

Structural insights into target DNA recognition and cleavage by the CRISPR-Cas12c1 system

Bo Zhang[†], Jinying Lin[†], Vanja Perčulija[†], Yu Li, Qiuhua Lu, Jing Chen and Songying Ouyang^{✉*}

The Key Laboratory of Innate Immune Biology of Fujian Province, Provincial University Key Laboratory of Cellular Stress Response and Metabolic Regulation, Biomedical Research Center of South China, Key Laboratory of OptoElectronic Science and Technology for Medicine of Ministry of Education, College of Life Sciences, Fujian Normal University, Fuzhou, 350117, China

Received April 08, 2022; Revised October 09, 2022; Editorial Decision October 12, 2022; Accepted October 20, 2022

ABSTRACT

Cas12c is the recently characterized dual RNA-guided DNase effector of type V-C CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein) systems. Due to minimal requirements for a protospacer adjacent motif (PAM), Cas12c is an attractive candidate for genome editing. Here we report the crystal structure of Cas12c1 in complex with single guide RNA (sgRNA) and target double-stranded DNA (dsDNA) containing the 5'-TG-3' PAM. Supported by biochemical and mutation assays, this study reveals distinct structural features of Cas12c1 and the associated sgRNA, as well as the molecular basis for PAM recognition, target dsDNA unwinding, heteroduplex formation and recognition, and cleavage of non-target and target DNA strands. Cas12c1 recognizes the PAM through a mechanism that is interdependent on sequence identity and Cas12c1-induced conformational distortion of the PAM region. Another special feature of Cas12c1 is the cleavage of both non-target and target DNA strands at a single, uniform site with indistinguishable cleavage capacity and order. Location of the sgRNA seed region and minimal length of target DNA required for triggering Cas12c1 DNase activity were also determined. Our findings provide valuable information for developing the CRISPR-Cas12c1 system into an efficient, high-fidelity genome editing tool.

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) are prokaryotic adaptive immune systems that protect the host

against bacteriophages by deploying Cas effector nucleases paired with CRISPR RNA (crRNA) to recognize and cleave foreign nucleic acid sequences complementary to the guide (also termed spacer) portion of crRNA (1–3). Based on characteristics of Cas effectors, CRISPR-Cas systems are grouped into classes 1 and 2 (4). Class 2 systems (type II, V and VI systems) use a single multidomain Cas effector guided by crRNA to carry out nucleic acid interference (5). Effectors of some class 2 systems also require an additional RNA molecule known as trans-activating crRNA (tracrRNA), which hybridizes with crRNA to form a functional RNA guide. Owing to their efficiency, targeting programmability and ease of use, a number of class 2 effectors have been successfully repurposed as nucleic acid-cleaving or binding tools for genome editing, transcriptome editing, nucleic acid diagnostics and imaging, and other applications (6–20). Cas9 from type II systems and Cas12a (Cpf1) from type V-A systems are widely used in genome editing, and type V effectors Cas12b (C2c1), Cas12e (CasX) and Cas12f (Cas14a) have also been harnessed for this purpose (21–29).

Although crRNA spacers can be designed to allow editing of virtually any genomic sequence, DNA-targeting Cas effectors also require the presence of an ortholog-specific short sequence flanking the targeted sequence known as the protospacer adjacent motif (PAM) (30–32). PAMs enable discrimination between self and non-self in the native environment, but limit target selection in genome editing (33,34). For instance, the commonly used *Streptococcus pyogenes* Cas9 (SpCas9) requires a 5'-NGG-3' PAM that restricts the availability of targetable sites in the human genome to approximately one site for every eight base pairs (7,8,35). Consequently, efforts are being made to loosen PAM-imposed targeting constraints by either engineering currently available Cas effectors or identifying new types and variants (18,34).

*To whom correspondence should be addressed. Tel: +86 591 22868072; Email: ouyangsy@fjnu.edu.cn

[†]The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.

Recently, new type V systems with functionally divergent effectors that could potentially expand the CRISPR-Cas toolbox have been discovered (4,26,36–40). Among them, Cas12c (formerly C2c3), the signature effector of type V-C systems, is a dual-RNA-guided DNA endonuclease similar in size to commonly used Cas9 and Cas12a variants (36,38). Analogously to other type V effectors, Cas12c contains a single RuvC nuclease domain at its C-terminal region, but the remainder of its primary sequence does not share significant similarity with any other Cas effector except Cas12d (36,37). Type V-C effectors process their own precursor crRNA (pre-crRNA) in the presence of a tracrRNA molecule termed short-complementarity untranslated RNA (scoutRNA) (41). Different from Cas12a, Cas12c effectors employ the RuvC active site to cleave pre-crRNA downstream of the spacer sequence in a metal ion-dependent manner (41,42). Most notably, Cas12c variants were found to be constrained by minimal PAMs — Cas12c1 and *Oleiphilus* sp. HI0009 Cas12c (OspCas12c) by a 5'-TG-3' PAM, and Cas12c2 by a 5'-TN-3' PAM (38). Therefore, Cas12c effectors could greatly expand the targeting space of CRISPR-Cas genome editing. However, molecular mechanisms underlying PAM recognition and target DNA cleavage by Cas12c remain elusive, thereby hindering efficient application.

Here, we report a crystal structure of Cas12c1 in complex with covalently fused crRNA and tracrRNA [i.e. single-guide RNA (sgRNA)] and double-stranded target DNA containing the 5'-TG-3' PAM. The structure reveals distinct structural features of Cas12c1 and the associated sgRNA, as well as certain similarities to other type V CRISPR-Cas systems. We find that Cas12c1 recognizes the 5'-TG-3' PAM mainly through a PAM-interacting (PI) loop and a mechanism that is interdependent on sequence identity and Cas12c1-induced conformational distortion of the PAM region. Another distinctive feature of Cas12c1 is the cleavage of both non-target DNA (NTD) and target DNA (TD) strands at a single and uniform site with indistinguishable capacity and order. Furthermore, we identify crucial structural elements and residues involved in target DNA duplex unwinding, TD-guide heteroduplex formation and recognition, and loading of the NTD strand into the DNase active site. Location of the seed region, minimal length of double-stranded and single-stranded DNA targets, and temperature and divalent metal ion preferences for the catalytic activity of Cas12c1 were also determined. The findings of this study lay the groundwork for effective utilization of the CRISPR-Cas12c1 system in genome editing and other applications.

MATERIALS AND METHODS

Expression and purification of Cas12c1

The gene encoding full-length Cas12c1 (residues 1–1302) was codon-optimized (Supplementary Table S1) and cloned between NdeI and XhoI sites of the pET-30b vector with a C-terminal 6× His-tag. This vector also served as a template for generating mutants of Cas12c1 using the site-directed mutagenesis protocol described below. All recombinant vectors were transformed into *Escherichia coli* Rosetta (DE3) competent cells (Novagen).

After the addition of 0.2 mM isopropyl-β-D-thiogalactopyranoside (IPTG) to the bacterial cultures that were first grown at 37°C until the OD₆₀₀ reached 0.6, wild-type Cas12c1 and its mutants were overexpressed at 18°C for 12 h. The cells were then collected, resuspended in buffer A1 [25 mM Tris-HCl (pH 7.5), 1 M NaCl, 5 μM ZnCl₂, 3.5 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride (PMSF)] and disrupted by ultrasonication. The supernatant was loaded onto a Ni-NTA Superflow column (QIAGEN), from which the target protein was eluted in buffer A2 [25 mM Tris-HCl (pH 7.5), 400 mM NaCl, 5 μM ZnCl₂, 3.5 mM β-mercaptoethanol, 1 mM PMSF] supplemented with 50 mM and 200 mM imidazole. The target protein was then purified with a HiTrap Heparin HP column (GE Healthcare) using a salt gradient between buffer B1 [25 mM Tris-HCl (pH 7.5), 400 mM NaCl, 2 mM MgCl₂, 2 mM dithiothreitol (DTT)] and buffer B2 [25 mM Tris-HCl (pH 7.5), 1 M NaCl, 2 mM MgCl₂, 2 mM DTT]. After pooling and concentrating the fractions containing purified target protein, the concentrated protein sample was dialyzed against buffer C [25 mM Tris-HCl (pH 7.5), 350 mM NaCl, 2 mM MgCl₂, 2 mM DTT] and purified with a Mono S 10/100 GL column (GE Healthcare) using a salt gradient between buffers C and B2. Lastly, fractions with purified target protein were pooled and concentrated, dialyzed against buffer D [25 mM Tris-HCl (pH 7.5), 200 mM NaCl, 2 mM MgCl₂, 2 mM DTT] and loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare) for the final purification step. Purified wild-type Cas12c1 and its mutants were concentrated and stored at -80°C until further use.

Selenomethionine (SeMet)-labeled Cas12c1 D969A mutant was expressed in *E. coli* Rosetta (DE3) cells grown in M9 minimal medium supplemented with SeMet, Lys, Phe, Thr, Val, Leu and Ile. Purification of SeMet-labeled Cas12c1 D969A mutant was carried out following the protocol described above.

In vitro transcription and purification of sgRNA

The DNA fragment encoding the 136 nt sgRNA sequence (Supplementary Table S2) was cloned into sites between StuI and HindIII of a modified pUC-119 vector. The recombinant vector was amplified in *E. coli* DH5α cells, extracted, linearized using HindIII and purified by phenol-chloroform extraction and ethanol precipitation. *In vitro* transcription was performed at 37°C for 3 h in a reaction mixture containing 100 mM HEPES-K (pH 7.9), 20 mM MgCl₂, 30 mM DTT, 3 mM of each NTP, 2 mM spermidine, 30 ng/μl linearized DNA template and 100 μg/ml homemade T7 RNA polymerase. The reaction mixture was resolved in a 10% denaturing polyacrylamide gel containing 8 M urea, following which the sgRNA band was excised from the gel and eluted using the Elutrap system (GE Healthcare). Purified sgRNA was desalted and concentrated by ethanol precipitation, dissolved in diethylpyrocarbonate (DEPC)-treated water and stored at -80°C.

