Special Topic: Gene Editing

BIOLOGY & BIOCHEMISTRY

Recent advances in structural studies of the CRISPR-Cas mediated genome editing tools

Yuwei Zhu and Zhiwei Huang

HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin, China

Corresponding author: HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China. Tel: +86-451-86403163; Fax: +86-451-86403163; Email: yuweizhu@hit.edu.cn or huangzhiwei@hit.edu.cn

ABSTRACT

Clustered regularly interspaced short palindromic repeat (CRISPR) and accompanying CRISPR-associated (Cas) proteins provide RNA-guided adaptive immunity for prokaryotes to defend viruses. The CRISPR-Cas systems have attracted much attention in recent years for their power in developing genome editing tools. Based on the composition of crRNA-effector complex, the CRISPR-Cas systems can be divided into two classes and six types. In this review, we summarize recent advances in the structural biology of the CRISPR-Cas mediated genome editing tools, which helps us to understand the mechanism of how the guide RNAs assemble with diverse Cas proteins to cleave target nucleic acids.

Keywords: CRISPR-Cas system, adaptive immune, viral infection, genome editing

Received: 06-Sep-2018; Revised: 21-Nov-2018; Accepted:26-Nov-2018

INTRODUCTION

An evolutionary struggle between prokaryotes and viruses has been going on for billions of years [1]. The selective pressures imposed by viruses drive the diversification of immune defense systems of prokaryotes [2, 3]. Clustered regularly interspaced short palindromic repeat (CRISPR) and accompanying (Cas) proteins constitute an RNA-based antiviral immune system, found in about 90% of archaea and 50% of bacteria [4]. A typical CRISPR locus consists of an array of short direct repeats and interspersed spacer sequences, which is flanked by diverse *cas* genes [5] (Fig. 1a). The repeats contain the same sequences within a CRISPR locus, vary in both length and sequence in different units. In contrast, the spacers present unique DNA sequences gained from invading viruses or plasmids. Adjacent to the repeat of a CRISPR locus, an A-T-rich 'leader' sequence is observed, which is vital the CRISPR transcription and spacer acquisition [6, 7] (Fig. 1a). The CRISPR-Cas adaptive immune systems are known to function through three distinct stages: spacer sequence acquisition (stage 1), CRISPR RNA (crRNA) biogenesis (stage 2), and RNA-guided interference (stage 3) [8, 9]. During infection, a short sequence (protospacer) from an invading virus or plasmid is inserted into the CRISPR locus as spacer [10, 11] (Fig. 1a). Biochemical and structural biology studies have shown that Cas1 and Cas2 form a stable complex, serving as a governor for the incorporation of new spacers into the CRISPR locus via a cut-and-paste mechanism [12, 13]. This acquisition machinery works in a sequence specific manner to avoid self-targeting, so that only the invading DNA flanked by a protospacer-adjacent motif (PAM) can be recognized and selected as a protospacer. In the crRNA biogenesis stage, the CRISPR locus is transcribed into a precursor CRISPR RNA (pre-crRNA), which is processed into mature crRNAs (Fig. 1b). Pre-crRNA cleavage is mediated by either Cas6 (Class 1 CRISPR-Cas systems) or RNase III (Class 2 CRISPR-Cas systems) 15]. Finally, crRNA-guided interference occurs. In this stage, mature crRNAs

with Cas proteins to form a surveillance complex, which recognizes and cleaves invading nucleic acids [16] (Fig. 1c).

Based on locus organization and gene conservation, the CRISPR-Cas systems can divided into two classes, six types and several subtypes [17, 18]. Class I CRISPR-Cas systems, consisting of types I, III, and IV, employ multisubunit crRNA-effector complexes for interference. Class II CRISPR-Cas systems, consisting of types II, V, and VI, are featured by the presence of a single subunit of crRNA-effector module. Type I system is defined by the signature *cas3* gene, and is currently divided into subtypes (I-A, I-B, I-C, I-D, I-E, I-F and I-U) [18]. During the interference stage, multiple Cas proteins assemble with a mature crRNA to form the Cascade complex, which recruits a nuclease-helicase protein Cas3 to degrade invading nucleic acids Unlike type I system, type II CRISPR locus displays a simplified composition, a single effector protein (Cas9) guided by a dual-RNA heteroduplex (crRNA:tracrRNA) [20, 21]. Type II system can be further divided into three subtypes (II-A, II-B and II-C) [18]. Type III system employs a multi-protein complex, which is similar to type I system. The signature gene of type III system is *cas10*, which a large multi-domain protein. Four subtypes of type III system have currently been identified, including III-A, III-B, III-C and III-D [18]. Type III system possesses two kinds of enzymatic activities (ssRNase and ssDNase) [22, 23]. This property confers type III system a versatile immune response against different types of foreign genetic elements, and an efficient fail-safe way for degradation of both invading DNA and its transcript [24]. Type IV system is putative and functionally uncharacterized, which shows a minimal multisubunit crRNA-effector complex differing from type I and type III systems [17]. *Csf1* is a hallmark gene of this system. Type V and VI systems a single subunit crRNA–effector complex. Three RuvC domain-containing proteins (Cpf1, C2c1 and C2c3) have currently been identified as the effectors of type V [25, 26]. Type VI system employs HEPN nuclease domain-containing effectors, including Cas13a, Cas13b, Cas13c and Cas13d [18].

Among all the CRISPR-Cas systems, the type I system accounts for 95%, and is the most widely distributed. In many cases, heterologous proteins, such as Cas9 and Cpf1, are hard to be transformed into bacteria and archaea due to their intrinsic toxicity, leading to low genome editing efficiency. Thus, the type I CRISPR-Cas system was harnessed as an endogenous RNA-guided machinery for multiplex genome editing in prokaryotes [27-29]. The type II CRISPR-Cas9 system is the most popular genome editing tool, and has been successfully applied in a broad range of organisms, such as bacteria, yeasts, plants, animals, and human cells [30-33]. The type V CRISPR-Cpf1 system emerges as an alternative to the CRISPR-Cas9 technology [34]. The genome editing activity of CRISPR-Cpf1 is not as robust as CRISPR-Cas9, but with higher targeting efficiency [35, 36]. Given the powerful RNA recognition and cleavage ability, the type VI CRISPR-C2c2 system has been harnessed as a toolkit for RNA base editing, RNA knockdown, nucleic acid detection, and transcript tracking [37]. Although lots of biochemical and structural studies have been reported concerning the composition and functional activities of these CRISPR-Cas systems, a comprehensive and systematic analysis of the diverse interference mechanisms of these genetic silencing systems is still lacking. In this review, we summarize the current knowledge related to these CRISPR-Cas effector complexes, which will deepen our understanding of the architecture of distinct types of CRISPR-Cas systems and how crRNA-guided Cas proteins recognize and cleave invading nucleic acids. Furthermore, it will enhance the application of CRISPR-Cas systems as genome editing tools.

TYPE I CRISPR-CAS SYSTEM: AN ENDOGENOUS TOOL FOR MULTIPLEX GENOME EDITING IN PROKARYOTES

The type I CRISPR-Cas complex is also named Cascade (CRISPR associated complex for antiviral defense), assembled by multiple Cas proteins and a crRNA [38, 39]. The recognition of target DNA is initiated by PAM scanning, which assists in the unzipping of the base pairs adjacent PAM [40]. Then, the target DNA strand

with the crRNA spacer to form a heteroduplex, while the non-target DNA strand is displaced. This unique conformation is called “R-loop” [41]. After the formation of Cascade/R-loop, Cas3 is recruited to degrade the target DNA [19, 42]. Up to now, the structures of type I-E and type I-F complexes have been determined, while the type I DNA interference complexes remain unknown.

Composition, structure and functional activities of the type I-E surveillance complex

The type I-E subtype is the most common and mostly studied type I CRISPR-Cas system. It has been utilized as a programmable gene expression regulator, enabling silence of both heterologous and endogenous genes [43]. Furthermore, it was engineered to be a genetically encoded device, termed DNAi, which could sense transcriptional inputs and make a direct degradation of user-defined DNAs [44]. The atomic structures of *E. coli* type I-E surveillance complex, and its complex with ssDNA or dsDNA were determined by X-ray diffraction [45, 46]. The *E. coli* type I-E complex has a molecular weight of 405 kDa, comprising eleven subunits from five Cas proteins ($CasA_1$:$CasB_2$:$CasC_6$:$CasD_1$:$CasE_1$), as well as a 61-nt crRNA (Fig. 2a). The 61-nt crRNA is processed from pre-crRNA by CasE [47], which could specifically recognize and cut the repeat sequences of pre-crRNA. The mature 61-nt crRNA is composed of 8-nt 5’ handle, a 32-nt spacer sequence and a 21-nt 3’ stem-loop (Fig. 2b).

