THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 283, NO. 29, pp. 20361–20371, July 18, 2008
Printed in the U.S.A.

A Novel Family of Sequence-specific Endoribonucleases Associated with the Clustered Regularly Interspaced Short Palindromic Repeats*[S]

Received for publication, April 28, 2008, and in revised form, May 14, 2008 Published, JBC Papers in Press, May 15, 2008, DOI 10.1074/jbc.M803225200

Natalia Beloglazova[‡§], Greg Brown[‡§], Matthew D. Zimmerman[¶‖], Michael Proudfoot[‡§], Kira S. Makarova[**], Marina Kudritska[‡§‖], Samvel Kochinyan[‡§], Shuren Wang[¶‖], Maksymilian Chruszcz[¶‖], Wladek Minor[¶‖], Eugene V. Koonin[1], Aled M. Edwards[‡§‖], Alexei Savchenko[‡§‖], and Alexander F. Yakunin[‡§2]**

From the [‡]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada, [§]Structural Proteomics in Toronto, Ontario Cancer Institute, Max Bell Research Centre 5R407, Toronto, Ontario M5G 2C4, Canada, the [¶]Department of Molecular Physiology and Biological Physics, University of Virginia, Charlottesville, Virginia 22908, the [‖]Midwest Center for Structural Genomics, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, and the [**]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

Clustered regularly interspaced short palindromic repeats (CRISPRs) together with the associated CAS proteins protect microbial cells from invasion by foreign genetic elements using presently unknown molecular mechanisms. All CRISPR systems contain proteins of the CAS2 family, suggesting that these uncharacterized proteins play a central role in this process. Here we show that the CAS2 proteins represent a novel family of endoribonucleases. Six purified CAS2 proteins from diverse organisms cleaved single-stranded RNAs preferentially within U-rich regions. A representative CAS2 enzyme, SSO1404 from *Sulfolobus solfataricus*, cleaved the phosphodiester linkage on the 3′-side and generated 5′-phosphate- and 3′-hydroxyl-terminated oligonucleotides. The crystal structure of SSO1404 was solved at 1.6 Å resolution revealing the first ribonuclease with a ferredoxin-like fold. Mutagenesis of SSO1404 identified six residues (Tyr-9, Asp-10, Arg-17, Arg-19, Arg-31, and Phe-37) that are important for enzymatic activity and suggested that Asp-10 might be the principal catalytic residue. Thus, CAS2 proteins are sequence-specific endoribonucleases, and we propose that their role in the CRISPR-mediated anti-phage defense might involve degradation of phage or cellular mRNAs.

Numerous prokaryotic genomes contain structures known as clustered regularly interspaced short palindromic repeats (CRISPRs),[3] composed of 25–50-bp direct repeats separated by intervening sequence spacers (or inserts) of similar length (1–3). CRISPRs represent the most widely distributed family of repeats in prokaryotes and are found in a great variety of bacteria and archaea, including many pathogens (such as *Mycobacterium tuberculosis, Streptococcus, Yersinia, Neisseria,* and *Corynebacterium*) (4–8). Most organisms possess multiple CRISPR clusters (2–20) that form tandem arrays containing from 4 to over 100 units (a repeat and a spacer) (1, 9–11). Sequence analysis of the CRISPR spacers in various organisms revealed that some of the spacer sequences are homologous (up to 100% sequence identity) to sequences in phage genomes, plasmids, or in the chromosome and have either sense or antisense orientations (5, 12–14). Remarkably, phages and plasmids fail to infect the specific strains carrying the cognate spacers, implying a relationship between CRISPRs and immunity against specific foreign DNA. Recent experiments on *Streptococcus thermophilus* have directly demonstrated that, after a phage challenge, this bacterium adds new CRISPR spacers derived from the phage genome (15, 16). Removal or addition of the phage-specific spacers correlated with the cell resistance to the respective phage implying that resistance specificity is determined by the sequence identity between the phage genome and spacers (15).

On the chromosome, CRISPR loci are flanked by a large number of *cas* (CRISPR-associated) genes encoding uncharacterized proteins. A comprehensive bioinformatic analysis of the CAS system in sequenced genomes resulted in a refined classification with 25 gene families and at least nine types of the *cas* operon organization (9, 17). Eight CAS protein families have been predicted to possess nuclease activity; nine families have been characterized as putative RNA-binding proteins (RAMP-domain proteins), and two families have been predicted to pos-

* This work was supported, in whole or in part, by National Institutes of Health Grants GM62414 and GM074942 from the Protein Structure Initiative (Midwest Center for Structural Genomics). This work was also supported by Genome Canada (through the Ontario Genomics Institute), the Ontario Research and Development Fund, and the Canadian Foundation for Innovation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
The atomic coordinates and structure factors (code 2i8e) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).
[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Table 1, Fig. 1, and additional references.
[1] Supported by the Intramural Research Program of the National Institutes of Health (National Library of Medicine, National Center for Biotechnology Information).
[2] To whom correspondence should be addressed. Tel.: 416-978-4013; Fax: 416-978-8528; E-mail: a.iakounine@utoronto.ca.

[3] The abbreviations used are: CRISPR, clustered regularly interspaced short palindromic repeats; CAS, CRISPR-associated; PAA, polyacryalamide; PAGE, polyacrylamide gel electrophoresis; PNK, polynucleotide kinase; ssRNA, single-stranded RNA; TA, toxin-antitoxin; nt, nucleotide; r.m.s.d., root mean square deviation; MES, 4-morpholineethanesulfonic acid; CAPS, 3-(cyclohexylamino)propanesulfonic acid.

TABLE 1
Nucleotide sequences of ssRNA substrates and SSO1404 cleavage sites

Substrate	Nucleotide sequence (5′ → 3′) and cleavage sites[a]
RNA1 (24 nt)	UUUCAAU ↓ UCCUUUUAGGAUUAAUC
RNA2 (25 nt)	CUUUCAA ↓ U ↓ UCCUUUUAGGAUUAAUC
RNA3 (38 nt)	UUGAAGAU ↓ AG ↓ AG ↓ UU ↓ AAAUGAACUUUUGAUGACACAAAA
RNA4 (41 nt)	GUCUACA ↓ C ↓ C ↓ UCUUGCGUCUAUUAUGAUAUCAUAUUUGUCAA
RNA5 (39 nt)	AAAUACG ↓ U ↓ U ↓ UUCUCCAUUGUCAUAUUGCGCAUAAGUUGA
RNA6 (39 nt)	CUAUUCU ↓ CCAAAU ↓ AAAAU ↓ U ↓ AU ↓ AUAACGCAUUAACAUCAA
RNA7 (36 nt)	GUUUUUGUACU ↓ CUCAAGA ↓ UUUAAGUAACUGUACAAC
RNA8 (30 nt)	GAGCUACCAGCU ↓ ACCCCGUAUGUCAGAGAG
RNA9 (30 nt)	CGUUCCUU ↓ UUUUCAAGGUAAUCUUUGAAAG
RNA10 (30 nt)	AAGUCCGUAAGCACCAG ↓ UUCCAAUCGUCAU
RNA11 (35 nt)	AUCA ↓ U ↓ UUUCAUGUCACUUUUAGUGACAUGUUUUAC
RNA12 (32 nt)	CAUUU ↓ UCUU ↓ AACACUG ↓ UACUUAACUUUUACUA
RNA13 (32 nt)	AUCA ↓ U ↓ UUUGUUAAGUACAGUGUUAAGU ↓ U ↓ UUCC
RNA14 (36 nt)	GUUCUCUCAAUG ↓ UCACGU ↓ CUCAUCACCUCUUCUUCG
RNA15 (37 nt)	AGAAAGU ↓ AGACGACAAGUUCGUUGAUUGUCUCGUCGC
RNA16 (36 nt)	ACGAGGAGUU ↓ U ↓ UACAUGUUUUAUGAGUCUAAUUUAG
RNA17 (36 nt)	GUUUAUAA ↓ UGGGA ↓ UA ↓ U ↓ UGGAGAAAAUGGAGAUAACA
RNA18 (16 nt)	AAAUACGUU ↓ UUCUCCA
RNA19 (16 nt)	CAUUUUCU ↓ UAACACUG
RNA20 (16 nt)	CGUUCCUU ↓ UUUUCAAG
RNA21 (13 nt)	UACGUU ↓ UUCUCCA
RNA22 (10 nt)	UACGUU ↓ UUCU
RNA23 (8 nt)	CGUUUUCU
RNA24 (6 nt)	GUUUUC
RNA25 (30 nt)	UUUCAAUU ↓ C ↓ CUUUUAGGAUUAAUCUUGAAG
RNA26 (40 nt)	UUUCAA ↓ U ↓ UCCUUUUAGGAUUAAUCU ↓ UGAAGAUAGAGUUAA

