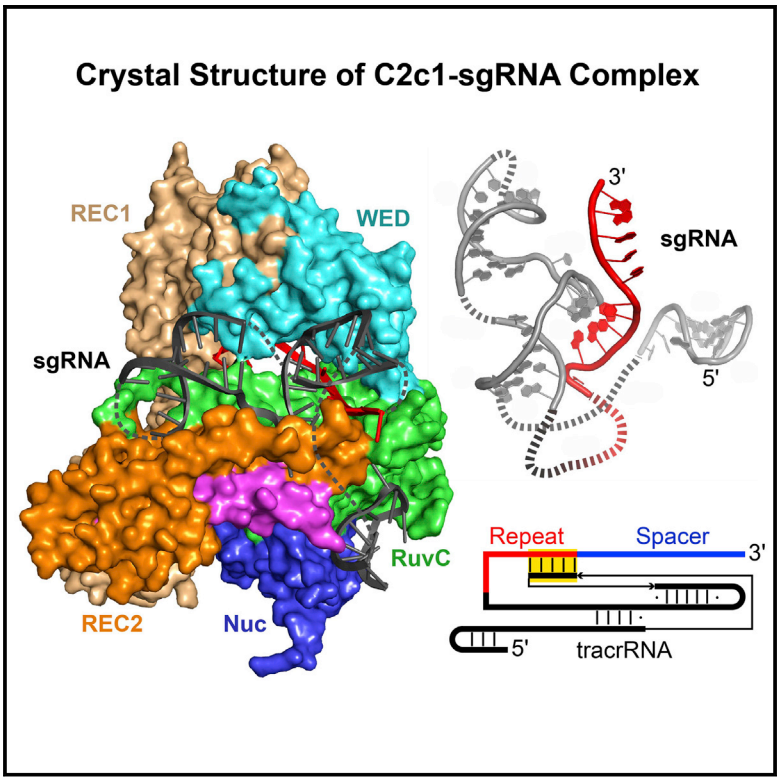
Article

Molecular Cell

C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism

Graphical Abstract



Authors

Liang Liu, Peng Chen, Min Wang, Xueyan Li, Jiuyu Wang, Maolu Yin, Yanli Wang

Correspondence

ylwang@ibp.ac.cn

In Brief

Liu et al. report the structure of C2c1-a type V-B CRISPR-Cas endonuclease, in complex with a chimeric single guide RNA, providing insights into the high specificity of DNA cleavage by C2c1. C2c1's low off-target rate makes it a valuable addition to the current arsenal of gene-editing tools.

Highlights

- Crystal structure of *Alicyclobacillus acidoterrestris* C2c1-sgRNA complex
- Mechanistic insights into dual-RNA-guided, C2c1-catalyzed, staggered dsDNA breaks
- High mismatch sensitivity of C2c1 reduces off-target cleavage
- Join 2/4-truncated sgRNA-guided DNA cleavage similar to that of wild-type sgRNA

Accession Numbers

5WQE

Liu et al., 2017, Molecular Cell *65*, 1–13
January 19, 2017
http://dx.doi.org/10.1016/j.molcel.2016.11.040

CellPress

Please cite this article in press as: Liu et al., C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.11.040

Molecular Cell
Article

C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism

Liang Liu,[1] Peng Chen,[1,2] Min Wang,[1] Xueyan Li,[1,2] Jiuyu Wang,[1] Maolu Yin,[1,2] 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.molcel.2016.11.040

SUMMARY

C2c1 is a type V-B CRISPR-Cas system dual-RNA-guided DNA endonuclease. Here, we report the crystal structure of *Alicyclobacillus acidoterrestris* C2c1 in complex with a chimeric single-molecule guide RNA (sgRNA). AacC2c1 exhibits a bi-lobed architecture consisting of a REC and NUC lobe. The sgRNA scaffold forms a tetra-helical structure, distinct from previous predictions. The crRNA is located in the central channel of C2c1, and the tracrRNA resides in an external surface groove. Although AacC2c1 lacks a PAM-interacting domain, our analysis revealed that the PAM duplex has a similar binding position found in Cpf1. Importantly, C2c1-sgRNA system is highly sensitive to single-nucleotide mismatches between guide RNA and target DNA. The resulting reduction in off-target cleavage renders C2c1 a valuable addition to the current arsenal of genome-editing tools. Together, our findings indicate that sgRNA assembly is achieved through a mechanism distinct from that reported previously for Cas9 or Cpf1 endonucleases.

INTRODUCTION

The bacterial CRISPR elements, together with their CRISPR-associated (Cas) proteins, form a prokaryotic defense system against invasive DNA elements (Barrangou et al., 2007; Barrangou and Marraffini, 2014; Marraffini, 2015). CRISPR-Cas systems establish immunity against bacteriophages and plasmids in a common three-step procedure, namely adaptation, CRISPR RNA (crRNA) biogenesis, and interference (van der Oost et al., 2014). The Cas proteins involved in the crRNA interference process, however, display great diversity, resulting in a classification system of CRISPR-Cas complexes consisting of classes and types, based on the traits of the Cas proteins involved (Makarova et al., 2011, 2015). Class 1 CRISPR-Cas systems, encompassing type I, type III, and the putative type IV systems, form multi-subunit effector complexes, where the crRNA is associated with a wide range of Cas proteins. In comparison, class 2 systems, encompassing types II, V, and VI CRISPR-Cas systems, are characterized by the fact that the crRNA binds to a single Cas protein.

Cas9, a class 2 type II protein, is the best-studied CRISPR-Cas endonuclease and has been explored as a genome-editing tool based on its RNA-guided endonuclease activity (Doudna and Sontheimer, 2014; Gasiunas et al., 2012; Jinek et al., 2012). Cas9 adopts a bi-lobed architecture comprised of recognition (REC) and nuclease (NUC) lobes, in which the guide RNA-target DNA duplex is bound within a positively charged central channel formed between the two lobes (Jinek et al., 2014; Nishimasu et al., 2014). The NUC lobe contains two nuclease domains: a RuvC-like domain and an HNH domain. To cleave double-stranded DNA, Cas9 associates with a dual-RNA-guide structure consisting of a crRNA and a *trans*-activating CRISPR RNA (tracrRNA). DNA cleavage by Cas9 is dependent on the presence of a protospacer adjacent motif (PAM) in the target DNA, restricting the choice of targetable sequences (Anders et al., 2014). Thus, Cas9 variants or new endonucleases, which recognize different PAM sequences, are required to address this inherent PAM sequence limitation and expand the spectrum of targetable genomic sites.

The recently identified class 2 type V CRISPR-Cas system has been further divided into three subtypes based on the type of Cas proteins present in the system. Cpf1, a V-A CRISPR endonuclease, is an RNA-guided protein that cleaves target DNA via a staggered cut, creating overhangs rather than a blunt end (Zetsche et al., 2015). Cpf1 has a bi-lobed architecture consisting of an α-REC lobe and a NUC lobe (Dong et al., 2016; Yamano et al., 2016). The α-REC lobe contains two REC domains at the N-terminal, and the NUC lobe consists of RuvC, wedge (WED), Nuc, and PAM-interacting (PI) domains. Cpf1 contains a RuvC domain, but not an HNH domain, suggesting that Cpf1 has a distinct molecular mechanism from Cas9. Recent structural studies showed that *Lachnospiraceae bacterium* Cpf1 bound to crRNA adopts a compact triangular shape and bi-lobed architecture. In the ternary complex of *Acidaminococcus sp.* Cpf1 bound to crRNA-DNA, the crRNA-target duplex binds at a central positively charged channel that is formed between the REC and NUC lobes.

C2c1 is a newly identified type V-B CRISPR-Cas system endonuclease, which may have potential in genome editing due to its

