Structural and Functional Characterization of an Archaeal Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Complex for Antiviral Defense (CASCADE) *

Received for publication, March 20, 2011, and in revised form, April 8, 2011 Published, JBC Papers in Press, April 20, 2011, DOI 10.1074/jbc.M111.238485

Nathanael G. Lintner 1,2, Melina Kerou 1, Susan K. Brumfield 3, Shirley Graham 4, Huanting Liu 4, James H. Naismith 4, Matthew Sdano 5, Nan Peng 5, Qunxin She 5, Valerie Copie 6, Mark J. Young 4, Malcolm F. White 4, and C. Martin Lawrence 5

From the 4Thermal Biology Institute, Departments of 6Chemistry and Biochemistry and 7Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana 59717, the 6Biomedical Sciences Research Complex, University of St. Andrews, North Haugh, St. Andrews, Fife KY 16 9ST, United Kingdom, and the **Department of Biology, Archaea Centre, University of Copenhagen, DK-2200 Copenhagen N, Denmark

In response to viral infection, many prokaryotes incorporate fragments of virus-derived DNA into loci called clustered regularly interspaced short palindromic repeats (CRISPRs). These loci are then transcribed, and the processed CRISPR transcripts are used to target invading viral DNA and RNA. The Escherichia coli “CRISPR-associated complex for anti-viral defense” (CASCADE) is central in targeting invading DNA. Here we report the structural and functional characterization of an archaeal CASCADE (aCASCADE) from Sulfolobus solfataricus. Tagged Csa2 (Cas7) expressed in S. solfataricus co-purifies with Cas5a-, Cas6-, Csa5-, and Cas6-processed CRISPR-RNA (crRNA). Csa2, the dominant protein in aCASCADE, forms a stable complex with Cas5a. Transmission electron microscopy reveals a helical complex of variable length, perhaps due to substoichiometric amounts of other CASCADE components. A recombinant Csa2-Cas5a complex is sufficient to bind crRNA and complementary ssDNA. The structure of Csa2 reveals a crescent-shaped structure unexpectedly composed of a modified RNA-recognition motif and two additional domains present as insertions in the RNA-recognition motif. Conserved residues indicate potential crRNA- and target DNA-binding sites, and the H160A variant shows significantly reduced affinity for crRNA. We propose a general subunit architecture for CASCADE in other bacteria and Archaea.

Adaptive and heritable immune systems have recently been recognized in Bacteria and Archaea. These systems consist of loci termed clustered regularly interspaced short palindromic repeats (CRISPRs) 6 and their associated protein encoding (cas) genes (1–5). CRISPR loci consist of a variable number of short direct repeats (21–47 nt) separated by short fragments (30–48 nt) of variable, invader-derived sequences that are incorporated in response to viral infection or other invading nucleic acid (5–7). CRISPR transcripts are processed into small RNAs (crRNA) (8, 9) in which the variable sequences, also known as spacers, serve as guide sequences for the recognition and neutralization of invading DNA (7, 8, 10–12) or RNA (9).

Co-occurrence patterns for cas genes within genomes and gene clusters suggest the Cas machinery takes on several different forms, referred to as subtypes (13) or CRISPR-associated systems (14). These systems share a set of core cas genes, which include cas 1–6, and eight sets of subtype specific genes (cse, csa, cst, csm, csn, csh, csu, and csd). A given CRISPR/Cas system will thus encode several of the core Cas proteins plus at least one of these eight subtypes. However, distant relationships across subtypes for several gene families have been recognized, such that some of the cas subtype gene families can be unified into superfamilies loosely based on clusters of orthologous groups (14). In addition, many CRISPR/Cas systems include a third cluster of genes that belong to the repeat associated mysterious protein superfamily and are named cmr1–6 (13).

Several activities of the CRISPR-associated protein machinery are now recognized. One function is the acquisition and insertion of new spacers into the CRISPR loci. Whereas little is known about this process, it is thought to involve Cas1 and Cas2. A second function is processing the CRISPR transcript to

6 The abbreviations used are: CRISPR, clustered regularly interspaced short palindromic (prokaryotic) repeats; CASCADE, CRISPR-associated complex for antiviral defense; aCASCADE, archaeal (Apemn) CASCADE; Cas, CRISPR-associated; crRNA, CRISPR-RNA; RRM, RNA-recognition motif; Csa, CRISPR-subtype Apemn; Cse, CRISPR-subtype E. coli; Csy, CRISPR-subtype Y pestis; Cst, CRISPR-subtype Tinea; Csd, CRISPR-subtype Dvulg; Csh, CRISPR-subtype Hmani; PAM, protospacer-adjacent motif; TEM, transmission electron microscopy; nt, nucleotide(s); ds, double-stranded; ss, single-stranded; NiNTA, nickel-nitrilotriacetic acid.
produce crRNA, and several endoribonucleases that process crRNA have now been identified, including *Pyrococcus furiosus* Cas6. A third function is the use of crRNA to guide neutralization of non-host RNA with the RNAi-like activity of the CMR complex, which has been demonstrated in *P. furiosus*. Interestingly, however, Cas6 and the CMR complex lack apparent homology to eukaryotic RNAi protein machinery in both primary sequence and three-dimensional structure (9, 15–17).

Last, and most relevant, a fourth function of CRISPR/Cas is the use of crRNA to guide neutralization of invading DNA (8). This activity is mediated by the CRISPR-associated complex for antiviral defense (CASCADE) (8).

The characterization of *Escherichia coli* CASCADE revealed a complex composed of the Cse1–4 subtype proteins and Cas5e. Collectively, these 5 proteins are also known as CasA–CasE. Together, they form a 405-kDa complex with crRNA that allows recognition of single- and double-stranded target DNAs complementary to the bound crRNA (8, 12). Following recognition by CASCADE, the nuclease/helicase activity of Cas3 is then recruited to degrade the invading DNA (18). Each component of *E. coli* CASCADE, along with Cas3, is required for viral resistance in vivo (8).

Transmission electron microscopy, small angle x-ray scattering, and non-covalent mass spectrometry reveal that *E. coli* CASCADE has an unusual quaternary structure, consisting of six copies of CasC(Cse4), which form the core or backbone of the CASCADE complex (8, 12). This core is complemented by single copies of CasA, CasD(Cas5e) and CasE(Cse3), with CasE exhibiting an endoribonuclease activity that specifically cleaves the CRISPR transcript (8, 19). In addition, CasB is also present with two copies per complex. The enzymatic activity of CasB is similar to *P. furiosus* Cas6 (8, 16, 20) and *Pseudomonas aeruginosa* Csy4 (21).