The overall structure of the type I-E complex adopts a sea-horse shape containing eleven subunits, which are arranged into two layers (Fig. 2a). CasD, six copies of and CasE constitute the outer layer, while CasA and two copies of CasB form the layer (Fig. 2a). The outer and inner layers wrap each other to form a DNA-like double-helix conformation. Within the outer layer, six copies of CasC subunits a symmetry-related helical alignment (Fig. 2a). The CasD and CasE locate at the two ends, respectively (Fig. 2a). The CasA subunit is a two-domain protein, containing an N-terminal globular fold and a C-terminal four-helix bundle (Fig. 2c). The CasB subunit comprises two helix-bundles connected by a loop (Fig. 2d). The structure of

CasC subunit resembles a right hand, consisting of a modified RRM domain (palm), a protruding β-hairpin (thumb), and a helical domain (fingers) (Fig. 2e). The CasD subunit also contains a modified RRM domain, with a β-hairpin protruding from the core (Fig. 2f). The CasE subunit contains two tandem RRM domains (Fig. 2g). Within the inner layer, CasA locates at one end, making extensive interactions with CasD. CasB dimer sits in the groove enclosed by CasA, CasC, and CasE (Fig. 2a). The 61-nt crRNA threads through the outer layer, connecting these subunits together. The 8-nt handle region of crRNA is sandwiched between the CasA, CasD and the adjacent $CasC^1$ (Fig. 2a). Six copies of CasC subunits oligomerize along the spacer region of crRNA (Fig. 2a). After processing of the pre-crRNA, CasE retains tightly binding to 3' stem-loop of the crRNA (Fig. 2a).

The structure of type I-E complex bound to target DNA reveals that the guide:target hybrid displays a ribbon-like conformation, in contrast to the double helix structure [48] (Fig. 2h). This is caused by the kinks occurring in every sixth base pair in both strands of the hybrid. In addition to the Watson-Crick hydrogen bonding with the spacer region of crRNA, the target DNA strand also makes interactions with the CasB dimer, and $CasC^1$ subunits [48] (Fig. 2h). The type I-E complex adopts a promiscuous PAM recognition mode [49]. At least five PAM sequences, such as 5'-ATG, AGG, AAG, TAG, and GAG could initiate the type I-E-mediated CRISPR interference [19, 42]. More recently, Ke's group reported the Cryo-EM structures of type I-E Cascade/R-loop and type I-E Cascade/R-loop/Cas3 from *Thermobifida fusca* at the atomic resolution [41, 50]. The R-loop formation induces severe dsDNA at the PAM-proximal side, as well as a series of conformational changes in type I-E Cascade [50]. Then, the type I-E Cascade/R-loop complex licenses Cas3 to bind (Fig. 2h-i). The recruitment of Cas3 mainly depends on the interactions between Cas3 and

CasA subunit in Cascade in a fashion of conformation capture [50] (Fig. 2h-i). After Cas3-medited ssDNA nicking, the severed non-target strand DNA relocates to Cas3 helicase [50]. Finally, the processive DNA degradation begins (Fig. 2i).

Structures of the type I-F surveillance complex bound with anti-CRISPRs

The structural studies of the type I-F CRISPR-Cas system are benefited from the identification of phage-encoded anti-CRISPR proteins. Overall, ten anti-CRISPR proteins targeting the type I-F genetic silencing machinery have been found [51, 52]. To investigate the inhibition mechanisms of these anti-CRISPR proteins, Cryo-EM structures of the type I-F surveillance complex bound to anti-CRISPR proteins, AcrF1, AcrF2, and AcrF10, were determined respectively [53-55]. The type I-F complex from *Pseudomonas aeruginosa* has a molecular weight of 350 kDa, comprising nine subunits from four Cas proteins ($Cas5f_1$:$Cas6f_1$:$Cas7f_6$:$Cas8f_1$), as well as a 60-nt crRNA (Fig. 2j). The type I-F complex shows a structural similarity to the previously described type I-E complex, with six copies of Cas7f as the backbone, one copy of Cas6f as the head, and one copy of the Cas8f-Cas5f heterodimer as the tail (Fig. 2j). However, structural differences between the type I-E and type I-F complex still exist. The head and tail of the type I-F complex is positioned in close proximity, which causes a nearly closed ring architecture. In addition, the $CasC^6$ subunit of the type I-E complex rotates 180 degrees to form a region for binding to dsDNA, which is not observed in the corresponding subunit of the type I-F complex.

The overall structure of the full-length crRNA in the type I-F complex resembles a string, which tethers distinct protein subunits together. Extensive inter-molecular interactions are formed between the crRNA and the protein subunits. The crRNA recognition modes between type I-F and I-E complexes are tremendously similar. Briefly, the 5' handle region of crRNA is sandwiched between the Cas5f, Cas8f and the adjacent $Cas7f^6$ subunits (Fig. 2j). The backbone region of crRNA threads through the multiple copies of Cas7f (Fig. 2j). The 3' stem-loop is recognized by the Cas6f subunit (Fig. 2j). As observed in these complex structures, all of these anti-CRISPR proteins (AcrF1, AcrF2, and AcrF10) locate in positions partially overlapped with the binding sites of target DNAs (Fig. 2j). These anti-CRISPR proteins adopt a similar inhibition strategy by interfering the type I-F silencing machinery to recognize the target DNAs.

TYPE II CRISPR-CAS9: A HIGHLY EFFICIENT GENOME EDITING TOOL IMPLEMENTED IN A BROAD RANGE OF ORGANISMS

Cas9 is the best-characterized member of the class II CRISPR-Cas system, which has been widely used as a tool for genome engineering and gene expression control [56, 57]. Interestingly, the CRISPR-Cas9 gene locus encodes another noncoding RNA except for the crRNA, named *trans*-activating crRNA (tracrRNA) [58]. The sequence of the tracRNA is partially complementary to the repeat segment of the crRNA, forming a tracrRNA:crRNA duplex. Cas9 is activated through assembling with this tracrRNA:crRNA duplex to form a Cas9-crRNA-tracrRNA surveillance complex (Table 1) [58]. The tracrRNA:crRNA duplex can be engineered to a chimeric structure by connecting the 5'-end of the tracrRNA to the 3'-end of crRNA, named sgRNA. The Cas9-sgRNA two-component system simplifies the applications of CRISPR-Cas9 technology in genome editing. The accurate selection of target DNA depends on a PAM sequence, as well as the base pairing between the target DNA strands with the "seed" sequence within the guide segment of crRNA [58]. Cas9 proteins are widespread among bacterial kingdom, differing in both sequence and size. Cas9 protein found in *Streptococcus pyogenes* (SpCas9) is the most common and widely studied one.

Domain organization, structure and functional activities of the type II CRISPR-Cas9 system

During the past a few years, several structural studies on SpCas9 have been including the structures of apo-form SpCas9, SpCas9-sgRNA binary complex, and SpCas9-sgRNA-target DNA ternary complex [59-62]. SpCas9 adopts a bi-lobed architecture, comprising a recognition (REC) lobe and a nuclease (NUC) lobe (Fig. The REC lobe is composed of a Bridge helix motif (BH), a REC1 domain (Helical-I), REC2 (Helical-II) domain and a REC3 (Helical-III) domain (Fig. 3a-b). The NUC consists of a RuvC domain, a HNH domain, and a PAM-interacting (PI) domain (Fig.

3a-b). The REC domain is composed of multiple helix bundles, showing no structural similarity to any known proteins. Upon sgRNA loading, the REC lobe undergoes substantial conformational changes, inducing a formation of a central channel for accommodating the guide RNA-target DNA heteroduplex [60]. The target DNA binding also causes a series of conformational changes of SpCas9. The HNH catalytic domain moves toward the target DNA strand, and a modest shift is observed in the lobe [60]. This substrate induced fit mechanism ensures the optimal positioning of target DNA for cleavage.

In the SpCas9-sgRNA binary complex, the sgRNA displays an L-shape, comprising a crRNA and a tracrRNA connected by an artificial tetraloop (Fig. 3c-d). The crRNA is composed of a guide and repeat segments (Fig. 3d). The tracrRNA consists of an anti-repeat segment and three stem loops (Fig. 3d). The repeat segment of crRNA and the anti-repeat segment of tracrRNA form the repeat:anti-repeat duplex (Fig. 3d). In the absence of target DNA, only the 10-nt seed sequence in the guide segment of crRNA is observed, which adopts an A-form conformation (Fig. 3c-d). In the SpCas9-sgRNA-target DNA ternary complex, full 20-nt guide sequence is present, hybridizing with the target DNA strand to form the guide:target duplex (Fig. 3e). The guide:target and repeat:anti-repeat duplexes, as well as the DNA duplex containing the PAM sequence locate in the channel formed by the REC and NUC lobes (Fig. 3e). The stem loops of tracrRNA are solvent exposed, making extensive interactions with the REC1, RuvC and PI domains.

