Dr. Matthew Schellenberg
Department of Biochemistry
Supervisor: Dr. Andrew MacMillan

I am originally from Williams Lake BC, and I began my undergraduate training at the University of Northern British Columbia. After two years I transferred to the University of Alberta where I obtained my B.Sc. in the Department of Biochemistry. My first independent research project helped me realize that I have a passion for scientific research, so I entered into the graduate program in the Department of Biochemistry and joined Dr. MacMillan's lab studying the field of pre-mRNA splicing. I chose to join his lab because the field of RNA biology is challenging with new and exciting findings that greatly add to our knowledge of cellular mechanisms.

I completed my Ph.D. in the Department of Biochemistry in November of 2010, and I was fortunate to be hired by Dr. MacMillan as a post-doctoral fellow immediately after my graduation. This allowed me to complete the research projects I began as a graduate student, including work describing the structure and function of the Cse3 endonuclease.

Recently, bacteria were shown to contain a genetic element called CRISPR, which serves to protect the bacteria against other invasive genetic elements such as bacterial viruses called bacteriophages. The CRISPR pathway encoded by the CRISPR element performs a similar function to our own immune system, by recognizing and eliminating threats from invasive biological agents, but the mechanism by which this system functions is not yet really known.

In our paper, we provide important details about RNA recognition and processing by the Cse3 protein, which contributes to our knowledge of RNA biology and the mechanism of the CRISPR pathway. The food and biotech industries are increasingly making use of bacteria to make products essential for our everyday life and treatment of disease. A greater knowledge of how this pathway functions may allow us to use this as a tool, to decrease losses of products due to contamination by viruses, or even to manipulate infectious bacteria in other ways such as reducing their ability to require antibiotic resistance. Furthermore, although higher eukaryotes do not contain CRISPR sequences, there is evidence of small RNA mediated genome regulation with an unidentified mechanism. Details of CRISPR function may help us understand pathways of gene regulation that are yet to be identified in humans, but may have arisen by convergent evolution.

I currently hold a position as a post-doctoral fellow in the lab of Dr. Williams at the National Institute of Health (NIH/NIEHS), in North Carolina USA, where I study cellular DNA repair processes. We believe our research will lead to the development of the next generation of chemotheraputics that can be used to treat cancer.

nature structural & molecular biology

Recognition and maturation of effector RNAs in a CRISPR interference pathway

Emily M Gesner[1,2], Matthew J Schellenberg[1,2], Erin L Garside[1], Mark M George[1] & Andrew M MacMillan[1]

In bacteria and archaea, small RNAs derived from clustered, regularly interspaced, short palindromic repeat (CRISPR) loci are involved in an adaptable and heritable gene-silencing pathway. Resistance to phage infection is conferred by the incorporation of short invading DNA sequences into the genome as CRISPR spacer elements separated by short repeat sequences. Processing of long primary transcripts (pre-crRNAs) containing these repeats by an RNA endonuclease generates the mature effector RNAs that interfere with phage gene expression. Here we describe structural and functional analyses of the *Thermus thermophilus* CRISPR Cse3 endonuclease. High-resolution X-ray structures of Cse3 bound to repeat RNAs model both the pre- and post-cleavage complexes associated with processing the pre-crRNA. These structures establish the molecular basis of a specific CRISPR RNA recognition and suggest the mechanism for generation of effector RNAs responsible for gene silencing.

© 2011 Nature America, Inc. All rights reserved.

Individual CRISPR loci in bacterial and archaeal genomes feature a cluster of repeats separated by short spacer elements[1,2]. They also typically include a set of CRISPR-associated (*cas*) genes that code for proteins, a number of which have been shown to be necessary for CRISPR-based gene interference[3–7]. Specific subtypes of CRISPR systems are defined by the structure and organization of the repeats as well as by the particular set of accompanying protein-coding *cas* genes[8–10]. For example, the CRISPR-7 locus of the *T. thermophilus* megaplasmid pTT27 contains 21 palindromic repeats 28 nucleotides (nt) in length, predicted to form stem-loop structures, separated by spacers of 30–34 nt, flanked by a set of eight upstream *cas* genes (**Fig. 1a**)[11,12].

In *Escherichia coli*, the products of the *cas1* and *cas2* genes are not required to effect CRISPR-based immunity using existing CRISPR spacers and are therefore likely to function in the acquisition of new repeats[4,13]. The Cas3 protein has been implicated as the ultimate effector protein in gene silencing, with evidence suggesting it targets mature crRNAs to invading DNA sequences (**Fig. 1a**)[4,14].

Analysis of both wild-type and mutant *E. coli* strains suggests that an initial pre-crRNA is processed to yield mature crRNA by endonucleolytic cleavage at the base of the repeat stem loop[4]. The enzymatic activity required for this processing has been shown to reside in the Cse3 protein, a component of the multiprotein CRISPR-associated complex for antiviral defense (Cascade) complex containing Cse1–Cse4 and Cas5e (**Fig. 1a**)[4,15].

