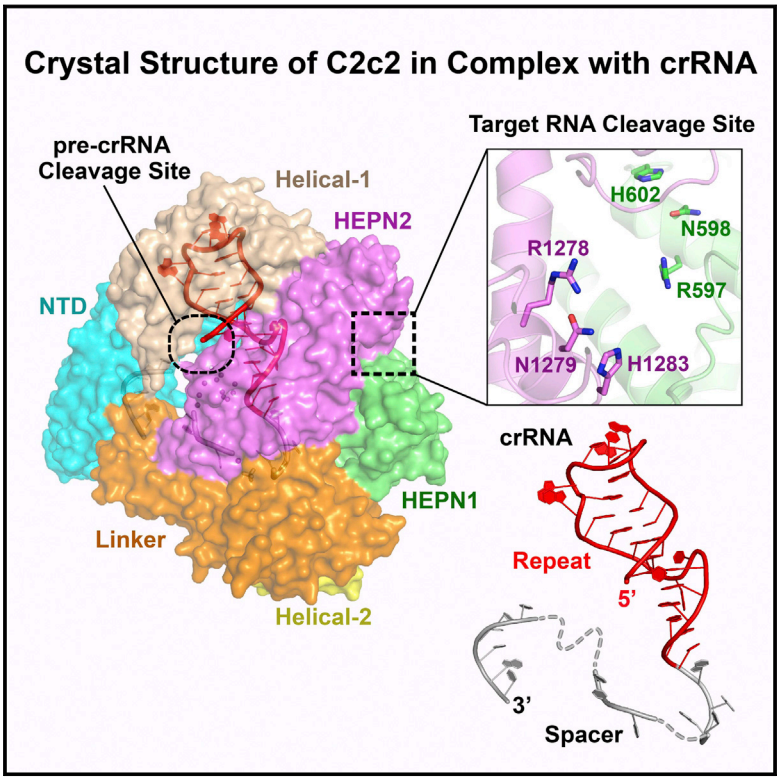
Article

Cell

Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities

Graphical Abstract



Authors

Liang Liu, Xueyan Li, Jiuyu Wang, ..., Jiazhi Li, Gang Sheng, Yanli Wang

Correspondence

ylwang@ibp.ac.cn

In Brief

The structural analysis of C2c2 reveals that its dual RNase catalytic sites locate independently in two physically distant domains, explaining its promiscuous cleavage of target RNA and specific processing of pre-CRISPR RNA.

Highlights

- Crystal structure of *Leptotrichia shahii* C2c2-crRNA binary complex and apo form
- A bulge-containing crRNA stem is essential for C2c2 RNase activities
- Helical-1 domain has the catalytic site for pre-crRNA processing
- Promiscuous RNA cleavage due to distance between HEPN catalytic site and crRNA guide

Data Resources

5WTJ
5WTK

Liu et al., 2017, Cell *168*, 121–134
January 12, 2017 © 2017 Elsevier Inc.
http://dx.doi.org/10.1016/j.cell.2016.12.031

CellPress

Cell

Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities

Liang Liu,[1] Xueyan Li,[1,2] Jiuyu Wang,[1] Min Wang,[1] Peng Chen,[1,2] Maolu Yin,[1,2] Jiazhi Li,[1,2] Gang Sheng,[1] and Yanli Wang[1,2,3,4,*]

[1]Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China

[4]Lead Contact

*Correspondence: ylwang@ibp.ac.cn

http://dx.doi.org/10.1016/j.cell.2016.12.031

SUMMARY

C2c2, the effector of type VI CRISPR-Cas systems, has two RNase activities—one for cutting its RNA target and the other for processing the CRISPR RNA (crRNA). Here, we report the structures of *Leptotrichia shahii* C2c2 in its crRNA-free and crRNA-bound states. While C2c2 has a bilobed structure reminiscent of all other Class 2 effectors, it also exhibits different structural characteristics. It contains the REC lobe with a Helical-1 domain and the NUC lobe with two HEPN domains. The two RNase catalytic pockets responsible for cleaving pre-crRNA and target RNA are independently located on Helical-1 and HEPN domains, respectively. crRNA binding induces significant conformational changes that are likely to stabilize crRNA binding and facilitate target RNA recognition. These structures provide important insights into the molecular mechanism of dual RNase activities of C2c2 and establish a framework for its future engineering as a RNA editing tool.

INTRODUCTION

Bacteria and archaea are protected against invading nucleic acids from phages and plasmids because of CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) systems, which are RNA-guided prokaryotic adaptive immune system (Barrangou et al., 2007; Barrangou and Marraffini, 2014; Marraffini, 2015; van der Oost et al., 2014). CRISPR-Cas systems are found in nearly half of all bacteria studied so far, as well as in the majority of archaea. The CRISPR-Cas system defends against foreign nucleic acids originating from viruses and plasmids in a three-step process (van der Oost et al., 2014). The first adaptation step involves the insertion of a fragment of the invading DNA into the CRISPR locus, thus generating a spacer that captures the previously invading nucleic acid. The second step involves crRNA biogenesis, whereby mature crRNAs are generated from a long pre-crRNA transcript. This pre-crRNA contains multi-repeat sequences flanking the spacers, which are processed into individual short mature crRNA consisting of a single spacer. The final step is interference. The crRNA assembles with Cas proteins to form the effector complex, in which the crRNA acts as a guide to recognize the target DNA/RNA sequence based on the complementarity and Cas proteins to act as nucleases to cleave the target, thus degrading the invading foreign nucleic acids.

Class 2 CRISPR-Cas systems, characterized by a single-component effector, comprise type II, type V, and type VI systems. Cas9 is a dual-RNA guided endonuclease of type II CRISPR-Cas systems (Makarova et al., 2015). Cas9 cleaves the target dsDNA by its RuvC and HNH nuclease domains (Gasiunas et al., 2012; Jinek et al., 2012). Type V CRISPR-Cas systems use a single RuvC domain-containing effector, including Cpf1, C2c1, and C2c3, to cleave the target DNA (Shmakov et al., 2015). Cpf1 is a single-RNA-guided DNA endonuclease of type V-A CRISPR-Cas system. C2c1 is a dual-RNA-guided DNA endonuclease of type V-B CRISPR-Cas system. Both Cpf1 and C2c1 cleave double-stranded target DNA through their RuvC domain and a putative nuclease domain, generating a product with staggered ends (Zetsche et al., 2015; Shmakov et al., 2015). All of these Class 2 CRISPR-Cas immune systems target DNA substrates.

C2c2 is a newly identified Class 2 type VI CRISPR-Cas system endonuclease. Sequence analysis revealed that C2c2 lacks an identifiable DNase catalytic site; instead, two HEPN domains containing highly conserved R-X-H motif were identified (Shmakov et al., 2015). Further functional studies revealed that C2c2 is a RNA-guided RNA-targeting CRISPR effector (Abudayyeh et al., 2016). Recent biochemical analysis found that C2c2 possesses a second RNase activity, which is responsible for crRNA maturation (East-Seletsky et al., 2016). These two RNase activities of C2c2 are mechanistically distinct from each other. While these studies have provided important biochemical insights of C2c2, the structural basis of the dual RNase activities remains unknown.

To understand how C2c2 recognizes its crRNA and target RNA and to elucidate the molecular mechanism of pre-crRNA processing, as well as crRNA-guided single stranded RNA cleavage, we determined the crystal structure of *Leptotrichia shahii* C2c2 (LshC2c2) in complex with a crRNA, together with the

structure of LshC2c2 in its RNA-free state (apo form). Our structures revealed that LshC2c2 recognizes the crRNA in a structure- and sequence-specific manner. We identified two independent catalytic sites for its two distinct RNA endonuclease activities. Furthermore, structural comparison of LshC2c2 in the crRNA-bound and in the crRNA-free states revealed that C2c2 undergoes significant conformational changes upon crRNA binding.

RESULTS

LshC2c2 Structure Reveals a Bilobed Architecture

To obtain structural insights into the architecture of C2c2, we first determined the crystal structure of full-length LshC2c2 in complex with a 56-nt crRNA at 2.65 Å resolution (Figures 1A and 1B, X-ray statistics summarized in Table S1). In the LshC2c2-crRNA binary structure, residues 103–174 within the NTD domain and residues 914–1005 within the Linker of LshC2c2 are disordered; in addition, 15 discontinuous nucleotides within the crRNA spacer are not visible, presumably due to the flexibility of this region.

The overall structure of the LshC2c2-crRNA complex, which represents the pre-target RNA-bound state, resembles a bilobed scaffold with an overall "clenched fist" shape. C2c2 consists of a crRNA recognition (REC) lobe and a nuclease (NUC) lobe. The REC lobe contains an N-terminal domain (NTD) and a Helical-1 domain. The NUC lobe contains two HEPN domains, a Linker that connects two HEPN domains, and a Helical-2 domain (Figure 1A). These two lobes are imperfectly stacked onto each other with each lobe being located at the opposing ends of the stack (Figures 1B and 1C). Interestingly, the bilobal structure of LshC2c2 is completely distinct from other Class 2 effectors, including Cas9, Cpf1, and C2c1, which also adopt a bilobed architecture (Dong et al., 2016; Gao et al., 2016; Jiang et al., 2016; Jinek et al., 2014; Nishimasu et al., 2014; Yamano et al., 2016; Liu et al., 2016). In addition, a Dali server (Holm and Rosenström, 2010) search with either the entire C2c2 protein or its individual domains failed to identify any structural homologies of LshC2c2.

The REC lobe forms the wrist-proximal part of the clenched fist structure (Figures 1B and S1A) and contains a cavity that is located between the NTD and the Helical-1 domains. The NTD domain, located at the N-terminal of C2c2, is the least conserved region of C2c2. This domain is separated into two subdomains (Figure S1C). The larger subdomain of NTD contains an ordered segment comprising seven α helices, as well as a disordered segment between residues 103–174. The smaller subdomain of NTD contains three α helices, one β-hairpin, and one β sheet. The Helical-1 domain contains six α helices, forming a V-shaped configuration (Figure S1C). The surface of the Helical-1 domain facing the NTD domain is positively charged, thus forming the crRNA-binding channel.

The C2c2 NUC lobe, which represents the top part of the clenched fist (Figure S1B), consists of two conserved catalytic domains, namely HEPN1 and HEPN2, a Helical-2 domain, and a Linker. The HEPN1 domain is further separated into two subdomains by the Helical-2 domain (shown in yellow in Figure 1). The HEPN1-I subdomain consists of four α helices and a short β-hairpin, and the HEPN1-II subdomain is composed of only three α helices. The HEPN2 domain, which is located at the C-terminal of C2c2, exhibits an entirely different structure to that of the HEPN1 domain. HEPN2 domain consists of seven α helices and one two-stranded β sheet.

The Helical-2 domain, which is located between the two subdomains of the HEPN1 domain, is composed of eight α helices and resembles a bean shape (Figure S1C). The Helical-2 domain contacts the single-stranded 3′ end of the crRNA repeat. In the crRNA-bound state, the Linker connecting the HEPN1 and HEPN2 domains is partially disordered, in contrast to the ordered structure in the crRNA-free state (discussed further below). Those residues that are ordered in the crRNA-bound structure form eight α helices, while residues 914 to 1005 are disordered.

Structure of the crRNA Repeat

The crRNA used for crystallization consists of 28-nt guide segment (G1-U28, in gray), a 28-nt repeat scaffold (C(−28)-C(−1), in red) (referred to as the 5′-handle), and two guanosines, which were introduced during the crRNA preparation via the in vitro transcription (Figure 2A, left). Clear electron density is observed for the 5′-handle (Figure S2A).

The 5′-handle of the crRNA adopts a stem-loop structure, consisting a stem of five Watson-Crick base pairs and a loop of 9 nucleotides (Figure 2B). The 9-nt loop is positioned in the cavity formed between the NTD and Helical-1 domains (Figure 2C), with the stem being located inside the highly basic groove between the NTD and the Helical-1 domains (Figure 2D). The nucleotides C(−25)-C(−22) form a duplex with nucleotides G(−9)-G(−12). Unexpectedly, the nucleotide A(−26), which is predicted to be single-stranded based on the sequence, forms a base pair with U(−6), thus extending the stem to five base pairs (Figure 2E). Notably, the nucleotides C(−7) and A(−8), located between the U(−6)-A(−26) base pair and G(−9)-C(−25) base pair, are flipped out, thus adopting a single-stranded bulge conformation (Figure 2E). As a result, the stem of the repeat contains five base pairs and a 2-nt bulge, which are not observed in the crRNA of Cpf1 and Cas9. To check whether this bulge-containing stem structure is present across species, we compared the C2c2 crRNA repeat sequences from Lsh, *Leptotrichia buccalis* (Lbu), and *Listeria seeligeri* (Lse). These three repeat sequences exhibit sequence identity of ~67% (Figure S2B). Similar to the structure of the Lsh crRNA repeat, the repeats of Lbu and Lse crRNA form a stem-loop structure, with an A-U base pair at the base of the stem and a 2-nt bulge comprising nucleotides A(−8) and C(−7) within the stem (Figure 2A), indicating that this structure is a universal feature of C2c2 guide RNA.

