



# Structural and biochemical insights into CRISPR RNA processing by the Cas5c ribonuclease SMU1763 from *Streptococcus mutans*

Received for publication, April 1, 2021, and in revised form, September 13, 2021. Published, Papers in Press, September 28, 2021.

<https://doi.org/10.1016/j.jbc.2021.101251>

Sofia Lemak<sup>1</sup>, M. Anca Serbanescu<sup>2</sup>, Anna N. Khusnutdinova<sup>1</sup>, Milosz Ruszkowski<sup>3</sup>, Natalia Beloglazova<sup>1</sup>, Xiaohui Xu<sup>1</sup>, Greg Brown<sup>1</sup>, Hong Cui<sup>1</sup>, Kemin Tan<sup>4</sup>, Andrzej Joachimiak<sup>4</sup>, Dennis G. Cvitkovitch<sup>2</sup>, Alexei Savchenko<sup>1</sup>, and Alexander F. Yakunin<sup>1,5,\*</sup>

From the <sup>1</sup>Department of Chemical Engineering and Applied Chemistry and <sup>2</sup>Faculty of Dentistry, Dental Research Institute, University of Toronto, Toronto, Ontario, Canada; <sup>3</sup>Synchrotron Radiation Research Section of MCL, National Cancer Institute, Argonne, Illinois, USA; <sup>4</sup>X-Ray Science Division, Midwest Center for Structural Genomics and Structural Biology Center, Argonne National Laboratory, Argonne, Illinois, USA; <sup>5</sup>Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, UK

Edited by Karin Musier-Forsyth

The cariogenic pathogen *Streptococcus mutans* contains two CRISPR systems (type I-C and type II-A) with the Cas5c protein (SmuCas5c) involved in processing of long CRISPR RNA transcripts (pre-crRNA) containing repeats and spacers to mature crRNA guides. In this study, we determined the crystal structure of SmuCas5c at a resolution of 1.72 Å, which revealed the presence of an N-terminal modified RNA recognition motif and a C-terminal twisted  $\beta$ -sheet domain with four bound sulphate molecules. Analysis of surface charge and residue conservation of the SmuCas5c structure suggested the location of an RNA-binding site in a shallow groove formed by the RNA recognition motif domain with several conserved positively charged residues (Arg39, Lys52, Arg109, Arg127, and Arg134). Purified SmuCas5c exhibited metal-independent ribonuclease activity against single-stranded pre-CRISPR RNAs containing a stem-loop structure with a seven-nucleotide stem and a pentaloop. We found SmuCas5c cleaves substrate RNA within the repeat sequence at a single cleavage site located at the 3'-base of the stem but shows significant tolerance to substrate sequence variations downstream of the cleavage site. Structure-based mutational analysis revealed that the conserved residues Tyr50, Lys120, and His121 comprise the SmuCas5c catalytic residues. In addition, site-directed mutagenesis of positively charged residues Lys52, Arg109, and Arg134 located near the catalytic triad had strong negative effects on the RNase activity of this protein, suggesting that these residues are involved in RNA binding. Taken together, our results reveal functional diversity of Cas5c ribonucleases and provide further insight into the molecular mechanisms of substrate selectivity and activity of these enzymes.

The complex process of the adaptive immunity implemented by the CRISPR and their associated (Cas) genes has been harnessed to revolutionize the fields of molecular biology,

biotechnology, and medicine (1–4). Three main stages of CRISPR immunity include adaptation, RNA maturation, and interference (5, 6). During the initial adaptation stage, new spacer sequences are excised from foreign nucleic acids and inserted into the host CRISPR array, separated by the “repeat” sequences of that locus. In the second stage, the long RNA transcript of the CRISPR locus (pre-crRNA) is specifically cleaved generating individual crRNAs consisting of a single spacer and flanking repeat sequences. These crRNAs bind to Cas proteins to form effector complexes that degrade the corresponding target sequences in the final stage of CRISPR immunity (interference) (7, 8). Although this is the general process of CRISPR immunity, the specific way in which it is achieved varies greatly among different types of CRISPR systems (9).

All CRISPR–Cas systems have been classified into two classes (1 and 2), six types (class 1: types I, III, and IV; class 2: types II, V, and VI), and further into 33 subtypes based on repeat sequences and *cas* operon composition (10–13). The processing of the pre-crRNA in the second step of CRISPR immunity involves more diverse families of proteins (14–19). In type II systems, an endogenous RNase III is responsible for the cleavage of the pre-crRNA/transactivating crRNA duplex, whereas the pre-crRNA processing is mediated by a multi-domain Cas protein in types V (Cas12) and VI (Cas13) (13, 14, 19, 20). In type III systems (A and B), Cas6 performs the crRNA maturation step by recognizing nonpalindromic ssRNA and cutting the pre-crRNA (21–23). This generates products 35 to 53 nt in length that are further processed into an active crRNA and used as part of a larger complex, which does not include Cas6, to target dsDNA or ssRNA (7, 24, 25). Likewise, most type I and IV systems use variants of the Cas6 protein for crRNA maturation. Types I-A and I-B have repeat RNAs similar to that in type III systems and therefore have similarly functioning Cas6a and Cas6b proteins (18, 20, 26, 27). However, the repeats of types I-D, I-E, and I-F form hairpin structures (27), and the Cas6 proteins that

\* For correspondence: Alexander F. Yakunin, [a.yakounine@utoronto.ca](mailto:a.yakounine@utoronto.ca).

## Structure and activity of the Cas5c ribonuclease SMU1763

cleave the pre-crRNA at the 3'-base of the stem-loop (Cas6d, Cas6e, and Cas6f) remain associated with the crRNA as they form the effector complexes. Structures of type I-E and type I-F crRNA-guided complexes demonstrated that these crRNAs are not processed further (9, 28–31).

Type I-C (previously Dvulg) is the only CRISPR subtype that employs a Cas5 variant, Cas5c (originally known as Cas5d), to process pre-crRNA (10). This protein family is unique as it performs the roles of both Cas5 and Cas6 proteins (18, 19, 22, 23). The repeat sequences in this subtype are also predicted to form stable RNA hairpin structures, and the biochemically and structurally characterized Cas5c proteins from *Bacillus halodurans* (BhaCas5c), *Streptococcus pyogenes* (SpyCas5c), *Xanthomonas oryzae* (XorCas5c), and *Desulfovibrio vulgaris* (DvuCas5c) have been shown to cleave at the 3'-base of the stem, with no further processing (32–35). Both Cas5 and Cas6 proteins belong to the repeat-associated mysterious protein (RAMP) superfamily containing one or more RAMP domains with the RNA recognition motif (RRM) fold (23). Cas6 proteins typically have two sequential RAMP domains, whereas Cas5 proteins contain a single N-terminal RAMP domain and a smaller C-terminal  $\beta$ -sheet domain with the RAMP (RRM) domain responsible for substrate binding and cleavage (in Cas5c) (32–34). RNA is cleaved through a metal-independent acid–base hydrolysis, generating 5'-OH and 2',3'-cyclic phosphate products (33). Based on sequence analysis, two subgroups have been proposed within the Cas5c family: the subgroup A (*B. halodurans* and *S. pyogenes*) contains a C-terminal extension (~30 residues), whereas the subgroup B (*X. oryzae*) has an insertion (~23 residues) between the  $\beta$ 3 and  $\beta$ 4 strands (34). The three biochemically characterized Cas5c proteins were also found to exhibit different functional properties. After RNA cleavage, the BhaCas5c protein from *B. halodurans* has been shown to assemble in a complex with the processed crRNA and the Cas8c and Cas7 proteins to form an interference complex (32). Furthermore, it has also been shown to bind and cleave both single-stranded (ss) and double-stranded (ds) DNA (36). Two other Cas5c proteins (SpyCas5c and XorCas5c) have been shown to bind dsDNA nonspecifically but not cleave it (34), underlining the diversity of the Cas5c protein family in both structure and function.

*Streptococcus mutans* is a Gram-positive cariogenic bacterium representing one of the major pathogens implicated in human dental caries (37). The *S. mutans* UA159 genome reveals the presence of two CRISPR loci, CRISPR1 (type II-A) and CRISPR2 (type I-C), with several spacers complementary or highly similar to sequences in genomes of the phage M102 or other *S. mutans* strains (38–40) (Fig. 1A). The CRISPR2 cluster contains only one spacer and a degenerate second repeat suggesting that it might be nonfunctional. However, both CRISPR loci appear to be rigorously maintained in *S. mutans* being present in most clinical and laboratory strains (50–70%) (40). Recently, the presence of both CRISPR1 and CRISPR2 loci in clinical isolates of *S. mutans* was also found to be correlated with thicker biofilm production (40). In our previous work, we demonstrated that the expression of both

CRISPR systems in the *S. mutans* UA159 cells is differentially modulated by the two-component signal transduction system VicR/VicK (41). Moreover, further studies with *cas* deletion mutants revealed that although deletions had no effect on *S. mutans* acid production or tolerance, both CRISPR systems appear to contribute to the stress response, whereas the CRISPR1 system is also involved in resistance against incoming plasmids containing target sequences (40, 41).

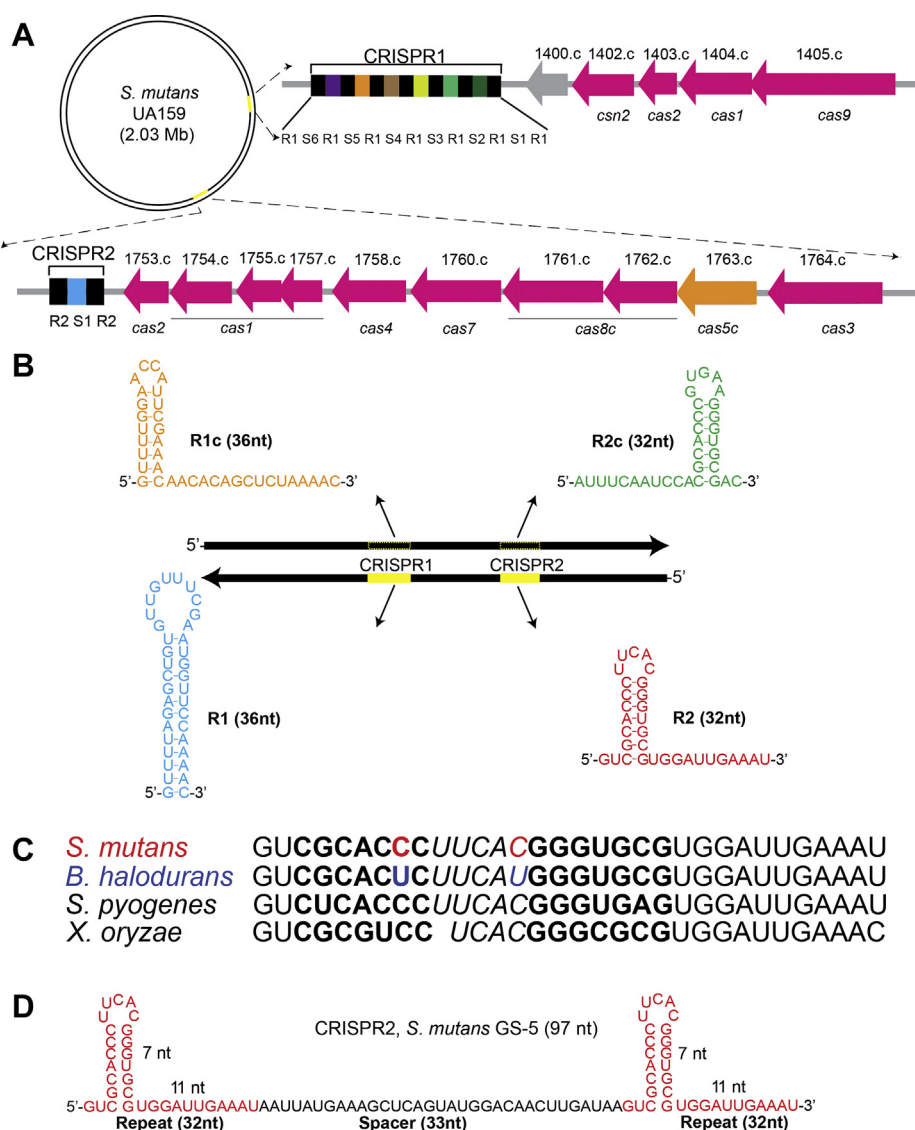
In the *S. mutans* genome, the Cas5c gene (SMU1763) is associated with the CRISPR2 locus (type I-C), which is characterized by the presence of a single spacer and several *Cas* genes: *cas1*, *cas2*, *cas3*, *cas4*, *cas5c*, *cas8c*, and *cas7* (Fig. 1A). The purified SmuCas5c protein has been shown to be catalytically active *in vitro* (41). However, the molecular mechanisms of substrate specificity and pre-crRNA cleavage of SmuCas5c remain unknown. In addition, significant variations have been reported recently regarding the biochemical and structural properties of BhaCas5c, SpyCas5c, and XorCas5c (19, 32–34, 36). To provide further insights into catalytic activities and functional diversity of the Cas5c family, we performed detailed biochemical, structural, and mutational studies on SmuCas5c. We have determined the crystal structure of SmuCas5c at a resolution of 1.7 Å and identified the catalytic triad and other active site residues using site-directed mutagenesis. Our work demonstrated that SmuCas5c is a metal-independent RNase that specifically recognizes the pre-crRNA stem-loop structure and cleaves it at the 3'-stem base. Compared with other Cas5c proteins (BhaCas5c), SmuCas5c exhibits significant substrate plasticity and tolerates sequence variations around the cleavage site.

