

Structural and mechanistic insights into the sequential dsDNA cleavage by SpCas12f1

Julene Madariaga-Marcos^{1,†}, Marius Baltramonaitis^{2,†}, Selgar Henkel-Heinecke^{1,†},
 Dominik J. Kauert¹, Patrick Irmisch¹, Greta Bigelyte-Stankeviciene², Arunas Silanskas²,
 Tautvydas Karvelis², Virginijus Siksnys², Giedrius Sasnauskas^{2,*}, Ralf Seidel^{1,*}

¹Peter Debye Institute for Soft Matter Physics, Universität Leipzig, Leipzig 04103, Germany

²Institute of Biotechnology, Vilnius University Life Sciences Center, Vilnius 10257, Lithuania

*To whom correspondence should be addressed. Email: ralf.seidel@physik.uni-leipzig.de

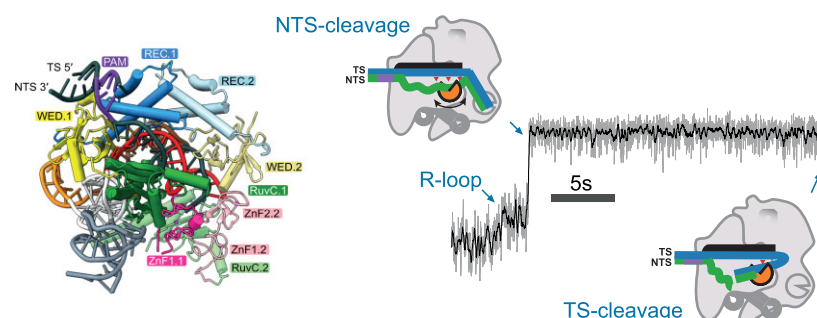
Correspondence may also be addressed to Giedrius Sasnauskas. Email: giedrius.sasnauskas@bti.vu.lt

[†]The first three authors should be regarded as Joint First Authors.

Abstract

Miniature CRISPR-Cas12f1 effector complexes have recently attracted considerable interest for genome engineering applications due to their compact size. Unlike other Class 2 effectors, Cas12f1 functions as a homodimer bound to a single ~200 nt RNA. While the basic biochemical properties of Cas12f1, such as its use of a single catalytic center for catalysis, have been characterized, the orchestration of the different events occurring during Cas12f1 reactions remained little explored. To gain insights into the dynamics and mechanisms involved in DNA recognition and cleavage by Cas12f1 from *Syntrophomonas palmitica* (SpCas12f1), we solved the structure of SpCas12f1 bound to target DNA and employed single-molecule magnetic tweezers measurements in combination with ensemble kinetic measurements. Our data indicate that SpCas12f1 forms 18 bp R-loops, in which local contacts of the protein to the R-loop stabilize R-loop intermediates. DNA cleavage is catalyzed by a single SpCas12f1 catalytic center, which first rapidly degrades a ~11 bp region on the nontarget strand by cutting at random sites. Subsequent target strand cleavage is slower and requires at least a nick in the nontarget strand.

Graphical abstract



Introduction

CRISPR-Cas systems, short for Clustered Regularly Interspaced Short Palindromic Repeats–CRISPR-Associated systems, serve as defense mechanisms of prokaryotic organisms against invading nucleic acids [1–3]. This protection is carried out by Cas effector complexes, which use CRISPR RNA as guides to selectively bind and cleave foreign DNA or RNA. Class 2 CRISPR-Cas systems encode single multidomain effector proteins, such as Cas9 (type II) and Cas12 (type V), which have been successfully repurposed for genome engineering applications [4–7].

CRISPR-Cas12 nucleases, which evolved from transposon-associated TnpB nucleases, form a highly diverse family [8, 9], with protein sizes ranging from 400 (TnpB) to 1500 (Cas12a) amino acids (aa), and their RNAs ranging from 50 (Cas12a) to 200 (Cas12f1, Cas12l) nucleotides. The type V-A Cas12a effectors are best understood with respect to structure, DNA recognition, and DNA cleavage mechanism and are used for a broad range of genome-editing applications [10–12]. Cas12f1 effector nucleases from the type V-F (formerly type V-U3) CRISPR-Cas systems have recently attracted considerable interest due to their compact size (400–600 aa) [13,

Received: March 28, 2025. Revised: May 28, 2025. Editorial Decision: June 5, 2025. Accepted: June 11, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License

(<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

14], which makes them compatible with adeno-associated viral delivery vectors, which have a limited packaging capacity [15].

