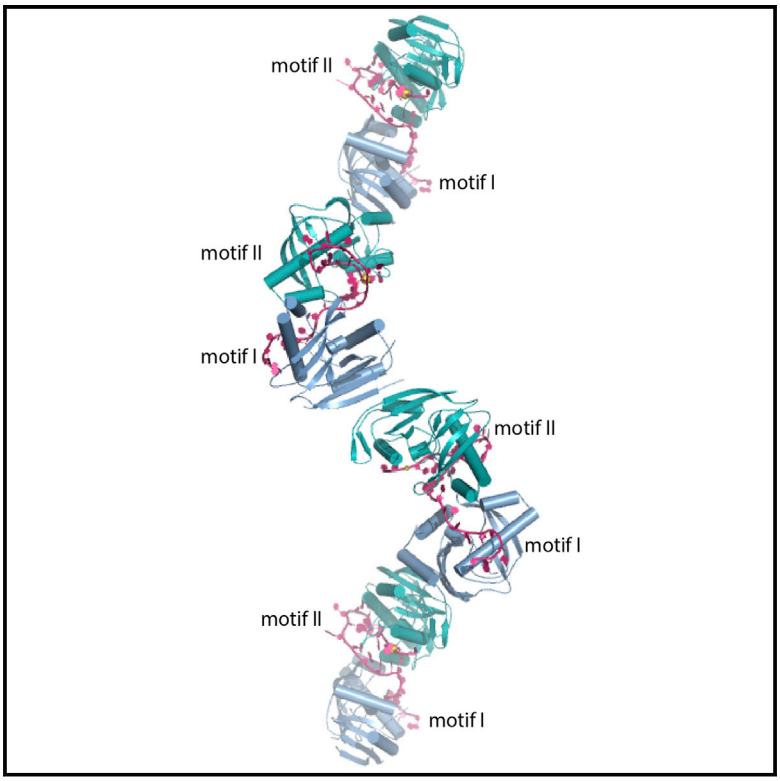
Article

Structure

A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6

Graphical Abstract



Authors

Yaming Shao, Hagen Richter, Shengfang Sun, Kundan Sharma, Henning Urlaub, Lennart Randau, Hong Li

Correspondence

hong.li@fsu.edu

In Brief

M. maripaludis Cas6 processes a long repeat RNA by binding both to the cleavage site and a distal site. Shao et al. show that the dual binding mode alleviates a non-productive secondary RNA structure for efficient processing and may represent a general method for Cas6 to process non-stem-loop repeat RNA.

Highlights

- Cas6 from *Methanococcus maripaludis* binds to two sites of the long CRISPR repeat
- *M. maripaludis* Cas6 recognizes a 2 base pair stem and an AAYAA loop
- Cas6 supplies a tryrosine residue as a nucleobase mimic that bases with an adenine
- The productive dual binding by Cas6 competes with a non-productive RNA structure

Accession Numbers

4Z7K
4Z7L

Shao et al., 2016, Structure *24*, 547–554
April 5, 2016
http://dx.doi.org/10.1016/j.str.2016.02.009

Structure
Article

A Non-Stem-Loop CRISPR RNA Is Processed by Dual Binding Cas6

Yaming Shao,[1] Hagen Richter,[3] Shengfang Sun,[1] Kundan Sharma,[5] Henning Urlaub,[5] Lennart Randau,[3,4] and Hong Li[1,2,*]
[1]Institute of Molecular Biophysics, Florida State University, Tallahassee, FL 32306, USA
[2]Department of Chemistry and Biochemistry, Florida State University, Tallahassee, FL 32306, USA
[3]Max-Planck-Institute for Terrestrial Microbiology, 35043 Marburg, Germany
[4]LOEWE Center for Synthetic Microbiology (Synmikro), 35043 Marburg, Germany
[5]Department of Cellular Biochemistry, Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, 37077 Göttingen, Germany
*Correspondence: hong.li@fsu.edu
http://dx.doi.org/10.1016/j.str.2016.02.009

SUMMARY

A subclass of recently discovered CRISPR repeat RNA in bacteria contains minimally recognizable structural features that facilitate an unknown mechanism of recognition and processing by the Cas6 family of endoribonucleases. Cocrystal structures of Cas6 from *Methanococcus maripaludis* (MmCas6b) bound with its repeat RNA revealed a dual site binding structure and a cleavage site conformation poised for phosphodiester bond breakage. Two non-interacting MmCas6b bind to two separate AAYAA motifs within the same repeat, one distal and one adjacent to the cleavage site. This bound structure potentially competes with a stable but non-productive RNA structure. At the cleavage site, MmCas6b supplies a base pair mimic to stabilize a short 2 base pair stem immediately upstream of the scissile phosphate. Complementary biochemical analyses support the dual-AAYAA binding model and a critical role of the protein-RNA base pair mimic. Our results reveal a previously unknown method of processing non-stem-loop CRISPR RNA by Cas6.

INTRODUCTION

RNAs possess an inherent structural flexibility that enables them to serve in a broad range of biological activities. Endoribonucleases are protein enzymes that cleave RNA, often at a specific phosphodiester bond, to yield functional forms of RNA. The principle of how endoribonucleases interact with and cleave intrinsically flexible RNA substrates remains only partially understood owing in part to the lack of 3D structures of endoribonucleases bound with their RNA substrates.

Cas6 superfamily endoribonucleases function in a small RNA-guided immunity pathway conferred by the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and associated proteins (Cas proteins) in prokaryotes (Hochstrasser and Doudna, 2015; Li, 2015). The CRISPR loci are segments of identical repeat sequences (repeats) interspersed by unique spacer sequences (spacers) (Barrangou and Marraffini, 2014; Gasiunas et al., 2014; Reeks et al., 2013; van der Oost et al., 2014; Wiedenheft et al., 2012). The repeat-spacer array encodes RNA substrates for Cas6. Cas6 recognizes features exclusively within the repeat and cleaves, in almost all cases, eight nucleotides away from the 3′ end of the repeat, resulting in the spacer RNA flanked by repeat sequences at both ends (CRISPR RNA or crRNA) (Hochstrasser and Doudna, 2015; Li, 2015). Processed crRNAs function as guides to the interference ribonucleoprotein particles (crRNPs) in identifying foreign nucleic acids (e.g., viral genomes or conjugative plasmids) to be degraded by Cas proteins (Barrangou and Marraffini, 2014; Gasiunas et al., 2014; Reeks et al., 2013; van der Oost et al., 2014; Wiedenheft et al., 2012).