In order to elucidate the basis of pre-crRNA recognition and processing, we have functionally and structurally characterized *T. thermophilus* Cse3, a member of the Ecoli Cas protein subtype (**Supplementary Fig. 1**). Here we report high-resolution X-ray structures of Cse3 bound to RNAs, representing both the cognate pre-crRNA substrate and a mimic of the cleaved product as well as a structure of the actual Cse3–product complex. These structures demonstrate the basis of sequence-specific RNA recognition by Cse3 and suggest a catalytic mechanism for endonucleolytic cleavage. Finally, a comparison of Cse3 with two different pre-crRNA processing enzymes suggests both common and distinct features of RNA binding and cleavage across a large family of CRISPR endonucleases.

RESULTS

RNA binding and cleavage by Cse3

We cloned, expressed and purified *T. thermophilus* Cse3 (**Supplementary Fig. 1**). The recombinant protein specifically bound a 28-nt RNA corresponding to the *T. thermophilus* repeat (**Fig. 1b**) but did not bind unrelated single-stranded RNAs, hairpins or double-stranded RNAs (data not shown).

T. thermophilus Cse3 also cleaved the 28-nt RNA, at a single position, yielding two RNA products consistent with cleavage after G21 (**Fig. 1b**,**c**). The 3′ product could be 5′ end-labeled, but the 5′ product was resistant to periodate oxidation and base-mediated elimination, indicating a modified 3′ end (**Supplementary Fig. 2**). The generation of modified 2′ and/or 3′ and free 5′-hydroxyl termini on the longer and shorter products, respectively, is consistent with a cleavage mechanism involving in-line attack of the G21 2′-hydroxyl group on the scissile phosphate, as observed in both protein and RNA-catalyzed RNA cleavage[16,17]. Also consistent with this mechanism, substitution of a 2′-deoxy residue at the G21 position abolished cleavage in the presence of the enzyme (**Fig. 1d**).

Structural basis for sequence-specific RNA recognition by Cse3

The X-ray structure of *T. thermophilus* Cse3 alone, which we solved at 1.85-Å resolution, has previously been reported and features tandem

[1]Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada. [2]These authors contributed equally to this work. Correspondence should be addressed to A.M.M. (andrew.macmillan@ualberta.ca).

Received 1 January; accepted 17 February; published online 15 May 2011; doi:10.1038/nsmb.2042

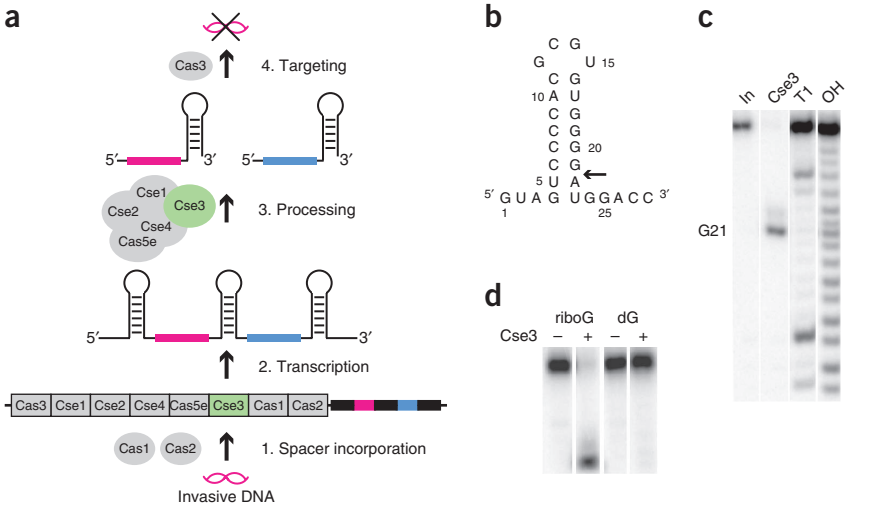

Figure 1 Processing of pre-crRNA by Cse3. (**a**) Schematic of the Ecoli subtype CRISPR-Cas system that includes *T. thermophilus* Cse3. The pre-crRNA transcript, processed by the Cascade complex consisting of Cse1–Cse4 and Cas5e, is endonucleolytically cleaved by Cse3 at the base of conserved stem-loop repeats to yield effector RNAs. Cas1 and Cas2 have been implicated in the incorporation of new spacers into the CRISPR loci, and Cas3 has been implicated in the targeting of invasive genetic elements. (**b**) Secondary structure of the *T. thermophilus* repeat RNA. The RNA is cleaved by Cse3 8 nt upstream of the spacer-repeat boundary (arrow). (**c**) Mapping of CRISPR repeat cleavage by Cse3. 5′-^{32}P-labeled RNA was incubated in the absence or presence of Cse3. The location of cleavage by Cse3 was mapped by comparison to RNase T1 and base hydrolysis ladders as resolved by denaturing PAGE. (**d**) Abolition of Cse3-mediated cleavage of a modified RNA. Substitution of the nucleotide immediately 5′ to the cleavage site with a deoxynucleotide prevents cleavage of the RNA in the presence of Cse3.

