Structural organization of a Type III-A CRISPR effector subcomplex determined by X-ray crystallography and cryo-EM

Bryan W. Dorsey, Lei Huang and Alfonso Mondragón *

Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

ABSTRACT

Clustered regularly interspaced short palindromic repeats (CRISPR) and their associated Cas proteins provide an immune-like response in many prokaryotes against extraneous nucleic acids. CRISPR–Cas systems are classified into different classes and types. Class 1 CRISPR–Cas systems form multiprotein effector complexes that includes a guide RNA (crRNA) used to identify the target for destruction. Here we present crystal structures of *Staphylococcus epidermidis* Type III-A CRISPR subunits Csm2 and Csm3 and a 5.2 Å resolution single-particle cryo-electron microscopy (cryo-EM) reconstruction of an *in vivo* assembled effector subcomplex including the crRNA. The structures help to clarify the quaternary architecture of Type III-A effector complexes, and provide details on crRNA binding, target RNA binding and cleavage, and intermolecular interactions essential for effector complex assembly. The structures allow a better understanding of the organization of Type III-A CRISPR effector complexes as well as highlighting the overall similarities and differences with other Class 1 effector complexes.

INTRODUCTION

Many prokaryotes, including most archaea and many bacteria, utilize immune-like response systems to defend themselves against phages and genomic invasions. One of these adaptive systems is comprised of genomic loci formed by clustered regularly interspaced short palindromic repeats (CRISPRs) and their CRISPR associated (Cas) proteins (1–5). Although originally CRISPR was thought to function as an antisense system targeting invading phages (6), it was later shown that it can also protect against horizontal gene transfer (7) and transformation (5). CRISPR loci contain several non-adjacent direct DNA repeats separated by variable sequences, called spacers, which correspond to segments of captured viral and plasmid sequences (8). These loci allow for immune memory against invading foreign genetic elements by integrating the foreign DNA, which is then transcribed and processed to form CRISPR RNA (crRNA) (6,9,10). The crRNA, in conjunction with Cas proteins, are utilized for the identification and eventual destruction of recurrent invaders (11).

There is a wide diversity in the organization and architecture of CRISPR–Cas systems. Two main classes (Class 1 and 2) have been identified from extensive comparative analysis of both the loci and proteins associated with each system (12). Class 1 CRISPR–Cas systems (Types I, III and IV) consist of multi-protein effector complexes that function together for target surveillance and defense, whereas Class 2 CRISPR–Cas systems (Types II, V and VI) have a single protein that incorporates both surveillance and defense stages into one protein, Cas9 or homologs (11,12). Within Class 1 systems, the number of proteins involved and the general architecture of the effector complexes varies and even within each type there are differences in subunit stoichiometry. Type I systems can be subdivided into seven subtypes, whereas Type III CRISPR systems are further divided into four subtypes, III-A to III-D (12). All Type III subtypes include the Type III-defining *cas10* gene (13), which encodes the signature Type III Cas10 protein that is responsible for DNA target degradation (14). There are two distinct effector complexes formed within Type III subtypes, known as the Csm (Types III-A/D) and Cmr (Types III-B/C) effector complexes (15–17). While originally it was not clear whether Type III systems targeted foreign DNA or RNA for destruction, it was later shown that in Type III-A systems, both RNA and DNA can be cleaved, but utilize different active sites and in a co-transcriptional manner (18). Further studies showed that both Type III-A and III-B effector complexes act as RNases and target RNA-activated DNA nucleases (19–21), unifying the previously incongruent mechanisms of the Type III systems into one cohesive mechanism.

The *Staphylococcus epidermidis* RP62a CRISPR–Cas system has served as the paradigm for the Type III-A sub-

---

*To whom correspondence should be addressed. Tel: +1 847 491 7726; Email: a-mondragon@northwestern.edu
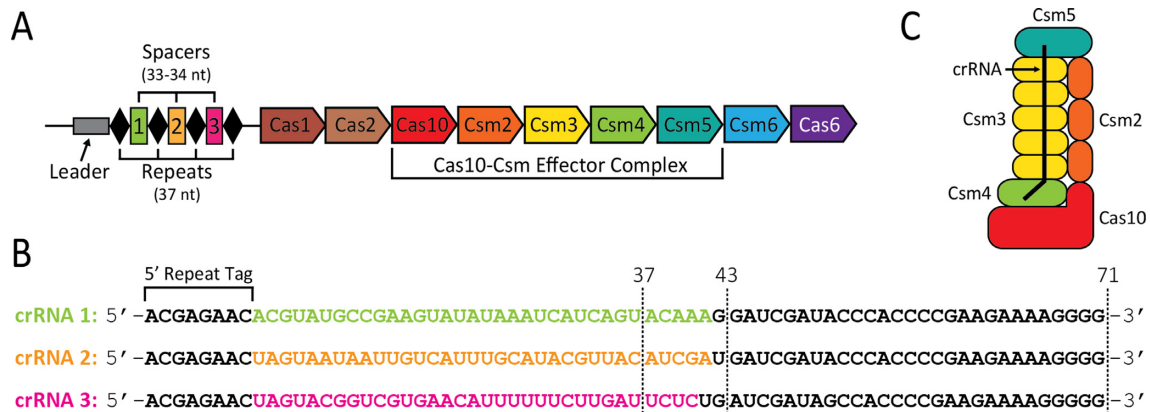Present address: Lei Huang, Donlen Corporation 3000 Lakeside Dr., 2nd Floor Bannockburn, IL 60015, USA.

Figure 1. *S. epidermidis* RP62a CRISPR–Cas locus and effector complex schematics. (**A**) Schematic of *S. epidermidis* RP62a CRISPR–Cas locus architecture. The locus is composed of the leader (gray rectangle) and tandem copies of the repeats (black diamonds) and spacers (numbered 1–3), followed by the nine associated protein-coding *cas* genes. Five proteins (Cas10, Csm2–5) associate to form the effector complex. (**B**) Sequences of the three crRNAs associated with the *S. epidermidis* RP62a repeat-spacer array shown in A. Repeat sequences are shown in black and spacer sequences for crRNA 1–3 are shown in green, orange, and magenta, respectively. Intermediate and mature lengths are denoted by a dotted line with the respective length listed above. (**C**) Schematic diagram of a proposed model of the Type III-A CRISPR effector complex with a 43 nucleotide (nt) CRISPR RNA (crRNA). The effector complex model is composed of five copies of Csm3 (yellow), three copies of Csm2 (orange), and one copy each of Csm4 (green), Csm5 (teal) and Cas10 (red), in addition to the 43 nt crRNA (black line) (28).

type. Only one CRISPR locus is encoded in *S. epidermidis* RP62a and contains three 33–34 base pair (bp) spacers flanked by four 37 bp direct repeats (Figure 1A) (22). The repeat-spacer array is processed as a continuous transcript that is initially processed into 71 nucleotide intermediate crRNAs by Cas6 (23,24). Cellular nucleases, such as polynucleotide phosphorylase (PNPase), have been shown in *S. epidermidis* RP62a to interact with the effector complex, generating two predominant mature crRNA species of 37 and 43 nucleotides (Figure 1B) (25). Previous studies have determined that the CRIPSR-Cas effector complex is composed of the Cas proteins Cas10, Csm2, Csm3, Csm4 and Csm5 in an undetermined stoichiometry and known as the Cas10–Csm effector complex (Figure 1A) (26). These proteins combine to form two helical protein arrangements, known as the major and minor filaments (Figure 1C) (27,28). The major filament is composed of multiple copies of Csm3, a repeat-associated mysterious protein (RAMP) containing an RNA recognition motif (RRM) fold. The filament is capped on each end by two other RAMPs, Csm4 and Csm5 (5,27,29). Csm3, in particular, has been implicated in binding the crRNA guide, as well as functioning as an endoribonuclease that acts to cleave the ssRNA target in a periodic manner (26). The minor filament is composed of multiple copies of Csm2, the smallest subunit in the Cas10–Csm effector complex. Its central function is thought to be in target strand binding (28). The protein component stoichiometry of both the major and minor filaments is dependent on the length of the crRNA, suggesting that the crRNA can modulate the assembly of the effector complex (27,30). The final and largest component of the Cas10–Csm effector complex is the signature Type III protein Cas10, an exonuclease implicated in DNA target strand degradation (31,32). Thus, the Cas10–Csm complex contains two different catalytic activities, with DNA degradation achieved by Cas10 and RNA degradation accomplished by Csm3, with both guided by the crRNA. Ad-

ditionally, recent studies have shown that Type III effector complexes generate secondary messengers that activate an accompanying CRISPR ribonuclease family protein, Csm6 (33). These results represent a newly discovered third interference mechanism in Type III CRISPR systems where the effector complex acts as a cyclic oligoadenlyate synthetase that converts ATP to cyclic oligoadenylates, activates Csm6, and allows for the targeting and destruction of invading phage transcripts (34).

Our understanding of the architecture and mechanism of Type III-A systems has been limited due to the lack of high resolution structural information on the Cas10–Csm complex and its components. Although there is a wealth of information on other CRISPR types, such as Type I, the inherent diversity in the mechanism and architecture of the CRISPR systems precludes extrapolating from one type to the other, limiting how much we can learn from one type about another. Recently, two Type III-A effector complex structures from *Thermococcus onnurineus* (35) and *Streptococcus thermophilus* (36) were reported. Structures of these effector complexes with crRNA alone, as well as ternary complexes with target RNA or anti-tag target RNA, show how the complexes are assembled and the changes that occur upon target RNA binding. Nevertheless, the extensive wealth of biochemical and functional information on Type III-A systems is largely based on the well characterized *S. epidermidis* system, making it important to obtain structural information on the *S. epidermidis* complex. We have studied the structure of the *S. epidermidis* Cas10–Csm effector complex by a combination of crystallographic and cryo-electron microscopy (cryo-EM) studies. The studies reveal the crystal structures of Csm2 and Csm3, the two most abundant subunits forming the backbone of the complex. In addition, a 5.2 Å resolution cryo-EM structure of a subcomplex, including most of the subunits and the crRNA, sheds light on the overall architecture of the Cas10–Csm complex, the binding of crRNA in the complex, and the ac-

tive site of Csm3. Furthermore, these structural studies help understand many of the previous biochemical observations on the mechanism of this particular CRISPR–Cas complex. Finally, there is a practical interest in understanding the mechanism of CRISPR–Cas immunity in *S. epidermidis*, as this gram positive bacteria are the most common of the coagulase-negative staphylococci found on human epithelia (37,38) and are among the most common pathogens that lead to nosocomial infections (38,39).

MATERIALS AND METHODS

SeCsm2 and SeCsm3 recombinant plasmid construction and expression

The coding region for the *csm2* and *csm3* genes were PCR-amplified from extracted *S. epidermidis* RP62a genomic DNA using PCR primers for *csm2* (Csm2-F and Csm2-R) and *csm3* (Csm3-F and Csm3-R) (Supplementary Table S1). PCR-amplified genes were inserted into the pET21-derived expression plasmid pMCSG7 (40) through the ligation-independent cloning (LIC) region. Utilizing the LIC region introduces an in-frame 6X Histidine tag and a tobacco etch virus (TEV) cleavage site at the N-terminus of the protein. The constructs were confirmed by DNA sequencing (ACGT Inc., Germantown, MD) and subsequently transformed into Rosetta competent cells (EMD Millipore) for protein expression and purification. Both constructs were grown in Terrific Broth (TB) media supplemented with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37°C to an OD_{600} of 0.8. Cells were cooled on ice for 1 h to reduce leaky gene expression before induction with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 18 h at 16°C. Cells were harvested by centrifugation at 5,500 × *g* for 15 min at 4°C. Cell paste was then flash frozen in liquid nitrogen and stored at −80°C until purification.

Expression of selenomethionine substituted SeCsm2 and SeCsm3 was performed by growth of a 5 ml pre-culture in Luria Broth (LB) rich media grown for 8 h at 37°C. A 10 ml M9 minimal media starter culture was inoculated with 1:1000 pre-culture and grown overnight at 37°C. One liter of M9 minimal media cultures were inoculated with 1:100 M9 starter and grown to an OD_{600} of 0.8. Cells were cooled on ice for 45 min before adding 100 mg lysine, 100 mg threonine, 100 mg phenylalanine, 50 mg leucine, 50 mg isoleucine, 50 mg valine and 50 mg L(+)-selenomethionine (ACROS Organics). After 15 min, cells were induced with 0.5 mM IPTG for 18 h at 16°C. Cells were harvested and frozen following the same protocol as for the native proteins.

Recombinant SeCsm2 protein purification

Pellets were thawed on ice for 30 min before adding 1:1 volume of resuspension buffer (100 mM Tris pH 8, 1M NaCl, 20 mM imidazole). Resuspended cells were incubated with 50 μM lysozyme, 0.12% Brij 58, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 1 mM benzamidine on a nutating mixer for 30 minutes at 4°C. Cells were lysed using a Misonix S-4000 sonicator with a 1/2" horn (Amplitude 20, 5 s on, 10 s off, 5 min total on-time). Cell lysate was clarified through ultracentrifugation at 160,000 × *g* for 45 min

at 4°C in a Ti-70 rotor (Beckman-Coulter). Insoluble cell debris was removed and supernatant was filtered through a 0.2 μm polyethersulfone (PES) membrane bottle-top filter. Filtered lysate was loaded onto a column containing 10 ml nickel nitrilotriacetic acid (Ni-NTA) Superflow resin (Qiagen) equilibrated with buffer 1 (50 mM Tris pH 8, 500 mM NaCl, 10 mM imidazole). The column was washed with 500 ml buffer 1, followed by protein elution in 10 ml fractions of elution buffer 1 (50 mM Tris pH 8, 500 mM NaCl, 250 mM imidazole). Elution fractions containing protein were dialyzed overnight at 4°C against 1.5 l dialysis buffer 1 (50 mM Tris pH 8, 500 mM NaCl). Dialyzed protein was concentrated using a Vivaspin 20 3 kDa molecular weight cutoff (MWCO) concentrator (GE Healthcare) to 2 ml. The protein was then diluted to 100 mM NaCl immediately before loading onto a column packed with 10 ml High-S cation-exchange resin (Bio-Rad) attached to a Gilson Minipuls 3 peristaltic pump at 1.2 ml/min to reduce the likelihood of precipitation in low salt. The column was washed with 50 ml of low salt buffer (50 mM Tris pH 8, 100 mM NaCl), followed by a 100 ml linear gradient from 100 mM to 1 M NaCl. Protein-containing fractions were pooled and dialyzed overnight at 4°C against 1.5 l dialysis buffer 1. Finally, the protein was concentrated to 0.7–1.2 mg/ml for crystallization experiments. Selenomethionine-labeled SeCsm2 (SeMet–SeCsm2) was purified using the same protocol.