The letter "P" in the CRISPR acronym reflects the "palindromic" feature detected in slightly more than 50% of the known repeat sequences (Kunin et al., 2007; Lange et al., 2013). The palindrome in CRISPR repeats results in stem-loop structures in their encoded RNA that are key features recognized by a number of known Cas6 endoribonucleases (Hochstrasser and Doudna, 2015; Li, 2015). In these cases, the Cas6 endoribonucleases depend on the hallmark stem-loop feature and cleave at the base of the stem. The 3′ cleavage product is released after cleavage but the 5′ cleavage product, which includes the stem loop, remains tightly bound to Cas6, which ultimately becomes part of the downstream interference crRNPs (Hochstrasser and Doudna, 2015; Li, 2015). Thus, the presence of the palindromic feature in repeats is indicative of the stem loop of the bound RNA substrate and therefore their mode of processing by Cas6.

Significantly, nearly half of the known repeats exhibit weak or no palindromic features, making prediction of the RNA processing mechanism difficult (Kunin et al., 2007; Lange et al., 2013). Most of these repeat RNAs are also recognized and cleaved by Cas6 from the same 8 nt distance to the 3′ end of the repeat. The Cas6 proteins associated with these repeat RNAs are predicted to have similar 3D structures as those associated with stem-loop RNAs. Biochemical studies showed that these Cas6 proteins tend to have weaker binding affinities for their 5′ cleavage product and therefore do not associate with the downstream interference crRNPs (Hochstrasser and Doudna, 2015; Li, 2015). In contrast to the well-accepted processing model for the stem-loop repeat RNA, our understanding of how non-stem-loop repeat RNAs are recognized by Cas6 is limited due to the large variations in RNA structures and currently inconsistent RNA

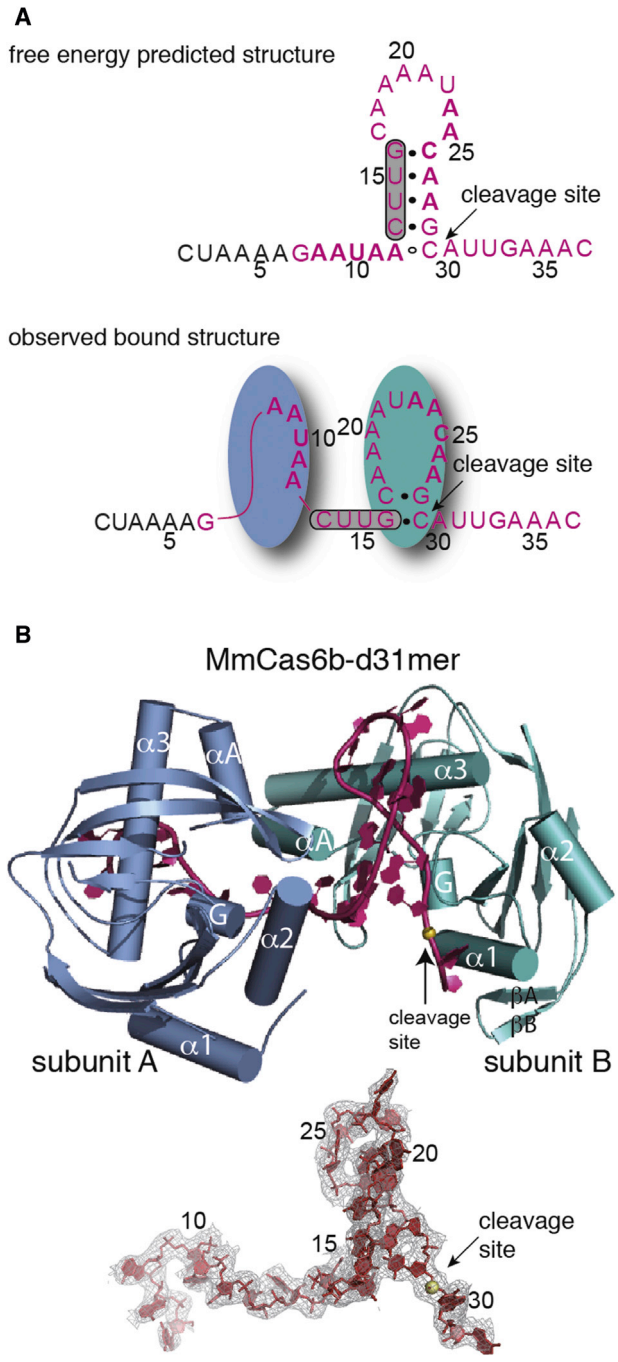


Figure 1. Overview of the MmCas6b-RNA Complex Structure

(A) The mFold-predicted (upper) and the actual bound (lower) secondary structure of the bound repeat RNA. The entire 37 nucleotides are listed and those colored in red are used in cocrystallization (d31mer). The four nucleotides enclosed in a gray frame would be base-paired in the predicted secondary structure but are otherwise unpaired in the actual bound structure. The cleavage site is identified by an arrow.

(B) Crystal structure of the MmCas6b-d31mer complex where the two subunits of protein are colored in teal and light blue and RNA in red, respectively. The bound RNA structure with the protein subunits removed is shown in the same orientation below, along with the $2F_o - 2F_c$ omit map computed using the phases from the refined protein molecules only. The orange sphere represents the scissile phosphate. Key secondary structure elements of MmCas6b are labeled where G stands for glycine-rich (G-rich loop).

binding results. Two examples of Cas6 binding to non-stem-loop repeat RNA have been studied. A short *Sulfolobus solfataricus* (Ss) repeat RNA is processed through stabilization of a 3 base pair stem loop by SsCas6 (Shao and Li, 2013), suggesting an intrinsic ability of Cas6 in reinforcing a stem loop near the cleavage site. However, *Pyrococcus furiosus* (Pf) repeat RNA is specifically recognized by its 5′ end as well as its cleavage site, leading to the wrap-around model in which the repeat RNA anchors its 5′ end on one side and the cleavage site on the other of a single Cas6 subunit (Wang et al., 2011). A thorough understanding of CRISPR RNA processing is not possible without a comprehensive knowledge associated with a wide range of repeat RNAs.

Formation of the stem immediately upstream of the cleavage site facilitates phosphodiester bond breakage catalyzed by Cas6. The Cas6 endoribonucleases known so far employ an RNase A-like cleavage mechanism (Cuchillo et al., 2011) in which a general base extracts the proton from the 2′-hydroxyl group immediately upstream of the scissile phosphate and a general acid donates a proton to the 5′ leaving oxygen associated with the scissile phosphate (Hochstrasser and Doudna, 2015; Li, 2015). A positively charged residue near the scissile phosphate stabilizes the developing negative charge of the transition state. The S_N2 reaction requires an alignment of the 2′-hydroxyl oxygen, the scissile phosphate, and the 5′ leaving oxygen of the RNA (inline conformation), which is incompatible with the A-form RNA structure. On the other hand, the base of a stem-loop RNA is ideal for inline conformation formation due to splaying of its downstream nucleotide. It is thus predicted that the non-stem-loop RNA substrates for Cas6 may also form analogous stem loops at the expense of Cas6 binding (Li, 2015; Shao and Li, 2013). However, experimental observations of non-stem-loop RNA bound with Cas6 remain scarce, raising the question of how Cas6 achieves specific recognition of the non-stem-loop repeat RNAs.