After target DNA unzipping, one of the DNA strands (target strand) hybridizes the guide region of crRNA to form the crRNA-DNA heteroduplex, whereas the other one (non-target DNA) is displaced. This represents a transient and pre-cleaved state, named R-loop conformation [63]. The formation of R-loop structure plays an role for placing each DNA strand for catalysis. Finally, Cas9 cleaves the target and non-target DNA strands using the HNH and RuvC nuclease domains, respectively, making a blunt double-stranded break (Fig. 3f). A near-atomic Cryo-EM structure of SpCas9-R-loop complex clearly shows that the displaced non-target DNA strand

protrudes into the active site of RuvC domain [63]. Another 5.2 Å Cryo-EM structure of SpCas9-sgRNA-target DNA captures a conformation of SpCas9 in which the HNH domain is close to the target DNA cleavage site [64]. These Cryo-EM structures strongly support the present understanding of the catalytic mechanism of type II CRISPR-Cas9 system. Besides SpCas9, lots of crystal structures of Cas9 homologs were determined, including Cas9 from *Actinomyces naeslundii*, *Campylobacter jejuni*, *Francisella novicida*, and *Staphylococcus aureus* [62, 65, 66]. These Cas9 homologs share similar domain composition and structural features. Though distinct sequence preference for PAMs and crRNA:tracrRNA scaffolds exist among these proteins, the mechanisms for substrate binding, PAM selection, target-DNA unzipping and cleavage are quite similar.

Structure of SpCas9 variants with broad PAM compatibility and enhanced specificity

SpCas9 specifically recognizes the 5'-NGG-3' PAM sequence through the PI domain. Two conserved residues (Arg^{1333} and Arg^{1335}) in the PI domain are inserted into the major groove of target DNA duplex forming hydrogen bonds with the two guanine bases in the PAM (Fig. 3e). Another two residues (Lys^{1107} and Ser^{1109}) in the same domain serving as a phosphate lock, recognize the phosphate group immediate upstream of the PAM and make a kink in the target DNA strand (Fig. 3e). Thus, Watson-Crick base pairs close to the PAM are separated. PAM recognition plays a role in preventing the CRISPR-Cas9 immune system to target the host's own genetic material, and facilitates the unzipping of PAM adjacent target DNA duplex. However, the specific PAM recognition pattern limits the applications of Cas9-mediated editing tool. To break this barrier, three SpCas9 variants were screened by utilizing a method named bacterial selection-based directed evolution, which could recognize the 5'-NGAN-3', 5'-NGNG-3', and 5'-NGCG-3' PAMs, respectively [67]. Structures of these Cas9 variants in complex with sgRNA and target DNAs containing

PAMs revealed that structural rearrangement occurs in the PAM region of target which allows the SpCas9 variants to form compact interactions with the altered PAM nucleotides through an induced fit mechanism. More recently, Liu's group screened a Cas9 variant (xCas9) through the phage-assisted continuous evolution [68]. The possesses the broadest PAM compatibility among Cas9 family proteins, and has a DNA specificity. Besides xCas9, several other SpCas9 variants with high fidelity and enhanced specificity also have been reported, including SpCas9-HF1, HypaCas9 and eSpCas9 [69-71]. These SpCas9 variants will improve the application of the CRISPR-Cas9 technology by reducing off-target cleavage and enhancing precision genome editing.

Structure of SpCas9 in complex with anti-CRISPR

Although CRISPR-Cas9 is the most powerful genome editing tool so far, and has been successfully applied in a broad range of organisms [31, 33]. The high off-target effect of CRISPR-Cas9 technology could not be ignored either [72]. AcrIIA4 is a currently identified anti-CRISPR protein, encoded by *L. monocytogenes* prophage. It was reported that AcrIIA4 could completely inhibit the activity of SpCas9 [73]. AcrIIA4 adopts a "triangle" fold, comprising three anti-parallel β-strands with three α-helices at one side (Fig. 3g). The structure of AcrIIA4 in complex with the sgRNA-loaded SpCas9 reveals that AcrIIA4 interacts with the REC, PI, and RuvC domains of SpCas9, sterically blocking the PAM binding site [74] (Fig. 3g). These studies facilitate the application of AcrIIA4 as an "off-swith" tool to control the activity of SpCas9.

TYPE V CRISPR-CPF1: AN ALTERNATIVE GENOME EDITING TOOL WITH HIGHER TARGETING EFFICIENCY

According to the distinct effector proteins, type V CRISPR-Cas system can be divided into three subtypes, including Cas12a-Cpf1 (V-A), Cas12b-C2c1 (V-B), and Cas12c-C2c3 (V-C). Cpf1 was first identified in 2015, which specifically cleaves both strands of target DNA [26]. Similar to Cas9, Cpf1 is a single RNA-guided endonuclease, showing robust genome editing activity in human cells [75]. However, the CRISPR-Cpf1 mediated DNA interference possesses four unique features. First, Cpf1 processes the pre-crRNA utilizing a divalent cation-independent endonuclease activity, and the mature crRNA does not require an additional trans-activating crRNA (tracrRNA) (Table 1) [76]. Second, Cpf1 recognizes T-rich PAMs and PAM-complementary nucleotides, whereas Cas9 recognizes G-rich PAMs (Table 1) [77]. Third, Cpf1 cleaves both strands of the target dsDNA with a staggered cut (4 or 5-nt 5' overhang), in contrast to the blunt ends produced by Cas9 [78]. Fourth, Cpf1 contains only a detectable endonuclease domain, RuvC, whereas Cas9 possesses another HNH endonuclease domain (Table 1) [78, 79]. Similar to Cpf1, C2c1 also recognizes the T-rich PAMs (Table 1) [25]. However, C2c1-mediated DNA cleavage requires both the crRNA and tracrRNA for activity (Table 1), and generates a staggered double-stranded break with 7-nt 5'overhang [25]. The cleavage activity of C2c1 is temperature-dependent, with 40-60℃ as the optimal cleavage temperature. This feature limits the utilization of C2c1 for genome editing application. C2c3 was reported with C2c1 at the same time, due to containing RuvC-like endonuclease domains [25]. However, the detailed domain composition, structure, and activity of C2c3 remain to be further investigated.

Domain organization and structure of CRISPR-Cpf1

The crystal structure of *Lachnospiraceae* bacterium ND2006 Cpf1 (LbCpf1) in complex with a 43-nt crRNA was first determined at a resolution of 2.38 Å [80]. Similar to the type II Cas9 effector, the LbCpf1 displays a bi-lobed architecture, consisting of a REC lobe and a NUC lobe (Fig. 4a). The REC lobe is composed of

two Helical domains, named Helical-I (REC1) and Helical-II (REC2) (Fig. 4a). The NUC lobe consists of an oligonucleotide-binding domain (OBD or WED), a looped-out helical domain (LHD or PI), a HLH domain (BH), a Nuc domain, and a RuvC domain (Fig. 4a). These domains enclose a triangle-like shape, with a large cavity at the center for placing the crRNA and the target dsDNA (Fig. 4b-d). In the complex structure of LbCpf1-crRNA, only the repeat sequence of the crRNA is well defined in the electron density, whereas the guide sequence is not observed (Fig. 4b). As shown in the complex structure, the repeat region of crRNA is highly distorted, adopting a stem-loop-like conformation (Fig. 4b). It makes extensive inter-molecular interactions with the WED and RuvC domains of LbCpf1 (Fig. 4b). It is worth to note that a $(Mg(H_2O)_6)^{2+}$ ion is observed in the center of the repeat region of crRNA, which functions as stabilizing its unique conformation (Fig. 4b).

Target DNA recognition and functional activity of CRISPR-Cpf1

Soon after the report of the complex structure of LbCpf1-crRNA, Nureki's group determined the crystal structure of *Acidaminococcus sp*. Cpf1 (AsCpf1) complexed with the crRNA and target DNA containing the 5'-TTTN-3' PAM at a resolution of 2.8 Å [77]. In this complex structure, we could clearly see that the crRNA-target DNA heteroduplex lies in the central cavity enclosed by the REC1, REC2, WED, and RuvC domains (Fig. 4c). The crRNA is composed of a 19-nt 5'-handle and a 24-nt guide segment. The PAM sequence is recognized by the REC1, WED and PI domains through both the base and shape readout mechanisms [77] (Fig. 4c). Combining the structural and biochemical information, the authors proposed that the Nuc domain is also an endonuclease domain, although showing no structural and sequence similarity to any identified nucleases. Subsequently, the crystal structure of *Francisella novicida* Cpf1 (FnCpf1)-R-loop complex was determined, which perfected the understanding of the process of recognition, unzipping, and cleavage of the target DNA [81, 82] (Fig. 4d). Putting all of these structures together, a working model for Cpf1 was

proposed. Firstly, the Cpf1-crRNA complex undergoes a conformational change upon target DNA binding to allow PAM scanning. The recognition of PAM induces the HLH domain to adopt a "flap-on" conformation, and the LKL helix to insert into the double-stranded DNA [81]. Then, the base pairs adjacent PAM are unzipped, which allows the hybridization of the target DNA strand with the crRNA. Finally, cleavage occurs on both strands of the target DNA to generate an overhang (Fig. 4e). In LbCpf1 and AsCpf1, mutations of the catalytic residues in the Nuc domain impact the cleavage of target DNA strand, whereas mutations of the catalytic residues in the RuvC domain disturb the cleavage of both strands of target DNA. Extensive mutational analysis of the putative active residues in FnCpf1 supports that a single active site located at the interface of the Nuc and RuvC domains takes in charge of cleaving both of the target and non-target DNA strands.