Recombinant SeCsm3 protein purification

Pellets were thawed on ice for 30 min before adding 1:1 volume of no salt buffer (50 mM Tris pH 8, 20 mM imidazole). Resuspended cells were incubated with 50 μM lysozyme, 0.12% Brij 58, 1 mM PMSF and 1 mM benzamidine on a nutating mixer for 30 min at 4°C. Additionally, 10 mM $MgCl_2$, 2 mM $CaCl_2$ and 16 μM bovine pancreas deoxyribonuclease I (Sigma Aldrich) were added and allowed to stir for 15 min at room temperature. Following this incubation, NaCl was added to a final concentration of 500 mM. Cells were further lysed using a Misonix S-4000 sonicator with a 1/2" horn (Amplitude 20, 5 s on, 10 s off, 5 min total on-time). Insoluble cell debris was removed by ultracentrifugation at 160,000 × *g* for 45 min at 4°C in a Ti-70 rotor (Beckman-Coulter). To remove unbound nucleic acids, a 5% solution of polyethyleneimine (PEI) pH 7.9 was added dropwise to a final concentration of 0.2% into the supernatant while stirring at 4°C (41). The solution was stirred for 5 min before removing the precipitated nucleic acids through centrifugation at 18,000 × *g* for 30 min in a FA-45-6-30 fixed-angle rotor (Eppendorf). The clarified lysate was filtered through a 0.2 μm PES membrane bottle-top filter. Filtered lysate was loaded onto a 10 ml Ni-NTA Superflow resin column equilibrated with buffer 1. The column was washed with 200 ml buffer 1, followed by 400 ml buffer 2 (50 mM Tris pH 8, 200 mM NaCl, 10 mM imidazole). Protein was eluted in 10 ml fractions with elution buffer 2 (50 mM Tris pH 8, 200 mM NaCl, 250 mM imidazole). Protein-containing fractions were incubated with 1:10 molar concentration of TEV protease while dialyzing overnight at 4°C against 1.5 l dialysis buffer 2 (50 mM Tris pH 8, 100 mM NaCl, 10 mM imidazole). The dialysis bag was transferred to fresh 1.5 l dialysis buffer 2 and allowed to equi-

librate again overnight at 4°C. Dialyzed, cleaved protein was subjected to another round of affinity chromatography to remove the cleaved His tag, uncleaved protein, and His-tagged TEV protease. The Ni-NTA column flow through was loaded onto a 10 ml High-S cation exchange resin column equilibrated with 50 mM Tris pH 8, 100 mM NaCl at 0.4 ml/min. Nucleic acids bind to the resin, allowing the nucleic acid-free protein to be collected in the flow through. Nucleic acid-free SeCsm3 was concentrated with a Vivaspin 20 10 kDa MWCO concentrator to 4–7 mg/ml to be used for crystallization experiments. SeMet–SeCsm3 was purified using the same protocol.

Crystallization conditions of SeCsm2 and SeCsm3

Crystallization trials of native and SeMet-derived SeCsm2 and SeCsm3 were conducted using a variety of commercial screens and utilizing both the hanging drop and sitting drop vapor diffusion methods. The best diffracting SeCsm2 and SeMet-SeCsm2 crystals were set up at 4°C in VDX hanging drop vapor diffusion plates (Hampton) and grown at 10°C in 100 mM Tris pH 7, 8% ethanol at 1.2 mg/ml. Crystals formed after 1–2 days of incubation. Crystals were cryo-protected by supplementing the crystallization condition with 20% ethylene glycol and 10% ethanol. SeMet–SeCsm3 crystals yielding the best diffraction data were set up at room temperature in Cryschem sitting drop vapor diffusion plates (Hampton) and grown at 14°C in 20% polyethylene glycol (PEG) 8000, 100 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 6.5, 200 mM calcium acetate at 6.8 mg/ml. Crystals formed after 24 h of incubation. SeMet–SeCsm3 crystals were cryo-protected by supplementing the crystallization condition with 20% ethylene glycol. For heavy atom derivatization, SeMet–SeCsm3 crystals were soaked for 2 min in the cryo-protectant supplemented with 10 mM samarium (III) chloride before freezing.

SeCsm2 structure determination

X-ray diffraction data were collected from native SeCsm2 and SeMet-SeCsm2 crystals at 100 K at the Life Sciences Collaborative Access Team (LS-CAT) beamlines at the Advanced Photon Source, Argonne National Laboratory, using a Rayonix MX300 CCD detector. All crystallographic data were processed with XDS (42) and Aimless (43). Positions of three Se atoms were obtained with the program Shake-and-Bake (44). A Single Anomalous Dispersion (SAD) electron density map was initially obtained using SHARP (45). The map showed clearly the fold of the protein, but the data were limited to 3.2 Å resolution. An experimental electron density map including both the SeMet data and a higher resolution native data set was calculated with SHARP (45) and was of better quality. The combined SeMet/native map allowed tracing of most of the molecule with the aid of Coot (46) followed by iterative model building and refinement with BUSTER (47), Phenix (48,49), and Refmac5 (50). The final model spans most of the molecule, aside from one short disordered loop (residues 29–36) and two amino acids missing at the C-terminus, and includes nine amino acids that form part of the His-tag used for purification. The stereochemistry of the model was validated with MolProbity (51) and Coot (46). The final model has an R_{work}/R_{free} of 24.47%/29.42% to 2.75 Å resolution with excellent stereochemistry. X-ray crystallographic statistics for the structure are summarized in Supplementary Table S2.

SeCsm3 structure determination

X-ray diffraction data were collected from SeMet–SeCsm3 crystals at 100 K at LS-CAT beamlines at the Advanced Photon Source, Argonne National Laboratory, using a Dectris Eiger 9M detector. Molecular Replacement calculations failed to give a solution and hence data from SeMet crystals soaked in 10 mM samarium (III) chloride were collected with the incident beam tuned to 11.5 keV (1.0781 Å) and 6.724 keV (1.8439 Å) for experimental phasing. Diffraction data were processed using XDS (42) and Aimless (43). An experimental electron density map was calculated using the automated pipeline CRANK-2 (52) in CCP4 (53) using the 1.8439 Å wavelength data set. Initial model building was performed with CRANK-2 (52) in CCP4 (53) followed by multiple iterations of automated model building using Buccaneer (54) and manual model editing using Coot (46). The model was refined against the 1.0781 Å data set using Refmac5 (50). The final model contains two molecules in the asymmetric unit (root-means-square-deviation [RMSD] of 0.15 Å) and spans most of the molecule (A monomer 2–20, 32–65, 74–124 and 137–214, B monomer 0–20, 31–65, 75–124, 137–214), with the missing amino acids residing in three disordered loops. The stereochemistry of the model was validated with MolProbity (51) and Coot (46). The final model has an R_{work}/R_{free} of 23.1%/26.7% to 2.40 Å resolution with excellent stereochemistry. X-ray crystallographic statistics for the structure are summarized in Supplementary Table S3.

S. epidermidis RP62a Cas10–Csm complex purification

S. epidermidis RP62a harboring p*crispr* with Cas10 N-terminally His-tagged was generously provided by Dr. Asma Hatoum-Aslan at the University of Alabama (26). Strains were grown in BBL brain heart infusion media (BD). The media was supplemented with neomycin (15 µg/ml) and chloramphenicol (10 µg/ml). Cultures (1 l culture per 2 l flask) were grown for 24 h at 37°C. Cells were harvested, frozen at −80°C, and stored until purification. Cell pellets from 6 l of culture volume were thawed at 4°C for 30 min and resuspended in 30 ml of resuspension buffer (30 mM MgCl₂, 35 µg/ml lysostaphin [Ambi Products]) supplemented with PMSF and benzamidine (1 mM final concentration). Cells were incubated at 37°C for 1 h, mixing every 20 min. Cell lysates were subsequently diluted with 1/3 lysate volume of lysis buffer (200 mM NaH₂PO₄ pH 7.4, 2 M NaCl, 80 mM imidazole, 0.4% Triton X-100, 2 mM PMSF, 2 mM benzamidine). Sonication was performed on the diluted lysates on ice (Amplitude 25, 5 s on, 10 s off, 10 min total on-time [Misonix S-4000, 1/2" horn]). Sonicated lysate was spun at 18,000 × *g* for 45 min at 4°C in a FA-45-6-30 fixed-angle rotor. The insoluble cell debris was removed and spun again at 18,000 × *g* for 30 min. The remaining cell debris was discarded and the clarified lysate was filtered through a 0.2 µm PES membrane bottle-top

filter. The filtered, clarified lysate was loaded onto a column packed with Ni-NTA Superflow resin (1 ml slurry per 1 l culture volume) equilibrated with equilibration buffer (50 mM NaH_2PO_4 pH 7.4, 500 mM NaCl, 10 mM imidazole). The lysate was passed over the resin twice before washing with 600 ml wash buffer (50 mM NaH_2PO_4 pH 7.4, 500 mM NaCl, 20 mM imidazole). Protein was eluted from the column with 25 ml elution buffer A (50 mM NaH_2PO_4 pH 7.4, 500 mM NaCl, 250 mM imidazole) in 5 ml fractions. Protein-containing fractions were pooled and dialyzed against 1.5 l dialysis buffer A (20 mM Tris pH 8, 200 mM NaCl) at 4°C overnight. Dialyzed complex was diluted to 100 mM NaCl immediately before loading onto a 10 ml column packed with Q Sepharose Fast Flow resin (GE Healthcare) equilibrated with buffer B (20 mM Tris pH 8, 100 mM NaCl) running at 1.4 ml/min. The column was washed with 25 ml of buffer B, followed by a 100 ml linear gradient from 100 mM to 2 M NaCl. Fractions of interest were pooled and dialyzed against 1.5 l dialysis buffer B (20 mM Tris pH 8, 500 mM NaCl) at 4°C overnight. Dialyzed complex was then concentrated to ∼300 μl and loaded onto a Superdex 200 10/300 GL (GE Healthcare) equilibrated with dialysis buffer B. The non-aggregated complex peak was collected. Peak fractions are stored at 4°C for up to 1 month without any noticeable degradation as assayed by SDS-PAGE.

Mass spectrometry analysis of the crRNA

crRNAs were isolated by multiple rounds of 1:1 phenol-chloroform extractions from the purified effector complex gel filtration peak fractions used in the EM structural studies. Following extraction, the crRNAs were ethanol precipitated with 1/10 volume 3 M sodium acetate, pH 5.2 and 3 volumes 100% ethanol and stored at −20°C. Samples were spun down and washed three times with 80% ethanol to remove excess salt. A final spin was performed, the ethanol was removed, and the crRNAs were dried. Once dry, the lyophilized pellets were resuspended in ddH_2O and submitted to the Integrated Molecular Structure Education and Research Center (IMSERC) at Northwestern University for mass spectrometry. Samples were run on an Agilent 6210A LC-TOF High Resolution Time of Flight Mass Spectrometer using ElectroSpray Ionization (ESI), connected to an Agilent 1200 series HPLC. Samples were analyzed using a Direct Loop injection, with water/acetonitrile (60/40) as carrier solvent. MassHunter Workstation Data Acquisition software was used for instrument operation and MassHunter Qualitative Analysis software for data analysis and processing.

in vitro target RNA cleavage assay

in vitro time-course RNA cleavage assays were performed by incubating 100 nM SeCas10–Csm complex with 50 nM 5′-fluorescein amidite (FAM)-labeled target RNA complementary to the Spacer 1 sequence (5′-FAM-AGCCUGACUGAUGAUUUAUAUACUUCG GCAUACGUUCCAG-3′; Integrated DNA Technologies [IDT]) in a 30 μl reaction. Reactions were performed in 20 mM Tris pH 7.5, 150 mM NaCl, 10 mM $MgCl_2$, 2

mM DTT and 3% glycerol. Control reactions replaced the $MgCl_2$ with 10 mM ethylenediaminetetraacetic acid (EDTA) to abolish cleavage activity. Reactions were stopped by addition of 2× formamide loading buffer (90% formamide, 0.025% SDS, 5 mM DTT), heated to 95°C for 5 min, analyzed on a 15% polyacrylamide denaturing 8 M urea gel, and visualized with a Typhoon 9400 scanner (GE Healthcare). In addition, cleavage assays under identical conditions but using a non-target RNA (bacterial RNase P) showed no cleavage, confirming that cleavage was not the result of contaminating nucleases.