Next, we analyzed the interactions between the crRNA stem and C2c2. This stem is stabilized by the electrostatic contacts and hydrogen bonds with the side chains of the positively charged residues (Figure 2D). The phosphates of the U(−6)-A(−26) base pair are stabilized through interactions with the side chains of Lys489 and Lys441, respectively (Figure 2F). The phosphate backbones of the nucleotides C(−25) to C(−22) and nucleotides G(−12) to G(−9) are stabilized through hydrogen bonds and charge-charge interactions with the NTD, Helical-1, and HEPN2 domains (Figures S2C and S2D). Positions of the positively charged residues that contact the RNA phosphate

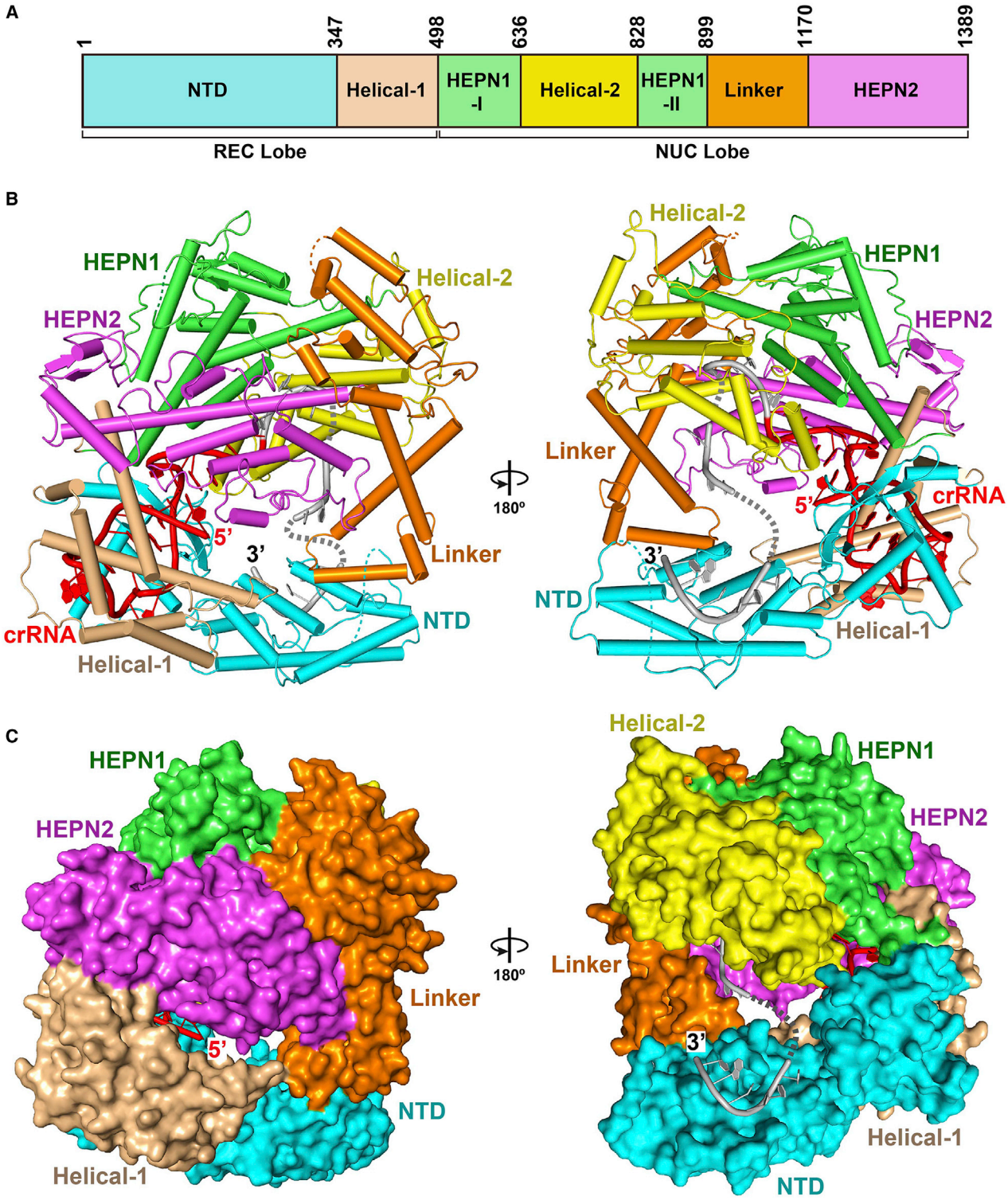


Figure 1. The Crystal Structure of LshC2c2-crRNA Binary Complex

(A) Domain organization of LshC2c2.

(B) Two views of a ribbon representation of LshC2c2-crRNA binary complex. Color-codes of LshC2c2 are defined as in Figure 1A. The crRNA repeat and spacer are shown in red and gray, respectively.

(C) Surface representations of the crystal structure LshC2c2 in complex with crRNA shown in the same view as in Figure 1B.

See also Figure S1 and Table S1.

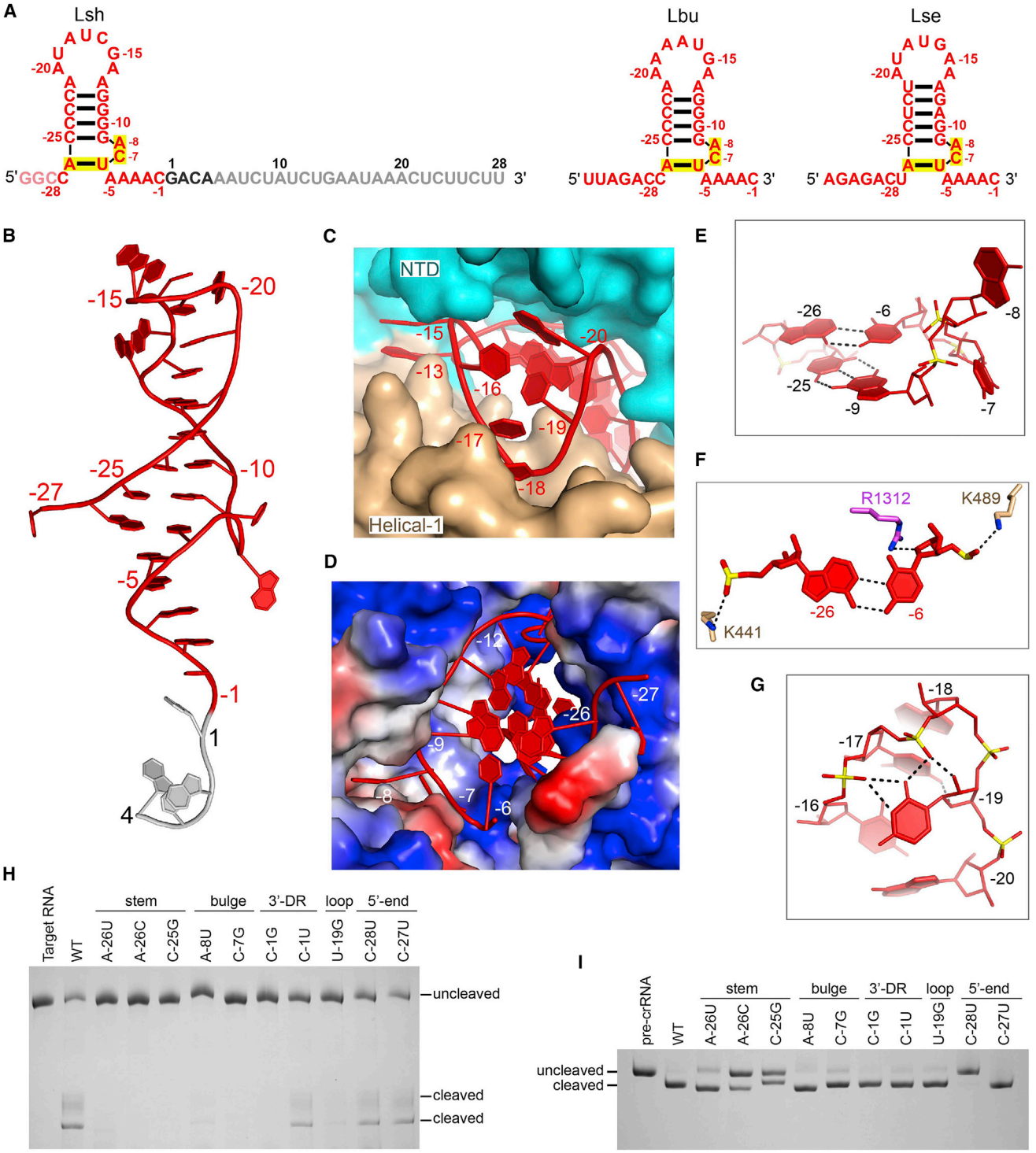


Figure 2. Structure of the crRNA

(A) Secondary structure diagram of the Lsh crRNA (left), the 5′-repeat of crRNA of Lbu (center), and of Lse (right). The repeat region is shown in red and the spacer in deep gray. The A(−26)-U(−6) base pair and 2-nt bulge within the stem are highlighted with yellow background. In the left panel, nucleotides not observed in the structure of LshC2c2-crRNA binary complex are shown in pink and light gray for the repeat and spacer regions, respectively.

(B) Structure of the crRNA in the LshC2c2-crRNA complex. The repeat region is shown in red and the spacer in gray.

(C) Binding of the crRNA 5′-handle in the channel between the NTD and Helical-1 domains.

(D) The crRNA stem binds in the positively charged channel formed between the NTD and Helical-1 domains.

(legend continued on next page)

backbone (Arg219, Lys489, Arg475, and Arg1312) are conserved among the five known C2c2 sequences (Figure S3), indicating that the positively charged groove is a universal feature of the C2c2 family. In addition to phosphate groups, the 2′-hydroxyl groups of the crRNA stem form extensive interactions with the NTD, Helical-1, and HEPN2 domains (Figures 3, S2C and S2D).

We then analyzed the structure of the crRNA loop, which contains nucleotides A(−21) to A(−13) (Figure S2E). The base of nucleotide A(−14) is sandwiched by the base pair C(−22)-G(−12) and the nucleotide A(−21). Nucleotides A(−13) and G(−15) are flipped out and are positioned within the narrow cleft between the NTD and the Helical-1 domain (Figure S2F). Nucleotides U(−17) and A(−18) stack onto each other and extrude out of the loop (Figure S2G). The base of nucleotide U(−19) stacks with the base of C(−16). Notably, the base and sugar of U(−19) form multiple hydrogen bonds with the nucleotides U(−17) and C(−16) (Figure 2G), indicating that nucleotide U(−19) plays a critical role in stabilizing the crRNA loop structure. This observation explains why mutation of U(−19) prevents cleavage of target RNA by C2c2, as reported earlier (Abudayyeh et al., 2016). In addition to the hydrogen bonds and aromatic stacking interactions formed within the crRNA loop, the 9-nt loop conformation is further stabilized by the extensive inter-molecular interactions between the crRNA and the REC lobe of the C2c2 protein (Figure S2H).

In addition to the stem-loop, the crRNA repeat contains a single-stranded 3′-directed repeat (3′-DR) region of five nucleotides (Figure 2B). Nucleotides A(−5)-A(−2) within this single-stranded 3′-DR maintain a nearly A-form conformation along the ribose-phosphate backbone, with a slight kink between nucleotides C(−1) and A(−2) (Figure S2I). To maintain this helical conformation, extensive hydrogen bonds are primarily formed between the HEPN2 domain and the backbone of the crRNA, including phosphate and 2′-hydroxyl groups (Figure S2J).

Both Structure and Sequence of the 5′-Handle Are Essential for C2c2 RNase Activities

To study the functional importance of both sequence and structure of the C2c2-bound crRNA, we tested crRNA-guided target RNA cleavage using mutant crRNAs (Figure 2H). Mutation of stem nucleotides A(−26) and C(−25) and bulge nucleotides A(−8) and C(−7) completely abolished cleavage. Furthermore, mutation of the loop nucleotide U(−19) also abolished target cleavage. These results suggest that the bulge-containing stem-loop structure is essential for crRNA-guided target cleavage. Furthermore, nucleotide C(−1) was mutated to either G or U, resulting in no or reduced activity, suggesting that this nucleotide is also important. In contrast, mutation of nucleotides at the 5′ end of the crRNA showed no effect on cleavage activity, indicating that this region plays no direct role for in vitro target cleavage.

Next, we used mutagenesis of our pre-crRNA to determine the importance of individual nucleotides for pre-crRNA processing (Figure 2I). Mutation of nucleotides either A(−26) or C(−25) in the stem region abolishing base-pairing adjacent to the base of the stem markedly reduced cleavage activity. In contrast, single mutation of nucleotides in the bulge showed little effect, suggesting that pre-crRNA processing is less affected by the type of nucleotide within the bulge. However, as discussed below, the presence of this bulge is absolutely essential for pre-crRNA cleavage. Mutation in either the loop region or the 3′-DR region showed no effect, suggesting that the precise sequences in these regions are not essential for pre-crRNA processing. Finally, mutation at the 5′ end of C(−28) completely abolished activity, whereas C(−27)U mutant lacked any effect of pre-cRNA cleavage. This observation is in agreement with a previous report for LbuC2c2, which showed that mutation of the two nucleotides adjacent to the scissile site abolished pre-crRNA processing (East-Seletsky et al., 2016).

Recognition of the 5′ Handle of the crRNA

The crRNA repeat stem-loop is recognized by LshC2c2 in a sequence-specific manner via extensive interactions (Figure 3). Arg1312, located at the HEPN2 domain, makes sequence-specific hydrogen-bonding contact in the minor groove of the crRNA stem to nucleotide C(−25). In addition, Arg1312 forms one hydrogen bond with the 2′-hydroxyl group of C(−25) (Figure 4A). The 2-nt bulge within the crRNA stem is also sequence-specifically recognized. Nucleotide A(−8) inserts into a cavity formed by the Helical-2 and HEPN1 domains. The base of A(−8) stacks on the side chain of the conserved Trp865, and its base forms hydrogen bonds with the side-chain of the conserved Gln759 (Figure 4B). The base of C(−7) is stabilized in a sequence-specific manner via the side chains of Thr502 and His340 (Figure 4C). Notably, nucleotide C(−1) is stabilized by forming hydrogen bonds with Arg858 and Asn857 via its base and its 2′-hydroxyl group (Figure 4D).

In addition to the crRNA stem and the 3′-DR region, the crRNA loop region is also recognized in a sequence-specific manner by C2c2. The nucleotides A(−21), A(−14), and A(−13) within the loop are recognized sequence specifically by Arg433, Ser333, and Lys405, respectively (Figure 4E). Together with a previous study reporting that a crRNA containing an A(−21)G mutation abrogated target RNA cleavage (Abudayyeh et al., 2016), our results suggest that these sequence-specific interactions are critical for C2c2-mediated RNA cleavage.