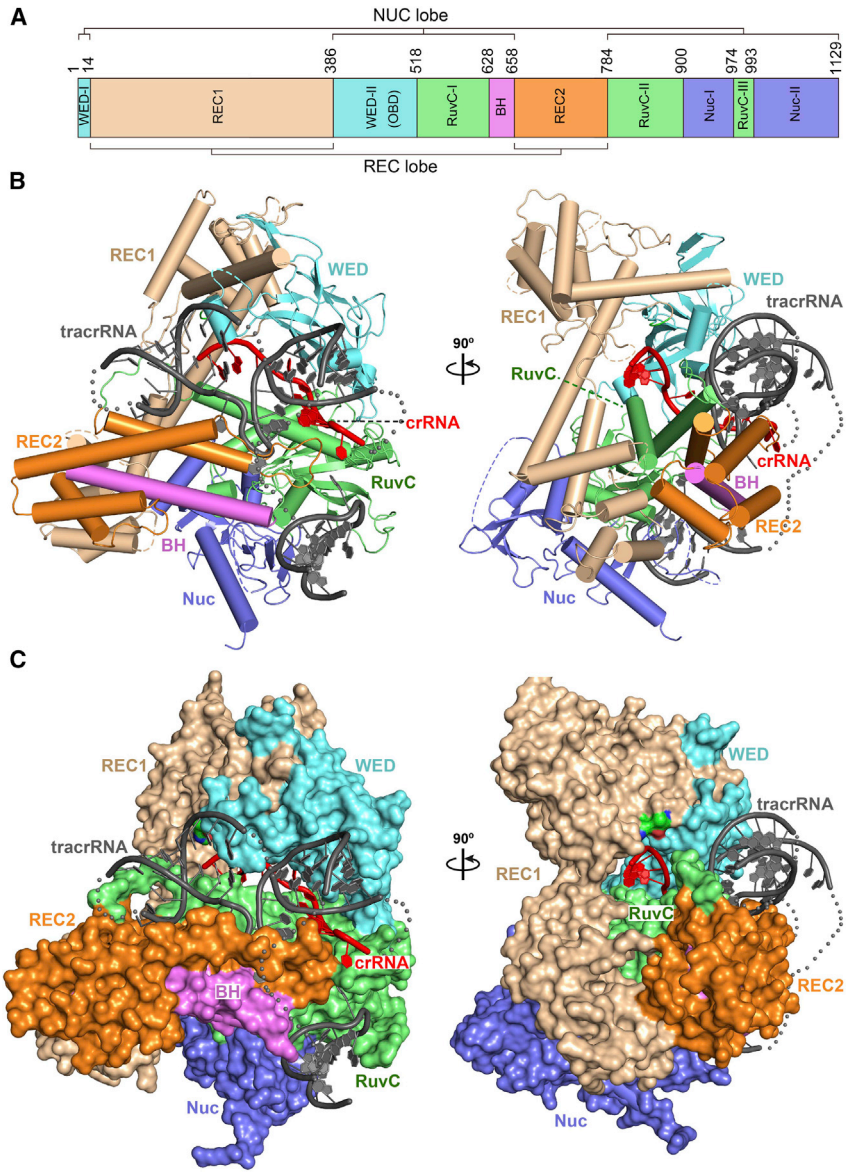


Figure 1. Overall Structure of the AacC2c1-sgRNA Complex

(A) Domain organization of the AacC2c1 protein. (B and C) Orthogonal views of the crystal structure of AacC2c1 bound to sgRNA in cartoon (B) and surface (C) presentation. The AacC2c1 domain color code is identical to that in (A). The crRNA segment is shown in red and the tracrRNA segment in black. See also Figure S1.

double-stranded DNA cleavage activity. In contrast to Cpf1, which is a single-RNA-mediated DNA endonuclease, C2c1 is a dual-RNA-guided DNA endonuclease (Shmakov et al., 2015). Recent biochemical studies have shown that either a crRNA in combination with a minimal 78 nt tracrRNA or a fused sgRNA is sufficient for C2c1-mediated DNA cleavage. A 14 nt direct repeat (DR) hybridizes with tracrRNA to form a crRNA:tracrRNA duplex, which is then loaded onto C2c1 to guide DNA recognition and cleavage. C2c1 recognizes the T-rich PAM at the 5′ end of the protospacer sequence to mediate DNA interference.

The role of C2c1 in bacterial immunity and its potential genome engineering applications rely on accurate DNA target selection. The target choice requires base-pairing between guide RNA and target DNA, as well as the presence of a PAM proximal to the target site. Thus, the binding of crRNA and tracrRNA to C2c1 is a pre-requisite for base-pairing between the guide and target strand. However, how C2c1 assembles the crRNA and tracrRNA or, alternatively, a chimeric sgRNA, is unclear, and how C2c1 cleaves its DNA target remains to be investigated. To this end, we solved the structure of C2c1 in complex with its sgRNA. Here, we provide a detailed description of the crystal structure of an AacC2c1-sgRNA binary complex, representing the pre-target bound state of the enzyme. We find that the AacC2c1 protein adopts a bi-lobed architecture that accommodates the guide region of the sgRNA in its central channel, with the tracrRNA bound at the surface. The present structure revealed unexpected structural divergence from the other two class 2 effectors Cas9 and Cpf1.

RESULTS

Overall Structure of the AacC2c1-sgRNA Complex

To determine how the C2c1 protein assembles with and positions the guide RNA prior to substrate recognition, we solved a 3.1 Å resolution crystal structure of AacC2c1 bound to a sgRNA (Figures 1A–1C). The complex structure was determined using the single-wavelength anomalous dispersion (SAD) method and a seleno-methionine (SeMet)-labeled protein (Table 1). Our crystal structure (Figure 1B) shows that AacC2c1 is a monomer and adopts a bi-lobed architecture consisting of an α-helical recognition lobe and a nuclease lobe. The REC lobe consists of REC1 and REC2 domains, and the NUC lobe consists of WED, RuvC, and Nuc domains. A Dali server search (Holm and Rosenström, 2010) identified Cpf1 as the protein with the most similar structure (*Z* score: 9.2; root-mean-square deviation [RMSD]: 4.8 Å over 405 amino acids, largely from the RuvC domain). This is consistent with the low sequence similarity outside the RuvC domain of C2c1 and Cpf1.

The REC1 domain consists of 11 α helices adopting a dumbbell configuration, in which two α-helical bundles, each containing five α helices, are connected by a long α helix (α8) (Figure S1A). Notably, the structural features of C2c1 REC1 domain differ significantly from that of Cpf1 and Cas9. Amino acids

Table 1. Crystallographic Data Collection and Refinement Statistics for Se_AacC2c1-sgRNA

Data Collection[a]	
Space group	*C*2
Cell Dimensions	
a, *b*, *c*	194.55 Å, 129.84 Å, 84.00 Å
α, β, γ	90.00°, 110.00°, 90.00°
Resolution	50.00–3.10 Å (3.15–3.10 Å)
R_{merge}	0.135 (0.658)
$I/\sigma I$	14.3 (2.2)
Completeness	99.6% (99.6%)
Redundancy	9.4 (6.5)
Refinement	
Resolution	3.1 Å
No. reflections	27,598
R_{work} / R_{free}	0.235/0.262
No. Atoms	
Protein	7,953
Nucleic Acid	1,277
Solvent	42
β Factors	
Protein	56.0 Å^2
Nucleic Acid	83.5 Å^2
Solvent	26.5 Å^2
R.m.s. Deviations	
Bond length	0.013 Å
Bond angles	1.876°
Ramachandran Plot	
Favored region	96.99
Allowed region	2.21
Outlier region	0.80

[a]Highest resolution shell is shown in parentheses.

135–173 in the REC1 domain, which are predicted to form two α helices, were disordered in our AacC2c1-sgRNA complex. The REC2 domain consists of a five α (α17–α21) helical bundle containing an Arg-rich bridge helix (α17) (shown in magenta). This helix serves as a connecting hinge in Cas9 and Cpf1 but does not appear to be positioned like a bridge in the C2c1 structure.

The AacC2c1 WED domain is assembled from two separate regions (WED-I and WED-II) in the C2c1 sequence (Figure 1A). It contains a core subdomain, which adopts an oligonucleotide-binding (OBD) fold. The core subdomain consists of eight β stranded distorted β sheets. The core subdomain is flanked by two α helices and one β-hairpin (Figure 1B; Figure S1B). These structural features suggest that, in C2c1, this domain fulfills a similar function as that in Cas9 and Cpf1.

The RuvC domain of C2c1 forms the structural core of the nuclease lobe and is located between the WED and Nuc domains. It comprises three RuvC-motifs (RuvC-I, RuvC-II, and RuvC-III) (Figure 1A) and contains a six-stranded β sheet surrounded by four α helices (Figure S1B). Apart from the WED and RuvC domains, the nuclease lobe also contains a putative Nuc domain, which is located at the C-terminal of the C2c1 sequence and consists of two Nuc regions separated by the RuvC-III motif (Figure 1A). The Nuc-I region consists of a central β-barrel formed from five β strands enclosed by three α helices. The Nuc-II region contains a three β stranded sheet and one α helix (Figure 1B; Figure S1B). In our structure, residues 916–947, 1,044–1,053, and 1,116–1,129 are disordered, indicating that the Nuc domain of AacC2c1 is flexible in the sgRNA-bound state.

Structure of the sgRNA

For crystallization experiments, we used a chimeric 111 nt sgRNA, a fusion product of a 35 nt crRNA (in red) and a 76 nt tracrRNA (in black) (Figure 2A). The crRNA segment contains a 20 nt guide segment (G92–A111) and a 15 nt 5′ DR (U77–C91). As shown in Figure S2A, we observed unambiguous electron density for 6 nt of 5′ DR, 5 nt of the guide region, and 48 nt of the tracrRNA. The global architecture of the sgRNA in the AacC2c1-crRNA binary complex comprises four helical stems (named P1–P4), connected by joining regions J1/2, J2/3, J3/4, and J2/4 with loops L1 and L3 (Figure 2B). As shown in Figures 2C, S2B, and S2C, the DR-tracrRNA joining region adopts a pseudoknot structure rather than a simple stem-loop structure as predicted from its nucleotide sequence. This suggests that the interaction between C2c1 and the sgRNA influences the conformation of the sgRNA, which is likely to be of functional consequence.