Of the *E. coli* CASCADE protein components, only the CasC(Cse4) and CasD(Cas5e) protein families are observed in the other Cas subtypes or Cas systems with recognizable CASCADE components (8, 13, 14), where their distant homologues are known under a variety of different names, and the nomenclature can be quite confusing. However, increased recognition of the central role of these protein families in CRISPR/Cas is leading to the adoption of unifying nomenclature, and the superfamilies that contain CasD and CasC are now increasingly referred to as Cas5 and Cas7, respectively. Interestingly, these superfamilies display conserved gene synteny, and thus, are likely to represent evolutionarily conserved elements that lie at the core of CASCADE structure and function.

*Sulfolobus* species represent an important model system for Archaea in general, and the Crenarchaeota in particular, and have been quite useful for investigating the function of CRISPR/Cas (15, 22–31). The *Sulfolobus solfataricus* P2 genome encodes six CRISPR loci designated CRISPRs A–F (32), and multiple paralogous of *cas*, *csa* (CRISPR subtype Apern), and *cmr* gene products, including three paralogs of the CASCADE-like *cas5/cas7* gene cassette. In *S. solfataricus* the Cas5 and Cas7 proteins have been generally referred to as Cas5a and Csa2, respectively. Here we report the identification of an archaeal CASCADE (aCASCADE), and confirm the central roles of archaeal Csa2 and Cas5a in this complex. We also report the functional characterization of Csa2 and Cas5a in the recognition of crRNA and invading DNA by aCASCADE, and the crystallographic structure of Csa2, the first structure for a member of the Cas7 superfamily.

**EXPERIMENTAL PROCEDURES**

*Protein Expression in S. solfataricus*—Tandem (strept. His<sub>6</sub>) tagged Sso1442 was expressed in a *S. solfataricus* PH1–16 uracil auxotroph strain (33) using Sulfolobus expression vector pSeSD1, which was developed in the laboratory. The pSeSD1 vector was derived from the *Sulfolobus-E. coli* shuttle plasmid vector pHZ2 reported previously (34), in which the expression of a target gene was under control of the *S. solfataricus* araS (arabinose-binding protein gene) promoter (35). Cells were grown in Brock’s minimal medium (36) supplemented with 0.1% tryptone and 0.2% sucrose or 0.2% arabinose for protein production. Additional details are provided under supplemental “Methods.”

Isolation and Characterization of aCASCADE Components—Sulfolobus cells (above) were disrupted by passage through a French press and clarified by centrifugation at 47,000 × *g* for 30 min. aCASCADE components were purified from the supernatant using Strep-Tactin resin (Novagen), Ni-NTA resin (Qiagen), and gel-filtration chromatography with a Superose-6 column. Detailed expression and purification protocols are provided under supplemental “Methods.”

Proteins were identified using SDS-PAGE with colloidal Coomassie staining (37), in-gel reduction, alkylation, and trypsin digestion followed by LC-MS/MS according to Ref. 38 and by in-solution trypsin digestion followed by LC-MS/MS according to Ref. 39. Nucleic acids were isolated from samples using basic phenol-chloroform extraction. TRizol-LS (Sigma) was used for the RNA-specific extractions. Small RNA was gel purified, adapter-ligated, reverse transcribed, polyadenylated, cloned, and sequenced according to Ref. 9.

**Ribonuclease Protection Assays**—200 μg of aCASCADE (total protein by Bradford assay (40)) was incubated without enzyme (lane 2) or with 1 μg of RNase A at room temperature for 1, 2, 18, or 24 h (lanes 3–6) followed by TRizol extraction and TBE-urea-PAGE analysis (supplemental Fig. 1C).

**Expression and Purification of Recombinant Proteins**—The genes encoding *S. solfataricus* P2 Cas6 (sso2004), Cas5a (sso1441), and Csa2 (sso1442) were amplified by PCR from genomic DNA. sso2004 was cloned into pDEST14 using the Gateway recombination cloning (41), and sso1441 and sso1442 were cloned into a modified pRSFDuet vector, pRSFDuetHISTEV.

Cas6 and the Csa2-Cas5a complex were expressed in *E. coli* BL21(DE3) cells by induction at *A*<sub>600</sub> = 0.6 with 0.4 mM isopropyl-1-thio-β-D-galactopyranoside and overnight incubation at 25 °C. Cells were harvested by centrifugation and disrupted by sonication. Cas6 was purified using nickel-chelate affinity and gel-filtration chromatography. The Csa2-Cas5a complex was purified by successive chromatography steps on 5-ml HisTrap HP, Superdex 200 10/300, and HiTrap heparin columns (GE Healthcare).

---

1. Q. She, unpublished data.
For structural studies, sso1442 was cloned with a minimal N-terminal non-cleavable His6 tag into pDest14 using site-specific recombination (Gateway, Invitrogen) and a nested PCR protocol as described previously (42). Sso1442 was expressed in BL21(DE3)-pLysS E. coli using ZYP-5025 autoinduction medium (43). Cells were harvested by centrifugation and disrupted by passage through a French Press. The protein was purified by metal-chelate affinity chromatography and gel-filtration chromatography on a Superdex S-75 column (GE Healthcare) equilibrated with 10 mM Tris (pH 8.0) and 200 mM NaCl. Detailed expression and purification protocols are provided under supplemental “Methods.” The oligomeric state and dissociation constant of recombinant Csa2 was determined using sedimentation velocity analytical ultracentrifugation (Center for Analytical Ultracentrifugation of Macromolecular Assemblies, University of Texas Health Sciences Center) at concentrations of 0.21, 0.13, and 0.08 mg/ml in a buffer consisting of 50 mM Na2HPO4, 100 mM NaCl and monitored at a wavelength of 230 nm. Data were analyzed using the UltraScan suite (49) and REFMAC5 (50, 51) using the native data set. Residue numbers in the model are consistent with the native protein structure. Residues 17–26, 165–172, 233–240, 303–305, and 321 of chain A, 15–22, 149–179, 234–242, and 302–321 of chain B, 16–25, 164–176, 233–241, 302–303, and 320–321 of chain C, 16–26, 149–177, 234–241, 301–303, and 321 of chain D were not modeled due to lack of interpretable electron density. TLS parameters (52) were included in the refinement, wherein the four Sso1442 chains were divided into 7 TLS groups with an average R-factor of 0.5 (4.7). Oligonucleotides—Synthetic DNA and RNA oligonucleotides for functional studies of the Csa2-Cas5a complex were purchased from Integrated DNA Technologies (IDT) and Eurofins MWG Operon, respectively. These substrates were purified on denaturing polyacrylamide/urea gels and 5′-end-labeled with T4 PNK (Fermentas) and [γ-32P]ATP (4500 Ci/mmol, MP Biomedicals). The sequences of the oligonucleotides used are provided under supplemental Table S3, where the CRISPR repeat-derived sequences are shown in bold.