[a] Cleavage sites are shown by arrows (boldface arrows indicate the preferred cleavage sites). The substrates correspond to CRISPR DNA sequences from the several organisms. *S. solfataricus* CRISPR cluster-2 is as follows: RNA1, repeat-1; RNA2, repeat-2; RNA3, spacer-1; RNA4, spacer-2; RNA5, spacer-3; RNA18, RNA21, RNA22, RNA23, and RNA24, 5′- and/or 3′-truncated RNA5; RNA25, RNA1 containing a 6-nt fragment of the spacer-1; RNA26, RNA1 containing a 16-nt fragment of RNA1. *A. fulgidus* CRISPR cluster-2 is as follows: RNA6, spacer-1. *S. thermophilus* LMG18311 CRISPR cluster-1: RNA7, repeat-1; RNA8, spacer-1; RNA9, spacer-2; RNA10, spacer-3; RNA20, the 5′-fragment (16 nt) of RNA9. *T. maritima* CRISPR cluster-8: RNA14, spacer-5; RNA15, spacer-6. *M. jannaschii* CRISPR cluster-6 is as follows: RNA16, spacer-11; RNA17, spacer-10. Scrambled RNA substrates are as follows: RNA11, RNA12, and RNA13; RNA19, 5′-fragment of RNA12.

sess helicase and DNA/RNA polymerase activity (17). This analysis, combined with the data on the homology of some CRISPR spacer sequences to phage genes, led to the hypothesis that the CRISPRs and *cas*-encoded proteins comprise a system of defense against invading phages and plasmids and that this system might operate analogously to the eukaryotic RNA interference systems (17). The wide distribution of the CRISPR-CAS system among archaea and bacteria, its apparent importance for immunity of prokaryotes against infectious agents, and the predicted novel mechanism have recently made this system a subject of intense interest (3).

Two *cas* genes (*cas1* and *cas2*) are always located near a CRISPR locus and are found only in species containing CRISPRs suggesting that these proteins play a central role in the CRISPR system (1, 17). The members of the CAS2 superfamily are small, uncharacterized proteins (80–120 residues), which belong to COG1343 and COG3512 groups of the COG protein classification system (18). CAS2 proteins contain several conserved sequence motifs, in particular an N-terminal motif that consists of a string of hydrophobic residues (a predicted β-strand), and typically ends with an aspartate (17). The CAS2 protein sequences show some similarity to the sequences of the VapD family of uncharacterized proteins that are functionally linked to the VapBC toxin-antitoxin (TA) operon (19). Based on the pattern of conserved amino acid residues, in particular the presence of a conserved aspartate after a predicted β-strand, and some functional clues on the TA systems, it has been hypothesized that both CAS2 and VapD might possess RNase activity (17).

Here we report for the first time the results of biochemical and structural characterization of a family of CRISPR-associated enzymes, the CAS2 family proteins from five prokaryotes. We show that CAS2 proteins are endoribonucleases that are specific to single-stranded (ss)RNAs and preferentially cleave them within U-rich regions. The crystal structure of a representative CAS2 protein, SSO1404 from *Sulfolobus solfataricus*, was solved to a 1.6 Å resolution and revealed a ferredoxin-like fold with the double split β-α-β motif, as well as the putative active site.

EXPERIMENTAL PROCEDURES

Protein Overexpression, Purification, and Site-directed Mutagenesis—The cloning of the genes encoding SSO1404 and other CAS2 proteins (SSO8090, TM1796, AF1876, MTH1083, and NE0845) into the modified pET15b was carried out as described previously (20). The proteins were expressed as a fusion with an N-terminal His_6 tag in *Escherichia coli* strain BL21 (DE3) and purified to more than 95% homogeneity using metal-chelate affinity chromatography on nickel affinity resin and gel filtration on a Superdex 200 26/60 column (Amersham Biosciences) as described before (20, 21). Site-directed mutagenesis of SSO1404 was performed as described previously (21) using a protocol based on the QuikChange site-directed mutagenesis kit (Stratagene).

Preparation of RNA Substrates—The short RNA substrates (Table 1) were purchased from IDT. The oligonucleotides were 5′-end-labeled with [γ-^{32}P]ATP (6,000 Ci/mmol; Amersham Biosciences) and T4 polynucleotide kinase (PNK) (Fermentas) and then purified by denaturing PAGE (15% polyacrylamide, 8 M urea gel). The labeled oligonucleotides were eluted from the gel, precipitated with 2% $LiClO_4$ in acetone, washed with acetone, dried, and dissolved in diethyl pyrocarbonate-treated Milli-Q water. The long RNA substrates were synthesized using the Ambion T7 RNA polymerase MAXIscript transcription kit.

For the synthesis of the mouse β-actin mRNA (304 nt) fragment, pTRI-actin-mouse DNA (MAXIscript kit, Ambion) was used as a template. To produce long CRISPR RNA substrates, a DNA template was prepared by cloning the 1,292-bp-long 5′-fragment of the *S. solfataricus* CRISPR cluster-2 DNA into the pBluescript SK+ vector (Stratagene). During the transcription reaction (MAXIscript kit, Ambion), this DNA template (linearized by EcoRV) directed the synthesis of a 270-nt-long RNA substrate containing the CRISPR cluster-2 upstream region, repeat-1, spacer-1, repeat-2, and a 27-nt fragment of spacer-2. Long RNA substrates were uniformly labeled during transcription using [α-^{32}P]UTP (3,000 Ci/mmol). The labeled RNA substrates were purified by denaturing 8% PAA, 8 M urea gel, eluted, ethanol-precipitated, and stored at −70 °C. Double-stranded RNA substrates were prepared by the incubation of equimolar amounts (10 μM) of the 5′-^{32}P-labeled sense and unlabeled antisense RNA oligoribonucleotides (RNA5, RNA25, and RNA26) in 50 mM Tris-HCl (pH 7.0), 100 mM KCl at 90 °C for 1 min and slow cooling at room temperature for 1.5 h. The formation of RNA duplexes was verified by electrophoresis in 10% nondenaturing PAA gels.

Enzymatic Assays—The reaction mixture for RNase assays (10 μl) contained 0.1 μM [^{32}P]RNA, 50 mM Tris-HCl (pH 8.5), 100 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, and 0.01–0.1 μg of enzyme. The pH dependence of SSO1404 was characterized using three buffers: MES-K (pH 5.5 to 6.5), Tris-HCl (pH 7.0 to 9.0), and CAPS-K (pH 9.4 to 11.0). The reaction mixture was incubated at 37 °C for the indicated period of time and quenched by the addition of equal volume of formamide loading buffer (80% formamide, 0.025% bromphenol blue, 0.025% xylene cyanol, and 10 mM EDTA (pH 8.0)). The reaction products were resolved by electrophoresis in 15% PAA, 8 M urea gels using TBE (10 mM Tris borate (pH 8.3), and 2 mM EDTA) as a running buffer. As nucleotide size markers, an imidazole ladder or a G-ladder produced by partial RNA cleavage by 2 M imidazole or RNase T1, respectively, was used (22, 23). For the analysis of the RNA product 5′-end, after RNase reaction, RNA products were precipitated by 2% $LiClO_4$, washed by acetone, dried, dissolved in Milli-Q water, and phosphorylated with [γ-^{32}P]ATP and T4 PNK using conditions for forward or phosphate exchange reaction according to the manufacturer's protocol (Fermentas). After the PNK reaction, RNA products were analyzed using denaturing 15% PAA, 8 M urea gels as described above for RNase assays. The reaction mixtures for DNase assays (40 μl) contained 50 mM HEPES-K buffer (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol, 0.3 μg of λ DNA (double-stranded DNA) or 0.75 μg of M13 DNA (ssDNA), and 1–4 μg of enzyme. After 1 h of incubation at 37 °C, the reactions were quenched by the addition of 6× DNA loading dye (Fermentas) and analyzed on EtBr-stained 1% agarose gels.

Protein Crystallization and Structure Determination—SSO1404 crystals were grown using the hanging drop vapor diffusion method with the drops containing a mixture of 2 μl of 10 mg/ml purified selenomethionine-incorporated SSO1404 protein and 2 μl of reservoir buffer (0.2 M NaI, 20% w/v PEG 3350, and 2% v/v isopropyl alcohol). For diffraction studies, the crystals were stabilized with the crystallization buffer supplemented with 20% ethylene glycol as a cryoprotectant and flash-frozen in liquid nitrogen. A single crystal of selenomethionine-incorporated SSO1404 was used to collect diffraction data at beamline 19-BM of the Structural Biology Center of the Advanced Photon Source (24) and was maintained at a temperature of 100 K. A single-wavelength anomalous diffraction dataset was collected at a wavelength of 0.9794 Å. Crystallographic data collection and model refinement statistics are summarized in Table 2. Reflection data were collected, indexed, integrated, and scaled with HKL-3000 (25).