According to previous reports using sequence-based structure prediction, the 5′ end of the DR (U79–U86) is paired with the 3′ end of the tracrRNA (A74–G67) forming a stem-loop containing eight Watson-Crick base pairs, while the other region of the tracrRNA forms four helices (Figure 2A) (Shmakov et al., 2015). Unexpectedly, base-pairing between the 5′ DR and the tracrRNA in our crystal lattice (Figure 2B) was different to that predicted earlier (Figure 2A). Instead of base-pairing between nucleotides U79–U86 of the 5′ DR and A74–G67 of the tracrRNA nucleotide as predicted previously (Figure 2A), we observed base-pairing between nucleotides G87–C91 of the 5′ DR and tracrRNA nucleotides C33–G29, forming a repeat:anti-repeat duplex (Figure 2B). The resulting sgRNA scaffold suggests that nucleotides C33–G29 represent an anti-repeat region. However, nucleotides U30–C33 of the anti-repeat region were previously predicted to form a duplex with nucleotides A55–G52 of the tracrRNA (Figure 2A) (Shmakov et al., 2015). Instead, the nucleotides A55–G52 form a stem-loop (P3/L3) with nucleotides G37–U43 in our structure (highlighted with a green background in Figure 2B). Helix P3 of the stem comprises five canonical Watson-Crick base pairs flanked by G–A and U–G mismatches at each end. Nucleotides G58–G62 form helix P2 by base-pairing with U28–C24. The 5′ end of the tracrRNA forms the stem-loop P1/L1 (Figures 2B and 2C). The joining regions J1/2, J2/3, and J2/4 and partial L3 are disordered in our structure.

To test whether this secondary sgRNA structure also occurs in other type V-B systems, we analyzed the structures of crRNA and tracrRNA from *Bacillus thermoamylovorans* (Bth). The in silico co-folding of the Bth crRNA and putative tracrRNA (Figure S2D) exhibits a stable secondary structure and complementarity

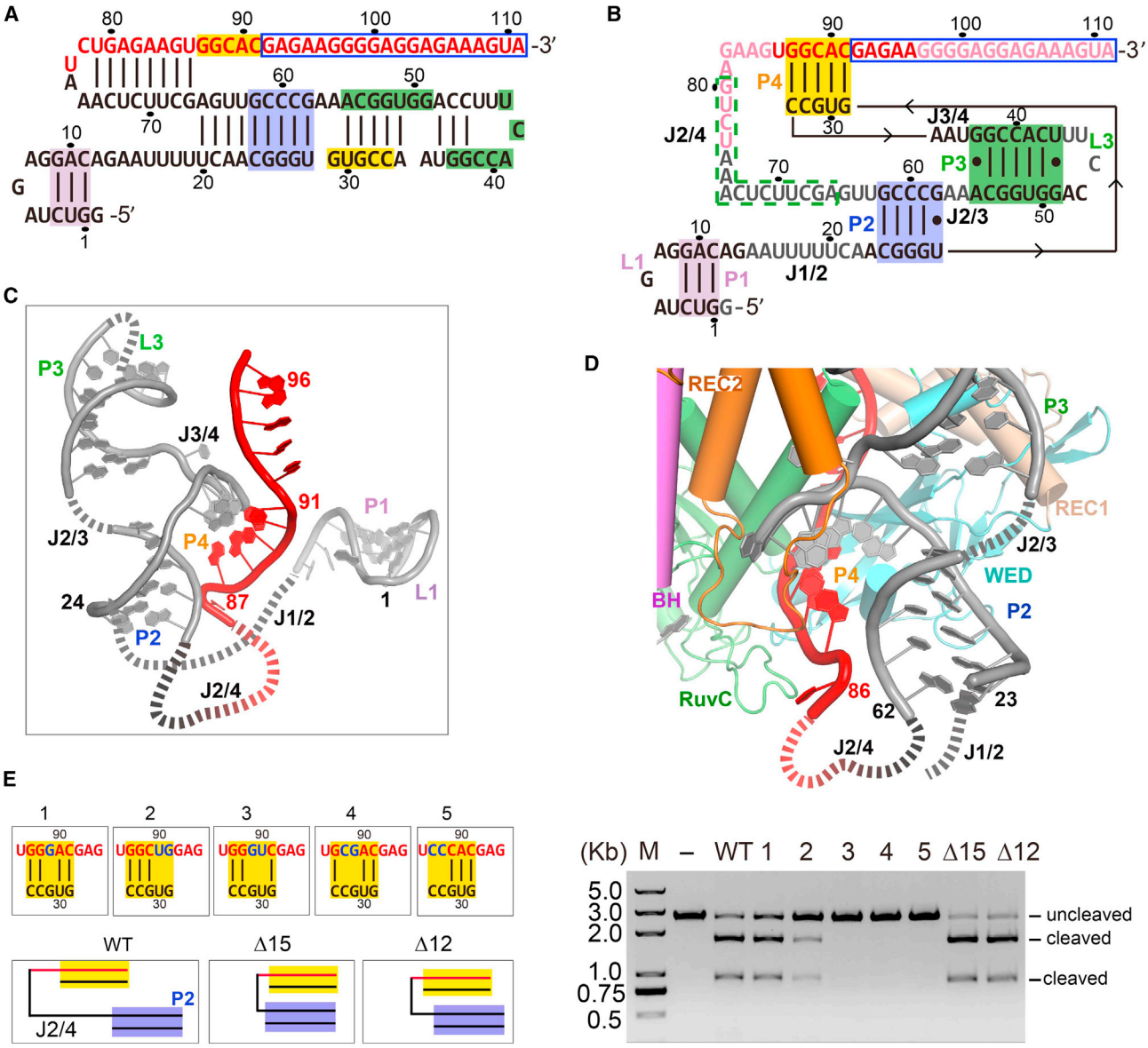


Figure 2. sgRNA Sequence and Structure

(A) Schematic presentation of the sgRNA used for crystallization. The secondary structure of sgRNA is predicated based on its sequence as previously reported. The tracrRNA segment is shown in black and the guide segment in red. The spacer region is outlined by a blue box. Nucleotides forming a duplex in the AacC2c1-sgRNA crystal structure are highlighted using backgrounds of identical color.

(B) Crystal structure-based schematic presentation of the sgRNA fold. The disordered nucleotides in the tracrRNA are shown in gray, the crRNA segments in light red. Nucleotides involved in forming the repeat:anti-repeat duplex P4 are shown with a yellow background, and nucleotides in helices P3, P2, and P1 are shown with green, light blue, and pink backgrounds, respectively.

(C) The structure of sgRNA when bound to the AacC2c1 protein.

(D) Magnified view of helix P2 and P4. Joining region 2/4 is disordered in the present structure and shown by a dashed line.

(E) AacC2c1-mediated DNA cleavage with mutant or truncated sgRNAs. –, uncleaved linear plasmid DNA with target DNA insertion; WT, wild-type sgRNA. The mutant nucleotides within sgRNA (lanes 1–5) are highlighted in blue in the left panel. Δ15: 15 nt deleted from the sgRNA J2/4 region (highlighted in the green dashed line box in B). Δ12: 12 nt have been removed from the sgRNA J2/4 region (in the green box in B).

See also Figure S2.

between the two RNAs similar to that found for AacC2c1 (Shmakov et al., 2015). Based on the secondary structure observed in our AacC2c1-sgRNA complex structure, we hypothesize that the repeat region immediately downstream of the spacer is complementary to the region highlighted in yellow of the tracrRNA in Figure S2D and therefore forms a scaffold structure similar to that found in our AacC2c1-sgRNA complex structure (Figure S2E). This suggests that this unusual AacC2c1-sgRNA scaffold may be

present in C2c1 proteins from different species. This insight should facilitate bioinformatics searches for tracrRNA sequences in other species.

The lack of electron density in our structure at the joining J2/4 region (nucleotides U63–G85) indicates that the spacer-distal end DR and the 3′ end of the tracrRNA in the AacC2c1-sgRNA complex are highly mobile. It was previously predicted that this region represents the crRNA repeat:tracrRNA anti-repeat duplex (Shmakov et al., 2015). However, our analysis of the AacC2c1-sgRNA complex indicates that this region forms a single-strand loop instead. Given the positions of nucleotides U86 and G62, it seems reasonable to speculate that the J2/4 region extends beyond the protein and does not interact directly with C2c1 (Figure 2D). On the basis of this finding, we hypothesized that the J2/4 region is of little importance for C2c1-mediated DNA cleavage. To test our hypothesis, we analyzed C2c1 DNA cleavage activity using truncated sgRNA and a plasmid DNA substrate. The sgRNA containing either 12 nt or 15 nt deletions within the J2/4 region exhibited an activity similar to that of wild-type sgRNA (Figure 2E). This finding is important for the development of improved RNA guides in combination with C2c1 for use in genome editing.

To further explore the importance of the crRNA repeat:tracrRNA anti-repeat duplex P4 for AacC2c1-meditated DNA cleavage, we conducted endonuclease activity assays using mutant guide RNAs. Single nucleotide substitutions within the region G87–C91 lacked any visible effects on the cleavage activity of C2c1 (Figure 2E; Figure S2F). In contrast, when nucleotides A90–C91 were replaced by U90–G91 to break the last 2 bp in helix P4, DNA cleavage activity was reduced. Notably, cleavage activity was completely abolished when a double mutation was introduced within the G87–A90 region (Figure 2E). Together, these observations suggest that proper base-pairing between the 5′ DR directly upstream of the spacer region and the anti-repeat region within the tracrRNA is essential for C2c1-mediated DNA cleavage and further confirmed that the specific structure of sgRNA found in our structure is functionally important.

