

# Structural basis of stringent PAM recognition by CRISPR-C2c1 in complex with sgRNA

*Cell Research* (2017) 27:705-708. doi:10.1038/cr.2017.46; published online 4 April 2017

**Dear Editor,**

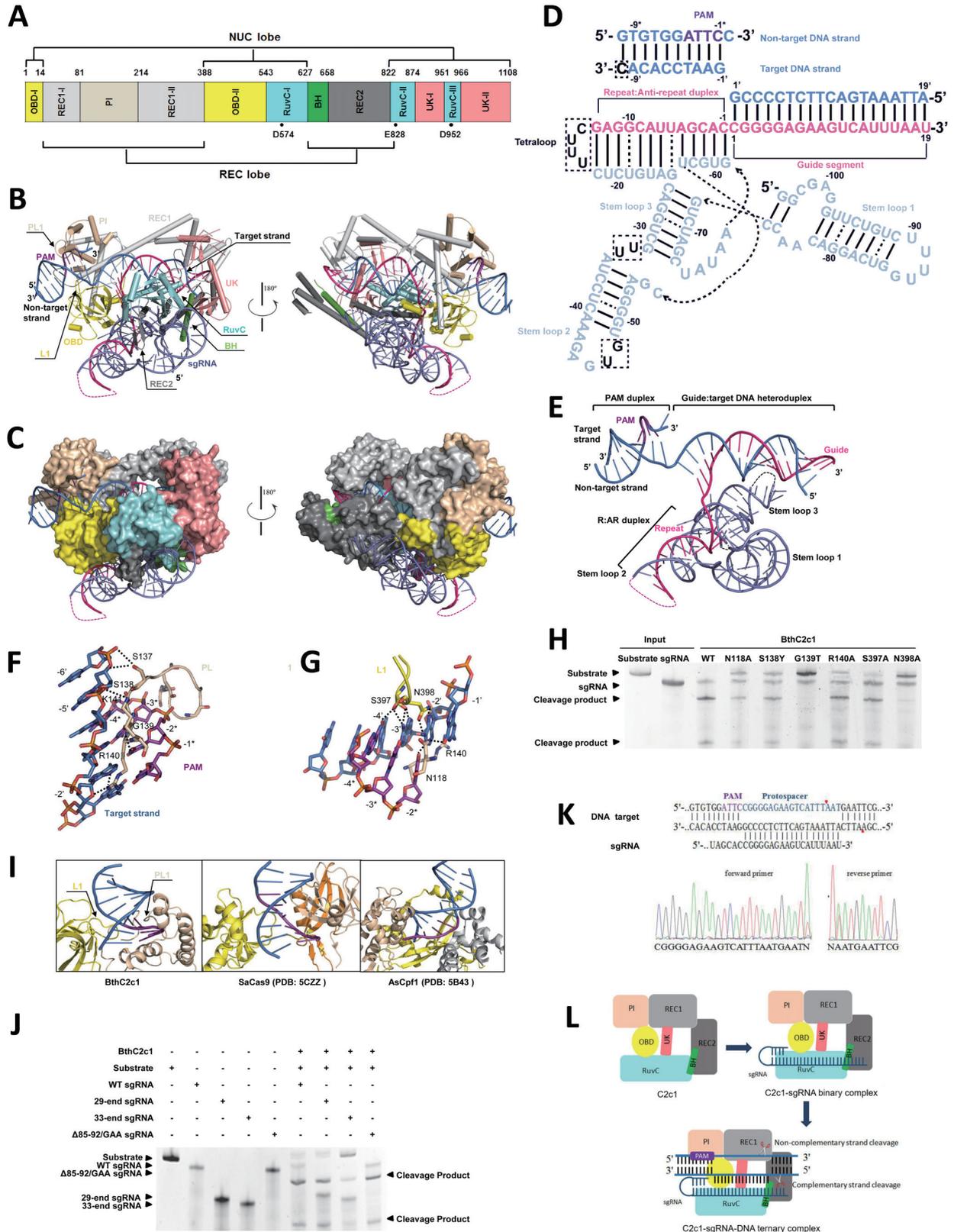
The clustered regularly interspaced short palindromic repeat (CRISPR)-Cas systems function as adaptive immune systems in bacteria [1, 2], which are used to defend against phages and invading nucleic acids. The CRISPR-Cas systems are broadly grouped into two classes: Class 1 systems contain a multi-subunit protein complex, whereas Class 2 systems use a single effector protein, as exemplified by the well-studied Cas9 [3]. Cas9 is an RNA-guided endonuclease, which targets and cleaves DNA bearing complementary sequences to the guide RNA. Protospacer adjacent motif (PAM) recognition by Cas9 and crRNA:tracrRNA complex is a critical prerequisite for substrate DNA melting and guide RNA:target DNA heteroduplex formation [4, 5]. Both catalytically active and inactive Cas9, combined with a single-strand guide RNA (sgRNA), have been widely used as programmable systems for various genetic manipulations [6, 7].