TABLE 2
Crystallographic data collection and refinement statistics

Data collection	SSO1404 (Protein Data Bank 2i8e)
Resolution range (Å)	50–1.59 (1.65–1.59)
Wavelength (Å)	0.9794
Space group	$P6_2$
Unit cell parameters	$a = b = 64.3$ Å, $c = 39.5$ Å $\alpha = \beta = 90°$, $\gamma = 120°$
No. of reflections	12,619 (1219)
R_{merge}	0.052 (0.309)
$R_{merge,a}$ [a]	0.038 (0.293)
Mean $I/\sigma I$	63.8 (8.2)
Completeness (%)	99.5 (96.2)
Redundancy	11.7 (10.3)
Wilson B factor (Å^2)	27.3
Refinement	
R (%)	18.8
R_{free} (%)	22.7
Protein atoms/AU	782
Heterogen (iodide) atoms/AU	4
Solvent atoms/AU	75
Mean B factor (Å^2)	20.9
r.m.s.d. bond lengths (Å)	0.016
r.m.s.d. angles (°)	1.69
Ramachandran plot	
Favored regions (%)	100.0
Allowed regions (%)	0.0
Outliers (%)	0.0

[a] $R_{merge,a}$ is calculated for unmerged anomalous pairs.

A two-site selenium substructure was determined; the structure was phased by single-wavelength anomalous diffraction, and an initial model was built. All structure solution and initial model building was performed by HKL-3000, which is integrated with SHELXD, SHELXE, MLPHARE, DM, O, COOT, SOLVE, RESOLVE, and ARP/wARP (25–32). The initial model was improved by iterative cycles of manual rebuilding in COOT, followed by maximum likelihood refinement with REFMAC5 (33). In later stages of refinement, a multigroup TLS model generated by the TLSMD web server was used to further improve the model. The final model was validated using Molprobity (34), SFCHECK (35), and PROCHECK. The atomic coordinates and structure factors for SSO1404 have been deposited in the Protein Data Bank with the accession code 2i8e.

RESULTS AND DISCUSSION

Enzymatic Activity of CAS2 Proteins—To characterize the biochemical activity of CAS2 proteins, we cloned and purified six members of this family from different organisms as follows: SSO1404 and SSO8090 from *S. solfataricus*, AF1876 from *Archaeoglobus fulgidus*, TM1796 from *Thermotoga maritima*, MTH1083 from *Methanobacterium thermoautotrophicum*, and NE0845 from *Nitrosomonas europaea*. Given the prediction that CAS2 proteins might possess nuclease activity (17),

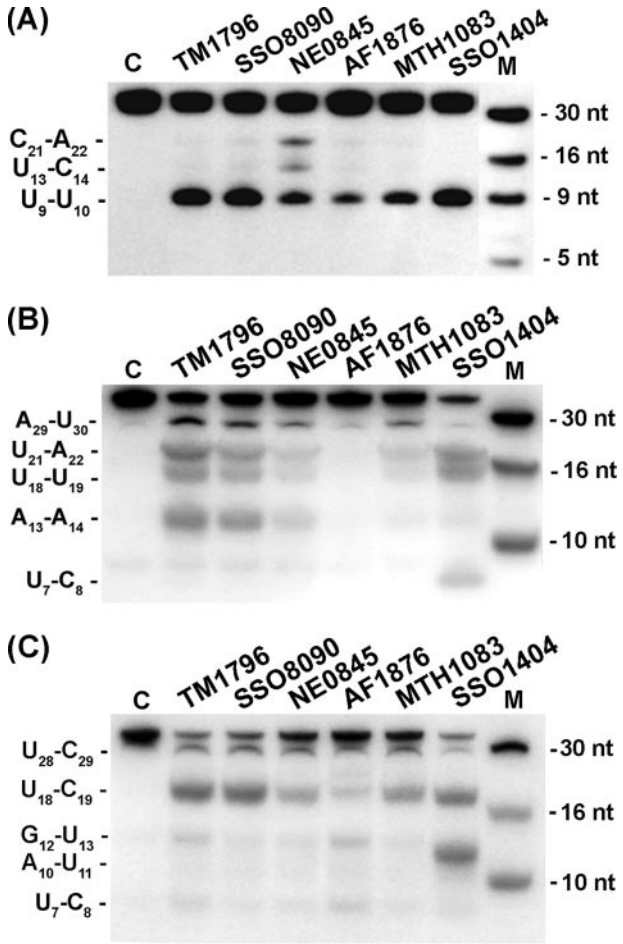


FIGURE 1. **Cleavage of ssRNAs by CAS2 proteins SSO1404, SSO8090, TM1796, NE0845, AF1876, and MTH1083.** Autoradiograph of the 15% PAA, 8 M urea gels showing the hydrolysis of RNA5 (*A*), RNA6 (*B*), and RNA14 (*C*) by the indicated CAS2 proteins. RNA substrates (0.1 μM) were incubated at 37 °C for 45 min with the CAS2 protein (50 μg/ml) in the presence of 50 mM Tris-HCl (pH 8.5), 5 mM $MgCl_2$, and 100 mM KCl. Reaction products were separated on a 15% PAAG, 8 M urea gel as described under "Experimental Procedures." *Lane C* represents the RNA sample incubated in the absence of protein; *lane M* shows the 5′-^{32}P-labeled marker oligonucleotides.

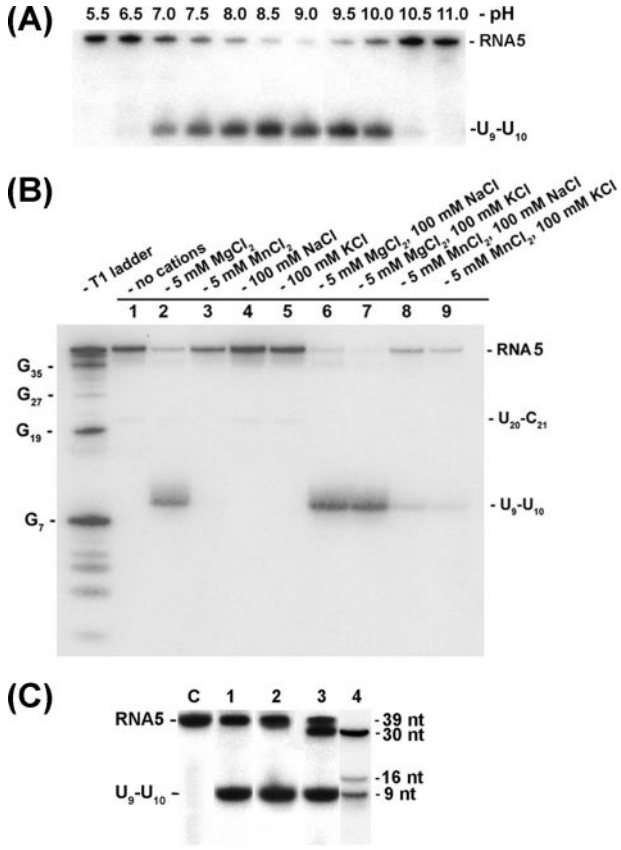


FIGURE 2. **Reaction requirements and RNA cleavage products of SSO1404.** Autoradiographs of 15% PAA, 8 M urea gels showing the effect of pH (*A*) or various metal cations (*B*) on the cleavage of RNA5 by SSO1404 (30 μg/ml). *A,* 5′-^{32}P-labeled RNA5 (0.1 μM) was incubated with SSO1404 at 37 °C for 10 min in a reaction mixture containing 50 mM buffer with the indicated pH (described under "Experimental Procedures"), 5 mM $MgCl_2$, 100 mM KCl, and 1 mM dithiothreitol. *B,* 5′-^{32}P-labeled RNA5 (0.5 μM) was incubated with SSO1404 in the absence or presence of 5 mM $MgCl_2$, 5 mM $MnCl_2$, 100 mM NaCl, or 100 mM KCl. *C,* analysis of the RNA5 cleavage products generated by SSO1404: *lane 1*, reaction mixture without the addition of PNK; *lane 2,* reaction mixture treated with PNK (5′-end labeling reaction); *lane 3,* reaction mixture treated with PNK (5′-phosphate exchange reaction). *Lane C* represents the RNA sample incubated in the absence of protein; *lane M* shows the 5′-^{32}P-labeled marker oligonucleotides. Reaction products were processed and visualized as described under "Experimental Procedures." The position of the main RNA5 cleavage product (U_9–U_{10}) is shown on the *right*. A T1 ladder represents the products of RNA5 cleavage by RNase T1.

the purified proteins were tested for the presence of DNase or RNase activity against single-stranded (ss) DNA (M13 DNA), double-stranded (ds) DNA (λ DNA), or ssRNA as follows: RNA5, RNA6, and RNA14 (Table 1), which are identical in sequence to the sense strands of the CRISPR repeats or spacers from *S. solfataricus*, *A. fulgidus*, or *T. maritima* (36 or 39 nt). No nuclease activity was found against either of the DNA substrates, but all proteins degraded the ssRNAs (Fig. 1). With the tested ssRNA substrates, the CAS2 proteins generated a limited number (one to five) of products of various lengths (7–29 nt) indicating that they cleave ssRNAs endonucleolytically. A similar but not identical pattern of products was observed, and most cleavage sites contained one or two U (Fig. 1). This observation suggests that the CAS2 ribonucleases recognize similar RNA sequences but also display some difference in substrate preference. SSO1404 also showed detectable cleavage of a long model RNA substrate (the 304-nt transcript of the 5′-fragment of the mouse β-actin gene) and generated several products of various lengths (18–200 nt), but the activity was lower than that against short oligoribonucleotides (data not shown). Thus, the CAS2 family proteins are ssRNA-specific endoribonucleases.