Similar to Cas12a, Cas12f1 recognizes double-stranded DNA (dsDNA) containing a protospacer-dependent motif (PAM) and introduces a double-strand break using a RuvC domain catalytic center [7, 14]. Atomic structures of Cas12f1 from an uncultured archaeon (UnCas12f1), *Acidibacillus sulfuroxidans* (AsCas12f1), and *Clostridium novyi* (CnCas12f1) surprisingly revealed that the active complex, in contrast to the other Cas12 nucleases, is a protein dimer with a composition of (Cas12f1)₂(RNA)₁, where the subunit 1 recognizes the PAM while subunit 2 stabilizes the formed R-loop and thus takes over the function of the REC1 and REC2 domains of Cas12a [16–19]. Furthermore, it has been demonstrated that a single catalytic center of the RuvC.1 domain (RuvC domain of the subunit 1) cleaves both the DNA target (TS) and the nontarget (NTS) strands, with the RuvC.2 domain (RuvC domain of the subunit 2) being catalytically inactive [16, 17]. It was additionally found that Cas12f, unlike other Cas12 nucleases, cleaves the NTS at two rather than one position—at ~13 and ~24 nt from the PAM [15, 20]. The TS is cleaved at ~22 nt from the PAM, i.e. similarly to the second NTS cut in the DNA duplex behind the R-loop [15, 20].

To date, a number of Cas12f1 variants have been successfully employed for genome editing [15, 18, 21–25]. However, despite the considerable progress in Cas12f1 applications, important mechanistic details of R-loop formation, as well as the dynamics and coordination of the three single-stranded DNA (ssDNA) cuts by a single RuvC site, are still missing [19, 18].

Since understanding the interplay of Cas12f1 structure and dynamics is imperative for further improving its performance as a genome editing tool, in this study we aimed to gain new insights into the structural and molecular mechanism of Cas12f1. To this end, we have solved the atomic structure of SpCas12f1 from *Syntrophomonas palmitatica* bound to sgRNA and target DNA. The structure revealed the atomic details of the PAM recognition as well as stabilization of the 18 nt R-loop and provided new insights for the design of sgRNA truncations. To characterize the dynamics of the multi-step DNA binding and cleavage by the SpCas12f1, we employed single-molecule magnetic tweezers experiments [26–28] in combination with ensemble DNA cleavage measurements [29]. In brief, we probed the length of untwisted DNA during R-loop formation [30–32] for different spacer lengths of the sgRNA, supporting a maximum R-loop length of 18 nt at which the R-loop is most stabilized, as well as R-loop intermediates on the path to full R-loop formation. Furthermore, we resolved that initially the NTS is nicked at ~10 s after R-loop formation and further rapidly trimmed such that a defined ~12 bp fragment is excised. Subsequent TS cleavage requires at least a single nick in the NTS and occurs on a minute time scale. Interestingly, TS cleavage is accompanied by the trapping of the PAM distal end of the DNA duplex (called clamping) by the Cas12f1 complex, although it is dispensable for the actual cleavage step [31, 32].

Materials and methods

Purification of SpCas12f1

For SpCas12f1, dSpCas12f1 (D228A), and SpCas12f1-PAM^{Mut} (N88A, Q89A, and N92A) protein purification,

