A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs

Asma Hatoum-Aslan, Pouam Samai1, Inbal Maniv, Wenyan Jiang, and Luciano A. Marraffini2

From the Laboratory of Bacteriology, The Rockefeller University, New York, New York 10065

Background: CRISPR immune systems protect prokaryotes from their viruses using small interfering RNAs (crRNAs), which require maturation events during their biogenesis.

Results: In Staphylococcus epidermidis, crRNAs undergo maturation in a Cas10-Csm ribonucleoprotein complex; Csm3 modulates the extent of maturation.

Conclusion: Csm3 acts as a ruler for crRNAs.

Significance: Investigating CRISPR immunity is important to understand prokaryotic ecology and to develop biotechnological applications.

Small RNAs undergo maturation events that precisely determine the length and structure required for their function. CRISPRs (clustered regularly interspaced short palindromic repeats) encode small RNAs (crRNAs) that together with CRISPR-associated (cas) genes constitute a sequence-specific prokaryotic immune system for anti-viral and anti-plasmid defense. crRNAs are subject to multiple processing events during their biogenesis, and little is known about the mechanism of the final maturation step. We show that in the Staphylococcus epidermidis type III CRISPR-Cas system, mature crRNAs are measured in a Cas10-Csm ribonucleoprotein complex to yield discrete lengths that differ by 6-nucleotide increments. We looked for mutants that impact this crRNA size pattern and found that an alanine substitution of a conserved aspartate residue of Csm3 eliminates the 6-nucleotide increments in the length of crRNAs. In vitro, recombinant Csm3 binds RNA molecules at multiple sites, producing gel-shift patterns that suggest that each protein binds 6 nucleotides of substrate. In vivo, changes in the levels of Csm3 modulate the crRNA size distribution without disrupting the 6-nucleotide periodicity. Our data support a model in which multiple Csm3 molecules within the Cas10-Csm complex bind the crRNA with a 6-nucleotide periodicity to function as a ruler that measures the extent of crRNA maturation.

CRISPR sequences (clustered regularly interspaced short palindromic repeats) are an essential component of a prokaryotic immune system that protects against phage infection and other invading genetic elements (1–3). CRISPR loci harbor an archive of short sequences (known as spacers) derived from past invaders that provide the specificity for CRISPR immunity. These sequences encode small CRISPR RNAs (crRNAs) that together with CRISPR-associated (Cas) proteins, can locate and destroy foreign nucleic acids by an antisense targeting mechanism (4–7). The CRISPR immune system can also build a memory of past infections by incorporating new invader-derived sequences into CRISPR loci (1, 8). Found in 40% of bacteria and nearly all archaea (9–12), CRISPR-Cas systems exhibit remarkable mechanistic and functional diversity. They can be classified in three main types (I–III) that are defined based upon cas gene content and differences in the mechanism of immunity (13).

crRNA biogenesis is the essential first step in the CRISPR immunity pathway. Spacers range from 24–48 nucleotides in length and are interrupted by similarly sized repeat sequences (see Fig. 1A). This repeat-spacer array is transcribed as a long precursor that is subsequently processed to liberate mature crRNAs. In all CRISPR-Cas systems, the first step in crRNA biogenesis, known as primary processing, entails endoribonucleolytic cleavage within repeats. In type I and III systems, Cas6 is considered the primary processing endonuclease (2, 13–15). One exception appears to be the type I-C system in Bacillus halodurans, where Cas5d was recently shown to catalyze the cleavage of repeat sequences (16). In contrast, primary processing in type II systems relies upon an antisense trans-encoded crRNA and the host RNase III to cleave within repeats (17). Primary processing generates crRNA intermediates that consist of a single spacer that is flanked on both ends by partial repeats. Whereas no further processing is known to occur in type I CRISPR-Cas systems, in type II and III systems, the crRNA intermediates are subject to a final maturation step that eliminates repeat and spacer sequences at the 5’ or 3’ end of the intermediate, respectively (17–19).

Type III CRISPR-Cas systems have been classified into two subtypes: III-A, containing the csn module of cas genes, and III-B, harboring the cmr module (13). In both subtypes, mature crRNAs display an invariant 5’ end containing 8 nt of repeat sequence (the crRNA tag) but variable 3’ ends that match the targeted sequence in the phage or plasmid genome (14, 19).

1 Supported by a Helmsley Postdoctoral Fellowship for Basic and Translational Research on Disorders of the Digestive System at The Rockefeller University.
2 Supported by the Searle Scholars Program, the Rita Allen Scholars Program, an Irma T. Hirschl Award, and National Institutes of Health Director’s New Innovator Award 1DP2AI10455-01. To whom correspondence should be addressed. Laboratory of Bacteriology, The Rockefeller University, 1230 York Ave., New York, NY 10065. Tel.: 212-327-7014; Fax: 212-327-8262; E-mail: marraffini@rockefeller.edu.

3 The abbreviations used are crRNA, CRISPR RNA; Cas, CRISPR-associated; nt, nucleotide(s); NEB, New England Biolabs.
Primary processing cleaves the repeat sequence immediately upstream of the crRNA tag and maturation occurs at the 3′-end of the intermediate crRNA. Importantly, the extent of maturation at the 3′-end determines the cleavage site within the target sequence (5), and its mechanism remains poorly understood.