Recently, a Class 2 CRISPR effector protein, C2c1 (classified as type V-B) [8], has been identified to cleave DNA under the guide of crRNA:tracrRNA, distinct from a type V-A effector protein Cpf1 that only requires a single crRNA [9]. Furthermore, C2c1 and Cpf1 recognize different PAM sequences. Like Cpf1, C2c1 contains a conserved RuvC endonuclease domain, though it harbors a second endonuclease domain that is not well defined by sequence. C2c1 has been proved to be endonuclease-active in human cell lysates. The mechanism underlying C2c1-mediated cleavage remains elusive. To reveal how C2c1 recognizes sgRNA and target DNA, we determined the crystal structure of *Bacillus thermoamylovorans* C2c1 (BthC2c1) in complex with a 123-nt sgRNA containing nearly full-length crRNA and tracrRNA, a 28-nt target DNA, and a 12-nt non-target DNA at 2.70 Å resolution by the single-wavelength anomalous dispersion method (Figure 1A-1C and Supplementary information, Table S1). The overall structure of the BthC2c1-sgRNA-DNA ternary complex is a bi-lobed architecture composed of an  $\alpha$ -helical recognition (REC) lobe and a nuclease (NUC)

lobe (Figure 1B). The REC lobe consists of a PAM-interacting (PI) domain, a REC1 domain, a REC2 domain, and a long  $\alpha$  helix referred to as the bridge helix (BH) (Figure 1A-1B). The NUC lobe contains an OBD domain, a RuvC domain, and a domain with unknown functions (termed “UK” domain) (Figure 1A-1B). The RuvC domain in the NUC lobe, composed by three split RuvC motifs (RuvC I-III), interfaces with the REC2 domain in the REC lobe to form a positively charged surface that interacts with the 3' tail of the sgRNA (Figure 1B). The interaction between the RuvC domain and REC1 domain is mainly mediated by the UK domain. The  $\alpha$  helix of BH forms  $\alpha$ -helical bundle with those of the REC2 domain to recognize the sgRNA and target DNA heteroduplex at one side. The other side of the heteroduplex is recognized by the REC2 domain. Dali search identified Cpf1 (PDB: 5B43 with an r.m.s.d. of 4.3 Å for 335 equivalent Ca atoms) as the most similar structure to that of BthC2c1, and the similarity is largely contributed by the RuvC domain.

The sgRNA in our structure consists of a guide segment (C1-U19), a repeat segment (C(-1)-G(-13)), a tetraloop (C(-14)-U(-17)), an anti-repeat segment (C(-18)-A(-24), and U(-57)-G(-61)), and three stem loops (stem loops 1-3) (Figure 1D and 1E). The guide segment and 19 nucleotides of the target DNA strand (dG(1')-dA(19')) form the guide:target heteroduplex, whereas the other 9 nucleotides of the target DNA strand (dG(-1')-dA(-9')) and the non-target DNA strand (dC(-1\*)-dT(-9\*)) form a PAM-containing duplex (PAM duplex) (Figure 1D and 1E; “'” indicates nucleotide in the target DNA strand and “\*” indicates nucleotide in the non-target DNA strand).