Reaction Requirements and RNA Cleavage Products of SSO1404—SSO1404 exhibited RNase activity over a broad pH range (7.0–10.0) with maximum activity at pH 8.5–9.0 (Fig. 2*A*). Very little cleavage of RNA5 by SSO1404 (between U20 and C21) was observed in the absence of both monovalent and divalent cations, as well as in the presence of Na^+, K^+, or Mn^{2+} (Fig. 2*B*). Cleavage between U9 and U10 was greatly enhanced by the addition of Mg^{2+}. The Mg^{2+}-dependent activity was further stimulated by K^+, whereas Na^+ had inhibitory effect. Thus, SSO1404 requires Mg^{2+} (5 mM), K^+ (100 mM), and pH 8.5–9.0 for optimal cleavage of ssRNA.

RNases can be subdivided into two groups depending on the position of the cleavage of the phosphodiester linkage (36). Enzymes of the first class cleave the bond on the 3′-side (producing a 5′-phosphorylated second product) and include

numerous intracellular endo- and exoribonucleases. The enzymes of the second class (*e.g.* RNase A, RNase T1, and barnases) cleave the linkage on the 5′-side releasing products containing a cyclic 2′,3′-phosphodiester bond. The products of ssRNA cleavage by a representative CAS2 protein, SSO1404, were characterized using the T4 PNK-catalyzed reactions of RNA phosphorylation (at 5′-hydroxyl termini) and the phosphate exchange between the 5′-phosphate groups of the oligonucleotide substrate and ATP (37). After hydrolysis of the RNA5 substrate by SSO1404, the reaction products were incubated with PNK and [γ-^{32}P]ATP and analyzed by denaturing PAGE and autoradiography (Fig. 2*C*). Incubation with PNK did not produce a labeled 3′-end product (no second band on the gel), and the gel showed only one labeled band corresponding to the 5′-end product (Fig. 2*C*, *lane 2*). This suggests that the 3′-end product is phosphorylated on its 5′-end. The presence of this 3′-end product in the reaction mixture with SSO1404 was confirmed using the PNK-catalyzed phosphate exchange reaction between the RNA 5′-phosphate and [γ-^{32}P]ATP in the presence of ADP excess (38). This experiment produced two labeled products on the gel with the expected length of 9 and 30 nt (Fig. 2*C*, *lane 3*). Thus, SSO1404 and apparently other CAS2 enzymes are the 5′-phosphomonoester-producing endoribonucleases.

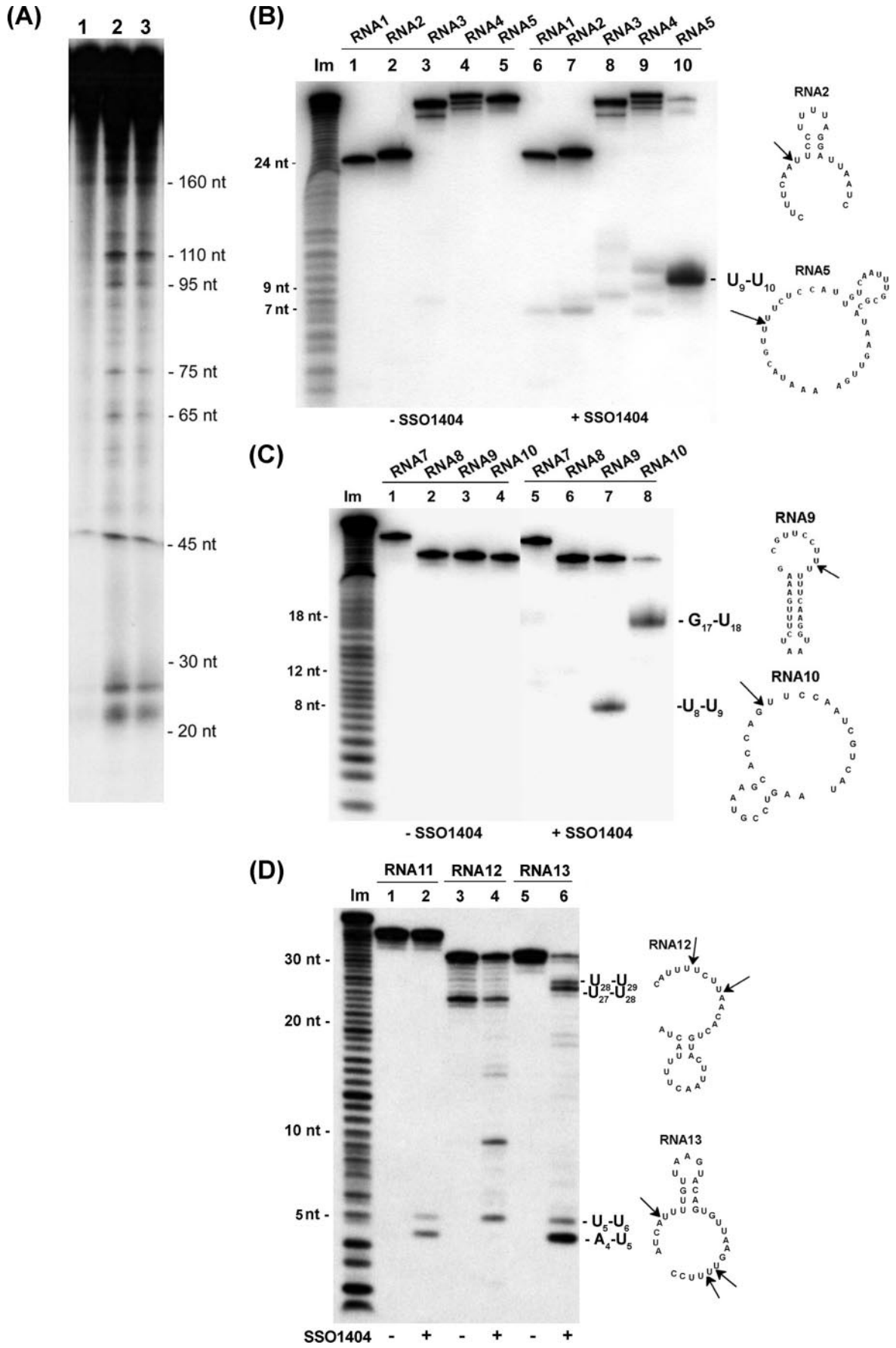
Endoribonuclease Activity of SSO1404 against Long CRISPR Substrates—Previous analyses of small RNAs produced in *S. solfataricus* and *A. fulgidus* revealed that the long CRISPR transcripts are processed into fragments with the size of one repeat and one spacer (62–75 nt) (39, 40). It has been proposed that these short CRISPR RNAs are produced by an unknown *cas* RNase that would cleave within the repeat sequence producing intact spacers flanked with the repeat fragments (3). The demonstration of ssRNase activity of SSO1404 suggested the possibility that CAS2 could be the RNase that is responsible for processing CRISPR RNAs. Thus, we tested SSO1404 for the ability to cleave the 270-nucleotide-long CRISPR transcript of the *S. solfataricus* CRISPR cluster-2. The uniformly labeled long CRISPR RNA substrate was prepared using T7 RNA polymerase and the 270-bp DNA fragment of the 5′-end of the *S. solfataricus* CRISPR cluster-2 containing the upstream region, repeat-1, spacer-1, repeat-2, and a 27-nt fragment of spacer-2 as the template. If SSO1404 were a CRISPR-processing endoribonuclease, it would be expected to cleave this substrate within the repeat sequences and produce a minimal product containing one spacer flanked by two repeat halves (63–66-nt-long), as well as a series of products containing two or more repeat + spacer units. SSO1404 showed detectable endoribonuclease activity with this long CRISPR RNA substrate, but it produced products that were smaller than expected (23, 27, and 45 nt long), as well as several longer fragments (65, 75, 95, 110, and 160 nt long) (Fig. 3*A*). This finding indicates that SSO1404 does not have the expected specificity toward CRISPR transcripts and thus is unlikely to be the CRISPR-processing endoribonuclease.

SSO1404 Cleaves ssRNAs Preferentially within U-rich Regions—The substrate specificity of SSO1404 was further characterized using an extended set of shorter synthetic ssRNA substrates, as well as several dsRNAs. The RNA1 to RNA5 substrates were identical in sequence to the sense strands of repeat-1 (24 nt), repeat-2 (25 nt), and three spacers (1–3) of the *S. solfataricus* CRISPR cluster-2 (Table 1). SSO1404 cleaved all these substrates, but the highest activity was observed with RNA5, which was preferentially cleaved between U9 and U10 (Fig. 3*B*). The hydrolysis of this substrate proceeded to near completion with the formation of one main labeled product.