Negative stain electron microscopy sample preparation and data acquisition

Carbon-coated copper Gilder grids (300 mesh, Electron Microscopy Sciences [EMS]) were glow-discharged for 7 s at 10 W in a Solarus 950 plasma cleaner (Gatan). Glow-discharged grids were placed carbon side down onto a 30 μl drop of 30 μg/ml purified complex for 5 min at room temperature. Grids were subsequently transferred twice to 30 μl drops of 20 mM Tris pH 8, 500 mM NaCl for 1 min per drop, with blotting occurring after the second transfer. The grids were then transferred twice to 30 μl drops of 2% uranyl acetate solution for 30 seconds per drop, followed by blotting after the second transfer and finally allowing the grid to air dry. Once dry, the grids were imaged on a JEOL 1400 transmission electron microscope (TEM) operating at 120 keV. 175 micrographs were collected using Leginon (55) at a magnification of 40,431× using an UltraScan4000 CCD camera with a pixel size of 3.71 Å at the specimen level and using a defocus range of –2 μm to –3 μm. Micrographs were processed using RELION-2.1 (56,57). A small group of micrographs were selected for manually picking particles for initial 2D classification and the 2D classes were used for subsequent auto-picking of the entire dataset (∼385,000 particles). Auto-picked particles were subjected to successive rounds of 2D classification to remove improperly picked particles yielding ∼175,000 particles. The ∼175,000 particles formed two distinct subsets of defined 2D classes. The first subset contained ∼63,000 particles detailing a more linear complex, whereas the second subset of ∼112,000 particles represented a double-helical complex. Each subset was subjected to 3D classification. An input model for each subset was generated *ab initio* from a subset of these particles. The model best representing the 2D classes was chosen as the initial model for cryo-EM studies.

Cryo-Electron microscopy sample preparation and data acquisition

Cryo-EM grids (C-flat, 1.2/1.3, 400 mesh, Cu; [EMS]) were glow-discharged for 10 s at 15 mA in an easiGlow glow discharger (Pelco). Grids were prepared under >95% humidity at 4°C using a Vitrobot Mark IV (FEI ThermoFisher). A glow-discharged cryo grid was placed in the Vitrobot sample chamber, onto which a 3 μl drop of protein (∼0.2 mg/ml) was deposited. After waiting 5 s, the grid was blotted for 4 s with a blot force of +5 and immediately plunged into liquid ethane. The grid was transferred to a grid box for

storage until data collection. Leginon (55) was used for automated EM image collection. Micrographs were collected using a JEOL 3200FS TEM operating at 300 keV equipped with an in-column energy filter (Omega filter) and a K2 Summit direct electron detector (Gatan). A magnification of 40,323 × was used, for a pixel size of 1.24 Å at the specimen level and a defocus range of –2 μm to –3.5 μm. Movies were collected in counting mode with a total dose of ∼37.5 e⁻/Å², fractionated into 0.2 s frames for a total of 8 s, corresponding to 40 frames per movie. A total of 1,037 movies were collected in this manner.

Cryo-EM image processing and model building

Cryo-EM movies were processed using RELION-3.0 (57–59). Motion correction was performed using the program MotionCor2 (60) and dose-weighted according to the relevant radiation damage curves (61) within RELION-3.0 (58) to allow for downstream post-processing. The Contrast Transfer Function (CTF) was estimated using CTFfind4 (62) on motion-corrected micrographs. 277,413 particles were auto-picked from templates generated from manual 2D classification. Auto-picked particles (277,413) were subjected to 3D classification, with the best class containing 107,620 particles. Additional RELION-3.0 post processing steps included per-particle CTF estimation and particle polishing to further improve the resolution. The particle polishing step showed a significant improvement on the final calculated volume, which had a 5.2 Å resolution according to the Fourier Shell Correlation (FSC) criterion (63). Initial inspection of the map showed that the quality was uneven, with the two extremes of the molecule at lower resolution than the central stem. The map was segmented into two pieces, the major stem (5× SeCsm3 and 1× SeCsm4), and the rest of the volume (putatively SeCas10), to use in multibody refinement using RELION-3.0 (59). Multi-body refinement did not improve the resolution of the stem, but did improve slightly the volume around the SeCas10 domain.

To position the SeCsm3 crystal structure in the cryo-EM map, a 6 dimensional search was done using the program Essens (64), which identified five copies of SeCsm3 along the stem. The positions corresponded to the same locations identified visually. Essens and visual searches using SeCsm2 as a template found no matches, indicating SeCsm2 was not part of the complex. Further visual inspection of the map showed the presence of an additional RRM motif at the base of the stem. Superposition of SeCsm3 on the RRM motif showed good agreement for the RRM region, but not for other regions, suggesting that the region corresponded to an RRM-containing protein distinct from SeCsm3. This was assigned as SeCsm4. The remaining density was assigned as SeCas10 based on its position and size. No density corresponding to SeCsm5 could be identified. Once all five SeCsm3 molecules had been identified, their position was refined using jigglefit (65) in Coot (46). The models showed excellent fit in most areas, but some loops and secondary structure elements needed adjustment to fit the density. To improve the placing of the crystal structures, the monomers were fit into the map using the MDFF routines (66) that are part of NAMD (67). No further refinement of the mod-

els was done. Unless noted, the figures were made using the crystal structures fit into the map.

Model superpositions, comparisons, and figures

Comparisons of the different models were all done using Coot (46) and Chimera (68). When superposing structures from different organisms, the superposition was based on secondary structure matching (SSM) as implemented in Coot (46) and Chimera (68).

Figures were drawn with Pymol (69) and Chimera (68). Conservation data were generated using the ConSurf server (70,71) and visualized in Pymol (69). Electrostatic potential calculations were done with APBS (72). Homology models were generated using the Phyre2 server (73) and visualized in Pymol (69) and Chimera (68).

RESULTS

The crystal structure of *S. epidermidis* Csm2 reveals a structural organization intermediate between Type III-A and III-B molecules

S. epidermidis Csm2 (SeCsm2) was purified using a His-tag to apparent homogeneity under high salt conditions to help solubilize the purified protein (Supplementary Figure S1A). Initial crystallization trials were performed with and without the affinity tag attached, with crystals of SeCsm2 only forming when the affinity tag was present (Supplementary Figure S1B). Structural determination of SeCsm2 was accomplished by experimental Single Anomalous Dispersion (SAD) from selenomethionine-incorporated crystals (Materials and Methods). The final model of SeCsm2 (R_{work}/R_{free} = 24.5%/29.4%) was obtained from data to 2.75 Å and consists of residues 14–28 and 37–139 with an additional 9 residues N-terminal to the start methionine identified as the TEV protease cleavage site (Supplementary Table S2 and Supplementary Figure S1C). This additional region provided essential crystal contacts in the crystal lattice (Supplementary Figure S1D). After the structure was solved, the gene annotation for *csm2* was updated (74), showing an additional 13 residues upstream of the original starting residue. The new N-terminal region is predicted to be unstructured (Supplementary Figure S1E), which could provide similar crystal contacts to the TEV protease cleavage site found in the crystal structure. It is presently unknown whether these additional N-terminal residues are relevant to the function of SeCsm2. All residue numbers stated in this work reflect the updated annotation for the *csm2* gene. This structure is identical to a recently published crystal structure of SeCsm2, with the exception of the TEV protease cleavage site being absent in that structure (75).

The overall structure of SeCsm2 reveals an exclusively α-helical fold composed of six helices (Figure 2A) forming a compact five-helix bundle with the sixth helix, α_4, connecting two of the helices in the bundle. The organization of the helices generally follows the established fold for the small subunits of Class 1 CRISPR effector complexes (76,77). There are two notable structural differences between the SeCsm2 structure and the only other Type III-A Csm2 crystal structure available, from *Thermotoga maritima* (TmCsm2; PDB: 5AN6) (76) (Figure 2B). The first
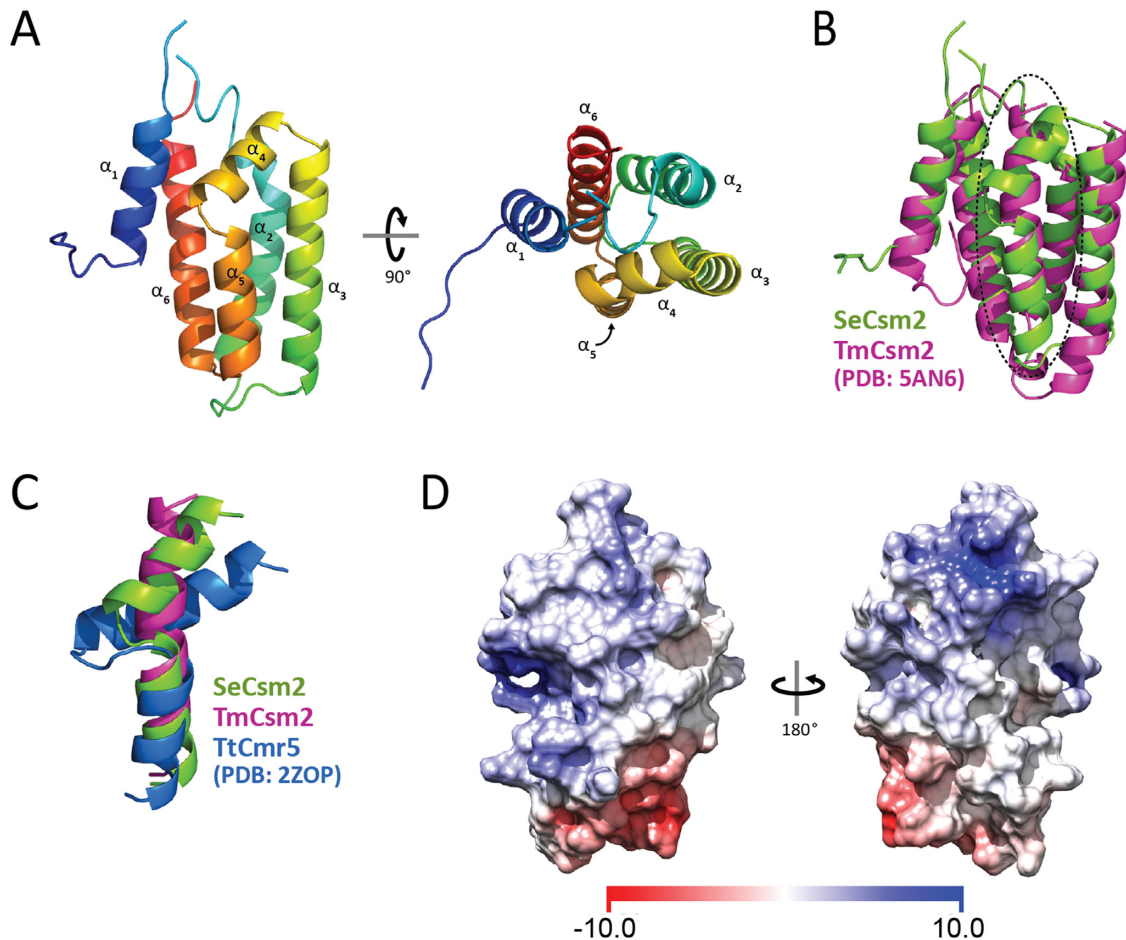
Figure 2. Molecular architecture of *S. epidermidis* Csm2. (**A**) Two perpendicular views of the crystal structure of SeCsm2 show an exclusively α-helical fold consisting of 6 helices. SeCsm2 is colored in a rainbow from the N- (blue) to C- (red) termini. A portion of the N-terminal affinity tag was present in the density and is included in the structure. (**B**) Structural alignment of SeCsm2 (green) and the putative monomeric version of TmCsm2 (76) (magenta). The overall architecture is similar with a root-means-square-deviation (RMSD) between structures of 1.98 Å. One significant difference is the rearrangement of TmCsm2 α3 to form two helices (α4–α5) in SeCsm2 (dashed oval). (**C**) Structural alignment of SeCsm2 α4–α5 (green) with Type III-A TmCsm2 α3 (magenta) and Type III-B TtCmr5 α4–α5 (78) (blue). The helix architecture of Type III-A and Type III-B Csm2 homologs reveals an intermediate organization in SeCsm2. (**D**) SeCsm2 accessible surface colored by electrostatic potential calculated with APBS (72). Two small positive regions (blue) are found on opposite sides of the molecule. One negative region (red) is shown at the opposite end from the positive regions. The bar at the bottom maps the electrostatic potential values from –10 k_BT/e_c [red] to 10 k_BT/e_c [blue].

difference occurs in the loop between the first two helices, α1 and α2. In SeCsm2, there is an extended loop (Loop 1) consisting of 20 amino acids (residues 25–44) with the middle 8 amino acids (residues 29–36) disordered in the crystal (Supplementary Figure S1C). This is much longer than the loop in TmCsm2, which only contains 6 amino acids (residues 30–35). In SeCsm2, Loop 1 contains many positively charged residues (K29, R31, K32, K34 and K36) and a pair of negatively charged residues (D30 and E35). Secondary structure predictions show an extension of α1 into Loop 1, however these residues are not part of α1 in the crystal (Supplementary Figure S1C). This suggests that this region is more flexible than predicted and is consistent with the absence of electron density for some of this loop. It is possible that this highly charged loop could play a functional role in target strand binding or protein-protein interactions within the effector complex, but its absence in the subcomplex structure (see below) precludes a more definite

assignment. Additionally, the crystal packing reveals that Loop 1 in each molecule faces the solvent space (Supplementary Figure S1D). The second difference involves the fourth helix in the bundle, α4, which rearranges in SeCsm2 to form two helices, α4 and α5, whereas it is a single helix in an unswapped monomer of TmCsm2 (Figure 2B) (77). The predicted secondary structure of SeCsm2 indicates that α4–α5 could form a single helix (Supplementary Figure S1C), which would be consistent with the helix topology described for Type III-A structures (Supplementary Figure S1F). Instead, it breaks and kinks in the vicinity of residue 103. Interestingly, the occurrence of two helices at this position is more prevalent in the Type III-B small subunit, Cmr5 (77). Comparison of SeCsm2 α4–α5 to TmCsm2 α3 and α4–α5 from the Type III-B *Thermus thermophilus* HB8 Cmr5 structure (TtCmr5; PDB: 2ZOP) (78) shows that SeCsm2 adopts an intermediate structural conformation between the single, linear helix in TmCsm2 and the pair of near per-

pendicular helices in TtCmr5 (Figure 2C). This 'intermediate structure' observation can be extended to the entire SeCsm2, with its relative helix orientations more similar to Type III-A small subunits, but including helix topologies reminiscent of those found in Type III-B small subunits. This suggests that the small subunits form a more diverse group of structures than previously thought.