Structural plasticity of PAM recognition by CRISPR-Cpf1

Besides the optimal canonical 5'-TTTN-3' PAM, Cpf1 recognizes the suboptimal non-canonical PAMs, including 5'-TCTA-3', 5'-TCCA-3', and 5'-CCCA-3' [83]. However, LbCpf1 recognizes the canonical PAM more efficiently than these non-canonical PAMs. Structures of LbCpf1 complexed with these non-canonical PAMs containing DNAs were determined [83]. Both of the canonical and non-canonical PAM duplexes are located in a channel formed by REC1, WED, and PI domains. Structural superposition of these four structures indicated that the PI domain moves outward in these non-canonical PAMs containing structures, which enlarged the non-canonical PAM binding channel [83]. The structural plasticity of the PAM binding channel renders the ability of Cpf1 to recognize both of the canonical and non-canonical PAMs.

Domain organization, structure and functional activities of CRISPR-C2c1

The structures of C2c1-sgRNA binary complex and C2c1-sgRNA-target DNA ternary complex were also determined [84-86]. Similar to Cas9 and Cpf1, C2c1 contains two lobes, displaying an overall "Crab Claw" fold. The domain organization of C2c1 resembles that of Cpf1, except for lacking the PI domain (Fig. 5a). In addition, the REC2 domain is in close proximity to REC1 domain in Cpf1, whereas it is inserted between BH and RuvC-II motifs in C2c1 (Fig. 5a). The Nuc domain of C2c1 is divided into two parts by the RuvC-III motif, and shows low structural similarity to that in Cpf1 (Fig. 5a). The sgRNA observed in the complex structure is a chimeric tracrRNA:crRNA duplex, which is engineered by connecting the 5'-end of the crRNA to the 3'-end of tracrRNA (Fig. 5b). The sgRNA makes extensive interactions with the OBD, REC, RuvC and Nuc domains of C2c1 (Fig. 5b). The guide region of crRNA hybridizes with the target DNA strand, located in a channel enclosed by REC, BH, OBD, and RuvC domains, whereas the tracrRNA is solvent exposed (Fig. 5c). In contrast to the relaxed PAM recognition pattern of Cas9 and Cpf1, C2c1 recognizes PAM with stringent specificity [84]. In addition, the cleavage site on the target strand locates within the guide:target duplex for Cas9 and Cpf1, whereas C2c1 cleaves the target strand at a site outside the guide:target duplex [84] (Fig. 5d). Interestingly, the structure of C2c1-crRNA-extended target DNA reveals that both the target and non-target strand extensions are inserted into the same RuvC catalytic pocket [85]. This provides an evidence that type V-Cas12 effectors may cleave both the target and non-target DNA strands using a single active site. However, a precise catalytic mechanism of how a RuvC active site of Cas12 nucleases cleaves both target and non-target DNA strands independently needs further investigation.

TYPE VI CRISPR-CAS13: A TOOL FOR RNA EDITING AND RAPID NUCLEIC ACID DETECTION

Type VI CRISPR-Cas system is a solely dedicated RNA-guided RNA-targeting adaptive immune system, characterized by the single HEPN domain-containing effector Cas13. In addition to being a tool for RNA base editing, Cas13 has been developed as a platform for rapid nucleic acid detection, named SHERLOCK [87]. Cas13 possesses two RNase activities, which are mechanistically distinct from each other. First, it can cleave and process the pre-crRNA to generate mature crRNA. Second, it can recognize and degrade target RNA under the guide of crRNA. Upon target RNA binding and activation, Cas13 also possesses the ability to cleave unrelated RNA molecules without any complementarity to the guide region of crRNA. Up to now, four Cas13 family proteins have been identified, including Cas13a, Cas13b, Cas13c and Cas13d. Among these proteins, Cas13a (C2c2) is the one firstly identified and mostly studied.

Domain organization and structure of CRISPR-C2c2

To understand the mechanism of per-crRNA processing and crRNA-guided ssRNA degradation, the structures of apo-C2c2, C2c2-crRNA, and C2c2-crRNA-ssRNA were determined. In 2016, Wang's group determined the crystal structures of *Leptotrichia shahii* C2c2 (LshC2c2) and its complex with a crRNA at a resolution of 2.65 Å and 3.5 Å, respectively. [88]. Similar to other class II effectors, the structure of LshC2c2 also displays a bi-lobed architecture, consisting of a REC lobe and a NUC lobe (Fig. 6a-b). The REC lobe is composed of an N-terminal domain (NTD) and a Helical-1 domain (Fig. 6a). The NUC lobe consists of two HEPN domains, a Helical-2 domain and a Linker domain (Fig. 6a). The crRNA is composed of a 5'-handle region and a guide segment, located in a groove enclosed by the REC lobe and NUC lobe (Fig. 6b-c). The REC lobe mainly recognizes the 5'-handle region, whereas the NUC lobe interacts with the guide segment (Fig. 6b). LshC2c2 recognizes the 5'-handle in a sequence-specific manner. Both structure and sequence of the 5'-handle region are vital for the dual RNase activities of LshC2c2 [88]. The crRNA guide segment in the

structure of LshC2c2-crRNA is incomplete. The 5'-end of the guide adopts a U-shape, embedded deeply in a hole enclosed by the HEPN2 domain and the Linker domain (Fig. 6b). The central part of the guide is unstructured and missing, indicating that it may be flexible and exposed to the bulk solvent (Fig. 6b). The 3'-end of the guide sits at the concave surface of the NTD domain (Fig. 6b). The central part and 3'-end of the guide work as a seed sequence to hybridize with the target RNAs [89]. Upon crRNA binding, the Helical-2 domain of LshC2c2 undergoes a large conformational change, which moves toward the Linker domain and HEPN2 domain to enclose a crRNA binding groove [88].

Target RNA recognition and functional activity of CRIPSR-C2c2

Subsequently, Wang's group reported the crystal structure of *Leptotrichia buccalis* C2c2 (LbuC2c2) in complex with a crRNA and a target RNA, as well as a Cryo-EM structure of the LbuC2c2-crRNA complex at a resolution of 3.08 Å and 3.2 Å, respectively [90], in which nearly all nucleotides of the crRNA are observed (Fig. 6d). The target RNA forms 28 base pairs with the guide region of crRNA in the structure of LbuC2c2-crRNA-target RNA, which leaves two nucleotides (one at the 5'-end and the other one at the 3'-end) flipping out of the target-guide RNA duplex (Fig. 6d). The nucleotide at the 5'-end of target RNA inserts into the catalytic pocket of HEPN1 domain of a neighboring LbuC2c2 molecule (Fig. 6d). The 3'-end nucleotide locates in a groove enclosed by the NTD domain and the Helical-1 domain (Fig. 6d). In addition to base pairing with the guide region of crRNA, the target RNA also interacts with the HEPN1, Helical-2, and Linker domains of LbuC2c2 (Fig. 6d). Upon target RNA binding, significantly conformational rearrangements occur in both LbuC2c2 and crRNA, which makes a suitable binding groove for the crRNA-target RNA duplex [90]. Meanwhile, the guide region of the crRNA changes its conformation from multiple turns to a regular A-form helix [90].

C2c2 has two separate catalytic centers for the dual RNase activities. The Helical-1 domain and HEPN2 domain are found to be responsible for pre-crRNA processing in

LshC2c2 and LbuC2c2, respectively [90]. While HEPN1 domain together with HEPN2 domain play the key role in target and collateral degradation [90]. The formation of guide-target RNA duplex causes the HEPN1 domain to move toward the HEPN2 domain. Two catalytic residues on each HEPN domain are brought together to create a composite HEPN catalytic center. The activated C2c2 can cleave any exposed ssRNAs, including the target RNA extending from the guide-target complementary region and the free RNAs in solution (Fig. 6e).

CONNECTION BETWEEN THE STRUCTURES AND POTENTIAL GENOME-EDITING USAGE

In biology, an important insight is that structure determines the function. Studying the structural information can better understand the functional activities of CRISPR-Cas systems, and promote the genome-editing application. Based on the 3D structures, Cas9 and Cpf1 variants with altered PAM specificity were designed. These variants enabled editing of gene sites that were not targeted by wild-type (wt) Cas9 or Cpf1 in human cells. The structural information can also guide the sgRNA design. For Cas9, truncated sgRNAs with 17 or 18 nt guide sequence showed much reduced off-target activity in human cells without reducing on-target genome editing efficiency [91]. Another strategy that can reduce off-target activity of Cas9 is mutating amino acid residues in charge of stabilizing the R-loop structure. In accordance with this principle, four Cas9 variants with high-fidelity and enhanced specificity were designed, including eSpCas9, SpCas9-HF, HypaCas9 and evoCas9 [69, 70, 92]. Especially, the evoCas9 showed a 79-fold higher fidelity than wtCas9 [92]. As mentioned above, Cas9 possesses two catalytic domains, HNH and RuvC. Inactivation of one of the catalytic residues generates a Cas9 variant (nickase) that can only cleave either the target DNA strand or non-target DNA strand. Cas9 nickases showed reduced off-target activity and facilitated overhang-based cloning [93, 94]. More recently, structure-based inactivation of Cas9 (dCas9), Cpf1 (dCpf1) and C2c2

(dC2c2) proteins were widely used, which were fused to specific effector proteins for base-specific genome editing [95].