The PI domain and the N-terminal region of the REC1 domain interact with the PAM-proximal region of the heteroduplex, whereas the C-terminal regions of the REC1 and REC2 domains interact with the PAM-distal region of the heteroduplex (Figure 1B and Supplementary information, Figure S2A). The negatively charged sgRNA:target DNA heteroduplex is accommodated in the positively charged channel at the interface formed by



REC and NUC lobes (Figure 1B and Supplementary information, Figure S1A). Recognition of the sgRNA:target DNA heteroduplex by BthC2c1 is mainly through interactions between sugar-phosphate backbone and the protein. The PAM-distal region (A13-U19) of the sgRNA interacts with the two REC domains (Lys752, Arg768, Val767, Gly765, Asp279, Tyr333, Gln323, and Lys320) (Supplementary information, Figure S2A), whereas the sugar-phosphate backbone of the target DNA sequence (dT(13′)-dA(19′)) complementary to that of PAM-distal guide segment is extensively recognized by the two REC domains (Arg769, Arg272, Thr280, Asn282, Arg294, and Arg328) and the RuvC domain (Arg841) (Supplementary information, Figure S2A). The repeat:anti-repeat duplex containing an anticipated base-pairing segment (U(-6):G(-25)-G(-13):C(-18)) and an unanticipated base-pairing segment (C(-1):G(-61)-A(-5):U(-57)), is recognized by OBD (Glu412, Lys415, Leu414, Lys413, Asn452, Try451, Arg448, Arg507, and Lys9) and REC2 (Lys813, Tyr808, Lys794, Trp815, Lys793, Asn743, His783, and Asp790) domains (Supplementary information, Figure S2A).

The 5′-ATTC-3′ PAM duplex is sandwiched between the OBD and PI domains. The OBD domain consists of a  $\beta$ -sheet barrel flanked by four short  $\alpha$ -helices, whereas the PI domain is composed of a bundle of four  $\alpha$ -helices connected by linkers and loop PL1 (Ser129-Arg143) (Figure 1B). The loop PL1 deeply inserts into the minor groove of PAM duplex and interacts with the target and non-target DNA strands (Figure 1B). Ser137, Lys141, and Arg140 from the loop PL1 hydrogen-bonds with the sugar-phosphate backbone of dC(-6′), dC(-5′), and dA(-2′), respectively (Figure 1F). The sugar-phosphate backbone of PAM in the non-target DNA strand is recognized by Ser211, Val212, Ser129, Gln130, Gly132, Trp162, and Arg143 via hydrogen-bonding interactions (Supplementary information, Figure S2A-S2B). The O2 and O4 of dT(-2′) and the O6 of dG(-1′) form hydrogen bonds with Arg140 and Asn118, respectively (Figure 1G and Supplementary information, Figure S2B), explaining

the requirement for dT(-2′) in the 5′-ATTC-3′ PAM [8]. In addition, the N3 of dA(-2′) is also recognized by the side chain of Arg140. Another loop (L1, residues Ser395-Asn400) from OBD recognizes the PAM duplex from the major groove side, through the hydrogen bonds between Ser397 and the N6 of dA(-4′), and N6 and N7 of dA(-3′), and those between Asn398 and N6 of dA(-3′), and N6 and N7 of dA(-2) (Figure 1G). Mutations of these PI residues largely reduced the DNA cleavage activity of BthC2c1 *in vitro* (Figure 1H), further supporting our structural observation. In addition, residues Ser138 and Gly139 from loop PL1 are located right at the bottom of the minor groove of PAM duplex (Figure 1F). Replacement of them by bulkier residues could cause steric repulsion between loop PL1 and PAM bases; indeed, the S138Y and G139T mutations significantly impaired the DNA cleavage activity of BthC2c1 (Figure 1H). These structural and biochemical data indicate that BthC2c1 has stringent specificity for PAM. This is in contrast with the relaxed PAM recognition mode seen in SaCas9 [10] and Cpf1 [11] (Figure 1I). While further verification by functional studies is needed, the stringent PAM recognition *in vitro* suggests a higher substrate cleavage specificity of BthC2c1.