SeCsm2 shows two positively charged electrostatic surfaces, with the first at the N-terminal end of α_2 and the other near where α_1 and α_6 overlap. Additionally, one negatively charged electrostatic surface is found at the opposite end of the molecule located near the N-terminal end of α_3 (Figure 2D). In contrast, TtCmr5 exhibits a large, positively charged electrostatic surface that has been implicated in target nucleic acid binding (77). When SeCsm2 and TtCmr5 are aligned, the positive patches are situated in different regions of the molecules. This could indicate alternate quaternary organizations between Type III-A and III-B effector complex minor filaments. The structural diversity within Type III-B Cmr5 subunits may play a larger role in producing alternate minor filament arrangements. Additionally, a positively charged surface in the TmCsm2 dimer has been observed (76), although it is not clear that the dimer represents a biologically relevant conformation. The packing in the crystal shows electrostatic interactions between adjacent SeCsm2 molecules in addition to the TEV cleavage site crystal contacts (Supplementary Figure S1D). However, these interactions do not suggest a model for SeCsm2 quaternary organization within the effector complex, as they are different from TmCsm2 and TtCmr5 models that show how oligomers could form that are suitable for RNA binding (77).

S. epidermidis Csm3 represents a minimized CRISPR backbone protein

Purification of nucleic acid-free *S. epidermidis* Csm3 (SeCsm3) was challenging due to its inherent ability to bind non-specifically to nucleic acids. Apparently pure SeCsm3 was contaminated with bound nucleic acids, which formed corkscrew-like aggregates of various lengths when visualized by negative stain electron microscopy (Supplementary Figure S2A). Nucleic acid-free SeCsm3 was only produced after very slow loading onto a strong cation exchange resin and collecting the flow through of the column, which contained pure SeCsm3 (Supplementary Figures S2B and S2C). Crystallization trials were conducted with nucleic acid-free SeCsm3, SeCsm3 bound to mature 37 nucleotide crRNA (Spacer 1 sequence), and SeCsm3 bound to intermediate 71 nucleotide crRNA (Spacer 1 sequence). However, only the nucleic acid-free SeCsm3 produced high resolution diffracting crystals (Supplementary Figure S2D). Molecular Replacement calculations gave no clear solution, despite using several models derived from many orthologs and hence experimental phasing was used to solve the structure. Although data were collected on both native and SeMet-substituted crystals, only the SeMet crystals were used in the final structure determination due to the crystals diffracting to higher resolution. Phasing using SeMet was not possible due to the weak anomalous signal, later explained by the location of many of the SeMet

in disordered regions of the structure (75% in molecule A and 50% in molecule B) (Supplementary Figure S2E). No cysteines are present in SeCsm3, eliminating the use of selenocysteine as an alternative labeling strategy. Structure determination was ultimately accomplished by experimental SAD phasing from SeMet–SeCsm3 crystals derivatized with samarium (III) chloride. The unit cell contains two molecules in the asymmetric unit (Supplementary Figure S3A). The final model of the SeCsm3 dimer (R_{work}/R_{free} = 23.0%/26.7%) was obtained from data to 2.40 Å and consists of residues 1–21, 31–65, 74–124 and 137–214 (Supplementary Figure S2E and Supplementary Table S3). The structure is identical to a recently published crystal structure of SeCsm3 (79). Experiments combining size-exclusion chromatography with multi-angle light scattering analysis (SEC-MALS) show a molecular weight consistent with a monomer of SeCsm3 (data not shown), confirming that the dimer in the crystal is a result of crystal packing and does not reflect the oligomerization state in solution.

SeCsm3 consists of three distinct domains: (i) an RRM fold, (ii) a β-hairpin and (iii) an isolated α-helix (Figure 3A). The RRM in SeCsm3 contains five anti-parallel β-strands and four α-helices ($\beta_1\alpha_1\alpha_2\beta_2\beta_5\alpha_4\alpha_5\beta_6\beta_7$) with an insertion in the middle, between β_2 and β_5, to form the β-hairpin (β_3 and β_4) and the additional helix (α_3) (Figure 3B). As expected for a molecule that binds nucleic acids, the electrostatic potential map shows a highly positively charged region spanning α_1, α_5, β_2, β_3, β_4 and the loops surrounding them (Supplementary Figure S3B). The crystallographic dimer shows the positive patches of each monomer facing one another, partially shielding them from the solvent (Supplementary Figure S3C). The monomers of the crystallographic dimer are identical, with an RMSD of 0.15 Å. To ascertain the importance of conserved regions in the protein, the sequence of SeCsm3 was compared against 150 homologous sequences using the ConSurf server (70,71) and the conserved regions were analyzed. The results show an abundance of conserved residues in the positive patch of the protein and low conservation in other regions (Supplementary Figure S3D), providing further evidence that the positively charged surface corresponds to a region of functional importance.

SeCsm3 is known to cleave target RNA (18) and mutating D32 abrogates this activity without impairing crRNA processing or complex formation (18), strongly suggesting that D32 is located in the RNA cleavage active site. Mapping of D32 in the structure shows that it is found just after a disordered loop and in the vicinity of many other highly conserved residues that have been shown to be important for modulating crRNA length (H18, D100, E120, K122, E124 and R129) (13,26). All these residues reside in the positive patch and many face in the same direction. Clustered together around D32 are H18, E124, R127, R137, R141 and R187. Based on sequence comparisons, H18 has been suggested to be part of the active site, although mutational studies of this residue in *S. epidermidis* have shown no effect on effector complex formation, crRNA binding, or target cleavage (26). Similar mutational studies of this conserved residue in other Type III Csm3 homologs have shown varied results, with no effect shown on CRISPR function in *Streptococcus thermophilus* Csm3 (80) and loss of target RNA
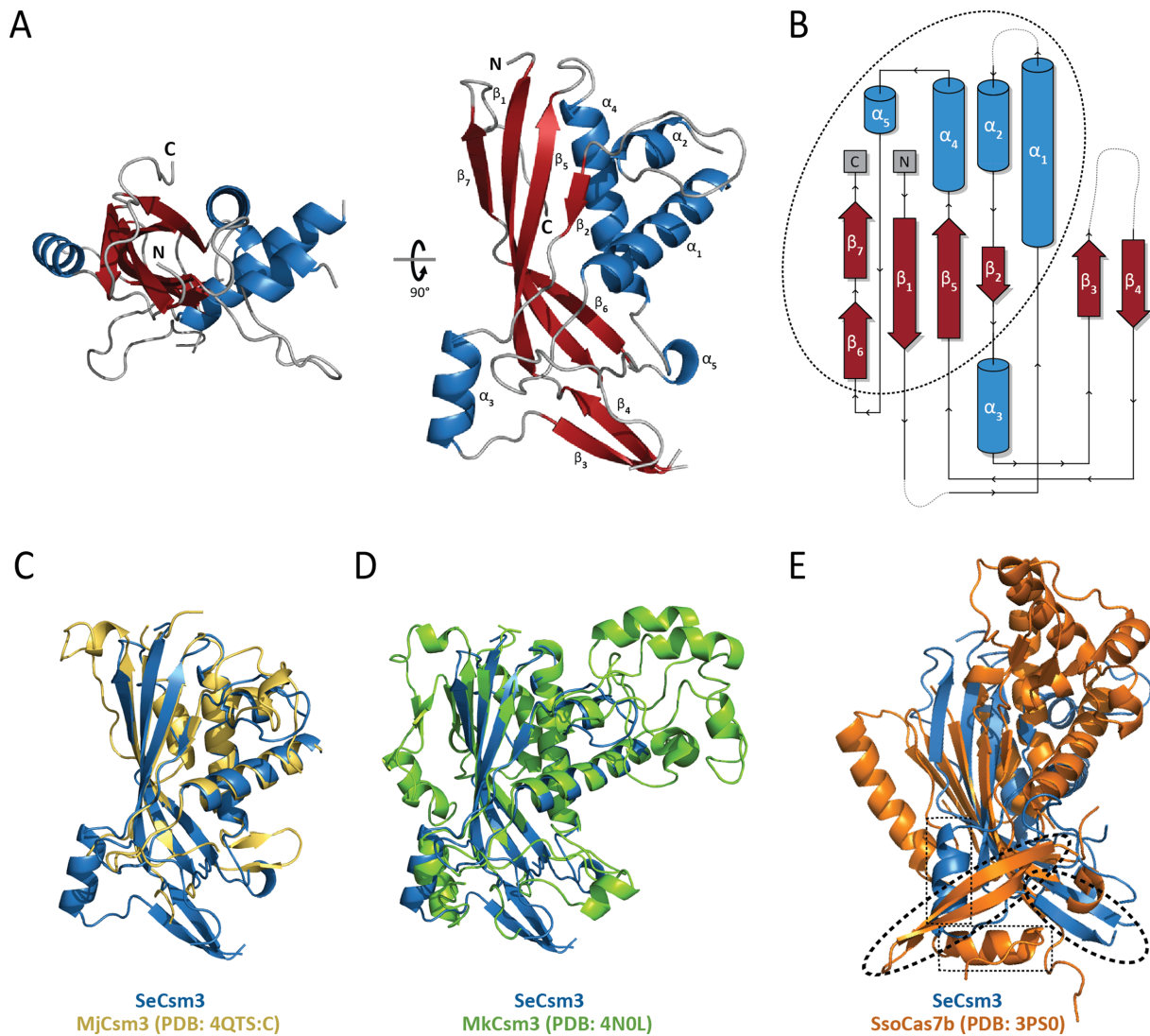
Figure 3. Molecular architecture of *S. epidermidis* Csm3. (**A**) Two perpendicular views of the crystal structure of SeCsm3 show that it is composed of an RRM fold consisting of a five-stranded anti-parallel β-sheet and four helices, a β-hairpin, and an isolated α-helix. Sheets are colored in red and helices in blue. (**B**) Topology diagram of SeCsm3 shows that the N- and C-terminal sequences form the RRM fold (dashed oval), with the other domains located at the center of the polypeptide. (**C-E**) Structural alignment of Cas7 family proteins. SeCsm3 is aligned with (**C**) MjCsm3 (82) (RMSD of 1.68 Å over 134 C_α atoms), (**D**) MkCsm3 (86) (RMSD of 1.98 Å over 153 C_α atoms), and (**E**) SsoCas7b (88) (RMSD of 2.99 Å over 122 C_α atoms). The β-hairpin (dashed oval) and helix (dashed rectangle) in SsoCas7b are rotated with respect to the ones in SeCsm3. The alignments show the wide variation in backbone subunit architecture of effector complexes of different types and subtypes.

binding in *Methanopyrus kandleri* Csm3 (81). This suggests that this basic residue has lost its function in some Csm3 homologs, but kept its functionality in others. The presence of acidic amino acids D32 and E124 suggests that they could be involved in cation binding. Finally, the presence of positively charged residues, such as K122 and R129 is also consistent with a region involved in binding nucleic acids. Overall, the clustering of the highly conserved amino acids H18, D32, E124, R127, R137, R141 and R187 strongly suggests that these amino acids are part of the active site involved in target RNA cleavage. Importantly, D100, E120 and K122, which have been shown to play a role in the regulation of crRNA length, are further away and not in the vicinity of the putative active site.

Multiple effector backbone subunit structures have been determined for Type III-A (81,82), Type III-B (14,83–85), Type I-D (86), Type I-E (87–90) and Type I-F (91,92) CRISPR systems, and each has shown the canonical RRM core region decorated with multiple, varying domains that may or may not play a role in its function. Type III-A Csm3 structures have been solved from *Methanocaldococcus jannaschii* (MjCsm3; PDB: 4QTS:C,D) (82) and *Methanopyrus kandleri* (MkCsm3; PDB: 4N0L) (81). Not surprisingly, SeCsm3 and MjCsm3, with 38% sequence identity, share the same core architecture of the RRM fold (Figure 3C). Unfortunately, in MjCsm3, the region corresponding to the β-hairpin and isolated α-helix in SeCsm3 is disordered, preventing structural comparison of these regions (Figure 3C). The main structural difference between SeCsm3 and

MjCsm3 occurs at the C-terminus, where an insertion from E213 to E224 in MjCsm3 accounts for roughly half of the residues that form the interface with MjCsm4 (82). When these residues were mutated to alanine, the interaction between the two subunits was lost. As SeCsm3 does not contain this region, it suggests that an alternate interaction surface must be present between SeCsm3 and SeCsm4. In contrast to MjCsm3, the structural differences between SeCsm3 and MkCsm3 are evident (Figure 3D). MkCsm3 contains a large N-terminal helical zinc-binding domain and a C-terminal helical domain that are completely absent from SeCsm3 (Figure 3D and Supplementary Figure S3E). When comparing the electrostatic potential maps of SeCsm3 and MkCsm3, there is a dramatic shift in the orientation of the putative RNA binding surface. SeCsm3 has the binding surface at the center of the molecule, whereas the MkCsm3 binding surface is on a more proximal region, classified as the lid domain (81). SeCsm3 also has a more occluded, channel-like binding region (Supplementary Figure S3B) compared to the flattened binding region on MkCsm3 (81). The variation of these binding region characteristics may change how successive Csm3 subunits interact to form the backbone of the effector complex and suggest that the structure of the effector complex may show differences even amongst members of the same subtype.