Formation of the Cas12c1 ternary complex

First, to obtain the SeMet-labeled Cas12c1 D969A–sgRNA binary complex, Cas12c1 D969A and sgRNA were mixed

in buffer D at a molar ratio of 1:1.6 and incubated at 16°C for 1 h. The mixture was passed through a Superdex 200 Increase 10/300 GL column. Fractions containing purified binary complex were pooled and concentrated for further experiments. Gel-purified NTD and TD strands used in this study (Supplementary Table S3) were purchased and dissolved in buffer D without 2 mM DTT to form a 100 μ M stock. The target duplex was prepared by mixing NTD and TD strands at a molar ratio of 1:1, denaturing the mixture at 95°C for 5 min and annealing the sample by gradual cooling to room temperature. The SeMet-labeled Cas12c1 D969A-sgRNA-target DNA ternary complex was assembled by mixing the prepared binary complex and annealed target duplex in buffer D at an approximate molar ratio of 1:1.1, followed by incubation at 16°C for 1 h. Fractions containing purified ternary complex were collected after purification with a Superdex 200 Increase 10/300 GL column and concentrated to $A_{280\text{ nm}} = 12.5$.

Crystallization

The SeMet-labeled Cas12c1 D969A-sgRNA-target DNA ternary complex was crystallized at 16°C using the hanging drop vapor diffusion technique. Diffraction-quality crystals were obtained by mixing 1 μ l of the ternary complex solution ($A_{280\text{ nm}} = 12.5$) with 1 μ l of reservoir solution [0.1 M Tris-HCl (pH 8.2), 16% polyethylene glycol 3350 (w/v), 0.02 M citric acid and 0.05 M lithium acetate dihydrate]. The crystals grew to full size within 12 days and were subsequently harvested, cryoprotected in reservoir solution containing 14% (v/v) glycerol and flash-frozen in liquid nitrogen.

Data collection and structure determination

X-Ray diffraction data were collected on beamline BL-17U1 of the Shanghai Synchrotron Radiation Facility (SSRF) by using the X-ray wavelength 0.9792 Å and the program Blu-Ice. The diffraction dataset was processed and scaled by the program HKL2000 (43). The crystal structure of the SeMet-labeled Cas12c1 D969A-sgRNA-target DNA ternary complex was solved by the single-wavelength anomalous dispersion method using programs AutoSol and AutoBuild in PHENIX (44). There are two ternary complexes contained in the crystal asymmetric unit. Model building was performed in Coot (45) and model refinement was carried out using REFMAC in CCP4 format (46). All structural figures were created in PyMOL (<http://pymol.org>), unless indicated otherwise. Data collection and refinement statistics are listed in detail in Supplementary Table S4.

Site-directed mutagenesis

Pairs of complementary oligonucleotide primers containing the desired nucleotide changes were designed (Supplementary Table S5) and synthesized (Sangon Biotech). Polymerase chain reactions (PCRs) using the mutagenic primers and the wild-type Cas12c1 vector as a template were carried out, and the parent template was afterward digested with DpnI. The PCR product was transformed into *E. coli*

DH5 α competent cells (Novagen), which were then cultured on LB agar plates containing kanamycin. Plasmids were isolated from bacterial colonies and verified for successful mutagenesis by DNA sequencing.

Plasmid cleavage assay

Oligonucleotides encoding desired NTD and TD strands (Supplementary Table S3) were annealed and cloned between EcoRI and HindIII sites of the pUC-19 vector, which was verified by DNA sequencing. Successfully constructed plasmids were linearized with ScaI. Plasmid cleavage assays were performed in buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM MgCl₂ and 1 mM DTT. Binary complexes used in the assays were prepared by incubating Cas12c1 (final concentration of 0.5 μ M) with sgRNA (final concentration of 0.5 μ M) at 25°C for 30 min. Cleavage reactions were carried out by adding 1000 ng of linearized target DNA plasmid to the mixture (final volume of the mixture adjusted to 30 μ l) and subsequent incubation at 45°C for 30 min. To perform the metal ion-dependent cleavage assays, 10 mM MgCl₂ in assay buffer was replaced by 10 mM of different divalent metal ions or 5 mM EDTA. Reactions were terminated by adding EDTA (final concentration of 100 mM) and Proteinase K (final concentration of 0.8 mg/ml) at 37°C for 30 min. Reaction samples were resolved on 1% agarose gels stained with Gel Stain (Transgene Biotech). All cleavage assays were performed at least three times independently.

Oligonucleotide cleavage assay

NTD strands containing 5'-6FAM fluorescent labels and TD strands containing 3'-6FAM fluorescent labels were purchased (Sangon Biotech) (Supplementary Table S6). Two strands were annealed with a molar ratio of 1:1 to prepare a 77 μ M stock. To prepare the binary complex, wild-type Cas12c1 (final concentration of 4 μ M) and sgRNA (final concentration of 4 μ M) were incubated to form the binary complex at 25°C for 30 min in assay buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM MgCl₂ and 1 mM DTT]. The cleavage assays were performed at 45°C for 30 min by adding fluorescently labeled double-stranded substrates (final concentration of 2 μ M) in reaction samples with total volume adjusted to 20 μ l. Reactions were terminated by adding EDTA (final concentration of 100 mM) and Proteinase K (final concentration of 0.8 mg/ml) at 37°C for 30 min. Reaction samples were run on 20% polyacrylamide gel electrophoresis (PAGE) TBE-urea denaturing gels and the results were visualized using the BIO-RAD Universal Hood II System. All cleavage assays were performed at least three times independently.

Cas12c1 activation and *trans* cleavage assay

Double-stranded DNA (dsDNA) activators containing 12–18 bp nucleotides after the PAM duplex and 12–18 nt single-stranded DNA (ssDNA) activators were assessed for their ability to trigger the Cas12c1 DNase activity (Supplementary Table S6). A non-specific ssDNA containing a 5'-6FAM fluorescent label was used as a substrate for detection

of the *trans* cleavage activity (Supplementary Table S6). The Cas12c1 binary complex was prepared following the protocol for the oligonucleotide cleavage assay. The *trans* cleavage assays were performed by adding dsDNA or ssDNA activator with a final concentration of 2 μ M and fluorescently labeled non-specific ssDNA with a final concentration of 2 μ M. Total volumes of reaction samples were adjusted to 20 μ l. The assays were performed at 45°C for 20 min in assay buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 10 mM MgCl₂ and 1 mM DTT]. Reactions were terminated by adding EDTA (final concentration of 100 mM) and Proteinase K (final concentration of 0.8 mg/ml) at 37°C for 30 min. Reaction samples were run on 20% PAGE TBE-urea denaturing gels and the results were visualized using the BIO-RAD Universal Hood II System. All cleavage assays were performed at least three times independently.

Electrophoretic mobility shift assay (EMSA)

To prepare target dsDNA, NTD and TD strands were mixed at a molar ratio of 1:1 and annealed to form a 50 μ M stock (Supplementary Table S6). To prepare the binary complexes, the Cas12c1 D969A or the five double mutants (K149A/D969A, R152A/D969A, K156A/D969A, N299A/D969A or R374A/D969A) were incubated with sgRNA using a molar ratio of 1:1 at 22°C for 30 min in assay buffer [50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 2 mM MgCl₂, 1 mM DTT]. To perform the EMSA, target dsDNA was incubated with the binary complexes at 37°C for 30 min in assay buffer. The final concentration of target dsDNA was 2.0 μ M, the three final concentrations of the binary complexes were 2.0, 4.0 and 8.0 μ M, and the total volume of the assay mixture was adjusted to 20 μ l. Samples were separated on a 6% native PAGE gel, and 0.5 \times TBE buffer was used for electrophoresis. The native PAGE gels were stained with Gel Stain (Transgene Biotech). The EMSAs were performed at least three times independently.

Precursor sgRNA cleavage assay

The 153 nt precursor sgRNA (pre-sgRNA) was designed by adding 17 nt sgRNA repeat sequence after the 3'-end of the 18 nt sgRNA spacer sequence (Supplementary Table S2). The pre-sgRNA was prepared using the same protocol as for the preparation of sgRNA. The cleavage assays were performed by incubating 3 μ M pre-sgRNA with 3 μ M wild-type Cas12c1 or the D969A mutant in the assay buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT] containing 10 mM MgCl₂ or 5 mM EDTA with total volumes adjusted to 20 μ l. Reactions were allowed to proceed at 45°C for 30 min and terminated as described above. Reaction samples were run on 10% PAGE TBE-urea denaturing gels and the results were visualized using Gel Stain (Transgene Biotech). All cleavage assays were performed at least three times independently.

RESULTS

Overall architecture of the Cas12c1 ternary complex

To understand how Cas12c1 recognizes sgRNA and double-stranded target DNA, interacts with the 5'-TG-3'

PAM and cleaves both strands of target DNA, we solved the crystal structure of the SeMet-labeled and catalytically inactive Cas12c1 in complex with sgRNA and double-stranded target DNA at 3.20 Å resolution (Supplementary Table S4). Cas12c1 was inactivated by mutating the catalytic residue D969 to alanine. The single-wavelength anomalous diffraction method was used to determine the crystal structure. Two copies of the ternary complex (denoted ternary complexes A and B) were found per crystal asymmetric unit (Supplementary Figure S1A). The ternary complex A will be analyzed herein unless stated otherwise.

The overall structure of Cas12c1 in the ternary complex evidently displays structural divergence compared with other type V Cas effectors, though it retains the typical bilobed architecture. The recognition (REC) lobe of Cas12c1 encompasses the Rec1, Rec2 and PI domains. The nuclease (NUC) lobe consists of the Wedge (WED), RuvC and Nuc domains (Figure 1A–C). The spatial distribution of each domain in Cas12c1 is comparable with that of Cas12e, except that the PI domain of Cas12c1 is replaced by the functionally distinct non-target strand-binding (NTSB) domain in Cas12e (Supplementary Figure S2). The Rec2, WED, RuvC and Nuc domains are located on a complicated scaffold formed by sgRNA. The guide region of sgRNA and the TD strand of the target DNA form a heteroduplex accommodated within a positively charged central channel between REC and NUC lobes (Figure 1B, C). The Rec1 domain consists of two α -helix bundles that constitute the PAM-proximal and PAM-distal regions of the domain, respectively. Due to flexibility, parts of the PAM-distal α -helix bundle are invisible in the present model (Figure 1B; Supplementary Figure S3A, B). The PI domain is made of seven α -helices and protrudes from the Rec1 domain (Figure 1B; Supplementary Figure S3A, C). The WED domain is divided into two motifs (WED-I and WED-II) that together form two β -sheets flanked by several α -helices (Figure 1B; Supplementary Figure S3A, D). The Rec2 domain comprises a bundle of eight α -helices and is connected to the Rec1 and WED domains (Figure 1B; Supplementary Figure S3A, E). The RuvC domain, which exhibits a well-conserved RNase H fold and contains the endonuclease active site, comprises two motifs (RuvC-I and RuvC-II) that fold into a central mixed β -sheet flanked by five α -helices (Figure 1B; Supplementary Figure S3A, F). The Nuc domain is positioned between the RuvC-I and RuvC-II motifs and contains a CCCC-type zinc-finger (ZF) motif (Figure 1B; Supplementary Figure S3A, G).