## Results and discussion

### Phylogenetic analysis of the Cas5c protein family

Most of the currently available Cas5c sequences (1635 genes; IPR010155; InterPro database) are present in bacterial genomes (1614 proteins) with only 35 Cas5c genes found in archaea and environmental metagenomes. Phylogenetic analysis of the Cas5c family revealed the presence of three major monophyletic groups: I, II, and III (Fig. 2). The largest group I contains Cas5c proteins from both Gram-negative and Gram-positive bacteria including the biochemically characterized DvuCas5c and XorCas5c proteins from Gram-negative bacteria (Fig. 2). The Cas5c groups II and III are phylogenetically close to each other and contain sequences mostly from Gram-positive bacteria (Firmicutes and Actinobacteria), including BhaCas5c (group II), SpyCas5c (group III), and SmuCas5c (group III) (Fig. 2). Based on sequence alignment, the group I proteins include the Cas5c sequences with an insertion (23 amino acids) in the RRM domain (DvuCas5c and XorCas5c), whereas the members of groups II and III have a C-terminal extension (26–29 amino acids) (BhaCas5c, SmuCas5c, and SpyCas5c) (Fig. S1). A recent study on the cryo-EM structure of the *D. vulgaris* cascade revealed that the RRM domain insertion of DvuCas5c clamps around the Cas7.7 protein providing additional interactions within the cascade complex (42). The phylogenetic Cas5c group I corresponds to

## Structure and activity of the Cas5c ribonuclease SMU1763



**Figure 1. CRISPR–Cas systems of *Streptococcus mutans* UA159.** A, organization of type II-A (CRISPR1) and type I-C (CRISPR2) loci along with their associated *Cas* genes. The *SmuCas5c* gene is highlighted in orange. B, secondary structure of CRISPR1 repeat (R1), CRISPR2 repeat (R2), and their complementary strand RNAs (R1c and R2c, respectively). C, sequences of the crRNA repeats from *S. mutans*, *Bacillus halodurans*, *Streptococcus pyogenes*, and *Xanthomonas oryzae*. Sequence variations between *S. mutans* and *B. halodurans* repeats are shown in red and blue, respectively. D, sequence and secondary structure of the CRISPR2 locus as seen in *S. mutans* GS-5. crRNA, CRISPR-RNA.

the previously proposed subgroup B (originally, Cas5d-B), whereas the groups II and III (*SmuCas5c*) are related to subgroup A (originally, Cas5d-A) (34).

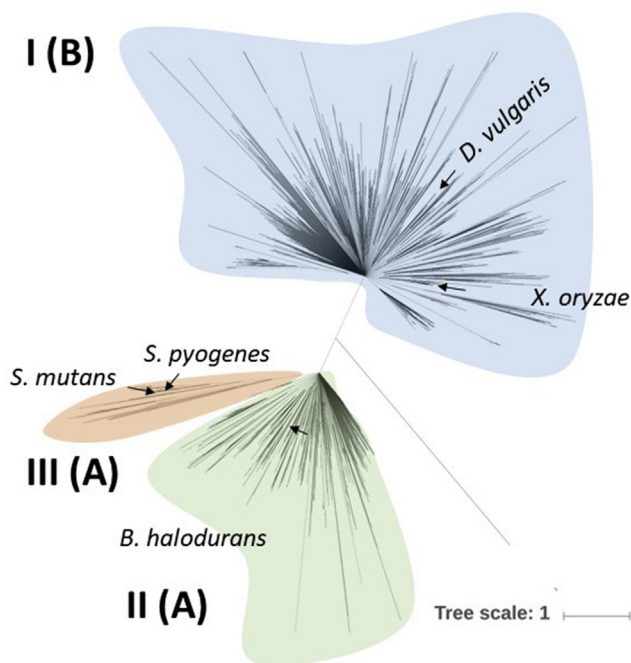
### Crystal structure of *SmuCas5c* reveals a modified RRM fold

Purified *SmuCas5c* was crystallized using the sitting-drop vapor diffusion method, and its structure was determined at a resolution of 1.72 Å using the selenomethionine (SeMet)-substituted protein and single-wavelength anomalous dispersion method (Table S1). The protomer structure revealed the presence of two domains, an N-terminal modified RRM domain (also known as the ferredoxin fold, Met1–Pro160) and a C-terminal twisted β-sheet domain (Tyr162–Thr214) (Figs. 3A and S2). RRM domains contain approximately 90 residues folded into an αβ sandwich structure with a

$\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$  topology (43). The canonical RRM fold is composed of one four-stranded antiparallel β-sheet flanked on one side by two α helices (α1 and α2). Cas6 proteins typically have two sequential RRM domains, whereas Cas5c contains an N-terminal RRM domain and a C-terminal domain with a three-stranded antiparallel β-sheet (23). Similar to crystal structures of BhaCas5c, SpyCas5c, and XorCas5c (32, 34), the topology of the *SmuCas5c* RRM domain deviates from the canonical βαβαβ fold and contains additional β-strands (within the twisted antiparallel β-sheet) (Figs. 3 and S2). The *SmuCas5c* RRM domain has the  $\beta_1\alpha_1\beta_2\beta_3\beta_4\alpha_2\alpha_3\beta_5\beta_6$  fold with the unresolved residues Leu78–Ser87 (Figs. 3 and S2). The α1, α2, and α3 helices of *SmuCas5c* are located on the concave side of the central β-sheet, whereas α4 is positioned on the other (convex) side (Fig. 3). A similar helix (labeled as α3) was also observed in the crystal structure of SpyCas5c (Protein



## Structure and activity of the Cas5c ribonuclease SMU1763



**Figure 2. Phylogenetic analysis of Cas5c proteins.** Unrooted phylogenetic tree of Cas5c proteins (IPR010155) showing the presence of three major monophyletic groups: I (previous subgroup B), II (previous subgroup A), and III (previous subgroup A). About 1548 Cas5c sequences were aligned using MAFFT online, version 7, and the tree was built using Geneious, version 8.1.9, with 60% threshold. The red circles indicate the structurally characterized Cas5c proteins: SmuCas5c (subgroup A) from *Streptococcus mutans*, XorCas5c (subgroup B) from *Xanthomonas oryzae*, BhaCas5c (subgroup A) from *Bacillus halodurans*, and SpyCas5c (subgroup A) from *Streptococcus pyogenes*.

Data Bank [PDB] code: 3VZH) but was missing in other Cas5c structures (BhaCas5c, DvuCas5c, and XorCas5c). The C-terminal domain of SmuCas5c (Asn164–Gly249) includes the twisted antiparallel  $\beta$ -sheet ( $\beta$ 7,  $\beta$ 8,  $\beta$ 9, and  $\beta$ 10),  $\alpha$ 5 helix (Thr167–Lys202), and C-terminal tail (Leu215–Gly249), which was disordered in the SmuCas5c crystal structure (Figs. 3 and S1).

Size-exclusion chromatography of purified SmuCas5c suggested that this protein exists as a dimer in solution (observed mass of 61.2 kDa and predicted subunit mass of 28.7 kDa) (Fig. S3). This is consistent with the prediction of a dimeric biological unit using the PDBePISA tool (44) formed mainly by interactions between the N-terminal strands (Asn2–Arg5) and  $\beta$ 4– $\alpha$ 2 loops of two protomers with buried surface area of 950  $\text{\AA}^2$  (compared with 2300  $\text{\AA}^2$  for the DvuCas5c–Cas7 interface area in the *D. vulgaris* cascade complex) (Fig. 3B) (42). The PDBePISA analysis also predicted a dimeric state for BhaCas5c, which was verified using size-exclusion chromatography using the purified BhaCas5c. As shown in Fig. S3, the major peak appears to represent a dimer (observed mass 50.5 kDa and predicted subunit mass of 26.9 kDa). The previous structural study of BhaCas5c also revealed the presence of two protein molecules in the asymmetric unit, but the protein behaved as a monomer based on size-exclusion chromatography and dynamic light scattering (32). Several

Cas6 proteins from the types I-A and I-B CRISPR systems have also been reported to form dimers (22).