The phosphate backbone of stem loop 1 (C(-74)-G(-104)) is recognized by the REC, BH, RuvC, and UK domains (Figure 1B and Supplementary information, Figure S2A). The flipped-out bases of A(-100) and G(-99) are recognized by Lys619 via hydrogen-bonding and Tyr808 via stacking interaction, respectively. G(-86) is extensively recognized by Arg613, His802, and Asn819. On the basis of the structural observation that stem loop 1 is bound to the backside surface of the catalytic center of RuvC, we reasoned that removal of stem loop 1 may not affect the cleavage activity of BthC2c1. Indeed, our *in vitro* cleavage assay confirmed that the DNA cleavage activity of BthC2c1 guided by a stem loop 1-truncated sgRNA (29-end; Supplementary information, Data S1) is comparable to that of full-length sgRNA, whereas BthC2c1 guided by an sgRNA with

**Figure 1** The stringent PAM recognition by BthC2c1-sgRNA complex. **(A)** Graphic representation of domain organization of BthC2c1. The putative catalytic residues Asp574, Glu828, and Asp952 are marked with a black dot. **(B)** Cartoon representation of the BthC2c1-sgRNA-DNA complex shown in two orientations. Disordered linkers are shown as dotted lines. Individual BthC2c1 domains are colored according to the scheme in **A**. **(C)** Surface representation of the BthC2c1-sgRNA-DNA complex. **(D)** Schematic representation of the sgRNA:PAM-containing DNA heteroduplex. **(E)** Cartoon representation of the sgRNA:PAM-containing DNA heteroduplex. **(F)** Recognition of the 5′-ATTC-3′ PAM by loop PL1 from the PI domain. The PAM sequence is highlighted in purple. Hydrogen bonds are shown as black dashed lines. **(G)** Recognition of the 5′-ATTC-3′ PAM by loop L1 from the OBD domain. **(H)** Cleavage activity analysis using wild-type or mutant BthC2c1 with the PAM-interacting residues mutated. Data shown are representative of three independent experiments. **(I)** Structural comparison of PAM recognition mode by BthC2c1, SaCas9 (PDB: 5CZZ), and AsCpf1 (5B43). The PAM sequence is highlighted in purple. **(J)** Cleavage activity analysis using distinct truncated sgRNAs. Data shown are representative of three independent experiments. **(K)** Sanger-sequencing traces of BthC2c1-digested *EMX1* DNA showing staggered overhangs. Two cleavage sites are highlighted by red triangles in the top panel. **(L)** Model of sgRNA-guided DNA cleavage by BthC2c1.

longer truncation (33-end; Supplementary information, Data S1) failed to efficiently cleave substrate DNA (Figure 1J). Based on the structural observation that the tetraloop is not bound to BthC2c1, we reasoned that the tetraloop may not be necessary for BthC2c1's cleavage activity; indeed, the DNA cleavage activity of BthC2c1 guided by a tetraloop-truncated-mutant sgRNA ( $\Delta 85-92/GAA$ ; Supplementary information, Data S1) is comparable to that of full-length sgRNA (Figure 1J).

To map the DNA cleavage site of BthC2c1, we performed Sanger sequencing to analyze the DNA ends of the cleaved products of *in vitro* cleavage reactions. We found that BthC2c1-cleaved DNA products had a 7-nt 5' overhang (Figure 1K), differing from the blunt DNA cleavage mode of Cas9 [3]. This staggered double-stranded cleavage occurred after the 16th nucleotide on the non-target strand and after the 23rd nucleotide on the target strand distal to the PAM sequence (Figure 1K). The BthC2c1 cleavage site on the target strand is located outside the guide:target heteroduplex segment. This is distinct from Cas9 and Cpf1, both of which cleave the target strand within the guide:target heteroduplex segment [3, 9]. Interestingly, the target strand cleavage mode of BthC2c1 resembles that of C2c2, although C2c2 digests crRNA-guided RNA substrates [12]. On the basis of these observations, we propose a model for C2c1-catalyzed RNA-guided DNA cleavage (Figure 1L).