Recognition of the crRNA repeat:tracrRNA Anti-repeat Duplex

The repeat:anti-repeat duplex P4, containing the nucleotides 87–91 within the repeat and nucleotides 29–33 within the tracrRNA, is located at the groove between the WED and RuvC domains (Figure 3A). The bound crRNA is continuously stacked from residues 87 to 91, with distinct breaks at the 86–87 and 91–92 steps (Figure S3A). The first nucleotide (G87) of helix P4 stacks onto the side chain of R746 (Figure 3B), resulting in the 5′ DR kinks at the G87 phosphodiester linkage. The last base pair (C91–G29) stacks onto the side chain of Q446, resulting in a kink at the junction between the spacer and repeat. These two stacking interactions also prevent further extension of the duplex P4.

The crRNA repeat:tracrRNA anti-repeat duplex is recognized in a sequence-specific manner and is further stabilized through extensive intermolecular interactions with the RuvC and WED domains (Figure 4). Our structure shows that the C89–G31 base pair in the 5′ handle is read out in the minor groove by the RuvC domain in a sequence-specific manner. The bases of C89 and G31 form hydrogen bonds with the side chains of K811 and D807, respectively (Figure 3C). The backbone phosphate groups between G87 and C91 within the repeat are mainly stabilized by the side chains of Y825, W835, D814, and Q882 within the RuvC domain via hydrogen bonds (Figure 3D). The complementary strand (nucleotides G29–C33) of the anti-repeat region forms an extensive network of interactions with the RuvC, WED, and REC2 domains. The side chains of S442, K9, Y501, and R484 of the WED domain and H800 and H803 of the RuvC domain form hydrogen bonds with the phosphate groups of nucleotides G29–C33 (Figure 3E). Together, these interactions appear to be important for overall stabilization of the RNA molecule in the C2c1 protein.

The Guide Segment Is Pre-organized Prior to Target Binding

In the AacC2c1-sgRNA structure, the nucleotides G92–A96 at the 5′ end of guide RNA segment are ordered and are positioned in the positively charged central channel formed by the REC and NUC lobes (Figure 3F). Nucleotides G92–A96 maintain a nearly A-form conformation along the ribose-phosphate backbone. To maintain this helical configuration, the WED and RuvC domains form extensive hydrogen-bonding interactions in a sequence-independent manner. The phosphate groups are stabilized by hydrogen bonding with the side chains of K4, S5, K7, N503, W391, and R485 of the WED domain and the side chains of K810 and N881 of the RuvC domain (Figure 3G). The O2′ atoms of G92 and A93 form hydrogen bonds with S5 and N503. In addition, the side chain of Q482 forms a van der Waals interaction with nucleotide G94. Notably, substitution of W391 by Ala significantly reduces the cleavage activity of C2c1, and Q482A and R485A mutants also exhibited reduced activity (Figure S3B). Together, these findings highlight the functional importance of the interaction between the duplex and the RuvC domain.

Given that the nucleotides (in the guide region) proximal to the repeat act as the seed region for nucleation with the target DNA in Cas9 and Cpf1 (Cong et al., 2013; Hsu et al., 2013; Jinek et al., 2012; Nishimasu et al., 2014; Yamano et al., 2016), we hypothesized that in C2c1, the 5 nt (92–96) immediately downstream of the repeat region are likely to also be part of the seed region. To test whether complementarity between nucleotides 92–96 and the relative DNA target is essential for C2c1 endonuclease activity, we conducted cleavage assays using a mutant DNA target. Mutation of any two consecutive nucleotides within the region complementary to nucleotides 92–96 on the target completely abolished the DNA cleavage of AacC2c1 (Figure S3C). This observation confirmed that nucleotides 92–96 function as the seed region. Additionally, nucleotides 92–96 of the guide (directly downstream of the 5′ handle) are facing the bulk solvent. This enables those nucleotides to be accessible for functional interaction with the target and further suggests that this region is the seed region. To test this hypothesis, we performed single-point mutation cleavage assay analysis with all 20 nt representing the entire spacer length. Our cleavage assay showed that mismatches at all positions affect cleavage activity, with the 18 nt proximal to the PAM sequence nearly completely abolishing cleavage, while the 2 nt at the 3′ end (distal the PAM sequence) significantly reduced cleavage activity (Figure 3H). This result suggests that the 5′ end of the guide RNA is the seed region, with a considerable

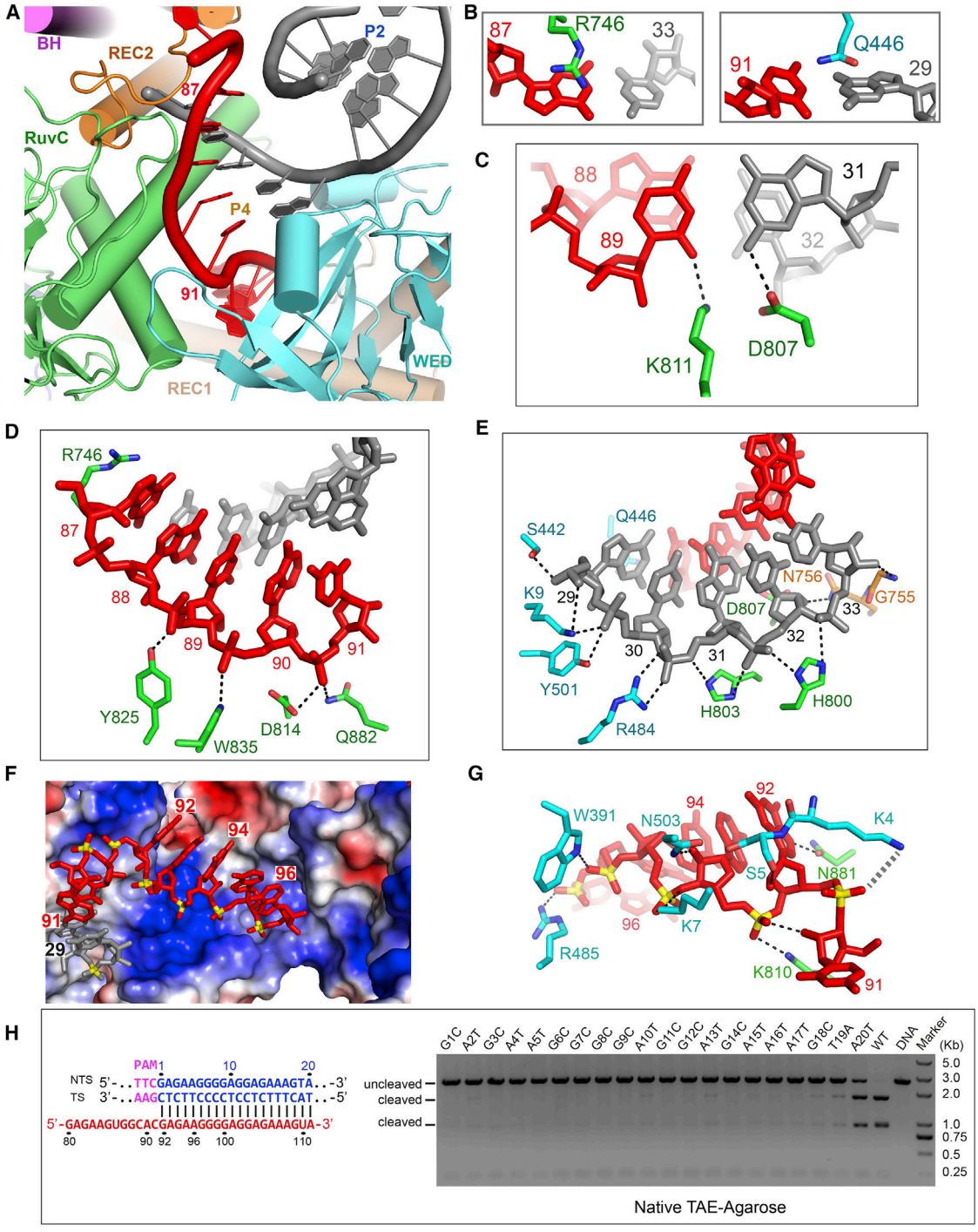


Figure 3. Binding of the crRNA

(A) The crRNA repeat:tracrRNA anti-repeat duplex P4 containing the RuvC and WED domains.

(B) Stacking interaction between R746 and the C33–G87 base pair (left panel) and between Q446 and the G29–C91 base pair (right panel).

(C) Recognition of the repeat:anti-repeat duplex by the RuvC domain.

(D) Intermolecular interactions between AacC2c1 and the repeat region adjacent to the spacer.

(E) Intermolecular interactions between AacC2c1 and the tracrRNA anti-repeat region.

(F) The crRNA seed region binds within the positively charged groove formed by the REC1 domain.

(G) Intermolecular interactions between AacC2c1 and the seed region.

(H) DNA cleavage using single-nucleotide mutant target DNA. The sequences of crRNA and wild-type target DNA are shown in the left panel.

See also Figure S3.