Eight ssRNA substrates (RNA7 to RNA10 and RNA14 to RNA17) corresponded to repeats and spacers from other CRISPR-containing organisms (*S. thermophilus, T. maritima*, and *Methanococcus jannaschii*), whereas RNA11, RNA12, and RNA13 were scrambled RNA substrates (Table 1). SSO1404 demonstrated endonucleolytic activity toward most of these substrates with the highest activity toward RNA9, RNA10, and RNA13 (Fig. 3*C*). The analysis of the cleavage sites showed that SSO1404 preferred U-rich sequences, and in many substrates (*e.g.* RNA9, RNA10, RNA12, and RNA13) the cleavage site was located between two Us (Table 1). Examination of the computer-predicted secondary structures of the cleavable RNA substrates revealed that SSO1404 cleaves RNA substrates in predicted single-stranded regions (Fig. 3). No cleavage was observed with the dsRNA substrates prepared by annealing ssRNA substrates (RNA5, RNA25, and RNA26) with their complementary RNA chains and containing blunt ends or a two-nucleotide 5′-overhang (not shown). Thus, SSO1404 is a single-stranded RNA-specific endoribonuclease.

Although several substrates shown in Fig. 3 (RNA2, RNA5, and RNA12) contain a potential SSO1404 cleavage site within the predicted loops, the enzyme cleaved these substrates only in close proximity to the 5′- or 3′-ends suggesting that it requires the presence of a free RNA end. To determine whether a stem-loop is required as an entry site for SSO1404, we tested its activity against three truncated oligoribonucleotides (16 nt) that lack the potential to form stem-loops and contain the SSO1404 cleavage site from RNA5 (RNA18), RNA11 (RNA19), or RNA9 (RNA20). SSO1404 efficiently cleaved all three substrates at the same position as in the corresponding longer substrates (Table 1) indicating that the primary sequences of these short oligoribonucleotides were sufficient to direct site-specific cleavage by SSO1404, and no specific secondary structure was required. To determine the minimal size of the SSO1404 substrate, the 39-nt-long RNA5 was further trimmed to 16-nt RNA18, 13-nt RNA21, 10-nt RNA22, 8-nt RNA23, and 6-nt RNA24 (Table 1). SSO1404 efficiently cleaved the four longer substrates (RNA5, RNA18, RNA21, and RNA22) but showed no activity against RNA23 (8 nt) and RNA24 (6 nt) (Fig. 4). Thus, oligoribonucleotides as short as 10 nt can serve as SSO1404 substrates.

An ssRNA endoribonuclease activity with a preference for U-rich regions also has been observed in the abortive phage infection determinant AbiB from *Lactococcus lactis* and in mRNA interferases MazF-mt1 and MazF-mt6 from *M. tuberculosis* (41, 42). The *L. lactis* AbiB prevented growth of the sensitive phage bIL170 through the selective degradation of phage mRNAs by endonucleolytic cleavage at U/U, A/U, and U/A sites (41). The mRNA interferases are toxin components of chromosomal TA modules that are abundant in free-living prokaryotes and induce reversible cell cycle arrest or programmed cell death in response to starvation or other


(A)
1 2 3
- 160 nt
- 110 nt
- 95 nt
- 75 nt
- 65 nt
- 45 nt
- 30 nt
- 20 nt
(B)
RNA1 RNA2 RNA3 RNA4 RNA5 RNA1 RNA2 RNA3 RNA4 RNA5
Im 1 2 3 4 5 6 7 8 9 10
24 nt -
9 nt -
7 nt -
- U9-U10
RNA2
RNA5
- SSO1404
+ SSO1404
(C)
RNA7 RNA8 RNA9 RNA10 RNA7 RNA8 RNA9 RNA10
Im 1 2 3 4 5 6 7 8
18 nt -
12 nt -
8 nt -
- G17-U18
-U8-U9
RNA9
RNA10
- SSO1404
+ SSO1404
(D)
RNA11 RNA12 RNA13
Im 1 2 3 4 5 6
30 nt -
20 nt -
10 nt -
5 nt -
- U28-U
-U27-U29
- U5-U6
- A4-U5
RNA12
RNA13
SSO1404 - + - + - +

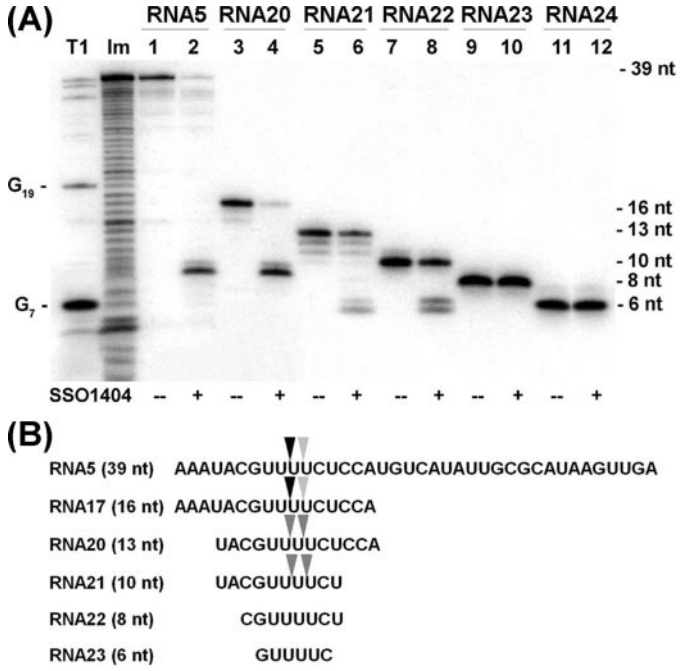


FIGURE 4. **Ribonuclease activity of SSO1404 against a series of the truncated RNA5 substrates.** *A,* autoradiograph of the 15% PAA, 8 M urea gel showing the activity of SSO1404 against the RNA5 substrate and its five truncated derivatives (the substrate lengths are shown on the *right*). *B,* sequences of the substrates used in the experiment and the positions of the cleavage sites of SSO1404. RNA substrates (0.1 μM) were incubated for 15 min (37 °C) in the absence (*lanes 1, 3, 5, 7, 9,* and *11*) or in the presence (*lanes 2, 4, 6, 8, 10,* and *12*) of SSO1404 (30 μg/ml). Reaction products were processed and visualized as described under "Experimental Procedures." *Lanes T1* and *Im* represent the products of the RNA5 cleavage by RNase T1 and 2 M imidazole (pH 7.0), respectively.

stress conditions (43–46). These enzymes have different mRNA cleavage specificities, and their expression in the cell causes the effective inhibition of protein synthesis leading to temporal cell growth arrest. In this regard, it might be relevant that U-rich and AU-rich regions have been identified upstream of the Shine-Dalgarno sequence in prokaryotic and phage mRNAs and shown to enhance translation (47, 48). The CAS2 proteins are unrelated to mRNA interferases or AbiB but, as mentioned above, appear to be homologous to VapD, a component of a distinct class of TA systems. Thus, CAS2 and VapD might represent a novel group of mRNA-specific endoribonucleases.

Crystal Structure of SSO1404—To elucidate the molecular basis for the ssRNase activity of CAS2, we determined the crystal structure of SSO1404 at 1.6 Å resolution. The structure demonstrated that SSO1404 is a homodimer (Fig. 5*A*). This is consistent with the results of our gel filtration experiments, which showed that the native protein has a molecular mass of 23.7 ± 0.6 kDa (predicted molecular mass 11.9 kDa). In the SSO1404 dimer, two monomers are joined by their β-sheets, and the α-helices are exposed on the surface. In the SSO1404 monomer, the four β-strands form an antiparallel β-sheet, and the α-helices are packed on one side, and the overall structure may be described as a double-split β-α-β fold. In addition, the last β-strand (β5) of each monomer interacts with the β-sheet of the other monomer creating two-joint, five-strand, antiparallel β-sheets (Fig. 5*A*).

A search for structural homologs of SSO1404 using the Dali (49) and secondary-structure matching (50) programs (Pro-Func data base (51) identified two uncharacterized proteins, TT1823 from *Thermus thermophilus* (Protein Data Bank code 1zpw, secondary-structure matching *Z*-score = 7.7, r.m.s.d. = 1.3 Å) and PF1117 from *Pyrococcus furiosus* (Protein Data Bank code 2i0x, secondary-structure matching *Z*-score = 7.0, r.m.s.d. = 1.7 Å) (Fig. 5, *B–D*). Like SSO1404, these two proteins belong to the CAS2 family (COG1343) (17). In addition, the Dali search recognized structures of domain III of the translation elongation factor G (EF-G) from *T. thermophilus* (Protein Data Bank code 1fnm) and elongation factor 2 from yeast (Protein Data Bank code 1n0v, *Z*-score = 6.1, r.m.s.d. = 2.3 Å) as significantly similar to SSO1404. All these structures have the double-split β-α-β fold, which is also observed in many ribosomal and other RNA-binding proteins (52). However, the ribosomal proteins and elongation factors do not have the conserved residues that are important for the RNase activity of SSO1404 (Tyr-9, Asp-10, Arg-17, Arg-19, Arg-31, and Phe-37) (see below).