Comparing the SeCsm3 structure to effector complex backbone subunits from other types shows a more diverged RRM architecture as well as variation in domain organization. When SeCsm3 is aligned to *Sulfolobus solfataricus* Cas7b (SsoCas7b; PDB: 3PS0) (88), the equivalent Type I-E backbone subunit, the most prominent feature is the reorganization of the β-hairpin (Figure 3E and Supplementary Figure S3E). The β-hairpin in SsoCas7b is flipped >90° with respect to the equivalent β-hairpin in SeCsm3, causing the isolated α-helix to be oriented almost perpendicular to the equivalent α-helix in SeCsm3 (Figure 3E). In addition, SsoCas7b α_2 (corresponding to α_1 in SeCsm3) is positioned such that the putative binding region of SeCsm3 would be occluded. These rearrangements relative to SeCsm3 create a more open binding surface, with the majority of the positively charged patch located around the β-hairpin (88). In SsoCas7b, the additional helical domain proximal to the RRM fold as well as the C-terminal helical domain coincide surprisingly well to the respective C- and N-terminal helical domains found in MkCsm3 (Supplementary Figure S3E). It is presently unknown how these rearrangements in SsoCas7b change the effector complex backbone quaternary architecture relative to the structure of the SeCsm3 major filament.

S. epidermidis cryo-EM structure elucidates the quaternary architecture of the major filament subcomplex

In order to study the overall architecture of the Cas10–Csm effector complex, the complex was purified directly from *S. epidermidis* RP62a cells using a Cas10 subunit with a His tag attached. Purification of the *S. epidermidis* effector complex yielded a homogeneous sample where all five subunits were present and of the expected molecular weight (Figures 4A and 4B). The purified effector complex was characterized both by negative stain (Supplementary

Figures S4A and S4B) and cryo-EM (Supplementary Figures S4C-E). To prevent aggregation of the complex, which caused particles to overlap in the EM grid, the salt concentration was increased from 200 to 500 mM NaCl, enriching the non-aggregated complex gel filtration peak (Figure 4B). The high salt species still showed the presence of five subunits and isolation of the crRNA from the purified complex by phenol-chloroform extraction yielded two mature-length species corresponding to the 37-nucleotide sequence from Spacer 1 and the 43-nucleotide sequence from Spacer 2 (Figures 1B and 4C), consistent with extracted crRNAs from native complexes as shown in previous studies (26). The crRNA masses correspond to RNAs that have both 5′ and 3′ terminal hydroxyl groups, which are consistent with the cleavage products produced by Cas6 (23,24) and polynucleotide phosphorylase (PNPase) (25,93), respectively. These data conclusively show that the effector complex was fully processed before purification. Target RNA cleavage activity was also assessed to verify that the purified complex was active (Figure 4D). Negative stain EM analysis was performed to generate an initial model for cryo-EM studies (Supplementary Figures S4A and S4B). Two distinct complex species were observed within the same dataset and were ultimately determined to be a complete effector complex (∼65% of all particles) (Supplementary Figure S4A) and an effector subcomplex (∼35% of all particles) (Supplementary Figure S4B). Initial analysis of the cryo-EM data showed detailed 2D class averages, with most corresponding to views where the effector complex lies with the long dimension parallel to the ice layer (Supplementary Figure S4D). The data showed the presence of only one particle species. A 3D density map was calculated from 107,620 particles using the complete effector complex negative stain model as a starting point (Supplementary Figure S4E). Post-processing steps were performed, including per-particle CTF estimation and particle polishing. Angular distribution analysis showed a slight preference for one orientation, but there were no preferred orientations (Supplementary Figure S5A). The final volume had an overall resolution of 5.2 Å, as estimated by the gold standard FSC criterion (Supplementary Figure S5B). Local resolution estimates (94) agreed with this resolution assignment (Supplementary Figure S5C). The structure shows a slightly helical stem with an extended domain at the base of the molecule (Figure 5A). The helical stem was clearly of higher resolution than the extended domain, whose resolution was estimated at around 7–9 Å (Supplementary Figure S5C). The cryo-EM data did not show the presence of the complete effector complex, either in the 2D or 3D classification steps, suggesting that the vitrification process partially disassembled the complete complex resulting in the smaller subcomplex.

In order to locate the positions of the SeCsm3 crystal structure in the cryo-EM map, a 6D search using Essens (64) was done. The search identified five separate SeCsm3 locations, in addition to an RRM fold associated at the base of the backbone stem, which was assigned as SeCsm4 (Figures 5B and 5C). A similar search with the crystal structure of SeCsm2 did not yield any solutions, consistent with visual inspection of the volume, and suggests that SeCsm2 is not found in this complex. The SeCsm3 subunits accounted
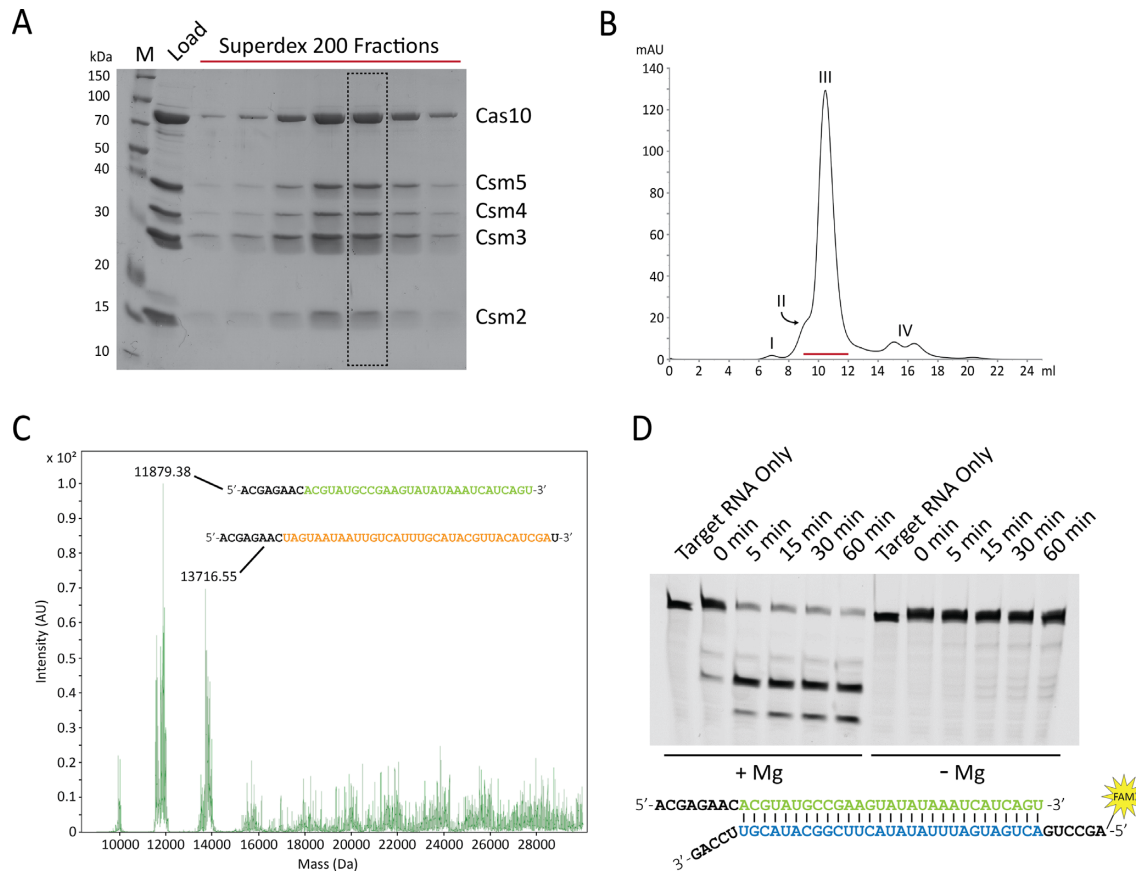
Figure 4. Purification and characterization of the *S. epidermidis* CRISPR effector complex. (**A**) Coomassie blue stained 12% SDS-PAGE of the purified SeCas10–Csm effector complex. The gel shows fractions collected around the peak eluted from the column. All five proteins expected in the complex are present in the peak. The identity of the proteins is shown on the right. The fraction enclosed by the dashed rectangle was used for EM structural studies. M, molecular weight markers. (**B**) Superdex 200 trace of the SeCas10–Csm effector complex purification. Indicated peaks correspond to the column void volume (I), SeCas10–Csm aggregate peak (II), SeCas10–Csm non-aggregate peak (III), and contaminants (IV). The horizontal red line marks the elution fractions loaded into the gel shown in A. (**C**) Mass spectrometry of crRNA isolated by phenol-chloroform extraction from the purified complex used for EM studies. The two crRNA species were determined to be a 37-nucleotide mature crRNA with the sequence from Spacer 1 (green sequence) and a 43-nucleotide mature crRNA with the sequence from Spacer 2 (orange sequence). The corresponding molecular weight and sequence are shown for each peak of interest. (**D**) Time course target RNA cleavage assay of the purified SeCas10–Csm effector complex. Cleavage was monitored using a 5′-FAM-labeled target RNA (blue sequence) complementary to the 37 nt Spacer 1 sequence (green sequence). Cleavage activity was assessed in the presence and absence of magnesium. Reactions were run on a 15% Urea-PAGE in order to visualize cleavage activity.

for most of the density in the stem, whereas the RRM fold assigned to SeCsm4 filled a small region at the base of the stem. No density corresponding to SeCsm5, which is expected to be located at the opposite end of the stem from the SeCsm4 RRM fold (Figure 1C), was identified in the structure. The remaining density at the base of the stem was generally assigned to SeCsm4 and SeCas10, which have been shown in other structures to interact directly (29,95–97) and which can be purified recombinantly as a binary complex (data not shown). The density at either end of the complex is of lower resolution and secondary structure elements are harder to discern. This could potentially be due to slightly different relative orientations between the base and stem. To address this, multi-body refinement of the base and stem as two separate bodies was performed (Supplementary Figure S5D). However, this did not improve the quality of the map.

The SeCsm3 crystal structure fits well at five locations (labeled as SeCsm3.1 to 3.5) (Figures 5B and 5D). A difference density map was generated by subtracting the fitted

crystal structure volumes of the five SeCsm3 subunits and one SeCsm4 RRM domain from the subcomplex volume (Figure 5C). Interestingly, the remaining density forms both a continuous volume traversing the length of the SeCsm3 stem and shows additional density for regions that are disordered in the crystal structure, such as an extension of the incomplete β-hairpin (Figures 5C and 5E). Flexible fitting of the structure using MDFF (66) into the cryo-EM map showed that the crystal structures fit without major changes, aside from slight re-orientations of some of the surface loops (Supplementary Figure S5E). SeCsm3.5 has the worst density, but overall it fits well. This may be due to the presence of two differently sized crRNAs (37 and 43 nucleotides), which generate complexes with four and five SeCsm3, respectively. SeCsm3.5 would therefore correspond to a complex with the longer crRNA, whose occupancy is ∼40% when compared to the relative abundance of the two crRNA species obtained from mass spectrometry (Figure 4C). However, sub-classification of the subcom-
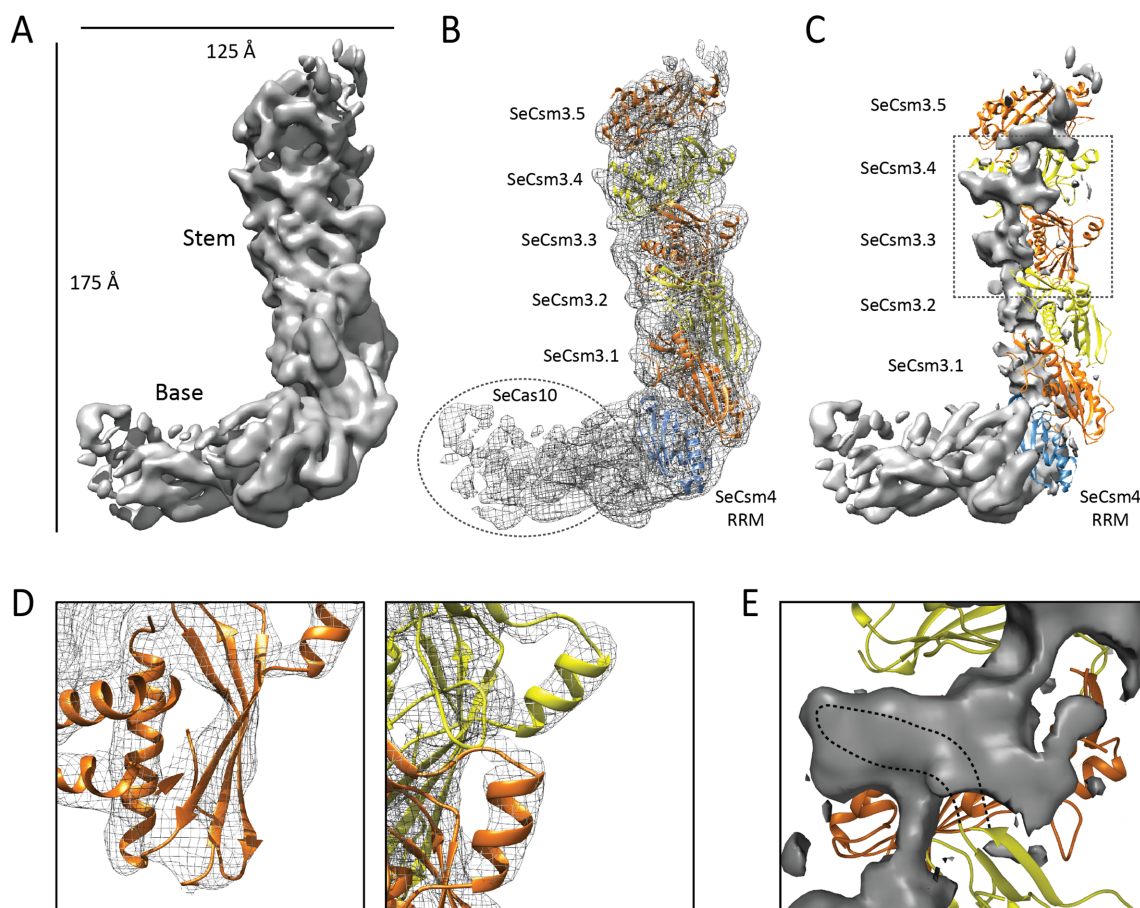
Figure 5. Overview of the cryo-EM structure of the major filament subcomplex. (**A**) A 5.2 Å cryo-EM reconstruction of the SeCas10–Csm subcomplex shows a helical stem region adjoining the non-helical base region. The stem is formed by five SeCsm3 subunits while the base includes SeCsm4 and SeCas10. (**B**) SeCsm3 and a putative SeCsm4 RRM are docked into the cryo-EM density map. Five SeCsm3 (alternately colored orange and yellow) and one SeCsm4 RRM (blue) were confidently placed in the density. Crystal structures of homologous structures or homology models for SeCas10 and SeCsm4 did not fit well into the density at the base. Additional density at the base was putatively assigned as SeCas10 (dashed oval). The SeCsm3 crystal structures and the SeCsm4 RRM form a helical arrangement, with the RRM β-strands facing the outside of the helix and the β-hairpin on the inside of the helix. (**C**) Difference density map between the experimental map and the model of the subcomplex reveals additional density along the SeCsm3 stem, which corresponds to the β-hairpin in SeCsm3 not present in the crystal structure as well as the crRNA traversing the length of the stem. Regions in the dashed box are shown in D and E. (**D**) Close-up views of the SeCsm3 secondary structure elements show good agreement with the cryo-EM density. (**E**) Difference density map shows that disordered regions in the SeCsm3 crystal structure have clear density in the cryo-EM map. These regions of density were putatively assigned as an extension of each SeCsm3 β-hairpin (dashed lines).