Nuclease Assays—To detect cleavage of the crRNA by Cas6, nuclease reactions (total volume 10 μl) were carried out in reaction buffer (20 mM MES, pH 6.0, 100 mM potassium glutamate, 0.5 mM DTT, 5 mM EDTA) at 45 °C for 15–30 min with 1 μm recombinant Cas6. The reactions were treated with 0.1 mg of Proteinase K for 15 min at 37 °C and the products were separated on 20% polyacrylamide, 7 M urea gels, visualized by phosphorimaging.

Gel Electrophoretic Mobility Shift Assays—The binding of DNA or RNA substrates by the Csa2-Cas5 complex was tested by gel electrophoretic mobility shift. Binding buffer (20 mM MES, pH 6.0, 50 mM potassium glutamate, 0.5 mM DTT, 5 mM EDTA, 5% glycerol) containing 100 mM radiolabeled substrate (crRNA or target DNA as indicated) was preincubated at 55 °C and reactions were initiated by the addition of the recombinant Csa2-Cas5a complex or Csa2 alone. After a 10-min incubation, at 55 °C, complexes were separated on native 10% polyacrylamide gels and visualized by phosphorimaging.

Transmission Electron Microscopy—Samples were negatively stained with 2% uranyl acetate on glow-discharged carbon-coated copper grids. Samples were viewed on a LE0912 AB TEM and photographed at ×37,500 and 40,000 with a Promscan 2,048 by 2,048 pixel charge-coupled device camera.

Crystallography—Crystals of Csa2 were grown in 2 + 2 μl hanging drops setup at 11 mg/ml in 1.9–2.1 M sodium formate, 0.1 M sodium acetate (pH 3.8–4.1), and 0.2 M sodium thiocyanate. Crystals were derivatized by soaking 22 h in 15 mM KAu(CN)2 then transferred to synthetic mother liquor containing 12.5% glycerol as a cryoprotectant for 30 s, and flash-frozen in liquid nitrogen. A 2.0-Å resolution three-wavelength anomalous diffraction data set centered on the Au-L3 edge was collected from a KAu(CN)2-soaked crystal at the Stanford Synchrotron Radiation Laboratory (SSRL beamline 9–2). A 2.0-Å resolution single-wavelength dataset from a native crystal was also collected (Table 1). Data were integrated, scaled, and merged with HKL2000 (45). Initial phases were determined with SOLVE/RESOLVE (46–48). The asymmetric unit contained four copies of Sso1442 and solvent content was 50%. Iterative model building and refinement were done with Coot (49) and REFMAC5 (50, 51) using the native data set. Residue numbers in the model are consistent with the native protein sequence. Residues 17–26, 165–172, 233–240, 303–305, and 321 of chain A, 15–22, 149–179, 234–242, and 302–303 of chain B, 16–25, 164–176, 233–241, 302–303, and 320–321 of chain C, 16–26, 149–177, 234–241, 301–303, and 321 of chain D were not modeled due to lack of interpretable electron density. TLS parameters (52) were included in the refinement, wherein the four Sso1442 chains were divided into 7 TLS groups with an average R-factor of 0.5 (4.7).
\textbf{crRNA Recognition by aCASCADE}

\begin{table}[ht]
\centering
\caption{Csa2 model refinement} \label{tab:1}
\begin{tabular}{|c|c|}
\hline
\textbf{Data} & \\
\hline
\textbf{R}_{work} \textsuperscript{a} (%) & 18.3 (20.0) \\
\textbf{R}_{free} \textsuperscript{b} (%) & 22.1 (24.6) \\
Real space CC \textsuperscript{c} (%) & 0.960 \\
Mean B value (overall; Å\textsuperscript{2}) & 22.155 \\
Coordinate error (based on maximum likelihood, Å) & 0.096 \\
\hline
\end{tabular}
\begin{flushright}
\begin{itemize}
\item Root mean square deviation from ideality
\item Bonds (Å)
\quad 0.006
\item Angles (°)
\quad 0.930
\item Ramachandran plot\textsuperscript{d}
\item Most favored (%)
\quad 98.08
\item Additional allowed (%)
\quad 1.92
\item PDB accession code
\quad 3PS0
\end{itemize}
\end{flushright}
\end{table}

\textsuperscript{a} \textbf{R}_{work} = \Sigma |\mathbf{F}_{obs} - \mathbf{F}_{calc}| / \Sigma |\mathbf{F}_{obs}|, where \mathbf{F}_{obs} and \mathbf{F}_{calc} are the observed and calculated structure factor amplitudes used in refinement.

\textsuperscript{b} \textbf{R}_{free} is calculated as \textbf{R}_{work} but using the test set of structure factor amplitudes that were withheld from refinement (4.0%). Numbers in parentheses refer to the fit to the reflections in the highest resolution bin (2.051–2.000 Å).

\textsuperscript{c} Correlation coefficient (CC) is agreement between the model and 2m\mathbf{F}_{o} – \mathbf{F}_{c} density map.

\textsuperscript{d} Calculated using Molprobity (53).

Most favored groups each (chain A, 1:1–29, 2:30–64, 3:65–105, 4:106–192, 5:193–226, 6:227–272, 7:273–320; chain B, 8:1–28, 9:29–63, 10:64–135, 11:136–193, 12:194–225, 13:226–296, 14:297–321; chain C, 15:1–44, 16:45–98, 17:99–132, 18:133–198, 19:199–260, 20:261–295, 21:296–319; chain D, 22:1–42, 23:43–99, 24:100–185, 25:186–228, 26:229–259, 27:260–295, 28:296–320). The final model yielded an \textbf{R}_{work} of 18.3% and an \textbf{R}_{free} of 22.1%. The geometry of the final model was validated with Molprobity (53) (Table 2). Coordinates and structure factors have been deposited with the Protein Data Bank (PDB code 3PS0). Three-dimensional structural homology searches were carried out using the DALI (54), VAST (55), and SSM (56) servers, and structural figures were generated with PyMOL (57).

\section*{RESULTS}

\textbf{Isolation of a CASCADE-like Protein-RNA Complex from S. solfataricus—}The \textit{S. solfataricus} genome encodes orthologs for two components of \textit{E. coli} CASCADE, Csa2 (Cas7) and Cas5a (14). To determine whether Csa2 and Cas5a participate in a CASCADE-like complex in \textit{S. solfataricus}, N-terminal tandem affinity (StrepII and His\textsubscript{8})-tagged Csa2 (Sso1442) was expressed in \textit{S. solfataricus}. To determine whether Csa2 and Cas5a participate in a

Because \textit{E. coli} aCASCADE binds processed crRNA and DNA (12), we examined the aCASCADE complex for co-purifying nucleic acid using basic phenol-chloroform extraction followed by RNase A or DNase I digestion. TBE-urea-PAGE with SYBR Gold staining revealed RNA species of 60–70 nt, plus low amounts of higher molecular weight RNA that included faint bands at two and three times the molecular weight of the major species. aCASCADE also co-purified with a smaller amount of high molecular weight (≥300 nt) DNA (supplemental Fig. S1B). It is not clear whether this represents a specific crRNA-DNA interaction, or like \textit{E. coli} CASCADE, the complex copurifies with nonspecifically bound DNA (8, 12). The RNA was found to co-purify with aCASCADE through the streptactin, Ni-NTA, and size exclusion chromatography (Fig. 1B). Furthermore, when subjected to ribonuclease protection assays the RNA in the complex showed no visible degradation, even after 24 h, indicating that the RNA is tightly bound and protected along its entire length (supplemental Fig. S1C).