Escherichia coli Arctic Express (DE3) strain was transformed with pMBP-SpCas12f1 (10.1038/s41467-021-26469-4), pTK411 (produced from pMBP-SpCas12f1 by introducing D228A mutation using Phusion Site Directed Mutagenesis Kit (Thermo Fisher Scientific)), and pTK1460 (produced from pMBP-SpCas12f1 by introducing N88A, Q89A, and N92A mutations), respectively. *E. coli* cells were grown in LB broth supplemented with ampicillin (100 µg/ml) and gentamicin (10 µg/ml) at 37°C. After reaching OD₆₀₀ of 0.6–0.8, the temperature was reduced to 16°C. The protein expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and cells were cultured for additional 16 h. Next, cells were pelleted by centrifugation, resuspended in 20 mM Tris-HCl (pH 8.0 at 25°C), 1.5 M NaCl, 5 mM 2-mercaptoethanol, 25 mM imidazole, 2 mM phenylmethylsulfonyl fluoride (PMSF) and 5% (v/v) glycerol buffer and disrupted by sonication. After centrifugation, the supernatant was loaded on Ni²⁺-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration in 20 mM Tris-HCl (pH 8.0), 0.5 M NaCl, and 5 mM 2-mercaptoethanol buffer. The fractions containing SpCas12f1 were pooled and loaded on a HiTrap heparin HP column (GE Healthcare). A linear gradient of increasing NaCl concentration was used for elution. The fractions containing the protein were pooled, and the 10× His-MBP tag was cleaved by incubating overnight with tobacco etch virus (TEV) protease at 4°C. The cleaved tag and TEV protease were removed with HiTrap heparin HP 5 column (GE Healthcare) using a linear gradient of increasing NaCl concentration. The collected fractions with SpCas12f1 were then dialyzed against 20 mM Tris-HCl (pH 8.0 at 25°C), 500 mM NaCl, 2 mM dithiothreitol (DTT), and 50% (v/v) glycerol and stored at –20°C. Protein sequences are provided in [Supplementary Table S1](#), links to the plasmid sequences are listed in [Supplementary Table S2](#).

sgRNA synthesis

DNA templates for *in vitro* transcription were generated by PCR using overlapping oligonucleotides or plasmids to add the T7 promoter sequence. RNAs were *in vitro* transcribed using the TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) and purified using the GeneJET RNA Purification Kit (Thermo Fisher Scientific). Sequences of sgRNAs used in this study are listed in [Supplementary Table S3](#).

Reconstitution of SpCas12f1-RNA complexes

To reconstitute SpCas12f1-sgRNA complexes, 1 µM SpCas12f1 was mixed with 1 µM sgRNA in assembly buffer [10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT] and incubated for 30 min at 37°C.

Cryogenic-electron microscopy sample preparation

The SpCas12f1-sgRNA-DNA complex was reconstituted by mixing SpCas12f1 and sgRNA with target DNA at a molar ratio of 1:1:1. First, SpCas12f1 protein was combined with sgRNA and incubated for 20 min on ice. Next, target DNA was added into the mix and incubated for additional 15 min on ice. Oligoduplex DNA was assembled by annealing target sequence containing oligonucleotides ([Supplementary Table S4](#)). The final reaction mixture contained 20 mM Tris-HCl

(pH 7.5 at 37°C), 230 mM NaCl, 10 mM MgCl₂, 1 mM DTT, and 1 mM EDTA. Finally, the sample (10 µM, 3 µl) was applied on a Quantifoil Cu 300 mesh R1.2/1.3 grid coated with graphene oxide (GO) using a waiting time of 15 s and a blotting time of 5 s under 95% humidity conditions at 4°C. GO coating was performed as described by Bokori-Brown *et al.* [33]. Briefly, GO dispersion (Merck) was diluted in ddH₂O to 0.2 mg/ml, spun down at 500 g for 60 s to remove aggregates, and applied on the carbon side of a freshly glow-discharged [50 mA, 1 min in a GloQube Plus device (Quorum Technologies)] grid. After incubation for 1 min, the excess of GO was removed using blotting paper. The grid was washed three times with ddH₂O, dried, and subsequently used for plunge-freezing into liquid ethane using Vitrobot Mark IV device (FEI).

Cryogenic-electron microscopy data collection, processing, model building, refinement, and analysis

Cryogenic-electron microscopy (cryo-EM) data collection on a Glacios microscope (Thermo Fisher Scientific) and processing in CryoSPARC [34, 35] are described in detail in the Supplementary data. An initial model for SpCas12f1 structure was obtained using AlphaFold 3 [36]. It was fitted into the cryo-EM maps using ChimeraX (v.1.7) [37]. Protein rebuilding and manual building of DNA and sgRNA were performed using Coot (v.0.9.8.1) [38]. Model refinement was performed using phenix.real_space_refine (v. 1.21.2-5419) [39]; refinement statistics are summarized in [Supplementary Table S5](#). Structure overlays and generation of structural images were performed using ChimeraX (v.1.7) [37]; protein interface surfaces were analyzed using the VoroContacts web server [40] with a 5.0 Å distance cut-off. The final model of the SpCas12f1:sgRNA:DNA ternary complex (PDB: 9i8y) covers aa 3–181, 187–329, 344–403, and 430–459 of the SpCas12f1 primary subunit; aa 5–144, 157–179, 189–218, 224–270, 292–329, 343–384, and 392–463 of the SCas12f1 secondary subunit; ribonucleotides –169 to –135, –121 to –72, and –9 to 18 of sgRNA; and deoxynucleotides –6 to 1 and –18 to 6 of the NTS and TS DNA strands, respectively.