plex did not produce different size models; only one complex was observed from the cryo-EM data, suggesting that the five SeCsm3 subunit complex can accommodate both the 37 and 43 nucleotide crRNAs. In addition, the negative stain data only showed one intact complex, again suggesting that the two different crRNAs can be accommodated within the same size complex. Homology models for SeCsm4 and SeCas10 were generated, but did not fit the density adequately and were therefore not included in the final structural model. Nevertheless, the homology models served to show that the density at the base of the stem was of the shape and size to accommodate both SeCas10 and SeCsm4. A partial RRM model for SeCsm4 was built using SeCsm3 as a template. The SeCsm4 RRM lies next to SeCsm3.1 in the stem and makes contacts with it (Figure 5C). Superposition of the MjCsm3 subunit in the MjCsm3-MjCsm4 complex (82) on SeCsm3.1 showed that SeCsm4 is positioned similarly to MjCsm4, confirming the assignment

of the density to the SeCsm4 subunit. Overall, it was possible to assign density for five SeCsm3 subunits, one SeCsm4 subunit and one SeCas10 subunit (Figure 6A). The SeCsm2 and SeCsm5 subunits are not present in the cryo-EM structure.

After assigning density to each of the three protein subunits (including the regions missing from the crystal structures), the continuous region of unaccounted for density was determined to be located within each SeCsm3 subunit (Figure 6A). This extra density was putatively assigned as the crRNA due to its shape and location (Figure 6B). To support the assignment, an electrostatic potential map was calculated for the docked SeCsm3 crystal structures forming the stem. A highly positively charged channel was observed traversing the length of the stem (Figure 6C). When the putative crRNA density was aligned with the electrostatic potential map, the density fit remarkably well into the positively charged channel (Figure 6D), allowing us to
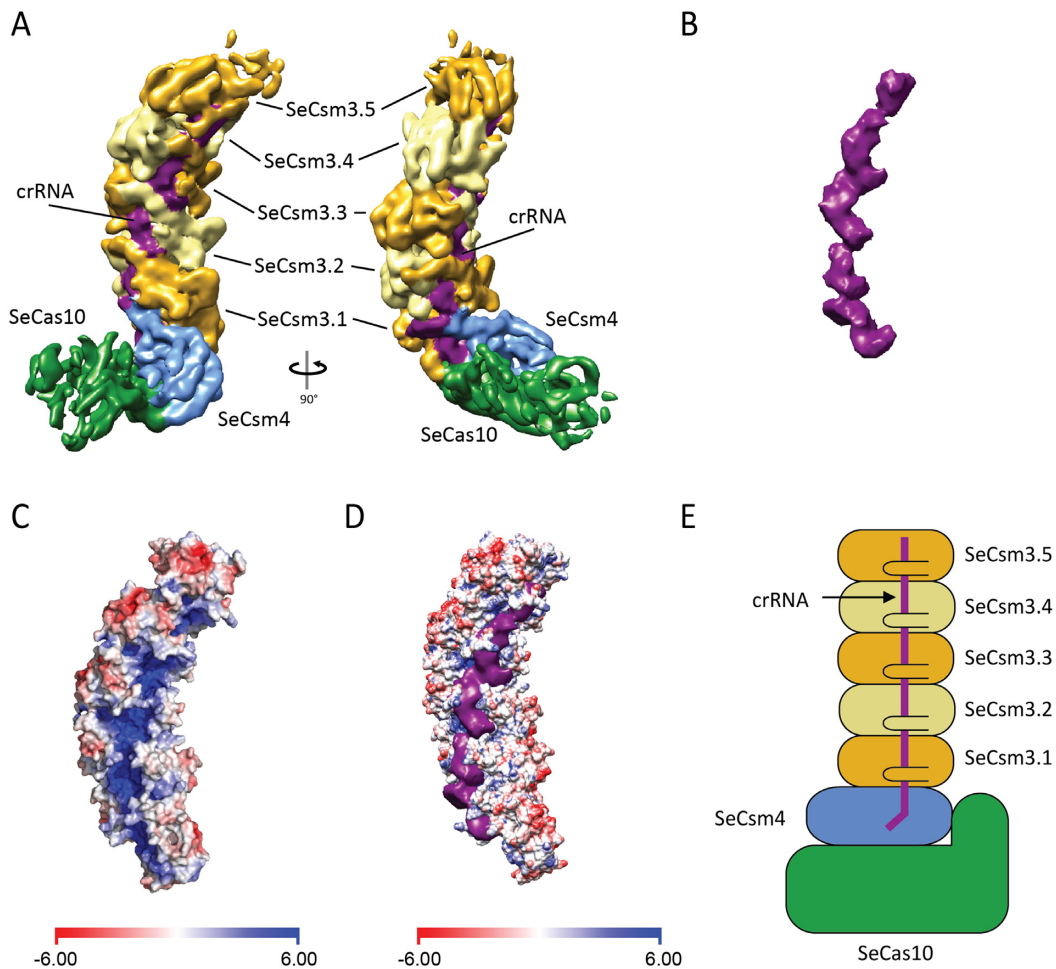
Figure 6. The effector subcomplex shows the crRNA in a conserved, positively charged channel. (**A**) Two orthogonal views of the cryo-EM map colored by subunit identity. Cryo-EM density was assigned to each subunit in the complex after crystal structure docking. Density is assigned for five SeCsm3 (gold and yellow), one SeCsm4 (blue), and one SeCas10 (green), with additional density within each SeCsm3 putatively assigned as the crRNA (purple). (**B**) Putative crRNA density isolated from the SeCas10–Csm subcomplex shown in A. The density follows a continuous path along the length of the stem. (**C**) The SeCsm3 stem is shown in accessible surface representation colored by electrostatic potential calculated using APBS (72). A highly positively charged channel (blue) is seen traversing the length of the stem. (**D**) Putative crRNA density (purple) overlaid on the electrostatic potential map of SeCsm3 stem. The positively charged channel surrounds the negatively charged crRNA, providing further evidence that the density corresponds to the crRNA. The bars at the bottom of C and D map the electrostatic potential values from $-6\ k_BT/e_c$ [red] to $6\ k_BT/e_c$ [blue]. (**E**) Schematic diagram showing the composition of the SeCas10–Csm subcomplex containing a 43 nt crRNA. Csm2 and Csm5 are not present in the subcomplex.

confidently assign this density to the crRNA. Taken together, the identification suggests that the structure corresponds to an effector subcomplex with a stoichiometry of SeCsm3$_5$SeCsm4$_1$SeCas10$_1$crRNA$_1$ (Figures 6A and 6E).

The overall architecture of the subcomplex (Figures 6A and 6E) is similar to that of other CRISPR–Cas complexes and is in agreement with the predicted overall arrangement (Figure 1C). The stem is formed by five SeCsm3 subunits with the SeCas10/SeCsm4 subunits at the base of the complex and almost perpendicular to the stem (Figure 6A). SeCsm4 makes clear contacts with SeCsm3.1, while Se-Cas10 also appears to contact SeCsm3.1. However, the potential interaction of SeCas10 and SeCsm3.1 is not as clear as for SeCsm4 and SeCsm3.1 since it is not possible to unambiguously assign density to SeCsm4 and SeCas10. The crRNA runs along the positively charged groove formed by SeCsm3.1 to 3.5 (Figure 6D) and appears to extend into

the RRM domain of SeCsm4 (Figure 6A). It is not clear whether it continues into SeCas10. SeCsm5 is expected to reside on the other end of the stem next to SeCsm3.5 (Figure 1C), but it is not possible to assign density for it.

The availability of an atomic model for the SeCsm3 stem helps locate many of the residues that have been implicated in different biochemical activities, such as forming the SeCsm3-SeCsm3 interface (25), crRNA binding and maturation (25,26,98), target RNA binding and cleavage (18,25,26), and protein stability (98). Biochemical and functional studies suggest that residues H18 and D32 are involved in target RNA cleavage, residues K52, K54, R56, D100, E120, E124 and E140 affect crRNA maturation, amino acids G182, G183 and G185 have an impact on protein stability, while residues K4 and D179 form part of the SeCsm3-SeCsm3 interface (Supplementary Table S4 and Supplementary Figures S6A and S6B). Many of these

residues are amongst the most highly conserved ones in the SeCsm3 backbone (Supplementary Figure S6C) and in general can be found in or near the highly positively charged channel (Supplementary Figure S6B), suggesting that their role is not confined to SeCsm3, but extends to all members of the Type III-A and possibly Type III-B subtypes. Analysis of both free SeCsm3 and SeCsm3 in the effector complex show that the four sets of residues are spatially separated from one another (Supplementary Figures S6A and S6B). As mentioned above, many highly conserved residues involved in target cleavage (26) cluster around H18 and D32 and form the putative active site in SeCsm3. The SeCsm3-SeCsm3 interface is extensive and the two residues biochemically identified as part of the interface (25) are facing each other in adjacent subunits, suggesting that K4 and D179 form a salt bridge that helps stabilize the stem in the effector complex. Residues involved in crRNA binding and maturation all cluster in the same region (Supplementary Figure S6B). Three basic amino acids, K52, K54 and R56 face the region where the crRNA is located in the effector complex (Supplementary Figure S7A). D100 is near a highly conserved lysine, K122, in the adjacent subunit and it is possible that they form an inter-subunit ion pair. E140 is further away and makes an intra-subunit salt bridge with R37 (Supplementary Figure S6B). Although R37 is not strictly conserved, it is either an arginine or a lysine, suggesting that the salt bridge is conserved and may play a role in stabilizing the tertiary structure of SeCsm3. E124 is also found near the interface between two subunits, but not the same subunits as for E140 and D100. Finally, the three glycines are part of a loop formed by four consecutive glycines and face the crRNA. Overall, it is possible to classify the mutated amino acids into three large groups: those involved in making contacts at the interface between adjacent subunits, those forming part of the putative active site, and those involved in structural stability (Supplementary Figures S6A and S6B).

The crRNA runs along the length of the SeCsm3 stem and travels near many conserved amino acids. Not surprisingly, the active site residues are not found abutting the crRNA density (Supplementary Figure S7A). Instead, the putative active site is found adjacent to the crRNA and facing a cavity that could easily accommodate the target RNA strand (Supplementary Figure S7B).

DISCUSSION

The structural studies presented here describe high-resolution crystal structures of SeCsm2 and SeCsm3 and a medium resolution cryo-EM structure of a Cas10–Csm subcomplex that shows the quaternary architecture of the major filament of the complex. The structure of SeCsm2 shows a small five-helix bundle protein similar to the orthologous Cmr5 found in Type III-B CRISPR systems (78), as well as to other Type III-A Csm2 proteins (76,77). However, the SeCsm2 structure shows important differences between both Type III-A and III-B Csm2 orthologs, such as the presence of a small helix that breaks a longer helix. This architecture shows SeCsm2 has some of the characteristics of other Type III-A Csm2 proteins, but also contains elements found in the Type III-B Cmr5 proteins. There are currently no known functions ascribed specifically to SeCsm2. However, other complexes have described orthologs acting to stabilize the binding of target DNA (99,100) by forming a second stem parallel to the SeCsm3 stem (Figure 1C) (21). Knockout studies have shown that SeCsm2 is one of the components required for crRNA maturation, but mutational studies do not implicate any of the conserved residues in this maturation role (25). Although the crystal structure helps understand the organization of SeCsm2, its absence in the subcomplex described here precludes any further identification of its role in the effector complex. In addition, it is not clear whether the differences observed between different Type III-A Csm2 proteins and their Type III-B orthologs may cause differences in the overall conformation of the effector complex.

In contrast to SeCsm2, the SeCsm3 crystal structure provides important information on the structure and function of the subunit and how it relates to its structural and functional roles within the effector complex. SeCsm3 represents a minimized version of proteins that form the backbone in CRISPR–Cas effector complexes. Unlike other Csm3 proteins and Csm3 orthologs, SeCsm3 is formed primarily by the core RRM fold and a β-hairpin. The SeCsm3 crystal structure shows the position of many important and highly conserved amino acids, as well as the position of the putative active site responsible for RNA target cleavage suggested by mutational studies to be located in the vicinity of D32 (18). Aside from H18 and D32, the putative active site contains many residues normally associated with active sites of proteins involved in nucleic acid cleavage, such as arginines, lysines, and additional acidic residues. However, unlike other ribonucleases, such as ribonucleases H1 and H2 (101), there is no clustering of acidic residues potentially involved in cation binding. Other endoribonucleases in CRISPR systems, such as *Thermus thermophilus* Cse3 (PDB: 2Y8W) (102) and *Methanococcus maripaludis* Cas6b (PDB: 4Z7K) (103), show different active site locations when aligned to SeCsm3, suggesting a different mechanism of cleavage for events associated with crRNA maturation and target RNA degradation. Determining whether SeCsm3 uses a two metal cleavage mechanism or other common ribonuclease catalytic strategy remains to be elucidated and may require structures of complexes with target RNA.