© 2011 Nature America, Inc. All rights reserved.

ferredoxin-like domains with characteristic βαββαβ folds, each consisting of a four-stranded β-sheet buttressed by two α-helices (**Supplementary Fig. 1**)[18]. Notable in both structures is a disordered loop extending from the adjacent β10-β11 strands; a disordered portion in the original structure forms the ordered β7–β8 hairpin in our structure. A similar overall domain structure characterizes the *Pyrococcus furiosus* Cas6 protein, a functional homolog of Cse3 (ref. 5). In contrast, the recently reported structure of Csy4, the Cse3 functional homolog from *Pseudomonas aeruginosa*, features a single N-terminal ferredoxin fold and a separate C-terminal domain that includes two α-helices joined to the main body by extended linker sequences[19]. RNA-binding by Sxl and PABP—each containing tandem RNA recognition motifs (RRMs), a subclass of the ferredoxin fold—is mediated across an extended β-surface, suggesting that the analogous surface of Cse3 might have a function in RNA binding[20,21].

We crystallized and solved the 3.2-Å structure of Cse3 bound to a noncleavable 19-nt RNA modeling a minimal CRISPR repeat and containing deoxyguanosine (deoxyG) at position 21 (**Fig. 2a**, **Table 1**, **Supplementary Figs. 3** and **4**); we also solved the 2.35-Å X-ray structure of Cse3 bound to the 18-nt RNA mimicking the 5′ cleavage product but lacking the 2′,3′ cyclic phosphodiester (**Table 1**, **Supplementary Fig. 3**). Finally, we incubated Cse3 with the 19-nt RNA containing riboguanosine (riboG) at position 21 under cleavage conditions, crystallized the resulting product complex and solved the 3.1-Å X-ray structure (**Table 1**, **Supplementary Fig. 3**). Although all three protein–RNA assemblies purified as 1:1 complexes, strand exchange on crystallization yielded a dimeric structure in the crystals of the precleavage and product complexes, but the product mimic complex lacked the loop nucleotides, presumably due to adventitious RNase activity (**Supplementary Fig. 3**). All three structures nevertheless include base-pairing of the predicted stem loop of the CRISPR repeat and thus represent the cognate complex. The overall architecture of the three complexes is very similar, specifically with respect to the details of protein-RNA recognition (**Fig. 2b**). The RNA is mounted on the protein opposite the extended β-surface. The largest number of protein-RNA contacts, including major-groove sequence-specific interactions, are made to the 3′ strand of the base-paired RNA that features an essentially A-form structure. This includes a series of interactions between side chains of the short β7–β8 hairpin represented by

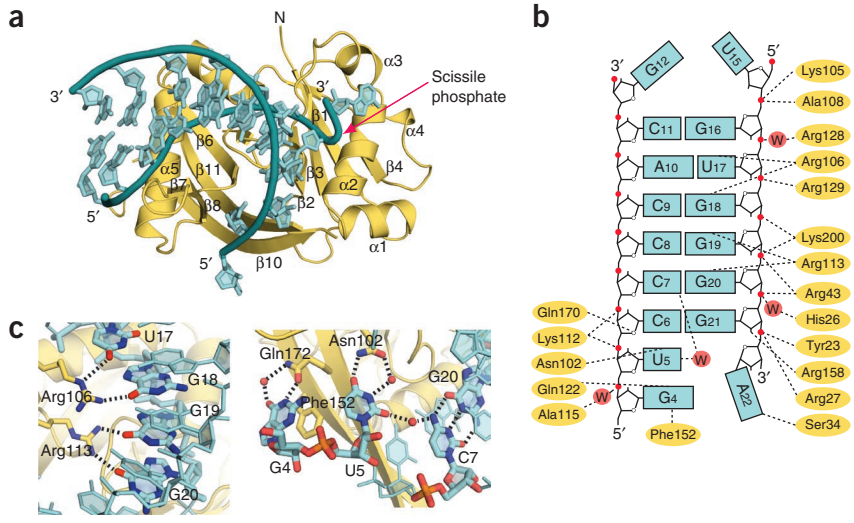

Figure 2 Structure of the Cse3 substrate–RNA complex. (**a**) Overview of the crystal structure of Cse3 bound to deoxyG, containing RNA. Cse3 is colored yellow and the RNA is colored cyan. (**b**) Summary of Cse3–RNA interactions. Represented are sequence-specific base interactions, contacts with ribose functionalities, and both direct and water-mediated contacts to the phosphodiester backbone (red). (**c**) Details of Cse3–RNA interactions. Left, sequence-specific recognition of the major-groove base pairs by arginine side chains on the β7–β8 hairpin. Right, sequence-specific recognition of the unpaired 5′ region of repeat RNA. Phe152 forms a stacking interaction with the purine base of G4 that is recognized by two hydrogen bonds to Gln172. U5 is recognized by hydrogen bonding to Asn102 and forms a water-mediated base triple with the C6–G21 base pair.