The type III-A system of \textit{Staphylococcus epidermidis} RP62a contains nine \textit{cas/csm} genes (see Fig. 1A). Previously, we showed that a ruler mechanism anchored at the primary processing site generates mature crRNAs of discrete lengths (37 and 43 nucleotides, see Fig. 1B) and that \textit{csm2}, \textit{csm3}, and \textit{csm5} are required for crRNA maturation (19). Here, we show that \textit{Csm2}, \textit{Csm3}, and \textit{Csm5}, along with \textit{Csm4} and the type III signature protein \textit{Cas10}, are part of a ribonucleoprotein complex analogous to the Cascade (CRISPR-associated complex for antiviral defense) complex described for \textit{Escherichia coli} and other organisms (2, 5, 16, 20, 21). This complex, here named \textit{Csm10-Csm}, is enriched with mature crRNAs that range from 31 to 67 nucleotides, measured precisely in 6-nucleotide increments. We show that \textit{Csm3} is essential for the formation of this complex, and demonstrate that mutating conserved residues or changing overall levels of \textit{Csm3} in vivo alters the size distribution of the crRNAs without disrupting their 6-nucleotide periodicity. Furthermore, recombinant \textit{Csm3} binds RNA in vitro in a sequence-independent manner, producing gel-shift patterns that suggest that each protein binds 6 nucleotides of substrate.

Our observations support a model in which multiple \textit{Csm3} molecules bind the crRNA with a 6-nucleotide periodicity to function as a ruler that measures the extent of crRNA maturation within the \textit{S. epidermidis} \textit{Cas10-Csm} complex.

**EXPERIMENTAL PROCEDURES**

**Bacterial Strains and Growth Conditions**—\textit{S. epidermidis} RP62a (22) and \textit{S. aureus} OS2 (23) strains were grown in brain heart infusion medium (Difco). When required, the medium was supplemented with antibiotics as follows: neomycin (15 \(\mu\)g/ml) for selection of \textit{S. epidermidis} LM1680; chloramphenicol (10 \(\mu\)g/ml) for selection of \textit{pcrispr} and \textit{pLM9}-based plasmids; and mupirocin (5 \(\mu\)g/ml) for selection of \textit{pG0400}. \textit{E. coli} BL21(DE3) Codon Plus cells (EMD Millipore) were grown in LB supplemented with chloramphenicol (34 \(\mu\)g/ml). When appropriate, the medium was also supplemented with kanamycin (50 \(\mu\)g/ml) to select for \textit{pET28b}-based plasmids or ampicillin (100 \(\mu\)g/ml) to select for \textit{pET23a}-based plasmids.

**Strain Construction**—\textit{S. epidermidis} LM1680 was isolated as a recipient clone that escaped CRISPR interference after the conjugative transfer of \textit{pG0400} from a donor strain \textit{S. aureus} RN4220 to the recipient \textit{S. epidermidis} RP62a. \textit{pG0400} was cured from \textit{S. epidermidis} LM1680 by multiple passages and replica plating on brain heart infusion agar with and without mupirocin. A large deletion flanking the \textit{pcrispr-cas} locus was mapped by PCR and Sanger sequencing: 257,871 bp are missing from coordinates 2,327,546 to 2,585,416 of the \textit{S. epidermidis} RP62a genome.

**Plasmid Construction**—\textit{pcrispr} was constructed by fusing plasmids \textit{pLM304} (19) and \textit{pC194} (24). Briefly, \textit{pC194} was amplified with primers W175 and W225 (supplemental Table). The PCR product and \textit{pLM304} were digested with AatII (New England Biolabs (NEB)) and BamHI (NEB), and both were gel-purified. The purified fragments were combined and ligated with T4 DNA ligase (NEB). The ligated product was transformed into \textit{S. aureus} OS2 and chloramphenicol-resistant colonies were selected. The \textit{pcrispr-cas} insert was confirmed by a functional test for CRISPR interference against the conjugative transfer of \textit{pG0400}. In-frame deletions, His\textsubscript{10} tags, and amino acid substitutions were introduced into \textit{pcrispr} by inverse PCR using the primers indicated in the supplemental table. Restriction cut sites were included on the primers for generation of \textit{Csm3} deletions (PspOMI and EagI), His\textsubscript{10} tags (Nhel), and amino acid substitutions (Csm3\textsubscript{D120A}, PspOMI; Csm3\textsubscript{T120A, E124A}, Agel) to facilitate ligation. Following inverse PCR, products were purified using a PCR purification kit (Qiagen) and were either 5′-phosphorylated using polynucleotide kinase (for blunt-end ligation) or cleaved with the appropriate restriction enzyme(s) (NEB). Restriction digests were heat-inactivated according to the manufacturer’s recommendations. Digested or phosphorylated PCR products were then circularized using T4 DNA ligase (NEB). All constructs were transformed into \textit{S. aureus} OS2 prior to transformation into \textit{S. epidermidis} LM1680. At least two chloramphenicol-resistant transformants were selected for each construct, and plasmids were purified. To confirm the intended mutations, plasmids were subject to PCR amplification using the appropriate primers flanking the mutation site (supplemental table). Deletion mutants were confirmed by sequencing of the entire CRISPR locus using primers L19, W13, W14, L35, W15, W16, W17, T17, and W19 (supplemental table), and His\textsubscript{10} tags and amino acid substitution mutations were confirmed by restriction digest with the appropriate enzyme and/or sequencing of the mutated region. At least two isolates of confirmed mutant plasmids were prepared from \textit{S. aureus} and transformed into \textit{S. epidermidis} LM1680.

The \textit{csm3-csm3} (tethered) mutant was constructed by the insertion of a region encoding a (GGGGS)\textsubscript{3} linker followed by a \textit{csm3} sequence downstream of the original copy of \textit{csm3} in \textit{pcrispr}. Plasmid and insert were ligated using Gibson assembly (25). The construct was transformed into \textit{S. aureus} OS2 prior to transformation into \textit{S. epidermidis} LM1680. At least two chloramphenicol-resistant transformants were selected, and plasmids were purified. To confirm the intended mutation, plasmids were subjected to PCR amplification and sequencing of the PCR product using the appropriate primers flanking the mutation site (supplemental table). \textit{pET23a-Cas10/Csm/csm3\textsubscript{216GN}} was constructed by amplification of \textit{Pcrispr/csm3\textsubscript{216GN}} and \textit{pET23a} (Novagen) with primers PS85/PS86 and PS87/PS88, respectively (supplemental table). The PCR products were ligated using Gibson assembly (NEB), and the identity of the insert was confirmed by sequencing with primers T7T and T7P (supplemental table).