DNA substrates for magnetic tweezers

The DNA substrate for magnetic tweezers was produced similarly as previously published [31, 41]. First, a single SpCas12f1 target site was cloned into the SmaI site of plasmid pUC19 using a DNA duplex, which was hybridized from the oligonucleotides SH_MT_SpCas12f1_NTS and SH_MT_SpCas12f1_TS ([Supplementary Table S4](#)). Next, the plasmid was used to produce a 2.04 kbp PCR fragment using primers CRISPR-SpeI-For and CRISPR-NotI-Rev, by which a SpeI or a NotI restriction site was introduced at either fragment end. The product was then purified using spin columns (Macherey-Nagel, PCR kit). The fragment was then digested with SpeI-HF and NotI-HF. Anchor fragments to flow cell and magnetic bead were synthesized by PCR using Taq polymerase, plasmid pBlue-script II SK, and primers pBlueFor and pBlueRev2 including biotinylated or digoxigenin-modified deoxyuridine triphosphate (dUTP) in the reaction mixtures. After purification, the modified fragments were cleaved at their centrally located SpeI (for biotin) and NotI (for digoxigenin) sites using the corresponding enzymes. After purification, the three cleaved DNA fragments were ligated overnight at 16°C using T4 ligase. The

final construct was separated from side products using a 0.7% agarose gel (7 V/cm for 150 min), and the corresponding bands were cut out by avoiding exposure to UV radiation or ethidium bromide [42].

Flow cells for magnetic tweezers experiments

Flow cells for the magnetic tweezers experiments were assembled from two cover slides and a cut parafilm spacer to form the final fluidic chamber. Before cell assembly, the bottom cover slide was spin-coated with polystyrene. Three micrometer carboxylated polystyrene reference beads (Invitrogen) in 1 M NaCl were added and incubated for 1 h to ensure adherence to the bottom. The fluidic cell was incubated with 50 µg/ml anti-digoxigenin in phosphate-buffered saline for at least 1 h, and afterwards for 1 h with 10 mg/ml bovine serum albumin to prevent nonspecific binding to the bottom of the cell.

Magnetic tweezers experiments

Magnetic tweezers measurements were performed in a custom-built microscope [28] equipped with a 100× objective (NA of 1.25, Olympus). The flow cell was mounted onto a piezo-actuated nanopositioning stage (P-517.3CD, PI; Physik Instrumente, Karlsruhe, Germany). The sample was illuminated using an LED emitting at 625 nm (CR5111AWY; Roithner Lasertechnik). Imaging was carried out with a TM1067 CL CCD camera (Pulnix) at 120 Hz. To generate the magnetic field gradient for the magnetic tweezers experiments, two permanent NeFeB magnets (W-05-N50-G; Supermagnete) were mounted onto a motorized stage above the sample. DNA molecules were bound to 1 µm magnetic beads (MyOne; Invitrogen) and flushed into the flow cell of the setup, allowing the tethering of the digoxigenin-modified DNA end. After removing unbound beads by flushing, unnicked DNA-tethered beads were selected. Using fast imaging of the beads with the camera and three-dimensional GPU-assisted real-time particle tracking [28], the length of the individual DNA molecules was determined from the position of the DNA-tethered beads with respect to an adsorbed reference bead. Forces were calibrated using bead fluctuations along the “long-pendulum” direction and power-spectral-density analysis [43]. During experiments, the force on the DNA molecules was set to 0.3–0.4 pN as indicated. Supercoiling of DNA was achieved by turning the magnets. Time trajectories of the DNA length were recorded at 120 Hz and smoothed with a sliding average to 6 Hz for analysis. For measurements with SpCas12f, reconstituted SpCas12f-sgRNA was added in measurement buffer (10 mM Tris-HCl, pH 7.5, 200 mM NaCl, 10 mM MgCl₂, 1 mg/ml bovine serum albumine (BSA)) at a concentration of 33 nM. After protein addition, DNA length changes were monitored in real-time.