CONCLUDING REMARKS

The CRISPR-Cas adaptive immune systems found in prokaryotes are thought to be one of the most significant discoveries in life science. Given the powerful applications in health care and agriculture, CRISPR-Cas systems have attracted much attention in recent years as a genome engineering tool. The recent elucidation of the biochemical researches, as well as structural studies of several Cas proteins and their complexes with nucleic acids highlight the understanding of the CRISPR-Cas genetic silencing machinery. In this review, we focus on the recent advances in the structural studies of these CRISPR-Cas mediated genome editing tools. The architecture of the type I Cascade complex shows similar to the type III Csm/Cmr complex, supporting the hypothesis that these two types of CRISPR immune systems may evolve from a common ancestor [96]. Especially, both of them contain a crRNA-binding platform composed of multiple copies of Cas7-family proteins. The Class II CRISPR-Cas effectors, such as type II Cas9, type V Cas12, and type VI Cas13, share low sequence similarity and adopt distinct domain organizations. A phylogenetic study suggests that these types of effectors may evolve independently from distinct members of the TnpB family nucleases [25]. The Class II CRISPR-Cas effectors recognize target nucleic acids dependently of the PAM sequence or 3'-PFS (protospacer-flank site). Cas9 and Cas12 recognize the PAMs in a sequence-specific manner, whereas Cas13 interacts with the 3'-PFS non-specifically (Table 1). In contrast to Cas9, which interacts only with the PAM in the non-target strand, Cas12 recognizes double DNA strands at the PAM region (Table 1). Cas13 cleaves both target and collateral RNAs in a non-specific manner (Table 1). Conversely, Cas9 and Cas12 cleave target DNA or RNA at specific sites (Table 1).

In the past decade, extensive research has built a framework for our understanding of the composition, structure and functional activities of distinct types of CRISPR-Cas systems. However, how to apply CRISPR-based technology to fulfil efficient and precise genome engineering still needs further exploration. We proposed that scientists need to devote more research enthusiasm and strength into at least two research fields. First is identifying novel proteins or small molecules that regulate the function of CRISPR-Cas machinery and understanding their mechanism of action. During the past a few years, scientists found that viruses and mobile genetic elements encode a type of proteins, named anti-CRISPR, which can destroy the highly prevalent CRISPR-Cas immune systems of prokaryotes. Lots of anti-CRISPR proteins targeting the type I Cascade complex and type II Cas9 have been identified [97-99]. These proteins are diverse in terms of sequence and structure, inhibiting the target CRISPR-Cas effectors with different strategies. In 2017, Huang's group determined the first structure of the Class II anti-CRISPR protein AcrIIA4 in complex with SpCas9 and a sgRNA, which provides a structural basis for developing tools to eliminate the genome-wide off-target activity of SpCas9 [74]. In contrast to repress the activity of CRISPR-Cas effectors, some other accessory proteins, such as Csx28 and WYL1, are found that could enhance the Cas13-mediated RNA interference [100, 101]. More recently, Zhu's group identified two small molecules (VE-822 and AZD-7762) through an unbiased drug-selection-based platform [102]. These two compounds can stimulate CRISPR-Cpf1-mediated precise genome editing. Second is structure based engineering or continuous directed evolution of the CRISPR-Cas effectors to improve their applications in genome editing, transcriptional activation, or clinically viral nucleic acid detection. Lots of scientists has succeeded in engineering the Cas9 and Cpf1 with altered PAM specificities to increase the genome targeting range [68, 103-105]. These studies will boost the use of CRISPR-Cas9/Cpf1 systems for applications in genome editing.

A close connection between the structural studies and potential genome-editing usage of CRISPR-Cas effector proteins exists. Learning the structural information

facilitates us to understand the domain architecture and conformational activation of these effectors, and improves the appplications for genome editing. Many strategies have been employed to overcome the limitations of off-target effects and stringent requirement for the protospacer adjacent motif (PAM) sequence. Structure-based engineering of the amino acid residues neighboring the PAM binding region yields lots of Cas9 and Cpf1 variants with expanded targeting space. A strategy that introduces non-base-specific interactions to compensate base-specific interaction is applied. In accordance with this principle, a SpCas9 variant SpCas9-NG was designed, which recognized relaxed NG PAMs [106]. In addition, structure-based engineering of the amino acid residues in charge of stabilizing the R-loop structure led to the discovery of evoCas9, which displayed a 79-fold higher fidelity than wtCas9 [92]. More recently, Cas9 and Cas12-directed DNA base editors, and Cas13-directed RNA base editor are developed, utilizing catalytically inactivated CRISPR-Cas effector proteins together with other enzymes [95]. Taken together, humans are gradually mastering the ability to efficiently and precisely edit genome of cells.

FUNDING

This work was supported by the National Natural Science Foundation of China [31422014, 31450001 and 21300605 to ZH.]

REFERENCES

1. Koskella, B, Brockhurst, MA. Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol Rev*. 2014; **38**(5): 916-31.

2. Stern, A, Sorek, R. The phage-host arms race: shaping the evolution of microbes. *Bioessays*. 2011; **33**(1): 43-51.

3. Oliveira, PH, Touchon, M, Rocha, EP. Regulation of genetic flux between bacteria by restriction-modification systems. *Proc Natl Acad Sci U S A*. 2016; **113**(20): 5658-63.

4. Marraffini, LA, Sontheimer, EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*. 2010; **11**(3): 181-90.

5. Bolotin, A, Quinquis, B, Sorokin, A*, et al.* Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005; **151**(Pt 8): 2551-61.

6. Jansen, R, Embden, JD, Gaastra, W*, et al.* Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002; **43**(6): 1565-75.

7. Pul, U, Wurm, R, Arslan, Z*, et al.* Identification and characterization of E. coli CRISPR-cas promoters and their silencing by H-NS. *Mol Microbiol*. 2010; **75**(6): 1495-512.

8. van der Oost, J, Jore, MM, Westra, ER*, et al.* CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci*. 2009; **34**(8): 401-7.

9. Wright, AV, Nunez, JK, Doudna, JA. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*. 2016; **164**(1-2): 29-44.

10. Fineran, PC, Charpentier, E. Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology*. 2012; **434**(2): 202-9.

11. Jackson, SA, McKenzie, RE, Fagerlund, RD*, et al.* CRISPR-Cas: Adapting to change. *Science*. 2017; **356**(6333).

12. Xiao, Y, Ng, S, Nam, KH*, et al.* How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration. *Nature*. 2017; **550**(7674): 137-41.

13. Wright, AV, Liu, JJ, Knott, GJ*, et al.* Structures of the CRISPR genome integration complex. *Science*. 2017; **357**(6356): 1113-8.

14. Deltcheva, E, Chylinski, K, Sharma, CM*, et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*. 2011; **471**(7340): 602-7.

15. Sashital, DG, Jinek, M, Doudna, JA. An RNA-induced conformational change required for CRISPR RNA cleavage by the endoribonuclease Cse3. *Nat Struct Mol Biol*. 2011; **18**(6): 680-7.

16. Garneau, JE, Dupuis, ME, Villion, M*, et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 2010; **468**(7320): 67-71.

17. Makarova, KS, Wolf, YI, Alkhnbashi, OS*, et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015; **13**(11): 722-36.

18. Koonin, EV, Makarova, KS, Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*. 2017; **37**: 67-78.

19. Westra, ER, van Erp, PB, Kunne, T*, et al.* CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell*. 2012; **46**(5): 595-605.

20. Barrangou, R, Doudna, JA. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol*. 2016; **34**(9): 933-41.

21. Hsu, PD, Lander, ES, Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell*. 2014; **157**(6): 1262-78.

22. Samai, P, Pyenson, N, Jiang, W*, et al.* Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity. *Cell*. 2015; **161**(5): 1164-74.

23. Liu, TY, Iavarone, AT, Doudna, JA. RNA and DNA Targeting by a Reconstituted Thermus thermophilus Type III-A CRISPR-Cas System. *PLoS One*. 2017; **12**(1): e0170552.

24. Pyenson, NC, Marraffini, LA. Type III CRISPR-Cas systems: when DNA cleavage just isn't enough. *Curr Opin Microbiol*. 2017; **37**: 150-4.

25. Shmakov, S, Abudayyeh, OO, Makarova, KS*, et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell*. 2015; **60**(3): 385-97.

26. Zetsche, B, Gootenberg, JS, Abudayyeh, OO*, et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015; **163**(3): 759-71.

27. Cheng, F, Gong, L, Zhao, D*, et al.* Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J Genet Genomics*. 2017; **44**(11): 541-8.

28. Pyne, ME, Bruder, MR, Moo-Young, M*, et al.* Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci Rep*. 2016; **6**: 25666.

29. Zhang, J, Zong, W, Hong, W*, et al.* Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metab Eng*. 2018; **47**: 49-59.

30. Friedland, AE, Tzur, YB, Esvelt, KM*, et al.* Heritable genome editing in C. elegans via a CRISPR-Cas9 system. *Nat Methods*. 2013; **10**(8): 741-3.