Overall architecture of sgRNA and target DNA

The sgRNA and target DNA duplex in the ternary complex displays a lying-down H-shaped architecture (Figure 1D, E; Supplementary Figure S4). The 136 nt sgRNA was generated by fusing the 3'-end of the 92 nt tracrRNA to the 5'-end of the 44 nt crRNA, whose sequences were based on the previous study (38). The crRNA is composed of a 21 nt repeat region and a 23 nt guide region, whereas the tracrRNA comprises three separate anti-repeat regions and three stem regions. The crRNA repeat region and the tracrRNA anti-repeat regions form a repeat:anti-repeat (R:AR) duplex consisting of three duplex regions. The R:AR duplex-1,

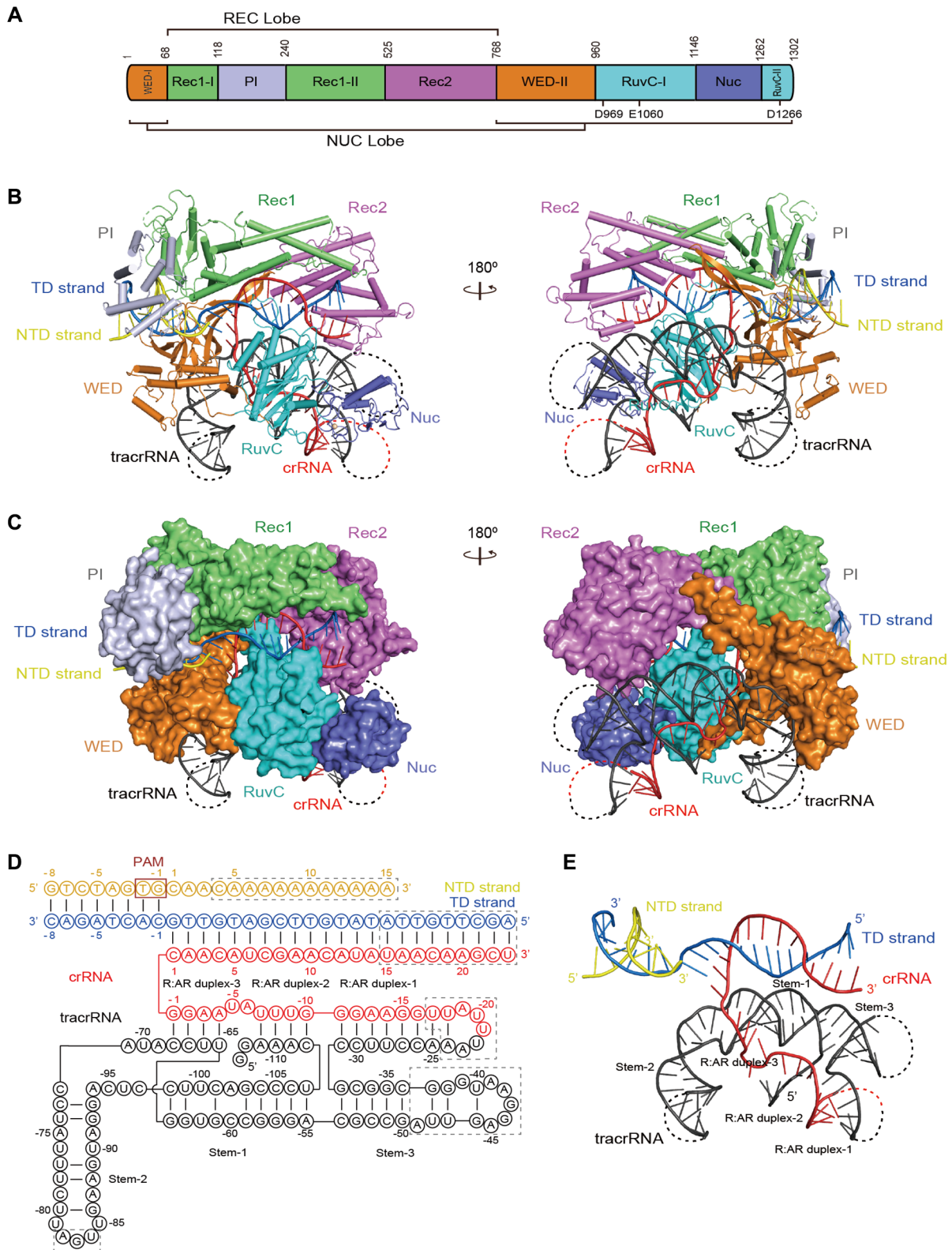


Figure 1. Crystal structure of the Cas12c1 ternary complex. (A) Domain organization of Cas12c1. Catalytic residues of the RuvC domain are annotated below the diagram. (B) Overall structure of the Cas12c1 ternary complex shown in two different orientations. (C) Surface representations of the Cas12c1 ternary complex shown in the same views as in (B). (D) Schematic representation of crRNA, tracrRNA, TD and NTD strands in the Cas12c1 ternary complex. Disordered regions are encircled with gray dashed lines. (E) Structures of crRNA, tracrRNA, TD and NTD strands in the Cas12c1 ternary complex. Domains of Cas12c1, crRNA, tracrRNA, TD and NTD strands are color-coded in subsequent figures as in (A) and (D) unless stated otherwise.

duplex-2 and duplex-3 regions are formed by pairing of the nucleotides U(-18)–G(-11) of crRNA with C(-31)–A(-24) of tracrRNA, G(-10)–U(-7) of crRNA with A(-111)–C(-108) of tracrRNA and A(-4)–G(-1) of crRNA with C(-68)–U(-65) of tracrRNA, respectively. Nucleotides A(-6) and U(-5) of crRNA are rotated away from R:AR duplex-2 and R:AR duplex-3, respectively, thus forming a kink between the two R:AR duplex regions. Stem-1 is formed by pairing of the nucleotides U(-107)–C(-98) with G(-64)–A(-55), and stem-2 and stem-3 are formed by base pairing within regions G(-93)–C(-73) and C(-54)–G(-32), respectively. In the target DNA, nucleotides dG(-8*)–dG(-1*) of the NTD strand and dC(-8)–dC(-1) of the TD strand constitute the PAM duplex. A 14 bp guide–target heteroduplex formed between the nucleotides C(1)–A(14) of the sgRNA guide region and dG(1)–dT(14) of the TD strand is identified in the present model.

Recognition of sgRNA

Most of R:AR duplex-1 is exposed to solvent (Figure 1B, E), except for it forming a hydrogen bond with residue R1204 in the Nuc domain (Supplementary Figure S5). The R:AR duplex-2 interacts with the RuvC and Nuc domains (Figure 1B, E). Residues S1238–R1242 of a loop region in the Nuc domain have several interactions with the sugar-phosphate backbone of the R:AR duplex-2 (Supplementary Figure S5). In addition, R:AR duplex-3 is recognized by residues from the WED and RuvC domains such as S12, R20 and K58 in the WED domain, and R1019 and N1037 in the RuvC domain (Supplementary Figure S5).

Stem-1 of tracrRNA is in contact with the RuvC and Rec2 domains. Residue R1005 in the RuvC domain forms a hydrogen bond with the sugar-phosphate backbone of stem-1. In addition, residues N630, R631 and N635 from the short helix $\alpha 5$ in the Rec2 domain interact with stem-1 (Figure 1B, E; Supplementary Figures S3E, S5). The stem-2 region of tracrRNA interacts solely with the WED domain (Figure 1B, E). Residues K14, R793 and K796 form hydrogen bonds or salt bridges with the sugar-phosphate backbone of stem-2, whereas the side chain of Y801 forms a hydrogen bond with the N3 atom of G(-92) (Supplementary Figure S5). The stem-2 nucleotides U(-90)–U(-76) are exposed to solvent. Stem-3 of tracrRNA protrudes from the ternary complex and is almost exposed to solvent (Figure 1B, E).

Recognition of the 5'-TG-3' PAM duplex

The PAM duplex is accommodated within the PAM-binding cleft, which is formed by the PI, WED and Rec1 domains (Figure 2A). Prior to PAM duplex binding, the PI domain is likely to exhibit a certain degree of flexibility owing to its connection with the Rec1 domain via two loops. For PAM recognition, Cas12c1 employs an important loop between helices $\alpha 2$ and $\alpha 3$ of the PI domain which we henceforth refer to as the PAM-interacting loop (PI loop, residues 148–160) (Figure 2A). The PI loop extends into the minor groove of the PAM duplex, and several of its residues (i.e. K149, R152, K153 and K156) together form a positively charged area that snugly accommodates the PAM duplex

(Figure 2B). In particular, residues K149 and R152 clamp the NTD strand, and R151 concurrently forms a hydrogen bond with E170 to reinforce the conformation of the PI loop (Figure 2C). Residues N105, K153 and K156 interact with the sugar-phosphate backbone of the PAM duplex through hydrogen bonds and salt bridges (Supplementary Figure S5).