Denaturing polyacrylamide gel electrophoresis experiments of dsDNA cleavage

For ensemble cleavage experiments, a TS and a corresponding NTS of 62 nt length (see [Supplementary Table S4](#) as well as the cartoons next to the gel images for the particular fluorescent and phosphorothioate modifications) were hybridized at a 1:1 ratio in annealing buffer (10 mM Tris-HCl, pH 8, 2 M NaCl, 50 mM MgCl₂, 1 mM EDTA) using a temperature ramp that decreased from 95°C to 10°C within 1 h. One hundred nanomolar preformed SpCas12f-sgRNA complex was then mixed with 5 nM target DNA for 30 min at 46°C to allow

prebinding in assembly buffer containing 10 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM EDTA, and 1 mM DTT. Subsequently, the reaction was initiated by the addition of 1 mM MgCl₂. The reaction was stopped at different time points by the addition of 47.5% formamide and heating to 95°C for 15 min. Products were then put on ice until further use. Products were analyzed using an 8.5 M urea polyacrylamide gel with an acrylamide/bisacrylamide ratio of 29:1, which was run at 55°C. Gel imaging and product quantification were done using a ChemiDoc MP imaging system. To obtain cleavage rates, we assumed a simple first-order reaction scheme with TS and NTS cleavage occurring independently such that both are described by single exponential decays. Weighted fitting of the exponential decay was performed using a custom Python (version 3.8.8) script utilizing the “curve_fit” of the SciPy package [44]. Polyacrylamide gel electrophoresis (PAGE) experiments were done in triplicates with errors of the cleavage rates and efficiencies corresponding to SD.

HEK293T cell culturing, transfection, and genome editing assay

HEK293T cells (ATCC CRL-3216) were cultivated at 37°C and 5% CO₂ in Dulbecco's Modified Eagle's Medium (Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco), penicillin (100 U/ml), and streptomycin (100 µg/ml). A day before transfection, cells were seeded in a 96-well plate at a density of 3.0×10^4 cells per well. After 24 h, cells were co-transfected with 150 ng of SpCas12f1 protein and 50 ng of sgRNA expression plasmids (Supplementary Table S2) using 0.5 µl of TurboFect transfection reagent (Thermo Fisher Scientific). The cells were grown for 72 h at 37°C and 5% CO₂ then washed with PBS (Thermo Fisher Scientific) and lysed using 30 µl QuickExtract DNA extraction solution (LGC Biosearch Technologies). Next, two rounds of PCR were performed. First, to amplify DNA target sites, 1 µl of cell lysate was used in 12-cycle PCR (primer sequences are provided in Supplementary Table S4) in final a volume of 20 µl, using Hot Start Phusion Polymerase (Thermo Fisher Scientific). Next, 1 µl of first-round PCR reaction mixture was used as a template for second round PCR consisting of 18 cycles to index and add P5 and P7 adapters required for Illumina sequencing. Amplicons were pooled and purified from agarose gel using GeneJET Gel extraction kit (Thermo Fisher Scientific). DNA library was quantified with Qubit 4 Fluorometer (Thermo Fisher Scientific) and pair-end sequenced (2×75 bp) using MiniSeq High Output Reagent Kit, 150 cycles (Illumina) on a MiniSeq System (Illumina). Genome editing efficiency was evaluated using CRISPResso2 [45]. Target sequences can be found in Supplementary Table S6.

Statistics and reproducibility

Relevant statistical parameters (sample sizes, resolutions, statistical errors) are given at the relevant spots in the text.