C2c2 Protein Mutagenesis Studies

To explore the C2c2-crRNA interactions required for RNA-guided target RNA cleavage, we performed in vitro cleavage assays using LshC2c2 protein mutants. In parallel, we performed

(E) Detailed view of the A(−26)-U(−6) and the C(−25)-G(−9) base pairs in the stem with a 2-nt (A(−8) and C(−7)) bulge in between. The phosphate groups are shown in yellow.
(F) Interactions between A(−26)-U(−6) base pair and LshC2c2. Intermolecular hydrogen bonds are shown as black dashed lines.
(G) Close-up view of the nucleotides A(−20)-C(−16). Interactions within this region are shown by the black dashed lines.
(H) LshC2c2 catalyzed crRNA-guided RNA cleavage using the mutant pre-crRNA.
(I) LshC2c2 catalyzed pre-crRNA cleavage assay using the mutant pre-crRNA.
See also Figure S2.

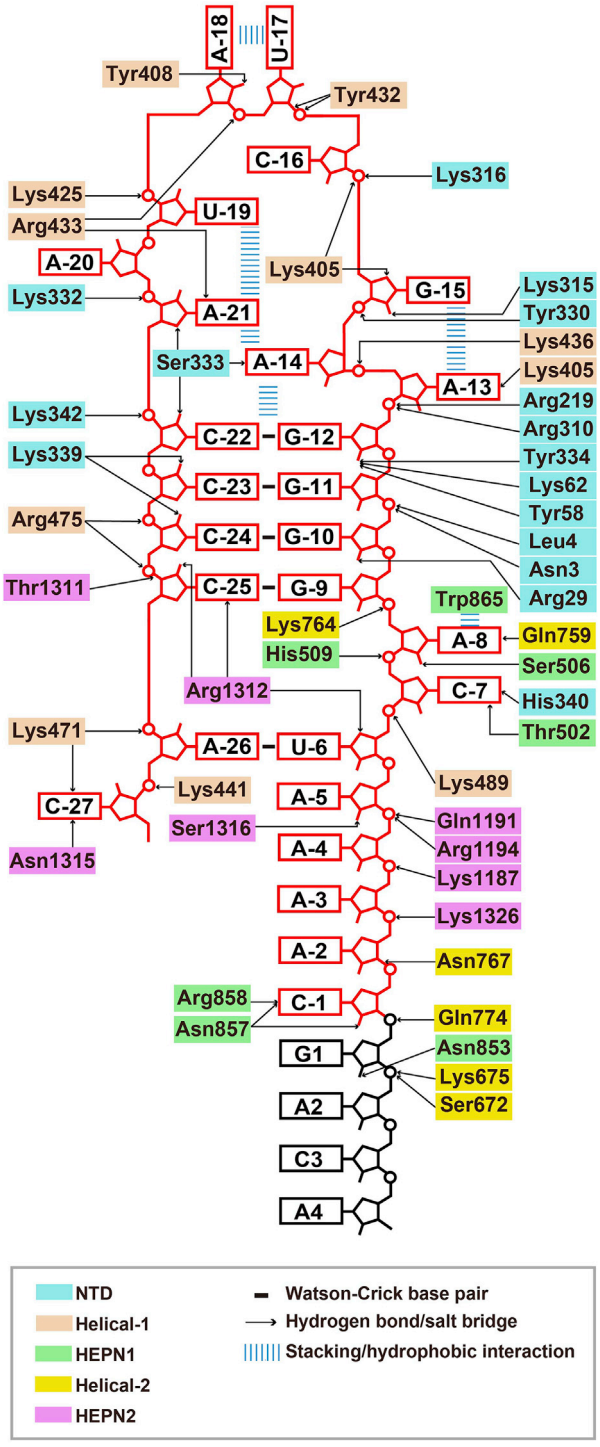


Figure 3. Schematic View of the Intermolecular Contacts between LshC2c2 and crRNA

The guide segment downstream nucleotide 4 is omitted. Hydrogen bond and salt bridge interactions are shown by black lines, and the stacking interactions are shown by dashed blue lines.

See also Figure S3.

sequence alignment of C2c2 proteins of different species to identify the degree of conservation of specific protein residues (Figure S3). Alanine substitution of the conserved Arg1312, which contacts the crRNA stem in a sequence-specific manner, abolished target RNA cleavage (Figure 4F). In addition, alanine substitution at either residue Trp865 or His509 abolished target RNA cleavage activity, suggesting that the contacts between the LshC2c2 and the 2-nt bulge are critical for target RNA cleavage. Furthermore, mutation of non-conserved Arg858, which interacts with C(−1) of the crRNA, abolished target RNA cleavage activity. This observation is consistent with the result of our target RNA cleavage assay using C(−1) crRNA mutants (Figure 2H). However, alanine substitution of either Ser1316 or Tyr330, both of which contact the phosphate backbone of the crRNA loop, showed little effect on target RNA cleavage. In addition, alanine substitution of the conserved Asn1315, which forms a hydrogen bond with the base of nucleotide C(−27), showed little effect on target RNA cleavage. These observations suggest that the interaction between C(−27) with LshC2c2 is less important for target RNA cleavage.

We next tested the functional importance of specific LshC2c2 residues for pre-crRNA cleavage. Arg1312Ala mutant showed little effect on the pre-crRNA cleavage (Figure 4G). In contrast, the mutation of either Asn1315 or Trp865 dramatically reduced pre-crRNA cleavage, suggesting that the spatial orientation and the stability of nucleotides C(−27) and A(−8) are critical for pre-crRNA processing. However, mutation of His509, Arg858, Ser1316, Tyr330, or Tyr334 lacked any effect on pre-crRNA cleavage. This result suggests that the interactions between C2c2 and either C(−7), C(−1), or the crRNA loop are not essential for pre-crRNA cleavage.

In summary, our mutagenesis studies of both C2c2 (Figures 4F and 4G) and crRNA (Figures 2H and 2I) provide strong evidence that the bulge-containing structure of the crRNA is critical for the dual RNase activities of C2c2 and strongly suggest that C2c2 recognizes its crRNA in a sequence- and structure-specific manner. For target RNA cleavage activity, the stem loop and the 3′-DR of the crRNA are essential, whereas the 5′ end, as well as the bulge-containing stem, are critical for precursor crRNA cleavage.

The crRNA Guide Strand Makes a U-Turn at the Concealed 5′ End

When analyzing the crRNA guide, we observed an unambiguous electron density for 7 nucleotides at the 5′ end and 6 nucleotides at the 3′ end of the 28-nt spacer segment. The nucleotides 1 to 4 extend the single-stranded form of the 3′-DR and bind at the cleft formed by the HEPN1 and Helical-2 domains. The 5′ end of the guide makes a U-turn at nucleotides 4-5, thus the subsequent 3-4 nucleotides bind inside the groove formed by the Linker and HEPN2 domain. Therefore, the 5′ end of the spacer containing 8 to 9 nucleotides is buried inside the NUC lobe. The subsequent central region of the spacer containing 8 to 10 nucleotides was disordered. However, we observed a discontinuous electron density following the nucleotides at the 5′ end, providing a clue for the central spacer-binding channel (Figures 4H and S4A). This suggests that the central spacer binds at the open groove formed by the NTD and the Helical-2 domain and

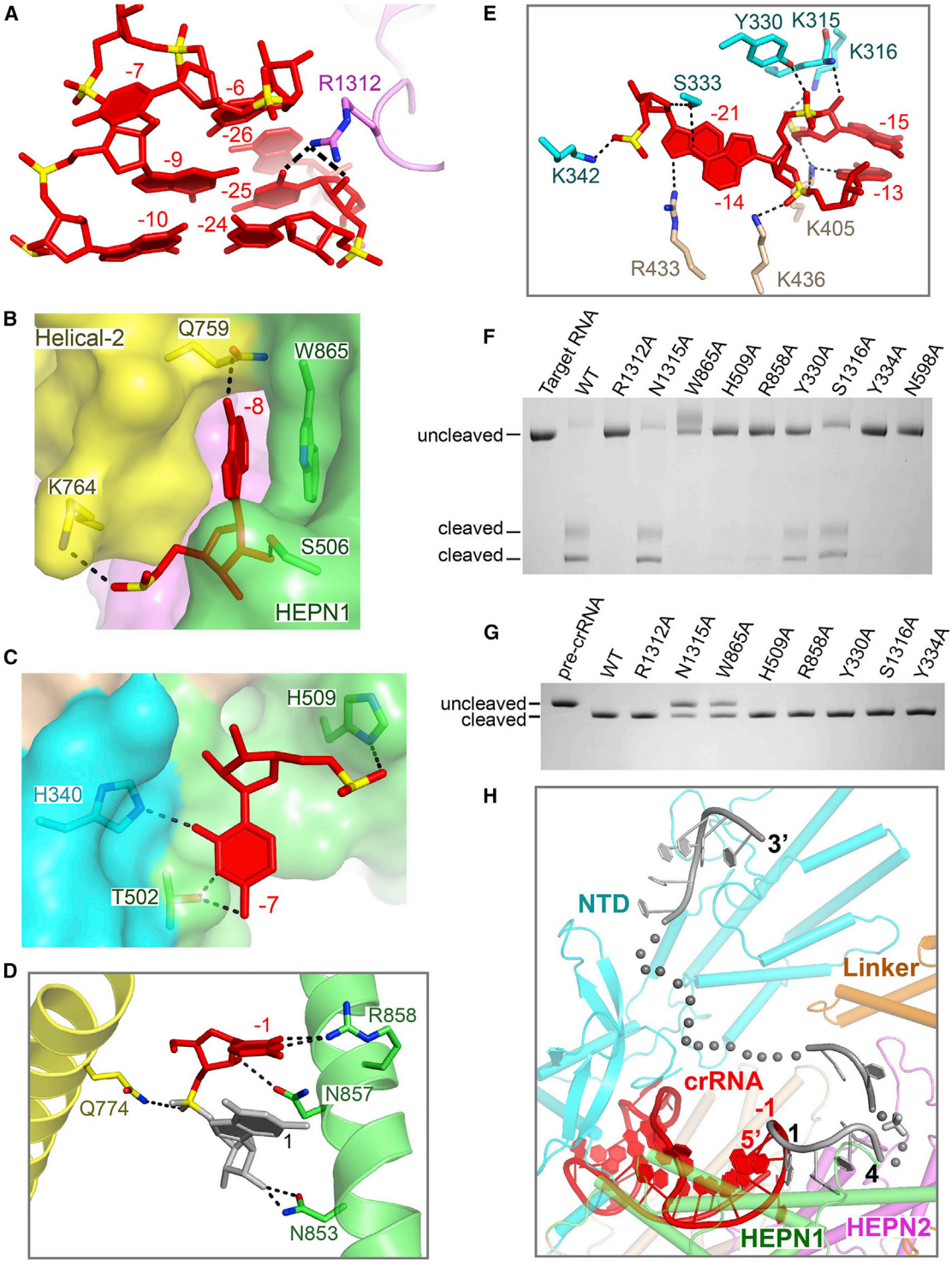


Figure 4. Recognition of the crRNA 5′-Handle

(A) The stem region within the 5′-handle is recognized by Arg1312 via its minor groove.

(B) Recognition of nucleotide A(−8).

(C) Recognition of nucleotide C(−7).

(D) Recognition of nucleotide C(−1).

(E) Recognition of the 5′-handle loop.

(legend continued on next page)

therefore is exposed to the bulk solvent (Figures S4B and S4C). Notably, the 3′ end of spacer is located at the concave surface within the larger subdomain of the NTD domain (Figure S4D). Thus, the 28-nt spacer is likely to adopt a non-helical conformation, with the 5′ end concealed by the C2c2 protein, with the central region and 3′ end being exposed. These features provide the structural basis for target RNA recognition and may allow the central, exposed region of the spacer to nucleate with the target (Figure 4H). This observation may explain a previous finding showing that the seed region of C2c2 is located in the central segment of the crRNA spacer rather than at the 5′ end (Abudayyeh et al., 2016). However, further structural studies are required to identify the accurate seed region and reveal the molecular mechanism for target recognition.

Lsh Pre-crRNA Cleavage Site

The RNA used for crystallization consists of two cytosines and two guanines upstream of the stem-loop (Figure 2A). However, we only observed ordered electron density for nucleotide C(−27), while the other three nucleotides at the 5′ end are removed during crystallization, as confirmed by denaturing gel (Figure S5A). Therefore, the crRNA observed in our structure is the mature form that functions as the guide strand in the subsequent crRNA interference step. Thus, the crystal structure of LshC2c2 bound to crRNA represents not only the pre-target bound state but also represents the product-bound state of the pre-crRNA processing.

When analyzing the LshC2c2-crRNA binary complex, we observed 27 nucleotides within the crRNA repeat. The nucleotide C(−27) is flipped away from the stem-loop, resulting in a kink between nucleotides C(−27) and A(−26) (Figure 5A). C(−27) is positioned at the positively charged cleft formed between the Helical-1 and HEPN2 domains. The base of C(−27) is sandwiched by Asn1314 and Lys1339 and is further stabilized by hydrogen bonds with Lys471 and Asn1315 via its O2 and N4 atoms, respectively (Figure 5B). The side chain of Arg438 lies adjacent to the ribose of C(−27), suggesting that this residue is also involved in pre-crRNA cleavage.

Two previous studies are in disagreement in terms of the exact position of the cleavage site. One study showed that the pre-cRNA is processed immediately upstream of nucleotide C(−28) (Abudayyeh et al., 2016), while a second study reported that LshC2c2 cleaves the pre-crRNA between the nucleotides C(−28) and C(−27), producing a mature crRNA with one nucleotide shorter than the previous study at the 5′-terminal (East-Seletsky et al., 2016). The LshC2c2-crRNA complex structure reveals that the mature crRNA of LshC2c2 contains a 27-nt repeat, which is in agreement with the latter finding (Figure 5C). Thus, LshC2c2 cleaves the pre-crRNA adjacent to the base of the stem-loop structure, reminiscent of the pre-crRNA cleavage site catalyzed by the type I CRISPR-Cas endonuclease Cas6 (Gesner et al., 2011; Haurwitz et al., 2010; Sashital et al., 2011). In addition, cleavage of the pre-crRNA during crystallization, together with our cleavage assay results (shown in Figure 2I), suggests that the full-length crRNA repeat is not required for successful pre-crRNA cleavage. Furthermore, our pre-crRNA containing GG at positions (−30)-(−29) instead of GA retains wild-type cleavage activity, strongly suggesting that the nucleotides (−30)-(−29) are not essential. In contrast, nucleotide C(−28) is critical for C2c2-catalyzed pre-crRNA processing.