The 5'-TG-3' PAM sequence is recognized by specific hydrogen bonds. On one side of the PAM duplex, the side chain of residue R152 forms hydrogen bonds with the O2 atom of dT(-2*), the O4' atom of dG(-1*) and the O2 atom of dC(-3) (Figure 2C, D; Supplementary Figure S6A). On the other side of the PAM duplex, the side chain of residue R374 forms a hydrogen bond with the O4 atom of dT(-2*) (Figure 2D; Supplementary Figure S6B). Moreover, the side chain of residue N299 in the Rec1 domain extends towards the nucleotide dG(-1*) and forms hydrogen bonds with the O6 atom of dG(-1*) and the N4 atom of dC(-1) (Figure 2D). These interactions distort both the orientation of dG(-1*) and the canonical base pairing between dG(-1*) and dC(-1). Consequently, the N2 atom of dG(-1*) forms an atypical hydrogen bond with the N3 atom of dA(-2) instead of the O2 atom of dC(-1) (Figure 2D; Supplementary Figure S6A). Such atypical hydrogen bonding between neighboring PAM nucleotides has not been reported in other type V CRISPR-Cas systems and should therefore be a prominent feature of the PAM determination by Cas12c1. It is noteworthy that dT(-2*) is recognized by the residues R152 and R374 in a base-specific manner, and dG(-1*) is recognized cooperatively by residue N299 and the nucleotide dA(-2). In line with the previous study (38), we also found that the replacement of dG(-1*) or dT(-2*) with any other nucleotide abrogated the target cleavage activity, whereas replacement of dG(-3*) did not produce a noticeable effect (Figure 2E). Moreover, the cleavage assays show that the K149A, R152A, N299A and R374A mutations almost abolish the cleavage activity, and the K156A mutation greatly decreases the activity (Figure 2F). Therefore, residues R152, N299 and R374 play critical roles in recognition of the 5'-TG-3' PAM, the readout of which is based on the interdependence of sequence identity and Cas12c1-induced distortion of the PAM region.

To verify whether residues (K149, R152, K156, N299 and R374) affecting PAM recognition compromise target dsDNA binding, we performed EMSAs using the catalytically inactive K149A/D969A, R152A/D969A, K156A/D969A, N299A/D969A and R374A/D969A double mutants. Compared with the Cas12c1 D969A mutation, the N299A and R374A mutations moderately decreased target dsDNA binding, whereas the K149A, R152A and K156A mutations greatly decreased target dsDNA binding (Supplementary Figure S7). In agreement with the structural analysis above, residues K149, R152 and K156 from the PI loop interacting with the PAM duplex greatly contribute to target dsDNA binding. In addition, residues N299 and R374, which are crucial for PAM recognition, also play a positive role in target DNA binding.

Of note, some mutations in this study may affect target cleavage indirectly by compromising the binding of target dsDNA or sgRNA to Cas12c1. The analysis of cleavage assays could not distinguish whether the results arise from de-

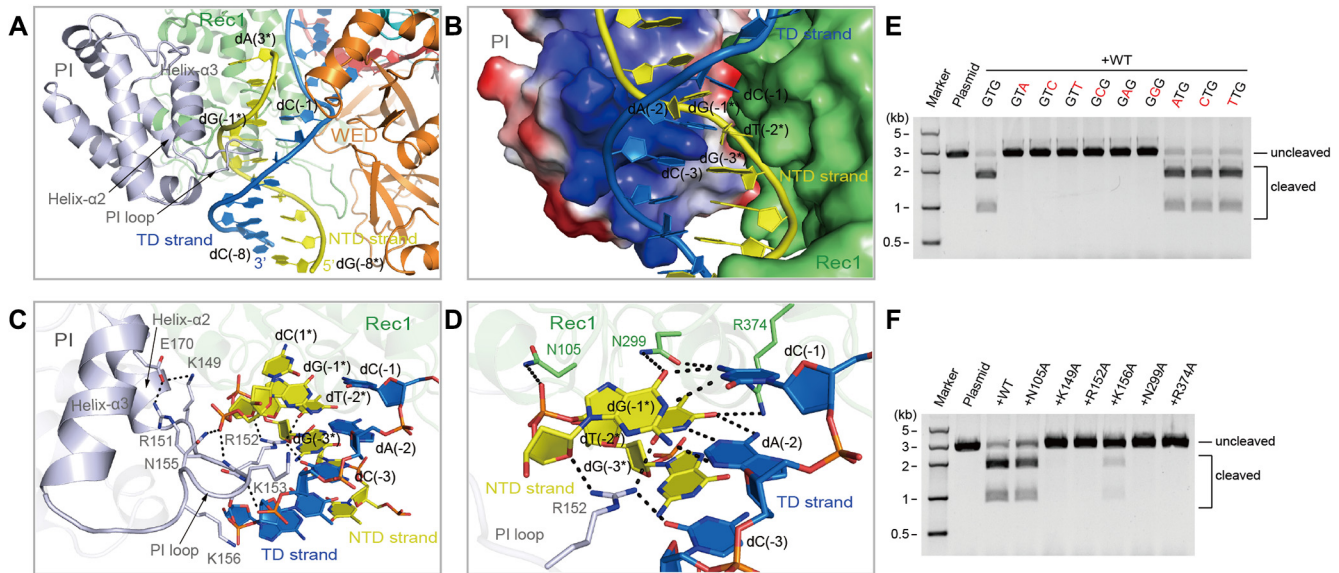


Figure 2. Recognition of the PAM duplex. (A) Recognition of the PAM duplex in the Cas12c1 ternary complex and the PAM-interacting cleft formed by the Rec1, PI and WED domains. The PI loop between helices $\alpha 2$ and $\alpha 3$ of the PI domain inserts into the minor groove of the PAM duplex. (B) Electrostatic potential surface of the PI domain. Red, white and blue indicate surfaces with negative, neutral and positive electrostatic potential, respectively. (C) Recognition of the PAM duplex by the PI domain. (D) Recognition of the PAM duplex by the Rec1 domain and residue R152 from the PI loop. (E) Agarose gel demonstrating the cleavage of the linearized plasmid containing the 5'-GTG-3' PAM sequence or its single-nucleotide mutants by wild-type Cas12c1 in complex with sgRNA. The mutated nucleotide is colored in red. (F) Agarose gel demonstrating the cleavage of the linearized plasmid by wild-type Cas12c1 and its mutants in complex with sgRNA.

creased target cleavage or decreased binding of target dsDNA or sgRNA.

Components facilitating the target duplex unwinding and heteroduplex formation

In the Cas12c1 ternary complex A, the TD strand is unwound from the NTD strand beyond the PAM duplex and forms a 14 bp heteroduplex with the sgRNA guide region. The possible U(15)–U(23):dA(15)–dA(23) base pairings are not visible in the model because of the flexibility of these nucleotides (Figure 1B; Supplementary Figure S4A). In contrast, a 13 bp heteroduplex and additional nucleotides dT(14)–dG(18) of the TD strand are observed in the ternary complex B (Supplementary Figures S1B and S4B). Of note, the side chain of the residue N297 in the Rec1 domain partially stacks with the pyrimidine ring of dC(-1) and disturbs the potential base pairing between dG(1) and dC(1*), thus facilitating the flipping of dG(1) for pairing with C(1) of the sgRNA guide region (Figure 3A). To investigate if residue N297 plays an important role in the target duplex unwinding at the PAM-proximal region, we prepared a N297G mutant and dsDNA substrates containing 1–4 bp mismatch bubbles neighboring the PAM sequence. The cleavage results showed that the N297G mutation obviously decreased the target cleavage activity, which was recovered by enlarging the mismatch bubble (Figure 3B). These findings indicate that residue N297 plays a significant role in target duplex unwinding and facilitating heteroduplex formation.

Other regions and residues of Cas12c1 also contribute to the heteroduplex formation. Strand $\beta 1$ of the WED domain is adjacent to the PAM-proximal region of the heteroduplex, greatly restricting the nucleotide orientation and sta-

bilizing base pairing in this region (Figure 3C). In particular, residue L52 appreciably limits the orientation of dG(1), and the side chain of T54 forms hydrogen bonds with the N2 atom of dG(1) and the O2 atom of C(1) (Figure 3C). Meanwhile, residues K875 and R915 form hydrogen bonds with the phosphate group between dC(-1) and dG(1) of the TD strand, which helps in maintaining the orientation of dG(1) for pairing with C(1) (Figure 3A). Moreover, helix $\alpha 6$ and strand $\beta 1$ of the WED domain are positioned above the R:AR duplex-3, which prompts the swinging of C(1) of sgRNA towards dG(1) of the TD strand (Figure 3C). The side chain of residue F898 is located above the nucleobases of the G(-1):C(-68) base pair, and residues K53 and Q1098 form hydrogen bonds with the phosphate group between G(-1) and C(1) (Figure 3C). The cleavage assays showed that the K53A, K875A and R915A mutations greatly decreased the cleavage activity, suggesting that these residues play important roles in facilitating the formation of the C(1):dG(1) base pair (Figure 3D).

Recognition of the sgRNA-target DNA heteroduplex

The pre-organized seed region of the crRNA plays a critical role in target recognition in CRISPR-Cas9, -Cas12a, -Cas12b and -Cas12i systems (47–52). To determine the possible seed region within the Cas12c1 sgRNA, we prepared 20 target duplex variants by tiling 1 nt mismatches across the sgRNA guide region and tested the effects of sgRNA-target duplex mismatches on Cas12c1 cleavage activity. The results showed that mismatches of the nucleotides C(1)–C(7) of the sgRNA guide region greatly impaired target cleavage, whereas mismatches of the nucleotides G(8)–A(20) were tolerated (Figure 4A), suggesting that the seed