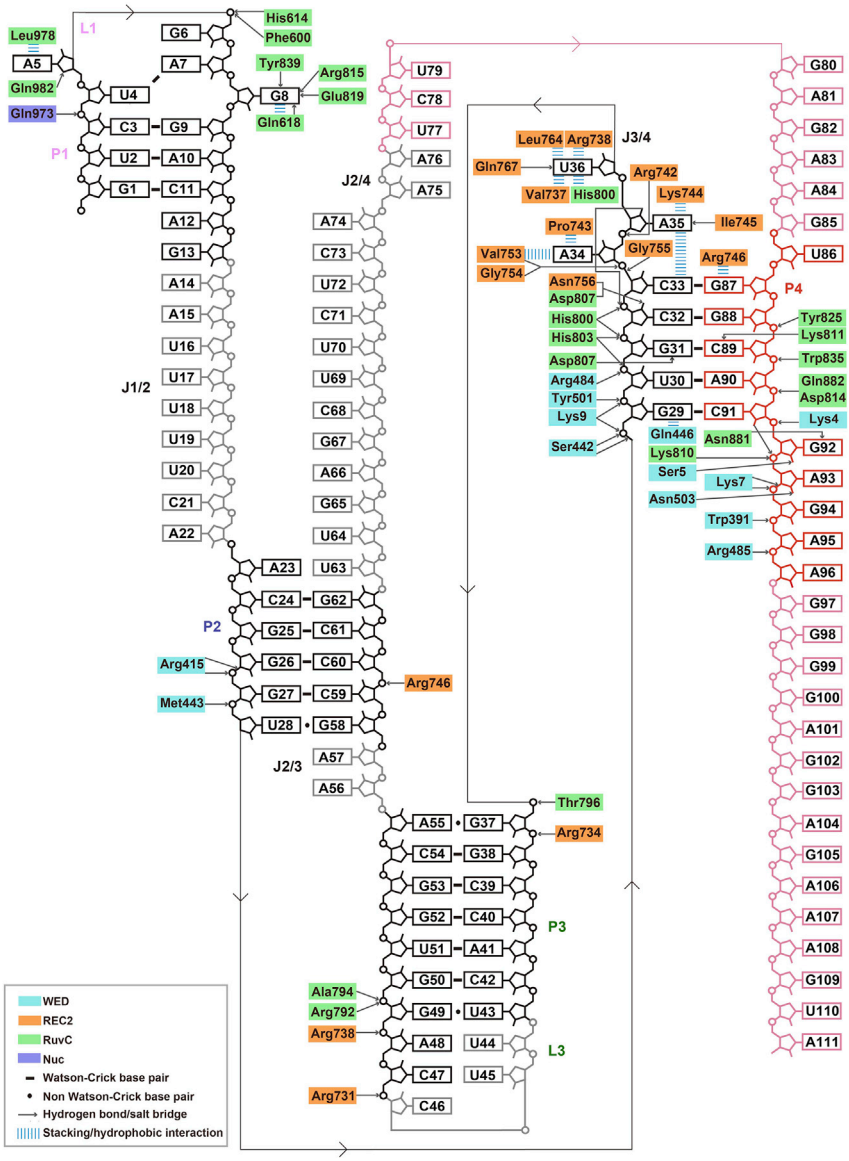


Figure 4. Schematic View of the Intermolecular Contacts in the AacC2c1-sgRNA Binary Complex

Hydrogen bond and salt bridge interactions are shown by black lines, stacking interactions by dashed blue lines.

extension toward the 3′ end. In contrast to the 5 nt seed region of Cpf1 guide RNA (Zetsche et al., 2015), the seed region of the AacC2c1 guide RNA appears to be within the first 18 nt on the 5′ end of the spacer sequence. Such sensitivity toward single-base mutation is in stark contrast to observations for Cas9 and Cpf1 (Jinek et al., 2012; Zetsche et al., 2015) and suggests that the C2c1 system might be an ideal editing tool, thanks to an expected minimal off-target cleavage activity.

Recognition of tracrRNA

As shown in Figure 2C, the tracrRNA is divided into two remotely located regions. The 5′ end of the tracrRNA, forming the stem loop P1/L1, is the smaller region and is located in a cleft formed by the RuvC and Nuc domains (Figure 5A). Nucleotides G1–C3 base pair with nucleotides C11–G9, and U4 pairs with A7, with one hydrogen bond in between (Figure S4A). Nucleotides A5 and G8 flip out and are stabilized by the RuvC domain via hydrogen bonds (Figures S4B and S4C). Notably, G8 is sequence-specifically recognized by R815 and E819 (Figure 5B). Together, these interactions between P1/L1 and C2c1 found in our structure confirm the critical roles of P1/L1, therefore providing a good explanation as to why a tracrRNA of minimum length 78 nt is essential for C2c1-mediated DNA cleavage, as observed previously (Shmakov et al., 2015). However, analysis of G8C and A5U mutants showed that these bases can be variable, without affecting AacC2c1-mediated DNA cleavage activity (Figure S4D).

The larger region of the tracrRNA is bound to C2c1 at a positively charged groove formed by the WED, RuvC, and REC2 domains (Figure 5C; Figure S5A). As shown in Figure 1B, the larger region is isolated from the small region by the REC2 and RuvC domains. Helix P2 connects with P4, with nucleotide 28 within P2 and nucleotide 29 within P4 being positioned in opposite orientations (Figure 5D; Figure S5B). Helix P2 binds to the groove formed between the WED and REC2 domains, with one strand being stabilized by residues S442, M443, R415, and R746 (Figure 4; Figure 5D; Figure S5C). Helix P3 is stabilized by the interactions between its phosphate backbone and the positively charged residues from the RuvC domain. Residues R731, R734, and R738 located in the α20 helix interact with the backbone of helix P3. In addition, helix α22 helix of the RuvC domain further stabilizes helix P3 (Figure S5D).

Interestingly, joining region J3/4 (nucleotides A34, A35, and U36), which connects helices P3 and P4 (Figure S5E), inserts into a cavity of the REC2 domain and is recognized in a base-specific manner (Figures 5E and 5F). All adjacent bases in J3/4 adopt a splayed conformation, with each base positioned orthogonally relative to their preceding base. Nucleotide A34 is flipped outward, relative to nucleotide 33, and the base of A34 is sandwiched between the P743 and V753 residues of the REC2 domain (Figure 5F; Figure S5F). The backbone phosphate group of A34 is further stabilized by an interaction with the main chains of the G754 and V753 residues of REC2. Nucleotide A35 stacks on nucleotide C33 on one side and the side chain of K744 on the other side. The N6 atom of A35 forms hydrogen