The cryo-EM structure shows a subcomplex formed by three of the five subunits: SeCsm3, SeCsm4 and SeCas10, along with the crRNA (Figure 6A). Neither SeCsm2 nor SeCsm5 could be located in the subcomplex. The absence of these subunits is intriguing as the purified complex showed a single migrating species in size-exclusion chromatography (Figure 4B) with all five protein components present (Figure 4A) as well as mature-length crRNAs (Figure 4C). One possible explanation for the loss of the subunits is the high salt concentration used to mitigate complex aggregation. SeCsm2 and SeCsm5 may also associate weakly in the complex without the presence of the target nucleic acid, enabling EM sample preparation to pull apart these components from the complex. Previous studies have shown that complex formation requires SeCsm3, SeCsm4, and SeCas10, while a subcomplex of these components forms when SeCsm2 or SeCsm5 are knocked out (98). These subcomplexes were shown to include only the intermediate

71 nucleotide crRNA ([98](#)), while crRNA isolated from the cryo-EM complex shows only mature-length crRNAs (Figure [4](#)C). This further supports the notion that the purified complex was intact, active, and containing all subunits, as it was capable of processing intermediate-length crRNA to the correct mature length, but lost some subunits during EM sample preparation. The subcomplex in the cryo-EM structure thus represents a partial complex formed by the loss of some of the subunits after formation, and not a partially assembled precursor complex.

The assignment of different subunits to the density in the cryo-EM map was aided by the crystal structure of SeCsm3 as well as the presence of an RRM fold in SeCsm4. Prior studies had determined that SeCsm3 is required for proper effector complex assembly ([26](#)), with multiple copies of Csm3 forming the backbone of Type III-A complexes ([28,29](#)). The structure confirms this and shows that SeCsm3 does form the backbone of the effector complex and that there are five SeCsm3 copies present (Figures [5](#)B and [5](#)C). Although SeCsm5 is expected to cap one of the ends of the effector complex, the assignment of the density at the end of the stem as SeCsm3.5, rather than SeCsm5, was based on the presence of an RRM fold that fits well into the map, density that accounts for the entirety of SeCsm3, the existence of crRNA density within this subunit, the size of the intact complex obtained by negative stain imaging, the length of the two crRNA contained in the complex, and comparison of this cryo-EM structure to the Type III-A effector complex structure from *Thermus thermophilus* ([29](#)), which shows six subunits in the backbone stem (Supplementary Figure S8A). At the other end of the stem, the location of SeCsm4 was aided by the presence of an RRM fold in the density. Although SeCsm3 has a similar RRM fold, the density did not fit the SeCsm3 subunit well, as there were clear differences in some regions described by either additional or missing density. In support of this assignment, the structure of a Csm3-Csm4 complex from *M. jannaschii* showed a similar interaction between the two proteins ([82](#)). Superposition of MjCsm3 on SeCsm3 in the complex positioned MjCsm4 in the same position as SeCsm4. Finally, the position of SeCas10 was deduced from the size and shape of the unaccounted for density, as well as comparisons of other orthologous Type III-B structures published describing the association of Cas10 (Cmr2) and Csm4 (Cmr3) ([31,95](#)). Given the resolution of the cryo-EM map, the boundaries between the SeCsm3, SeCsm4 and SeCas10 are not well defined, but the general organization is.

The resolution of the cryo-EM map is uneven; the SeCsm3 stem is the most ordered region, while the two ends are less ordered, particularly the SeCas10 region (Supplementary Figure S5C). This observation suggests that SeCas10 is either moving or does not have a tightly held position with respect to the SeCsm3 stem. It is not clear if this difference in relative orientation is due to the absence of SeCsm2, which could play a role in stabilizing the interaction, or simply reflects the flexibility of the complex. It is possible that the conformation of SeCas10 is not rigid and thus is part of its function. During target DNA binding, SeCas10 may undergo a conformational change to help capture and orient target DNA properly in the active site. Recent studies of Type III-A effector complexes from

other species have shown that upon target RNA binding, either minor rearrangements ([35](#)) or dramatic conformational changes ([36](#)) in the complex can occur. It is therefore uncertain how the SeCas10–Csm effector complex would behave under similar functional situations.

The position of the crRNA in the complex follows the expected path, winding its way along the positively charged and highly conserved groove created by the SeCsm3 subunits. Previously, it was found that each SeCsm3 subunit binds six nucleotides ([26](#)). The structure shows that the distance between identical residues in adjacent SeCsm3 subunits is around 25 Å (Supplementary Figure S7B), which is consistent with the distance spanned by six nucleotides in an extended conformation. It appears that the crRNA phosphate backbone is protected by the protein, where the β-hairpin loop perpendicularly transverses the crRNA while the bases of the crRNA are exposed to the solvent, where it could easily interact with complementary target DNA or RNA. Residues that have been implicated in crRNA binding ([25,26,98](#)) are located in the vicinity (Supplementary Figure S7A). The active site location is adjacent to the crRNA and exposed and, as expected, the crRNA does not come close to the putative SeCsm3 active site (Supplementary Figure S7A). It is possible that once the crRNA recognizes and binds its target RNA, the target RNA enters the SeCsm3 active site where it can be cleaved. In the structure, there appears to be space to accommodate the target RNA (Supplementary Figure S7B), although it is possible that the stem has to move in a concerted fashion, as has been observed in other CRISPR–Cas effector complexes ([36,97,100](#)).

Thermus thermophilus harbors various types and subtypes of CRISPR systems, including both Type III-A ([29](#)) and III-B ([96](#)) CRISPR systems. The Type III-A *T. thermophilus* effector complex comprises a full complex with a putative $Cas10_1Csm2_3Csm3_6Csm4_2Csm5_1crRNA_1$ stoichiometry. Aligning the two structures shows similar major filament architectures (Supplementary Figure S8A), even though the SeCas10–Csm structure represents a subcomplex missing SeCsm2 and SeCsm5 (Figures [6](#)A and [6](#)E). Superposition of the density maps from both structures show that the *S. epidermidis* subcomplex has a very similar overall structure to the Type III-A *T. thermophilus* complex (Supplementary Figure S8A), with one major difference being the number of Csm3 and Csm4 subunits. The higher resolution cryo-EM structure of the SeCas10–Csm complex allows unambiguous assignment of the crystal structure of SeCsm3 to the cryo-EM map, where it is clear that there are only five SeCsm3 subunits. In addition, the volume also suggests the presence of only one SeCsm4 subunit. The superposition of the maps shows a similar size volume for the region assigned to one SeCsm4 and one SeCas10 as well as additional density to place one SeCsm5 at the other end of the stem. The Type III-A *T. thermophilus* structure also suggests a path for the SeCsm2 subunits, forming a helical stem running parallel to the SeCsm3 stem. To test this, the *S. epidermidis* negative stain model of the full complex (Supplementary Figure S4A) was aligned with the cryo-EM subcomplex reconstruction (Supplementary Figure S8B). Extra density was observed, as expected, in equivalent locations as for the Csm2 and Csm5

subunits in the *T. thermophilus* structure. Four SeCsm2 crystal structure monomers and one homology-modeled Csm5 using Phyre2 (73) were docked into the negative stain model based on homologous structures (35,96). Including these docked structures gives a putative stoichiometry of $Cas10_1Csm2_4Csm3_5Csm4_1Csm5_1crRNA_1$, which is similar to the stoichiometry found in the *T. thermophilus* effector complex. Interestingly, this stoichiometry seems to accommodate both the 37 and 43 nucleotide mature crRNAs present in the purified effector complex, as attempts at subclassification of the cryo-EM data yielded only one complex stoichiometry. This seems to suggest that shorter-length crRNAs can be found and function in complexes that would normally bind a longer crRNA. Previous research has shown that the abundance of SeCsm3 dictates the maximal crRNA length, with higher amounts resulting in longer crRNA species (26). However, this does not preclude shorter length crRNAs from being utilized within an effector complex of identical stoichiometry. The effector subcomplex shown here is an example of this phenomenon due to the presence of two distinct crRNA lengths and sequences that can be isolated from a single complex stoichiometry. This contrasts with the Type III-A effector complex structures from *T. onnurineus* (35) and *S. thermophilus* (36), which have a $Cas10_1Csm2_2Csm3_3Csm4_1Csm5_1crRNA_1$ stoichiometry and are smaller than the *S. epidermidis* subcomplex, even though the length of the crRNAs, 38 nucleotides (35) and 36 nucleotides (36), are similar to the 37 nucleotide and 43 nucleotide crRNA present in the *in vivo* assembled *S. epidermidis* subcomplex. However, the quaternary architecture of these structures is quite similar to each other and to the *S. epidermidis* subcomplex. This suggests that, while all Type III-A systems are analogous in their overall architecture, the organization and other characteristics within each effector complex, such as crRNA length and protein stoichiometry, differ in important ways, presenting an example for how the CRISPR machinery has been adapted differently within each organism.

The structure of the Type III-B *T. thermophilus* effector complex is also known (96) and has a $Cmr1_1Cmr2_1Cmr3_1Cmr4_4Cmr5_3Cmr6_1crRNA_1$ stoichiometry (the corresponding orthologs are Cmr1:Csm3, Cmr2:Cas10, Cmr3:Csm4, Cmr4:Csm3, Cmr5:Csm2, Cmr6:Csm5). Type III-B systems are an evolutionary offshoot from the Type III-A systems, where a Cas7 family gene duplication occurred. This led to the incorporation of three Cas7 family proteins (Cmr1, Cmr4, and Cmr6) into Type III-B effector complexes, whereas only two Cas7 family proteins (Csm3 and Csm5) are found in Type III-A effector complexes (13). The two subtypes have similar effector complex stoichiometries when compared by gene family, $Cas10_1Cas5_1Cas7_6Cas11_3crRNA_1$, where the Cas5 family includes Csm4/Cmr3 and the Cas11 family includes Csm2/Cmr5. However, when the structures are compared, it is clear that they are very different in overall architecture (Supplementary Figure S8C). While the *S. epidermidis* effector complex has a boot-shaped overall architecture, the Type III-B *T. thermophilus* complex is much more rod-shaped and without a marked protrusion. Both complexes have in common the presence of a helical

stem and a similar subunit arrangement, but overall the two complexes are quite different.

The same is true when the *S. epidermidis* complex is compared to Type I structures, such as the Type I-E *Escherichia coli* CASCADE complex (104). The most dramatic difference between the two structures is the overall organization of the backbone subunits. The Type I-E *E. coli* structure has an exaggerated helical arrangement, generating a long, extended quaternary structure (104). A similar architecture to the *E. coli* complex is seen in the Type I-F effector complex from *Pseudomonas aeruginosa* (91). While these different Type I complexes show strikingly similar overall architectures, a Type I-F variant from *Shewanella putrefaciens* details a drastically different effector complex structure (92). This effector complex looks more structurally similar to Type III complexes with their relatively straight crRNA topology. However, there is still an underlying helical arrangement of the backbone subunits. The main commonality amongst all these complexes is the presence of a helical arrangement of subunits to bind and present the crRNA. Overall, the diversity in subunit composition of the different types and subtypes is reflected in a wide array of effector complex structures, with the major commonality occurring in the arrangement of the subunits that bind the crRNA.

CONCLUSIONS

The structures presented help to understand better the overall architecture of the Type III-A Cas10–Csm effector complex. The structures help to identify the location and organization of functionally relevant residues identified by biochemical experiments, to understand the way that the Cas10–Csm complex binds the crRNA, and to detail the organization of the RNA cleavage site. This new structural information provides a structural foundation to determine the molecular mechanisms responsible for the function of the Type III-A CRISPR system.