\textit{pET28b-His\textsubscript{10}Smt3-csm3} was constructed by inserting an \textit{S. epidermidis} \textit{csm3} PCR product into the \textit{pET28b-His\textsubscript{10}Smt3} multiple cloning site. Briefly, \textit{csm3} was amplified from \textit{pcrispr-cas} with primers A132 and A133 (supplemental table), and both the PCR product and \textit{pET28b-His\textsubscript{10}Smt3} were digested with BamHI (NEB) and XhoI (NEB). The digested PCR product and linearized vector were gel-purified, combined, and ligated by T4 DNA Ligase (NEB) and then transformed into \textit{E. coli} BL21 (DE3) codon plus cells (EMD Millipore). The identity of the
cloned DNA fragment was confirmed by sequencing using primers T7P and T7T (see supplemental table).

pLM9 was constructed by adding the lacI repressor gene and the pSpac promoter region of pMutin-HA (26) into the HindIII site of pC194 (24). pLM9/csm3 and pLM9/csm3DIGO were created by ligating the different csm alleles into the multiple cloning site of pLM9 using Gibson assembly (25).

Conjugation—Conjugation was carried out by filter mating as described previously (3). Confirmation of the presence of the desired plasmids in transconjugants was achieved by extracting DNA of at least two colonies and performing PCR with suitable primers (L70/L71 to confirm pG0400 and various primers as specified in the supplemental table to confirm persir mutations).

**Cas10-Csm Purification from S. epidermidis—*S. epidermidis* LM1680 strains harboring persir with the indicated His8 tag were grown to A660 of 2 and harvested. Cell pellets were frozen at −80 °C overnight and then thawed on ice for ~1 h. Cell pellets were resuspended in 10 ml of lysis buffer (20 mM MgCl2, 35 μg/ml lysozyme) and incubated at 37 °C for 30 min. Cell lysates were then diluted 1:1 with 2× resuspension buffer (600 mM NaCl, 100 mM NaH2PO4, 20 mM imidazole, 0.1% Triton X-100, and 2 complete EDTA-free protease inhibitor tablets (Roche Applied Science). Diluted lysates were sonicated on ice (large probe, power of 10), with four 30-s pulses and 30 s of rest in between pulses. Lysates were centrifuged at 15,000 × g for 20 min (twice) to pellet insoluble cell debris, and cleared lysates were filtered through a 0.2-μm bottle top filter. A column (1-cm diameter) was packed with nickel-nitrilotriacetic acid resin (Thermo, 1 ml of slurry per 1 liter of starting culture volume) and pre-equilibrated with 10 ml pre-equilibration buffer (100 mM NaCl, 50 mM NaH2PO4). Cleared lysate was applied to the column, and the column was washed with 10 ml of wash buffer 1 (100 mM NaCl, 50 mM NaH2PO4, 20 mM imidazole), and 10 ml of wash buffer 2 (100 mM NaCl, 50 mM NaH2PO4, 20 mM imidazole, 10% glycerol). Protein was eluted from the column with 4 ml of elution buffer (300 mM NaCl, 50 mM NaH2PO4, 250 mM imidazole, and 10% glycerol), and 500-μl fractions were collected. Fractions containing Cas10-Csm were pooled and either dialyzed in 1 liter of dialysis buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, and 25% glycerol) at 4 °C overnight, or where indicated, subjected to a second analytical affinity purification as follows: streptavidin-coated magnetic beads (15 μl per sample, Thermo) were equilibrated by washing three times (100 μl per wash) with wash buffer 3 (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, and 5% glycerol). Beads were incubated with 2 ng of 5′-biotinylated oligonucleotide antisense to *spe1* (see the supplemental table for sequence) for 30 min at room temperature. Beads were washed three times with wash buffer 3 (100 μl per wash) and mixed with 10 μg of eluted Cas10-Csm complex. Complexes were allowed to anneal to the antisense oligonucleotide for 30 min at room temperature, and the beads were washed three times with wash buffer 3 (100 μl per wash) and resuspended in a final volume of 25 μl of wash buffer 3. Beads were boiled in an equal volume of 2× protein loading buffer (4% SDS, 20% glycerol, 120 mM Tris-HCl, pH 7, and 2% bromophenol blue) and resolved by SDS-PAGE on 4−15% gradient gels (Bio-Rad) in 1× protein running buffer (25 mM Tris, 192 mM glycine).

**Conjugation and Csm Purification from E. coli—** Cultures (4 liters) of *E. coli* BL21 (DE3) codon plus cells (EMD Millipore) containing pET23a-Cas10/Csm2His6 or pET28b-His6Smt3-Csm3 were grown at 37 °C in Luria-Bertani medium containing 34 μg/ml chloramphenicol and either 100 μg/ml ampicillin or 50 μg/ml kanamycin, respectively. When the A660 reached 0.6, the cultures were adjusted to 0.3 mM isopropyl-1-thio-β-D-galactopyranoside, and incubation was continued for 16 h at 17 °C with constant shaking. The cells were harvested by centrifugation, and the pellets were stored at −80 °C. All subsequent steps were performed at 4 °C.