Methanococcus maripaludis C5 (Mm) contains a type I-B Cas6, MmCas6b (MmarC5_0767), responsible for processing crRNA (Richter et al., 2012, 2013). The RNA substrate of MmCas6b, the NC_009135 repeat, contains 37 nucleotides with a calculated minimum free energy stem-loop structure (13–28) one nucleotide upstream of the cleavage site at position 29 (Figure 1). To understand the biochemical mechanism of Cas6b, we determined two cocrystal structures of MmCas6b bound with its RNA substrate analogs. The complex structure with a non-cleavable (2′-deoxy modification at position 29), near full-length RNA substrate (d31mer) at 3.0 Å resolution reveals two unexpected dual binding motifs optimal for cleavage site recognition. The complex structure with a minimal recognition motif RNA (14mer) at 3.5 Å resolution reveals negligible structural changes in the active site despite the lower stability of the complex, suggesting a role of the distal minor motif in aiding substrate recognition rather than catalysis. Dual recognition by Cas6 reconciles the previous inconsistent binding data among Cas6 proteins and provides a strategy for Cas6 endoribonucleases to process CRISPR repeats lacking stable stem-loop structures.

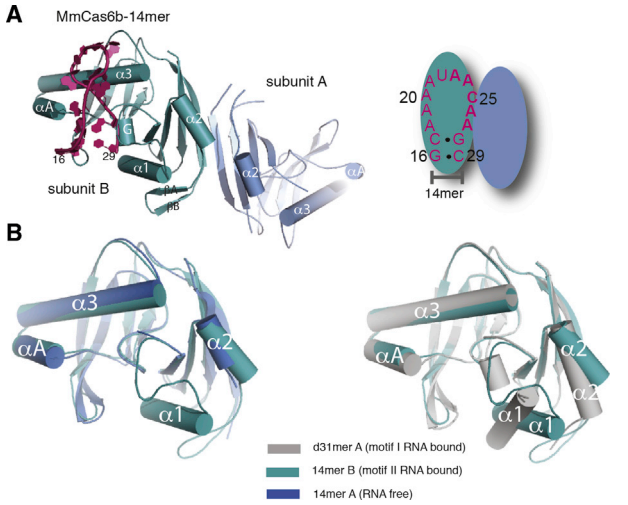


Figure 2. Structure of the MmCas6b-14mer RNA Complex and Structural Comparison

(A) The crystal structure of the MmCas6b-14mer complex where the two subunits of protein are colored in teal and light blue and RNA in red, respectively. Secondary structure of the 14mer and the schematic MmCas6b dimer are shown on the right.

(B) Structural comparison among the four known MmCas6b structures. Note that the entire α1 helix and part of the α2 helix of the RNA-free MmCas6b (blue) are missing, and the same helices have different conformations when bound with the two different motifs of RNA (gray and teal).

RESULTS

MmCas6b Recognizes Dual RNA Motifs

Although free energy minimization predicts a stem-loop structure for the NC_009135 repeat RNA in isolation (Figure 1A), it forms two recognition motifs (I and II) when bound to MmCas6b (Figures 1A and 1B). One of the two bound MmCas6b (subunit A) binds at the minor recognition site comprising nucleotides 7–12 (motif I) and the other (subunit B) binds at the major recognition site comprising nucleotides 16–30 including the cleavage site between nucleotides 29 and 30 (motif II) (Figures 1A and 1B). The two motifs are linked by three extended nucleotides 13–15 (~40 Å), which results in few contacts between the two bound MmCas6b subunits with only 235 Å^2 buried solvent-accessible surface area.

To assess the impact of the distal motif I on the structure of the MmCas6b-RNA complex, we also analyzed a crystal structure of MmCas6b bound with a 14mer RNA containing only motif II nucleotides excluding the cleavage site (nucleotides 16–29) (Figure 2). MmCas6b binds to one 14mer molecule as an isologous homodimer where one subunit contains no bound RNA (Figure 2A). The homodimer buries a 598 Å^2 solvent-accessible area averaged among the three complexes in the asymmetric unit, suggesting a weak but discernible dimerization interaction. The moderate interface of the 14mer-bound dimer observed in the crystal structure likely explains the small proportion of dimer revealed in a gel filtration experiment with the RNA-free MmCas6b (Figure S1). The 14mer-bound dimer is unrelated to the two subunits bound with the d31mer RNA that interact only with the d31mer and not with each other. To compare MmCas6b structures when free and bound with different RNAs, we superimposed the structures of the four available MmCas6b subunits (two from each structure). The 14mer-bound subunit is nearly identical to that bound to the motif II in the d31mer complex (Figure S2). Thus, motif I has no significant impact on the structure of the bound motif II. The RNA-free subunit is also in general similar to the motif-II-bound subunit except that its α1 and part of α2 helices are disordered (Figure 2B). Interestingly, the 14mer- or the motif-II-bound subunit differs from the motif-I-bound subunit in α1 and α2 orientations (Figure 2B), suggesting that the short stem in motif II induces the conformational changes in α1 and α2 that are likely required for catalysis.

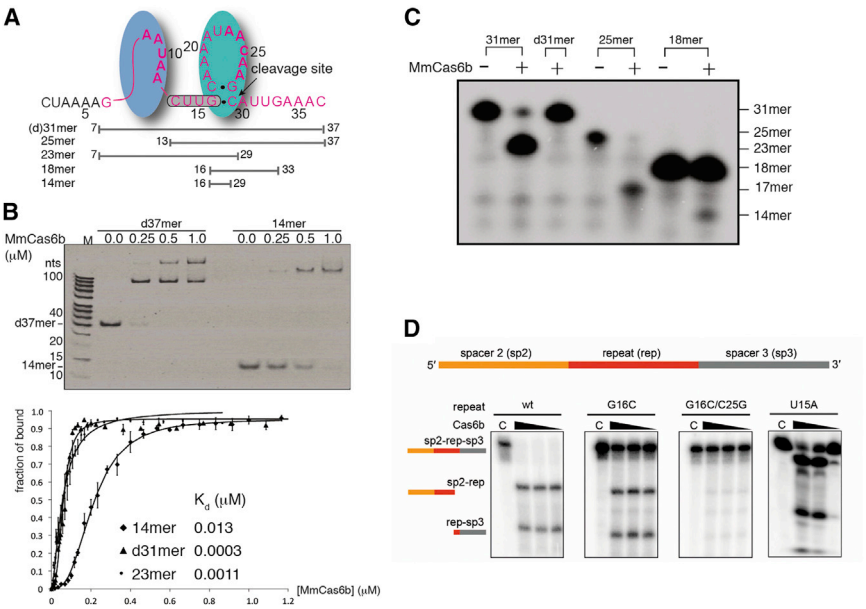


Figure 3. RNA Binding and Cleavage Assay Results

(A) Schematics of different RNA oligomers used for cleavage and binding studies.