During our preparation of this manuscript, the structures of *Alicyclobacillus acidoterrestris* C2c1 (AacC2c1) in complex with sgRNA and target DNA [13], and AacC2c1 in complex with sgRNA [14] were reported. The BthC2c1 possesses 33% sequence identity with AacC2c1 (Supplementary information, Figure S1B). Structural comparison of the C2c1-sgRNA-DNA ternary complex between *B. thermoamylovorans* and *A. acidoterrestris* indicates that the overall structure of BthC2c1 adopts a similar fold as that of AacC2c1, and sgRNA and target DNA also display a similar conformation in these two structures (Supplementary information, Figure S2B). The overall main chain r.m.s.d between BthC2c1 and AacC2c1 is 1.4 Å for 701 comparable C $\alpha$  atoms. In addition, these two studies also revealed a mode of staggered double-stranded DNA breaks in C2c1-cleaved products [13, 14].

In summary, the data presented here reveal the mechanism of recognition of sgRNA and PAM-duplex by BthC2c1, which is different from those of Cas9 and Cpf1. Our study provides insights into generation of engineered C2c1 family proteins with better efficiency and specificity for genome manipulation applications.

**Accession number:** The atomic coordinates and structure factors of the BthC2c1-crRNA-DNA complex have been deposited to the Protein Data Bank under the accession code of 5WTI.

## Acknowledgments

We thank J He at the Shanghai Synchrotron Radiation Facility (SSRF) and D Yao at beamline BL19U1 for help with data collection. We thank JJ Chai for critical reading of the manuscript. This research was funded by the National Natural Science Foundation of China (31422014, 31450001, and 31300605 to ZH).

Dan Wu<sup>1,\*</sup>, Xiaoyu Guan<sup>1,\*</sup>, Yuwei Zhu<sup>1,\*</sup>, Kuan Ren<sup>1</sup>, Zhiwei Huang<sup>1</sup>

<sup>1</sup>HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China

\*These three authors contributed equally to this work.

Correspondence: Zhiwei Huang

E-mail: huangzhiwei@hit.edu.cn

## References

- 1 Marraffini LA. *Nature* 2015; **526**:55-61.
- 2 Makarova KS, Haft DH, Barrangou R, *et al. Nat Rev Microbiol* 2015; **13**:722-736.
- 3 Jinek M, Chylinski K, Fonfara I, *et al. Science* 2012; **337**: 816-821.
- 4 Anders C, Niewoehner O, Duerst A, *et al. Nature* 2014; **513**: 569-573.
- 5 Jiang F, Zhou K, Ma L, *et al. Science* 2015; **348**: 1477-1481.
- 6 Wright AV, Nuñez JK, Doudna JA. *Cell* 2016; **164**: 29-44.
- 7 Sternberg SH, Doudna JA. *Mol Cell* 2015; **58**: 568-574.
- 8 Shmakov S, Abudayyeh OO, Makarova KS, *et al. Mol Cell* 2015; **60**: 385-397.
- 9 Zetsche B, Gootenberg JS, Abudayyeh OO, *et al. Cell* 2015; **163**: 759-771.
- 10 Nishimasu H, Cong L, Yan WX, *et al. Cell* 2015; **162**: 1113-1126.
- 11 Yamano T, Nishimasu H, Zetsche B, *et al. Cell* 2016; **165**: 949-962.
- 12 Abudayyeh OO, Gootenberg JS, Konermann S, *et al. Science* 2016; **353**: aaf5573.
- 13 Yang H, Gao P, Rajashankar KR, Patel DJ. *Cell* 2016; **167**: 1814-1828.
- 14 Liu L, Chen P, Wang M, *et al. Mol Cell* 2017; **65**:310-322.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)