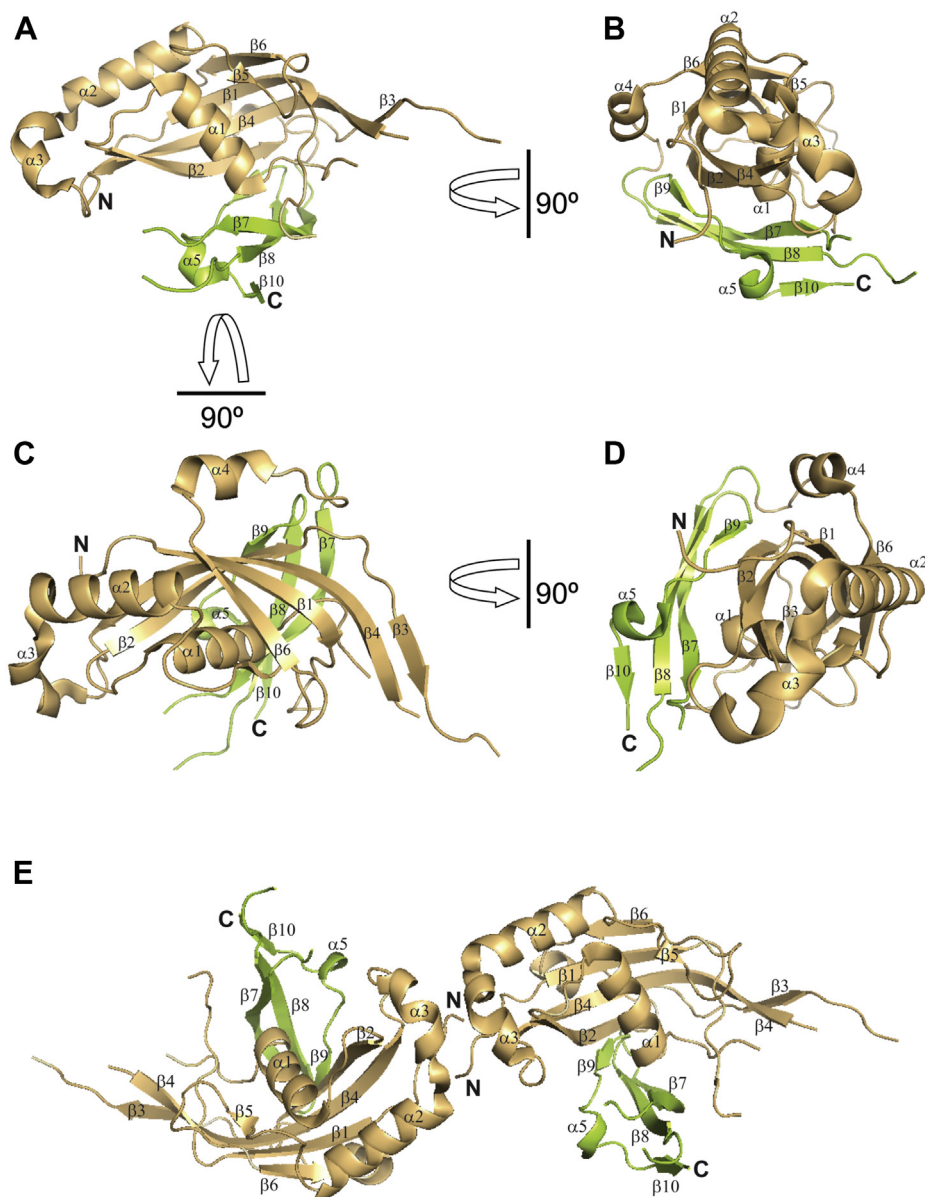
A Dali search (45) for structurally homologous proteins in the PDB database revealed several similar structures of characterized and uncharacterized Cas5c proteins from different bacteria with SpyCas5c from *S. pyogenes* as the top structural homologue (PDB code: 3VZH; Z-score: 31.5; rmsd: 1.4  $\text{\AA}$ ; and 80% sequence identity). The other three homologous Cas5c structures include BhaCas5c from *B. halodurans* (PDB code: 4F3M; Z-score: 26.0; rmsd: 1.9  $\text{\AA}$ ; and 46% sequence identity), XorCas5c from *X. oryzae* (PDB code: 3VZI; Z-score: 22.9; rmsd: 2.5  $\text{\AA}$ ; and 26% sequence identity), and MsuCas5c (MS0988) from *Mannheimia succiniciproducens* (PDB code: 3KG4; Z-score: 21.6; rmsd: 2.1  $\text{\AA}$ ; and 27% sequence identity).

Surface charge analysis of the SmuCas5c structure revealed that positively charged residues dominate in a shallow groove formed by the RRM domain, which is likely to represent the RNA-binding site (Fig. 4). The positively charged area extends into the dimer interface, which represents a twofold symmetry structure with the two positively charged grooves exposed on one dimer side (Fig. 4G). Surface sequence conservation analysis of the SmuCas5c dimer using the ConSurf web server (46) demonstrated strong conservation of positively charged residues located in both shallow grooves (e.g., Lys52, Arg134, Lys120, Arg142) suggesting functional importance of these residues for RNA binding (Fig. 4H).

### Sequence analysis of *S. mutans* crRNA repeats

As shown in Figure 1A, the *S. mutans* CRISPR1 cluster (type II-A) contains six spacers (30 bp each) separated and flanked by seven 36 bp repeats. The CRISPR2 locus (type I-C) has just one spacer (34 bp) matching (100% identity) a sequence in the *S. mutans* LJ23 genome and flanked by two repeats (32 bp), which show no sequence similarity to the CRISPR1 repeats. Analysis of potential secondary structures of the *S. mutans* CRISPR repeats (both coding and complementary strands) using the Mfold server (47) suggested the presence of a hairpin (stem–loop) structure with a long 3'-tail in the CRISPR2 repeat (type I-C) and a noncanonical long (13 nt) stem and 10 nt loop structure with neither 5'-tail nor 3'-tail in the CRISPR1 repeats (Fig. 1B). This analysis predicted the formation of hairpins with different stem lengths (13 and 8 nt) in both CRISPR1 crRNAs, whereas both CRISPR2 repeats were predicted to form similar stem–loop structures with a 7-nt stem and 5-nt loop flanked by the 2-nt and 11-nt overhangs (Fig. 1B).

Interestingly, the overall structure and lengths of the stem (7-nt) and overhangs (2-nt and 11-nt) of the predicted crRNA repeat structures are identical for the crRNA repeats from *S. pyogenes*, *B. halodurans*, *X. oryzae*, and *S. mutans* (R2 from CRISPR2) (Fig. 1B). The four crRNA repeats (*B. halodurans*, *X. oryzae*, *S. pyogenes*, and *S. mutans*) also share high sequence similarity with just 2 to 4 different nucleotides (88–94% identity) and identical 5' (2 nt) and almost identical 3' (11 nt) overhangs (Fig. 1C). In contrast, the corresponding Cas5c



**Figure 3. Crystal structure of SmuCas5c.** A–D, overall fold of the SmuCas5c protomer: four views related by a 90° rotation. The protein RRM domain is colored *light orange*, whereas the C-terminal twisted  $\beta$ -sheet domain is represented in a *lime green* color. E, SmuCas5c dimer (shown in the orientation related to those in Fig. 4, G and H). The protein subunits are shown as *ribbon* diagrams with secondary structure elements labeled. RRM, RNA recognition motif.

proteins have low overall sequence similarity (26–80% sequence identity) (Fig. S1).

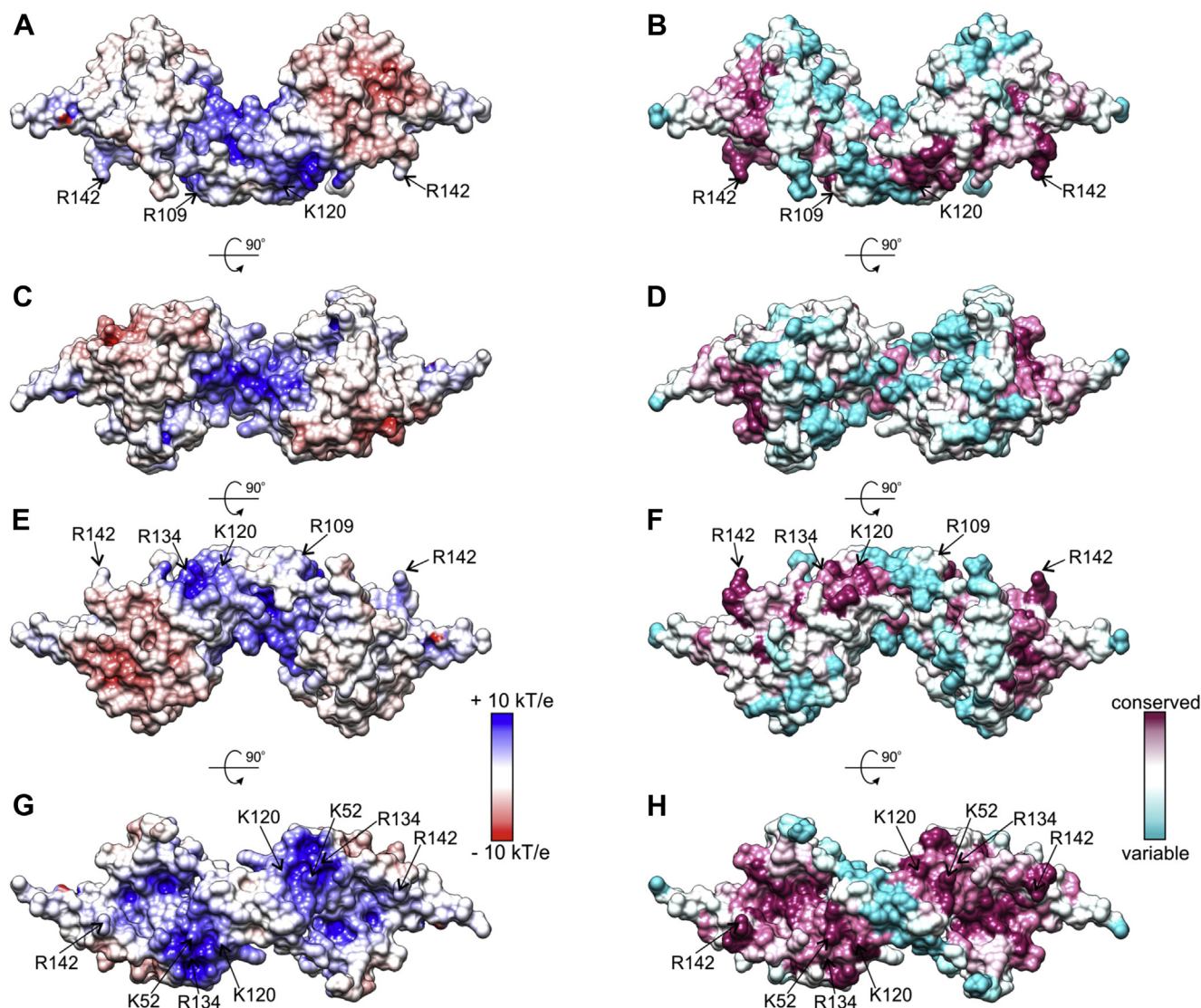
#### Nuclease activity of purified SmuCas5c

The hallmark of all Cas5c proteins biochemically characterized to date is the specific cleavage of palindromic repeat crRNA at the 3'-base of the stem-loop region (22, 32, 34). In addition, the repeat sequence must belong to the CRISPR locus associated with the Cas5c protein, making the ribonuclease activity both sequence and structure dependent (32, 34). To characterize specific RNase activity of SmuCas5c against CRISPR-related RNA substrates, we designed four synthetic RNA substrates (R1, R1c, R2, and R2c) using the repeat

sequences of the two *S. mutans* CRISPR loci, both coding (R1 and R2) and complementary strands (R1c and R2c) (Fig. 1B). Nuclease assays with purified SmuCas5c using 5'-[<sup>32</sup>P]-labeled repeat crRNAs, as well as a scrambled RNA substrate (RC), revealed the presence of endoribonuclease activity against the CRISPR2 repeat RNA (R2), whereas the other four substrates were not cleaved (Fig. 5A). Cleavage of the R2 substrate by SmuCas5c produced a single [<sup>32</sup>P]-labeled product with the length approximately 20 nt indicating that the cleavage site is located close to the 3'-side of the R2 stem-loop base. Endoribonuclease activity of SmuCas5c was found to be metal independent with no inhibition by divalent metal ions or EDTA except for Cu<sup>2+</sup> (Fig. 5C). This is in line with a metal-independent crRNA cleavage reported for the *B. halodurans*



## Structure and activity of the Cas5c ribonuclease SMU1763



**Figure 4. Crystal structure of SmuCas5c: surface analysis of the dimer.** A, C, E, and G, distribution of surface potential shown in four orientations with a blue (positively charged) to red (negatively charged) gradient. B, D, F, and H, conservation of surface residues shown in four orientations with a magenta (conserved) to cyan (variable) gradient. The scale bars show coloring schemes of the surface-exposed residues. The positions of catalytic residues are indicated by arrows. Dimer orientations shown in panels G and H are related to that shown in Figure 3E.