To confirm that the RNA was CRISPR-derived, the 60–70 nt band was gel-extracted, cloned, sequenced, and compared with the \textit{S. solfataricus} P2 genome and two available \textit{S. solfataricus} P1 CRISPR sequences (32). Fifteen of 16 sequenced clones were CRISPR derived with fragments of direct repeat sequence on the 5′ and 3′ ends that were separated by variable spacer sequence (Fig. 1C). All three CRISPR repeat sequences found in \textit{S. solfataricus} P2 were represented among the clones, indicating this aCASCADE complex binds each type of crRNA. Clone 7 contained the shortest spacer, 38 bases, whereas the clone 6 spacer was the longest at 44 bases. Twelve clones contained spacers present in strain P2 CRISPRs and thus could be assigned to individual CRISPR loci (B, C, D, and F) (Fig. 1C). Three additional clones had spacers that were not present in the sequenced \textit{S. solfataricus} P1 or P2 CRISPRs and based on the direct repeat sequence could belong to either CRISPR A or B. Eight of the clones represented a complete repeat-spacer unit with 8 nt of repeat sequence at the 5′ end and 16–17 nt of repeat at the 3′ end, reminiscent of CRISPR transcript processing in \textit{P. furiosus} by Cas6 (16). Several clones also had single nucleotide mismatches with the P2 CRISPR repeat sequences, potentially due to differences between the PH1–16 and P2/P1 strains or the use of an error-prone polymerase to amplify the cDNA.

To identify additional proteins that bind weakly or are present in lower abundance, purification of aCASCADE was limited to the streptactin resin and analyzed by in-solution tryptic digest and LC-MS/MS in three independent experiments (supplemental Table S2). In addition to the expected Csa2 (Sso1442) and Cas5a (Sso1441), we also identified the crRNA processing endonuclease Cas6 (Sso1437) and Cas5 (Sso1443). Csa5 is a 150-residue protein of unknown function encoded immediately upstream from Csa2 in many archaeal genomes.
and may thus represent an archaeal-specific CASCADE subunit. We also identified the Csa2 paralogue, Sso1399, the Cas5a paralogue Sso1400, and in one experiment, Csa4 (Sso1401/Cas8a2). After size exclusion chromatography, Csa2 (Sso1442) and Cas5a (Sso1441) co-purified as expected, but the Csa5 (Sso1443), Cas6 (Sso1437), and Csa4 (Sso1401) proteins were not detectable by mass spectrometry, suggesting that they are more weakly associated in aCASCADE, or might associate with aCASCADE via an incompletely processed CRISPR transcript.

Recombinant *S. solfataricus* Cas6 Generates Fragments Identical to Those in aCASCADE—Cas6 from *P. furiosus* has been shown to cleave the CRISPR transcripts specifically, 8 bases upstream from the 3’ end of the direct repeats (16), generating products similar to those identified in *S. solfataricus* aCASCADE, although an equivalent activity has not been demonstrated in *S. solfataricus*. The annotated Cas6 orthologs in *S. solfataricus* (Sso1381, Sso1406, Sso1437, and Sso2004) share negligible sequence similarity with *P. furiosus* Cas6 and even structure based threading using the Phyre server (59) does not

![Figure 1](image-url). **A.** Isolation and characterization of aCASCADE from *S. solfataricus*. A, colloidal Coomassie-stained SDS-PAGE gel showing the co-purification of Csa2 and Cas5a through streptactin (second lane), Ni-NTA (third lane), and size exclusion chromatography (fourth and fifth lanes) steps. MASCOT identification statistics are listed in supplemental Table S1. **B.** SYBR Gold-stained UREA-PAGE gel showing RNA co-purification with aCASCADE through all three purification steps. **C.** Alignment of non-redundant cDNA sequences derived from aCASCADE-associated RNA. The co-purifying RNA is CRISPR derived and hails from each of the three *S. solfataricus* CRISPR types. The labels indicate the CRISPR locus from which the clone is derived. Several clones could not be definitively assigned. The underlined spacer (clone 10) appears to be derived from *Sulfolobus icelandicus* Rod-shaped virus.
find a match with *P. furiosus* Cas6. Accordingly, it was important to ascertain whether the putative *S. solfataricus* Cas6 orthologs cleaved crRNA *in vitro*. We cloned the *sso2004* gene and expressed it in *E. coli*, allowing purification of the recombinant protein. Recombinant *S. solfataricus* Cas6 showed metal-independent ribonuclease activity, cleaving an *in vitro* transcript comprising the first two repeat-spacer units of the *S. solfataricus* P2 CRISPR A locus, yielding a pattern consistent with cleavage at a single position within the repeat at the same position cleaved by *P. furiosus* Cas6 (Fig. 2, A and C). This was confirmed by the cleavage pattern generated from an RNA oligonucleotide comprising a single 25-nt repeat sequence with a 15 unit 5’ extension, which was cut at a single site (5’-AGGA/AUUG) (Fig. 2, B and D), yielding the 8-nt 5’ tag (or 5’ handle) identified previously (16). Thus, Cas6 is physically associated with the Csa2-Cas5a complex and can generate the crRNA products found in this complex. This is reminiscent of *E. coli* CASCADE where the crRNA cleaving subunit CasE (cse3) is a constituent of the complex (8).

**Investigation of the Core aCASCADE Subunits**—For structural studies and activity assays, recombinant Csa2 (Sso1442) was expressed in *E. coli*, both alone and with Cas5a (Sso1441). For His-tagged Csa2 alone, size exclusion chromatography was unable to distinguish between monomer and dimer. Analytical ultracentrifugation revealed a monomer-dimer equilibrium with a dissociation constant of 4.5 μM, indicating that in the absence of other CASCADE components, recombinant Csa2 is predominantly monomeric at physiologically relevant concentrations. In contrast, Csa2 and Cas5a co-expressed in *E. coli* formed a stable complex that could be purified to homogeneity. The Coomassie staining suggested an excess of Csa2 over Cas5a (Fig. 3A), in agreement with the data for the complex isolated from *S. solfataricus*, although the relative amounts of Cas5a are higher for the complex expressed in *E. coli*.