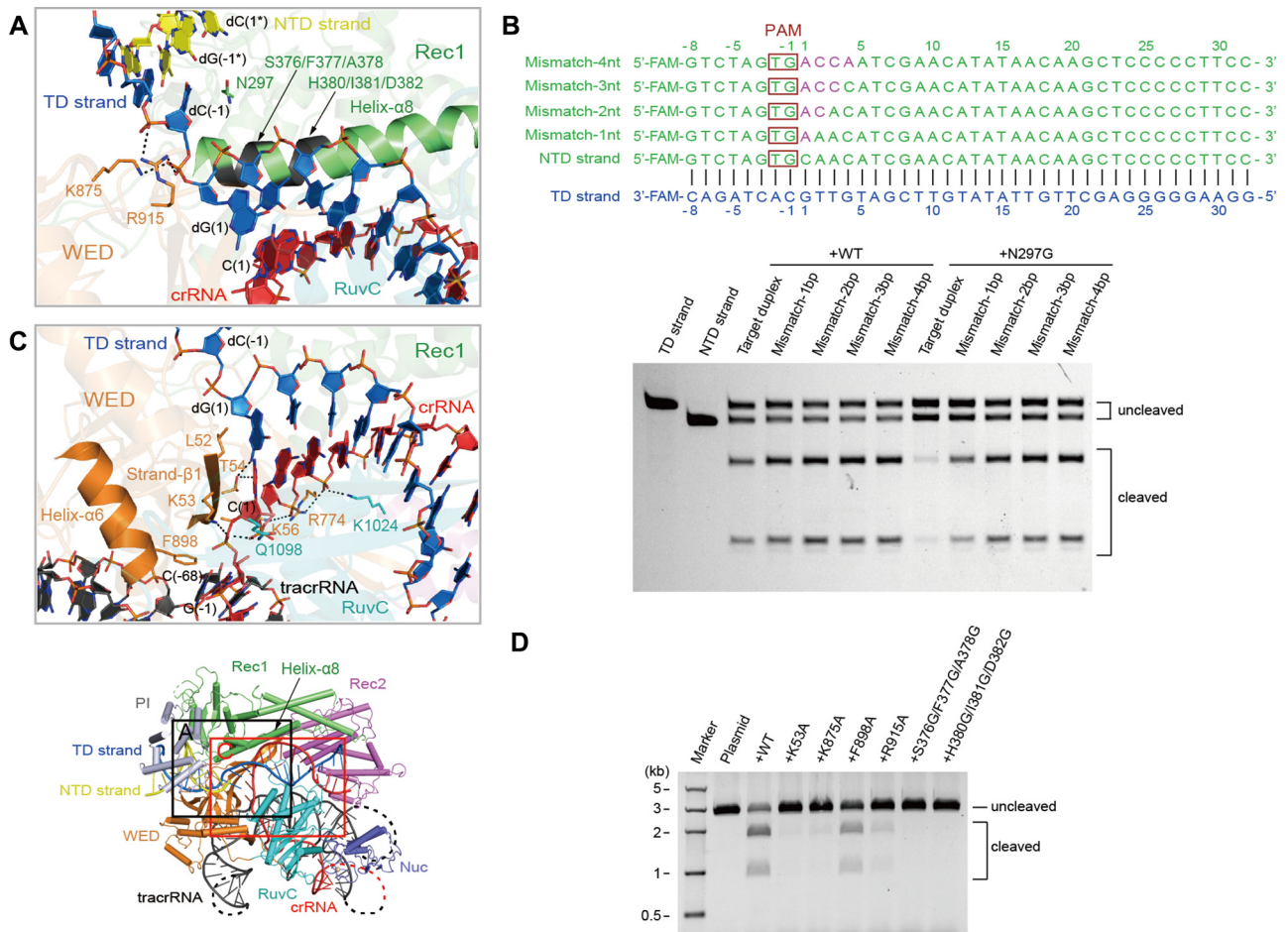


Figure 3. Target DNA duplex unwinding and heteroduplex formation. (A) Detailed interactions between the unzipped target duplex and Cas12c1 at the PAM-proximal region of the Cas12c1 ternary complex. Residue N297 facilitates unwinding of the target duplex. The long helix $\alpha 8$ of the Rec1 domain is labeled. (B) Top: schematic representations of the fluorescently labeled target duplex and 1–4 nt mismatched NTD strands designed to form 1–4 bp mismatched duplexes with the TD strand. Bottom: denaturing gel demonstrating the cleavage of the fluorescently labeled target duplex and mismatched duplexes by wild-type Cas12c1 and the N297G mutant in complex with sgRNA. (C) Detailed interactions between the PAM-proximal region of the heteroduplex and Cas12c1. (D) Agarose gel demonstrating the cleavage of the linearized plasmid by wild-type Cas12c1 and its mutants in complex with sgRNA.

region of the Cas12c sgRNA is within the first seven nucleotides at the 5'-end of the sgRNA guide region. Residues K56, R774, K1024 and K516 help in stabilizing the sugar-phosphate backbone of the seed region (Figure 3C; Supplementary Figure S5).

The 14 bp heteroduplex in the ternary complex A is sandwiched by the two Rec domains on one side and the RuvC domain on the other side (Figure 1B, C). Helix $\alpha 8$ of the Rec1 domain lies above the groove formed by the sgRNA guide and the TD strand (Figure 3A). The PAM-proximal region of helix $\alpha 8$ helps with stabilizing the conformation of the nucleotides dG(1)–dG(4) that pair with the complementary sgRNA seed nucleotides. The cleavage assays demonstrated that the S376G/F377G/A378G or H380G/I381G/D382G triple mutation in the PAM-proximal region of helix $\alpha 8$ abolished the cleavage activity of Cas12c1 (Figure 3D). A plausible explanation for such a result is that these mutations could disrupt the conformation of the PAM-proximal region of helix $\alpha 8$ and thereby undermine correct base pairing between the sgRNA

seed nucleotides and the PAM-proximal nucleotides of the TD strand. Moreover, a glycine/lysine-rich loop (residues K509–Q527, contains four glycine and six lysine residues) connects the Rec1 and Rec2 domains. This loop region runs in parallel and has several interactions with the sgRNA guide region (Supplementary Figure S5).

A helix-loop-helix (HLH) motif (residues P1001–F1053) of the RuvC domain is situated between the heteroduplex and the sgRNA R:AR duplex-3 (Figure 4B). The loop region (residues S1011–S1029) of the HLH motif contains several positively charged residues and extends towards the Rec2 and WED domains. The HLH motif should undergo certain conformational rearrangement upon binding of the sgRNA and target duplex. In the ternary complex, we can discern that residues D969, E1060 and D1266 form the Cas12c1 active site. The side chain of R1233 is pointing towards the active site. Single mutations D969A, E1060A, D1266A or R1233A abolished the Cas12c1 cleavage activity (Figure 4C). Moreover, the lid motif [residues Y1061–F1084; may also be termed the helix-loop (HL) motif] of

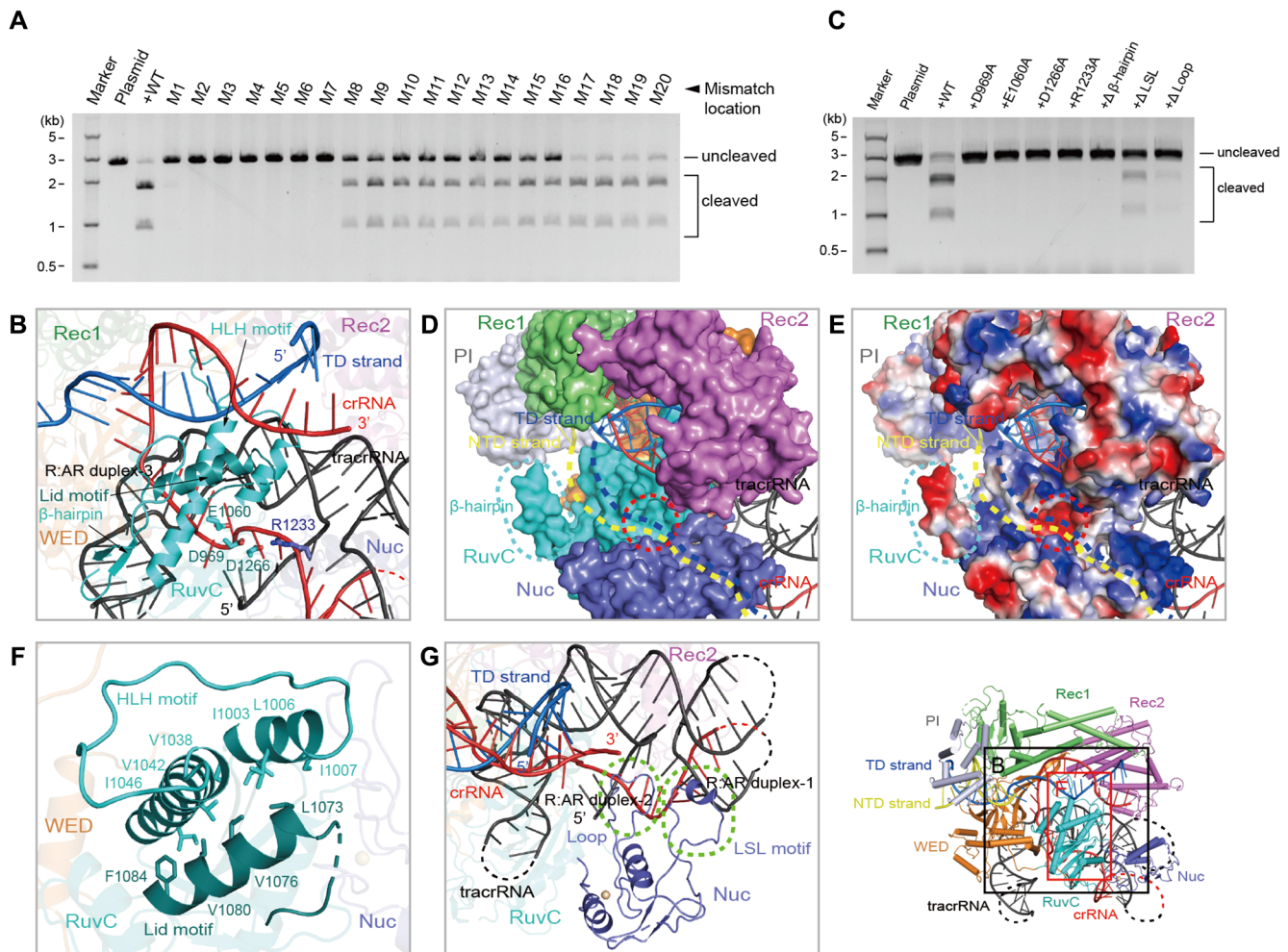


Figure 4. Recognition of the sgRNA-target DNA heteroduplex. (A) Agarose gel demonstrating the cleavage of the linearized plasmid containing target DNA sequence or its single-nucleotide mutants by wild-type Cas12c1 in complex with sgRNA. (B) The heteroduplex formed by the sgRNA guide region and the TD strand. The helix-loop-helix (HLH; residues P1001–F1053) and lid (residues Y1061–F1084) motifs, the β -hairpin (residues I1108–L1132) and critical residues for the DNase activity are labeled. Residue A969 of the catalytically inactive Cas12c1 ternary complex was virtually mutated back to D969. (C) Agarose gel demonstrating the cleavage of the linearized plasmid by wild-type Cas12c1 and the mutants in complex with sgRNA. (D, E) The possible loading pathways for the NTD and TD strands (colored in yellow and blue dashed lines, respectively) in surface and electrostatic potential surface representations of the Cas12c1 ternary complex. The β -hairpin and the active site of the RuvC domain are encircled with cyan and red dashed lines, respectively. Red, white and blue in (E) indicate negative, neutral and positive electrostatic potential surfaces, respectively. (F) The HLH and lid motifs neighboring each other in the RuvC domain. Hydrophobic interactions between two motifs are labeled. (G) The LSL motif (residues D1193–L1210) and loop region (residues P1236–S1247) of the Nuc domain interacting with the R:AR duplex-1 and duplex-2, respectively.

the RuvC domain is in a conformation in which the RuvC catalytic residues D969, E1060 and D1266 are almost entirely unblocked from accessing the substrate nucleotides (Figure 4B, D, E). Unblocking the conformation of the lid motif has been reported to be a prerequisite for releasing the DNase activities of Cas12a, Cas12i, and Cas12j (51, 53–55), implying that the lid motif of Cas12c1 should have a similar function. Of note, the HLH and lid motifs neighbor each other and maintain contacts through a number of hydrophobic interactions (Figure 4F), indicating that the motion of these two motifs may be coupled during the conformational activation of the Cas12c1 DNase activity.