(B) Gel mobility shift assay (upper) and binding isotherm plot obtained from filter-binding assays (lower) for the 14mer, d37mer, and 23mer (filter-binding only). The MmCas6b-RNA complexes were resolved on a 13% native polyacrylamide gel. The concentrations used for the gel shift experiments were 0.25 μM, 0.5 μM, and 1.0 μM for MmCas6b, 0.5 μM for the d37mer, and 1.0 μM for the 14mer. "M" indicates nucleotide molecular weight markers (low molecular weight marker from Affymetrix). The binding parameters, K_d, and Hill coefficients were obtained from fitting the isotherms to the Hill equation. SDs of data points are the results from three replicas of binding measurements.

(C) RNA cleavage assay results. 5′-labeled RNA substrates are indicated on the top. For cleavage reactions, 1 μM MmCas6b was mixed with the RNA and incubated for 1 hr.

(D) RNA cleavage results of a wild-type (wt), G16C, G16C/C25C, and U15A mutant RNA substrate (schematics shown). The RNA is internally labeled.

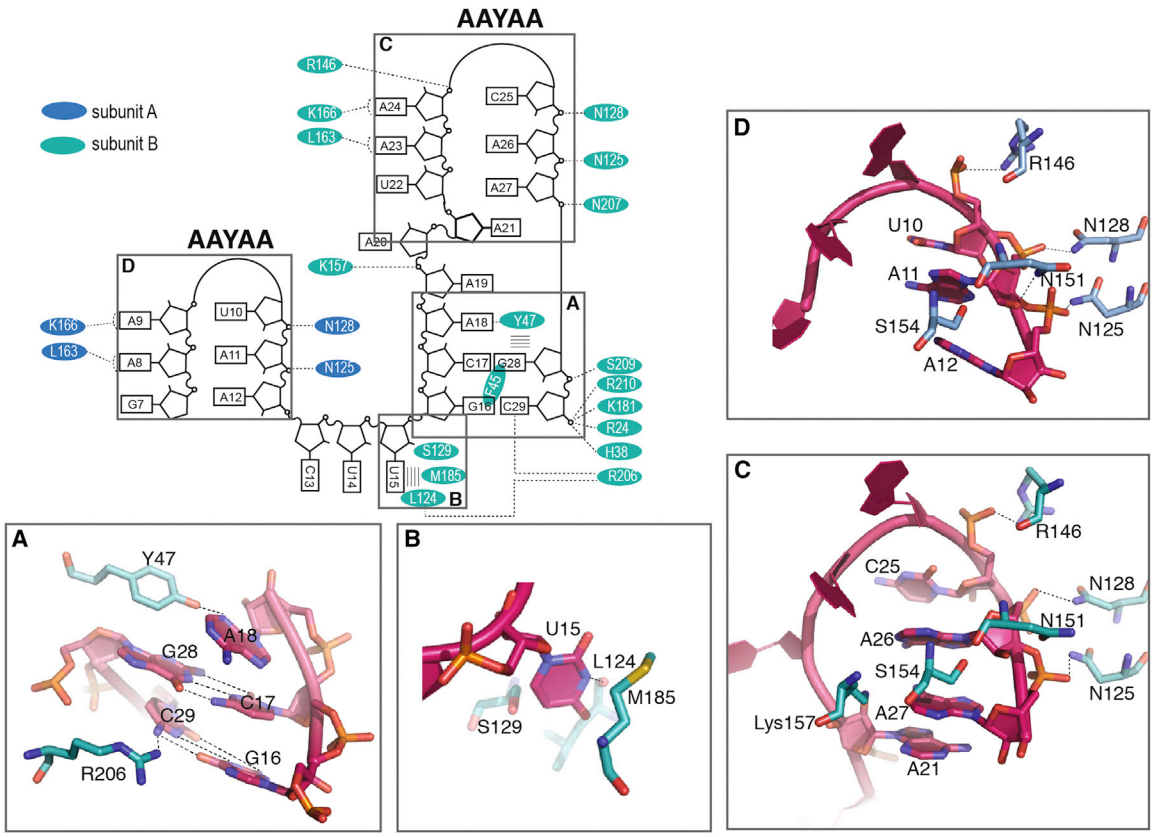


Figure 4. Details of MmCas6b-RNA Interactions

Four regions of the protein-RNA interactions are highlighted for detailed descriptions (A–D). Note that motifs I and II share similar interactions involving the AAYAA (Y = U, C) sequence.

(A) Interactions around the short stem region where Tyr47 forms a base pair mimic with A18 and Arg206 forms hydrogen bonds with the G16-C29 pair. Dotted lines indicate close contacts.

(B) Stabilization of U15 by Met185, L124, and Ser129.

(C) Interactions of the AAYAA sequence in motif II with an asparagine cluster (Asn125, Asn128, Asn151) of subunit B.

(D) Interactions of the AAYAA sequence in motif I with the same asparagine cluster of subunit A as that of subunit B.

To assess how protein:RNA stoichiometry depends on the two motifs in solution, we carried out electrophoretic mobility shift assays of MmCas6b with the repeat RNA in the presence (d37mer) and absence (14mer) of motif I. The binding of the d37mer RNA substrate showed a single retarded species at low MmCas6b concentrations that was further retarded as additional MmCas6b was added (Figures 3A and 3B). A similar stepwise shift of the wild-type repeat RNA was previously observed (Richter et al., 2013). These observations suggest a sequential binding of motif II and motif I by MmCas6b. In contrast, binding of the 14mer RNA resulted in a single retarded species for all MmCas6b concentrations (Figures 3A and 3B), consistent with the presence of only a single binding site in this RNA. Interestingly, the retarded species of the 14mer lacking motif I migrated slower than that of the lower shifted species of the d37mer. It is plausible that the additional retardation represents the cooperative dimerization of MmCas6b as observed for the MmCas6b-14mer complex.

Measurement of binding constants revealed notable differences between MmCas6b binding to the dual- and the single-motif RNA oligomers. Fitting filter-binding isotherms to a Hill equation showed ~40× increase in K_d when binding the 14mer than the d31mer (Figures 3A and 3B), suggesting an impact of motif I on MmCas6b binding thermodynamics. In contrast, removal of the last eight nucleotides from the 3′ end (23mer) had only a minor impact on the binding constant (Figure 3B). The non-hyperbolic binding isotherms for both RNA oligomers is consistent with the observed 2:1 protein:RNA stoichiometry ratio in both the 31mer and the 14mer complexes. Although a truncated RNA lacking the AAUAA of motif I was cleaved by MmCas6b (25mer), further truncation of two nucleotides toward motif II (18mer) severely affected cleavage by MmCas6b (Figure 3C). Taken together, these results suggest that motif I greatly facilitates but is not required for cleavage at motif II by MmCas6b.