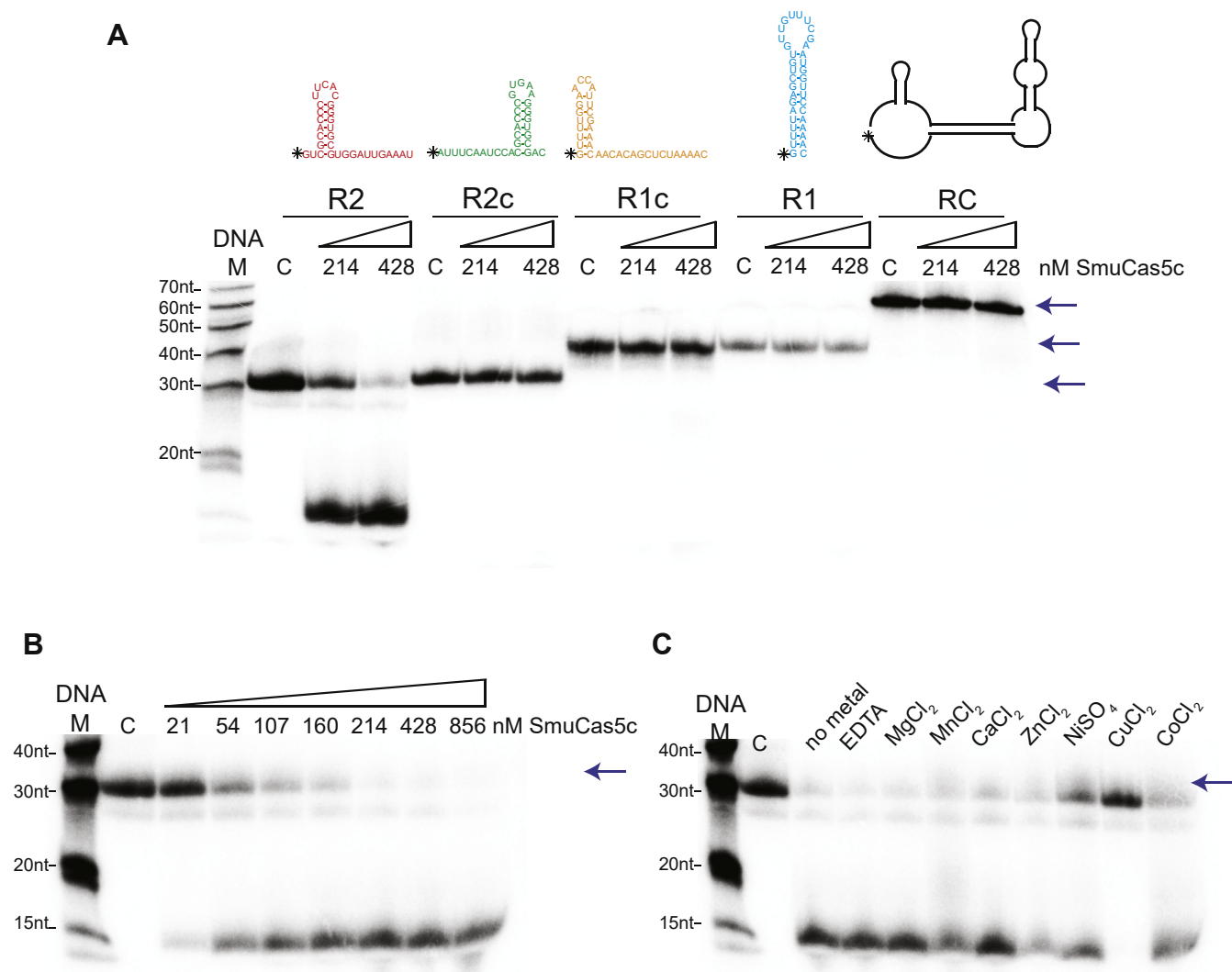
Cas5c (32). This activity was also maximal at pH 7 to 10 and was insensitive to KCl concentrations up to 200 mM (Fig. S3, G and H). Thus, our results with SmuCas5c confirmed a strong preference of Cas5c proteins for the CRISPR repeat RNA as well as the metal-independent cleavage of RNA at neutral-alkaline pH and the absence of inhibition by KCl. Likewise, while it is possible for Cas proteins to be shared between CRISPR loci within an organism, it is uncommon for the RNA-processing enzymes to mature different kinds of pre-crRNA (19).

Although Cas5c proteins have been shown to function in crRNA processing as structure-specific and sequence-specific RNases, the three biochemically characterized Cas5c enzymes also exhibited DNA binding (SpyCas5c and XorCas5c) or DNA cleavage (BhaCas5c) activities (34, 36). Therefore, the purified SmuCas5c (Fig. S3A) was tested for the presence of DNA binding and cleavage activities

toward ssDNA and dsDNA. In accordance with our previous study (41), no nuclease or DNA-binding activities were found using ssDNA (circular M13mp18 phage DNA; DNA1) or dsDNA (circular pUC19 plasmid; DNA1/DNA2) as substrates in the absence/presence of metal ions (Fig. S3, B–E).

### SmuCas5c activity: Analysis of crRNA cleavage site and products

Metal-independent RNases usually cleave the phosphodiester bond of their RNA substrates with the formation of 2',3'-cyclic phosphate and 5'-OH groups (48). This was also true for the biochemically characterized BhaCas5c from *B. halodurans* (32). The crRNA cleavage products of SmuCas5c were analyzed using T4 polynucleotide kinase (PNK) in the reactions of direct phosphorylation of cleavage products at



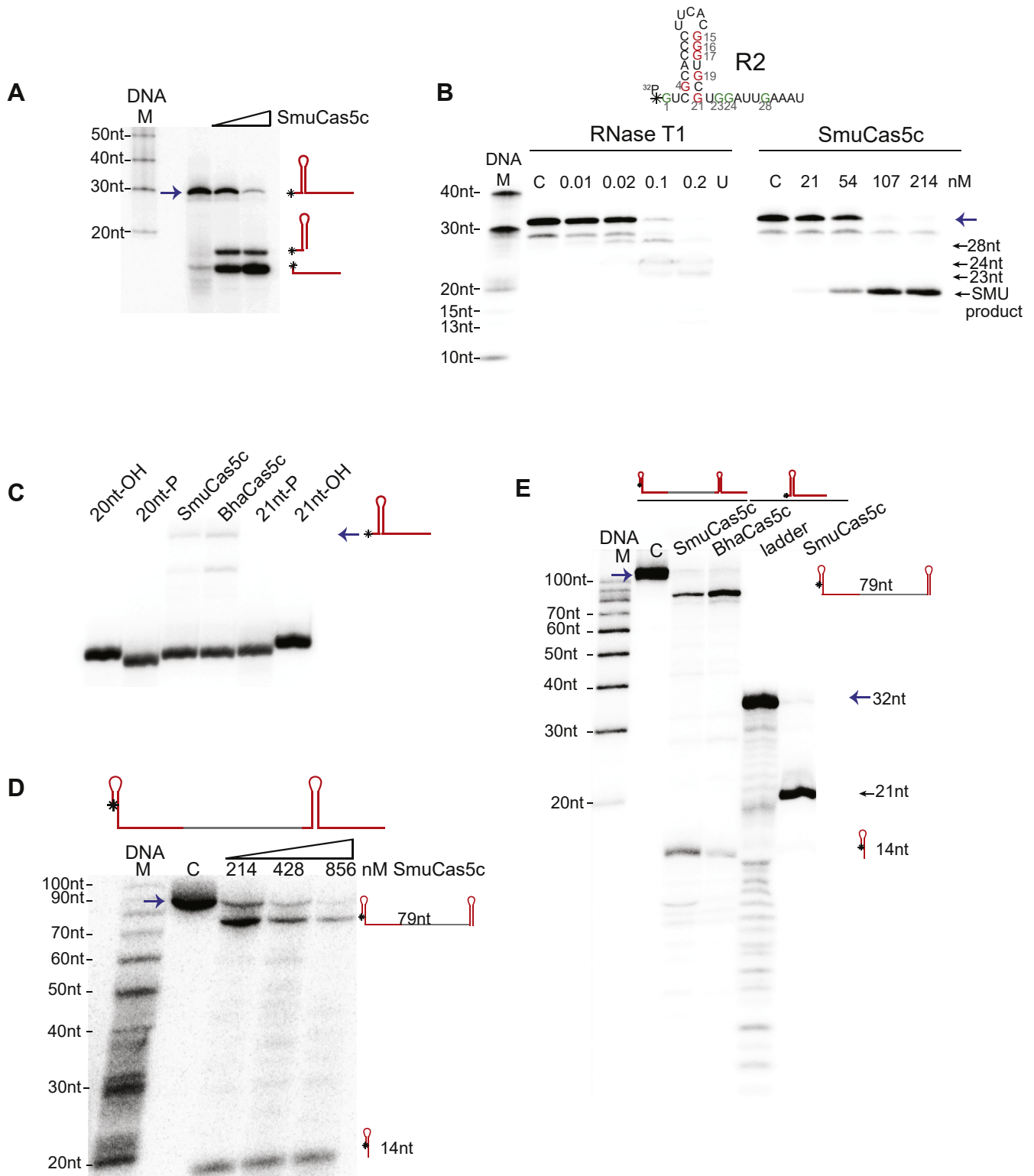
**Figure 5. RNase activity of SmuCas5c against different RNA substrates.** A, screening of purified SmuCas5c against five different RNA substrates. 5'-[<sup>32</sup>P]-labeled repeat and complementary repeat RNAs from CRISPR1 and CRISPR2 or scrambled RNA were incubated with 214 nM or 428 nM SmuCas5c at 37 °C for 15 min. B, cleavage of R2 RNA in the presence of increasing amounts of SmuCas5c (21–856 nM; 15 min incubation at 37 °C). C, effect of divalent metal ions on R2 cleavage by 214 nM SmuCas5c (15 min incubation at 37 °C). All gels were analyzed using denaturing PAGE alongside a DNA marker, and blue arrows indicate the positions of uncleaved substrate bands.

the 5'-end (requires a 5'-OH end) and phosphate exchange at the 5'-end (requires a 5'-phosphate end). These experiments revealed high product labeling by PNK in the forward phosphorylation reaction (Fig. 6A) and no product labeling in the phosphate exchange reaction (data not shown) suggesting the formation of a 5'-OH as one of the SmuCas5c products (and supposedly 2',3'-cyclic phosphates as the second product). The formation of these product termini is consistent with a cleavage mechanism based on the attack of the G21 2'-hydroxyl group on the scissile phosphodiester. Similar RNA cleavage product ends (5'-OH and 2',3'-cyclic phosphate) have been reported for Cas5c proteins from *Thermus thermophilus* and *B. halodurans* as well as for several Cas6 proteins (19, 22, 32, 33, 49). To confirm the integrity of the dsRNA region of the R2 stem structure, we used the monospecific RNase T1, which is known to cleave after guanine in ssRNA areas (50). The formation of reaction products with the lengths

23, 24, and 28 nt and the absence of shorter reaction products (4–21 nt) after incubation with RNase T1 suggested that the stem structure of the R2 RNA remained intact during the reaction (Fig. 6B).

The exact size of the 5'-[<sup>32</sup>P]-labeled cleavage product of the R2 substrate was determined using truncated versions of the R2 substrate (20 and 21 nt long) containing either a 3'-OH or 3'-phosphate as markers. As shown in Figure 6C, these assays revealed that the length of the SmuCas5c product of R2 cleavage is 21 nt and confirmed the formation of a phosphorylated 3'-end of the product. This suggests that processing of pre-crRNA by SmuCas5c produces mature crRNAs (65 nt) containing a single spacer (33 nt) with the 5'-handle (11 nt) and 3' stem-loop part (21 nt). Purified SmuCas5c and BhaCas5c also produced identical products (21 nt) during cleavage of the *B. halodurans* repeat as substrate (Fig. S4). Thus, Cas5c proteins from different

## Structure and activity of the Cas5c ribonuclease SMU1763



**Figure 6. Product analysis of RNA cleavage by SmuCas5c.** A, PNK catalyzed forward phosphorylation labeling of reaction products of the R2 RNA cleavage by purified SmuCas5c. A schematic of major products is indicated on the side. B, cleavage of the R2 substrate by RNase T1 (Ambion) or SmuCas5c. The sequence and predicted structure of R2 are shown with the  $^{32}\text{P}$ -label indicated with an asterisk. Green G letters in the substrate indicate potential cleavage sites for RNase T1, whereas red G letters indicate nucleotides located within the stem structure (which are not supposed to be cleaved). Major products formed by both enzymes are indicated with arrows. C, cleavage products of 5'- $^{32}\text{P}$ -labeled R2 substrate by SmuCas5c or BhaCas5c (640 nM each) alongside 20 and 21 nt R2 RNA markers containing either 3'-OH or 3'-phosphate ends (20 min at 37 °C). D, cleavage of 5'- $^{32}\text{P}$ -labeled repeat-spacer-repeat RNA by SmuCas5c (214–856 nM) after 45 min at 37 °C analyzed by denaturing gel electrophoresis alongside a 20 to 100 nt DNA marker (IDT). E, cleavage of 5'- $^{32}\text{P}$ -labeled repeat-spacer-repeat RNA by SmuCas5c or BhaCas5c (856 nM each) analyzed by denaturing gel electrophoresis alongside a 20 to 100 nt DNA marker, alkaline hydrolysis ladder of the R2 substrate, and a 21 nt R2 marker (30 min at 37 °C). R2\_SMU is the product of 640 nM SmuCas5c incubated with 5'- $^{32}\text{P}$ -labeled R2 substrate (20 min at 37 °C). A schematic of full-size substrate and major products of estimated size is indicated. All gels were analyzed using denaturing PAGE, and blue arrows indicate the positions of uncleaved substrate bands. PNK, polynucleotide kinase.



phylogenetic groups exhibit high conservation of the mechanisms of pre-crRNA cleavage with the cleavage site located after the nucleotide-21 at the hairpin base.