For purification of the Cas10-Csm(Csm2His6) complex, thawed bacteria were resuspended in 75 ml of buffer A1 (50 mM Tris-HCl, pH 7.5, 350 mM NaCl, 200 mM Li2SO4, 20% sucrose, and 10 mM imidazole) containing two complete EDTA-free protease inhibitor tablets (Roche Applied Science). Triton X-100 and lysozyme were added to final concentrations of 0.1% and 0.1 mg/ml, respectively. After 1 h, the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation for 30 min at 15,000 rpm in a Beckman JA-3050 rotor. The soluble extract was mixed for 1 h with 5 ml of Ni2+-nitrilotriacetic acid-resin (Qiagen) that had been pre-equilibrated with buffer A1. The resin was recovered by centrifugation and then first washed with 50 ml of buffer A1, followed by washing with 50 ml of IMAC buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 10% glycerol) containing 15 mM imidazole. The resin was subsequently resuspended in 10 ml of IMAC buffer containing 50 mM imidazole and then poured into a column. The column was then eluted stepwise with 10 ml aliquots of IMAC buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 10% glycerol) containing 100, 200, 350, and 500 mM imidazole. The 100 mM imidazole elutes containing the complex was pooled together. Subsequently, 0.5 ml of the fraction was analyzed by gel filtration chromatography with a Superdex 200 10/300 GL (GE Healthcare) using buffer B (50 mM Tris-HCl, pH 7.5, 5% glycerol, 150 mM NaCl). The protein complex elution profile was monitored continuously by UV absorbance, and A280 was plotted as a function of elution volume. The void volume of the column was measured by tracking the elution peak of blue dextran, and the column was calibrated by tracking the elution profiles of marker proteins of known native size (Bio-Rad gel filtration standards).

For purification of Csm3, thawed bacteria were resuspended in 30 ml of buffer A2 (50 mM Tris-HCl, pH 7.5, 1.25 mM NaCl, 200 mM Li2SO4, 10% sucrose, 15 mM imidazole) containing one complete EDTA-free protease inhibitor tablet (Roche Applied Science). Triton X-100 and lysozyme were added to final concentrations of 0.1% and 0.1 mg/ml, respectively. After 1 h, the lysate was sonicated to reduce viscosity. Insoluble material was removed by centrifugation for 30 min at 15,000 rpm in a Beckman JA-3050 rotor. The soluble extract was mixed for 1 h with 2 ml of Ni2+-nitrilotriacetic acid-agarose resin (Qiagen) that had been pre-equilibrated with buffer A2. The resin was recovered by centrifugation and then first washed with 40 ml of buffer A2, followed by washing with 5 ml of 3 M KCl solution. The resin was subsequently resuspended in 25 ml of buffer A2 and then poured into a column. The column was then eluted stepwise with 3-ml aliquots of IMAC buffer (50 mM Tris-HCl, pH 7.5, 250 mM NaCl, 10% glycerol) containing 50, 100, 200,
and 500 mM imidazole. The 200 mM imidazole elutes containing the His$_6$-Smt3-Csm3 polypeptide were pooled and dialyzed for 3 h in dialysis buffer (25 mM Tris-HCl, pH 7.5, 150 mM NaCl, 15 mM imidazole, 5% glycerol) containing 5 units of SUMO protease (27). Dialysis was then incubated with Ni$^{2+}$-nitriilo-triacetic acid-agarose resin pre-equilibrated in the dialysis buffer for 1 h at 4 °C. The beads and dialysate mixture were passed through a column, and the flow-through containing purified Csm3 was collected. Protein concentration was determined by using the Bio-Rad dye reagent with BSA as the standard and confirmed by densitometry after resolving Csm3 and the BSA standards on a gel.

**Csm3 Binding Assay**—Purified Csm3 (0–384 pmol) and trace amounts of 5’-end labeled PAGE-purified RNA substrates were combined in RNA binding buffer (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 5 mM DTT, and 2% glycerol) and incubated at room temperature for 10 min. The samples were resolved at 4 °C on an 8% native polyacrylamide gel in TBE buffer (90 mM Tris-borate, 2 mM EDTA, pH 8.3).

**Western Blotting**—Dialyzed Ni$^{2+}$ affinity-purified proteins (10 μg) were resolved by SDS-PAGE on 4–15% gradient gels (Bio-Rad) and electroblotted onto Immobilon-P PVDF membranes (Millipore) in ice-cold transfer buffer (25 mM Tris, 192 mM glycine, and 20% methanol) at 4 °C by applying 100 V for 1 h. Membranes were blocked for 1 h in TBS-T (50 mM Tris, 150 mM NaCl, 0.1% Tween-20) containing 2.5% BSA, nickel-horse-radish peroxidase (Ni-HRP, Thermo) was added (1:5000 dilution), and the membrane was incubated for an additional 1 h. Membranes were washed three times with TBS-T, and nickel HRP was visualized using ECL-based chemiluminescence.

**Cas10-Csm Nuclease Assay**—The 71-nt crRNA substrates (sp1, 5’-ACGAGAACACGUAUGCGAGAUAUAAAUCAUCAGAAGAAAAGGAUCGUAUACCACCCCGAAGAAGGGG-3’; sp2, 5’-ACGAGAACTAGTAATAATTGTC-ATTCTACGATCTATATCGATGCUAGUACUCCACCCCGAAGAAGGGG-3’; and sp3, 5’-ACGAGAACTAGTACCGGTCGTGAACATTTTTTCTTGATTCTCTGAUCGAU-ACCCACCCCGAAGAAGGGG3’) were 5’-end-labeled, PAGE-purified, and combined with the Cas10-Csm complex (500 nM) in nuclease buffer (50 mM Tris-HCl, pH 7.5, 50 mM KCl, and 10 mM DTT) containing the indicated metal (10 mM) or EDTA (5 mM). The nuclease reaction was allowed to proceed at 37 °C for 20 min, and the RNA was ethanol-precipitated and resolved using denaturing PAGE.

**CRISPR RNA Capture**—Sp1 crRNAs were captured from 50 μg of total RNA extract using a 5’-biotinylated PAGE-purified oligonucleotide antisense to sp1 (supplemental table) as described previously (19).