The Csa2-Cas5a Complex Binds crRNA and Forms Ternary Complexes with Target DNA—*E. coli* CASCADE utilizes bound crRNA to target viral DNA, forming a ternary complex that is thought to result in cleavage of the DNA target by other CAS proteins, most likely Cas3 (8, 12). To determine whether Csa2-Cas5a had similar functionality, we carried out electrophoretic mobility shift assays (EMSA) to visualize crRNA and DNA binding using the sequence of CRISPR locus A, spacer 1. We first tested the ability of Csa2 and the Csa2-Cas5a complex to bind radiolabeled crRNA (Fig. 3B). Both bound the crRNA with roughly similar affinities, suggesting that Csa2 is the major RNA binding subunit of the complex. In the absence of crRNA-A1, the Csa2-Cas5a complex showed very little binding to a labeled target DNA species (Target-A1f) (supplemental Table S3, and Fig. 3C, lanes 1–5). However, when the protein complex was preincubated with crRNA-A1, which can base pair with the central region of the tA1 oligonucleotide, an RNA-DNA heteroduplex was formed that was gel-shifted efficiently by the Csa2-Cas5a complex (lanes 6–10). The Csa2-Cas5a-crRNA complex did not bind the reverse complementary DNA strand (Target-A1r), which cannot form a heteroduplex with crRNA-A1 (lanes 11–15). These data demonstrate that the Csa2-Cas5a complex has a crRNA-dependent DNA binding activity that is consistent with its presumed function in CRISPR-mediated antiviral defense, analogous to *E. coli* CASCADE. Recent data in *S. solfataricus* suggests that target DNAs are only cleaved if they include a “protospacer adjacent motif” (PAM) sequence, typically CCN, at the 5’ end of the protospacer (28, 32). This may be a mechanism to allow dis-
crRNA Recognition by aCASCADE

The Structure of Csa2—To better understand the role of the Cas7 family of proteins in CASCADE in general, and the role of Csa2 in aCASCADE in particular, the structure of recombinant Csa2 was solved using x-ray crystallography. Csa2 crystallized in space group P212121 with four copies per asymmetric unit (chains A-D). However, the protein-protein contacts in the Csa2 crystal appear unlikely to recapitulate protein-protein interactions in aCASCADE. Although there are substantial contacts at the A/B and C/D interfaces, the surfaces are discontinuous and exhibit improper, closed symmetry that is inconsistent with the apparent open symmetry of aCASCADE.

The structure of the Csa2 protomer reveals a 3-domain, crescent-shaped structure that is 65 Å in length, tip to tip (Fig. 4A). The central domain is comprised of a five-stranded anti-parallel β-sheet (β6, β7, β1, β8, and β9), flanked by four α-helices. The first four strands of the central β-sheet along with helices α2 and α8 unexpectedly display the βαβαβ topology of the RNA-binding domain or RNA recognition motif (RRM), a ferredoxin-like fold that frequently serves in RNA recognition. We thus refer to these structural elements as the RRM-like subdomain (purple, Fig. 4A). In Csa2, the RRM is elaborated upon by a C-terminal addition comprised of residues 256–320 that begins with an extended 13-residue connection leading into helix α9. This is followed by a short

crimination between foreign DNA and the chromosomal CRISPR loci, which lack PAM sequences and are therefore not targeted by the CRISPR system (10). The target oligonucleotide A1 used for gel shifting included a PAM sequence, however, binding to an alternative oligonucleotide lacking a PAM sequence gave similar results (data not shown). This suggests that there is no discrimination based on PAM presence or absence in the minimal recombinant system tested here.

Structural Studies of aCASCADE—To investigate the overall structure of the aCASCADE complex purified from S. solfataricus, the purified material was visualized using negative stain transmission electron microscopy (TEM). Following elution from the streptactin column, and at all subsequent stages of the purification including the peak fraction from the size exclusion column, we observed protein filaments with a width of ~6 nm. Interestingly, the filaments are present as extended right-handed helices of variable length, with an average helical width of 11.5 nm and a pitch of 14 nm (Fig. 3D). These particles are clearly larger than that suggested by size exclusion chromatography, probably due to their non-spherical nature.

Importantly, the observed right-handed helical assemblies indicate that Csa2, the major protein in aCASCADE, is capable of forming oligomers with open, as opposed to closed symmetry. Although these assemblies are significantly larger than those seen in E. coli CASCADE, we note that this unusual open oligomeric assembly is consistent with the relative abundance of the Cas7 protein in both S. solfataricus (Csa2) and E. coli (CasC) CASCADE. In addition, the observation that recombinant Csa2 expressed in E. coli is predominantly monomeric in the absence of Cas5a and crRNA suggests that Cas5a and crRNA may be responsible for nucleation and/or stabilization of this helical assembly. For these, and other reasons discussed in more detail below, we believe that many features of this extended complex are likely to be relevant to endogenous aCASCADE.

The structure of aCASCADE reveals a 3-domain, crescent-shaped structure that is 65 Å in length, tip to tip (Fig. 4A). The central domain is comprised of a five-stranded anti-parallel β-sheet (β6, β7, β1, β8, and β9), flanked by four α-helices. The first four strands of the central β-sheet along with helices α2 and α8 unexpectedly display the βαβαβ topology of the RNA-binding domain or RNA recognition motif (RRM), a ferredoxin-like fold that frequently serves in RNA recognition. We thus refer to these structural elements as the RRM-like subdomain (purple, Fig. 4A). In Csa2, the RRM is elaborated upon by a C-terminal addition comprised of residues 256–320 that begins with an extended 13-residue connection leading into helix α9. This is followed by a short
connection to β9 that adds as a fifth antiparallel strand to the β-sheet of the RRM, and α10, which sits “underneath” the β-sheet (Fig. 4A). Interestingly, Csa2 lacks the conserved sequence motifs in strands β1 and β7 that, in the canonical RRM, recognize single-stranded nucleic acid (Fig. 4D) (60), although a solvent-exposed aromatic residue is conserved C-terminal to β6 (Tyr141). Consistent with the lack of a canonical ssRNA-binding motif, the extended β8-α9 loop and helix α9 lie on “top” of the antiparallel β-sheet, partially occluding the RNA binding face of the typical RRM-fold. Thus, whereas the central domain of Csa2 may have evolved from the RRM, it is likely that RNA recognition by Csa2 will
crRNA Recognition by aCASCADE

differ in detail from that of other RRM containing proteins currently in the Protein Data Bank.

Two additional bipartite domains, one at each tip of the crescent-shaped protein, are formed by four insertions into the RRM-fold (Fig. 4A and supplemental Fig. S2). The “1–3” domain is found “above” the RRM domain as it is pictured in Fig. 4A, and is formed from insertions 1 and 3. The first insertion (residues 27–46) includes α1, followed by a disordered loop that connects to the β2–β3 hairpin, which extends to the upper tip of the “crescent.” The 1–3 domain is then completed by insertion 3 (residues 145–180), which contributes helix α7. However, this helix is ordered only in chains A and C, where it sits against the face of the β2–β3 hairpin, whereas at least 7 residues in the α7–β7 loop are disordered in all chains.