Both Dali and VAST searches also recognized various proteins with the ferredoxin-like fold as being structural relatives of each of the three CAS2 structures with similar, moderate scores and r.m.s.d. values (supplemental Table 1), suggesting that CAS2 proteins comprise a new superfamily within the ferredoxin-like fold. Indeed, this is how the available CAS2 structures have been classified in the SCOP data base.

The ferredoxin-like domains consist of one of the most populated protein folds, with numerous structural and functional derivatives (53). In particular, this fold is present in numerous RNA-binding proteins, including the RNA-binding domain, the anticodon-binding domain of PheRS, ribosomal proteins S6 and S10, and also the prominent components of the CAS system, the RAMP superfamily proteins, which possess a duplicated ferredoxin-like domain (17, 54). However, to our knowledge, no nucleases with the ferredoxin-like fold have been characterized with the possible exception of the IS200 transposase of *S. solfataricus* (SSO1474, Protein Data Bank code 2f5g). Like many ribosomal proteins (55), SSO1404 exposes several aromatic and hydrophobic residues to the solvent (Ile-32, Tyr-34, Ile-75, Ile-84, Val-85, Ile-86, Phe-80, and Tyr-88), and some of these residues are conserved in large subsets of the CAS2 family proteins (supplemental Fig. 1), suggesting their involvement in the interaction with the bases of RNA. The extended loop α2-β4 of SSO1404 also could be a candidate for interaction with RNA.

FIGURE 3. **RNase activity of SSO1404 with CRISPR-related substrates.** *A,* cleavage of the 5′-fragment (270 nt) of the *S. solfataricus* CRISPR cluster-2 containing the cluster-2 upstream region, repeat-1, spacer-1, repeat-2, and a 27-nt fragment of spacer-2. Autoradiograph of the 8% PAA, 8 M urea gel showing the cleavage of the uniformly labeled [^{32}P]RNA (0.1 μM) by SSO1404 (30 μg/ml). RNA was incubated in the absence of SSO1404 (30 min, *lane 1*) or in the presence of SSO1404 for (15 min, *lane 2*) or (30 min, *lane 3*). *B,* RNA1 to RNA5; *C,* RNA6 to RNA9; *D,* RNA10 to RNA12. The *right part* of the of *B–D* shows predicted secondary structures and the positions of the SSO1404 cleavage sites in selected RNA substrates. *Im*, imidazole.

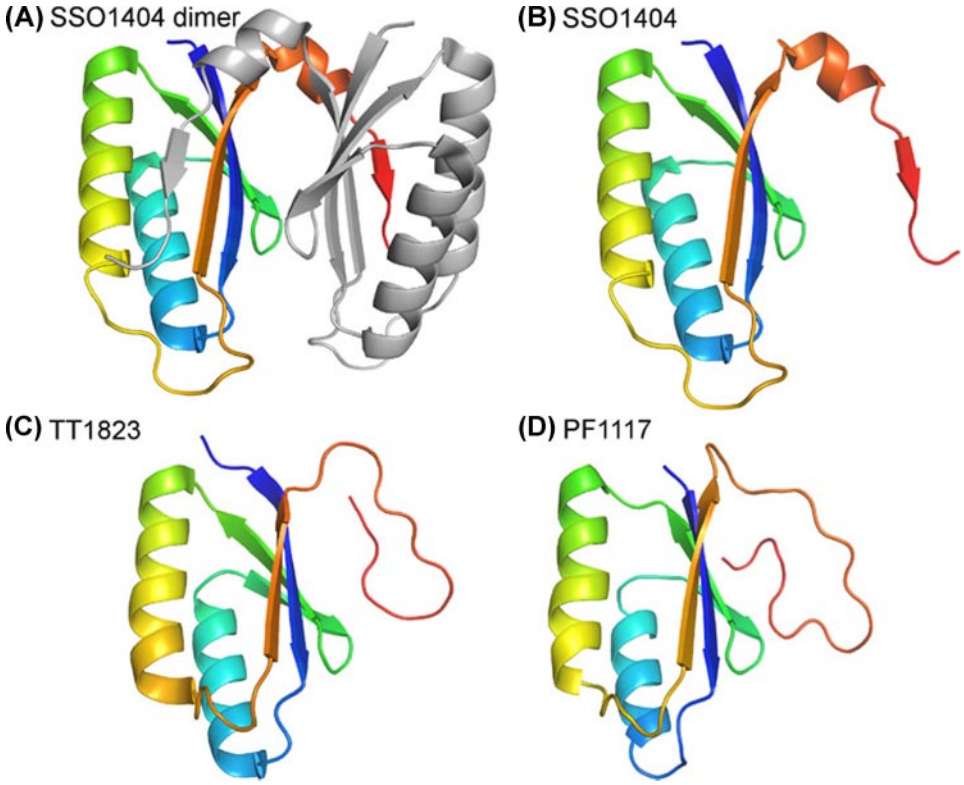


FIGURE 5. **Structures of SSO1404 and other CAS2 proteins.** *A,* SSO1404 dimer. One of the two monomers is colored under "rainbow" representation (*blue* at the N terminus to *red* at the C terminus). *B,* SSO1404 monomer. *C,* TT1823 monomer (Protein Data Bank code 1zpw). *D,* PF1117 monomer (Protein Data Bank code 2i0x). All of the polypeptides are shown in equivalent orientations and are colored under rainbow representation.

Mutational Analysis of SSO1404 and the Potential Catalytic Site—The structure of the SSO1404 dimer revealed several major cavities or grooves located at the interface of the two monomers, which might represent potential catalytic sites (Fig. 6*A*). The most prominent cavity consists of the deep cleft between two monomers accommodating the side chains of conserved Asp-10, which is bracketed by two α2-β4 loops (Fig. 6*A*). To identify the active site residues involved in RNA cleavage, we performed alanine replacement mutagenesis of the conserved residues of SSO1404. A multiple sequence alignment of CAS2 proteins (17) identified several conserved amino acids (Fig. 7; supplemental Fig. 1). The single conserved aspartate, Asp-10, was of special interest as the putative principal catalytic residue. In the SSO1404 structure, Asp-10 is located at the end of the first β-strand and is preceded by the conserved tyrosine Tyr-9. Several other partially conserved residues are spatially close to Asp-10, including Arg-17, Arg-19, and Arg-31. Alanine replacement mutagenesis of SSO1404 revealed that Asp-10, Arg-31, and Phe-37 are strictly required for the catalytic activity, whereas Y9A, R17A, and R19A showed very low activity (Fig. 8). The T12A and D65A mutants showed an ~2-fold drop in activity, whereas the other mutant proteins exhibited slightly reduced (Q33A, Y34A, and S35A) or wild-type level (D13A, D14A, N18A, and R67A) activity.

In the SSO1404 structure, five residues that are important for activity (Tyr-9, Asp-10, Arg-17, Arg-31, and Phe-37) are located in the long cavity formed by the α1 helix on one side, the β2 and β3 strands on the other side, and the β1 strand at the bottom (Fig. 6*B*). SSO1404 requires a divalent metal cation for activity, Mg^{2+} being the optimal metal (Fig. 2*B*). In most known RNases, two or three conserved acidic residues are critical for catalysis and are involved in the coordination of one or two metal cations (Mg^{2+} or Mn^{2+}), which activate a nucleophilic water molecule for hydrolysis of the phosphodiester bond or stabilize the transition state in cleavage reactions (56). The structures of SSO1404 and two other CAS2 proteins (TT1823 and PF1117) revealed no metal atoms bound to the protein. Nevertheless, given that SSO1404 required the addition of Mg^{2+} or Mn^{2+} for activity and the conserved Asp-10 was critical for catalysis, we suggest that this residue is involved in the coordination of the metal cation in SSO1404 and other CAS2 proteins. Interestingly, the side chains of Asp-10 in the two monomers are oriented toward the active site of the other monomer and are positioned 6.5 Å apart (Fig. 6*B*) suggesting that these residues do not form two independent active sites. In addition, these aspartates are located inside a deep, narrow cleft making them inaccessible for direct interaction with the RNA substrate (Fig. 6*A*). Therefore, we propose that the side chains of the two Asp-10 bind one Mg^{2+} atom, which might coordinate a nucleophilic water molecule. As shown here (Fig. 2*C*), SSO1404 generates products terminating in a 3′-hydroxyl and a 5′-phosphate suggesting that, like in the *E. coli* RNase E, the reaction involves an attack of a water molecule on the susceptible phosphodiester bond followed by scission of the 3′-oxygen–phosphorus bond (57). Therefore, the catalytic mechanism of SSO1404 probably involves the in-line nucleophilic attack on the scissile phosphate by an activated hydroxyl group generated from a water molecule that is coordinated to magnesium.