Substrate specificity of SmuCas5c was further characterized using a synthetic 5'-[<sup>32</sup>P]-labeled RNA with the repeat-spacer-repeat (R-S-R) sequence from the *S. mutans* GS-5 (type I-C) locus with the R2 repeat (Fig. 1D). Given synthesis restrictions and the fact that the 5'-end of the repeat sequence was previously shown to be not critical for crRNA processing (32), the first seven nucleotides of the repeat were omitted from the 5'-end of the substrate (Table S2). This RNA substrate contains two potential cleavage sites for SmuCas5c, which is expected to generate two <sup>32</sup>P-labeled products (79 and 14 nt) given the cleavage site location at the bottom of the stem loop. As expected, incubation of the R-S-R RNA substrate with SmuCas5c produced two 5'-labeled products with the long product around 79 nt and the short product less than 21 nt in length (Fig. 6, D and E). The cleavage of the R-S-R construct by SmuCas5c confirmed the site-specific mechanism of crRNA cleavage by this enzyme. Although the first seven nucleotides of the 5'-terminal repeat were missing, the major cleavage product was still around 79 nt in length, corresponding to a cleavage site of 1 bp into the second repeat sequence. A smaller product that is less than 20 nt in length (expected size of 14 nt) indicates that the cleavage of the first repeat is not based on a ruler mechanism but rather on crRNA sequence and structure. Purified BhaCas5c was found to generate the same products (79 and 14 nt) after cleavage of the *S. mutans* R-S-R crRNA substrate (Fig. 6E). Thus, the cleavage of the R-S-R construct further confirmed the structure-specific and site-specific mechanism of pre-crRNA cleavage by Cas5c proteins.

Previous studies with BhaCas5c revealed that the crRNA recognition was focused mostly on the trinucleotide sequence (U22-G23-G24) immediately downstream of the cleavage site with crRNA processing inhibited by mutations near the cleavage site (at the stem-loop base) and in the 3'-overhang but not in the pentaloop (UUCAU) or 5'-overhang (32). This is in contrast to Cas6 proteins recognizing the crRNA stem-loop part and upstream nucleotides (22). Given high similarity of the *S. mutans* and *B. halodurans* repeat sequences and low similarity between SmuCas5c and BhaCas5c (46% sequence identity), we thought to compare the tolerance of these enzymes to sequence variations around the crRNA R2 cleavage site. As shown in Figure 7, SmuCas5c and BhaCas5c showed comparable cleavage activities with the wildtype *S. mutans* crRNA substrate (R2). With mutated R2 substrates, SmuCas5c showed slightly reduced cleavage of the U22G variant but exhibited close to wildtype activities against the C20G, G21U, and G23U substrates. In contrast, the activity of BhaCas5c toward U22G was almost abolished (Fig. 7). These results indicate that SmuCas5c exhibits a greater tolerance to substrate sequence variations compared with BhaCas5c including mutations located downstream of the cleavage site.

Crystal structures of Cas5c proteins revealed that the active site of SmuCas5c includes His51, which is replaced by Trp47 in BhaCas5c, Trp51 in DvuCas5c, or Tyr46 in DvuCas5c (Figs. 8A, S1, and S4). Most bacterial and archaeal Cas5c

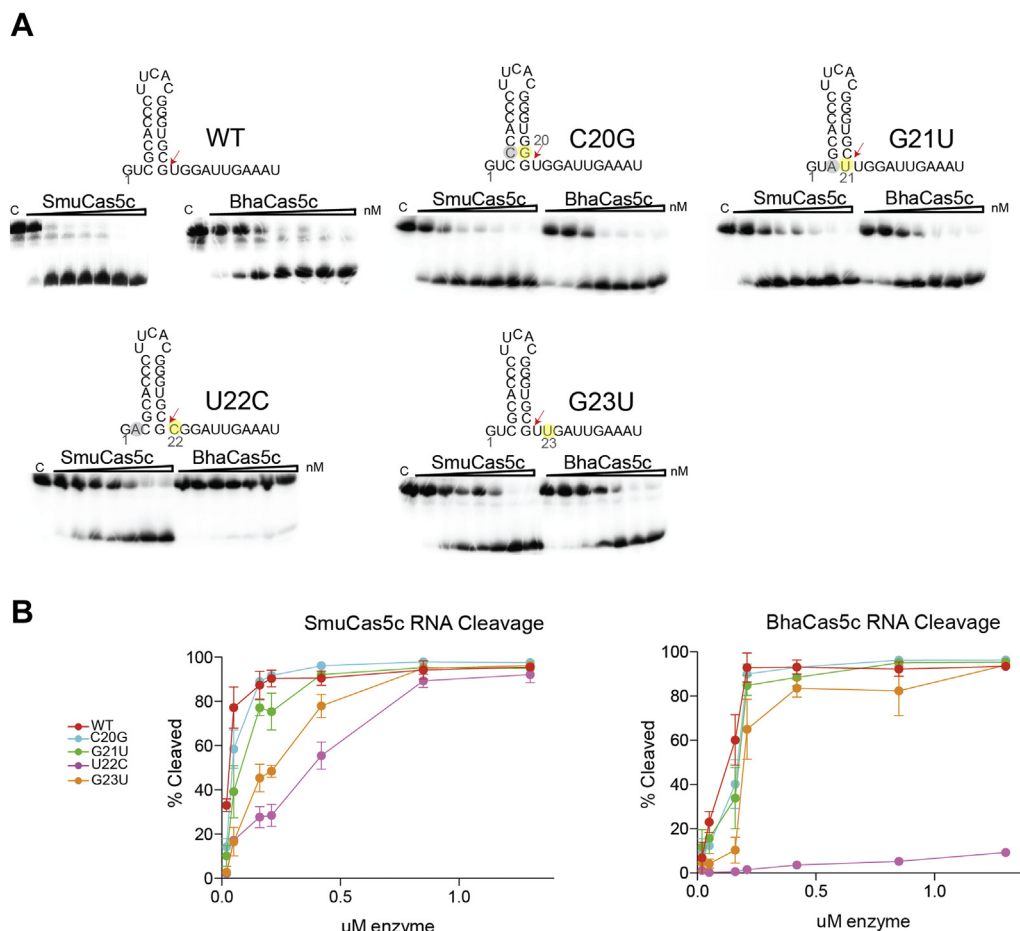
sequences show the presence of Trp, Tyr, or Phe at this position, whereas His is present mainly in various *Streptococcus* strains (*Streptococcus anginosus*, *Streptococcus constellatus*, *Streptococcus equi*, *Streptococcus gordonii*, etc). As revealed by the cryo-EM structure of the *D. vulgaris* cascade-crRNA complex, the DvuCas5c Trp51 is involved in crRNA binding with the side chain positioned between the first two bases (U1 and G2) of the crRNA 5'-handle (42). In the SmuCas5c active site, the smaller side chain of His51 can also interact with crRNA bases, but it has different steric properties compared with Trp regarding substrate coordination (Fig. S4D). Therefore, His51 is one of the residues potentially contributing to a higher substrate plasticity of SmuCas5c compared with BhaCas5c. Other potential SmuCas5c residues include Arg39, Gln43, Asp47, Glu122, and Lys131 (Figs. 8A and S1).

### Mutational studies of the SmuCas5c active site

The crystal structure of SmuCas5c revealed the presence of three sulfate molecules (presumably from the crystallization mother liquor) bound in the positively charged groove, which likely mimics phosphate groups of RNA (Fig. 8A). The S1 sulfate is bound near the side chains of Arg127 (2.7 Å), Lys131 (4.3 Å), and Arg134 (3.2 Å), whereas S2 is coordinated by Arg135 (2.7–3.1 Å), and S3 is close to Arg39 (2.6–2.8 Å) and Tyr179 (2.6 Å) (Fig. 8A). The fourth sulfate molecule (S4) is bound nonspecifically outside the predicted RNA-binding groove close to Tyr163 (3.9 Å). The bound sulfate S1 is located in proximity to the side chains of conserved SmuCas5c residues Tyr50 (6.8 Å), Lys120, and His121, which are closely clustered in the space (2.8–4.3 Å) (Fig. 8A).

Cas5c and Cas6 enzymes have been proposed to use a general acid-base mechanism based on a catalytic triad of His, Tyr, and Lys, which may function as a general base, a general acid, and in intermediate stabilization, respectively (22, 23, 32, 49). Previous studies on DvuCas5c suggested that the catalytic triad of this enzyme includes Tyr46, Lys116, and His117, which are located close to the 3'-end of crRNA in the *D. vulgaris* cascade-crRNA complex (32, 42). Metal ion-independent RNase activity and RNA cleavage products of SmuCas5c also suggest that this enzyme cleaves crRNA substrates using a general acid-base mechanism similar to that of RNase A and RNA-splicing endonucleases (51, 52). To gain further insight into the SmuCas5c mechanism, Tyr50, Lys120, and His121 were mutated to Ala, and RNase activity of purified mutant proteins was analyzed using the R2 RNA substrate. As shown in Figure 8B (and Fig. S5), alanine replacement mutagenesis produced catalytically inactive proteins suggesting that the catalytic triad of SmuCas5c includes Tyr50 (located at the end of  $\alpha$ 1), Lys120, and His121 (both are located at the beginning of  $\alpha$ 3). The catalytic triad Lys120 is the only triad residue that is invariant in Cas5c proteins, whereas the Tyr50 and His121 can be replaced by other amino acids (His, Phe, Leu, and Tyr) (23, 33, 34). This is in line with the absence of homologous residues in crystal structures of noncatalytic Cas5 proteins (e.g., *Escherichia coli* Cas5; PDB: 4QYZ). According to the proposed mechanism for SmuCas5c, His121 (the general