The 2–4 domain, found “below” the RRM domain, is likewise composed of insertions 2 and 4. The second insertion, residues 68–136, contributes a mixed α/β structure consisting of α3–α6, which lies along one face of the short antiparallel β4–β5 hairpin, whereas the opposite face of the β4–β5 hairpin remains solvent exposed on the concave face of the crescent. The fourth insertion, residues 192–216, contains another extended connection followed by the N-terminal half of α8. The α8 helix is kinked at residue 216 where it leaves the 2–4 domain and forms the second helix of the βββαβ RRM-like-fold.

To identify functionally important residues, including potential sites for RNA recognition, or interactions with other aCASCADE subunits, we examined the locations of conserved surface features. Makarova et al. (14) identified 3 conserved sequence motifs in CRISPR-associated Cas7 proteins, specifically: 1) s-h-Asn; 2) Arg; and 3) (Phe/Pro/His/Gly)-Gly, where s and h indicate small and hydrophobic residues, respectively. In Csa2 these correspond to: 1) Ser14-Leu15-Asn16; 2) Glu58; and 3) Gly121–Gly122. All six residues are solvent exposed and found on the concave surface of the Csa2 crescent (Fig. 4B). Among these, Asn16 is poorly ordered (chain A) or disordered (chains B–D).

Because few residues are generally conserved among all Cas7 proteins, we also examined the location of surface residues conserved just among the Csa2 orthologs (Fig. 4B and supplemental Fig. S3). Most of the strictly conserved, solvent-exposed residues are found in two closely spaced clusters, which are coincident with the Cas7 motifs discussed above, and are thus also found on the concave surface of the crescent. The first cluster is on the surface of the 1–3 domain near Asn16, whereas the second cluster lies at the interface between the RRM-like subdomain and the 2–4 domain, near Gln58 and Gly121–Gly122. We thus refer to these clusters as the asparagine (Asn) and glycine (Gly) clusters, respectively (Fig. 4B). These surfaces are hydrophilic and somewhat basic (Fig. 4C). Although either cluster might indicate a surface involved in subunit interactions, the identities of several conserved residues (His160, Arg162, His55, and Asn16) are more suggestive of nucleic acid rather than protein recognition. In agreement with this, mutation of His160 to alanine resulted in a significant reduction in the affinity of the variant Csa2 for crRNA (Fig. 4E).

These conserved surface features are adjacent to three disordered loops, two of which contain additional residues that are strictly conserved among Csa2 orthologs; Gly22 and Asn23 are found in the α1–β2 loop, and Arg240 is present in the α8–β8 loop (Fig. 4B). Further evidence for flexibility is also seen in the orientation of the β2–β3 hairpin of the 1–3 domain, which is shifted by 10.4 Å in chains B and D, relative to that in chains A and C (supplemental Fig. S4). The conformational change is accompanied by the loss of additional ordered density in the α7–β7 loop in chains B and D, including strictly conserved His160 and Arg162 (supplemental Fig. S4). The presence of conserved, flexible loops in Csa2 and the lowered affinity of the H160A variant for crRNA (above, Fig. 4E) suggests these flexible loops may be involved in the recognition of crRNA or in crRNA-directed DNA recognition.

DALLI and SSM searches (54, 56) identified P. furiosus Cas6 (16, 17) as the closest structural homolog to Csa2, but the similarity is limited to the RRM-like subdomain (Cas6 PDB code 3I4H, 2.9-Å root mean square deviation over 87 residues for chain A). Cas6, a member of the repeat-associated mysterious protein superfamily, displays tandem ferrodoxin- or RRM-like domains, with the N-terminal domain of Cas6 showing the greatest similarity to Csa2 (Fig. 5). Cas6 is a metal-independent ribonuclease, in which His46, Tyr41, and Lys52 form a putative catalytic triad (16, 17). The SSM superposition places these residues in the vicinity of the Gly cluster on Csa2. Although the Gly cluster lacks a recognizable catalytic triad and is thus unlikely to represent a nuclease active site, the structural alignment with the Cas6 active site further suggests the Gly cluster may function in nucleic acid recognition. In contrast to the RRM domain, Dali queries of the Protein Data Bank with the 1–3 and 2–4 domains alone did not yield statistically significant matches, these domains thus appear to be unique to the Csa2 structure.

We next asked how the structure of the Csa2 protomer might relate to the helical assemblies observed by TEM, and began by considering the relative scale of these two structures. Interestingly, the width of the helical assemblies (6 nm) observed by TEM is approximately equal to the tip-to-tip diameter of the crescent-shaped Csa2 protomer (65 Å). The right-handed helical assembly can thus be crudely modeled by placing multiple copies of the Csa2 protomer in a right-handed helical arrangement such that the long axis of the Csa2 protomer runs perpendicular to the direction of the protein filament, and by requiring the pitch and width of the model helix to be consistent with the TEM images. In addition, in our models, we also chose to require the conserved crescent-shaped face of Csa2, including His160, which appears to be involved in RNA recognition, to remain solvent exposed. We emphasize that there are additional ways these helices might be modeled, and that there is no reason to believe that the Csa2/Cas6 interface employed in our models corresponds to that in the real complex. However, the exercise is valuable in that it suggests 6–12 protomers per turn of helix.

DISCUSSION

Here we report the first structure of a Cas7 protein, and the isolation and characterization of a complex that bears many of the hallmarks expected of an archaeal CASCADE. Similar to E. coli CASCADE (8, 12), aCASCADE includes a Cas7/CasC protein (Csa2), a CasD/Cas5e ortholog (Cas5a), and the com-
plex co-purifies from *S. solfataricus* with processed crRNA. Furthermore, the recombinant Csa2–Cas5a complex produced in *E. coli* specifically binds crRNA, which in turn, recognizes single-stranded "target" DNA *in vitro*. Finally, the complex isolated from *S. solfataricus* also co-purifies with the more weakly interacting or lower abundance components, Cas6, Csa5, and perhaps Csa4. In *S. solfataricus* aCASCADE, Cas6 appears to serve a function analogous to that of *E. coli* CasE/Cse3 (8). Thus, there are clear orthologs in aCASCADE for each component of *E. coli* CASCADE except CasA and CasB, components that appear to be limited to the *E. coli* CRISPR/Cas subtype (13, 14), although the possibility that Csa4 (Sso1401/Cas8a2) is functionally similar to CasA might be considered. Importantly, the presence of core CASCADE components in *S. solfataricus* aCASCADE (Csa2/Cas5a), as well as Cas6, suggests that structural and functional studies of aCASCADE are relevant not only to the Apern subtype (Csa) CASCADE, but are also generally relevant to orthologs in other CRISPR/Cas subtypes, especially the Cst, Csh, and Csm subtypes that associate with Cas6 (13, 17, 61, 62). The potentially loose association of Cas6 in the *S. solfataricus* aCASCADE may allow it to function in the initial processing of the CRISPR transcript for both CRISPR systems.