We can also observe that a groove existed on the surface of the RuvC and Nuc domains, which should serve as a pathway guiding the unwound NTD strand into the RuvC active site (Figure 4D, E). The negative charge-rich β -

hairpin (residues I1108–L1132) in the RuvC domain is likely to act as a barrier and generate charge repulsion against the sugar-phosphate backbone of the NTD strand, which would further facilitate proper loading of the NTD strand into the RuvC active site. In line with this analysis, the cleavage assays showed that the Cas12c1 construct in which this β -hairpin is replaced with a GGSG linker lost its target cleavage activity (Figure 4C).

In the Nuc domain, a loop-short helix-loop (LSL) region (residues D1193–L1210) interacts with the sgRNA R:AR duplex-1, and another loop region (residues P1236–S1247) extends towards the sgRNA R:AR duplex-2 (Figure 4G). These two structural elements may restrict the relative positions between sgRNA and the Nuc domain, and facilitate cooperation between the Nuc and RuvC domains in the cleavage of the NTD and TD strands. Correspondingly,

substituting either of these two regions with a GGSG linker results in a significant decrease in the cleavage activity (Figure 4C).

Notably, the Rec2 domain generates a physical obstacle for further extension of the heteroduplex at its PAM-distal region (Figure 1B, C), so that the PAM-distal region of the heteroduplex has to make a sharp turn towards the space between the REC and NUC lobes. This anticipatedly leads to redirection of the TD strand towards the RuvC active site as observed in the ternary complex B (Supplementary Figure S1B, C). Besides generating a physical obstacle, the Rec2 domain most probably distorts the base pairing of the heteroduplex at its PAM-distal region (Figure 1B, E; Supplementary Figure S1B, C), which may facilitate the heteroduplex unwinding around this region.

The *cis* cleavage of the NTD and TD strands

To determine the cleavage site on each strand of the target duplex, the NTD and TD strands were labeled with 5'-6FAM and 3'-6FAM fluorescent dyes, respectively, and annealed. The cleavage results showed that the NTD and TD strands were cleaved at 14 nt and 23 nt after the PAM duplex, respectively (Figure 5A). Compared with the multiple cleavage sites reported to be generated by Cas12a, Cas12b, Cas12e, Cas12i and Cas12j on the NTD and/or TD strand (25,48,50,52,54,56,57), Cas12c1 appears to produce a single and uniform cleavage site on the NTD and TD strands. This feature of Cas12c1 would be highly useful for high-fidelity genome editing.

The sequential *cis* cleavage of the NTD and TD strands has been reported in detail for Cas12a and Cas12i. Each of the two enzymes displays different cleavage capabilities for the NTD and TD strands, with the NTD strand generally cleaved faster than the TD strand (53,54,58). Interestingly, our time-course cleavage assay indicated that Cas12c1 cleaves both NTD and TD strands with similar capabilities (Figure 5A), and the phenomenon of higher cleavage capacity for the NTD strand than that for the TD strand was not observed. A possible explanation is that Cas12c1 unwinds the PAM-distal heteroduplex earlier and/or more easily than other type V effectors such as Cas12a and Cas12i. This would allow the RuvC active site of Cas12c1 to access and cleave the unwound PAM-distal end of the TD strand and the NTD strand with similar probabilities.

To investigate the minimal length of target DNA capable of activating the DNase activity of Cas12c1, we performed cleavage assays with different lengths of target dsDNA or ssDNA. The results indicated that the target dsDNA containing a 16 bp protospacer sequence or 16 nt target ssDNA can induce the full activation of the DNase activity of Cas12c1 (Figure 5B).

Temperature- and metal ion-dependent cleavage

To understand how temperature and divalent metal ions affect the target cleavage activity of Cas12c1, we performed cleavage assays at different temperatures or with the addition of different metal ions. The results showed that Cas12c1 can cleave target DNA between 33°C and 49°C, and the optimal cleavage activity was observed at ~45°C (Figure

5C). Among the tested metal ions, Mg²⁺ or Ca²⁺ ions were required by Cas12c1 for substrate cleavage (Figure 5D). The two-Mg²⁺-ion catalysis is utilized by Cas effectors such as Cas9, Cas12a and Cas12i for DNA cleavage (52,59,60). Ca²⁺ has also been found to support substrate cleavage in Cas12a from *Francisella novicida* and Cas12b from *Alicyclobacillus acidoterrestris* (50,60). However, Ca²⁺ usually does not support two-metal-ion-dependent phosphoryl transfer reactions because of its large atomic radius and coordination geometry (61). Although coordination patterns of Mg²⁺ and Ca²⁺ involved in the substrate catalysis may vary slightly, which requires further investigation, both ions support the substrate catalysis in Cas12c1. In addition, Mn²⁺ supports a very weak substrate cleavage in Cas12c1 (Figure 5D).

DISCUSSION

Structural comparison between Cas12c1 and other type V effectors of similar size reveals that the overall architecture of Cas12c1 differs from those of Cas12b and Cas12i but resembles those of Cas12a and Cas12e (Supplementary Figure S2). Even so, secondary structure compositions and ternary folds of all Cas12c1 domains except the RuvC domain are markedly different from their counterparts in Cas12a and Cas12e. This is in line with the lack of significant primary sequence similarity outside the RuvC domain. Moreover, the domain arrangement of Cas12c1 matches that of Cas12e, as the PI domain of Cas12c1 and the functionally distinct NTSB domain of Cas12e are connected to respective Rec1 domains, whereas the PI domain of Cas12a is connected to its WED domain. Altogether, despite the dissimilarities in domain structure, Cas12c1 displays a considerable degree of structural convergence with Cas12a and Cas12e.

Consistent with the hypothesis of the independent evolution of different type V systems (4,5), the sequence and architecture of Cas12c1 sgRNA are quite distinct from those of its Cas12b and Cas12e analogs. Interestingly, all three sgRNAs constitute a notably larger portion of the mass in their binary complexes with the cognate Cas effectors [~23% in Cas12c1, ~26% in *Deltaproteobacteria* Cas12e (DpbCas12e) and ~22% in *Alicyclobacillus acidoterrestris* Cas12b (AacCas12b)] than in the single RNA-guided effectors *Francisella novicida* Cas12a (FnCas12a) (~8%) and Cas12i1 (~10%), or in the dual RNA-guided type II effector SpyCas9 (~17%). Such large, solvent-exposed sgRNA scaffolds are in stark contrast to crRNAs of Cas12a and Cas12i effectors. Furthermore, there is a strong possibility that the solvent-exposed regions of Cas12c1 sgRNA (such as the R:AR duplex-1, stem-2 and stem-3) can be simplified without negatively affecting the target cleavage activity of Cas12c1.

The mechanism of PAM readout by Cas12c1 is special among studied type V effectors because Cas12c1 reshapes the conformation of the PAM region, which results in the involvement of the neighboring nucleotide dA(-2) in the recognition of the PAM nucleotide dG(-1*). In Cas12a, the main element mediating PAM recognition, i.e. the loop–lysine helix–loop (LKL) element of the PI domain, acts as a plough that inserts into the PAM duplex to mediate PAM