**Mass Spectrometry Analysis**—Pooled fractions of freshly purified Cas10-Csm complexes (~100 μg) were submitted to the Proteomics Resource Center at Rockefeller University for Mass Spectrometry, and data analysis was done with a nano-ESI LTQ Orbitrap XL instrument.

**RESULTS**

**Mature crRNAs Are Measured in a Cas10-Csm Ribonucleoprotein Complex**—Our previous genetic characterization of the *S. epidermidis* CRISPR-Cas system showed that isogenic strains lacking csm2, csm3, or csm5 accumulated intermediate crRNAs but were lacking in the mature species (19). These genes were candidates to encode proteins involved in crRNA maturation. In types I and III CRISPR-Cas systems, different cas genes were candidates to encode proteins involved in crRNA maturation. In types I and III CRISPR-Cas systems, different ribonucleoprotein complexes (2, 5, 16, 20, 21). Mass spectrometry also confirmed the identity of its members. Mass spectrometry also revealed the presence of different crRNA species with lengths corresponding to the 71-nt intermediate crRNA as well as a series of mature crRNA species separated by 6-nt intervals (31, 37, 43, 50).

**Table 1**

| Strain$^a$ | Recipients$^b$ | Transconjugants$^b$ | Conjugation efficiency$^p$ |
|---|---|---|---|
| pcrispr | 1.8 × 10$^4$ | 2.0 × 10$^3$ | 1.1 × 10$^{-6}$ |
| pcrispr/ΔR/S | 4.0 × 10$^4$ | 3.3 × 10$^3$ | 8.3 × 10$^{-5}$ |
| pcrispr/cas10H11001 | 5.0 × 10$^4$ | 1.0 × 10$^3$ | 2.0 × 10$^{-6}$ |
| pcrispr/csm2H11002 | 1.3 × 10$^5$ | 1.0 × 10$^3$ | 7.8 × 10$^{-7}$ |
| pcrispr/csm4H11002 | 2.0 × 10$^5$ | 5.0 × 10$^3$ | 2.5 × 10$^{-7}$ |
| pcrispr/csm5H11002 | 3.0 × 10$^5$ | 2.0 × 10$^4$ | 6.7 × 10$^{-6}$ |
| pcrispr/csm6H11002 | 2.5 × 10$^5$ | 1.0 × 10$^4$ | 4.0 × 10$^{-6}$ |
| pcrispr/csm7H11002 | 2.0 × 10$^5$ | 4.0 × 10$^4$ | 2.0 × 10$^{-6}$ |
| pcrispr/csm31103A | 6.0 × 10$^5$ | 1.0 × 10$^5$ | 1.7 × 10$^{-7}$ |
| pcrispr/csm41103A | 7.0 × 10$^5$ | 1.0 × 10$^5$ | 1.4 × 10$^{-7}$ |
| pcrispr/csm51103A | 3.0 × 10$^5$ | 3.0 × 10$^4$ | 1.0 × 10$^{-6}$ |
| pcrispr/csm3-csm3 | 2.4 × 10$^5$ | 1.0 × 10$^4$ | 4.2 × 10$^{-5}$ |

$^a$H6N indicates a His$_6$-N-terminal tag; H6C indicates a His$_6$-C-terminal tag.

$^b$Results from one representative trial are shown. Conjugation efficiency is expressed as the ratio transconjugants/recipients. pcrispr and pcrispr/ΔR/S constitute positive and negative controls, respectively, for CRISPR immunity against plasmid conjugation.
Csm3 Works as a Ruler to Determine crRNA Length

FIGURE 1. Mature crRNAs are measured in a Cas10-Csm complex. A, organization of the type III-A CRISPR system in S. epidermidis RP62A. This system contains 9 CRISPR-associated (cas and csm) genes, 4 direct repeats (black boxes), and 3 spacers (colored boxes), the first of which targets the nickase gene in staphylococcal conjugative plasmids. B, a ruler mechanism determines the length of mature crRNAs. Transcription of the repeat-spacer array generates a precursor crRNA that is subject to two cleavage events: primary processing within repeats to yield ~71-nt intermediates (filled triangles), and maturation through trimming of the 3' end of the intermediate (empty triangles). A ruler mechanism anchored at the primary processing site determines the extension of maturation to generate 37- and 43-nt-long mature crRNA. C, the type III-A Cas10-Csm complex. His6 tags were placed on the indicated (N or C) terminus of each of the genes involved in crRNA biogenesis. Constructs were expressed in S. epidermidis LM1680, and whole cell lysates were subject to Ni²⁺ affinity chromatography, followed by size exclusion with a Superdex 200 column. UV absorbances at A₂₈₀ (red) and A₅₅₀ (blue) are shown in miliabsorbance units (mAU) as a function of elution volume. Molecular weight estimates for each peak are indicated. B, protein content for the load (L) and indicated 1-ml fractions (12, 17, and 19) resolved by SDS-PAGE and visualized using Coomassie G-250 staining. C, crRNAs associated with Cas10-Csm (Csm2H6N) complexes extracted from S. epidermidis (S) and E. coli (E) shown. crRNAs extracted from purified complexes were radio-labeled on their 5' end and resolved using denaturing PAGE.

TABLE 2

| Protein* | Peak area* | Theoretical mass |
|----------|------------|------------------|
| Cas10    | 17 ± 0.7   | 87,581           |
| Csm2     | 12 ± 0.9   | 15,411           |
| Csm3     | 32 ± 7.5   | 24,001           |
| Csm4     | 5 ± 0.5    | 34,512           |
| Csm5     | 4 ± 0.3    | 39,390           |
| Cas6     | 0.03 ± 0.009 | 28,935        |

* CRISPR-associated protein hits are indicated.