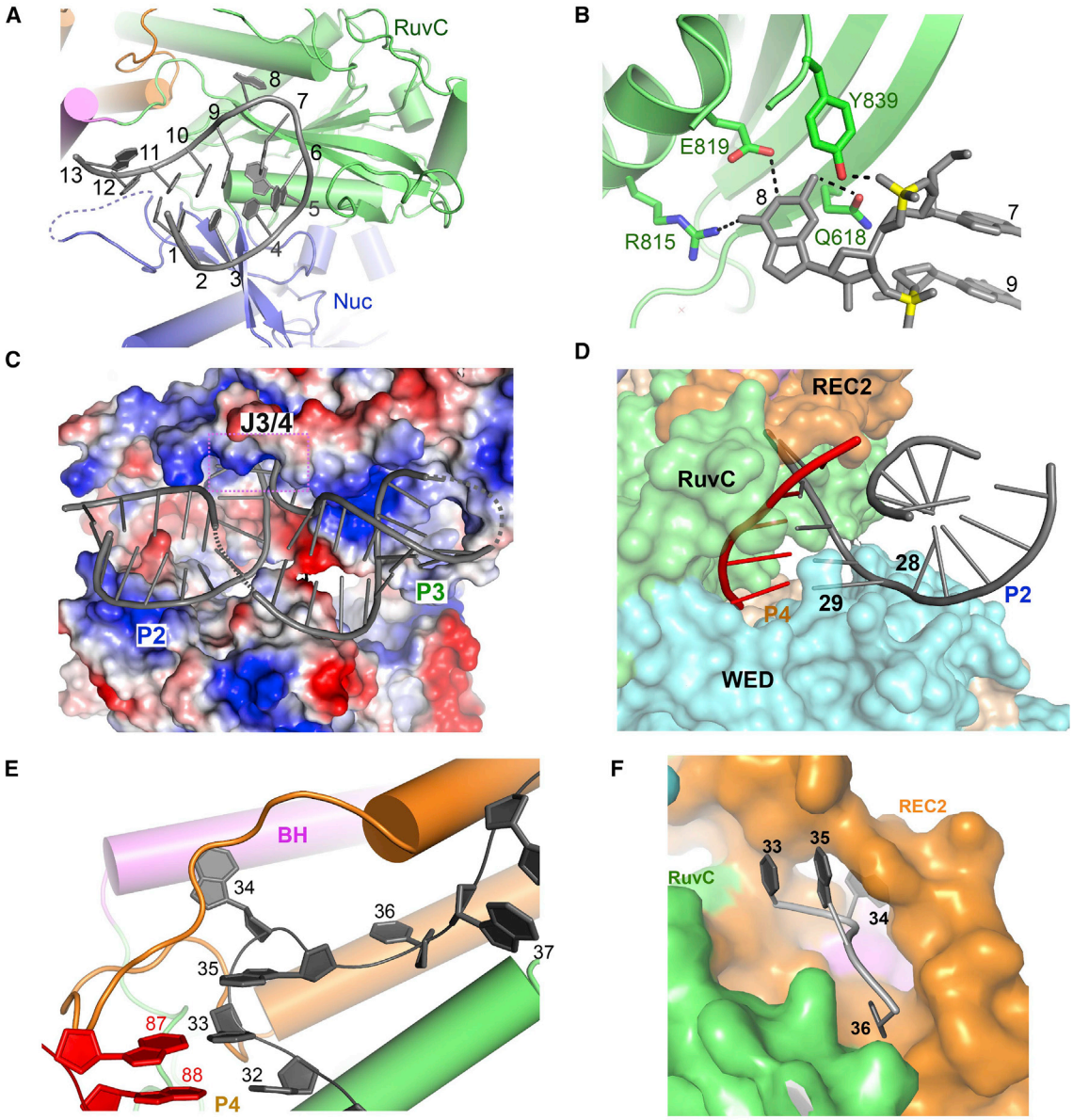


Figure 5. Intermolecular Interactions between AacC2c1 and the Bound tracrRNA Segment

(A) Binding of the stem-loop P1/L1 of the tracrRNA to the groove between the RuvC and Nuc domains.

(B) Sequence-specific interactions between nucleotide G8 and the RuvC domain.

(C) Binding of tracrRNA helix P2 and P3 to the positively charged groove formed by the RuvC, WED, and REC2 domains.

(D) Binding of crRNA repeat:tracrRNA anti-repeat duplex P4 and tracrRNA helix P2 with the RuvC, REC2, and WED domains.

(E) Binding of the joining J3/4 region with the RuvC and REC2 domains.

(F) Close-up view of nucleotides 34–36 within the J3/4 region.

See also Figures S4 and S5.

bonds with the main chain of I745. Base U36 inserts into a narrow cleft formed by RuvC and REC2 and stacks on H800 and L764 on one side and on the peptide plane of V737–R738 on the other side (Figure S5F). Together, these observations suggest that in order to maintain the conformation of the tracrRNA, a significant number of interactions with multiple domains of C2c1 are required.

RuvC and Nuc Domains

Like Cpf1, the RuvC and Nuc domains are juxtaposed in the AacC2c1 structure (Figure 6A). To investigate the relative importance of the RuvC and Nuc domains for catalytic activity of C2c1, we first performed a Dali database search (Holm and Rosenström, 2010) to identify possible homologous structures. We found that the RuvC domain of AacC2c1 is structurally

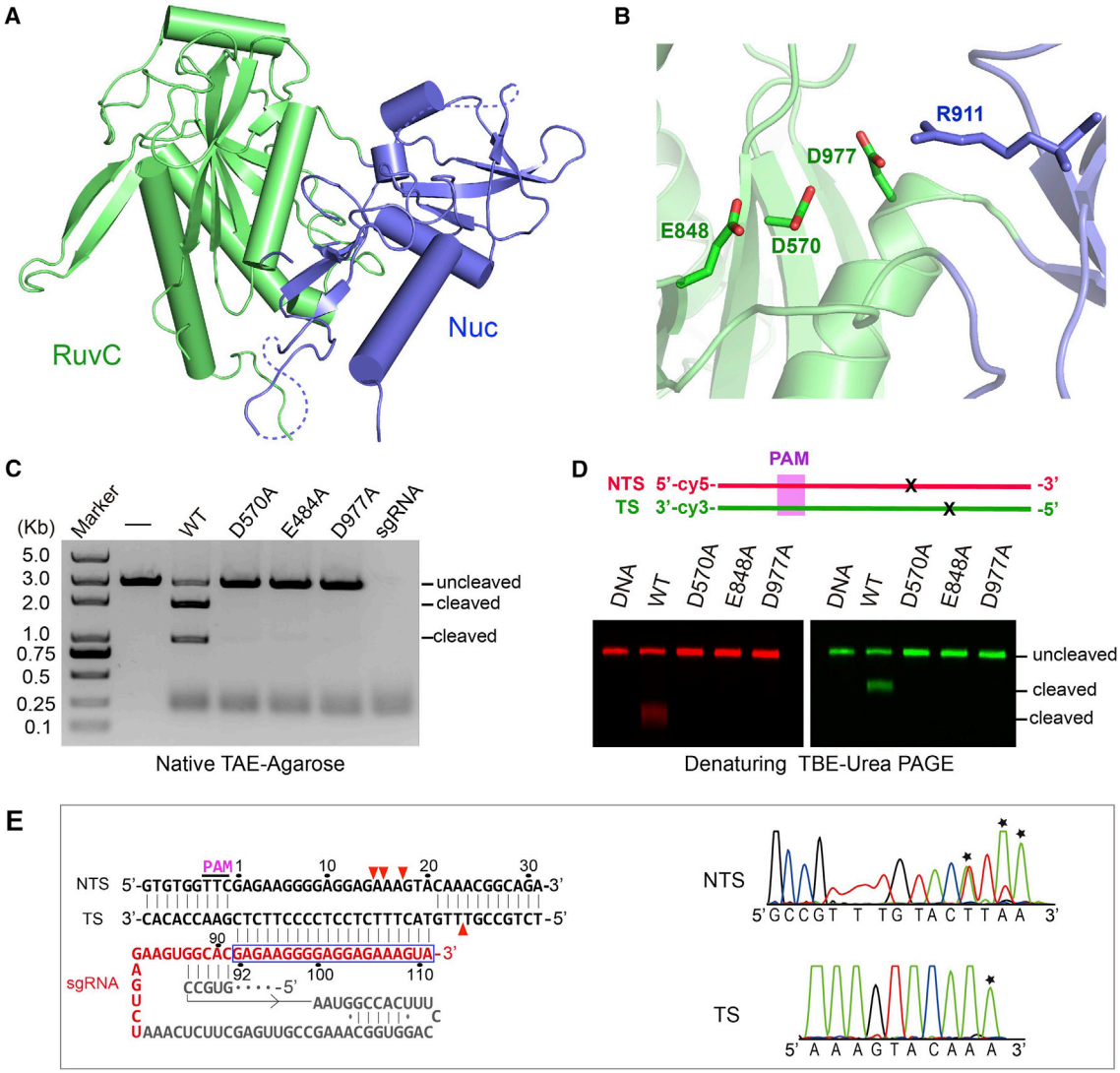


Figure 6. RuvC and Nuc Domains

(A) Structure of the RuvC and Nuc domains.

(B) Active site of the RuvC domain.

(C) Mutational analysis of RuvC catalytic residues using plasmid DNA.

(D) Mutational analysis of RuvC catalytic residues using a labeled target synthetic short double-stranded DNA. The target and non-target strands are labeled by cy5 at the 5′ end and cy3 at the 3′ end. The cleavage sites on the target strand (TS) and the non-target strand (NTS) are indicated by two X symbols.

(E) Sanger-sequencing traces from AacC2c1-cleaved target show staggered overhangs. The cleavage sites are highlighted by red triangles in the left panel. The non-templated addition of an additional adenine, indicated by black star, is an artifact of the polymerase used in sequencing (right panel).

See also Figure S6.

homologous with the Holliday junction resolvase protein (Li et al., 2016) (PDB: 5E6F; *Z* score: 9.0; RMSD: 2.4 Å over 110 aligned Cα atoms) and LbCpf1 (Dong et al., 2016) (PDB: 5ID6; *Z* score: 8.8; RMSD: 3.5 Å over 151 aligned Cα atoms). The RuvC domain is the most conserved domain of the C2c1 protein, suggesting that this domain uses similar molecular mechanisms for DNA cleavage as the RuvC domain of Cpf1.

The RuvC domain of C2c1 possesses all the catalytic residues found in this family of endonucleases (Shmakov et al., 2015). As shown in Figure 6B, the conserved negatively charged residues D570, E848, and D977 form an active triad similar to that found in the Cpf1 RuvC domain and the Cas9 RuvC domain. Therefore, we generated mutants of three catalytic residues (D570, E848, and D977) to test whether these conserved residues are essential for the nuclease activity of C2c1. Complete loss of DNA cleavage activity confirmed that D570, E848, and D977 are essential for C2c1-mediated DNA cleavage (Figure 6C). Further mutational analysis showed that single-point mutations at D570, E848, and D977 abolished DNA cleavage activity of both target and non-target strands, suggesting that RuvC domain are involved in the

cleavage of both the target strand and non-target strands (Figure 6D).