The SSO1404 D65A protein had a 2-fold reduced activity (Fig. 8). This conserved aspartate is located on the long α2-β4 loop (Fig. 6, *A* and *B*) that varies in length among CAS2 proteins and accommodates several partially conserved charged residues (Glu-64, Asp-65, Glu-66, and Arg-67 in SSO1404) (Fig. 7). Based on the structure of the SSO1404 dimer (Fig. 6, *A* and *B*), we suggest that the α2-β4 loops are responsible for the recognition of RNA substrates and might determine the substrate selectivity of different CAS2 proteins. We propose that like the *E. coli* RNase E S1 domain (58), the SSO1404 α2-β4 loop might function to clamp down the RNA substrate in the active site orienting its phosphate backbone for nucleophilic attack by the magnesium-coordinated activated water molecule.

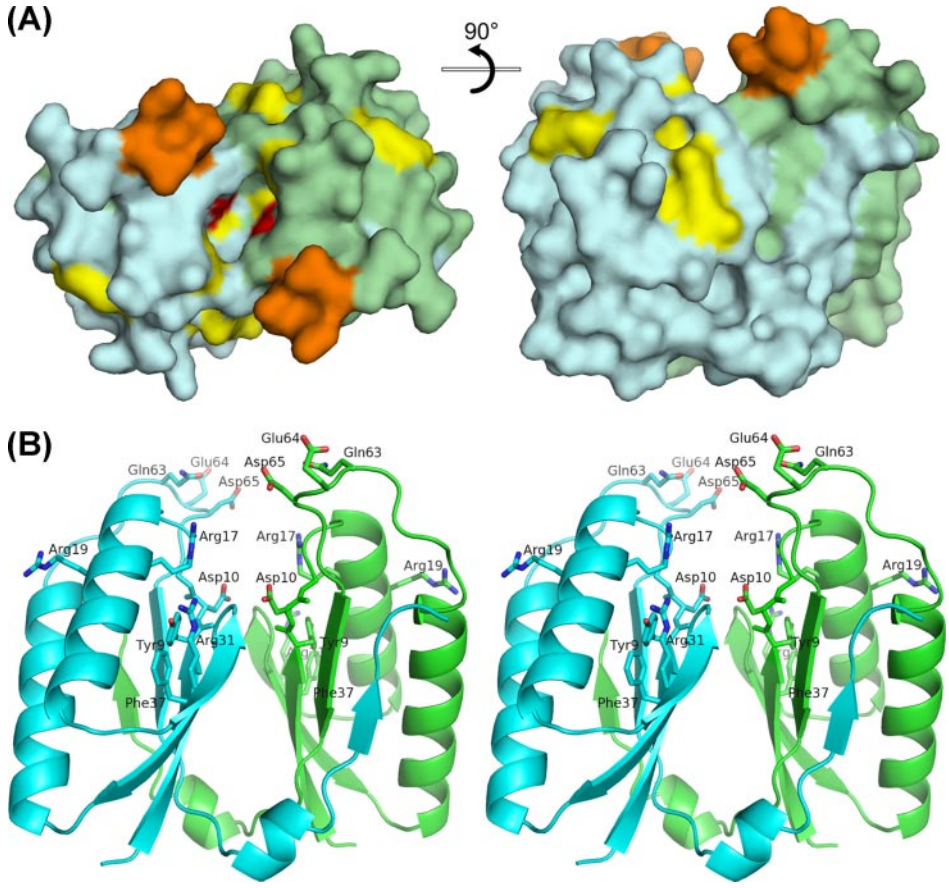


FIGURE 6. **Potential active site of SSO1404.** *A,* surface representation (two orientations) of the SSO1404 dimer showing the potential active site: the deep cavity between two monomers harboring two conserved Asp-10 (colored *red*). One of the monomers is colored in *cyan*, and the other is in *green*. Residues colored *orange* (Gln-63, Glu-64, and Asp-65) are located on the two $\alpha 2$-$\beta 4$ loops, which are predicted to be involved in RNA binding. Other conserved residues important for activity are shown in *yellow* (Tyr-9, Arg-17, Arg-19, Arg-31, and Phe-37). *B,* stereo view of the SSO1404 dimer showing the positions of the residues important for activity (Tyr-9, Asp-10, Arg-17, Arg-19, Arg-31, and Phe-37). One of the monomers is colored in *cyan*, and the other is in *green*. The orientation of the SSO1404 dimer in *B* corresponds to the right model of *A*. Note the proximity of the two principal catalytic Asp-10 residues in the dimer (6.5 Å).

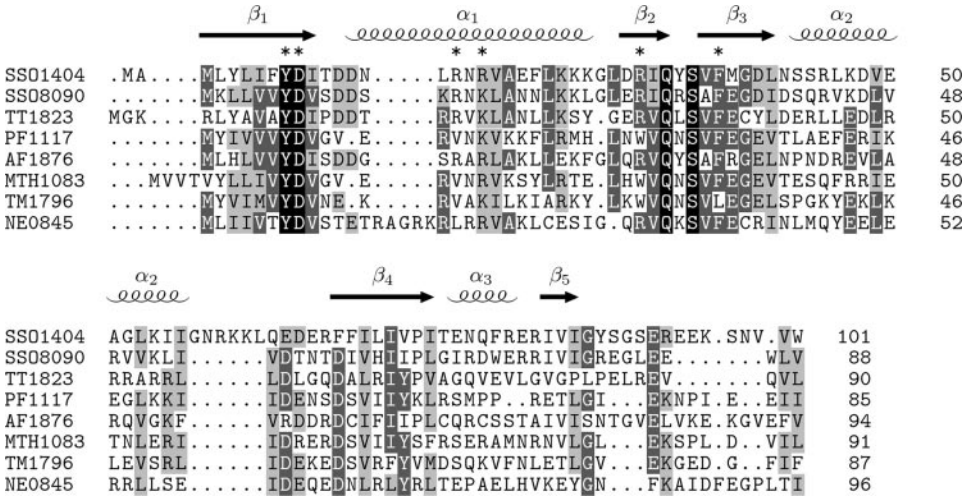


FIGURE 7. **Structure-based sequence alignment of SSO1404 and selected CAS2 proteins.** Residues conserved in all CAS2 proteins are highlighted in *black*, highly conserved residues in *dark gray*, and similar residues in *light gray*. The secondary structure elements derived from the structure of SSO1404 (Protein Data Bank code 2i8e) are shown above the alignment. The *asterisks* designate the residues important for the catalytic activity of SSO1404. The compared proteins are as follows: SSO1404 (Q97YC2), SSO8090 (Q97Y85), TT1823 from *T. thermophilus* (Q746F4), PF1117 from *P. furiosus* (Q8U1T8), AF1876 from *A. fulgidus* (O28403), MTH1083 from *M. thermoautotrophicum* (O27155), TM1796 from *T. maritima* (Q9X2B6), and NE0845 from *N. europaea* (Q82W51). The sequences were aligned using both sequence and structural information by 3DCOFFEE (63), and the figure was generated using TEXSHADE (64).

Conclusions—This work is the first step toward a comprehensive biochemical and structural characterization of CAS proteins. Recent bioinformatic analyses suggest that the CRISPR and CAS proteins might use several dissimilar mechanisms to abrogate phage infection and that at least one of these mechanisms could resemble eukaryotic RNA interference (3, 17). This study shows that SSO1404 and other CAS2 proteins represent a novel family of endoribonucleases, which possess a ferredoxin-like fold and are specific to ssRNA substrates. SSO1404 cleaves ssRNAs preferentially within U-rich regions and, in this respect, resembles the phage abortive infection endoribonuclease AbiB from *L. lactis* (41). A similar mechanism of selective degradation of phage transcripts might be proposed for SSO1404 and other CAS2 proteins as well. Under this hypothesis, CAS2 would be the functional analog of the eukaryotic slicer nuclease, a function that is performed by the PIWI domains that are unrelated to CAS2 (59). Considering the apparent mechanism of CRISPR-associated anti-phage defense that involves integration of sequences homologous to fragments of phage mRNAs into the CRISPR loci, the CAS2 endoribonuclease activity might contribute to this process as well. Another possible role of CAS2 proteins in anti-phage defense might be associated with the global inhibition of translation by mRNA cleavage, a mechanism that has been proposed for RelBE and several other TA systems (MazEF, PemIK, and ChpBIK) that contain RNase components (mRNA interferases) (45, 60, 61). These TA systems play important roles in stress response to nutritional limitations or DNA damage, and their expression results in growth arrest or programmed cell death (43–45). Moreover, it has been shown that the *E. coli* MazEF TA module prevents multiplication of the phage P1 by promoting cell death (62). A similar mechanism of anti-phage response appears to be a possibility

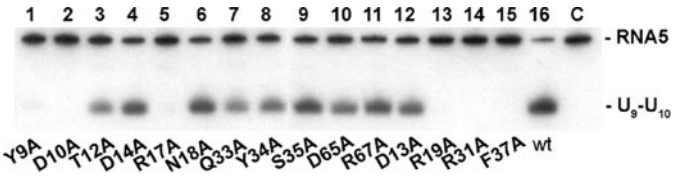


FIGURE 8. **Site-directed mutagenesis of SSO1404, RNase activity of purified mutant proteins with the RNA5 substrate.** Autoradiograph of the 15% PAA, 8 M urea gel showing the cleavage of 5-^{32}P-labeled RNA5 (0.1 μM) by purified proteins (30 μg/ml). Reaction mixtures were incubated at 37 °C for 10 min, and the reaction products were processed as described under "Experimental Procedures." *Lane C* represents the RNA5 sample incubated without protein addition.

for CAS2, especially in light of the previously described relationship with VapD, an uncharacterized protein that is functionally linked to the VapBC toxin-antitoxin system (17). *S. solfataricus* and many other CRISPR-containing organisms have at least two CAS2 proteins that might have different sequence specificities and could target distinct sets of mRNAs.