The Catalytic Sites for Pre-crRNA Processing

As LshC2c2 cleaves pre-crRNA directly upstream of nucleotide C(−27), we hypothesized that the amino acids proximal to the 5′-OH of nucleotide C(−27) represent the catalytic residues for the pre-crRNA cleavage. To identify the exact residues representing the catalytic center, we performed point-mutation of amino acids located proximal to the 5′-OH group of nucleotide C(−27). Specifically, we replaced each polar amino acid within the Helical-1 and HEPN2 domains with an alanine residue and assessed pre-crRNA cleavage activity. Mutation of Lys471 severely impaired crRNA processing, suggesting that this modest conserved residue plays an important role in substrate orientation (Figure 5D). Alanine substitution of Arg438 and Lys441 abolished pre-cRNA processing in vitro, indicating that Arg438 and Lys441 are critical for pre-crRNA processing. In contrast, we showed that mutations of other amino acids in the Helical-1 domain, including Arg433, Lys448, Glu470, Lys474, Lys452, Arg475, and Glu482, had little effect on activity. Next, we tested the effect of several mutations within the HEPN2 domain. Alanine substitution of the conserved Asn1315, which makes base-specific contact with C(−27), clearly reduced cleavage activity (Figures 4G and 5B), suggesting that residue Asn1315 is essential, even though it is not directly involved in catalysis. However, mutations of other residues proximal to nucleotide C(−27) in the HEPN2 domain, including Asp1332, Asp1334, Glu1335 Lys1337, Lys1338, Lys1339, and Arg1312, showed little effect on the C2c2-catalyzed pre-crRNA processing. These observations indicate that the catalytic pocket for the crRNA maturation is not located within the HEPN2 domain.

Together, our results show that the Helical-1 domain plays a key role in C2c2-catalyzed crRNA maturation and that the conserved Arg438 and non-conserved Lys441 are both critical for pre-crRNA processing. The lack of conservation of residue Lys441 might be related to the fact that the cleavage position varies between pre-crRNAs associated with C2c2 of different species. Therefore, the REC lobe is not only essential for crRNA recognition but also critical for pre-crRNA cleavage.

External HEPN Catalytic Site

Bioinformatics analysis suggests that the two HEPN domains of C2c2 are responsible for cleavage of target RNA (Shmakov et al., 2015). Another previous study using site-directed mutagenesis confirmed that the highly conserved Arg597 and His602 within the HEPN1 domain, and Arg1278 and His1283 in the HEPN2 domain are critical for target RNA cleavage (Abudayyeh et al.,

(F and G) Mutational analysis of crRNA-interacting residues. Effects of mutations on the crRNA-guided RNA cleavage (F) and on the pre-crRNA cleavage (G). (H) The guide region-binding channel. The gray dashed line indicates the position of the disordered nucleotides within the guide region. The repeat and guide segments are shown in red and gray, respectively.
See also Figure S4.

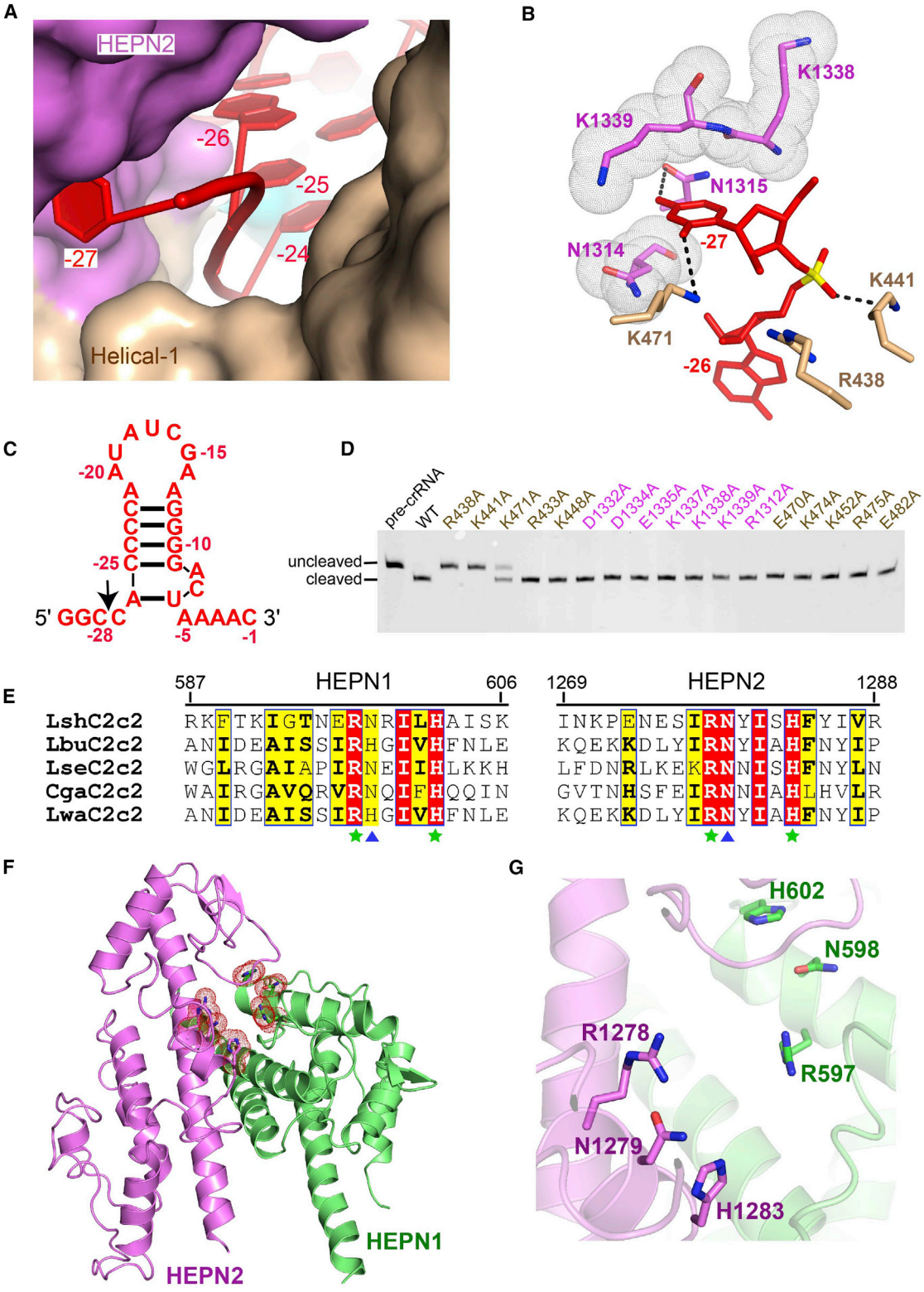


Figure 5. Two Distant LshC2c2 Catalytic Sites for Its Two RNase Activities

(A) The magnified view of 5′ end of the crRNA binding in the cleft formed between the HEPN2 and Helical-1 domains.

(B) The interaction between nucleotide C(−27) and LshC2c2.

(C) LshC2c2 cleaves the pre-crRNA between the nucleotides C(−28) and C(−27) (cleavage site indicated by a black arrow).

(legend continued on next page)

2016). Here, we found that the residues Asn598 in the HEPN1 domain and Asn1279 in the HEPN2 domain are conserved within the C2c2 family (Figure 5E). The Asn residue subsequent to the conserved Arg in the R-X_{4-6}-H motif is also weakly conserved within the HEPN domain of most Cas proteins (Shmakov et al., 2015). We therefore hypothesized that this residue is essential for target RNA cleavage. In agreement with our hypothesis, when either Asn598 or Asn1279 was substituted by alanine, cleavage of target RNA was completely abolished (Figures 4F and S5B).

Next, we analyzed the structure of the catalytic site within the two HEPN domains. As shown in Figure 5F, the conserved HEPN1 and the HEPN2 domains are positioned side by side and resemble an X shape, resulting in the juxtapositioning of the catalytic R-X_{4-6}-H motifs from the HEPN1 and HEPN2 domains to form a composite symmetric active pocket. Unexpectedly, the four catalytic residues in the HEPN domains are located at the external concave surface (Figures 5G and S5C). Previously, the HEPN domain was also identified in other CRISPR-associated endonucleases, including *Thermus thermophilus* Csm6 in type III-A system (Niewoehner and Jinek, 2016), *Pyrococcus furiosus* Csx1 (Kim et al., 2013), and *Sulfolobus solfataricus* Csx1 (Lintner et al., 2011) in type III-B system. All of these endonucleases possess one HEPN domain and form a composite ribonuclease active site via dimerization. However, the relative orientation of the two HEPN domains of LshC2c2 differs to that of Csm6 dimer and Csx1 dimer (Figure S5D).

The Crystal Structure of C2c2 in the apo State

To obtain further structural insights into the architecture of LshC2c2 before binding to the crRNA, we solved the crystal structure of LshC2c2 in the absence of crRNA at 3.5 Å resolution (Figure 6A). Unexpectedly, the initial crystal of this apo form lacked the N-terminal NTD domain. This suggested that the NTD domain has been removed by a putative protease during crystallization, probably because this domain is more flexible in the apo state than in the crRNA-bound state.

To increase the quality of the LshC2c2 crystal in the apo state, we truncated the NTD domain with thermolysin prior to setting the crystallization trays and solved the crystal structure of truncated LshC2c2 with each asymmetric unit contains two C2c2 molecules in the absence of the NTD domain (Table S1). Two HEPN domains are similar to those of C2c2 in the LshC2c2-crRNA complex. Notably, almost all residues of the Linker, which connects the HEPN1 and HEPN2 domains, are visible in the crRNA-free structure. This observation is in contrast to our finding that the Linker region is partially disordered in the crRNA-bound state. The Linker in the RNA-free state consists of three regions, with each region containing four α helices (Figure S6A). In the RNA-bound state, the first region of the Linker is disordered, whereas this region forms a four-helical bundle and extends away from the other parts of C2c2 apo structure.

To compare the stability of LshC2c2 in different states, we performed limited proteolysis of C2c2 with and without crRNA and of C2c2 containing both crRNA and target RNA (Figure S6B). Both the LshC2c2-crRNA and LshC2c2-crRNA-target RNA complexes were more resistant to limited proteolysis than crRNA-free LshC2c2, indicating that, in the crRNA bound state, LshC2c2 has undergone a higher degree of compaction.

Conformational Change of C2c2 upon crRNA Binding

Lastly, our comparison between crRNA-free and crRNA-bound C2c2 structures revealed that C2c2 undergoes a significant conformational change upon crRNA binding (Figures 6B and S6C). Specifically, the Helical-2 domain and the Linker within the NUC lobe undergo substantial rearrangement relative to the two HEPN domains (Figure 6C). The Helical-2 domain and the Linker adopt a tweezers-like configuration, with the two domains remaining close to each other at the distal end of the NTD domain in either state. In the crRNA-free state (indicated by the black arrow in Figure 6D, left panel), the two domains are separated from each other at the proximal end of the NTD domain. Upon crRNA binding, the Helical-2 domain rotates toward the Linker and the HEPN2 domain, with an α-helix-α-helix (residues 750–780) covering the crRNA-binding channel (Figure 6D, right panel). Therefore, the movement of the Helical-2 domain results in a nearly-closed crRNA-binding channel, which is further stabilized by the interactions between the crRNA and the Helical-2 domain (Figure 3). Thus, these conformational changes are likely to play an important role in both stabilization of the crRNA and facilitating target RNA binding. Interestingly, Cas9 and Cpf1 have also been shown to undergo a dramatic conformational rearrangement upon sgRNA binding (Jiang et al., 2015; Jinek et al., 2014).

DISCUSSION

Our structures presented here of the crRNA-bound and crRNA-free LshC2c2 provide mechanistic insights for a more detailed understanding of pre-crRNA processing and RNA-guided RNA cleavage. To date, C2c2 remains the only single-RNA guided ribonuclease identified in Class 2 CRISPR-Cas systems, whereas Cas9, Cpf1, and C2c1 are RNA-guided DNA endonucleases.

C2c2 Possesses Two Independent Catalytic Sites for Its Dual RNase Activities

C2c2 has two distinct RNase activities, one for pre-crRNA processing and the other for crRNA-guided RNA cleavage

(D) Functional analysis of possible catalytic residues of LshC2c2 for pre-crRNA cleavage.
(E) Sequence alignment of conserved R-X-H motifs in the HEPN1 and HEPN2 domains. The highly conserved Arg and His are highlighted by green starts, and the conserved Asn are highlighted by blue triangles.
(F) The structure of the HEPN1 and HEPN2 domains. The conserved residues, including Arg, Asn, and His within the R-X_{4-6}-H motif, are shown in sticks and red dots.
(G) The closed-up view of the HEPN catalytic site, which is responsible for the crRNA-guided RNA cleavage.
See also Figure S5.