RNA Recognition by MmCas6b

Motif II of the repeat RNA is the major site recognized by MmCas6b. It comprises a 2 base pair stem (G16-C29/C17-G28) and an adenine-rich loop (A18-A27). The major groove of the short stem faces the interior of MmCas6b, and together with the major groove edge of U15, forms a polar interface with the protein at which Arg206 of the G loop interacts directly

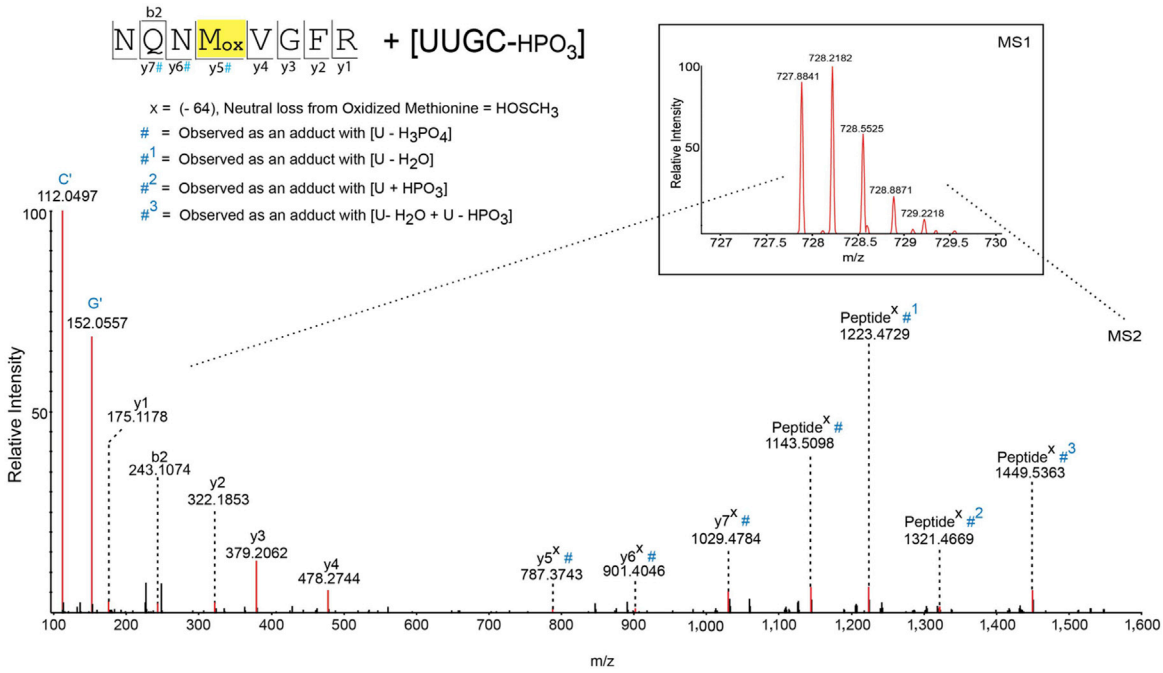


Figure 5. MS and MS/MS Spectra of Cas6b Peptide NQNMoxVGFR Crosslinked to RNA Oligonucleotide UUGC-PO3

The crosslinked peptide with its cross-linked RNA oligonucleotide was identified and sequenced with the help of RNPxl software applied on MS raw data (Kramer et al., 2014). The crosslinked peptide sequence and its corresponding y- and b-type fragment ions are indicated at the top (left), with the crosslinked amino acid highlighted in yellow. The MS1 spectrum for crosslinked species is shown at the top (right). y- and b-type fragment ion peaks with their corresponding m/z values are marked in the MS/MS spectrum. y-type ions with a mass shift of #, #1, #2, and #3 correspond to the crosslinked nucleotides U-H_3PO_4, U-H_2O, U + HPO_3 and U-H_2O + U-H_3PO_4, respectively. These y-type ions enable the unambiguous assignment of methionine in its oxidized state (y5#) as crosslinked amino acid. The crosslinked nucleotide is one of the uracil residues as the bases of nucleotides G and C are present as marker ions in the lower m/z regime of the spectrum. Of note, UV-induced crosslinking of proteins to RNA takes place between the side chains of the amino acid and the bases of the nucleotides. U, uracil; G′, base of G with marker ion of 152.05 m/z, C′, base of C with marker ion of 112.05 m/z (Kramer et al., 2014).

with the N4 atom of C29 and, through the carbonyl oxygen of Leu124, with U15 (Figure 4). Other residues do not form base-specific hydrogen bonds but pack around the two base pairs. These include Gln183 of loop β6-β7, Asn39, His40, and Phe45 of loop α1-β2, and Asn207 and Ser209 of the G loop (Figure 4). The adenine-rich loop wraps mostly around the α3 helix and partially around the αA helix and contacts several asparagine residues via its curved phosphosugar backbone (Figures 1 and 4). Only A26 and A27 of the loop are involved in base interactions with the main chain atoms of the MmCas6b residues (Figure 4). Motif I comprises a similar adenine-rich loop and its A_8AUAA_{12} loop is nearly identical to $A_{23}ACAA_{27}$. Consequently, the two motifs share the AAYAA (Y=C, U) sequence and use the same surface of the subunit A to interact with the adenine-rich loop (Figure 4).

The most unusual feature of MmCas6b-RNA interaction involves a protein-RNA base pair mimic between Tyr47 and A18. The aromatic ring of Tyr47 inserts into the RNA stem and pairs with the minor groove edge of A18, forming an RNA base pair mimic (Figure 4). The Tyr47-A18 base pair mimicry is enabled by unstacking and splaying of the A-rich loop that gives away space for Tyr47 to stack on top of the 2 base pair stem. In addition, A19 stacks on the Tyr47-A18 pair while its base interacts with the major groove edge of A21. This mode of MmCas6b-RNA interaction creates a hybrid stem structure upstream of the cleavage site.

Consistent with the observed structural features, mutation of G16 to C greatly impaired the cleavage, and double mutation G16C/C25G abolished the cleavage activity (Figure 3D). Furthermore, Tyr47Ala mutation greatly reduced cleavage activity (Richter et al., 2012). In contrast, mutation U15A did not appreciably affect the cleavage activity (Figure 3D). To further confirm the spatial positioning of U15, we performed protein-RNA UV crosslinking followed by mass spectrometry analysis. The base of U15 was indeed observed to crosslink with Met185 after UV irradiation of the Cas6b-repeat RNA complex in solution (Figure 5), providing convincing evidence for the observed binding mode of U15 (Figure 4B).