**Structural Models for aCASCADE**—The overexpression of Csa2 in *S. solfataricus* results in the production of extended right-handed helical assemblies of variable length. The ability of the extended Csa2 assembly to bind crRNA and to protect the crRNA from RNase digestion suggests that the crRNA is bound by protein along its entire length. That the assembly copurifies with Cas5a, Csa5, and Cas6, in addition to the crRNA, also suggests that many aspects of the assembly are physiologically relevant. However, the extended helical filaments observed in preparations from *S. solfataricus* are longer than needed to accommodate a single crRNA; one turn of the Csa2 helix should be more than sufficient to accomplish this. Thus, if these helices are physiologically relevant, they are likely to harbor multiple crRNAs and could potentially be used in succession to screen potential target DNA for a match to the collection of bound crRNA.

Alternatively, we note that the open symmetry of the Csa2 assembly coupled with high concentrations of Csa2 from over-

---

**FIGURE 5.** The structural similarity between Csa2 and Cas6 is limited to the RNA-recognition motif. A, Csa2 and Cas6 are shown in equivalent orientations based on an SSM Structural alignment. The RRM-like subdomains are colored violet in both structures. The Csa2 1–3 domain is colored red, the 2–4 domain is colored orange, and the C-terminal subdomain is colored yellow. The portions of the Cas6 N-terminal domain that do not exhibit similarity to the Csa2 RRM subdomain are depicted in green and the C-terminal domain in light cyan. The conserved clusters on Csa2 and the putative active site residues on Cas6 are shown as “sticks” colored in dark cyan.
expression in *S. solfataricus* and substoichiometric amounts of endogenous Cas5a, Cas5, or Cas6, might allow the assembly to grow to physiologically irrelevant lengths, particularly in the presence of stabilizing crRNA. Thus, we also consider a shorter model, in which native aCASCADE includes a single crRNA and a limited number of Csa2 subunits, resulting in an arch-shaped structure corresponding to less than one turn of a helix. Indeed, this second model is consistent with the most recent model for *E. coli* CASCADE, which binds a single crRNA and is observed as a smaller arch-shaped particle (12). We suggest that the arch-shaped backbone of *E. coli* CASCADE is explained by the presence of a half-turn or more of CasC helix (supplemental Fig. S5), and that the helical, open symmetry of aCASCADE is also present in the CasC backbone of *E. coli* CASCADE.

In either model, the major function of Csa2 appears to be the construction of an extended assembly that functions to support the crRNA spacer sequence along its entire length (thus protecting it from RNase digestion). At the same time, we anticipate that the Watson-Crick edges of the bases must remain solvent exposed and available for interaction with target DNA, such that aCASCADE also effectively templates or presents the crRNA spacer sequence for DNA recognition. This suggests that some of the conserved surface features on Csa2, including His160, will tightly interact with RNA in a sequence-independent manner. Collectively, the interactions with the Csa2 protomers might also serve to stabilize a hybrid RNA/target-DNA complex, and/or destabilize bound dsDNA, allowing DNA within the cell to be surveyed for homology to the bound crRNA spacer.

How is crRNA specifically recruited to aCASCADE? The data are consistent with a cooperative CASCADE assembly process that is dependent on the presence of crRNA, as we do not see extended helices of recombinant Csa2 or Csa2/Cas5a in the absence of crRNA. In addition, whereas the Csa2 backbone of aCASCADE is expected to bind the variable CRISPR spacer in a sequence-independent manner, the complex clearly distinguishes crRNA from other cellular RNAs, most likely through sequence-specific interactions with the 5′ and 3′ handles. For these reasons, it is attractive to consider roles for the additional aCASCADE components (Cas5a, Cas6, Cas4, and Cas5) in specific crRNA recognition. Such interactions might also serve to nucleate (crRNA induced oligomerization) or terminate growth of the Csa2 helix, and thus govern the length of the Csa2 backbone in aCASCADE. Termination of helix growth could, in turn, limit the length of the nucleoprotein filament to a single crRNA, giving rise to an *E. coli*-like aCASCADE assembly, rather than the extended nucleoprotein filaments seen when Csa2 alone is overexpressed in *S. solfataricus*. Thus, guided by our own data, and by the current model for *E. coli* CASCADE (8, 12), we propose the model for aCASCADE presented in Fig. 6. The structural core of aCASCADE is modeled as a partial turn of Csa2 helix, with crRNA running along the length of the Csa2 assembly, and Cas5a, Cas6, Cas4, Cas5, or other unidentified proteins at the 5′- and 3′-ends, where they may serve to initiate and terminate growth of the complex.

Once assembled, CASCADE must probe DNA within the cell for sequences complimentary to the bound crRNA spacers. Although we cannot predict with certainty which surface CASCADE might be utilized for this process, we note that the concave surface formed by a partial turn of the Csa2 helix is large enough to accommodate, or wrap around dsDNA, or a hybrid RNA-DNA complex. Furthermore, several of our TEM-based models indicate that the partial Csa2 helix can be docked to dsDNA in a coaxial arrangement, that is, with the Csa2 helical axis coincident with that of the DNA double helix, such that the Csa2 helix wraps around the dsDNA. In such an arrangement, the Csa2 protomers would be positioned along the dsDNA, each equidistant from the DNA, potentially facilitating DNA surveillance. For these reasons, we tentatively place the conserved surfaces of Csa2 and crRNA along the concave surface of the oligomeric Csa2 arch. However, we must emphasize that there is, as yet, no experimental evidence to support the position of the DNA or crRNA within the proposed model. On the other hand, this model is similar to the current model for RecA, where RecA binds ssDNA to form a helical nucleoprotein filament, which in turn catalyzes recognition of homologous dsDNA and strand exchange to produce the heteroduplex. In particular, crystallographic and EM studies indicate RecA and its eukaryotic homologs do indeed wrap around dsDNA such that the helical axes of the protein and nucleic acid are coincident with each other (63, 64), similar to the tentative model for CASCADE that we propose here.