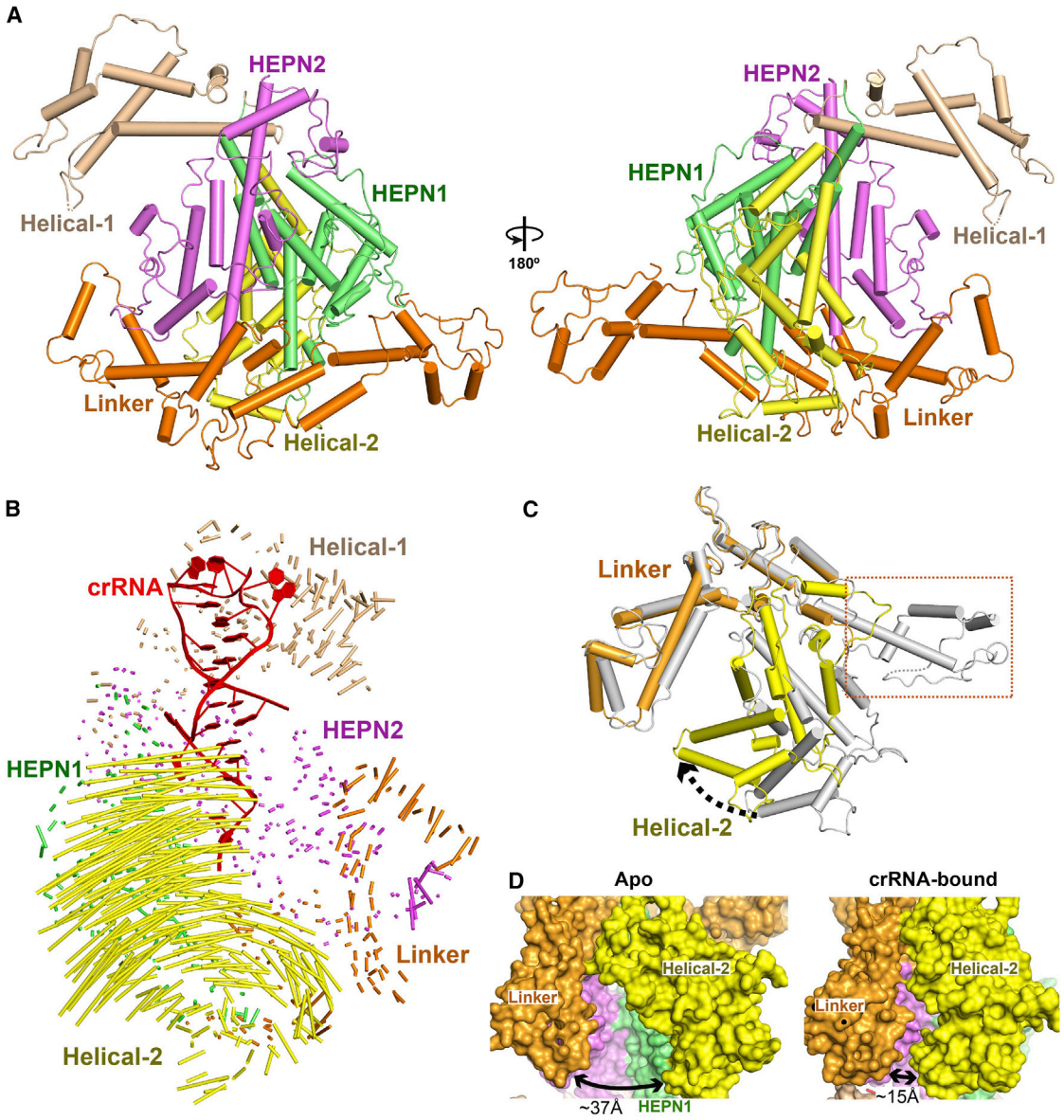


Figure 6. Structural Rearrangement of LshC2c2 upon crRNA Binding

(A) Two views of the crystal structure of C2c2 in the apo state (color coded as defined in Figure 1A).

(B) Structural comparison between C2c2-crRNA binary complex and C2c2 in the apo state. Vector length correlates with the domain motion scale (color coded as defined in Figure 1A).

(C) Superposition of the LshC2c2-crRNA complex onto the LshC2c2 in the absence of crRNA shows that LshC2c2 undergoes conformational change upon crRNA binding. For clarity, only the Helical-2 domain and the Linker are shown. The black dashed arrow indicates domain movement of C2c2 upon crRNA binding. The orange dashed line box highlights the order Linker in the apo state. The color code of LshC2c2-crRNA complex is identical to that used in Figure 1A, with C2c2 in the apo state indicated in gray.

(D) The movement of the Helical-2 domain upon the crRNA binding results in a closed crRNA-binding channel. The "open" conformation of crRNA-binding channel in the C2c2 apo state (left). The "closed" conformation of crRNA-binding channel in the crRNA-bound state (right). The black double-headed arrows indicate the distance between the Linker and the Helical-2 domain.

See also Figure S6.

(East-Seletsky et al., 2016). Our findings suggest that the Helical-1 domain has the catalytic pocket for pre-crRNA processing. Lys441, Arg438, and Lys471 on the Helical-1 domain are essential for pre-crRNA cleavage. In addition, previous work revealed that two conserved HEPN domains catalyze crRNA-guided RNA degradation, and the conserved R-X-H motif is critical for target RNA cleavage (Abudayyeh et al., 2016; Shmakov et al., 2015). These findings strongly suggested that C2c2 possesses two distinct RNase catalytic sites, responsible for its two distinct RNase activities: crRNA maturation and crRNA-guided RNA

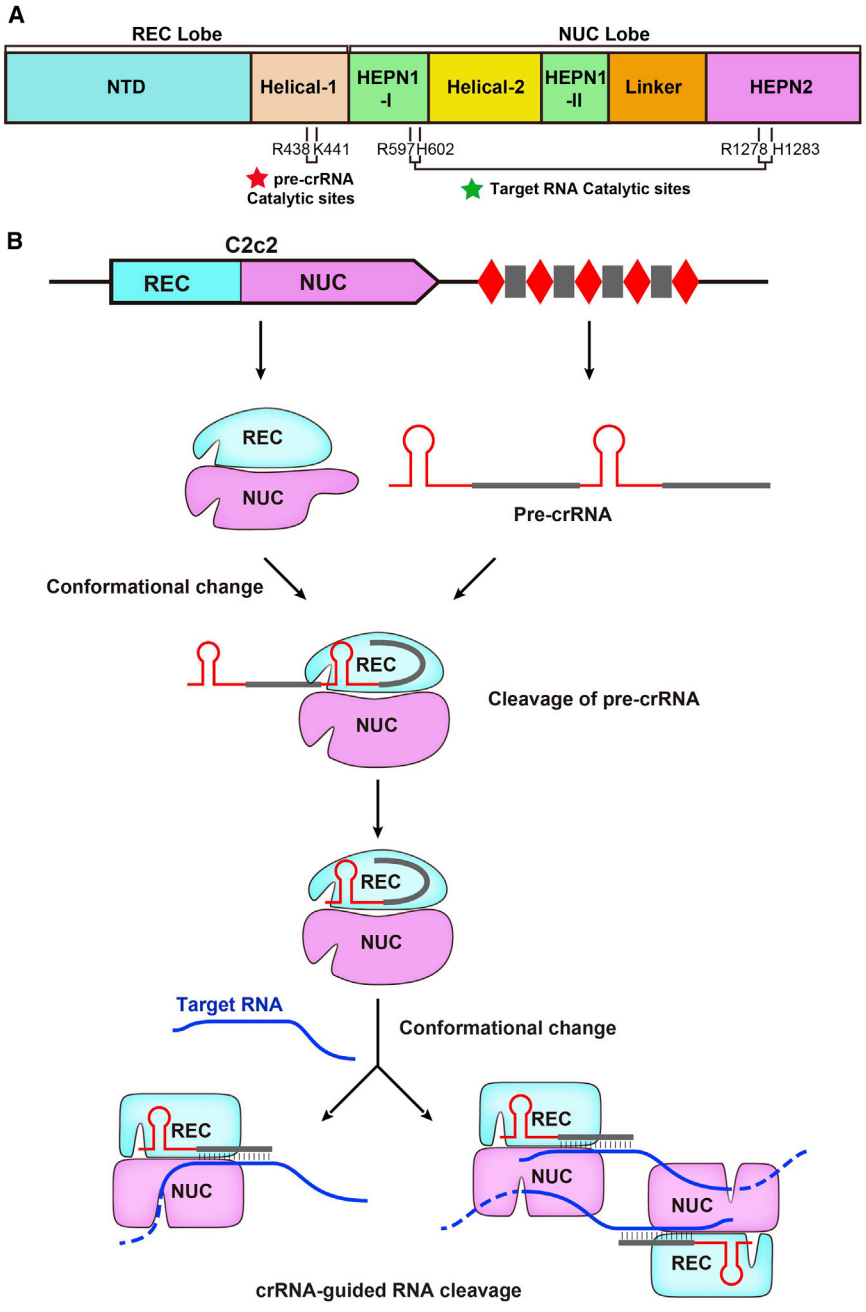


Figure 7. Model for C2c2 Catalyzed crRNA-Guided RNA Cleavage

(A) LshC2c2 has two catalytic sites, which are highlighted by the red and green stars for the pre-crRNA cleavage and target RNA cleavage. The possible catalytic residues are indicated.

(B) Model of pre-crRNA processing and crRNA-guided RNA cleavage by C2c2. The REC lobe (shown in cyan) is involved in cleavage of pre-crRNA. The NUC lobe (shown in magenta) undergoes a conformational change upon crRNA binding. Upon the target RNA being recognized and forming duplex with the guide, the crRNA guide and C2c2 protein may undergo further conformational changes. C2c2 cleaves the bound target RNA by its catalytic site within the HEPN domains in a monomer or dimer configuration. See also Figure S7.

cleavage (Figure 7A). The mutation in one catalytic pocket of C2c2 showed no effect on the activity of the second site. Together, our studies revealed that the two RNase functions are separated due to the fact that they are located at different domains of C2c2.

Comparisons between Class 2 Effectors

C2c2, Cas9, Cpf1, and C2c1 are the single effector proteins associated with Class 2 CRISPR-Cas systems, for which structures are available. The shared trait of theses enzymes is that they possess RNA-guided endonuclease activity, with the ability to cleave target nucleic acids complementary to the guide RNA sequence. However, their structures and catalytic mechanisms exhibit significant difference.

Structural comparison with the Class 2 effectors Cas9, Cpf1, and C2c1 reveals that, despite C2c2 sharing a similar size and a two-lobed architecture (Figure S7), its structure differs both at the architectural level of the entire protein and at the level of its domain organization. At the architectural level, Cas9 adopts an open-clam shape (Jiang et al., 2015; Jinek et al., 2014), while Cpf1 adopts a compact triangular shape in the crRNA-bound state (Dong et al., 2016). In comparison, our C2c2 structure resembles a rhombohedron, similar to a clinched fist. At the domain level, although the domains of Cas9, Cpf1, C2c1, and C2c2 have been assigned similar nomenclatures, both their actual structures and functions differ significantly, with none of the C2c2 domains possessing any structural similarity to those of Cas9, Cpf1, or C2c1. Furthermore, C2c2 endonuclease lacks the RuvC domain found in Cas9, Cpf1, and C2c1; instead, it contains two HEPN domains function as the catalytic domain (Shmakov et al., 2015).

Next, we compared the functional regions of the guide strand. The crRNA of C2c2 possesses a central seed region (Abudayyeh et al., 2016), whereas Cas9, Cpf1, and C2c1 contain either 3′- or 5′-end seed regions proximal to the repeat (Jinek et al., 2012; Zetsche et al., 2015). While the bases at the 5′ end of the Cpf1 guide strand and 3′ end of the Cas9 guide strand are exposed, the 5′ end spacer of the LshC2c2-associated crRNA is concealed inside the protein. In contrast, the central region of C2c2 crRNA spacer is exposed to the bulk solvent, allowing target RNA to bind to this region and form the guide RNA-target

RNA duplex. Upon base pairing with the target RNA, the guide strand of C2c2 must undergo a conformational change to form the rigid guide-RNA-target RNA duplex. To accommodate RNA duplex, it is likely that the C2c2 protein also changes its conformation.

External HEPN Catalytic Pocket and the RNA Cleavage

Interestingly, the positions of the catalytic sites of C2c2 differ from those of the other Class 2 effectors. In Cas9, Cpf1, and C2c1, the guide RNA-target DNA heteroduplexes bind in the central channel formed by two lobes, with all catalytic pockets located between the two lobes and face the inside of the protein. In contrast, the RNA-guided RNA target cleavage site within two HEPN domains is located on the outer surface of the C2c2 (Figure S5C).

The external location of the catalytic pocket may affect the crRNA-guided RNA cleavage pattern of C2c2. In contrast to Cas9, Cpf1 and C2c1, which cleave their targets at specific sites (Gasiunas et al., 2012; Jinek et al., 2012; Shmakov et al., 2015; Zetsche et al., 2015), C2c2 cuts its RNA target more promiscuously. Given that target cleavage occurs at the outer surface of C2c2 protein and away from the guide-target RNA duplex region (Figure S5C), we speculate that, following the basing paring between the guide and target strands, the target RNA reaches the HEPN catalytic site through an external binding channel of C2c2, which can accommodate the RNA with various length and various structure. A second possibility is that the positioning of catalytic site on the outside allows for cooperative cleaving of target RNA by multiple C2c2 molecules. While our preliminary biochemical analysis yielded no meta-stable dimeric or trimeric structures, it is feasible to imagine that target RNA itself may have a transiently stabilizing effect on such oligomers.

Non-G PFS Requirement for Target RNA Cleavage

Unlike the Cas9, Cpf1, and Cascade complexes, which require the presence of a PAM in the target DNA for their RNA-guided DNA cleavage (Deveau et al., 2008; Jinek et al., 2012; Zetsche et al., 2015), C2c2-catalyzed crRNA-guided target cleavage depends on a non-G 3′ protospacer-flanking site (PFS) (Abudayyeh et al., 2016). Our structure of the LshC2c2-crRNA complex clearly shows that 3′-flanking region of the crRNA repeat, together with the downstream spacer, adopts a single-stranded form with a modest kink between nucleotides (−1) and (+1). We therefore hypothesize that the crRNA nucleotide C(−1) base pairs with the G if the target RNA carries a nucleotide G at the PFS site. The base pair between nucleotide C(−1) and the nucleotide G at PFS site may break the interactions of C(−1) with the HEPN1 domain, as well as eliminate the kink between nucleotides (−1) and (+1), thus affecting target cleavage. This may explain the requirement of non-G PFS for crRNA-guided RNA cleavage and suggests that the single-stranded form of nucleotide (−1) is critical for target cleavage.