Intriguingly, a Dali search (Holm and Rosenström, 2010) revealed that the Nuc domain, one of the least conserved regions across the C2c1 family, lacks any structural similarity with other known proteins, indicating that it is a C2c1-specific functional domain. The Nuc domain of C2c1 protein shows no structural similarity to Cpf1, even though both of these two proteins belong to the type-V system. To identify the critical residues within the Nuc domain, we performed cleavage assay using single mutant proteins. The R911A mutant showed dramatically reduced activity, indicating that R911 is critical for DNA cleavage (Figure S6A). The R911 is located close to the RuvC catalytic pocket (Figure 6B). Similarly, R1226 of AsCpf1, which is also located near the RuvC catalytic site, is critical for the DNA cleavage (Yamano et al., 2016), indicating that while the structure of C2c1 Nuc domain is different from the Nuc domain of Cpf1, the functions of the two Nuc domains might be similar. The alanine substitution of R1000 and R1015, which are located on the surface of the Nuc domain, also reduced cleavage activity, suggesting that these two residues are also important in cleavage. However, none of the remaining mutants affected cleavage activity (Figure S6A).

To determine the type of product generated by AacC2c1, we performed endonuclease assays using Cy3/Cy5-labeled DNA. As shown in Figure 6D, AacC2c1-catalyzed DNA cleavage generated the product containing 5′ overhang like Cpf1. We then mapped the cleavage site of AacC2c1 using Sanger sequencing of the cleaved plasmid DNA ends. As shown in Figure 6E, AacC2c1-mediated cleavage of plasmid DNA resulted in a 6–8 nt 5′ overhang. The target DNA strand that is complementary to the guide segment of the sgRNA is cleaved 23 bp upstream of the PAM sequence. The non-complementary DNA strand is cleaved at 1–3 sites within 14 to 17 bp upstream of the PAM sequence.

As DNA cleavage activity of the C2c1 RuvC domain is metal dependent (Ariyoshi et al., 1994), we conducted cleavage assays in the presence of different metals. As shown in Figure S6B, AacC2c1 cleaves DNA in the presence of Mg^{2+}, Mn^{2+}, or Ca^{2+}, with the highest activity being observed for Mn^{2+}. However, cleavage activity was absent in the presence of Ni^{2+}, Zn^{2+}, or Co^{2+}, showing that C2c1 requires identical metals as Cas9 and Cpf1 for cleavage activity.

Interestingly, RuvC is not only an endonuclease domain, but also plays critical roles in sgRNA binding. Helix α22, with loops at both ends, interacts with both the tracrRNA and guide segment (Figure S6C). It is perpendicular to the repeat:anti-repeat duplex and recognizes this duplex by sequence-specific interactions via residues K811 and D807. T796 and H800 are involved in the stabilization of helix P3 (Figure 4). Furthermore, residues E819, R815, and Y839 interact with the loop of stem-loop P1/L1 within the tracrRNA in a sequence-specific manner.

DISCUSSION

The present structure of the AacC2c1-sgRNA binary complex provides insights into how C2c1 in complex with its crRNA and tracrRNA provide immunity against invading DNA elements. Our structural analysis has shown that, even though this enzyme fulfills an immune function similar to Cas9 and Cpf1, its structural features differ markedly from those two proteins. While C2c1 possesses a bi-lobed architecture, the structure and organization of its functional domains is distinct from related proteins in other CRISPR systems. The most intriguing difference is the absence of a PAM-interacting domain, which is essential for target DNA binding in Cas9 and Cpf1 (Anders et al., 2014; Jiang et al., 2016; Yamano et al., 2016). Furthermore, the structure of the sgRNA is considerably more complex than that found for Cas9 or Cpf1 (Figure 7A).

Our structural comparison of C2c1, Cpf1, and Cas9, the only available structures of class 2 CRISPR-Cas system effectors, revealed considerable similarities in their overall architectures. Despite the lack of sequence similarity outside the RuvC domain, the three proteins are of approximately the same size and adopt a bi-lobed structure. In all three proteins, the crRNA is located in the central channel between the two lobes (Dong et al., 2016; Jiang et al., 2015). Previous structural studies of Cas9 and Cpf1 revealed that the crRNA-target DNA heteroduplex is located in the positively charged central channel between the REC and Nuc lobes (Nishimasu et al., 2014; Yamano et al., 2016). A structural comparison between AacC2c1 and AsCpf1 suggests that in C2c1, the crRNA-target DNA heteroduplex is also positioned inside the central positive-charged channel (Figure S7A). The second α-helical bundle of the REC1 and REC2 domains contains several Arg residues, generating a positively charged internal channel surface on the opposite direction of the crRNA seed region. This observation indicates that these regions interact with the crRNA-target DNA duplex. In our sgRNA-bound structure, the channel between the two lobes is too small to harbor an RNA-DNA duplex, suggesting that the two REC domains need to be rearranged upon target DNA binding.

Both Cas9 and Cpf1 contain a PI domain, which is essential for PAM recognition during the target DNA binding step. Mutation of the PI domain has been shown to alter the PAM specificity of Cas9 (Anders et al., 2014, 2016; Nishimasu et al., 2014). In Cpf1, the PAM duplex of the double-stranded target DNA binds to the groove that is formed by the WED, REC1, and PI domains. The WED-REC1 and PI domains recognize the PAM duplex from the major and minor groove side, respectively (Yamano et al., 2016). In Cas9, the PAM sequence is recognized by the residues of the PI domain in a sequence-specific manner, whereby the side chains of two Arg from PI form hydrogen bonds with the bases of the PAM sequence (Anders et al., 2014). Surprisingly, the AacC2c1 protein lacks any identifiable PI domain (Figures 1A, 1B, and 7A), suggesting that C2c1 recognizes the PAM-containing target DNA duplex in a manner distinct from that of Cpf1 or Cas9. Our structural comparison of C2c1 and Cpf1 in complex with both crRNA and target DNA suggests that the PAM duplex of C2c1 is located at a similar position to that of Cpf1. In addition, the presence of a positively charged surface of the wider groove between the REC1 and WED domains in C2c1 further suggests that the PAM duplex binds to this groove. The WED and REC1 domains may be responsible for the PAM recognition to compensate the lack of a PI domain. On the basis of these results, we modeled the PAM DNA duplex substrate into the

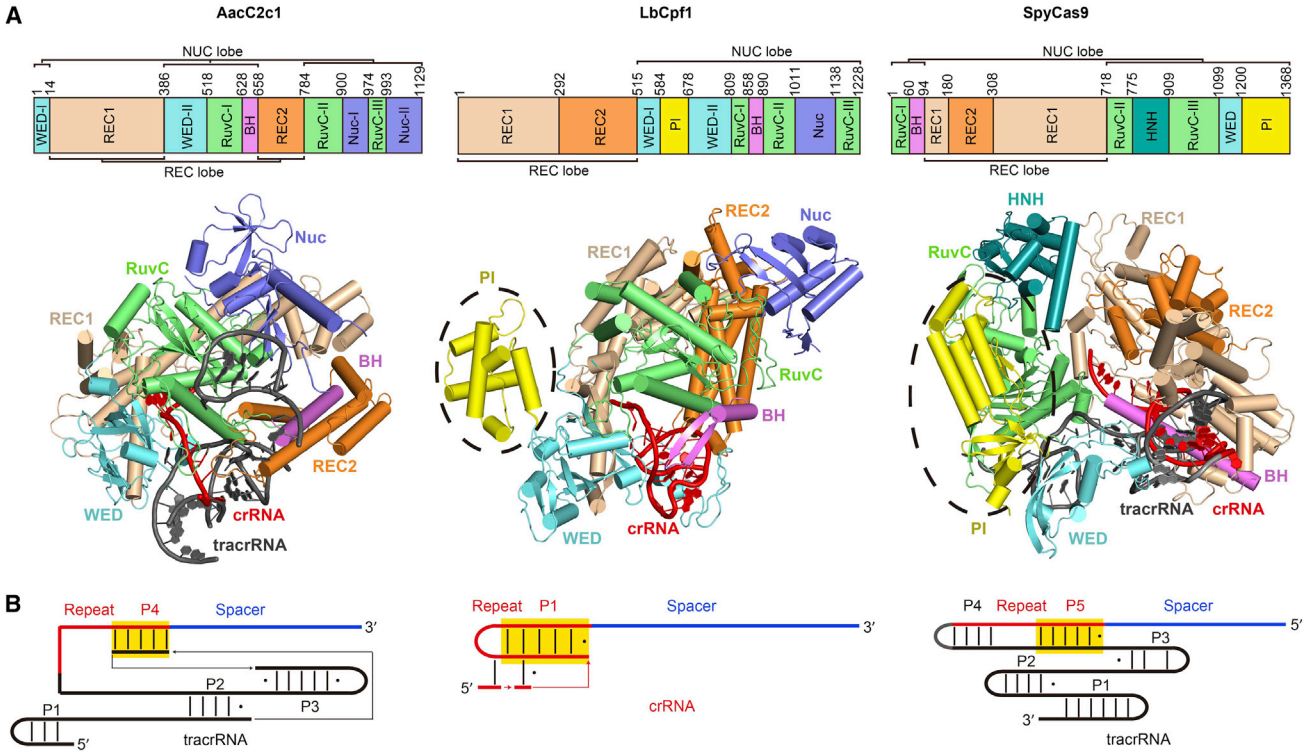


Figure 7. Comparison of C2c1, Cpf1, and Cas9

(A) Comparison of the domain organizations and overall structures of AacC2c1, LbCpf1 (PDB: 5ID6), and SpyCas9 (PDB: 4ZT0). The PI domain in LbCpf1 and SpyCas9 are shown in yellow and highlighted by a dashed oval.