## Structure and activity of the Cas5c ribonuclease SMU1763



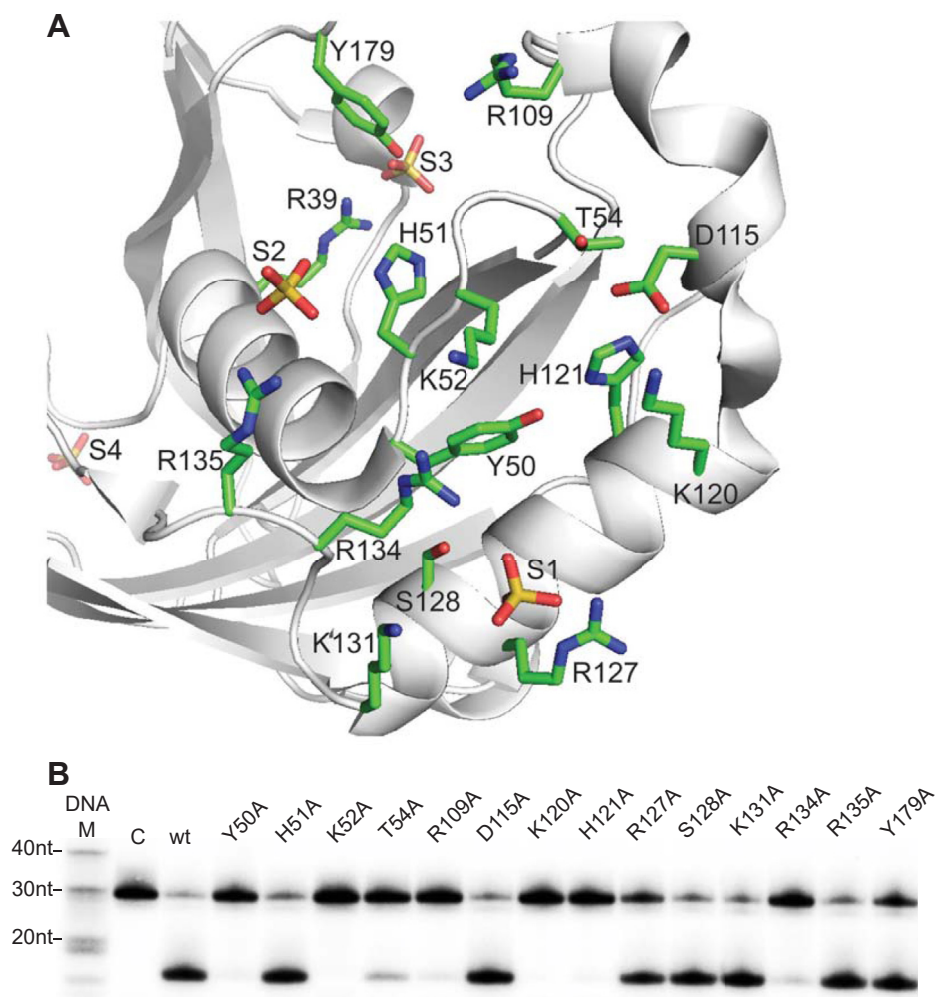
**Figure 7. Substrate specificity of SmuCas5c (SMU1763) and BhaCas5c (BH0337).** A, substrate cleavage of 5'-[<sup>32</sup>P]-labeled wildtype or modified R2 RNAs under stringent conditions by 20 nM to 1.3 μM SmuCas5c or BhaCas5c enzymes at 37 °C for 30 min. The cleavage site is indicated using a red arrow, with cut-site proximal nucleotide modifications highlighted in yellow, and noncleavage site modifications made to maintain structure stability highlighted in gray. The repeat sequences of *Bacillus halodurans* and *Streptococcus mutans* are shown in Figure 1C. B, quantification of RNA cleavage gels showing protein amount versus percentage of substrate cleaved for each of the wildtype or modified R2 RNA substrates. The analysis is shown as a mean ± SD of the gels from panel A and Figure 4C.

base) is expected to deprotonate the nucleophilic 2'-oxygen of crRNA ribose, whereas Tyr50 (the general acid) apparently protonates the leaving 5'-hydroxyl oxygen, and Lys120 stabilizes the developing negative charge of the transition state (15, 24, 32, 53).

Based on the surface sequence conservation and sequence alignment (Figs. 4 and S1), the single RRM domain of SmuCas5c accommodates a large number of conserved charged and polar residues, which are presumably involved in crRNA binding. This is consistent with a recent study on the cryo-EM structure of the *D. vulgaris* type I-C cascade showing that homologous residues of DvuCas5c interact with the crRNA 5'-handle (nucleotides U1–G12) (42). Accordingly, R2 cleavage assays with SmuCas5c mutant proteins revealed greatly reduced activity in K52A, T54A, R109A, R127A, and R134A suggesting that these residues are involved in binding of the crRNA 5'-handle (Fig. 8B). Furthermore, the bound sulfate S3 likely indicates a potential binding site for the crRNA stem-loop part, which is in line with reduced R2 cleavage activity of Y179A (Fig. 8B). The *E. coli* cascade structure (30) revealed a similar (hooked) conformation of the crRNA 5'-handle bound to Cas5e indicating that interactions with the crRNA

5'-handle are critical for the assembly of cascade complexes. In contrast, Cas6e recognizes the crRNA substrate mainly within the stem-loop part and remains tightly associated with the stem-loop part of the mature crRNA (30).

Our previous study reported the presence of promiscuous ssRNA cleavage activity in SmuCas5c (41). Nonspecific nuclease activity of this enzyme was further tested against the repeat RNAs of the two UA19 CRISPR loci (both sense and antisense strand sequences), an R1c RNA with two additional nucleotides at the 5'-end (R1c+), and a scrambled RNA (RC) using relaxed reaction conditions (with high protein levels and extended incubation times). As shown in Fig. S6, SmuCas5c cleaved all RNA substrates with no specific pattern with varying size of dominant products. We also tested the SmuCas5c mutant proteins for nonspecific cleavage activity against several RNA substrates (Fig. S6). Interestingly, while the active site residues Lys120 and His121 were still important for promiscuous RNA cleavage, alanine replacement of the third triad residue, Tyr50, had no negative effect on this activity (even improving the activity compared with the wildtype protein). Furthermore, mutations in other residues located in the active site (Thr54, Ser128, Lys131, Arg135, and Tyr179) produced



**Figure 8. Active site and mutational studies of SmuCas5c.** *A*, close-up view of the SmuCas5c active site (indicated by the red rectangle in panel *A*) showing the active site residues selected for mutagenesis. The protein ribbon diagram is colored gray, whereas the active site residues mutated in this study are shown as sticks with green carbons and labeled. S1, S2, S3, and S4 are sulphate molecules bound to the protein. *B*, RNase activity of purified wildtype and mutant proteins against R2 RNA. 5'-[<sup>32</sup>P]-labeled R2 RNA was incubated with SmuCas5c proteins (214 nM) for 20 min at 37 °C. These assays were performed in triplicate with a representative gel presented in this figure alongside a DNA marker.

significant reduction or abolished RNA cleavage suggesting that these residues contribute to substrate binding.

Promiscuous RNA cleavage activity has also been demonstrated in purified BhaCas5c and two Cas6 enzymes (Cas6-1 and Cas6-2a) from *Synechocystis* sp. PCC 6803 (36, 54). Since both Cas6 proteins were highly specific for cognate crRNA transcripts *in vivo*, it has been proposed that the *in vivo* cleavage of crRNAs might involve additional Cas protein(s) (54). Furthermore, our experiments with SmuCas5c suggest that promiscuous RNA cleavage by purified Cas5c and Cas6 proteins *in vitro* can be observed under relaxed reaction conditions (using high concentrations of substrate and enzymes).

#### Effect of SmuCas5c deletion on the growth and stress tolerance of *S. mutans* cells

Recently, Serbanescu *et al.* (41) demonstrated that the deletion of the *S. mutans* CRISPR2 locus (which also includes the *SmuCas5c* gene) results in a higher sensitivity to heat shock. To determine if SmuCas5c is essential for *S. mutans*

growth and stress tolerance, we constructed a SmuCas5c deletion strain and compared its growth with the wildtype (UA159) strain. Growth experiments using THYE (rich) medium revealed no apparent growth defects in the SmuCas5c deletion strain under normal growth conditions (pH 7.0 and 37 °C) as well as in the presence of different stressors, such as low pH (5.5), elevated temperature (50 °C), extracellular (0.003% H<sub>2</sub>O<sub>2</sub>) and intracellular (25 mM paraquat) oxidative stress, cell membrane stress (0.004% SDS), NaCl (0.4 M), and ethanol (2%) (Fig. S7). These results suggest that there appears to be no direct role for SmuCas5c in growth and stress tolerance in *S. mutans*, and the increased sensitivity of the *S. mutans* CRISPR2 deletion mutant is related to the loss of this CRISPR locus, rather than Cas5c.

#### Conclusions

In the CRISPR–Cas system, Cas5c proteins represent a multifunctional component, which is involved in crRNA maturation, cascade complex assembly, and CRISPR



## Structure and activity of the Cas5c ribonuclease SMU1763

interference (22, 49). A unique feature of Cas5c proteins is a combination of metal-independent endoribonuclease activity involved in crRNA maturation (as in Cas6 proteins) and ability to bind the crRNA 5'-handle and assemble with other Cas proteins forming an integral part of cascade (like noncatalytic Cas5 proteins) (32, 35, 49). Cas5c and Cas6 are structurally similar proteins with metal-independent RNase activity generating RNA products with 5'-OH and 2',3'-cyclic phosphates (23, 24, 29, 33). Furthermore, these enzymes use a similar catalytic mechanism and catalytic residues to cleave pre-crRNA substrates at the base of the stem-loop structure suggesting a similar mode of crRNA binding (22). However, interactions with the stem-loop structure and upstream nucleotides appear to be critical for Cas6 activity, whereas Cas5C proteins seem to recognize the cut site and downstream nucleotides of crRNA (5'-handle) (22, 32). Accordingly, Cas5c and Cas6 proteins were found to be located at opposite ends of elongated seahorse-shaped cascade complexes with Cas5c bound near the 5'-crRNA handle at the tail of the *D. vulgaris* cascade (like noncatalytic Cas5 proteins in type I-E systems), whereas Cas6 is positioned at the 3'-end of the crRNA at the head of the *E. coli* cascade (30, 42).

Exquisite binding specificity and affinity of Cas5c and Cas6 proteins are based on extensive contacts made by these proteins with their substrate and product crRNAs, which position the scissile phosphate in the enzyme active site for cleavage (22). The crystal structure of SmuCas5c with bound sulfate molecules implies that S1 and S2 molecules are bound at the crRNA 5'-handle binding site, whereas S3 might indicate the binding site for the crRNA stem-loop part (Fig. 8A). Site-directed mutagenesis of SmuCas5c suggested that Tyr50, Lys120, and His121 represent the catalytic triad of this enzyme, whereas crRNA binding presumably involves Lys26, Arg31, Gln43, His51, Lys52, and Arg142 (crRNA 5'-handle) and Tyr179 (crRNA 3'-stem loop) (Fig. 8A). Similar catalytic triads are also present in other Cas5c proteins (Fig. S1), but they are absent in noncatalytic Cas5 proteins. SmuCas5c preferentially recognizes and cleaves pre-crRNAs in a sequence-specific and structure-specific manner, which relies on base-specific contacts between the nucleotides around the cleavage site and substrate-binding site residues located on  $\alpha$ 1,  $\beta$ 2,  $\alpha$ 3, and connecting loops. Mutational analysis of RNA substrate nucleotides around the cleavage site demonstrated a higher tolerance of SmuCas5c (compared with BhaCas5c) to nucleotide substitutions both upstream and downstream of the cleavage site (Fig. 7). Thus, our work suggests that there is an even greater biochemical diversity within the Cas5c family than it was anticipated before.

Cas5c proteins act as a multifunctional component of small type I-C cascade complexes, which represent a promising CRISPR-Cas tool for genome manipulations (35). Type I systems are the most widespread CRISPR-Cas systems in nature (11), which enable the use of endogenous cascade-Cas3 complexes for genetic engineering *via* self-targeting (55–58). Furthermore, type I systems have been optimized for heterologous genome editing in bacteria as well as for DNA cleavage and transcriptional modulation in human cells and

plants (59–63). Repurposing the naturally streamlined and minimal type I-C CRISPR system for programmable genetic manipulations will create a powerful tool both for basic research and for a wide range of applications in genome engineering including the targeted removal of large genomic regions and genome minimization.