Model of C2c2-Mediated RNA Cleavage

On the basis of our structural findings, we propose a model for type VI CRISPR system to protect the bacteria from RNA phage (Figure 7B). The two RNase activities of C2c2 are involved in crRNA biogenesis and crRNA interference. On the one hand, C2c2 cleaves the long pre-crRNA at the location upstream the stem through its Helical-1 domain, generating the mature crRNA. The mature crRNA remains bound to C2c2, with the 5′ end spacer region binding inside the positively charged channel. The central spacer segment functions as the seed region to recognize its target RNA containing the complementary sequences and non-G PFS. The external catalytic site, formed by the two conserved HEPN domains, cleaves the target RNA.

Two Class 2 CRISPR-Cas effectors, Cas9 and Cpf1, have been successfully harnessed for genome engineering based on their RNA-guided dsDNA cleavage activities. The use of CRISPR-Cas effectors for RNA studies, however, awaits further development. Thanks to its efficient RNA-guided RNA cleavage activity, C2c2 has the potential to serve as a promising tool for RNA studies. In addition, its pre-cRNA cleavage activity enables C2c2 to generate a set of crRNA guides, simultaneously targeting multiple sites, thus raising the efficiency of target recognition.

In conclusion, the crRNA recognition mechanism identified here provides the molecular details for the precursor RNA processing and target RNA cleavage by the C2c2 endonuclease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Strains
- METHOD DETAILS
  - Protein Expression and Purification
  - LshC2c2-crRNA Complex Reconstitution
  - Crystallization, Data Collection and Structure Determination
  - Limited Proteolysis Assays
  - Pre-crRNA Processing Assays
  - Target Cleavage Assays
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY
  - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2016.12.031.

AUTHOR CONTRIBUTIONS

L.L. and X.L. expressed, purified, and grew crystals of LshC2c2 and LshC2c2-crRNA complex. L.L., J.W., P.C., J.L., and G.S. performed the biochemical assays. M.W. carried out all cloning. L.L., M.Y., and J.W. collected X-ray diffraction data. L.L. and Y.W. solved LshC2c2 and LshC2c2-crRNA complex structures, and Y.W. wrote the manuscript and supervised all of the research.

ACKNOWLEDGMENTS

We thank the staff of the BL-17U1, BL-18U, and BL-19U1 beamlines at the National Center for Protein Sciences Shanghai (NCPSS) at SSRF and the staff of BL41XU beamline at SPring-8. This work was supported by grants from the Natural Science Foundation of China (31630015, 91440201, 31571335, and 31400640), the Chinese Ministry of Science and Technology

(2014CB910102), and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB08010203). We thank Dr. Torsten Juelich for critical reading of the manuscript and linguistic assistance and Dr. Yang Yu for scientific discussion.

Received: December 2, 2016
Revised: December 20, 2016
Accepted: December 20, 2016
Published: January 12, 2017

REFERENCES

Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. Science *353*, aaf5573.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. *58*, 1948–1954.

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

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Deveau, H., Barrangou, R., Garneau, J.E., Labonté, J., Fremaux, C., Boyaval, P., Romero, D.A., Horvath, P., and Moineau, S. (2008). Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. J. Bacteriol. *190*, 1390–1400.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D., et al. (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature *532*, 522–526.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270–273.

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

Gao, P., Yang, H., Rajashankar, K.R., Huang, Z., and Patel, D.J. (2016). Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res. *26*, 901–913.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA *109*, E2579–E2586.

Gesner, E.M., Schellenberg, M.J., Garside, E.L., George, M.M., and Macmillan, A.M. (2011). Recognition and maturation of effector RNAs in a CRISPR interference pathway. Nat. Struct. Mol. Biol. *18*, 688–692.

Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J.A. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. Science *329*, 1355–1358.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W559.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J.A. (2015). STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science *348*, 1477–1481.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science *351*, 867–871.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. Science *343*, 1247997.

Kim, Y.K., Kim, Y.G., and Oh, B.H. (2013). Crystal structure and nucleic acid-binding activity of the CRISPR-associated protein Csx1 of Pyrococcus furiosus. Proteins *81*, 261–270.

Lintner, N.G., Frankel, K.A., Tsutakawa, S.E., Alsbury, D.L., Copié, V., Young, M.J., Tainer, J.A., and Lawrence, C.M. (2011). The structure of the CRISPR-associated protein Csa3 provides insight into the regulation of the CRISPR/Cas system. J. Mol. Biol. *405*, 939–955.

Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M., and Wang, Y. (2016). C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. Mol. Cell *65*, 1–13.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Niewoehner, O., and Jinek, M. (2016). Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. RNA *22*, 318–329.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935–949.

Otwinowski, Z., and Minor, W. (1997). Processing of X-Ray Diffraction Data Collected in Oscillation Mode. Methods Enzymol. *276*, 307–326.

Sashital, D.G., Jinek, M., and Doudna, J.A. (2011). An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. Nat. Struct. Mol. Biol. *18*, 680–687.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell *60*, 385–397.

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.

Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.M., Li, Y., Fedorova, I., Nakane, T., Makarova, K.S., Koonin, E.V., et al. (2016). Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell *165*, 949–962.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
L(+) - Selenomethionine	Affymetrix	122243
Sodium malonate	Sigma-Aldrich	63409-100G
HEPES	Sigma-Aldrich	H3375-250G
L-Proline	Sigma-Aldrich	P5607-100G
PEG8000	Sigma-Aldrich	81268
PEG3350	Sigma-Aldrich	88276
Thermolysin	Sigma-Aldrich	T7902-25MG
Spermindine	Sigma-Aldrich	S2626-1G
ATP	Sigma-Aldrich	A2383
UTP	Sigma-Aldrich	U6625
CTP	Sigma-Aldrich	C1506
GTP	Sigma-Aldrich	G8877
Pure Link HiPure Plasmid Gigagrep Kit	Invitrogen	K210009
AxyPrep Plasmid Maxiprep Kit 25-prep	Axygen	AP-MX-P-25G
Critical Commercial Assays		
Chelating Sepharose Fast Flow	GE Healthcare	Cat# 17-0575-02
HiTrap Heparin HP	GE Healthcare	Cat# 17-0407-03
Superdex 200 Increase, 10/300 GL	GE Healthcare	Cat# 28-9909-44
Deposited Data		
The coordinates of LshC2c2-crRNA complex	This paper	PDB: 5WTK
The coordinates of LshC2c2 in apo state	This paper	PDB: 5WTJ
The coordinates of TtCsm6	Niewoehner and Jinek, 2016	PDB: 5FSH
The coordinates of PfCsx1	Kim et al., 2013	PDB: 4EOG
The coordinates of SsCsx1	Lintner et al., 2011	PDB: 2I71
The coordinates of SpyCas9-sgRNA complex	Jiang et al., 2015	PDB: 4ZT0
The coordinates of LbCpf1-crRNA complex	Dong et al., 2016	PDB: 5ID6
Experimental Models: Organisms/Strains		
Escherichia coli Rosetta (DE3)	Novagen	Cat# 70954-3
Recombinant DNA		
pET30b-C2c2, various mutants	This paper	N/A
Sequence-Based Reagents		
Pre-crRNA sequence: GGCCACCCCAAUAUCGAAGGGG ACUAAAACGACAAAUCUAUCUGAAUAAACUCUUCUU	This paper	N/A
Target RNA sequence for Limited proteolysis assays: GGAAGAAGAGUUUAUUCAGAUAGAUUUGUC	This paper	N/A
Target RNA sequence for Target cleavage assays: GGAA GAAGAGUUUAUUCAGAUAGAUUUGUCCUCGACCUUU UUAGGUCGAGUCGACGGAGUCUAGACUCCGUCCUG AUGAGUCCGUGAGGACGAAAGACUCGACCUGAUGU AUAGAGUAGA	This paper	N/A
Primer sequences	Table S2	N/A
Software and Algorithms		
Phenix	Adams et al., 2002	http://www.phenix-online.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Coot	Emsley et al., 2010	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
Pymol	Molecular Graphics System, Version 1.8 Schrödinger, LLC	http://www.pymol.org/
Other		
Amicon Ultra Centrifugal Filters Ultracel-30K	Merck Millipore Ltd	UFC903096
3 Well Midi Crystallization Plate (Swissci)	HAMPTON RESEARCH	Cat# HR3-125

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by Lead Contact Yanli Wang (ylwang@ibp.ac.cn).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains

LshC2c2 protein for crystallographic studies was expressed in *E. coli* Rosetta (DE3) cells. *E. coli* Rosetta (DE3) cells were cultured at 37°C and 16°C in LB medium (OXOID), supplemented with 50 μg/mL kanamycin (AMRESCO) and 34 μg/mL chloramphenicol (AMRESCO).

METHOD DETAILS

Protein Expression and Purification

The full-length LshC2c2 gene (encoding residues 1-1389) was synthesized from Sangon Biotech and cloned into pET-30b vector (Novagen) containing a His_6 tag at C-terminal of LshC2c2. The LshC2c2 protein was overexpressed in *E. coli* Rosetta (DE3) (Novagen) cells that were induced with 0.1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) at OD_{600} = 0.6 at 16°C for 14 hr. Cells were collected and lysed by sonication in buffer containing 20 mM Tris-HCl, pH 7.5, 500 mM NaCl and 5% glycerol. After centrifugation, supernatant was incubated with Ni Sepharose (GE healthcare). The bound protein was eluted with buffer containing 100 mM imidazole and was further purified by Heparin column (GE healthcare), and eluted with buffer containing 20 mM Tris-HCl, pH 7.5, 1 M NaCl, 5% glycerol. LshC2c2 protein was further purified by size-exclusion chromatography (Superdex 200 Increase 10/300, GE Healthcare) in a buffer containing 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% glycerol and 1 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP). Finally, LshC2c2 protein was concentrated to 20 mg/ml. The seleno-methionine (SeMet) substituted LshC2c2 was purified with an identical protocol.

LshC2c2-crRNA Complex Reconstitution

The 58-nt crRNA was synthesized via in vitro transcription with T7 RNA polymerase and linearized plasmid DNA as template. To refold the purified in vitro transcribed RNA, the crRNA was heated at 75°C for 5 min and cooling it slowly to room temperature. The C2c2-crRNA complex was reconstituted by incubating C2c2 and crRNA at the molar ratio of 1:1.4 at room temperature for 30 min. Then, the complex was further purified on a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare) in buffer containing 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol and 1 mM TCEP. Samples used for crystallization were measured for final absorbance of ~10 using a Nanodrop spectrophotometer.

Crystallization, Data Collection and Structure Determination

Crystals of C2c2-crRNA complex were grown by hanging drop vapor diffusion method at 16°C. Crystals were obtained by mixing 1 μL complex solution and 1 μL of reservoir solution. Crystals of C2c2-crRNA complex were grown from 0.1 M Sodium malonate and 12% PEG 3350. All crystals were cryo-protected using corresponding reservoir buffers supplemented with 20% glycerol and flash frozen in liquid nitrogen.

Prior to setting trays for the apo state, the LshC2c2 protein was digested with thermolysin protease for 30 min at 25°C in a ratio of 2000:1 (weight: weight). The protease-resistant core was subjected to size exclusion chromatography on a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare) in buffer containing 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol and 1 mM TCEP. The complex fractions were collected and concentrated. Samples used for crystallization were measured for final absorbance of ~8 using a Nanodrop spectrophotometer. Crystals of the apo-C2c2 thermolysin-resistant core were grown by hanging drop vapor diffusion method at 16°C. Crystals were obtained by mixing 1 μL protein solution and 1 μL of reservoir solution. Crystals of the

apo-C2c2 thermolysin-resistant core were grown from 0.4 M L-Proline, 0.1 M HEPES pH 7.5 and 10% PEG 3350. All crystals were cryo-protected using corresponding reservoir buffers supplemented with 20% glycerol and flash frozen in liquid nitrogen.

All diffraction datasets were collected at beamline BL-19U1 at Shanghai Synchrotron Radiation Facility (SSRF), and processed with HKL2000 (Otwinowski and Minor, 1997). The phases of LshC2c2 were solved by Se single wavelength anomalous dispersion method using PHENIX Autosol (Adams et al., 2002). The model was manually built and adjusted using the program Coot (Emsley et al., 2010). Iterative cycles of crystallographic refinement were performed using PHENIX. All statistics of data processing and structure refinement of LshC2c2 are summarized in Table S1. The structure figures were prepared using PyMOL (http://www.pymol.org/).

Limited Proteolysis Assays

Thermolysin was used for limited proteolysis assays. Limited proteolysis assays were performed at 25°C in the proteolysis buffer (50 mM Tris-HCl, pH 7.5, 400 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP). 2 μM of C2c2, C2c2-crRNA (molar ratio, 1:1.4) and C2c2-crRNA-targetRNA (molar ratio, 1:1.4:1.8) were incubated with 0.25 ng/μl thermolysin for different time (0 min, 10 min, 30 min, 60 min). The reactions were stopped by adding 2 × SDS loading buffer and 95°C quenched for 5 min. Samples were analyzed on a 12% SDS polyacrylamide gel with Tris-Glycine buffer.

Target RNA: 5′-GGAAGAAGAGUUUAUUCAGAUAGAUUUGUC-3′.

Pre-crRNA Processing Assays

8 μM of pre-crRNA were incubated at 37°C for 3 min with 8 μM C2c2 in 10 μL reactions containing 50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP. Reactions were stopped by adding 2 × loading buffer and 75°C quenched for 5 min. Samples were analyzed on a 20% urea denaturing polyacrylamide gel with TBE buffer.

Target Cleavage Assays

Target cleavage assays were carried out using target RNA. 1 μM of crRNA were incubated at 25°C for 30 min with 1 μM C2c2 in cleavage buffer (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP). Then the resulting samples were incubated at 37°C for 1 hr with 1 μM target RNA in 10 μL reactions. Reactions were stopped by adding 2 × loading buffer and 75°C quenched for 5 min. Samples were analyzed on a 20% urea denaturing polyacrylamide gel with TBE buffer. The target RNA used in this experiment has the sequence: 5′-GGAAGAAGAGUUUAUUCAGAUAGAUUUGUCCUCGACCUUUUUAGGUCGAGUCGACGGAGUCUAGACUCCGUCCUGAUGAGUCCGUGAGGACGAAAGACUCGACCUGAUGUAUAGAGUAGA-3′.