The Active Site Structure of MmCas6b

Phosphodiester bond cleavage by MmCas6b takes place between C29 and A30. The nucleotide A30 splays away from the helical track of the stem toward the α1 helix (Figure 6A). The rotation of A30 around the scissile phosphate bond is critical to formation of the conformation favorable for catalysis. Due to the use of 2′-deoxy modification of C29 in crystallization, the 2′-hydroxyl oxygen is absent from our structure. However, if it were present, the three reactive atoms (2′-hydroxyl oxygen, phosphate, and 5′-oxygen) would form the inline conformation (Figure 6A). Two histidine residues (His38 and His40) and four positively charged residues (Arg24, Lys29, Lys181, and Arg210) surround the cleavage site (Figure 6A). His40 is in a

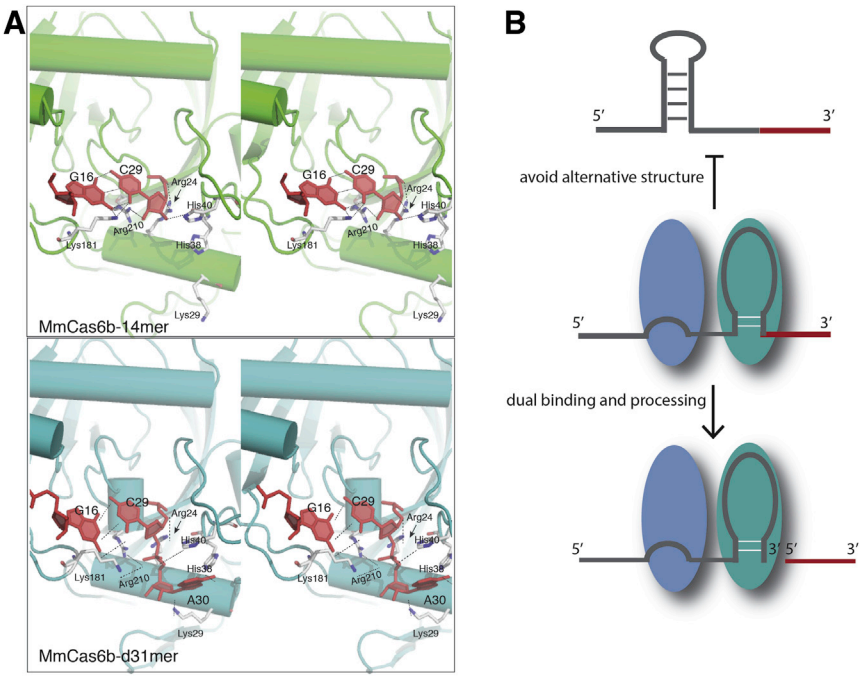


Figure 6. Structure of the MmCas6b Active Site and the Model of Dual Binding Processing by Cas6

(A) The active site structures for the 14mer- (upper) and d31mer-bound (lower) MmCas6b in cross-eyed views. Red sticks indicate RNA nucleotides G16, C29, and A30. Interactions among and between the RNA and MmCas6b are shown by dashed lines. Note that the 2′-hydroxyl oxygen is absent due to the 2′-deoxy modification of the d31mer RNA and that the 5′ leaving oxygen, the scissile phosphate, and the 2′-hydroxyl oxygen (if present) would form the inline conformation.

(B) The model of dual binding processing. In addition to binding of one Cas6 to the cleavage site, capturing the free 5′ end by a second Cas6 prevents alternative but inhibitory secondary structures often found in repeat RNA. The Cas6 subunit bound at the cleavage site stabilizes a short stem loop, allowing the inline conformation and cleavage of the 5′ handle (red).

position to deprotonate the 2′-hydoxyl group and Arg24 or Arg210 is in a position to donate proton to the leaving 5′-oxygen. Lys181 or Arg210 probably stabilizes the negative charge of the transition state. The imidazole ring of His38 interacts with the splayed A30 nucleotide (Figure 6A). Consistently, mutations of His40 and His38 both reduced the cleavage activity, and their double mutation abolished cleavage (Richter et al., 2012). Interestingly, mutation of Arg24 or Lys29 to alanine did not reduce RNA cleavage activity (Richter et al., 2012), suggesting the redundant roles of these residues in RNA cleavage. Alternatively, MmCas6b relies primarily on the general base His40 and less so on a general acid for protonating the leaving oxygen.

DISCUSSION

We determined the X-ray crystal structure of a Cas6 crRNA processing endonuclease (MmCas6b) bound with its RNA substrate analogs. MmCas6b cleaves within the repeat region of the precursor crRNA 8 nt from its 3′ end. The structure of MmCas6b-RNA complex revealed a surprising secondary structure of the repeat RNA that is stabilized by a base pair mimic and a second MmCas6b subunit bound to a minor recognition site. This secondary structure permits proper recognition of the cleavage site in spite of a competing, potentially more stable, structure. The 2 base pair stem immediately upstream of the cleavage site facilitates the inline conformation of the scissile phosphate bond required for cleavage. This mode of Cas6-RNA recognition illustrates the power of protein enzymes in shaping structurally flexible RNA for catalysis.

MmCas6b is a member of Cas6 RNA processing endoribonucleases found in bacteria and archaea whose RNA substrates have a wide range of structural features. While recognition and processing of stable stem-loop RNA are well understood due to several high-resolution structures of Cas6 bound with their respective stem-loop RNA substrates (Hochstrasser and Doudna, 2015; Li, 2015), these processes are not clear for repeat RNA lacking stable stem loops. One method of processing is for Cas6 to impose a stem-loop structure of the RNA that is otherwise unstable in isolation. *Sulfolobus solfataricus* Cas6 bound with its short repeat RNA (24mer) illustrates this principle (Shao and Li, 2013). However, the extent to which Cas6 stabilizes a short stem-loop structure is limited for long repeat RNA containing potentially competing secondary structures. MmCas6b structures reported here provide a second method of recognition in which the free 5′ end of the repeat is stabilized by one Cas6 subunit so the cleavage site can be shaped by another Cas6 subunit (Figure 6B). Although in this case the 5′ distal recognition was shown to be non-essential to cleavage, other Cas6 may have evolved to depend more critically on sites beyond the cleavage site to ensure a catalytically competent conformation at the cleavage site. The added binding site(s) for Cas6 may have an evolutionary advantage in mesophilic organisms where it is difficult to dissolve competing structures in long pre-crRNA transcripts. *Pyrococcus furiosus* (Pf) Cas6 may also adopt this method of recognition with a stringent requirement for the presence of the 5′ terminus of the repeat RNA (Carte et al., 2008). Consistently, crystal structures of PfCas6 and its close homolog *Pyrococcus horikoshii* (Ph) Cas6 reveal their specific interactions with the 5′ terminus of the repeat RNA (Wang et al., 2011, 2012). The dual binding model allows Cas6 to overcome the challenge presented in the repeat RNA containing competing and non-productive secondary structures and may thus have general applicability in more Cas6 processing systems.