Peak area value represents an average of three independent measurements.

To further characterize the complex, we overexpressed and purified the His₆-Csm2 complex in E. coli and subjected it to size-exclusion chromatography (Fig. 2, A and B). The elution profile demonstrated a strong association between the members of the complex and estimates a molecular mass of 331 kDa. Assuming each subunit of the complex exists in a single copy (the theoretical mass of each protein is shown in Table 2), and allowing for a maximum of 25 kDa additional for the crRNA, the mass of such complex will add up to 255 kDa (i.e. over 70 kDa of mass remains unaccounted for). The higher molecular weight estimated after gel filtration could be due to a particular stoichiometry or to a non-globular shape that impacts on the apparent molecular weight of the complex. crRNAs extracted from the E. coli-expressed complex maintain their 6-nucleotide periodicity (Fig. 2C); however, the overall length of mature crRNAs in the complex are shortened by 2 nucleotides when compared with those produced in S. epidermidis. We hypothesize that different host-encoded nucleases might be responsible for the maturation cleavage event.
Csm3 Works as a Ruler to Determine crRNA Length

Nonetheless, we tested whether the Cas10-Csm complex itself performs maturation of intermediate crRNAs. We purified complexes from *S. epidermidis* using a Δcas6 construct (Fig. 3A) because Cas6 is considered the primary processing endonuclease for type I and III CRISPR-Cas systems (2, 13–15) and therefore the obtained complexes do not contain any endonuclease for type I and III CRISPR-Cas systems (2, 13–15) that are capable of processing the mature crRNAs. To test this, we overexpressed and purified Csm3 from *S. epidermidis* sequence alignment using BLAST. Highly conserved residues are highlighted with asterisks. B, mutations in conserved Csm3 residues do not affect CRISPR interference against the conjugal plasmid pG0400. *S. epidermidis* LM1680 strains harboring wild-type and mutant *pcrisr* plasmids with the indicated alanine substitutions in Csm3 were used as recipients for the transfer of pG0400. Conjugation was carried out in duplicate; the values (in cfu/ml; mean ± S.D.) obtained for recipients and transconjugants are shown. C, wild type and mutant Cas10-Csm3 complexes were purified from *S. epidermidis* cas6 construct in *S. epidermidis* LM1680 strains harboring wild type and Δcas6 strains, respectively. Asp-100, Glu-120, and Glu-124 (Fig. 4B) served acidic residues: Asp-100, Glu-120, and Glu-124 (Fig. 4B) expected to have a catalytic role (28), as well as three highly conserved acidic residues: Asp-100, Glu-120, and Glu-124 (Fig. 4B). Mutation of these residues either individually or in combination had no effect on CRISPR function (Fig. 4B and Table 1). We purified complexes carrying these mutations and examined the crRNA distributions. Whereas the H18A complexes produced wild-type levels and sizes of crRNAs, the E120A,E124A substitutions caused the appearance of crRNA species between 31 and 37 nucleotides, and the D100A substitution led to the elimination of all crRNAs except the 31-nt species.

**Multiple Copies of Csm3 Bind the crRNA with a 6-Nucleotide Periodicity**—A direct role of Csm3 in the modulation of crRNA size distribution, the D100A substitution leads to the elimination of all crRNAs except the 31-nt species. The 6-nt periodicity of crRNAs, however, is inherent to the Cas10-Csm complex, indicating that one or more of its components acts as a ruler to determine the length of the mature crRNAs.

**Csm3 Modulates the Length of Mature crRNAs**—Complexes carrying the His₉ tag in different proteins showed the same crRNA pattern with the exception of the His₉-Csm3 complex, which displayed a crRNA size distribution that is shifted to shorter lengths (37 and 31 nt) (Fig. 1E). This observation suggested that the addition of the tag affects Csm3 activity or association with the rest of the members of the complex in a manner that impacts the extent of crRNA maturation. To investigate this possibility, we performed alanine substitutions in conserved residues in Csm3 and looked for mutants that affect the mature crRNA sizes. We mutated a conserved histidine (His-18) suspected to have a catalytic role (28), as well as three highly conserved acidic residues: Asp-100, Glu-120, and Glu-124 (Fig. 4A). Mutation of these residues either individually or in combination had no effect on CRISPR function (Fig. 4B and Table 1). We purified complexes carrying these mutations and examined the crRNA distributions. Whereas the H18A complexes produced wild-type levels and sizes of crRNAs, the E120A,E124A substitutions caused the appearance of crRNA species between 31 and 37 nucleotides, and the D100A substitution caused mature crRNAs to collapse down to the smallest detectable species, 31 nt (Fig. 4C). These results indicate that Csm3 has an important role in the measurement of the length of mature crRNAs.
Csm3 Works as a Ruler to Determine crRNA Length

We tested RNA molecules of 12, 18, and 33 nucleotides in length and obtained 2, 3, and 5 shifted bands, respectively. This suggests that each Csm3 protein interacts with 6 nt of the crRNA. This interaction did not depend on the sequence of the RNA substrate, as similar results were obtained for an RNA with a random sequence (Fig. 5, B and D). We were unable to purify enough quantities of Csm3D100A to perform substrate-binding studies. This protein was highly insoluble, suggesting that the D100A mutation might affect the folding and/or stability of Csm3 when overexpressed in E. coli.