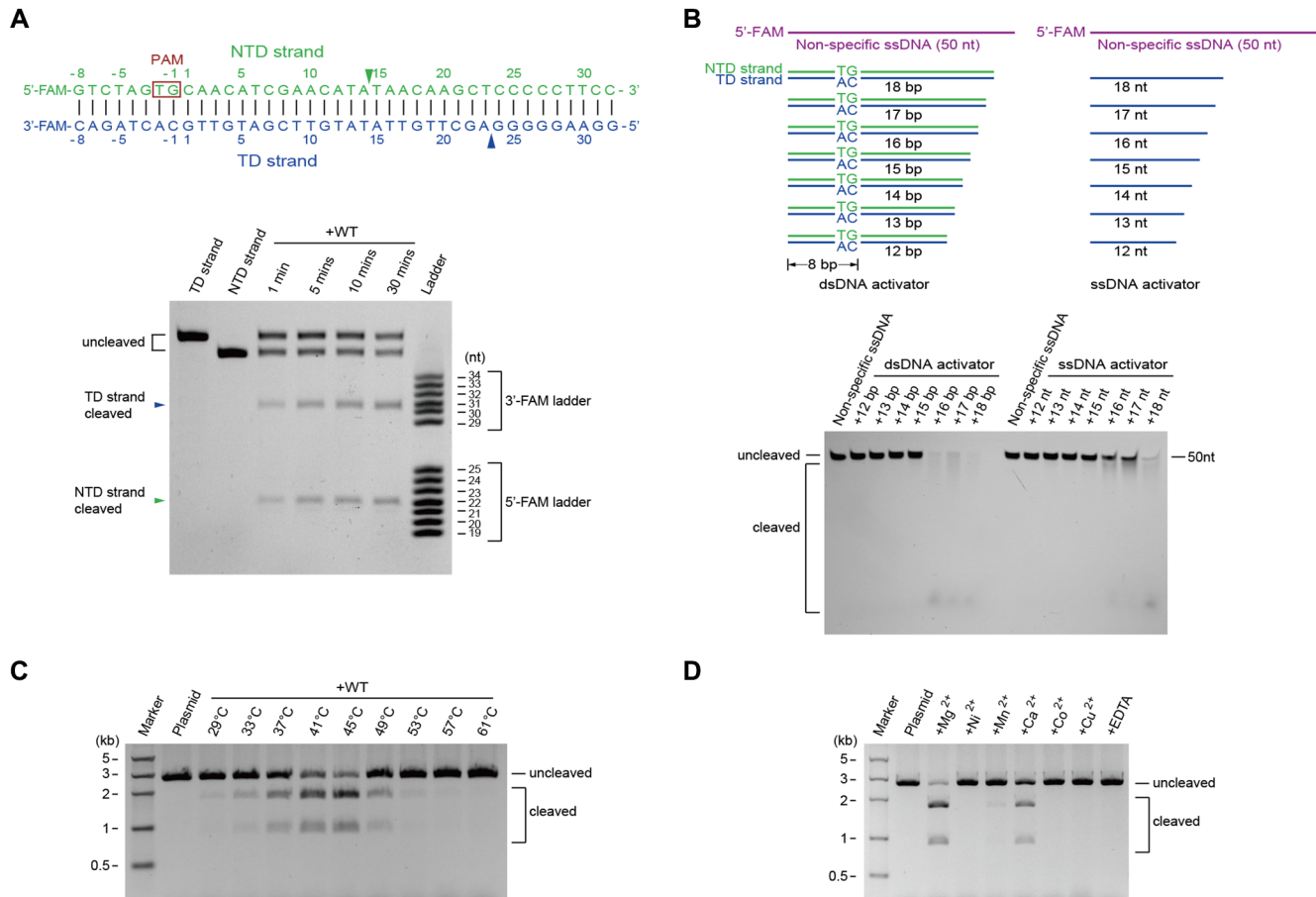


Figure 5. The *cis* cleavage of the target DNA duplex and minimal length of the target. (A) Top: schematic representations of the fluorescently labeled target duplex used in the cleavage assays. Bottom: denaturing gel demonstrating the time-course cleavage of the fluorescently labeled target duplex by wild-type Cas12c1 in complex with sgRNA. The *cis* cleavage products of the NTD and TD strands are indicated with a green and a blue triangle, respectively. (B) Top: schematic representations of dsDNA activators and ssDNA activators with different lengths. Bottom: denaturing gel demonstrating the cleavage of the fluorescently labeled non-specific ssDNA by wild-type Cas12c1 in complex with sgRNA and in the presence of a dsDNA or ssDNA activator. (C) Agarose gel demonstrating the cleavage of the linearized plasmid by wild-type Cas12c1 in complex with sgRNA at different temperatures. (D) Agarose gel demonstrating the cleavage of the linearized plasmid by wild-type Cas12c1 in complex with sgRNA and in the presence of different divalent metal ions or EDTA.

recognition and target duplex unwinding in concert with other regions of Cas12a (53,56). The role of ‘pins’ in the LKL loop is performed by three highly conserved lysine residues, of which the middle lysine inserts into the PAM region similarly to R152 of Cas12c1 for base- and shape-specific PAM readout (56). However, LKL is structurally and mechanistically distinct from the PI loop of Cas12c1, and is directly involved in target duplex unwinding.

During the submission of this manuscript, Kurihara *et al.* reported cryo-electron microscopy structures of Cas12c2-sgRNA-target DNA and Cas12c2-sgRNA complexes (42). Our experimental results show that, like Cas12c2 (42), Cas12c1 processes pre-sgRNA autonomously using the RuvC catalytic site via a Mg²⁺ ion-dependent mechanism (Supplementary Figure S8). The length and location of seed regions in both systems also appear largely similar — while the experimental data in this study indicate that the seed region in the CRISPR-Cas12c1 system is within the first seven nucleotides of the sgRNA guide, the overlay of Cas12c2 binary and ternary complexes demonstrated that the first six

nucleotides of the sgRNA guide region are pre-organized for interrogation of target DNA sequence, suggesting that these six nucleotides function as the seed region in Cas12c2 sgRNA (42). Current findings also suggest that certain similarities exist in the modes by which Cas12c1 and Cas12c2 recognize the PAM nucleotide dT(-2*). Both effectors similarly utilize two arginine residues that have been shown to be crucial for this task, namely R152 (R137 in Cas12c2) in the PI loop and R374 (R351 in Cas12c2) in the Rec1 domain (42). However, while the PAM nucleotide dG(-1*) in Cas12c1 is specifically recognized by the residue N299 and the nucleotide dA(-2), dG(-1*) in Cas12c2 lacks a specific interaction. Moreover, the sequences and lengths of PI loops are greatly divergent between Cas12c1 and Cas12c2.

Interestingly, Cas12c2 lacks target DNA cleavage activity under the tested conditions even though its RuvC active site is similar to those in other type V CRISPR-Cas effectors, including Cas12c1. In this study, Cas12c1 in complex with sgRNA was demonstrated to be competent for target DNA cleavage, which is in line with findings of a pre-

vious study on Cas12c1 (62). A recent study by Huang *et al.* further demonstrated that Cas12c2 and another Cas12c ortholog (Cas12c.4) carry out interference not by cleaving but by binding target DNA (63). The same study also showed that Cas12c.4 exhibits RNase but not DNase activity (63). Cas12c2 may behave like Cas12c.4 given that they differ in only seven amino acid residues, compared with 28.56% sequence identity between Cas12c2 and Cas12c1. Analysis of Cas12c1 and Cas12c2 structures in their respective ternary complexes indicates that their RuvC domains adopt a similar fold. However, the difference is clearly reflected in distinct electrostatic potential surfaces surrounding respective RuvC active sites, which appears to be more positive in Cas12c1 than in Cas12c2 (Supplementary Figure S9A, B). It is worth noting that, due to the absence of the Nuc domain in the Cas12c2 ternary complex, the analysis of electrostatic potential surfaces surrounding the RuvC active sites of Cas12c1 and Cas12c2 is incomplete. Furthermore, the overlay shows that overall architectures of the Cas12c1 and Cas12c2 ternary complexes resemble each other. The root-mean-square deviation (RMSD) value is 2.62 Å between 807 aligned backbone C α atoms of Cas12c1 and Cas12c2. The Rec2 domain appears to interact differently with the PAM-distal heteroduplex in the two effectors (Supplementary Figure S9C). In Cas12c1, the Rec2 domain redirects the TD strand towards the RuvC catalytic site (Supplementary Figures S1C and S9C). In contrast, the Rec2 domain of Cas12c2 appears to guide the TD strand away from the RuvC catalytic site and towards the space between two Rec domains (Supplementary Figure S9B, C). How precisely these and/or other factors may result in competent DNase activity in Cas12c1 and its absence in Cas12c2 remains to be investigated.

Thus far, the conformational changes which Cas12c1 undergoes between binary and ternary complexes as well as between intermediate states in ternary complex remain elusive. Whether the conformational changes are regulated by checkpoints such as Rec linker, lid and finger regions in Cas12a also needs further investigation (53). The conformational changes in Cas12c1 may be at least to some degree comparable with Cas12c2. During transition from a binary to a ternary complex, Cas12c2 changes from a closed to an open conformation, and the largest conformational changes occur in its Rec2 domain (42). Moreover, the PI domain is disordered in Cas12c2 binary complex and becomes stabilized upon interaction with PAM duplex, which is also likely to be the case with the Cas12c1 PI domain given its connection to the rest of the protein via two loops.

One distinctive feature of Cas12c1 observed in this study is the indistinguishable order and cleavage capacity of NTD and TD strand *cis* cleavage, which is different from the sequential NTD–TD strand cleavage with higher NTD strand cleavage capacity observed for Cas12a and Cas12i (53,54,58). As Cas12c1 employs its Rec2 domain to prevent heteroduplex elongation and possibly distort base pairing within the PAM-distal heteroduplex markedly earlier than in Cas12a and Cas12i, it is likely that the route by which the TD strand reaches the RuvC active site of Cas12c1 is more direct and thereby more effective than in the other type V CRISPR-Cas effectors, explaining why cleavage of the TD strand is executed with capacity comparable with that of the

NTD strand. This suggests that the specific ways in which NTD and TD strands are loaded into the RuvC active site are also important factors determining the cleavage order and capacity for the two target DNA strands.

In summary, we determined the crystal structure of the Cas12c1-sgRNA-target dsDNA ternary complex and reveal the molecular basis for PAM recognition, target dsDNA unwinding, heteroduplex formation and recognition, as well as cleavage of NTD and TD strands, and characterized temperature and metal ion preferences for the cleavage activity of Cas12c1. Our findings suggest that the CRISPR-Cas12c1 system displays great potential as an efficient and high-fidelity DNA-targeting tool and provide useful information for future studies aimed at engineering and adapting it for genome editing and other applications.

DATA AVAILABILITY

The Cas12c1 ternary complex has been deposited in the Protein Data Bank under the accession code 7VYX. Other data are available from the corresponding author upon reasonable request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The diffraction data were collected at the beamline BL-17U1 of Shanghai Synchrotron Radiation Facility (SSRF). *Author contributions:* B.Z. and S.O. designed the experiments. S.O. and B.Z. supervised the study. B.Z., J.L. and V.P. prepared Cas12c1, its mutants and the Cas12c1 complexes. B.Z., V.P. and J.L. prepared the *in vitro* transcribed RNA. J.L., V.P. and B.Z. performed the cleavage assays. B.Z., J.L. and V.P. performed the crystallization screening and optimized the crystallization conditions. B.Z. collected X-ray diffraction data and solved the crystal structure. Y.L. and Q.L. helped with the cleavage assays. Q.L. and J.C. helped with the preparation of the *in vitro* transcribed RNA. B.Z. and V.P. wrote the manuscript. B.Z. prepared the figures. S.O., B.Z. and V.P. revised the manuscript.