Table 1 Data collection and refinement statistics (molecular replacement)

	3′OH RNA	deoxyG RNA	Cut RNA
Data collection			
Space group	$P2_12_12_1$	$I222$	$I222$
Cell dimensions			
a, *b*, *c* (Å)	37.06, 72.41, 100.23	61.50, 68.91, 158.06	62.17, 68.75, 155.84
α, β, γ (°)	90, 90, 90	90,90,90	90,90,90
Wavelength (Å)[a]	0.97946	0.97945	1.11588
Resolution (Å)[b]	50–2.35	50–3.2	50–3.1
R_{sym}	0.112 (0.502)	0.168 (0.412)	0.176 (0.499)
$I / \sigma I$	9.9 (2.0)	9.8 (3.8)	9.2 (2.7)
Completeness (%)	95.5 (75.1)	97.6 (89.4)	98.4 (98.9)
Redundancy	3.5	7.8	4.2
Refinement			
Resolution (Å)	50–2.35	50–3.2	50–3.1
No. reflections	11135	5,452	5,979
R_{work} / R_{free}	0.175/.217	0.223/0.266	0.245/0.300
No. atoms	2,211	2,027	2,017
Protein	1,709	1,623	1,631
Ligand/ion	350	402	386
Water	152	0	0
B-factors			
Protein	23.6	26.0	24.2
Ligand/ion	25.7	38.8	32.3
Water	27.8	–	–
R.m.s. deviations			
Bond lengths (Å)	0.013	0.006	0.007
Bond angles (°)	1.52	1.16	0.657

[a]Each dataset was collected from a single crystal. [b]Values in parentheses are for the highest-resolution shell (2.43–2.35 Å for the 3′-OH structure, 3.31–3.20 Å for the deoxyG RNA structure and 3.21–3.10 Å for the cut RNA structure).

residues 105–115 and the major-groove functionalities of U17 and G18–G20 (**Fig. 2c**, left). This contrasts with recognition of the *P. aeruginosa* CRISPR repeat where major-groove recognition involves the extended linker and adjoining α-helix[19]. A recent structural analysis of Cas6 RNA interaction suggests that the 5′ end of a single-stranded repeat is anchored in a groove between the ferredoxin folds and traverses the protein to position the site of cleavage at the active site on the opposing surface of the protein[22].

The A-form duplex in the Cse3–RNA structures is disrupted at the base of the RNA stem by the insertion of the β10–β11 hairpin. At the 5′ end of the spacer, a base-triple interaction is formed, in which the exocyclic amine of C7 that participates in the C7-G20 base pair interacts with the O2 carbonyl of U5 through a water molecule (**Fig. 2c**, right). A series of protein contacts to nucleotides 4–6 of the repeat includes direct and indirect interaction with the phosphodiester backbone as well as specific recognition of the base functionalities of G4 and U5 (**Fig. 2b**,**c**).

Organization of the Cse3 active site

The structure of the Cse3–RNA complex at the 3′ end of the RNA has clear implications for the catalytic mechanism of RNA cleavage by the enzyme (**Fig. 3**). In the deoxyG-containing structure, although the C6-G21 base pair is contained within an A-form duplex, a sharp turn in the backbone following the deoxyG residue disrupts the predicted U5-A22 base pair (**Figs. 1b** and **2a**). Instead, the adenine base is held in position by hydrogen-bonding between the Watson-Crick face and the side chain of Ser34 (**Fig. 3a**). The position occupied by Ser30, below the plane of the base, is a tryptophan in most Cse3 homologs (**Supplementary Fig. 1**) and thus may further stabilize this conformation of the RNA by π stacking. An essential feature of the observed kinking of the backbone is the positioning of the scissile phosphate in a conformation consistent with in-line displacement of the A22 5′-oxy leaving group by the G21 2′-hydroxyl group (**Fig. 3a**).

The backbone conformation at the scissile phosphate suggests that the structure of the complex models an assembled active site primed for catalysis. Analysis of the homologous Cas6 protein structure suggested that a catalytic triad of tyrosine, lysine and histidine residues might function in the catalysis of repeat cleavage, in analogy to the mechanism used by the tRNA-splicing endonuclease[16,17]. The X-ray structure of this endonuclease bound to RNA suggests that tyrosine and histidine side chains act as general base and acid, respectively, in the deprotonation of the 2′-hydroxyl nucleophile and in the protonation of the 5′ leaving group; the lysine side chain is positioned to stabilize the transition-state intermediate in the cleavage reaction[17].

In the Cse3 active site, the conserved Tyr23 is positioned proximal to the scissile phosphate to stabilize the charged intermediate; the absolutely conserved His26 does not appear to have a direct function in catalysis