31. Jiang, W, Bikard, D, Cox, D*, et al.* RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013; **31**(3): 233-9.

32. Hwang, WY, Fu, Y, Reyon, D*, et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol*. 2013; **31**(3): 227-9.

33. Cho, SW, Kim, S, Kim, JM*, et al.* Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol*. 2013; **31**(3): 230-2.

34. Fagerlund, RD, Staals, RH, Fineran, PC. The Cpf1 CRISPR-Cas protein expands genome-editing tools. *Genome Biol*. 2015; **16**: 251.

35. Kleinstiver, BP, Tsai, SQ, Prew, MS*, et al.* Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol*. 2016; **34**(8): 869-74.

36. Kim, H, Kim, ST, Ryu, J*, et al.* CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat Commun*. 2017; **8**: 14406.

37. East-Seletsky, A, O'Connell, MR, Knight, SC*, et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*. 2016; **538**(7624): 270-3.

38. Brouns, SJ, Jore, MM, Lundgren, M*, et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 2008; **321**(5891): 960-4.

39. Jore, MM, Lundgren, M, van Duijn, E*, et al.* Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*. 2011; **18**(5): 529-36.

40. Westra, ER, Semenova, E, Datsenko, KA*, et al.* Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet*. 2013; **9**(9): e1003742.

41. Xiao, Y, Luo, M, Hayes, RP*, et al.* Structure Basis for Directional R-loop Formation and Substrate Handover Mechanisms in Type I CRISPR-Cas System. *Cell*. 2017; **170**(1): 48-60 e11.

42. Hochstrasser, ML, Taylor, DW, Bhat, P*, et al.* CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A*. 2014; **111**(18): 6618-23.

43. Luo, ML, Mullis, AS, Leenay, RT*, et al.* Repurposing endogenous type I CRISPR-Cas systems for programmable gene repression. *Nucleic Acids Res*. 2015; **43**(1): 674-81.

44. Caliando, BJ, Voigt, CA. Targeted DNA degradation using a CRISPR device stably carried in the host genome. *Nat Commun*. 2015; **6**: 6989.

45. Jackson, RN, Golden, SM, van Erp, PB*, et al.* Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from Escherichia coli. *Science*. 2014; **345**(6203): 1473-9.

46. Zhao, H, Sheng, G, Wang, J*, et al.* Crystal structure of the RNA-guided immune surveillance Cascade complex in Escherichia coli. *Nature*. 2014; **515**(7525): 147-50.

47. Niewoehner, O, Jinek, M, Doudna, JA. Evolution of CRISPR RNA recognition and processing by Cas6 endonucleases. *Nucleic Acids Res*. 2014; **42**(2): 1341-53.

48. Mulepati, S, Heroux, A, Bailey, S. Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science*. 2014; **345**(6203): 1479-84.

49. Hayes, RP, Xiao, Y, Ding, F*, et al.* Structural basis for promiscuous PAM recognition in type I-E Cascade from E. coli. *Nature*. 2016; **530**(7591): 499-503.

50. Xiao, Y, Luo, M, Dolan, AE*, et al.* Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science*. 2018; **361**(6397).

51. Bondy-Denomy, J, Pawluk, A, Maxwell, KL*, et al.* Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*. 2013; **493**(7432): 429-32.

52. Pawluk, A, Staals, RH, Taylor, C*, et al.* Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol*. 2016; **1**(8): 16085.

53. Chowdhury, S, Carter, J, Rollins, MF*, et al.* Structure Reveals Mechanisms of Viral Suppressors that Intercept a CRISPR RNA-Guided Surveillance Complex. *Cell*. 2017; **169**(1): 47-57 e11.

54. Guo, TW, Bartesaghi, A, Yang, H*, et al.* Cryo-EM Structures Reveal Mechanism and Inhibition of DNA Targeting by a CRISPR-Cas Surveillance Complex. *Cell*. 2017; **171**(2): 414-26 e12.

55. Peng, R, Xu, Y, Zhu, T*, et al.* Alternate binding modes of anti-CRISPR viral suppressors AcrF1/2 to Csy surveillance complex revealed by cryo-EM structures. *Cell Res*. 2017; **27**(7): 853-64.

56. Larson, MH, Gilbert, LA, Wang, X*, et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc*. 2013; **8**(11): 2180-96.

57. Cong, L, Ran, FA, Cox, D*, et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science*. 2013; **339**(6121): 819-23.

58. Jinek, M, Chylinski, K, Fonfara, I*, et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012; **337**(6096): 816-21.

59. Jiang, F, Zhou, K, Ma, L*, et al.* STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*. 2015; **348**(6242): 1477-81.

60. Anders, C, Niewoehner, O, Duerst, A*, et al.* Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature*. 2014; **513**(7519): 569-73.

61. Nishimasu, H, Cong, L, Yan, WX*, et al.* Crystal Structure of Staphylococcus aureus Cas9. *Cell*. 2015; **162**(5): 1113-26.

62. Jinek, M, Jiang, F, Taylor, DW*, et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*. 2014; **343**(6176): 1247997.

63. Jiang, F, Taylor, DW, Chen, JS*, et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science*. 2016; **351**(6275): 867-71.

64. Huai, C, Li, G, Yao, R*, et al.* Structural insights into DNA cleavage activation of CRISPR-Cas9 system. *Nat Commun*. 2017; **8**(1): 1375.

65. Yamada, M, Watanabe, Y, Gootenberg, JS*, et al.* Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Mol Cell*. 2017; **65**(6): 1109-21 e3.

66. Hirano, H, Gootenberg, JS, Horii, T*, et al.* Structure and Engineering of Francisella novicida Cas9. *Cell*. 2016; **164**(5): 950-61.

67. Kleinstiver, BP, Prew, MS, Tsai, SQ*, et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015; **523**(7561): 481-5.

68. Hu, JH, Miller, SM, Geurts, MH*, et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*. 2018; **556**(7699): 57-63.

69. Slaymaker, IM, Gao, L, Zetsche, B*, et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science*. 2016; **351**(6268): 84-8.

70. Chen, JS, Dagdas, YS, Kleinstiver, BP*, et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature*. 2017; **550**(7676): 407-10.

71. Kleinstiver, BP, Pattanayak, V, Prew, MS*, et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*. 2016; **529**(7587): 490-5.

72. Kim, D, Bae, S, Park, J*, et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*. 2015; **12**(3): 237-43, 1 p following 43.

73. Rauch, BJ, Silvis, MR, Hultquist, JF*, et al.* Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell*. 2017; **168**(1-2): 150-8 e10.

74. Dong, Guo, M, Wang, S*, et al.* Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature*. 2017; **546**(7658): 436-9.

75. Kim, D, Kim, J, Hur, JK*, et al.* Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol*. 2016; **34**(8): 863-8.

76. Fonfara, I, Richter, H, Bratovic, M*, et al.* The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*. 2016; **532**(7600): 517-21.

77. Yamano, T, Nishimasu, H, Zetsche, B*, et al.* Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell*. 2016; **165**(4): 949-62.

78. Garcia-Doval, C, Jinek, M. Molecular architectures and mechanisms of Class 2 CRISPR-associated nucleases. *Curr Opin Struct Biol*. 2017; **47**: 157-66.

79. Stella, S, Alcon, P, Montoya, G. Class 2 CRISPR-Cas RNA-guided endonucleases: Swiss Army knives of genome editing. *Nat Struct Mol Biol*. 2017; **24**(11): 882-92.

80. Dong, D, Ren, K, Qiu, X*, et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*. 2016; **532**(7600): 522-6.

81. Stella, S, Alcon, P, Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*. 2017; **546**(7659): 559-63.

82. Swarts, DC, van der Oost, J, Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol Cell*. 2017; **66**(2): 221-33 e4.

83. Yamano, T, Zetsche, B, Ishitani, R*, et al.* Structural Basis for the Canonical and Non-canonical PAM Recognition by CRISPR-Cpf1. *Mol Cell*. 2017; **67**(4): 633-45 e3.

84. Wu, D, Guan, X, Zhu, Y*, et al.* Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA. *Cell Res*. 2017; **27**(5): 705-8.

85. Yang, H, Gao, P, Rajashankar, KR*, et al.* PAM-Dependent Target DNA Recognition and Cleavage by C2c1 CRISPR-Cas Endonuclease. *Cell*. 2016; **167**(7): 1814-28 e12.

86. Liu, L, Chen, P, Wang, M*, et al.* C2c1-sgRNA Complex Structure Reveals RNA-Guided DNA Cleavage Mechanism. *Mol Cell*. 2017; **65**(2): 310-22.

87. Myhrvold, C, Freije, CA, Gootenberg, JS*, et al.* Field-deployable viral diagnostics using CRISPR-Cas13. *Science*. 2018; **360**(6387): 444-8.

88. Liu, L, Li, X, Wang, J*, et al.* Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell*. 2017; **168**(1-2): 121-34 e12.

89. Abudayyeh, OO, Gootenberg, JS, Konermann, S*, et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science*. 2016; **353**(6299): aaf5573.

90. Liu, L, Li, X, Ma, J*, et al.* The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell*. 2017; **170**(4): 714-26 e10.

