter J, Westra ER, Brouns SJ, van der Oost J, Terwilliger TC, Read RJ and Wiedenheft B (2014) Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479.
9 Mulepati S, Heroux A and Bailey S (2014) Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* **345**, 1479–1484.
10 Zhao H, Sheng G, Wang J, Wang M, Bunkoczi G, Gong W, Wei Z and Wang Y (2014) Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150.
11 Osawa T, Inanaga H, Sato C and Numata T (2015) Crystal structure of the CRISPR-Cas RNA silencing Cmr complex bound to a target analog. *Mol Cell* **58**, 418–430.
12 Jackson RN and Wiedenheft B (2015) A conserved structural chassis for mounting versatile CRISPR RNA-guided immune responses. *Mol Cell* **58**, 722–728.
13 Spilman M, Cocozaki A, Hale C, Shao Y, Ramia N, Terns R, Terns M, Li H and Stagg S (2013) Structure of an RNA silencing complex of the CRISPR-Cas immune system. *Mol Cell* **52**, 146–152.
14 Staals RH, Agari Y, Maki-Yonekura S, Zhu Y, Taylor DW, van Duijn E, Barendregt A, Vlot M, Koehorst JJ, Sakamoto K *et al*. (2013) Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of *Thermus thermophilus*. *Mol Cell* **52**, 135–145.
15 Staals RH, Zhu Y, Taylor DW, Kornfeld JE, Sharma K, Barendregt A, Koehorst JJ, Vlot M, Neupane N, Varossieau K *et al*. (2014) RNA targeting by the type III-A CRISPR-Cas Csm complex of *Thermus thermophilus*. *Mol Cell* **56**, 518–530.
16 Rouillon C, Zhou M, Zhang J, Politis A, Beilsten-Edmands V, Cannone G, Graham S, Robinson CV, Spagnolo L and White MF (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell* **52**, 124–134.
17 Agari Y, Yokoyama S, Kuramitsu S and Shinkai A (2008) X-ray crystal structure of a CRISPR-associated protein, Cse2, from *Thermus thermophilus* HB8. *Proteins* **73**, 1063–1067.
18 Nam KH, Huang Q and Ke A (2012) Nucleic acid binding surface and dimer interface revealed by CRISPR-associated CasB protein structures. *FEBS Lett* **586**, 3956–3961.
19 Reeks J, Graham S, Anderson L, Liu H, White MF and Naismith JH (2013) Structure of the archaeal Cascade subunit Csa5: relating the small subunits of CRISPR effector complexes. *RNA Biol* **10**, 762–769.
20 Park JH, Sun J, Park SY, Hwang HJ, Park MY, Shin M and Kim JS (2013) Crystal structure of Cmr5 from *Pyrococcus furiosus* and its functional implications. *FEBS Lett* **587**, 562–568.
21 Sakamoto K, Agari Y, Agari K, Yokoyama S, Kuramitsu S and Shinkai A (2009) X-ray crystal structure of a CRISPR-associated RAMP module [corrected] Cmr5 protein [corrected] from *Thermus thermophilus* HB8. *Proteins* **75**, 528–532.
22 Zhu X and Ye K (2012) Crystal structure of Cmr2 suggests a nucleotide cyclase-related enzyme in type III CRISPR-Cas systems. *FEBS Lett* **586**, 939–945.
23 Gallo G, Augusto G, Rangel G, Zelanis A, Mori MA, Campos CB and Wurtele M (2016) Structural basis for dimer formation of the CRISPR-associated protein Csm2 of *Thermotoga maritima*. *FEBS J* **283**, 694–703.
24 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.
25 Katoh K, Misawa K, Kuma K and Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30**, 3059–3066.
26 Söding J, Biegert A and Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* **33**, W244–W248.
27 Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292**, 195–202.
28 Kabsch W and Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637.
29 Holm L and Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* **16**, 566–567.
30 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC and Ferrin TE (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.
31 Baker NA, Sept D, Joseph S, Holst MJ and McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci USA* **98**, 10037–10041.

32 Robert X and Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* **42**, W320–W324.
33 Tamulaitis G, Kazlauskiene M, Manakova E, Venclovas Č, Nwokeoji AO, Dickman MJ, Horvath P and Siksnys V (2014) Programmable RNA shredding by the type III-A CRISPR-Cas system of *Streptococcus thermophilus*. *Mol Cell* **56**, 506–517.
34 Jung TY, An Y, Park KH, Lee MH, Oh BH and Woo E (2015) Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure* **23**, 782–790.
35 Murzin AG, Brenner SE, Hubbard T and Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536–540.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Fig. S1. The minor protein filament involving Cmr5 and the Cmr2 D4 domain is shown in the context of entire structure of the Cmr complex. Cmr2 D4 domain, two Cmr5 subunits, crRNA (orange) and the target oligonucleotide (yellow), are shown in solid colors. Remaining Cmr subunits are shown in semitransparent gray color.

Table S1. DaliLite pairwise comparison of structures (RMSD Å/superimposed residue pairs).