## Experimental procedures

### Protein expression, purification, and mutagenesis

Gene cloning and affinity protein purification as well as site-directed mutagenesis of the 6His-tagged SmuCas5c from *S. mutans* UA159 (SMU\_1763c; UniProt ID: Q8DSL7) and BhaCas5c from *B. halodurans* (UniProt ID: Q9KFY3) were performed as described previously (64). The oligomeric state of purified proteins (2.5 mg/ml, 5 mg per load) was analyzed using size-exclusion chromatography in 50 mM Hepes-potassium buffer (pH 7.5) containing 150 mM NaCl on a Superdex 200 HiLoad 16/60 GL or Superdex 200 10/300 GL column (Amersham Biosciences) as reported previously (65). Molecular weight standards used were from Bio-Rad (catalog no.: 1511901): thyroglobulin (670 kDa),  $\gamma$ -globulin (158 kDa), ovalbumin (44 kDa), myoglobin (17 kDa), and vitamin B12 (1.4 kDa). The molecular weight was calculated using the Unicorn 4.12 software (Amersham Biosciences).

### Preparation of nucleic acid substrates

The RNA oligonucleotides used in this work (Table S2) were purchased from IDT. The oligonucleotides were [ $^{32}$ P]-labeled at the 5'-end using the T4 PNK (BioLabs) and purified as described previously (66). The synthetic dsDNA substrates were prepared by annealing oligonucleotides DNA1 and DNA2 as described in Table S2.

### Nuclease and DNA-binding assays

The reaction mixture for RNase assays with Cas5c proteins contained 50 mM Hepes-potassium (pH 7.0), 50 mM KCl, 1 mM DTT, 0.1  $\mu$ M 5'-[ $^{32}$ P]-labeled RNA substrate, and indicated amounts of Cas5c proteins (higher protein amounts were used in experiments for product analysis, *e.g.*, Fig. 6). The solutions were incubated at 37 °C and analyzed by non-denaturing PAGE and autoradiography as previously described (66). Reactions with purified BhaCas5c were carried out using the same reaction mixture incubated at 20 °C. The synthetic dsDNA substrates were prepared by annealing oligonucleotides DNA1 and DNA2 by incubating at 90 °C for 4 min followed by 10 min at 37 °C in a thermocycler and a slow cooldown to room temperature on benchtop (Table S2). DNA cleavage assays contained 641 nM SmuCas5c in 50 mM Hepes-potassium (pH 7.0), 50 mM KCl, 2 mM metal ion, and 5 nM of the M13mp18 ssDNA or 20 nM pUC19 dsDNA. Reactions were incubated for 30 min at 37 °C and analyzed by agarose gel electrophoresis and SYBR Green staining as previously described (66). DNA-binding assays with 107 nM to 2  $\mu$ M SmuCas5c contained 50 mM Tris-HCl (pH 7.0), 50 mM KCl, 2 mM metal ion, and 0.1  $\mu$ M 5'-[ $^{32}$ P]-labeled substrate. Reactions were incubated for 1 h at 37 °C and analyzed by

nondenaturing PAGE and autoradiography as previously described (66). The reaction mixture for RNase assays with RNase T1 contained 1× RNA Structure Buffer (Ambion), 1 μg of yeast RNA, 0.1 μM 5'-[<sup>32</sup>P]-labeled substrate, and 0.01 to 0.2 U of RNase T1 (Ambion). The solutions were incubated at room temperature for 15 min, and reaction products were analyzed by denaturing PAGE and autoradiography. The analysis of the ssRNA product ends was carried out using T4 PNK as previously described (66).

### Crystallization and structure determination of SmuCas5c

SmuCas5c was crystallized at room temperature using the hanging drop vapor diffusion protocol by mixing 1 μl of the SeMet-substituted protein (10 mg/ml in 10 mM Hepes-potassium, pH 7.5, and 500 mM NaCl) with 1 μl of the crystallization solution containing 0.1 M Tris-HCl (pH 8.5), 12% glycerol, 1.5 M ammonium sulphate, 0.02 mg/ml of subtilisin (temperature of 16 °C). The crystals were stabilized with the crystallization buffer containing 3 M NaCl followed by cryoprotection in Paratone-N prior to flash freezing in liquid nitrogen. A set of single-wavelength diffraction data was collected near the selenium absorption peak (12.66 keV) at 100 K from one SeMet-labeled SmuCas5c crystal at the 19-ID beamline of the Structural Biology Center at the Advanced Photon Source at Argonne National Lab using the program SBCcollect (67).

The data were integrated, scaled, and merged with the HKL3000 program suite (68). The heavy atom Se sites were located using the program SHELXD (69), and they were used for phasing with the program MLPHARE (70). After density modification (70), a partial model was built in three cycles of Arp/Warpmodel building (71). All the aforementioned programs are integrated within the HKL3000 program suite (68). The structural model of SmuCas5c was completed after several iterative cycles of model building using the program COOT (72) and restrained refinement using the program Refmac (70). Final refinements including translation/libration/screw refinement were carried out with the program Phenix.refine (73). Structural validation was performed using the program Molprobability (74). Sulphate groups were identified based on the match of electron densities, anionic groups, and potential interactions between anions and positively charged side chains. Data collection and structure refinement statistics for the SmuCas5c structure are summarized in Table S1. The atomic coordinates and structure factors of SmuCas5c have been deposited in the PDB under the accession code 4R0J.

### Sequence and structural analyses

Phylogenetic analysis of the Cas5c family was performed using 1548 Cas5c sequences (InterPro database; IPR010155) aligned using the MAFFT-online server (version 7) (<https://mafft.cbrc.jp/alignment/server/>). The phylogenetic tree of Cas5c proteins was built the FastTree 2.1.5 algorithm built in Geneious 8.1.9 with 60% threshold (<https://www.geneious.com/download/previous-versions/>). Analysis of potential secondary structures of the *S. mutans* crRNAs was performed

using the Mfold web server (<http://www.unafold.org>) (47). Surface electrostatic potential of SmuCas5c was calculated using the PDB2PQR and APBS servers (<https://server.poissonboltzmann.org/>) (75, 76), whereas surface conservation was determined based on the alignment of 1548 Cas5c sequences using ConSurf (<https://consurf.tau.ac.il>) (46).

### S. mutans growth and stress resistance assays

The preparation of the *S. mutans* Cas5c deletion strain ( $\Delta$ Cas5c) was described previously (Table S3) (41). Growth analysis of *S. mutans* wildtype and mutant cells was performed using a microplate reader (Bioscreen C LabSystems) equipped with BioLink software (LabSystems) programmed to monitor an absorbance at 600 nm at 37 °C every 20 min for 24 h, with moderate shaking every 10 min (41). Overnight cultures were diluted (20×) in fresh THYE (rich) medium (Todd-Hewitt broth supplemented with 0.3% yeast extract) and grown to an absorbance of approximately 0.4. About 20 μl at 600 nm of midlogarithmic phase cells for the mutant and wildtype strains were inoculated in triplicate into microtiter plate wells containing 350 μl of THYE. Wells containing uninoculated THYE were used as controls. To generate growth curve, measurements of an absorbance at 600 nm were plotted against time. The following stressors affecting growth rates were tested as described previously (77): acid (pH 5.5), ethanol (5%, v/v), NaCl (0.4 M), paraquat (25 mM), and H<sub>2</sub>O<sub>2</sub> (0.006%, v/v). For heat shock resistance assays, overnight cultures grown in THYE medium were diluted 1:20 using prewarmed THYE, and two culture aliquots were then grown until midlogarithmic phase (absorbance at 600 nm = ~0.4). One aliquot of each culture was incubated at 50 °C for 60 min, and the other aliquot was used to assay survival at time 0. Samples were gently sonicated and serially diluted in PBS, and sensitivity was quantitatively assessed by plating cells before and after heat shock incubation. Each dilution was then spotted in triplicate (20 μl each) onto THYE agar plates and incubated at 37 °C for 2 days, and colonies (colony-forming unit [CFU]) were counted. Heat shock resistance was calculated by dividing the number of CFUs obtained after incubation at 50 °C for 60 min by the number of CFUs present at time 0 and multiplying the result by 100.

### Data availability

All data presented in this article are contained within the article. The atomic coordinates and structure factors of SmuCas5c (SMU1763) have been deposited in the PDB under the accession code 4R0J.

*Supporting information*—This article contains supporting information.

*Acknowledgments*—The authors thank all members of the BioZone Centre for Applied Science and Bioengineering and Biozone MS facility for help in conducting the experiments. This work was supported by the National Sciences and Engineering Research Council Strategic Network grant IBN and National Sciences and

## Structure and activity of the Cas5c ribonuclease SMU1763

Engineering Research Council Discovery grant. The Structural Biology Center beamlines are supported by US Department of Energy, Office of Biological and Environmental Research, under contract DE-AC02-06CH11357.

**Author contributions**—S. L., M. A. S., N. B., D. G. C., and A. F. Y. conceptualization; S. L., M. A. S., A. N. K., M. R., N. B., X. X., G. B., and H. C. methodology; S. L., M. A. S., A. N. K., M. R., N. B., X. X., and G. B. formal analysis; S. L., M. A. S., A. N. K., M. R., N. B., X. X., G. B., H. C., and K. T. investigation; A. J., D. G. C., and A. S. resources; S. L. and M. A. S. data curation; S. L., M. A. S., A. N. K., M. R., G. B., K. T., D. G. C., A. S., and A. F. Y. writing – original draft; G. B., D. G. C., and A. F. Y. writing – review and editing; M. R., N. B., X. X., and K. T. visualization; A. J., D. G. C., A. S., and A. F. Y. supervision; A. J. and A. S. project administration; A. J., D. G. C., and A. F. Y. funding acquisition.

**Conflict of interest**—The authors declare that they have no conflicts of interest with the contents of this article.

**Abbreviations**—The abbreviations used are: BhaCas5c, Cas5c from *Bacillus halodurans*; CFU, colony-forming unit; crRNA, CRISPR-RNA; DvuCas5c, Cas5c from *Desulfovibrio vulgaris*; PDB, Protein Data Bank; PNK, T4 polynucleotide kinase; R–S–R, repeat–spacer–repeat; RAMP, repeat-associated mysterious protein; RRM, RNA recognition motif; SeMet, selenomethionine; SmuCas5c, Cas5c from *Streptococcus mutans*; SpyCas5c, Cas5c from *Streptococcus pyogenes*; XorCas5c, Cas5c from *Xanthomonas oryzae*.