91. Fu, Y, Sander, JD, Reyon, D*, et al.* Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol*. 2014; **32**(3): 279-84.

92. Casini, A, Olivieri, M, Petris, G*, et al.* A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol*. 2018; **36**(3): 265-71.

93. Ran, FA, Hsu, PD, Wright, J*, et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013; **8**(11): 2281-308.

94. Ran, FA, Hsu, PD, Lin, CY*, et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*. 2013; **154**(6): 1380-9.

95. Rees, HA, Liu, DR. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat Rev Genet*. 2018; **19**(12): 770-88.

96. Jackson, RN, Wiedenheft, B. A Conserved Structural Chassis for Mounting Versatile CRISPR RNA-Guided Immune Responses. *Mol Cell*. 2015; **58**(5): 722-8.

97. Zhu, Y, Zhang, F, Huang, Z. Structural insights into the inactivation of CRISPR-Cas systems by diverse anti-CRISPR proteins. *BMC Biol*. 2018; **16**(1): 32.

98. Pawluk, A, Davidson, AR, Maxwell, KL. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol*. 2018; **16**(1): 12-7.

99. Borges, AL, Davidson, AR, Bondy-Denomy, J. The Discovery, Mechanisms, and Evolutionary Impact of Anti-CRISPRs. *Annu Rev Virol*. 2017; **4**(1): 37-59.

100. Smargon, AA, Cox, DBT, Pyzocha, NK*, et al.* Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Mol Cell*. 2017; **65**(4): 618-30 e7.

101. Yan, WX, Chong, S, Zhang, H*, et al.* Cas13d Is a Compact RNA-Targeting Type VI CRISPR Effector Positively Modulated by a WYL-Domain-Containing Accessory Protein. *Mol Cell*. 2018; **70**(2): 327-39 e5.

102. Ma, X, Chen, X, Jin, Y*, et al.* Small molecules promote CRISPR-Cpf1-mediated genome editing in human pluripotent stem cells. *Nat Commun*. 2018; **9**(1): 1303.

103. Hirano, S, Nishimasu, H, Ishitani, R*, et al.* Structural Basis for the Altered PAM Specificities of Engineered CRISPR-Cas9. *Mol Cell*. 2016; **61**(6): 886-94.

104. Gao, L, Cox, DBT, Yan, WX*, et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol*. 2017; **35**(8): 789-92.

105. Anders, C, Bargsten, K, Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell*. 2016; **61**(6): 895-902.

106. Nishimasu, H, Shi, X, Ishiguro, S*, et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*. 2018; **361**(6408): 1259-62.

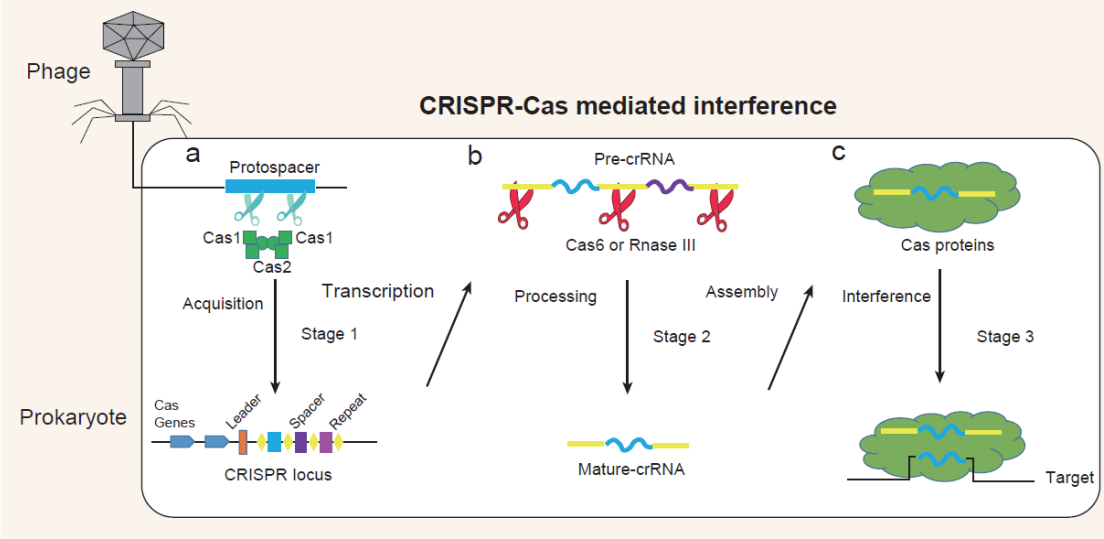

Figure 1. A cartoon depicting the general organization of a CRISPR-Cas locus and three stages of CRISPR-Cas activity. (a) In the stage of spacer sequence acquisition, Cas1 and Cas2 heterocomplex recognize invading DNA (protospacer), and insert it into CRISPR array as a spacer sequence via a cut-and-paste mechanism. (b) In the stage of RNA (crRNA) biogenesis, Cas6 or RNase III family nuclease processes the long transcript (pre-crRNA) from CRISPR locus to a mature crRNA. (c) In the stage of RNA-guided interference, mature crRNAs associate with Cas proteins to form a surveillance complexes, which recognize and cleave invading nucleic acids.

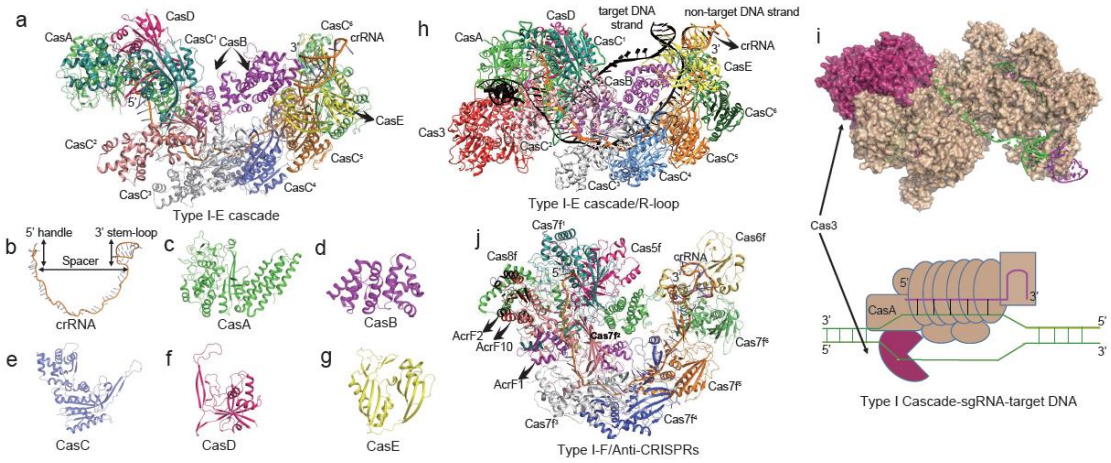

Figure 2. Structures of type I Cascade and its complexes. (a) Crystal structure of the RNA-guided type I-E CRISPR surveillance complex (PDB: 4U7U). The crRNA and five Cas proteins (CasA-E) are labeled. (b-g) Enlarged view of the 61-nt crRNA, as well as five Cas proteins CasA-E. (h) Cryo-EM structure of the type I-E Cascade/R-loop/Cas3 from *Thermobifida fusca* (PDB: 6C66). Cas3 and target dsDNA are colored red and black, respectively. The other subunits are colored the same as Fig. 2a. (i) Cartoon show of the working model of type I Cascade (PDB: 6U66). (j) Cryo-EM structure of type I-F CRISPR surveillance complex bound with anti-CRISPRs, AcrF1, AcrF2, and AcrF10 (PDB: 5UZ9 and 6B48). AcrF1, AcrF2, and AcrF10 are labeled and colored magenta, black, and red, respectively.