QUANTIFICATION AND STATISTICAL ANALYSIS

Pre-crRNA processing and target RNA cleavage experiments were repeated three times, and representative results were shown.

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession numbers for the atomic coordinates and structure factors reported in this paper are PDB: 5WTJ and PDB: 5WTK (see Table S1).

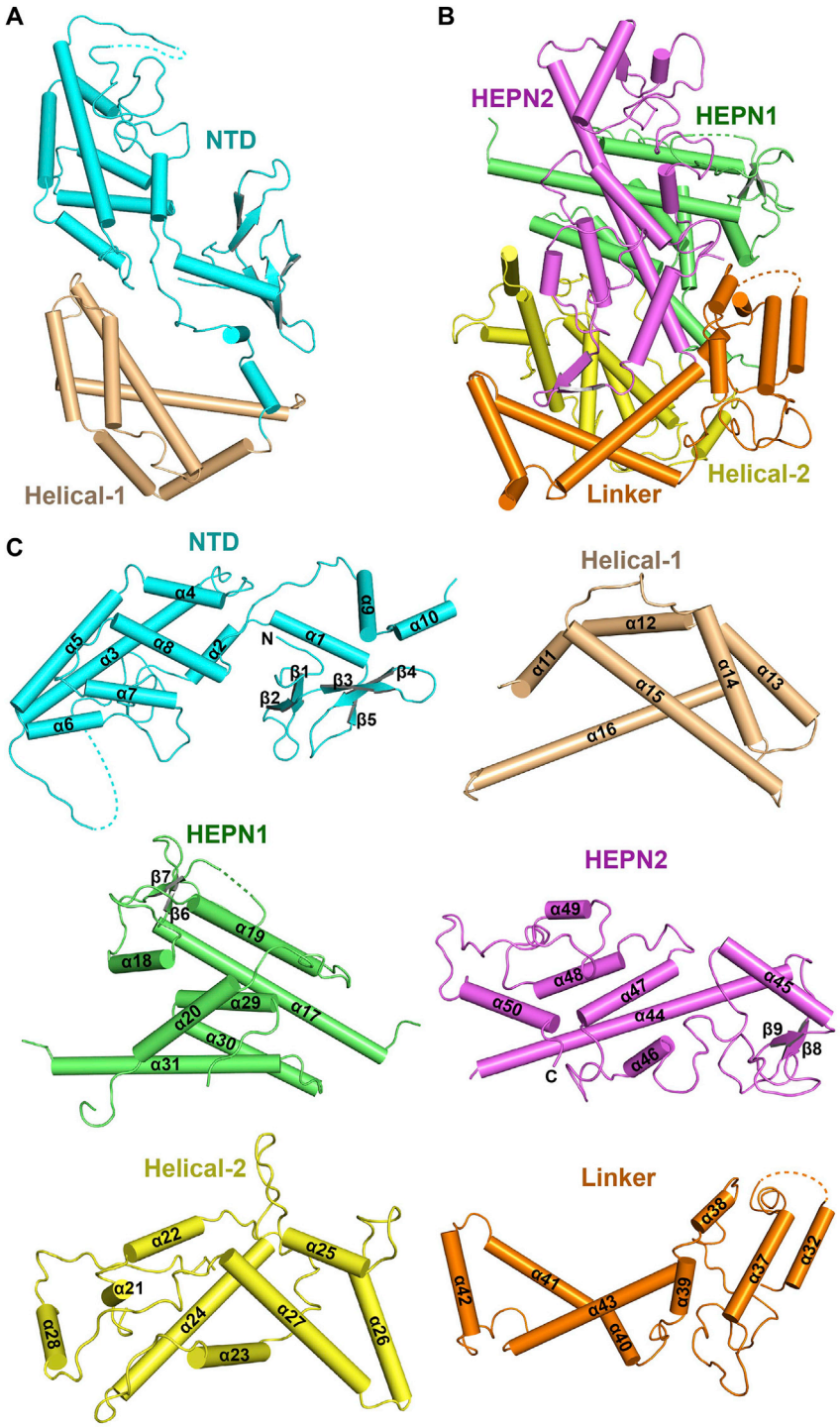
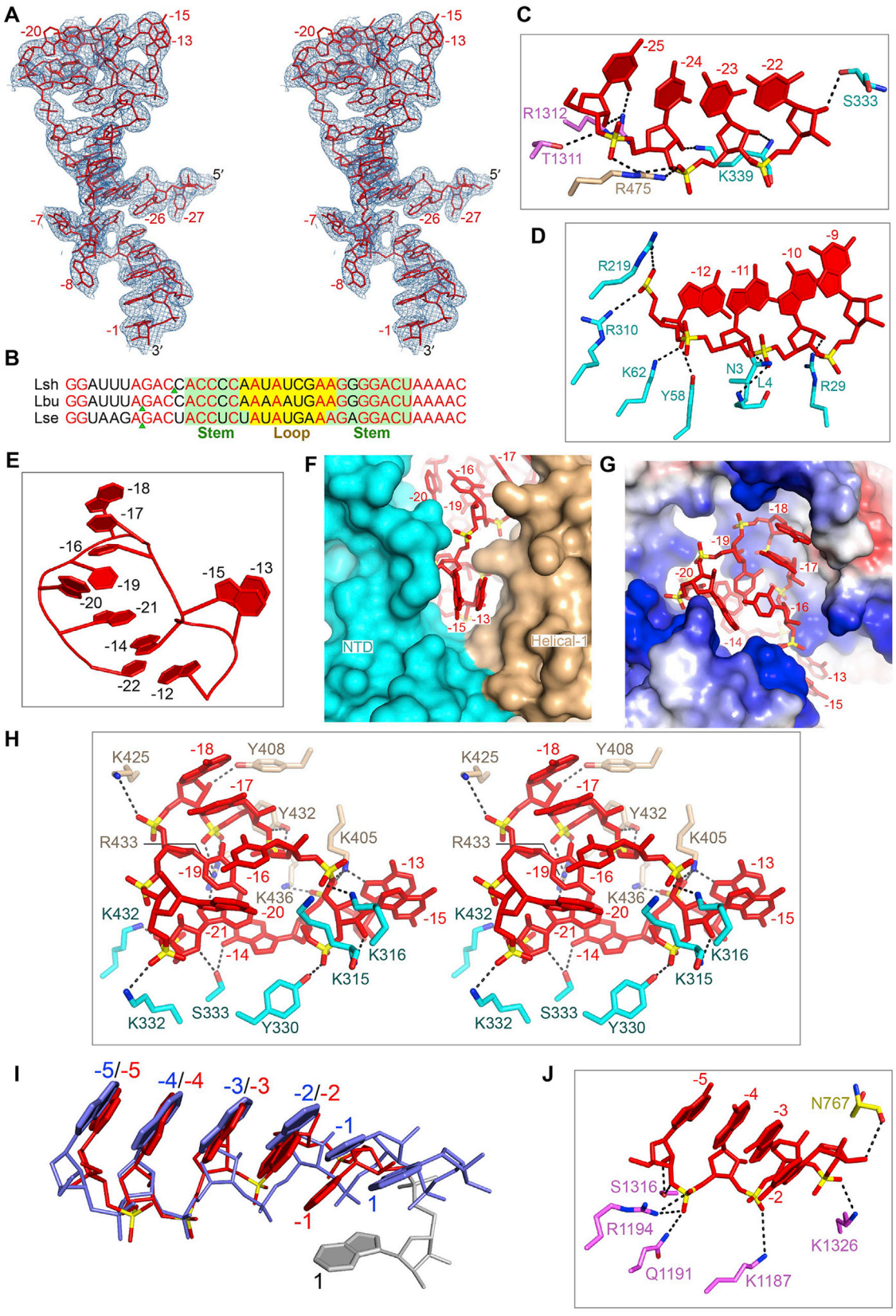


Figure S1. Crystal Structure of LshC2c2-crRNA Binary Complex, Related to Figure 1

(A and B) Structures of the REC lobe (A), and the NUC lobe (B).

(C) The structures all individual domains of LshC2c2 (color-coded as defined in Figure 1A).



(legend on next page)

Figure S2. The Structure of crRNA Repeat, Related to Figure 2

(A) Electron Density of LshC2c2-crRNA Complex. 2Fo-Fc omit map (blue color, contoured at 1.0 σ) of the crRNA repeat in the LshC2c2-crRNA binary complex. The crRNA is shown in stick representation.

(B) Sequence alignment of the crRNA repeats from Lsh, Lbu and Lse. Identical nucleotides are in red and non-conserved nucleotides are in black. The stem and the loop are highlighted by green and yellow background, respectively. The catalytic sites are indicated by green triangles.

(C) Detailed view of the interaction between the nucleotides (−25)-(−22) and LshC2c2.

(D) Detailed view of the interaction between the nucleotides (−12)-(−9) and LshC2c2.

(E) Structure of the 9-nt crRNA loop.

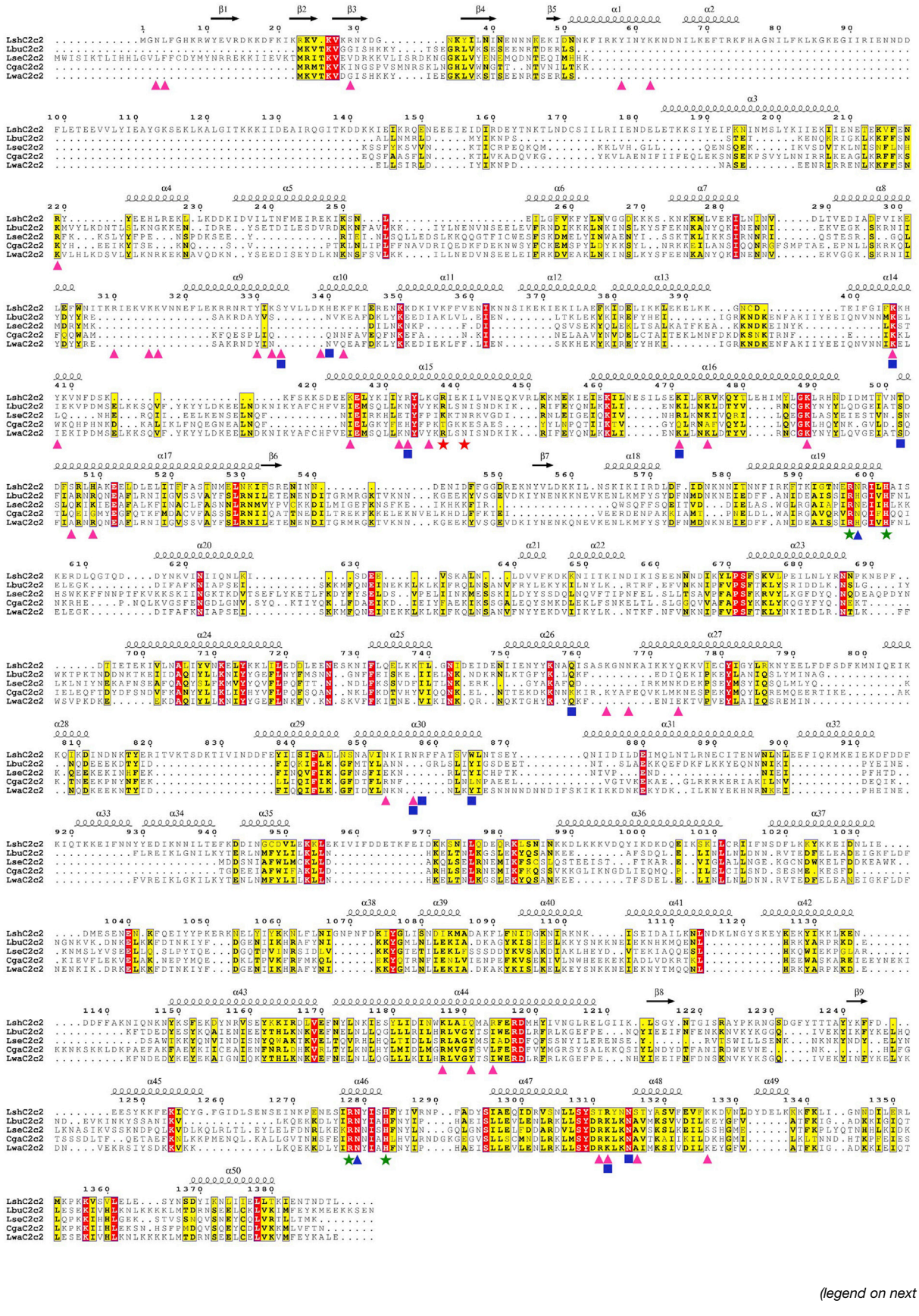
(F) Nucleotides (−15) and (−13) within the 9-nt crRNA loop insert into a groove formed between the NTD and the Helical-1 domains.

(G) The 9-nt crRNA loop binds in a positively charged groove.

(H) Stereo view of the interaction between the 9-nt crRNA loop and LshC2c2.

(I) Superposition of nucleotides (−5)-1 of the crRNA (red and gray) on a 6-nt A-form single stranded RNA (light blue) after best-fit superposition of the segment of nucleotides (−5)-(−2).

(J) Detailed view of the interaction between the nucleotides (−5)-(−2) and LshC2c2.



(legend on next page)

Figure S3. Sequence Alignment of C2c2 Proteins, Related to Figure 3

The *Leptotrichia shahii* C2c2 protein was aligned with its respective homologs. The aligned sequences (Genebank ID and designated gi) are in the order of LshC2c2 (*Leptotrichia shahii* C2c2, gi: 517262777), LbuC2c2 (*Leptotrichia buccalis* C2c2, gi: 506250229), LseC2c2 (*Listeria seeligeri* C2c2, gi: 502750493), CgaC2c2 (*Carnobacterium gallinarum* C2c2, gi: 736550717), and LwaC2c2 (*Leptotrichia wadei* C2c2, gi: 545623306). The secondary structure of *Leptotrichia shahii* C2c2 is shown on top. Conserved residues are shaded in yellow, whereas essentially invariant residues are shown in red. The pre-crRNA catalytic residues are marked by red stars, target RNA catalytic residues are marked by green stars, residues involved in crRNA phosphate binding are shown by magenta triangles, and residues involved in sequence specific interaction with crRNA are shown in blue squares. Conserved residues (Asn598 and Asn1279) of R-X-H motifs in the HEPN1 and HEPN2 domains are shown in blue triangles.