EXPERIMENTAL PROCEDURES

Protein Purification and Crystallization

The *M. maripaludis cas6* gene, MmarC5_0767, was cloned into the vector pET20b with a C-terminal 6× his affinity tag. The MmCas6b protein was overexpressed in *Escherichia coli* Rossetta strain and purified by a similar method

Table 1. Crystallographic Data Collection and Refinement Statistics[a]

	MmCas6b-d31mer	MmCas6b-14mer
Data Collection Statistics		
Space group	$P3_1$	$P6_4$
a	91.983	180.57
b	91.983	180.57
c	96.637	119.022
α	90	90
β	90	90
γ	120	120
Resolution range (Å)	46.06–3.00 (3.05–3.00)	50–3.50 (3.56–3.50)
No. of observed unique reflections	15,603	27,918
Redundancy	9.1 (2.6)	11.1 (10.1)
Completeness (%)	93.0 (58.7)	100 (100)
I/σ(I)	23.78 (1.64)	7.7 (2.25)
R_{sym}(%)	16.0 (70.4)	29.8 (84.8)
$CC_{1/2}$	0.993 (0.716)	0.994 (0.752)
Refinement Statistics		
Resolution range (Å)	50–3.0 (3.05–3.0)	50–3.50 (50–3.50)
R_{work}(%) (with twinning operation)	22.4 (22.6)	24.3
R_{free}(%) (with twinning operation)	27.8 (25.4)	29.0
Model information		
No. of protein/RNA complexes	2/1	6/3
No. of protein/RNA atoms	3,392/531	9,804/897
Root-mean-square deviations		
Bond length (Å)	0.010	0.012
Bond angle (°)	1.432	1.536
Ramachandran plot of protein residues		
Preferred regions (%)	96.0	94.0
Allowed regions (%)	4.0	6.0
Disallowed region (%)	0	1.0

[a]Values in parentheses are for the last resolution shell.

as that previously described (Richter et al., 2012). Briefly, the log phase *E. coli* cells were induced by adding isopropyl β-D-1-thiogalactopyranoside (IPTG) to 0.5 mM and harvesting following overnight growth at 18°C. The cell pellets were re-suspended in a lysis buffer (10 mM Tris-HCl [pH 7.5], 500 mM NaCl, 10% glycerol, and 0.5 mM DTT) and disrupted by sonication. The supernatant was loaded onto an Ni^{2+}-NTA affinity column and eluted by the same buffer supplemented with 350 mM imidazole. The elutant was further purified by a Superdex 200 gel filtration column equilibrated with 500 mM NaCl, 20 mM Tris-Cl (pH 7.5), 5% glycerol. The peak fractions corresponding to the monomer species (Figure S1) were pooled and concentrated to 50 mg/ml and stored in −80°C until crystallization.

The RNA oligomers were ordered from Integrated DNA Technologies. MmCas6b and the RNA were combined in a 1:1.2 ratio, and the complexes were crystallized at 30°C using the hanging drop vapor diffusion method by mixing equal volumes of well solution and protein-RNA complexes. MmCas6b with the 31mer RNA substrate analog was crystallized in 100 mM KCl, 10 mM $MgCl_2$, 50 mM sodium cacodylate (pH 6.5), 12% PEG4000. Crystals were cryo-protected in a well solution supplemented by 25% glycerol before being flash-frozen in liquid nitrogen. MmCas6b bound with 14mer was crystallized in 20 mM magnesium acetate, 50 mM sodium cacodylate (pH 5.5), and 1.7 M ammonium sulfate. Crystals were soaked overnight in well solution supplemented with 20 mM $HgCl_2$ followed by cryo-protection in well solution supplemented with 25% glycerol before being flash-frozen in liquid nitrogen.

Data Collection and Structure Determination

Single wavelength anomalous X-ray diffraction (SAD) data of the MmCas6b-14mer complex were collected at the mercury absorption edge and processed using HKL2000 (Otwinowski and Minor, 1997), which allowed determination of experimental phases with Phenix.autosol (Adams et al., 2010). Most of the protein residues and the 14mer RNA could be unambiguously built from the solvent flattened density map, and the structure was refined with Phenix.refine (Adams et al., 2010). The crystal structure of the MmCas6b-d31mer complex was determined by molecular replacement using MmCas6b from the 14mer complex as the probe with Phenix.Phaser (Adams et al., 2010). The structural models were improved by Coot (Emsley et al., 2010) and refined by Phenix.refine (Adams et al., 2010) (Table 1). Throughout refinement of both structures, torsion-angle non-crystallographic symmetry restraints were applied. A small fraction of twinning by merohedry was detected by using Wilson intensity ratio analysis. Three possible merohedral twin operators (−h,−k,l; h, −h−k,−l; and –k,−h,−l) were identified that resulted in 5%, 6%, and 18% twinning fractions, respectively. The molecular replacement solution was successful without any detwin operation. Similarly, refinement and model building proceeded satisfactorily without detwin. However, inclusion of the (−k,−h,−l) twin operation without alternation of the structure at any point of the refinement significantly reduced the calculated *R* factors. We thus report the final *R* factors for with (−k,−h,−l) and without twin operation in Table 1.

RNA Binding and Cleavage Assay

The RNA cleavage activity assay was carried out in a volume of 20 μl containing 1 μM MmCas6b, 16 nM 5′-^{32}P-labeled RNA, 200 mM NaCl, 20 mM Tris-HCl (pH 7.5). The cleavage reaction was terminated by adding 98% formamide at various time points. The cleavage products were separated on a 14% polyacrylamide gel and visualized by phosphorimaging.

The electrophoretic mobility shift assays were carried out with a 13% native polyacrylamide gel loaded with binding reaction mixtures containing 0.25–1.0 μM MmCas6b, 0.5 μM (d37mer) or 1 μM (14mer) RNA oligo, 200 mM KCl, 5% glycerol; 0.5 mM DTT; 0.5 mM EDTA, and 10 mM Tris-HCl (pH 8.0). The unbound RNA and retarded species were visualized by staining the gel with SYBR Gold (Life Technologies) for 30 min in a Tris acetate-EDTA buffer.