Upon reflection, additional advantages of the unusual open symmetry of the Csa2 oligomer become apparent, particularly in light of the variable spacer sequences seen both within and between different organisms utilizing CRISPR-Cas. *E. coli* K12 shows a limited range of spacer lengths, generally 32–33 bases (supplemental Fig. S6A). Assuming the spacer sequence is largely bound by the 6 CasC subunits in the backbone of *E. coli* CASCADE, each CasC subunit would accommodate about 5–6 bases of the crRNA spacer. However, spacer lengths are generally longer in Archaea. Thus, for *S. solfataricus*, spacer lengths vary between 34 and 44 bases, with a 39-base spacer the most common (Fig. 1C and supplemental Fig. S6A). Although there are other ways the helix parameters might be adjusted, the additional length of these archaeal spacers might be accommodated by extending the Csa2 helix from 6 to 7, or even 8 subunits, depending on the size of the particular spacer. Similarly, the possibility that crRNA length may define the number of CasC subunits has also been considered for *E. coli CASCADE* (12). Thus, a role for the crRNA spacer in governing the number of Cas7 subunits in the CASCADE “backbone” may be a general feature of CASCADE architecture, allowing crRNA of variable lengths to be accommodated, both within and across species.

The architecture of CASCADE might also impose constraints on the size of a functional CRISPR spacer. Specifically, the majority (greater than 99.6%) of spacers in the CRISPR data base (65) are 50 nt or shorter (supplemental Fig. S6B). This may, in part, reflect the need for growth of the Csa2 oligomer, and Cas7 oligomers in general, to terminate before completing a full turn of helix, allowing the inner surface of the helical Csa2 backbone to remain accessible to dsDNA.
CRISPR-mediated Viral Defense in the Archaea—The data presented here for *S. solfataricus*, which to our knowledge includes the first structure for a member of the Cas7 superfamily, coupled with previous work on the CMR complex in *P. furiosus* (9), allows the construction of an emerging model for the CRISPR system in many Archaea (Fig. 6E). CRISPR transcripts are processed by Cas6 to produce crRNA and incorporated into aCASCADE, which has a stable core composed of Csa2-Cas5, with the former present at a higher copy number like its *E. coli* homolog CasC/cse4. Additional subunits may include Csa5, which appears unique to aCASCADE, and Cas6. This complex can form a ternary complex with target DNA that is presumably cleaved by the Cas3 helicase-HD nuclease protein(s) (18), again in line with the situation in
E. coli. Alternatively, the crRNA can be further processed by an unknown nuclease to remove the 3’ repeat-derived RNA, generating smaller crRNAs that are loaded into the CMR complex and used to target viral RNA (9). The two systems, aCASCADE and CMR, may work in parallel in a “belt and braces” approach to maximize the utility of the CRISPR system.

Acknowledgments—We thank Paul Talbot and Donald L. Alsbury for technical assistance and Dr. Christophe Rouillon and Dr. Jing Zhang for scientific discussion. Portions of this work were carried out at the Stanford Synchrotron Radiation Laboratory, a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research and by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. The Macromolecular Diffraction Laboratory at Montana State University was supported, in part, by a grant from the Murdock Science. The Macromolecular Diffraction Laboratory at Montana State University was supported, in part, by a grant from the Murdock Foundation. We thank the Mass Spectrometry facility in St. Andrews supported by the Wellcome Trust.

REFERENCES

1. van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., and Brouns, S. J. (2009) Trends Biochem. Sci. 34, 401–407
2. Horvath, P., and Barrangou, R. (2010) Science 327, 167–170
3. Sorek, R., Kunin, V., and Hugenholtz, P. (2008) Nat. Rev. Microbiol. 6, 181–186
4. Tyson, G. W., and Banfield, J. F. (2008) Environ. Microbiol. 10, 200–207
5. Karginov, F. V., and Hannon, G. J. (2010) Mol. Cell 37, 7–19
6. Andersson, A. F., and Banfield, J. F. (2008) Science 320, 1047–1050
7. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) Science 315, 1709–1712
8. Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V., and van der Oost, J. (2008) Science 321, 960–964
9. Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graevey, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009) Cell 139, 945–956
10. Marraffini, L. A., and Sontheimer, E. J. (2008) Science 322, 1843–1845
11. Garneau, J. E., Dupuis, M. É., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A. H., and Moineau, S. (2010) Nature 468, 67–71
12. Jore, M. J., Lundgren, M., van Duijn, E., Butema, J. B., Westra, E. R., Waghmare, S. P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., Beijer, M. R., Barendregt, A., Zhou, K., Snijders, A. P., Dickman, M. J., Doudna, J. A., Boekema, E. J., Heck, A. J., van der Oost, J., and Brouns, S. J. J. (2011) Nat. Struct. Mol. Biol., in press
13. Haft, D., Selengut, J., Morgindorf, E., and Nelson, K. (2005) PloS Comp. Biol. 1, 474–483
14. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006) Biol. Direct 1, 7
15. Belousova, N., Brown, G., Zimmerman, M. D., Proudfoot, M., Makarova, K. S., Kudritska, M., Kudritska, M., Kochinyan, S., Wang, S., Chruszcz, M., Minor, W., Koonin, E. V., Edwards, A. M., Savchenko, A., and Yakunin, A. F. (2008) J. Biol. Chem. 283, 20361–20371
16. Ke, A., and Delisa, M. P. (2011) Mol. Microbiol. 79, 584–599
17. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M., and Li, H. (2011) Structure 19, 257–264
18. Siksnys, V. (2011) EMBO J. 30, 1335–1342
19. Perez-Rodriguez, R., Haitjema, C., Huang, Q., Nam, K. H., Bernardis, S., Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. D Biol. Crystallogr. 60, 2126–2132
20. Bailey, S. (1994) Acta Crystallogr. D Biol. Crystallogr. 50, 760–763
21. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr.


**crRNA Recognition by a CASCADE**

D 53, 240–255

52. Painter, J., and Merritt, E. A. (2006) *Acta Crystallogr. D Biol. Crystallogr.* **62**, 439–450

53. Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437–450

54. Holm, L., and Sander, C. (1993) *J. Mol. Biol.* **233**, 123–138

55. Gibrat, J. F., Madej, T., and Bryant, S. H. (1996) *Curr. Opin. Struct. Biol.* **6**, 377–385

56. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268

57. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, Schrödinger, LLC, New York

58. Hügler, M., Krieger, R. S., Jahn, M., and Fuchs, G. (2003) *Eur. J. Biochem.* **270**, 736–744

59. Kelley, L. A., and Sternberg, M. J. (2009) *Nat. Protoc.* **4**, 363–371

60. Maris, C., Dominguez, C., and Allain, F. H. (2005) *FEBS J.* **272**, 2118–2131

61. Lawrence, C. M., and White, M. F. (2011) *Structure* **19**, 142–144

62. Kunin, V., Sorek, R., and Hugenholtz, P. (2007) *Genome Biol.* **8**, R61

63. Chen, Z., Yang, H., and Pavletich, N. P. (2008) *Nature* **453**, 489–494

64. Sheridan, S. D., Yu, X., Roth, R., Heuser, J. E., Sehorn, M. G., Sung, P., Egelman, E. H., and Bishop, D. K. (2008) *Nucleic Acids Res.* **36**, 4057–4066

65. Grissa, I., Vergnaud, G., and Pourcel, C. (2007) *BMC Bioinformatics* **8**, 172