Acknowledgments—We thank all members of the Structural Proteomics in Toronto (SPiT) Centre, the Structural Biology Center at the Advanced Photon Source and Midwest Center for Structural Genomics, and Andrzej Joachimiak for help in conducting experiments and discussions. Crystallography results were derived from work performed at Argonne National Laboratory, Structural Biology Center, at the Advanced Photon Source. Argonne is operated by University of Chicago Argonne, LLC, for the United States Department of Energy, Office of Biological and Environmental Research, under Contract DE-AC02-06CH11357.

REFERENCES

1. Jansen, R., Embden, J. D., Gaastra, W., and Schouls, L. M. (2002) *Mol. Microbiol.* **43,** 1565–1575
2. Mojica, F. J., Diez-Villasenor, C., Soria, E., and Juez, G. (2000) *Mol. Microbiol.* **36,** 244–246
3. Sorek, R., Kunin, V., and Hugenholtz, P. (2008) *Nat. Rev. Microbiol.* **6,** 181–186
4. Groenen, P. M., Bunschoten, A. E., van Soolingen, D., and van Embden, J. D. (1993) *Mol. Microbiol.* **10,** 1057–1065
5. Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., and Soria, E. (2005) *J. Mol. Evol.* **60,** 174–182
6. Masepohl, B., Gorlitz, K., and Bohme, H. (1996) *Biochim. Biophys. Acta* **1307,** 26–30
7. Hoe, N., Nakashima, K., Grigsby, D., Pan, X., Dou, S. J., Naidich, S., Garcia, M., Kahn, E., Bergmire-Sweat, D., and Musser, J. M. (1999) *Emerg. Infect. Dis.* **5,** 254–263
8. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987) *J. Bacteriol.* **169,** 5429–5433
9. Haft, D. H., Selengut, J., Mongodin, E. F., and Nelson, K. E. (2005) *Plos Comput. Biol.* **1,** e60
10. Grissa, I., Vergnaud, G., and Pourcel, C. (2007) *BMC Bioinformatics* **8,** 172
11. Lillestol, R. K., Redder, P., Garrett, R. A., and Brugger, K. (2006) *Archaea* **2,** 59–72
12. Pourcel, C., Salvignol, G., and Vergnaud, G. (2005) *Microbiology* **151,** 653–663
13. Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S. D. (2005) *Microbiology* **151,** 2551–2561
14. Horvath, P., Romero, D. A., Coute-Monvoisin, A. C., Richards, M., Deveau, H., Moineau, S., Boyaval, P., Fremaux, C., and Barrangou, R. (2008) *J. Bacteriol.* **190,** 1401–1412
15. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007) *Science* **315,** 1709–1712
16. Deveau, H., Barrangou, R., Garneau, J. E., Labonte, J., Fremaux, C., Boyaval, P., Romero, D. A., Horvath, P., and Moineau, S. (2008) *J. Bacteriol.* **190,** 1390–1400
17. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I., and Koonin, E. V. (2006) *Biol. Direct* **1,** 7
18. Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2003) *BMC Bioinformatics* **4,** 41
19. Katz, M. E., Wright, C. L., Gartside, T. S., Cheetham, B. F., Doidge, C. V., Moses, E. K., and Rood, J. I. (1994) *J. Bacteriol.* **176,** 2663–2669
20. 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* (*Lond.*) **9,** 1095–1106
21. Proudfoot, M., Sanders, S. A., Singer, A., Zhang, R., Brown, G., Binkowski, A., Xu, L., Lukin, J. A., Murzin, A. G., Joachimiak, A., Arrowsmith, C. H., Edwards, A. M., Savchenko, A. V., and Yakunin, A. F. (2008) *J. Mol. Biol.* **375,** 301–315
22. Mironova, N. L., Pyshnyi, D. V., Shtadler, D. V., Fedorova, A. A., Vlassov, V. V., and Zenkova, M. A. (2007) *Nucleic Acids Res.* **35,** 2356–2367
23. Nishikawa, S., Morioka, H., Kim, H. J., Fuchimura, K., Tanaka, T., Uesugi, S., Hakoshima, T., Tomita, K., Ohtsuka, E., and Ikehara, M. (1987) *Biochemistry* **26,** 8620–8624
24. 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., Walsh, M. A., Westbrook, E., and Joachimiak, A. (2006) *J. Synchrotron Radiat.* **13,** 30–45
25. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **62,** 859–866
26. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. Sect. A* **47,** 110–119
27. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **55,** 849–861
28. Terwilliger, T. C. (2002) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **58,** 1937–1940
29. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat. Struct. Biol.* **6,** 458–463
30. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **60,** 2126–2132
31. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) *Methods Enzymol.* **374,** 229–244
32. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **58,** 1772–1779
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Sect. D. Biol. Crystallogr.* **53,** 240–255
34. Lovell, S. C., Davis, I. W., Arendall, W. B., III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50,** 437–450
35. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 191–205
36. Saida, F., Uzan, M., and Bontems, F. (2003) *Nucleic Acids Res.* **31,** 2751–2758
37. Novogrodsky, A., Tal, M., Traub, A., and Hurwitz, J. (1966) *J. Biol. Chem.* **241,** 2933–2943
38. Berkner, K. L., and Folk, W. R. (1980) *Methods Enzymol.* **65,** 28–36
39. Tang, T. H., Bachellerie, J. P., Rozhdestvensky, T., Bortolin, M. L., Huber, H., Drungowski, M., Elge, T., Brosius, J., and Huttenhofer, A. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 7536–7541
40. Tang, T. H., Polacek, N., Zywicki, M., Huber, H., Brugger, K., Garrett, R., Bachellerie, J. P., and Huttenhofer, A. (2005) *Mol. Microbiol.* **55,** 469–481
41. Parreira, R., Ehrlich, S. D., and Chopin, M. C. (1996) *Mol. Microbiol.* **19,** 221–230
42. Zhu, L., Zhang, Y., Teh, J. S., Zhang, J., Connell, N., Rubin, H., and Inouye, M. (2006) *J. Biol. Chem.* **281,** 18638–18643
43. Hayes, F. (2003) *Science* **301,** 1496–1499
44. Gerdes, K. (2000) *J. Bacteriol.* **182,** 561–572

45. Gerdes, K., Christensen, S. K., and Lobner-Olesen, A. (2005) *Nat. Rev. Microbiol.* **3,** 371–382
46. Inouye, M. (2006) *J. Cell. Physiol.* **209,** 670–676
47. Zhang, J., and Deutscher, M. P. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89,** 2605–2609
48. Olins, P. O., and Rangwala, S. H. (1989) *J. Biol. Chem.* **264,** 16973–16976
49. Holm, L., and Sander, C. (1994) *Proteins* **19,** 165–173
50. Krissinel, E., and Henrick, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2256–2268
51. Laskowski, R. A., Watson, J. D., and Thornton, J. M. (2005) *Nucleic Acids Res.* **33,** W89–W93
52. Ramakrishnan, V., and White, S. W. (1998) *Trends Biochem. Sci.* **23,** 208–212
53. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature* **372,** 631–634
54. Ebihara, A., Yao, M., Masui, R., Tanaka, I., Yokoyama, S., and Kuramitsu, S. (2006) *Protein Sci.* **15,** 1494–1499
55. Draper, D. E. (1999) *J. Mol. Biol.* **293,** 255–270
56. Worrall, J. A., and Luisi, B. F. (2007) *Curr. Opin. Struct. Biol.* **17,** 128–137
57. Redko, Y., Tock, M. R., Adams, C. J., Kaberdin, V. R., Grasby, J. A., and McDowall, K. J. (2003) *J. Biol. Chem.* **278,** 44001–44008
58. Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., and Luisi, B. F. (2005) *Nature* **437,** 1187–1191
59. Hutvagner, G., and Simard, M. J. (2008) *Nat. Rev. Mol. Cell Biol.* **9,** 22–32
60. Zhang, J., Zhang, Y., Zhu, L., Suzuki, M., and Inouye, M. (2004) *J. Biol. Chem.* **279,** 20678–20684
61. Zhang, Y., Zhu, L., Zhang, J., and Inouye, M. (2005) *J. Biol. Chem.* **280,** 26080–26088
62. Hazan, R., and Engelberg-Kulka, H. (2004) *Mol. Genet. Genomics* **272,** 227–234
63. Poirot, O., Suhre, K., Abergel, C., O'Toole, E., and Notredame, C. (2004) *Nucleic Acids Res.* **32,** W37–W40
64. Beitz, E. (2000) *Bioinformatics* (*Oxf.*) **16,** 135–139