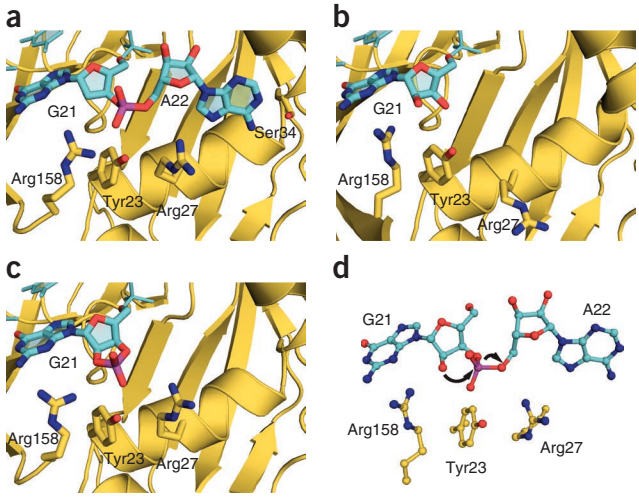

Figure 3 Structural basis for cleavage of pre-crRNA by Cse3. (**a**) Detail of the scissile phosphate environment in the structure of Cse3 bound to deoxyG-containing RNA. The conformation of the G21–A22 backbone is consistent with an in-line displacement mechanism by the 2′-hydroxyl of G21 that is absent in this substrate. Tyr23 is positioned to stabilize the charged transition-state intermediate through interaction with nonbridging phosphate oxygens. The side chain of Arg27 is positioned to stabilize the A22 5′-hydroxyl leaving group. The scissile phosphate is colored magenta. (**b**) Active-site environment in structure of Cse3 bound to RNA mimicking the 5′ cleavage product. The side chain of Arg158 is ordered and positioned to interact with the 2′-hydroxyl of G21. (**c**) Active-site environment in structure of Cse3 bound to RNA product. The side chain of Arg158 is ordered and positioned to interact with the 2′-oxygen of G21. The side chain of Tyr23 is positioned to stabilize the transition-state intermediate. (**d**) Catalytic model of the mechanism of RNA cleavage by Cse3 based on composite of the three Cse3–RNA structures reported here. Arg158 interacts with the 2′-hydroxyl nucleophile and Arg27 stabilizes the 5′ leaving group. The side chain of Tyr23 interacts with the nonbridging oxygens of the transition-state intermediate, which may also be stabilized by Arg27.

© 2011 Nature America, Inc. All rights reserved.

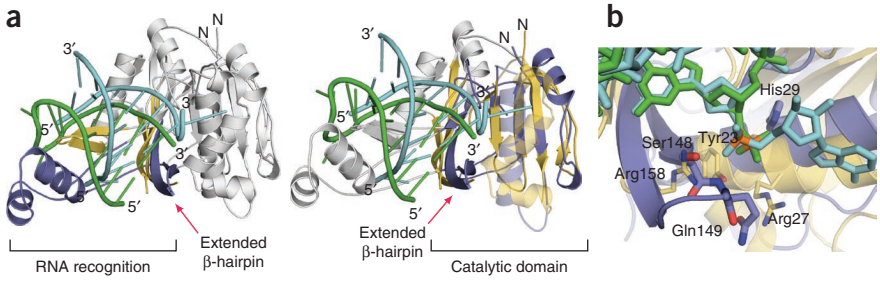

Figure 4 Modular organization of pre-crRNA recognition and processing revealed by comparison of Cse3 and Csy4 RNA complexes. (**a**) Superposition of cognate *P. aeruginosa* Csy4–RNA (PDB 2XLK[19]) and *T. thermophilus* Cse3–deoxyG RNA X-ray structures, highlighting RNA recognition by distinct C-terminal modules that specifically recognize the RNA major groove as well as by a common extended β-hairpin interacting with the base of the duplex proximal to the cleavage site (left), and the N-terminal catalytic domains (right). Csy4 and Cse3 are colored blue and yellow, respectively. The Csy4 RNA is colored green and the Cse3 RNA is colored cyan. The β-hairpin contributes to both RNA binding and the architecture of the active site. (**b**) Alignment of nucleotides proximal to the scissile phosphates of the Csy4 and Cse3 RNA substrates. RNA-dependent ordering of the extended hairpin (β10–β11 in Cse3) includes the positioning of the Cse3 Arg158 side chain to interact with the 2′-hydroxyl nucleophile. In Csy4, the analogous position in the corresponding hairpin formed by β7 and β8 is occupied by Ser148, which is proposed to have a similar role[19]. The main chain N-H of Csy4 Gln149 is positioned to stabilize the cleavage transition-state intermediate analogous to the role proposed here for Tyr23 of Cse3, and the side chain of Csy4 His29 is positioned to stabilize the leaving group, filling the proposed role of Arg27 in Cse3. Csy4 and Cse3 are colored blue and yellow, respectively. The Csy4 RNA is colored green, the Cse3 RNA is colored cyan and the scissile phosphates are colored orange.

© 2011 Nature America, Inc. All rights reserved.

but interacts through a bound water molecule with the phosphate 1 nt 5′ to the cleavage site. This is reminiscent of the direct interaction of an invariant histidine in the active site of topoisomerase II with the DNA backbone[23]. The side chain of Arg27 is positioned to both stabilize the transition-state intermediate and the departing leaving group (**Fig. 3a**).

In the product mimic Cse3–RNA complex, the side chain of Arg158 is positioned to interact with the 2′ oxygen of G21. This side chain is less ordered in the deoxyG-containing structure, presumably because of the lack of a substituent at the G21 2′ position (**Fig. 3b**). One potential consequence of the Arg158-RNA interaction is the ordering of residues 158–170 to form an extended β-hairpin that is disordered in the structure of Cse3 alone (**Supplementary Figs. 1** and **3**). Because processing of pre-crRNA by Cse3 takes place within the context of the multiprotein Cascade complex, the RNA-dependent ordering and extension of the β10–β11 hairpin may facilitate downstream events in the CRISPR pathway.

The crystallized Cse3 product complex contains the 18-nt 5′ product of cleavage (**Supplementary Fig. 3**). Consistent with the expected mechanism and the results of our chemical modification experiment, we observe a cyclic 2′,3′-phosphodiester in this structure. The disposition of Tyr23 and Arg158 is also consistent with their playing a direct role in the cleavage chemistry (**Fig. 3c**).