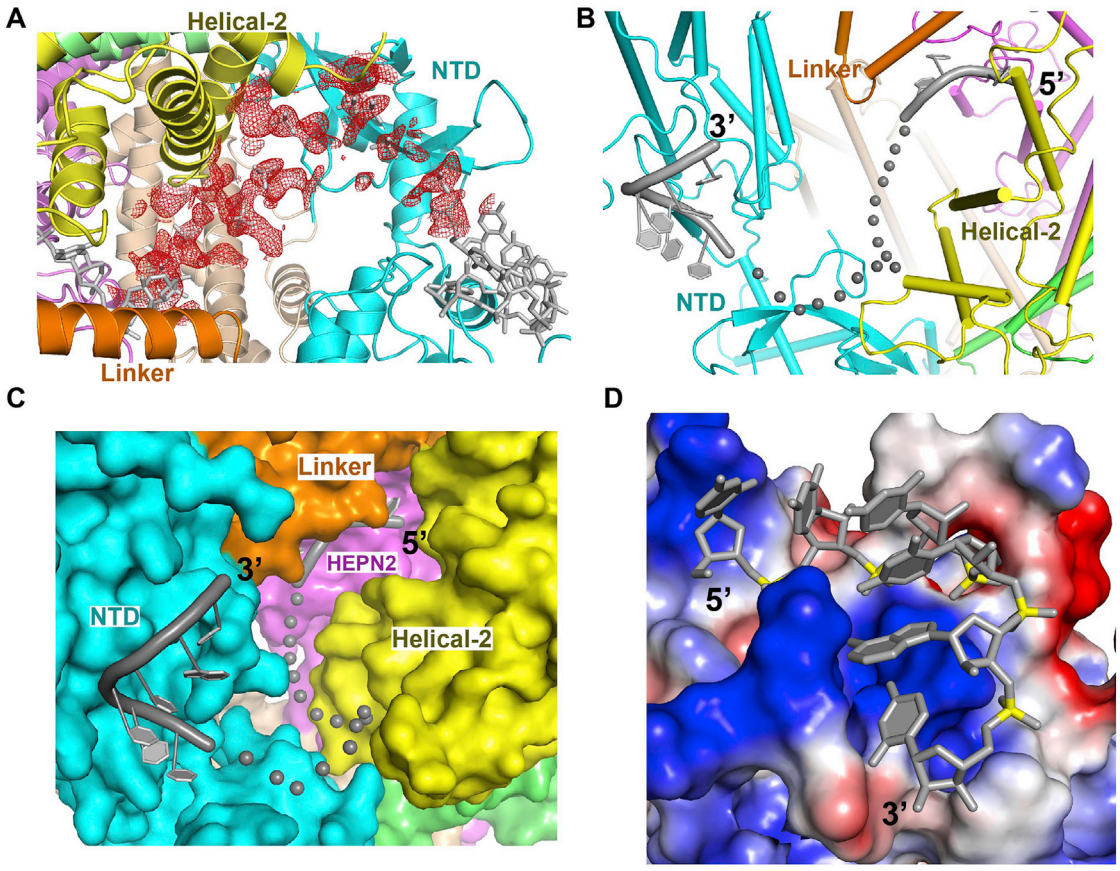


Figure S4. crRNA Guide Segment, Related to Figure 4

(A) The discontinuous electron density is shown in red, suggestive of the binding channel for the central spacer. The gray phosphate groups inside the electron density map indicate the binding position of the nucleotides within the central spacer.

(B) The central spacer binds in the open groove formed by the Linker, the NTD and the Helical-2 domains.

(C) Positioning of the central spacer within a groove formed by the Linker, the NTD and the Helical-2 domains. C2c2 is shown in surface.

(D) Detailed view of the positioning of the 3′ end guide segment within an external groove of the NTD domain.

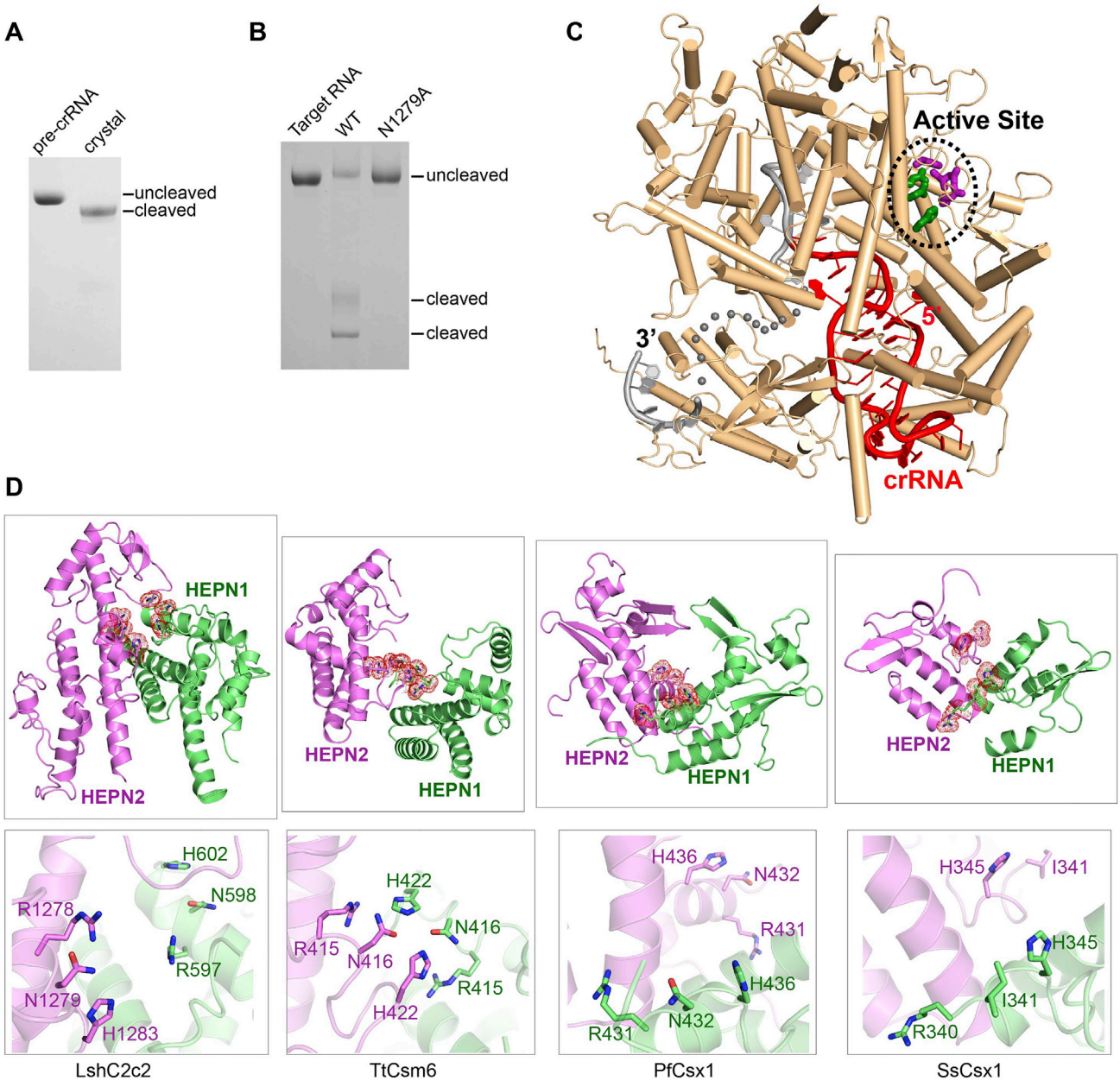


Figure S5. The Catalytic Sites of LshC2c2, Related to Figure 5

(A) Analysis using a denaturing gel of the RNA in the crystal indicates that the pre-crRNA has been cleaved during crystallization.

(B) The crRNA-guided RNA cleavage assay using an Asn1279Ala mutant.

(C) The catalytic pocket within two HEPN domains is located on the outer surface of the LshC2c2. The catalytic residues from HEPN1 and HEPN2 are shown by green and magenta sticks. For clarity, the entire LshC2c2 is shown in wheat. The active site located on the HPEN domains is highlighted by black dashed circle.

(D) Comparison of the HEPN domains from LshC2c2, TtCsm6, PfCsx1 and SsCsx1.

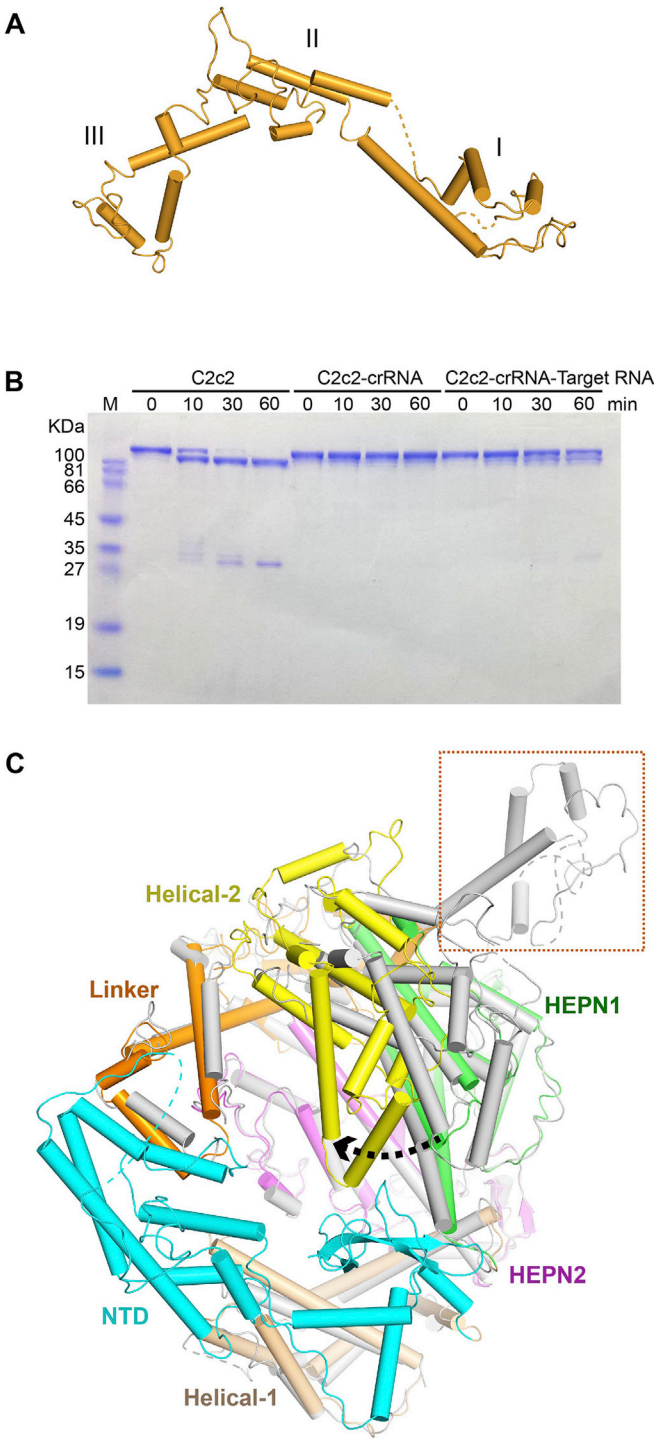


Figure S6. The Crystal Structure of LshC2c2 in its apo State, Related to Figure 6

(A) The structure of the Linker of C2c2 in the apo state.

(B) Limited proteolysis of LshC2c2 with and without RNA substrates by thermolysin to test for large-scale conformational changes of LshC2c2 upon cRNA binding and target RNA recognition.

(C) Superposition of the LshC2c2-crRNA complex onto the LshC2c2 in the absence of crRNA shows that C2c2 changes its conformation upon crRNA binding.

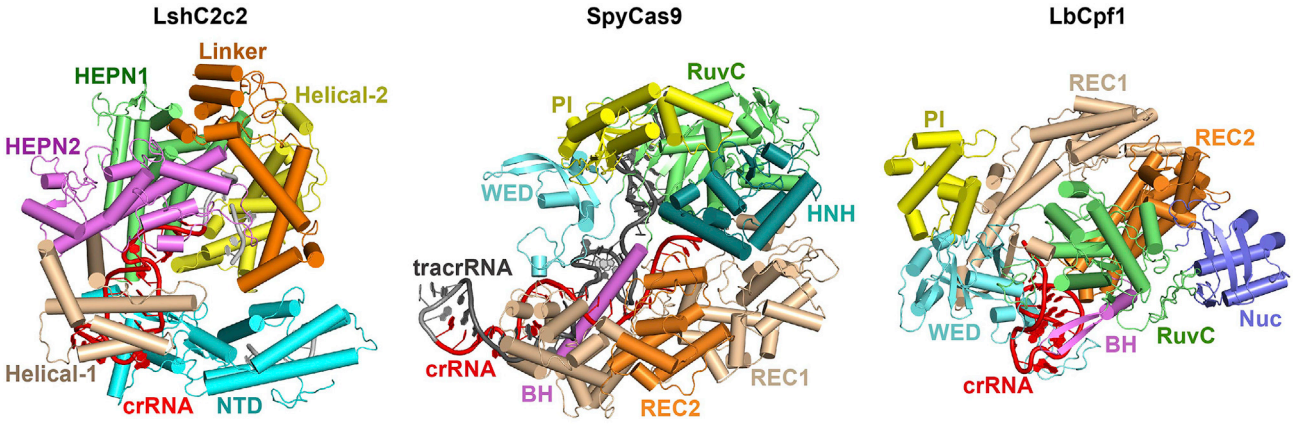


Figure S7. Comparison of C2c2, Cas9, and Cpf1, Related to Figure 7

Comparison of the overall structures of LshC2c2, SpyCas9 (PDB ID: 4ZT0) and LbCpf1 (PDB ID: 5ID6).