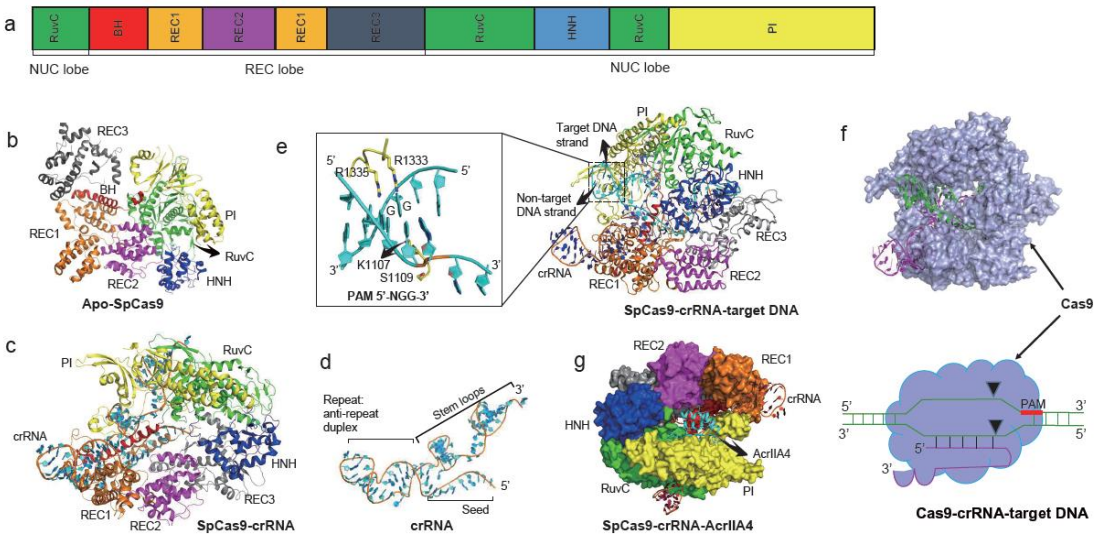

Figure 3. Domain organization and structures of Cas9 and its complexes. (a) Domain organization of Cas9. (b) Crystal structure of Apo-SpCas9 (PDB: 4CMP). (c) Crystal structure of SpCas9 in complex with sgRNA (PDB: 4ZT0). (d) Cartoon show of the sgRNA observed in the structure of SpCas9-sgRNA binary complex. (e) Crystal structure of SpCas9 in complex with sgRNA and target DNA (5'-NGG-3' PAM) (PDB: 4UN3). (f) Cartoon show of the working model of type II-A SpCas9 (PDB: 4UN3). (g) The complex structure of SpCas9-sgRNA-AcrIIA4 (PDB: 5XBL). sgRNA and AcrIIA4 are colored orange and cyan, respectively. In panels (b-c, e-f), SpCas9 domains are colored the same as panel (a).

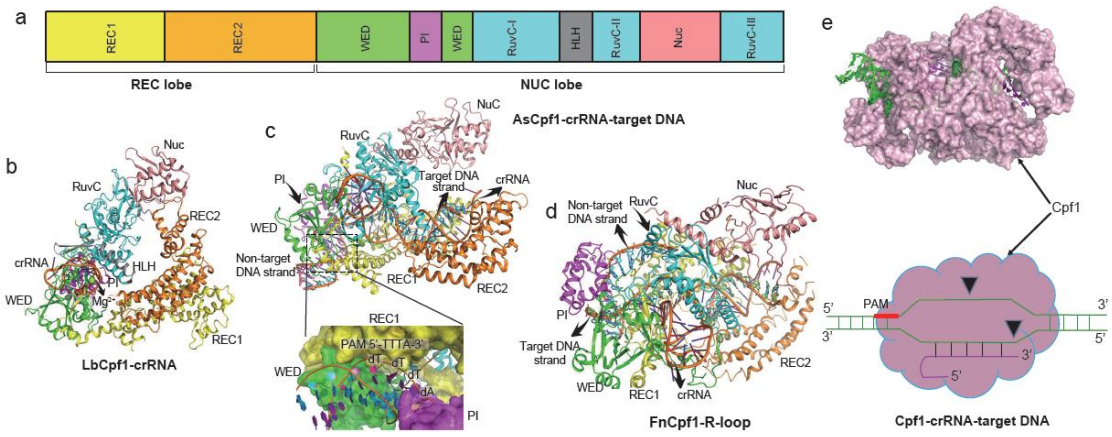

Figure 4. Domain organization and structures of Cpf1 and its complex with nucleic acids. (a) Domain organization of Cpf1. (b) Crystal structure of LbCpf1-crRNA binary complex (PDB: 5ID6). (b) Crystal structure of AsCpf1 in complex with crRNA and target DNA (5'-TTTA-3' PAM) (PDB: 5B43). (d) Crystal structure of FnCpf1-R-loop complex (PDB: 5MGA). In panels (b-d), Cpf1 domains are colored the same as panel (a). (e) Cartoon show of the working model of type V-A Cpf1 (PDB: 5B43).

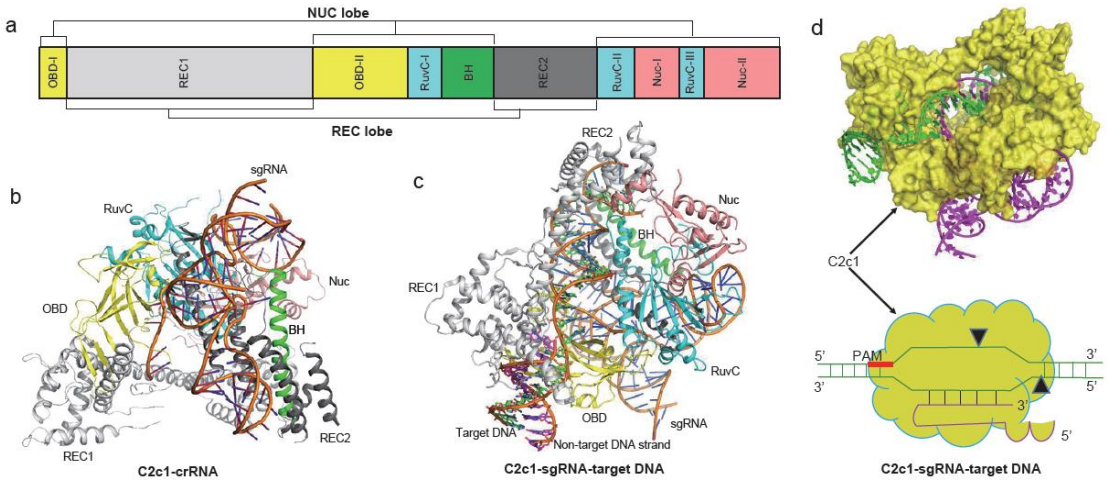

Figure 5. Domain organization and structures of C2c1 and its complex with nucleic acids. (a) Domain organization of C2c1. (b) Crystal structure of AacC2c1-sgRNA binary complex (PDB: 5U34). (c) Crystal structure of AacC2c1 in complex with crRNA and target DNA (5'-TTC-3' PAM) (PDB: 5B43). In panels (b,c), C2c1 domains are colored the same as panel (a). (d) Cartoon show of the working model of type V-B C2c1 (PDB: 5B43).

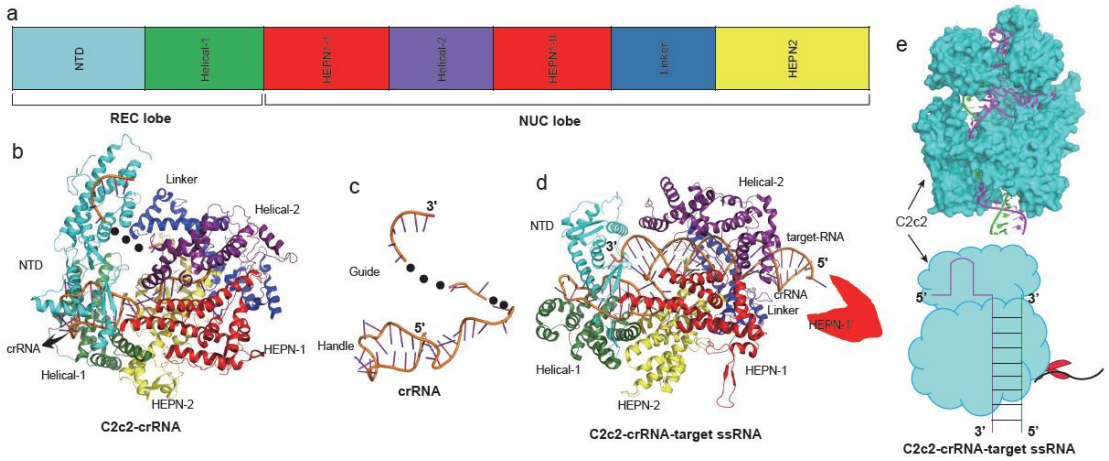

Figure 6. Domain organization and structures of C2c2 and its complex with nucleic acids. (a) Domain organization of C2c2. (b) Crystal structure of LshC2c2-crRNA binary complex (PDB: 5WTK). (c) Cartoon show of the crRNA observed in the structure of C2c2-crRNA binary complex. (d) Crystal structure of LbuC2c2-crRNA-target RNA ternary complex (PDB: 5XWP). In panels (b,d), C2c2 domains are colored the same as panel (a). (e) Cartoon show of the working model of type VI-A C2c2 (PDB: 5XWP).

Table 1. Comparison of distinct types of CRISPR-Cas effectors

	Type I-E Cascade	Type II Cas9	Type V-A Cfp1	Type V-B C2c1	Type VI C2c2
Protein Composition	multiple subunits	single subunit	single subunit	single subunit	single subunit
Pre-crRNA processing	mediated by accessory protein	mediated by accessory protein	self-processing	mediated by accessory protein	self-processing
RNA Composition	crRNA	tracrRNA/ crRNA	crRNA	tracrRNA/ crRNA	crRNA
Substrate	dsDNA	dsDNA	dsDNA	dsDNA	ssRNA
PAM	promiscuous PAMs	G-rich	T-rich	T-rich	Non-G PFS
PAM recognition pattern	both DNA strands	NT strand	both DNA strands	both DNA strands	T strand
Length of guide-target duplex	32	20	20	20	24
Catalytic domain	HD (Cas3)	HNH and RuvC	RuvC-Nuc	RuvC-Nuc	2*HEPN