DATA AVAILABILITY

Atomic coordinates and structure factors for the reported crystal structures have been deposited with the Protein Data Bank under accession numbers 6NBT and 6NBU. The electron density map for the cryo-EM structure has been deposited in the EMDataBank under accession number EMD-0442.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Horvath,P. and Barrangou,R. (2010) CRISPR/Cas, the immune system of bacteria and archaea. *Science*, **327**, 167–170.
2. Karginov,F.V. and Hannon,G.J. (2010) The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol. Cell*, **37**, 7–19.
3. Marraffini,L.A. (2015) CRISPR–Cas immunity in prokaryotes. *Nature*, **526**, 55–61.
4. Marraffini,L.A. and Sontheimer,E.J. (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.*, **11**, 181–190.
5. van der Oost,J., Westra,E.R., Jackson,R.N. and Wiedenheft,B. (2014) Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **12**, 479–492.
6. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
7. Marraffini,L.A. and Sontheimer,E.J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845.
8. Jansen,R., van Embden,J.D.A., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.
9. Bolotin,A., Quinquis,B., Sorokin,A. and Ehrlich,S.D. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, **151**, 2551–2561.
10. Pourcel,C., Salvignol,G. and Vergnaud,G. (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, **151**, 653–663.
11. Koonin,E.V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
12. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
13. Makarova,K.S., Haft,D.H., Barrangou,R., Brouns,S.J., Charpentier,E., Horvath,P., Moineau,S., Mojica,F.J., Wolf,Y.I., Yakunin,A.F. *et al.* (2011) Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **9**, 467–477.
14. Ramia,N.F., Spilman,M., Tang,L., Shao,Y., Elmore,J., Hale,C., Cocozaki,A., Bhattacharya,N., Terns,R.M., Terns,M.P. *et al.* (2014) Essential structural and functional roles of the Cmr4 subunit in RNA cleavage by the Cmr CRISPR–Cas complex. *Cell Rep.*, **9**, 1610–1617.
15. Rouillon,C., Zhou,M., Zhang,J., Politis,A., Beilsten-Edmands,V., Cannone,G., Graham,S., Robinson,C.V., Spagnolo,L. and White,M.F. (2013) Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol. Cell*, **52**, 124–134.
16. Zhang,J., Graham,S., Tello,A., Liu,H. and White,M.F. (2016) Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res.*, **44**, 1789–1799.
17. Zhang,J., Rouillon,C., Kerou,M., Reeks,J., Brugger,K., Graham,S., Reimann,J., Cannone,G., Liu,H., Albers,S.V. *et al.* (2012) Structure and mechanism of the CMR complex for CRISPR-mediated antiviral immunity. *Mol. Cell*, **45**, 303–313.
18. Samai,P., Pyenson,N., Jiang,W., Goldberg,G.W., Hatoum-Aslan,A. and Marraffini,L.A. (2015) Co-transcriptional DNA and RNA cleavage during Type III CRISPR–Cas immunity. *Cell*, **161**, 1164–1174.
19. Elmore,J.R., Sheppard,N.F., Ramia,N., Deighan,T., Li,H., Terns,R.M. and Terns,M.P. (2016) Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev.*, **30**, 447–459.
20. Estrella,M.A., Kuo,F.T. and Bailey,S. (2016) RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes Dev.*, **30**, 460–470.
21. Kazlauskiene,M., Tamulaitis,G., Kostiuk,G., Venclovas,C. and Siksnys,V. (2016) Spatiotemporal control of Type III-A CRISPR–Cas Immunity: Coupling DNA degradation with the target RNA recognition. *Mol. Cell*, **62**, 295–306.
22. Grissa,I., Vergnaud,G. and Pourcel,C. (2007) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics*, **8**, 172.
23. Carte,J., Wang,R., Li,H., Terns,R.M. and Terns,M.P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.*, **22**, 3489–3496.
24. Carte,J., Pfister,N.T., Compton,M.M., Terns,R.M. and Terns,M.P. (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA*, **16**, 2181–2188.
25. Walker,F.C., Chou-Zheng,L., Dunkle,J.A. and Hatoum-Aslan,A. (2017) Molecular determinants for CRISPR RNA maturation in the Cas10–Csm complex and roles for non-Cas nucleases. *Nucleic Acids Res.*, **45**, 2112–2123.
26. Hatoum-Aslan,A., Samai,P., Maniv,I., Jiang,W. and Marraffini,L.A. (2013) A ruler protein in a complex for antiviral defense determines the length of small interfering CRISPR RNAs. *J. Biol. Chem.*, **288**, 27888–27897.
27. Staals,R.H.J., Agari,Y., Maki-Yonekura,S., Zhu,Y., Taylor,D.W., van Duijn,E., Barendregt,A., Vlot,M., Koehorst,J.J., 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.
28. Tamulaitis,G., Venclovas,C. and Siksnys,V. (2017) Type III CRISPR–Cas Immunity: Major differences brushed aside. *Trends Microbiol.*, **25**, 49–61.
29. Staals,R.H., Zhu,Y., Taylor,D.W., Kornfeld,J.E., Sharma,K., Barendregt,A., Koehorst,J.J., 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.
30. 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.
31. Jung,T.Y., An,Y., Park,K.H., Lee,M.H., Oh,B.H. and Woo,E. (2015) Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure*, **23**, 782–790.
32. Ramia,N.F., Tang,L., Cocozaki,A.I. and Li,H. (2014) Staphylococcus epidermidis Csm1 is a 3′-5′ exonuclease. *Nucleic Acids Res.*, **42**, 1129–1138.

33. Jiang,W., Samai,P. and Marraffini,L.A. (2016) Degradation of phage transcripts by CRISPR-Associated RNases enables Type III CRISPR–Cas immunity. *Cell*, **164**, 710–721.

34. Niewoehner,O., Garcia-Doval,C., Rostol,J.T., Berk,C., Schwede,F., Bigler,L., Hall,J., Marraffini,L.A. and Jinek,M. (2017) Type III CRISPR–Cas systems produce cyclic oligoadenylate second messengers. *Nature*, **548**, 543–548.

35. Jia,N., Mo,C.Y., Wang,C., Eng,E.T., Marraffini,L.A. and Patel,D.J. (2019) Type III-A CRISPR–Cas Csm complexes: assembly, periodic RNA cleavage, DNase activity regulation, and autoimmunity. *Mol. Cell*, **73**, 264–277.

36. You,L., Ma,J., Wang,J., Artamonova,D., Wang,M., Liu,L., Xiang,H., Severinov,K., Zhang,X. and Wang,Y. (2019) Structure studies of the CRISPR-Csm complex reveal mechanism of Co-transcriptional interference. *Cell*, **176**, 239–253.

37. Otto,M. (2009) Staphylococcus epidermidis–the 'accidental' pathogen. *Nat. Rev. Microbiol.*, **7**, 555–567.

38. Otto,M. (2012) Molecular basis of Staphylococcus epidermidis infections. *Semin. Immunopathol.*, **34**, 201–214.

39. von Eiff,C., Peters,G. and Heilmann,C. (2002) Pathogenesis of infections due to coagulase-negative staphylococci. *Lancet Infect. Dis.*, **2**, 677–685.

40. Stols,L., Zhou,M., Eschenfeldt,W.H., Millard,C.S., Abdullah,J., Collart,F.R., Kim,Y. and Donnelly,M.I. (2007) New vectors for co-expression of proteins: structure of Bacillus subtilis ScoAB obtained by high-throughput protocols. *Protein Expr. Purif.*, **53**, 396–403.

41. Burgess,R.R. (1991) Use of polyethyleneimine in purification of DNA-Binding proteins. *Method Enzymol.*, **208**, 3–10.

42. Kabsch,W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.

43. Evans,P.R. and Murshudov,G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.*, **69**, 1204–1214.

44. Miller,R., DeTitta,G.T., Jones,R., Langs,D.A., Weeks,C.M. and Hauptman,H.A. (1993) On the application of the minimal principle to solve unknown structures. *Science*, **259**, 1430–1433.

45. de la Fortelle,E. and Bricogne,G. (1997) Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.*, **276**, 472–494.

46. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

47. Blanc,E., Roversi,P., Vonrhein,C., Flensburg,C., Lea,S.M. and Bricogne,G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2210–2221.

48. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J., Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

49. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Echols,N., Headd,J.J., Hung,L.W., Jain,S., Kapral,G.J., Grosse Kunstleve,R.W. *et al.* (2011) The Phenix software for automated determination of macromolecular structures. *Methods*, **55**, 94–106.

50. Murshudov,G.N., Vagin,A.A. and Dodson,E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 240–255.

51. Chen,V.B., Arendall,W.B. 3rd, Headd,J.J., Keedy,D.A., Immormino,R.M., Kapral,G.J., Murray,L.W., Richardson,J.S. and Richardson,D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 12–21.

52. Pannu,N.S., Waterreus,W.J., Skubak,P., Sikharulidze,I., Abrahams,J.P. and de Graaff,R.A. (2011) Recent advances in the CRANK software suite for experimental phasing. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 331–337.

53. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

54. Cowtan,K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.*, **62**, 1002–1011.

55. Suloway,C., Pulokas,J., Fellmann,D., Cheng,A., Guerra,F., Quispe,J., Stagg,S., Potter,C.S. and Carragher,B. (2005) Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.*, **151**, 41–60.

56. Kimanius,D., Forsberg,B.O., Scheres,S.H. and Lindahl,E. (2016) Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife*, **5**, e18722.

57. Scheres,S.H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.*, **180**, 519–530.

58. Zivanov,J., Nakane,T., Forsberg,B.O., Kimanius,D., Hagen,W.J., Lindahl,E. and Scheres,S.H. (2018) New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife*, **7**, e42166.

59. Nakane,T., Kimanius,D., Lindahl,E. and Scheres,S.H. (2018) Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife*, **7**, e36861.

60. Zheng,S.Q., Palovcak,E., Armache,J.P., Verba,K.A., Cheng,Y. and Agard,D.A. (2017) MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods*, **14**, 331–332.

61. Grant,T. and Grigorieff,N. (2015) Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.*, **192**, 204–208.

62. Rohou,A. and Grigorieff,N. (2015) CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.*, **192**, 216–221.

63. Rosenthal,P.B. and Henderson,R. (2003) Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.*, **333**, 721–745.

64. Kleywegt,G.J. and Jones,T.A. (1997) Template convolution to enhance or detect structural features in macromolecular electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.*, **53**, 179–185.

65. Brown,A., Long,F., Nicholls,R.A., Toots,J., Emsley,P. and Murshudov,G. (2015) Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. *Acta Crystallogr. D Biol. Crystallogr.*, **71**, 136–153.

66. Trabuco,L.G., Villa,E., Mitra,K., Frank,J. and Schulten,K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure*, **16**, 673–683.

67. Phillips,J.C., Braun,R., Wang,W., Gumbart,J., Tajkhorshid,E., Villa,E., Chipot,C., Skeel,R.D., Kale,L. and Schulten,K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.

68. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

69. Pymol. Version 1.7. Schrödinger, LLC.

70. Ashkenazy,H., Erez,E., Martz,E., Pupko,T. and Ben-Tal,N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529–W533.

71. Landau,M., Mayrose,I., Rosenberg,Y., Glaser,F., Martz,E., Pupko,T. and Ben-Tal,N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–W302.

72. Baker,N.A., Sept,D., Joseph,S., Holst,M.J. and McCammon,J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 10037–10041.

73. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.

74. Tatusova,T., Ciufo,S., Fedorov,B., O'Neill,K. and Tolstoy,I. (2015) RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res*, **43**, 3872.

75. Takeshita,D., Sato,M., Inanaga,H. and Numata,T. (2019) Crystal structures of Csm2 and Csm3 in the Type III-A CRISPR–Cas effector complex. *J. Mol. Biol.*, j.jmb.2019.01.009.

76. Gallo,G., Augusto,G., Rangel,G., Zelanis,A., Mori,M.A., Campos,C.B. and Wurtele,M. (2016) Structural basis for dimer formation of the CRISPR-associated protein Csm2 of Thermotoga maritima. *FEBS J.*, **283**, 694–703.

77. Venclovas,C. (2016) Structure of Csm2 elucidates the relationship between small subunits of CRISPR–Cas effector complexes. *FEBS Lett.*, **590**, 1521–1529.

78. Sakamoto,K., Agari,Y., Agari,K., Yokoyama,S., Kuramitsu,S. and Shinkai,A. (2010) X-ray crystal structure of a CRISPR-associated RAMP superfamily protein, Cmr5, from Thermus thermophilus HB8 (vol 75, pg 528, 2009). *Proteins*, **78**, 1611–1611.

79. Zhao,Y.Q., Wang,J.J., Sun,Q., Dou,C., Gu,Y.J., Nie,C.L., Zhu,X.F., Wei,Y.Q. and Cheng,W. (2018) Structural insights into the CRISPR–Cas-associated ribonuclease activity of Staphylococcus epidermidis Csm3 and Csm6. *Sci. Bull.*, **63**, 691–699.

80. Tamulaitis,G., Kazlauskiene,M., Manakova,E., Venclovas,C., Nwokeoji,A.O., Dickman,M.J., 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.

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

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

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

84. 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.

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

86. Hrle,A., Maier,L.K., Sharma,K., Ebert,J., Basquin,C., Urlaub,H., Marchfelder,A. and Conti,E. (2014) Structural analyses of the CRISPR protein Csc2 reveal the RNA-binding interface of the type I-D Cas7 family. *RNA Biol.*, **11**, 1072–1082.

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

88. Lintner,N.G., Kerou,M., Brumfield,S.K., Graham,S., Liu,H., Naismith,J.H., Sdano,M., Peng,N., She,Q., Copie,V. *et al.* (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol .Chem.*, **286**, 21643–21656.

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

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

91. Guo,T.W., Bartesaghi,A., Yang,H., Falconieri,V., Rao,P., Merk,A., Eng,E.T., Raczkowski,A.M., Fox,T., Earl,L.A. *et al.* (2017) Cryo-EM structures reveal mechanism and inhibition of DNA targeting by a CRISPR–Cas surveillance complex. *Cell*, **171**, 414–426.

92. Pausch,P., Muller-Esparza,H., Gleditzsch,D., Altegoer,F., Randau,L. and Bange,G. (2017) Structural variation of Type I-F CRISPR RNA Guided DNA surveillance. *Mol. Cell*, **67**, 622–632.

93. Nurmohamed,S., Vaidialingam,B., Callaghan,A.J. and Luisi,B.F. (2009) Crystal structure of Escherichia coli polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly. *J. Mol. Biol.*, **389**, 17–33.

94. Kucukelbir,A., Sigworth,F.J. and Tagare,H.D. (2014) Quantifying the local resolution of cryo-EM density maps. *Nat. Methods*, **11**, 63–65.

95. Shao,Y., Cocozaki,A.I., Ramia,N.F., Terns,R.M., Terns,M.P. and Li,H. (2013) Structure of the Cmr2-Cmr3 subcomplex of the Cmr RNA silencing complex. *Structure*, **21**, 376–384.

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

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

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

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

100. 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.

101. Tadokoro,T. and Kanaya,S. (2009) Ribonuclease H: molecular diversities, substrate binding domains, and catalytic mechanism of the prokaryotic enzymes. *FEBS J.*, **276**, 1482–1493.

102. Sashital,D.G., Jinek,M. and Doudna,J.A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.*, **18**, 680–687.

103. Shao,Y., Richter,H., Sun,S., Sharma,K., Urlaub,H., Randau,L. and Li,H. (2016) A non-stem-loop CRISPR RNA is processed by dual binding Cas6. *Structure*, **24**, 547–554.

104. Mulepati,S., Heroux,A. and Bailey,S. (2014) Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*, **345**, 1479–1484.