Nitrocellulose filter-binding experiments were carried out to determine the binding constant. A trace amount of (1 nM) 5′-^{32}P-labeled RNA was incubated with increasing amounts of protein in 100 μl of binding buffer containing 20 mM Tris-Cl (pH 7.5), 100 mM NaCl, and 5% glycerol for 1 hr. The mixture was loaded into the wells of a MINIFOLD II Slot-blot system (Schleicher & Schuell) and passed through nitrocellulose and a nylon filter membrane sequentially after applying a vacuum. The filter was then washed with binding buffer and visualized by phosphorimaging. The bound fraction was calculated from the intensity of the spot on the nitrocellulose membrane divided by the sum of the intensity of spots on the nitrocellulose and nylon membrane. The apparent dissociation constant was determined by fitting the binding isotherm to the Hill equation:

$$\theta = \frac{[P]^n}{K_d + [P]^n},$$

where θ is the bound fraction, n is the Hill coefficient, K_d is the binding constant, and [P] is the concentration of MmCas6b.

Photocrosslinking and High-Resolution Mass Spectrometry

UV-induced protein-RNA crosslinking was performed essentially as described (Kramer et al., 2014). Briefly, 1 nmol of the Cas6b protein was mixed with 1 nmol of the repeat RNA and incubated on ice for 30 min for complex formation. The complexes were transferred to black polypropylene microplates (Greiner Bio-One) and UV irradiated at 254 nm for 10 min on ice at a distance of 1 cm from the UV lamp. The samples were precipitated with ethanol and the

pellet was dissolved in 4 M urea and 50 mM Tris-HCl (pH 7.9), which was then diluted to 1 M with 50 mM Tris-HCl (pH 7.9). The RNA was hydrolyzed using 1 μg of Benzonase for 2 hr at 37°C in the presence of 1 mM $MgCl_2$. Following RNA digestion, the sample was digested with trypsin (Promega) with an enzyme to substrate ratio of 1:20 at 37°C overnight. To remove non-cross-linked RNA fragments, the sample was desalted using in-house prepared C18 (Dr. Maisch) columns, and the crosslinked peptides were enriched using an in-house prepared TiO_2 column (GL Sciences). The samples were then vacuum dried and dissolved in 12 μl of 5% v/v acetonitrile in water, 1% v/v formic acid in water for liquid chromatography (LC) tandem mass spectrometry (MS) analysis. 8 μl were injected onto a nano-LC system (Agilent 1100 series; Agilent Technologies) coupled with an LTQ-Orbitrap Velos instrument (Thermo Fisher Scientific). Electrospray ionization-MS was performed in data-dependent mode using a TOP10 HCD method. The precursor ions as well as fragment ions were scanned in Orbitrap, and the resulting spectra were measured with high accuracy (<5 ppm) both at the MS1 and MS2 level. Data analysis was performed as described previously (Kramer et al., 2014) and the high-scoring crosslinked peptides were manually annotated for confirmation.

ACCESSION NUMBERS

The crystal structures of the MmCas6b-RNA complexes have been deposited in the PDB database under the accession number PDB: 4Z7K for the d31mer complex and PDB: 4Z7L for the 14mer complex.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2016.02.009.

AUTHOR CONTRIBUTIONS

Y.S., S.S., and H.L. designed and performed crystal structure determination and biochemical analysis. H.R. and L.R. cloned MmCas6b and performed the mutagenesis analyses. K.S. and H.U. performed the mass spectrometry analyses. Y.S. and H.L. wrote and all other authors edited the paper.

ACKNOWLEDGMENTS

This work was supported by NIH grant R01 GM99604 to H.L. and the Deutsche Forschungsgemeinschaft (DFG FOR 1680) to L.R. and H.U.. X-Ray diffraction data were collected from the Southeast Regional Collaborative Access Team (SER-CAT) 22-ID beamline at the Advanced Photon Source, Argonne National Laboratory. Supporting institutions for APS beamlines may be found at http://necat.chem.cornell.edu/ and http://www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the United States Department of Energy Office of Science, Office of Basic Energy Sciences, under contract no. W-31-109-Eng-38.

Received: October 4, 2015
Revised: February 5, 2016
Accepted: February 17, 2016
Published: March 17, 2016

REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D. Biol. Crystallogr. *66*, 213–221.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. Mol. Cell *54*, 234–244.

Carte, J., Wang, R., Li, H., Terns, R.M., and Terns, M.P. (2008). Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. Genes Dev. *22*, 3489–3496.

Cuchillo, C.M., Nogues, M.V., and Raines, R.T. (2011). Bovine pancreatic ribonuclease: fifty years of the first enzymatic reaction mechanism. Biochemistry *50*, 7835–7841.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D. Biol. Crystallogr. *66*, 486–501.

Gasiunas, G., Sinkunas, T., and Siksnys, V. (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol. Life Sci. *71*, 449–465.

Hochstrasser, M.L., and Doudna, J.A. (2015). Cutting it close: CRISPR-associated endoribonuclease structure and function. Trends Biochem. Sci. *40*, 58–66.

Kramer, K., Sachsenberg, T., Beckmann, B.M., Qamar, S., Boon, K.L., Hentze, M.W., Kohlbacher, O., and Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. Nat. Methods *11*, 1064–1070.

Kunin, V., Sorek, R., and Hugenholtz, P. (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. *8*, R61.

Lange, S.J., Alkhnbashi, O.S., Rose, D., Will, S., and Backofen, R. (2013). CRISPRmap: an automated classification of repeat conservation in prokaryotic adaptive immune systems. Nucleic Acids Res. *41*, 8034–8044.

Li, H. (2015). Structural principles of CRISPR RNA processing. Structure *23*, 13–20.

Otwinowski, Z., and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode, *vol. 276* (Academic Press).

Reeks, J., Naismith, J.H., and White, M.F. (2013). CRISPR interference: a structural perspective. Biochem. J. *453*, 155–166.

Richter, H., Zoephel, J., Schermuly, J., Maticzka, D., Backofen, R., and Randau, L. (2012). Characterization of CRISPR RNA processing in *Clostridium thermocellum* and *Methanococcus maripaludis*. Nucleic Acids Res. *40*, 9887–9896.

Richter, H., Lange, S.J., Backofen, R., and Randau, L. (2013). Comparative analysis ofCas6b processing and CRISPR RNA stability. RNA Biol. *10*, 700–707.

Shao, Y., and Li, H. (2013). Recognition and cleavage of a nonstructured CRISPR RNA by its processing endoribonuclease Cas6. Structure *21*, 385–393.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR-Cas systems. Nat. Rev. Microbiol. *12*, 479–492.

Wang, R., Preamplume, G., Terns, M.P., Terns, R.M., and Li, H. (2011). Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. Structure *19*, 257–264.

Wang, R., Zheng, H., Preamplume, G., Shao, Y., and Li, H. (2012). The impact of CRISPR repeat sequence on structures of a Cas6 protein-RNA complex. Protein Sci. *21*, 405–417.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