The above results lead us to hypothesize that Csm3 could bind the crRNAs associated with the Cas10/Csm complex and determine their length. Csm3 could have two possible functions: to protect the 3’ end of the crRNA from nuclease activity or to specify a cleavage site within the crRNA. To distinguish between these situations, we engineered complexes carrying a Csm3-Csm3 dimer in which both proteins are tethered by a (GGGGS)₃ flexible linker. The presence of a dimer will change the periodicity of crRNA lengths from 6 to 12 in the protection scenario, but will remain 6 nt if Csm3 specifies the cleavage site. We purified complexes containing the dimer (Fig. 6A) and looked for their associated crRNA species (Fig. 6B). Consistent with Csm3 protection of the crRNA, the 31-nt species disappeared, and the smallest mature crRNA species in the Csm3-Csm3 complex was 37 nt. Similar results were obtained by capturing spc1 crRNAs from total RNA extracts (Fig. 6C).

According to the protection hypothesis, the next crRNA species in the Csm3/Csm3 complexes should be 49 nt (37 + 12 nt). CrRNA species larger than 37 nt, however, were not observed, and complexes containing the dimer were not able to execute full CRISPR immunity against the pG0400 plasmid (Table 1). We believe that although these complexes can be formed and purified, they might adopt an artificial configuration or stoichiometry, precluding an unequivocal interpretation of these experiments. Despite these caveats, the in vitro and in vivo results taken together confirm that each Csm3 molecule inter-
acts with 6 nt of the crRNA and likely protects it from the nuclease involved in maturation.

Csm3 Levels Determine the Extent of crRNA Maturation—Next, we wanted to test the role of Csm3 in the generation of multiple crRNA species. Csm3 is an ortholog of CasC (13), which has been shown to exist in six copies per Cascade complex in the type I-E system of E. coli (29). We hypothesized that multiple copies of Csm3 might exist within the Cas10-Csm complex and could bind the crRNA with a periodicity of 6-nt to protect it from the nuclease that cleaves or degrades the 3’ end of the intermediate crRNA (Fig. 7A). If this is the case, complexes lacking Csm3 should contain very short or no crRNAs, whereas extra copies of Csm3 should result in larger sizes for mature crRNAs. We attempted the purification of Cas10-Csm complexes from strains lacking Csm3. Absence of Csm3 not only prevented the formation of the complex but also the detection of the other members of the Cas10-Csm complex. Regardless of which complex member carried the His6 tag, we were unable not only to purify a soluble subcomplex but also to detect the tagged protein in cell extracts by Western blot (Fig. 7B). In Pseudomonas aeruginosa, the crRNA is required for the assembly of the Csy complex (20). Therefore, we hypothesized that because Csm3 can bind the crRNA, the lack of Csm3 could prevent the recruitment of the crRNA into the Cas10-Csm complex, thus preventing its assembly. To test whether the crRNA is indeed required for the formation of the complex, we purified it from a strain deleted for the repeat/spacer sequences. The Cas10-Csm complex did form in the absence of crRNAs (Fig. 7, C and D), a result that rejected our hypothesis. These findings suggest Csm3 is an essential member of the core complex and that in its absence, the Cas10-Csm complex is unstable, and its members are degraded. Alternatively, although unlikely, Csm3 could be required for the expression of the cas/csm genes.

To test the effect of additional copies of Csm3 within the Cas10-Csm complex, we cloned the csm3 and csm3D100A mutant genes under an isopropyl-β-D-galactopyranoside-inducible promoter in the S. epidermidis plasmid pLM9, creating pLM9/csm3 and pLM9/csm3D100A, respectively. We then induced expression of the different csm3 alleles in wild-type S. epidermidis RP62A (a wild-type strain expressing all the cas/csm genes from the chromosomal operon) and captured spc1 crRNAs from cellular extracts with a biotinylated antisense probe. Consistent with our model, overexpression of Csm3 resulted in a shift in spc1 crRNA sizes toward larger species (49, 55, and 61 nt) and in the disappearance of the smallest detectable size in this pulldown assay (37 nt, Fig. 7E). In contrast, overexpression of Csm3D100A caused no detectable shift and was undistinguishable from an empty vector control. Altogether, these data demonstrate that Csm3 plays a fundamental role in determining the extent of crRNA maturation. Higher levels of this protein result in longer crRNAs, suggesting that Csm3 acts as a ruler that sets the lengths of the Cas10-Csm complex-bound crRNAs. We believe that alterations of Csm3 that perturb the folding and/or result in a reduced ability to bind the crRNA, such as the addition of an N-terminal His tag or the D100A mutation, result in shorter mature crRNA species.

DISCUSSION

Here, we show that a Cas10-Csm ribonucleoprotein complex, similar to the E. coli Cascade (2) and Pyrococcus furiosus and Sulfolobus solfataricus (5, 21) Cmr complexes, mediates type III-A CRISPR immunity in staphylococci. This complex is com-
posed of Cas10, Csm2, Csm3, Csm4, and Csm5 and mature crRNAs that, through base pair interactions with a cognate DNA sequence, guide the complex to its target on the genome of plasmids and bacteriophages. Unexpectedly, multiple mature crRNA species differing by 6-nt increments at the 3’ end are present in this complex. Previously, we demonstrated that the final length of mature crRNAs is determined by a ruler mechanism that measures from the 5’ end primary processing cleavage site (19). Here, we demonstrate that Csm3 is the ruler that determines the mature crRNA length. Csm3 is essential for the formation of the complex and seems to be present in multiple copies. In vitro, multiple copies of Csm3 bind the RNA in a sequence-independent manner, one protein every 6 nt of substrate. Alanine substitution of a conserved aspartate residue, Asp-100, impacts Csm3 folding and/or ability to bind the crRNA and prevents the accumulation of longer mature crRNA species. However, overexpression of Csm3 leads to the accumulation of longer species. Altogether, these results allow us to propose that Csm3 binds to the crRNAs in the complex at multiple sites, once every 6 nt, with each additional copy extending the crRNA length by 6 nt.