A composite based on the three Cse3–RNA structures yields a model of the assembled active site in which Arg158 and Arg27 stabilize the nucleophile and leaving group, respectively, and Tyr23 interacts with the backbone nonbridging oxygens of the transition-state intermediate (**Fig. 3d**). We confirmed the importance of the proposed active site residues by functional analysis of the Y23F, R27A and R158A Cse3 mutants. Although none of these was seriously compromised in binding the CRISPR repeat, the observed rate of cleavage was ~1% that of the wild type for each of the mutants (**Supplementary Fig. 5**), in contrast to the 15% of wild-type activity observed upon mutation of Arg157—which is involved in stabilization of the β10–β11 hairpin—to alanine.

DISCUSSION

The model of the Cse3 active site derived from our structural analysis places two arginine residues, Arg158 and Arg27, in the position expected to be occupied by a general base or acid. The side chain of Arg27 may act as a general acid in the protonation of the leaving group. With respect to the role of Arg158, although there are some examples of arginine residues implicated as general bases in enzyme active sites[24], the high pK_a of the arginine side chain and the solvent exposure of Arg158 argue against such a mechanism in this case. A plausible alternative is that both residues function as metal surrogates, with a role similar to that of divalent metal ions in the classic 'two-metal mechanism' common to many protein- and RNA-based phosphoryl transfers[25].

Genetic and structural evidence suggest an evolutionary relationship between the pre-crRNA processing activity in distinct CRISPR systems[5,8,9,19]. Superposition of the Cse3 structures reported here with the high-resolution structures of both the Csy4–RNA complex and Cas6 endonuclease (**Fig. 4**, **Supplementary Fig. 6**) suggest the evolution of distinct RNA binding mechanisms from a common platform, provide a basis for modeling the active site in the latter proteins and also indicate a coupling of specific RNA recognition to catalysis of RNA cleavage.

All three proteins, Cse3, Csy4 and Cas6, have a bimodal structural organization consisting of N-terminal ferredoxin-like folds that include elements of the endonuclease active site and distinct C-terminal RNA-binding modules (**Fig. 4a**)[5,19]. In Cse3 and Csy4, highly specific RNA binding is largely mediated through major-groove recognition by distinct modules—a short β-hairpin and an α-helix, respectively—that extend from within the C-terminal regions of the proteins (**Fig. 4a**). Although the active site is not clearly defined in the Csy4–RNA complex, alignment with the Cse3–deoxyG RNA structure supports the proposed interaction of the Ser148 side chain with the 2′-hydroxyl nucleophile and suggests stabilization of the leaving group by His29. This analysis also indicates that in Csy4, stabilization of the cleavage transition-state intermediate, mediated in Cse3 by Tyr23, is probably a function of the main chain amide group of Gln149 (**Fig. 4b**). Alignment of the Cse3 and Cas6 structures suggests that the catalytic triad proposed for Cas6 may indeed function to promote RNA cleavage but with a disposition and distinct function of catalytic residues from that reported here for Cse3 (**Supplementary Fig. 6**). A final point of similarity between Cse3 and Csy4—and, on the basis of homology, Cas6—is the extended β-hairpin at the base of the duplex, which forms specific contacts to the RNA (**Figs. 2** and **4**). We find it intriguing that this element presents the Arg158 side chain that we suggest forms part of the processing active site (Ser148 in Csy4); structural homology suggests a similar role for Lys195 of Cas6 (**Supplementary Fig. 6**). Because the extended hairpin is fully ordered only in the Cse3 and Csy4 RNA complexes, we propose that the final assembly of the enzyme active site is a function of specific RNA binding and thus acts as a regulatory mechanism.

Our comparison of three diverse CRISPR endonucleases alone and complexed to RNA shows that although these proteins have evolved both distinct RNA recognition modes and catalytic mechanisms, they nevertheless share important common features governing both of these steps in pre-crRNA processing. The documented evolutionary relationship between CRISPR subtypes suggests that the features described here may broadly be generalized to a large family of processing endonucleases. An important goal of future structural and functional studies will be to relate these features of the individual pre-crRNA processing complexes to subsequent steps in the CRISPR targeting mechanism.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession Codes. Protein Data Bank: Coordinates for the Cse3–product mimic RNA, Cse3–deoxyG RNA and Cse3–product complexes have been deposited under the accession codes 3QRP, 3QRQ and 3QRR respectively.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work was supported by an operating grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) to A.M.M. We are also grateful for the support of the Alberta Synchrotron Institute during the early stages of this work.

AUTHOR CONTRIBUTIONS

E.M.G. cloned, expressed and biochemically characterized Cse3 and Cse3 mutants, purified the proteins and protein–RNA complexes and grew the crystals. M.J.S. solved the structures. E.L.G. and M.M.G. cloned, expressed and purified Cse3 mutants. E.M.G., M.J.S., E.L.G. and A.M.M. designed the experiments and contributed to writing the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

© 2011 Nature America, Inc. All rights reserved.

1. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).
2. Nakata, A., Amemura, M. & Makino, K. Unusual nucleotide arrangement with repeated sequences in the *Escherichia coli* K-12 chromosome. *J. Bacteriol.* **171**, 3553–3556 (1989).
3. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
4. Brouns, S.J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
5. Carte, J., Wang, R., Li, H., Terns, R.M. & Terns, M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
6. Jansen, R., Embden, J.D., Gaastra, W. & Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).
7. Marraffini, L.A. & Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
8. Haft, D.H., Selengut, J., Mongodin, E.F. & Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1**, e60 (2005).
9. Kunin, V., Sorek, R. & Hugenholtz, P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* **8**, R61 (2007).
10. Makarova, K.S., Grishin, N., Shabalina, S., Wolf, Y. & Koonin, E. 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**, 7–33 (2006).
11. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8**, 172–182 (2007).
12. Agari, Y. *et al.* Transcription profile of *Thermus thermophilus* CRISPR systems after phage infection. *J. Mol. Biol.* **395**, 270–281 (2010).
13. Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904–912 (2009).
14. Han, D. & Krauss, G. Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* **583**, 771–776 (2009).
15. Pougach, K. *et al.* Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol. Microbiol.* **77**, 1367–1379 (2010).
16. Xue, S., Calvin, K. & Li, H. RNA recognition and cleavage by a splicing endonuclease. *Science* **312**, 906–910 (2006).
17. Calvin, K., Xue, S., Ellis, C., Mitchell, M. & Li, H. Probing the catalytic triad of an archaeal RNA splicing endonuclease. *Biochemistry* **47**, 13659–13665 (2008).
18. Ebihara, A. *et al.* Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein Sci.* **15**, 1494–1499 (2006).
19. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 (2010).
20. Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–845 (1999).
21. Handa, N. *et al.* Structural basis for recognition of the *tra* mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
22. Wang, R., Preamplume, G., Terns, M.P., Terns, R.M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**, 257–264 (2011).
23. Schmidt, B.H., Burgin, A.B., Deweese, J.W., Osheroff, N. & Berger, J.M. A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. *Nature* **465**, 641–644 (2010).
24. Guillén Schlippe, Y.V. & Hedstrom, L. A twisted base? The role of arginine in enzyme-catalyzed proton abstractions. *Arch. Biochem. Biophys.* **433**, 266–278 (2005).
25. Steitz, T.A. & Steitz, J.A. A general two-metal ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 6498–6502 (1993).

ONLINE METHODS

Cloning, expression and purification of *Thermus thermophilus* Cse3. The *T. thermophilus* Cse3 gene *TTHB192* (NC_006462.1) was PCR amplified from genomic DNA (ATCC 27634D-5) using primers containing EcoRI and BamHI sites and was then cloned into pET-30a(+) (WT, Y23F) or pACYC-duet (R24A, R157A, R158A) (Novagen). Mutagenesis was carried out by PCR and confirmed by sequencing. *E. coli* Rosetta cells were transformed with the plasmid, grown to an OD_{600nm} of ~0.8 and induced with 1 mM IPTG for 12 h at 24 °C. Cells were lysed at 4 °C for 30 min (100 mM NaCl, 20 mM Tris-HCl, pH 9.0, 1 mM 2-mercaptoethanol, 20 mM imidazole, 1 μg ml^{-1} lysozyme, 1 mM PMSF) and then sonicated. Lysate was cleared by centrifugation at 40,000*g* for 30 min, bound to Ni Sepharose 6 Fast Flow resin (GE Healthcare) and eluted with lysis buffer containing 200 mM imidazole. The resultant His_6-tagged Cse3 fusion proteins were purified by Superdex 75- and anion-exchange chromatography. Dialyzed protein was aliquoted and stored at −20 °C in a buffer containing 15% (v/v) glycerol, 100 mM NaCl, 10 mM Tris-HCl, pH 9.0, 0.5 mM 2-mercaptoethanol and 0.5 mM EDTA.

RNA preparation. RNAs purchased from IDT were designed to model a full or partial CRISPR repeat: 5′-GUA GUC CCC ACG CGU GUG GGG AUG GAC C-3′ (28 nt modeling the full repeat), 5′-GUA GUC CCC ACG CGU GUG GGdG AUG GAC C-3′ (28-nt noncleavable repeat), 5′-GUC CCC ACG CGU GUG GGdG A-3′ (22-nt noncleavable minimal repeat), 5′-GUC CCC ACG CGU GUG GGG A-3′ (22-nt cleavable minimal repeat), and 5′-GUC CCC ACG CGU GUG GGG-3′ (21-nt mimic of the 5′ cleavage product).

Gel mobility shift and cleavage assays. 5′-^{32}P-radiolabeled RNA substrate was preheated to 55 °C for 2 min, then incubated in 10-μl reaction solutions containing 0–10 μM Cse3 protein (100 mM NaCl, 10 mM Tris-HCl, pH 9.0, 100 μM EDTA, 500 μM 2-mercaptoethanol and 20 μg $μl^{-1}$ yeast tRNA) for 30 min and immediately loaded onto a 6% tris-glycine (w/v) polyacrylamide gel. For cleavage assays, reactions were quenched at various time points in an equal volume of loading dye containing 2% SDS, 7 M urea, before resolution on a 20% 29:1 (w/v), 8 M urea-sequencing PAGE. Initial cleavage rates, determined in triplicate at 22 °C, used a five-fold excess of 5′-^{32}P-radiolabeled 28-nt RNA. Gels were exposed to a phosphor screen (Molecular Dynamics), scanned with a Storm 840 Phosphorimager and analyzed using ImageQuant software (Molecular Dynamics).