### References

- 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/Cas systems. *Science* **339**, 819–823
- DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., and Church, G. M. (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343
- 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
- Mali, P., Aach, J., Stranges, P. B., Esvelt, K. M., Moosburner, M., Kosuri, S., Yang, L., and Church, G. M. (2013) CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838
- Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006) A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 7
- Wiedenheft, B., Sternberg, S. H., and Doudna, J. A. (2012) RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338
- Hale, C. R., Zhao, P., Olson, S., Duff, M. O., Graveley, B. R., Wells, L., Terns, R. M., and Terns, M. P. (2009) RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell* **139**, 945–956
- Garneau, J. E., Dupuis, M. E., 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
- Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuys, 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
- Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F. J., Wolf, Y. I., Yakunin, A. F., van der Oost, J., and Koonin, E. V. (2011) Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477
- Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S. J., Charpentier, E., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J., Terns, R. M., Terns, M. P., et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736
- Chylinski, K., Makarova, K. S., Charpentier, E., and Koonin, E. V. (2014) Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105
- Makarova, K. S., Wolf, Y. I., Iranzo, J., Shmakov, S. A., Alkhnbashi, O. S., Brouns, S. J., Charpentier, E., Cheng, D., Haft, D. H., Horvath, P., Moineau, S., Mojica, F. J. M., Scott, D., Shah, S. A., Siksnys, V., et al. (2020) Evolutionary classification of CRISPR-cas systems: A burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83
- Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M. R., Vogel, J., and Charpentier, E. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607
- Sashital, D. G., Jinek, M., and Doudna, J. A. (2011) An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat. Struct. Mol. Biol.* **18**, 680–687
- Jore, M. M., Lundgren, M., van Duijn, E., Bultema, 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., et al. (2011) Structural basis for CRISPR RNA-guided DNA recognition by cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536
- Sternberg, S. H., Haurwitz, R. E., and Doudna, J. A. (2012) Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**, 661–672
- Richter, H., Lange, S. J., Backofen, R., and Randau, L. (2013) Comparative analysis of Cas6b processing and CRISPR RNA stability. *RNA Biol.* **10**, 700–707
- Charpentier, E., Richter, H., van der Oost, J., and White, M. F. (2015) Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**, 428–441
- Behler, J., and Hess, W. R. (2020) Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods* **172**, 12–26
- Wang, R., Preamplume, G., Terns, M. P., Terns, R. M., and Li, H. (2011) Interaction of the Cas6 ribonuclease with CRISPR RNAs: Recognition and cleavage. *Structure* **19**, 257–264
- Hochstrasser, M. L., and Doudna, J. A. (2015) Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* **40**, 58–66
- Reeks, J., Naismith, J. H., and White, M. F. (2013) CRISPR interference: A structural perspective. *Biochem. J.* **453**, 155–166
- Carte, J., Wang, R., Li, H., Terns, R. M., and Terns, M. P. (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496
- Carte, J., Pfister, N. T., Compton, M. M., Terns, R. M., and Terns, M. P. (2010) Binding and cleavage of CRISPR RNA by Cas6. *RNA* **16**, 2181–2188
- Lintner, N. G., Kerou, M., Brumfield, S. K., Graham, S., Liu, H., Naismith, J. H., Sdano, M., Peng, N., She, Q., Copie, V., Young, M. J., White, M. F., and Lawrence, C. M. (2011) Structural and functional characterization of an archaeal clustered regularly interspaced short palindromic repeat (CRISPR)-associated complex for antiviral defense (CASCADE). *J. Biol. Chem.* **286**, 21643–21656
- Kunin, V., Sorek, R., and Hugenholtz, P. (2007) Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61
- Scholz, I., Lange, S. J., Hein, S., Hess, W. R., and Backofen, R. (2013) CRISPR-Cas systems in the cyanobacterium *Synechocystis* sp. PCC6803 exhibit distinct processing pathways involving at least two Cas6 and a Cmr2 protein. *PLoS One* **8**, e56470
- Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358



30. Jackson, R. N., Golden, S. M., van Erp, P. B., Carter, J., Westra, E. R., Brouns, S. J., van der Oost, J., Terwilliger, T. C., Read, R. J., and Wiedenheft, B. (2014) Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* **345**, 1473–1479
31. Chowdhury, S., Carter, J., Rollins, M. F., Golden, S. M., Jackson, R. N., Hoffmann, C., Nosaka, L., Bondy-Denomy, J., Maxwell, K. L., Davidson, A. R., Fischer, E. R., Lander, G. C., and Wiedenheft, B. (2017) Structure reveals mechanisms of viral suppressors that intercept a CRISPR RNA-guided surveillance complex. *Cell* **169**, 47–57.e11
32. Nam, K. H., Haitjema, C., Liu, X., Ding, F., Wang, H., DeLisi, M. P., and Ke, A. (2012) Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* **20**, 1574–1584
33. Garside, E. L., Schellenberg, M. J., Gesner, E. M., Bonanno, J. B., Sauder, J. M., Burley, S. K., Almo, S. C., Mehta, G., and MacMillan, A. M. (2012) Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. *RNA* **18**, 2020–2028
34. Koo, Y., Ka, D., Kim, E. J., Suh, N., and Bae, E. (2013) Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *J. Mol. Biol.* **425**, 3799–3810
35. Hochstrasser, M. L., Taylor, D. W., Kornfeld, J. E., Nogales, E., and Doudna, J. A. (2016) DNA targeting by a minimal CRISPR RNA-guided cascade. *Mol. Cell* **63**, 840–851
36. Punetha, A., Sivathanu, R., and Anand, B. (2014) Active site plasticity enables metal-dependent tuning of Cas5d nuclease activity in CRISPR-Cas type I-C system. *Nucleic Acids Res.* **42**, 3846–3856
37. Burne, R. A. (1998) Oral streptococci... products of their environment. *J. Dent. Res.* **77**, 445–452
38. van der Ploeg, J. R. (2007) Genome sequence of *Streptococcus mutans* bacteriophage M102. *FEMS Microbiol. Lett.* **275**, 130–138
39. van der Ploeg, J. R. (2009) Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* **155**, 1966–1976
40. Chen, J., Li, T., Zhou, X., Cheng, L., Huo, Y., Zou, J., and Li, Y. (2017) Characterization of the clustered regularly interspaced short palindromic repeats sites in *Streptococcus mutans* isolated from early childhood caries patients. *Arch. Oral Biol.* **83**, 174–180
41. Serbanescu, M. A., Cordova, M., Krastel, K., Flick, R., Beloglazova, N., Latos, A., Yakunin, A. F., Senadheera, D. B., and Cvitkovitch, D. G. (2015) Role of the *Streptococcus mutans* CRISPR-Cas systems in immunity and cell physiology. *J. Bacteriol.* **197**, 749–761
42. O'Brien, R. E., Santos, I. C., Wrapp, D., Bravo, J. P. K., Schwartz, E. A., Brodbelt, J. S., and Taylor, D. W. (2020) Structural basis for assembly of non-canonical small subunits into type I-C cascade. *Nat. Commun.* **11**, 5931
43. Maris, C., Dominguez, C., and Allain, F. H. (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J.* **272**, 2118–2131
44. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
45. Holm, L., and Rosenstrom, P. (2010) Dali server: Conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549
46. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., and Ben-Tal, N. (2016) ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350
47. Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415
48. Yang, W. (2011) Nucleases: Diversity of structure, function and mechanism. *Q. Rev. Biophys.* **44**, 1–93
49. van der Oost, J., Westra, E. R., Jackson, R. N., and Wiedenheft, B. (2014) Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **12**, 479–492
50. Donis-Keller, H., Maxam, A. M., and Gilbert, W. (1977) Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res.* **4**, 2527–2538
51. Calvin, K., and Li, H. (2008) RNA-splicing endonuclease structure and function. *Cell. Mol. Life Sci.* **65**, 1176–1185
52. Raines, R. T. (1998) Ribonuclease A. *Chem. Rev.* **98**, 1045–1066
53. Gesner, E. M., Schellenberg, M. J., Garside, E. L., George, M. M., and Macmillan, A. M. (2011) Recognition and maturation of effector RNAs in a CRISPR interference pathway. *Nat. Struct. Mol. Biol.* **18**, 688–692
54. Reimann, V., Alkhnbashi, O. S., Saunders, S. J., Scholz, I., Hein, S., Backofen, R., and Hess, W. R. (2017) Structural constraints and enzymatic promiscuity in the Cas6-dependent generation of crRNAs. *Nucleic Acids Res.* **45**, 915–925
55. Vercoe, R. B., Chang, J. T., Dy, R. L., Taylor, C., Gristwood, T., Clulow, J. S., Richter, C., Przybilski, R., Pitman, A. R., and Fineran, P. C. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet.* **9**, e1003454
56. Hidalgo-Cantabrana, C., Goh, Y. J., Pan, M., Sanozky-Dawes, R., and Barrangou, R. (2019) Genome editing using the endogenous type I CRISPR-Cas system in *Lactobacillus crispatus*. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15774–15783
57. Xu, Z., Li, M., Li, Y., Cao, H., Miao, L., Xu, Z., Higuchi, Y., Yamasaki, S., Nishino, K., Woo, P. C. Y., Xiang, H., and Yan, A. (2019) Native CRISPR-cas-mediated genome editing enables dissecting and sensitizing clinical multidrug-resistant *P. aeruginosa*. *Cell Rep.* **29**, 1707–1717.e3
58. Zheng, Y., Han, J., Wang, B., Hu, X., Li, R., Shen, W., Ma, X., Ma, L., Yi, L., Yang, S., and Peng, W. (2019) Characterization and repurposing of the endogenous type I-F CRISPR-Cas system of *Zymomonas mobilis* for genome engineering. *Nucleic Acids Res.* **47**, 11461–11475
59. Csorgo, B., Leon, L. M., Chau-Ly, I. J., Vasquez-Rifo, A., Berry, J. D., Mahendra, C., Crawford, E. D., Lewis, J. D., and Bondy-Denomy, J. (2020) A compact cascade-Cas3 system for targeted genome engineering. *Nat. Methods* **17**, 1183–1190
60. Dolan, A. E., Hou, Z., Xiao, Y., Gramelspacher, M. J., Heo, J., Howden, S. E., Freddolino, P. L., Ke, A., and Zhang, Y. (2019) Introducing a spectrum of long-range genomic deletions in human embryonic stem cells using type I CRISPR-cas. *Mol. Cell* **74**, 936–950.e5
61. Morisaka, H., Yoshimi, K., Okuzaki, Y., Gee, P., Kunihiro, Y., Sonpho, E., Xu, H., Sasakawa, N., Naito, Y., Nakada, S., Yamamoto, T., Sano, S., Hotta, A., Takeda, J., and Mashimo, T. (2019) CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* **10**, 5302
62. Young, J. K., Gasior, S. L., Jones, S., Wang, L., Navarro, P., Vickroy, B., and Barrangou, R. (2019) The repurposing of type I-E CRISPR-cascade for gene activation in plants. *Commun. Biol.* **2**, 383
63. Cameron, P., Coons, M. M., Klompe, S. E., Lied, A. M., Smith, S. C., Vidal, B., Donohoue, P. D., Rotstein, T., Kohrs, B. W., Nyer, D. B., Kennedy, R., Banh, L. M., Williams, C., Toh, M. S., Irby, M. J., et al. (2019) Harnessing type I CRISPR-Cas systems for genome engineering in human cells. *Nat. Biotechnol.* **37**, 1471–1477
64. Zhang, R. G., Skarina, T., Katz, J. E., Beasley, S., Khachatryan, A., Vyas, S., Arrowsmith, C. H., Clarke, S., Edwards, A., Joachimiak, A., and Savchenko, A. (2001) Structure of *Thermotoga maritima* stationary phase survival protein SurE: A novel acid phosphatase. *Structure* **9**, 1095–1106
65. Snider, J., Gutsche, I., Lin, M., Baby, S., Cox, B., Butland, G., Greenblatt, J., Emili, A., and Houry, W. A. (2006) Formation of a distinctive complex between the inducible bacterial lysine decarboxylase and a novel AAA+ ATPase. *J. Biol. Chem.* **281**, 1532–1546
66. Beloglazova, N., Petit, P., Flick, R., Brown, G., Savchenko, A., and Yakunin, A. F. (2011) Structure and activity of the Cas3 HD nuclease MJ0384, an effector enzyme of the CRISPR interference. *EMBO J.* **30**, 4616–4627
67. Rosenbaum, G., Alkire, R. W., Evans, G., Rotella, F. J., Lazarski, K., Zhang, R. G., Ginell, S. L., Duke, N., Naday, I., Lazarz, J., Molitsky, M. J., Keefe, L., Gonczy, J., Rock, L., Sanishvili, R., et al. (2006) The structural biology center 19ID undulator beamline: Facility specifications and protein crystallographic results. *J. Synchrotron Radiat.* **13**, 30–45
68. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: The integration of data reduction and structure solution–

## Structure and activity of the Cas5c ribonuclease SMU1763

- from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 859–866
69. Schneider, T. R., and Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779
70. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763
71. Cohen, S. X., Morris, R. J., Fernandez, F. J., Ben Jelloul, M., Kakaris, M., Parthasarathy, V., Lamzin, V. S., Kleywegt, G. J., and Perakakis, A. (2004) Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2222–2229
72. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
73. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
74. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383
75. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667
76. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–10041
77. Senadheera, M. D., Guggenheim, B., Spatafora, G. A., Huang, Y.-C. C., Choi, J., Hung, D. C. I., Treglown, J. S., Goodman, S. D., Ellen, R. P., and Cvitkovitch, D. G. (2005) A VicRK signal transduction system in *Streptococcus mutans* affects *gtfBCD*, *gbpB* and *fff* expression, biofilm formation and genetic competence development. *J. Bacteriol.* **187**, 4064–4076