(B) Schematic representation of sgRNA bound to AacC2c1, LbCpf1, and SpyCas9. The spacer and repeat of the crRNA are shown in blue and red, respectively. The tracrRNA segment is in black.

See also Figure S7.

C2c1-sgRNA complex structure based on the structure of the Cpf1-crRNA-target DNA ternary complex (Figure S7B). However, to identify the precise molecular mechanisms of DNA substrate recognition of C2c1, further structural studies that include a target DNA molecule will be required.

Cas9 contains RuvC and HNH nuclease domains that cleave the non-target and target strands, respectively, creating blunt ends proximal to the PAM site. In contrast, C2c1 and Cpf1 contain a RuvC and a putative Nuc domain, generating a staggered double-stranded break distal to the PAM site. While the RuvC domains of Cas9, C2c1, and Cpf1 adopt similar folds (Figure S7C), the HNH domain of Cas9 has no structural similarity to the Nuc domain of either C2c1 or Cpf1. Furthermore, the putative Nuc domain of C2c1 shows no sequence or structural similarity to that of Cpf1 (Figure S7D).

The HNH domain of Cas9 cuts target DNA strands within the RNA-DNA heteroduplex (Jinek et al., 2012), whereas the Nuc domain of Cpf1 cleaves the single-stranded region of the target DNA outside the crRNA-target DNA heteroduplex (Zetsche et al., 2015). The present structure, together with data from our own and previous biochemical studies (Shmakov et al., 2015), suggests that C2c1 cleaves the target strand DNA outside the heteroduplex in a manner similar to that observed for the Cpf1 Nuc domain. Previous reports have shown that inactivation of the non-target cutting RuvC domain of Cas9 has no effect on the cleavage of the target strand cut by the HNH domain. In contrast, non-target strand cleavage by the Cpf1 RuvC domain is a prerequisite for the cutting of the target strand by its Nuc domain. Our biochemical analysis of C2c1 activity has shown that mutation of the catalytic residues of RuvC also abolishes target strand cleavage (Figures 6C and 6D), suggesting that an active RuvC domain is required for the cleavage of both strands, similar to Cpf1.

A comparison of the guide RNA structures for C2c1, Cpf1, and Cas9 shows that the nucleotides adjacent to the spacer region form a duplex with the anti-repeat regions in all three proteins (Figure 7B). Specifically, in the AacC2c1-sgRNA, 5 nt adjacent to the spacer form an RNA duplex with the anti-repeat region within the tracrRNA. Both Cas9 and C2c1 are dual-RNA-guided endonucleases. The sgRNA bound to Cas9 and C2c1 forms a duplex with the tracrRNA, whereas in Cpf1, these nucleotides form a stem-loop within the 5′ repeat of the crRNA. Interestingly, mutations that break the base-pairing within this crRNA repeat duplex reduce cleavage activity of Cpf1 (Zetsche et al., 2015). In all three endonucleases, the duplex adjacent to the seed-region-containing guide sequence may facilitate the proper positioning of the seed region within the binding channel, facilitating target DNA binding.

In summary, our study provides structural understanding of the mechanistic principles of C2c1 function in CRISPR-dependent bacterial immunity. Importantly, given the restrictions imposed by the PAM sequence on the design of genome-editing tools, these insights should open avenues in expanding our gene-editing repertoire across different species.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The full-length AacC2c1 gene (encoding residues 1–1,129) was synthesized by Sangon Biotech and cloned into the pET30b vector (Novagen) containing a His_6 tag at the AacC2c1 C-terminal. The AacC2c1 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 for 14 hr at 16°C. 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, supernatants were incubated with Ni Sepharose (GE Healthcare). Bound proteins were eluted with buffer containing 100 mM imidazole and further purified on a Heparin column (GE Healthcare) and eluted with buffer containing 20 mM Tris-HCl (pH 7.5), 1 M NaCl, and 5% glycerol. Target proteins were collected and concentrated by ultrafiltration. Finally, concentrated protein solutions were loaded onto a Superdex 200 Increase 10/300 gel filtration column (GE Healthcare Life Sciences) equilibrated with 20 mM Tris-HCl buffer of pH 7.5 (with 200 mM NaCl, 5% glycerol, and 1 mM Tris [2-carboxyethyl] phosphine hydrochloride [TCEP]). Fractions containing AacC2c1 proteins were collected and then concentrated to a final concentration of 20 mg/mL. The SeMet-substituted AacC2c1 was purified using an identical protocol.

AacC2c1-sgRNA Complex Reconstitution

The 111 nt sgRNA was synthesized by in vitro transcription with T7 RNA polymerase using linearized plasmid DNA as a template. To re-fold the purified in vitro transcribed RNA, we heated the sgRNA at 75°C for 5 min and slowly cooled it to room temperature. The AacC2c1-sgRNA complex was reconstituted at room temperature for 30 min by incubating AacC2c1 and sgRNA at a molar ratio of 1:1.4. The complex was then 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. The complex fractions were collected and concentrated. Samples used for crystallization were measured for final absorbance of ~20 using a Nanodrop spectrophotometer.

Crystallization, Data Collection, and Structure Determination

SeMet crystals of the AacC2c1-sgRNA complex were grown by the hanging-drop vapor diffusion method at 16°C. Crystals were obtained by mixing 1 μL complex solution and 1 μL of reservoir solution. SeMet Crystals of the AacC2c1-sgRNA complex were grown from 0.1 M Tris-HCl (pH 8.0), 0.2 M LiCl, and 16% PEG 8000. All crystals were cryo-protected using corresponding reservoir buffers supplemented with 10% (v/v) glycerol and then flash-frozen in liquid nitrogen.

All diffraction datasets were collected at beamline BL-19U1 at the Shanghai Synchrotron Radiation Facility (SSRF) and processed with HKL2000 (Otwinowski and Minor, 1997). The phases of AacC2c1 were solved with the 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 data processing and structure refinement statistics for AacC2c1 are summarized in Table 1. The structure figures were prepared using PyMOL (http://www.pymol.org/).

In Vitro Cleavage Assays

5′-Cy5- and 3′-Cy3-labeled target DNA (59 nt) or pUC19 target DNA (52 nt target DNA cloned into the pUC19 vector) were used for in vitro cleavage assays. The pUC19 target DNA construct was linearized by ScaI digestion prior to the cleavage reaction. 5 μmol of the AacC2c1-sgRNA complex was incubated for 2 hr (4 hr for Figure 3H) at 48°C with 1 μmol 5′-Cy5- and 3′-Cy3-labeled target DNA or 200 ng pUC19 target DNA in 10 μL reaction buffer containing 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. 5′-Cy5-and 3′-Cy3-labeled target DNA reactions were mixed 1:1 with loading buffer and separated on a 20% urea denaturing polyacrylamide gel. DNA was visualized using a FluorChem system (Sangon Biotech). pUC19 target DNA reactions were run on 1% agarose gels (Life Technologies).

DNA Oligonucleotides

The sequences of DNA oligonucleotides used in the study are provided in Table S1. All DNA oligonucleotides used for biochemical experiments were purchased from Sangon Biotech. DNA oligonucleotides used for cleavage assays contained 5′-Cy5 and 3′-Cy3 labels.

ACCESSION NUMBERS

The accession number for the atomic coordinates of the AacC2c1-sgRNA complex reported in this paper is PDB: 5WQE.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.11.040.

AUTHOR CONTRIBUTIONS

L.L. and P.C. expressed, purified, and grew crystals of the AacC2c1-sgRNA complex. L.L. and X.L. 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 the AacC2c1-sgRNA complex structure. 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. This work was supported by grants from the Natural Science Foundation of China (91440201, 31630015, 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 and Dr. Joy Fleming for critical reading of the manuscript and linguistic assistance.

Received: October 13, 2016
Revised: November 21, 2016
Accepted: November 29, 2016
Published: December 15, 2016

REFERENCES

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.

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *513*, 569–573.

Anders, C., Bargsten, K., and Jinek, M. (2016). Structural plasticity of PAM recognition by engineered variants of the RNA-guided endonuclease Cas9. Mol. Cell *61*, 895–902.

Ariyoshi, M., Vassylyev, D.G., Iwasaki, H., Nakamura, H., Shinagawa, H., and Morikawa, K. (1994). Atomic structure of the RuvC resolvase: a holliday junction-specific endonuclease from E. coli. Cell *78*, 1063–1072.

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.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

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.

Doudna, J.A., and Sontheimer, E.J. (2014). Methods in Enzymology. The use of CRISPR/Cas9, ZFNs, and TALENs in generating site-specific genome alterations. Preface. Methods Enzymol. *546*, xix–xx.

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., 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.

Holm, L., and Rosenström, P. (2010). Dali server: conservation mapping in 3D. Nucleic Acids Res. *38*, W545–W549.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O., et al. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. *31*, 827–832.

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.

Li, H., Hwang, Y., Perry, K., Bushman, F., and Van Duyne, G.D. (2016). Structure and metal binding properties of a Poxvirus Resolvase. J. Biol. Chem. *291*, 11094–11104.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR-Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., 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.

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.

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.