Results

Overall structure of SpCas12f1 nuclease

To understand how the SpCas12f1 nuclease achieves PAM recognition and RNA-guided target DNA cleavage, we determined the structure of WT SpCas12f1:RNA:DNA ternary complex using cryo-EM. The complex was assembled using a 210 nt sgRNA, formed by fusing the 50 nt CRISPR RNA

(crRNA, comprising a 32 nt CRISPR-repeat part and an 18 nt spacer-derived guide) and the 156 nt trans-activating crRNA (tracrRNA) and via the 5'-GAAA tetranucleotide linker [21], as well as target DNA mimicking a cleavage product, which includes a 14 bp duplex region with a 5'-TTC-3' PAM sequence, an 18 nt TS, and a 5 nt NTS (Fig. 1A). In the 2.8 Å cryo-EM map, we resolved two protein monomers (412/497 and 392/497 aa modeled), 113 nt sgRNA, 7 bp of the PAM-containing DNA duplex, 1 nt of the NTS, and 18 nt of the TS annealed to the guide-RNA (Fig. 1B).

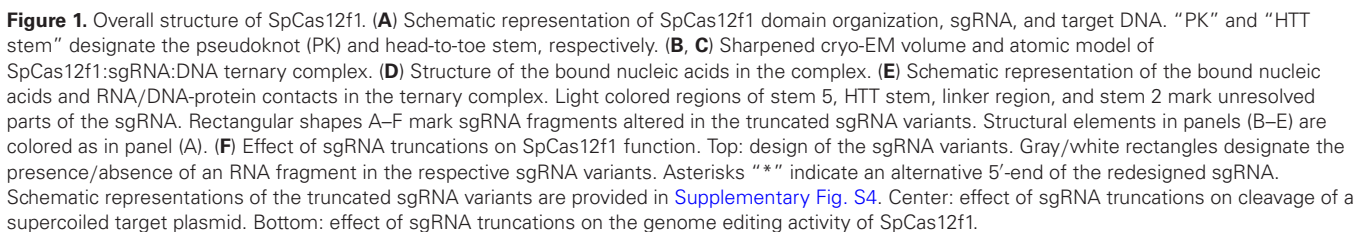
The SpCas12f1 protein monomer adopts a bi-lobed structure, characteristic of other Cas12/TnpB family proteins, comprising an N-terminal recognition (Rec) and a C-terminal nuclease (Nuc) lobe (Fig. 1A). The Rec lobe includes wedge (WED) and REC domains, while the Nuc lobe contains a RuvC domain with a two zinc finger (ZnF) domain insertion (partially resolved in both subunits) and the third ZnF domain fused to the C-terminus (unresolved in both subunits, Supplementary Fig. S1).

In comparison to ~400 aa TnpB, which comprises the minimal functional core of the Cas12 nuclease family [9], the 497 aa SpCas12f1 has several minor insertions in the REC, WED, and RuvC domains, along with two extra ~30 aa ZnF domains (Supplementary Fig. S1).

The overall architecture of the SpCas12f1 ternary complex closely resembles the structures of AsCas12f1 [15, 18], Un-Cas12f1 [16, 17], and CnCas12f1 [19], suggesting a conserved DNA recognition and cleavage mechanism across the Cas12f1 protein family (Supplementary Fig. S2A). The two SpCas12f1 subunits assemble mainly through their REC domains (708 Å² out of 943 Å², the total interaction surface), forming a symmetric dimerization surface from each subunit involving helix α1 (aa 27–56) and the loop (aa 108–121) connecting α3 helix and β2 strand (Supplementary Fig. S1). The REC.1–REC.2 interaction surface comprises mostly polar residues and includes multiple hydrogen (H) bonds (Supplementary Fig. S3A). The relatively small and nonhydrophobic contact area explains why SpCas12f1 protein in the absence of sgRNA is a monomer in solution [21].

The “primary” SpCas12f1 subunit (SpCas12f1.1) makes most of the interactions with both sgRNA (~3214 Å² contact area) and DNA (~1577 Å² contact area), including all contacts with the PAM region of the DNA target and the PAM-proximal end of TS-DNA:sgRNA heteroduplex (Fig. 1D). The “secondary” subunit SpCas12f1.2 mainly interacts with the PAM-distal region of the DNA:RNA heteroduplex (~1204 Å² contact area with sgRNA and ~474 Å² contact area with DNA).