Characterization of Cse3-dependent 3′ RNA cleavage product. The 28-nt repeat RNA (~1 pmol) was incubated with 30 ng Cse3 as described above for 30 min at 55 °C followed by phenol/chloroform extraction and ethanol precipitation. RNA was radiolabeled using T4 kinase (Invitrogen) and [γ-^{32}P]ATP (PerkinElmer) according to the manufacturer's instructions and analyzed as described above.

Characterization of Cse3-dependent 5′ RNA cleavage product. Oxidation-elimination reaction experiments were conducted essentially as described[26]. The 5′ radiolabeled 28-nt repeat or 21-nt repeat RNAs were incubated for 30 min at 55 °C in the presence or absence of Cse3 and then subjected to phenol-chloroform extraction and ethanol precipitation. Resuspended reactions were incubated in 60 mM borate–boric acid buffer, pH 8.6, and 25 mM sodium periodate at 0 °C for 60 min. Reactions were quenched with 10 μl of glycerol before phenol-chloroform extraction and ethanol precipitation. Resuspended reactions were incubated in 1 M lysine-HCl, pH 9.3, for 90 min at 45 °C and then subjected to phenol-chloroform extraction, ethanol precipitation and analysis on a 20% 29:1 (w/v), 8 M urea sequencing gel.

Purification of Cse3–RNA complexes. Purified Cse3 protein and RNA oligonucleotides were preheated to 55 °C before being mixed at a 1:1.2 protein:RNA ratio in 100 mM NaCl, 10 mM Tris-HCl buffer, pH 9.0, 100 μM EDTA, 500 μM 2-mercaptoethanol, and incubated for 30 min at 55 °C. The reactions were purified on a Superdex 200 column (GE) and dialyzed into storage buffer containing 100 mM NaCl, 10 mM Tris-HCl, pH 9.0, 100 μM EDTA, 100 μM 2-mercaptoethanol.

Crystallization. Crystals of Cse3 were grown at 23 °C by hanging-drop vapor diffusion, in which 1 μl of 5 mg ml^{-1} protein solution was mixed with 1 μl of precipitant. Crystallization conditions were as follows: Cse3 crystals (50 mM bicine, pH 9.0, 15–18% PEG 2000); Cse3–product mimic crystals (100 mM Tris, pH 8.0, 10–15% (w/v) PEG 8000, 100 mM KCl, 5 mM $MgSO_4$, 1 mM spermidine); Cse3–deoxyG RNA crystals (100 mM HEPES buffer, pH 7.5, 10–15% PEG 3350, 100 mM KOAc, 1 mM spermidine); Cse3–product crystals (50 mM sodium succinate, pH 6.0, 100 mM NaCl, 1 mM spermidine). Crystals were transferred to precipitant containing 20% (v/v) glycerol and frozen in liquid nitrogen for data collection.

Data collection and processing. Data were collected at 100 K at beamline 9-2 of the Stanford Synchrotron Radiation Lightsource and beamline 12.3.1 at the Advanced Light Source of the Lawrence Berkeley National Laboratory. Data were collected at a single wavelength for each crystal: 0.97946 Å for the Cse3–product mimic RNA complex, 0.97945 Å for the Cse3–deoxyG RNA complex and 1.11588 Å for the Cse3–product RNA complex. Data were processed and scaled using the HKL package[27].

Model building and refinement. The structures were solved by molecular replacement using the program PHASER[28]. For the Cse3 structure and the Cse3–product mimic RNA structure, the search model was PDB entry 1WJ9 (ref. 18). The refined model of Cse3–product mimic RNA was used as the search model for the Cse3–deoxyG RNA and the Cse3–product RNA structure. The initial molecular replacement model was improved by iterative cycles of manual model building using COOT and by refinement using REFMAC[29,30]. Refinement statistics for structures are summarized in Table 1. Final refined models occupied the following regions of a Ramachandran plot: Cse3–product mimic RNA, 99.5% favored, 0.5% allowed and 0% outliers; Cse3–deoxyG RNA, 96.0% favored, 3.5% allowed and 0.5% outliers; and Cse3–product RNA, 97% favored, 3.0% allowed and 0% outliers. Ramachandran plots were generated using RAMPAGE[31]. All structural representations were created using PyMOL (Delano Scientific).

26. Igloi, G.L. & Kossel, H. Affinity electrophoresis for monitoring terminal phosphorylation and the presence of queuosine in RNA. Application of polyacrylamide containing a covalently bound boronic acid. *Nucleic Acids Res.* **13**, 6881–6898 (1985).
27. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
28. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
29. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
30. Read, R.J. Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1373–1382 (2001).
31. Lovell, S.C. *et al.* Structure validation by Cα geometry: φ,Ψ and Cβ deviation. *Proteins* **50**, 437–450 (2003).

© 2011 Nature America, Inc. All rights reserved.