The nuclease involved in the biogenesis of crRNAs remain to be determined. In other type III and type I CRISPR-Cas systems, primary processing is carried out by Cas6 (2, 15, 30), which cleaves within repeats at the base of a hairpin and defines the 5’ end for all crRNAs. Therefore, S. epidermidis Cas6 is a strong candidate to cleave the crRNA precursor into 71-nt intermediates. The identity of the nuclease responsible for crRNA maturation in type III CRISPR-Cas systems is less clear. The fact that the same type III-A complex, when expressed in S. epidermidis and E. coli, produce mature species of different lengths and that this complex is unable to direct the maturation of a crRNA substrate in vitro, suggests that a host-encoded endo- or exoribonuclease might be responsible for the degradation of the 3’ end of crRNA intermediates. Several ribonucleases are annotated in the S. epidermidis RP62a genome (22), RNase H-II and -III, RNase III, RNase II/VacB and RNase R, RNase BN, RNase P, CbfI, and YigF. Any of these could participate in crRNA maturation. Future biochemical and genetic experiments will determine the role of host RNases in crRNA maturation.

The three different types of CRISPR-Cas systems can be distinguished according to their crRNA biogenesis pathways. Type II systems display a distinct mechanism for the generation of crRNAs that requires the pairing of the precursor crRNA with an antisense trans-encoded crRNA to ensure RNase III cleavage of the precursor (17). Type I and III systems, however, have comparable crRNA biogenesis pathways; the main difference being the lack of further processing of the intermediate crRNAs in the former. This may be attributed to the different properties of the type I Cas6 homologs. In these systems, the endonuclease is part of the Cascade complex (2, 7, 29) and remains tightly bound to the 3’ end of its product after cleavage (15). The strong and continuous protection of the crRNA by the type I Cas6 homologs may prevent further nucleolysis of the 3’ end. In contrast, Cas6 is not part of the S. epidermidis type III-A ribonucleoprotein complex or of the well characterized type III-B P. furiosus and S. solfataricus Cmr complex, which also contain mature crRNAs differing by 6-nt increments at the 3’ end (5, 18, 21, 31). Due to its homology to Csm3, we suspect that Cmr6 acts as a ruler of the mature crRNA length in P. furiosus.

It remains to be known whether the precise measurement of crRNA length by Csm3 is important for CRISPR immunity. One possibility is that the extra trimming of crRNAs is caused by rearrangements in the crRNA:Cas-Csm complex that expose these RNAs to host nucleases and has no consequences for CRISPR immunity. Alternatively, crRNA maturation could be necessary for CRISPR-Cas function. In this scenario, it is possible that repeat sequences at the 3’ end of an intermediate crRNA, which do not anneal with the target and form stem-loop structures, would interfere with target recognition and/or cleavage. In fact, in the type III-B system of P. furiosus, the 3’ end of the mature crRNA determines the cleavage site on the target sequence exactly 14 nt upstream of this end (5); such precise cleavage may not be possible with a guide crRNA containing extra sequences at the 3’ end. Equally unknown is the importance, if any, of the presence of multiple crRNA species. If indeed the target cleavage site of type III CRISPR-Cas systems is determined by the mature 3’ end of the crRNA, then multiple mature crRNA sizes could direct multiple target cleavage events. The advantage of this would be that targets cleaved by complexes containing longer crRNAs could be subjected to a second cleavage event directed by shorter crRNAs, thus providing less chance for repair and stronger immunity. Our results argue that this may not be the case since we demonstrated that the Csm3D100A complex, which produces only the shortest crRNA (31 nt), confers full immunity against plasmid conjugation. In this system, a perfect match exists between the spc1 crRNA and its targeted sequence, the protospacer. The possibility exists that longer crRNAs may be required when there is an imperfect match between the crRNA and the protospacer. In the latter scenario, longer crRNAs would increase the likelihood of establishing sufficient length of complementarity to facilitate interference. The 31 nt-long crRNA contains 8 nt of repeat sequences at the 5’ end plus 23 nt of spacer sequence at the 3’ end. Because the full-length spacer is 36 nt, this result also shows that the 13 nt at the 3’ end of the spacer sequence are not required to specify the target of CRISPR immunity. This suggests that if there is a “seed sequence” in type III systems, a region of homology between the crRNA and the target absolutely required for CRISPR immunity (20, 32), it must be located at the 5’ end of the crRNA target. Deciphering the mechanisms of crRNA biogenesis and targeting will be important both to understand how this immune system prevents infection in prokaryotes as well as to exploit it for biotechnological applications (33–37).

Acknowledgments—We thank members of our laboratory for critical discussion of the paper.

REFERENCES

1. 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

2. 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) Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964

3. Marrafini, L. A., and Sontheimer, E. J. (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843–1845

4. Garneau, J. E., Dupuis, M. É., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A. H., and Moineau, S. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71

5. Cady, K. C., Bondy-Denomy, J., Heussler, G. E., Davidson, A. R., and Achman, H. (2006) The repetitive DNA elements called piRNA in bacterial genomes. Cell 127, 257–267

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

7. Doudna, J. A. (2013) CRISPR as a tool for the study of epigenetics. Science 341, 1223–1226

8. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas9 system. Science 339, 819–823

Csm3 Works as a Ruler to Determine crRNA Length

9. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) Multiplex genome editing of bacterial genomes using CRISPR-Cas systems. Mol. Cell 49, 993–1003

10. Haurwitz, R. E., Jinek, M., Doudna, J. A., and Charpentier, E. (2012) Cas9 protein processes pre-crRNA and assembles into a CRISPR RNA-guided endonuclease. Science 337, 1691–1695

11. Godde, J. S., and Bickerton, A. (2006) The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. 62, 718–729

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

13. Doudna, J. A. (2013) CRISPR as a tool for the study of epigenetics. Science 341, 1223–1226

14. Kong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819–823

15. Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. 31, 233–239

16. Mali, P., Yang, L., Essel, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. Science 339, 823–826

17. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173–1183