FUNDING

The National Nature Science Foundation of China [82225028 and 82172287]; the National Key Research and Development Program of China [2021YFC2301403]; the Special Funds of the Central Government Guiding Local Science and Technology Development [2020L3008]; and the High-level personnel introduction grant of Fujian Normal University [Z0210509]. Funding for open access charge: The National Nature Science Foundation of China [82172287 and 82225028].

Conflict of interest statement. None declared.

REFERENCES

1. 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.

2. Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) The biology of CRISPR-Cas: backward and forward. *Cell*, **172**, 1239–1259.
3. Koonin, E.V. and Makarova, K.S. (2019) Origins and evolution of CRISPR-Cas systems. *Philos. Trans. R. Soc. B: Biol. Sci.*, **374**, 20180087.
4. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P. *et al.* (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.*, **18**, 67–83.
5. Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I. *et al.* (2017) Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
6. 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.
7. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
8. Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
9. Wang, H., La Russa, M. and Qi, L.S. (2016) CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.*, **85**, 227–264.
10. Wright, A.V., Nunez, J.K. and Doudna, J.A. (2016) Biology and applications of CRISPR systems: harnessing nature's toolbox for genome engineering. *Cell*, **164**, 29–44.
11. Jiang, F. and Doudna, J.A. (2017) CRISPR-Cas9 structures and mechanisms. *Annu. Rev. Biophys.*, **46**, 505–529.
12. 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.
13. Kim, D., Kim, J., Hur, J.K., Been, K.W., Yoon, S.H. and Kim, J.S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.*, **34**, 863–868.
14. Kleinstiver, B.P., Tsai, S.Q., Prew, M.S., Nguyen, N.T., Welch, M.M., Lopez, J.M., McCaw, Z.R., Aryee, M.J. and Joung, J.K. (2016) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.*, **34**, 869–874.
15. Knott, G.J. and Doudna, J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*, **361**, 866–869.
16. Pickar-Oliver, A. and Gersbach, C.A. (2019) The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.*, **20**, 490–507.
17. Smargon, A.A., Shi, Y.J. and Yeo, G.W. (2020) RNA-targeting CRISPR systems from metagenomic discovery to transcriptomic engineering. *Nat. Cell Biol.*, **22**, 143–150.
18. Anzalone, A.V., Koblan, L.W. and Liu, D.R. (2020) Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.*, **38**, 824–844.
19. Tong, B., Dong, H., Cui, Y., Jiang, P., Jin, Z. and Zhang, D. (2020) The versatile type V CRISPR effectors and their application prospects. *Front. Cell Dev. Biol.*, **8**, 622103.
20. Perculija, V., Lin, J., Zhang, B. and Ouyang, S. (2021) Functional features and current applications of the RNA-targeting type VI CRISPR-Cas systems. *Adv. Sci. (Weinh.)*, **8**, 2004685.
21. Komor, A.C., Badran, A.H. and Liu, D.R. (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, **168**, 20–36.
22. Paul, B. and Montoya, G. (2020) CRISPR-Cas12a: functional overview and applications. *Biomed. J.*, **43**, 8–17.
23. Teng, F., Cui, T., Feng, G., Guo, L., Xu, K., Gao, Q., Li, T., Li, J., Zhou, Q. and Li, W. (2018) Repurposing CRISPR-Cas12b for mammalian genome engineering. *Cell Discov.*, **4**, 63.
24. Strecker, J., Jones, S., Koopal, B., Schmid-Burgk, J., Zetsche, B., Gao, L., Makarova, K.S., Koonin, E.V. and Zhang, F. (2019) Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.*, **10**, 212.
25. Liu, J.J., Orlova, N., Oakes, B.L., Ma, E., Spinner, H.B., Baney, K.L.M., Chuck, J., Tan, D., Knott, G.J., Harrington, L.B. *et al.* (2019) CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, **566**, 218–223.
26. Harrington, L.B., Burstein, D., Chen, J.S., Paez-Espino, D., Ma, E., Witte, I.P., Cofsky, J.C., Kyrpides, N.C., Banfield, J.F. and Doudna, J.A. (2018) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, **362**, 839–842.
27. Xu, X., Chemparathy, A., Zeng, L., Kempton, H.R., Shang, S., Nakamura, M. and Qi, L.S. (2021) Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol. Cell*, **81**, 4333–4345.
28. Wu, Z., Zhang, Y., Yu, H., Pan, D., Wang, Y., Wang, Y., Li, F., Liu, C., Nan, H., Chen, W. *et al.* (2021) Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat. Chem. Biol.*, **17**, 1132–1138.
29. Bigelyte, G., Young, J.K., Karvelis, T., Budre, K., Zedaveinyte, R., Djukanovic, V., Van Ginkel, E., Paulraj, S., Gasior, S., Jones, S. *et al.* (2021) Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nat. Commun.*, **12**, 6191.
30. Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
31. Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J. and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology (Reading)*, **155**, 733–740.
32. Leenay, R.T. and Beisel, C.L. (2017) Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.*, **429**, 177–191.
33. Marraffini, L.A. and Sontheimer, E.J. (2010) Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*, **463**, 568–571.
34. Collias, D. and Beisel, C.L. (2021) CRISPR technologies and the search for the PAM-free nuclease. *Nat. Commun.*, **12**, 555.
35. Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
36. 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.
37. Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A. and Banfield, J.F. (2017) New CRISPR-Cas systems from uncultivated microbes. *Nature*, **542**, 237–241.
38. Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V. *et al.* (2019) Functionally diverse type V CRISPR-Cas systems. *Science*, **363**, 88–91.
39. Strecker, J., Ladha, A., Gardner, Z., Schmid-Burgk, J.L., Makarova, K.S., Koonin, E.V. and Zhang, F. (2019) RNA-guided DNA insertion with CRISPR-associated transposases. *Science*, **365**, 48–53.
40. Pausch, P., Al-Shayeb, B., Bisom-Rapp, E., Tsuchida, C.A., Li, Z., Cress, B.F., Knott, G.J., Jacobsen, S.E., Banfield, J.F. and Doudna, J.A. (2020) CRISPR-CasPhi from huge phages is a hypercompact genome editor. *Science*, **369**, 333–337.
41. Harrington, L.B., Ma, E., Chen, J.S., Witte, I.P., Gertz, D., Paez-Espino, D., Al-Shayeb, B., Kyrpides, N.C., Burstein, D., Banfield, J.F. *et al.* (2020) A scoutRNA is required for some type V CRISPR-Cas systems. *Mol. Cell*, **79**, 416–424.
42. Kurihara, N., Nakagawa, R., Hirano, H., Okazaki, S., Tomita, A., Kobayashi, K., Kusakizako, T., Nishizawa, T., Yamashita, K., Scott, D.A. *et al.* (2022) Structure of the type V-C CRISPR-Cas effector enzyme. *Mol. Cell*, **82**, 1865–1877.
43. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, **276**, 307–326.
44. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D: Biol. Crystallogr.*, **66**, 213–221.
45. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D: Biol. Crystallogr.*, **60**, 2126–2132.
46. Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F. and Murshudov, G.N. (2004) REFMAC5

- dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D: Biol. Crystallogr.*, **60**, 2184–2195.
47. Jiang, F., Zhou, K., Ma, L., Gressel, S. and Doudna, J.A. (2015) A Cas9–guide RNA complex preorganized for target DNA recognition. *Science*, **348**, 1477–1481.
 48. Swarts, D.C., van der Oost, J. and Jinek, M. (2017) Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol. Cell*, **66**, 221–233.
 49. Yang, H., Gao, P., Rajashankar, K.R. and Patel, D.J. (2016) PAM-dependent target DNA recognition and cleavage by C2c1 CRISPR-Cas endonuclease. *Cell*, **167**, 1814–1828.
 50. Liu, L., Chen, P., Wang, M., Li, X., Wang, J., Yin, M. and Wang, Y. (2017) C2c1–sgRNA complex structure reveals RNA-guided DNA cleavage mechanism. *Mol. Cell*, **65**, 310–322.
 51. Zhang, H., Li, Z., Xiao, R. and Chang, L. (2020) Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat. Struct. Mol. Biol.*, **27**, 1069–1076.
 52. Huang, X., Sun, W., Cheng, Z., Chen, M., Li, X., Wang, J., Sheng, G., Gong, W. and Wang, Y. (2020) Structural basis for two metal-ion catalysis of DNA cleavage by Cas12i2. *Nat. Commun.*, **11**, 5241.
 53. Stella, S., Mesa, P., Thomsen, J., Paul, B., Alcon, P., Jensen, S.B., Saligram, B., Moses, M.E., Hatzakis, N.S. and Montoya, G. (2018) Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity. *Cell*, **175**, 1856–1871.
 54. Zhang, B., Luo, D., Li, Y., Perculija, V., Chen, J., Lin, J., Ye, Y. and Ouyang, S. (2021) Mechanistic insights into the R-loop formation and cleavage in CRISPR-Cas12i1. *Nat. Commun.*, **12**, 3476.
 55. Pausch, P., Soczek, K.M., Herbst, D.A., Tsuchida, C.A., Al-Shayeb, B., Banfield, J.F., Nogales, E. and Doudna, J.A. (2021) DNA interference states of the hypercompact CRISPR-CasPhi effector. *Nat. Struct. Mol. Biol.*, **28**, 652–661.
 56. Stella, S., Alcon, P. and Montoya, G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*, **546**, 559–563.
 57. Carabias, A., Fuglsang, A., Temperini, P., Pape, T., Sofos, N., Stella, S., Erlendsson, S. and Montoya, G. (2021) Structure of the mini-RNA-guided endonuclease CRISPR-Cas12j3. *Nat. Commun.*, **12**, 4476.
 58. Swarts, D.C. and Jinek, M. (2019) Mechanistic insights into the cis- and trans-acting DNase activities of Cas12a. *Mol. Cell*, **73**, 589–600.
 59. 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.
 60. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. and Charpentier, E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517–521.
 61. Yang, W., Lee, J.Y. and Nowotny, M. (2006) Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
 62. Wang, Z. and Zhong, C. (2021) Cas12c-DETECTOR: a specific and sensitive Cas12c-based DNA detection platform. *Int. J. Biol. Macromol.*, **193**, 441–449.
 63. Huang, C.J., Adler, B.A. and Doudna, J.A. (2022) A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. *Mol. Cell*, **82**, 2148–2160.