The overall symmetry of the SpCas12f1 protein homodimer is disrupted by the Nuc-lobe RuvC domains, which in the two subunits adapt different orientations relative to their Reclobes (Supplementary Fig. S3B). The RuvC domain of the primary subunit (RuvC.1) inserts its helical region (aa 250–300) between the DNA:RNA heteroduplex, stems 1–2, and the pseudoknot (PK), its distal region (aa 270–273) contacting the RuvC domain of the secondary subunit (RuvC.2) at the structure's core. In contrast, the RuvC domain of the secondary subunit contacts sgRNA from the exterior side of the complex, with its distal part (unresolved in the structure) pointing away from the complex center (Supplementary Fig. S3B). The SpCas12f1 catalytic centers are formed by the RuvC domain residues D228, E327, and D434 (Fig. 1A and B). The lid motif



(Supplementary Fig. S1), which plays a key role in regulating the activity of the RuvC catalytic centers (SpCas12f1 residues 330–344) [46], is poorly resolved in both subunits, likely due to high mobility, suggesting that both catalytic centers are in the “open” conformation.

RNA structure

Similar to other characterized Cas12f1 proteins, SpCas12f1 is guided by a long ~200 nt sgRNA, which adopts a complex 3D structure anchored to the WED and RuvC domains of the primary subunit, and RuvC domain of the secondary subunit (Fig. 1C and D). The core part of the sgRNA structure consists of stem 1 and the adjacent PK, while stem 2, stem 3, and stem 4 side elements extend in different directions from the core (Fig. 1D). Stem 5 is formed by annealing of repeat and anti-repeat regions of sgRNA (parts of crRNA and tracrRNA, respectively).

Moreover, the relative proximity of stem 5 and stem 1 in the structure and observed complementarity in unresolved sgRNA parts between sequences in stem 5 and 5'-terminus of tracrRNA imply formation of an additional “head-to-toe” stem that was predicted previously [24] (Fig. 1D). A comparison of SpCas12f1 sgRNA structure to those of UnCas12f1, AsCas12f1, and CnCas12f1 revealed conservation of the core part of the structure, including stem 1, PK, stem 2, and the 3'-terminal guide region forming the heteroduplex with TS DNA (Supplementary Fig. S2B). These elements also align with the structural core of rRNA of TnpB, the evolutionary predecessor of Cas12f1 (Supplementary Fig. S2B). However, the positions of other structural elements are highly divergent, resulting in unique sets of interactions with the respective proteins.

To investigate whether unresolved and potentially flexible sgRNA regions are essential for SpCas12f1 complex activity, we engineered a set of sgRNA variants containing truncations (or combinations of several truncations) in the predicted “head-to-toe” stem, stem 5, and stem 2 elements (v1.1–1.11) (Fig. 1E and F top panel and Supplementary Fig. S4A). Additionally, the “head-to-toe” stem was replaced with an equivalent 5'-terminal stem (except v1.4) (Fig. 1F top panel and Supplementary Fig. S4A). The SpCas12f1 complexes were assembled with engineered sgRNAs and subjected to plasmid DNA cleavage assays *in vitro*. The results revealed that all tested variants, despite substantial sgRNA truncations or redesign of the “head-to-toe” stem, retained significant plasmid DNA cleavage activity, suggesting that unresolved elements of sgRNA in the SpCas12f1 complex structure are not critical for the nuclease function (Fig. 1F middle panel and Supplementary Fig. S4B). However, the observed DNA cleavage efficiency was lower for most of the SpCas12f1 complexes assembled with engineered sgRNAs except the stem 2 truncation (v1.4), indicating that “head-to-toe” and stem 5 regions unresolved in the structure might enhance complex formation and/or activity *in vitro*.

The guide RNA truncations are a commonly used strategy to improve the performance of CRISPR-Cas12f1 nucleases in genome editing applications [22, 24]; therefore we tested the engineered sgRNAs in HEK293T cell genome editing assays. Comparison of indel generation frequencies that indicate genomic DNA cleavage by the SpCas12f1 complex revealed improved indel formation of up to two-fold for sgRNA variants containing truncations of the “head-to-toe” stem with redesigned 5'-termini and stem 5 (Fig. 1F bottom

panel and Supplementary Fig. S4C). These results are in agreement with the previous study where truncations of sequences corresponding to predicted “head-to-toe” stem and stem 5 elements significantly improved the activity of SpCas12f1 [24]. Altogether, our findings provide a set of constraints for further SpCas12f1 sgRNA optimization.

DNA–RNA heteroduplex formation and PAM recognition

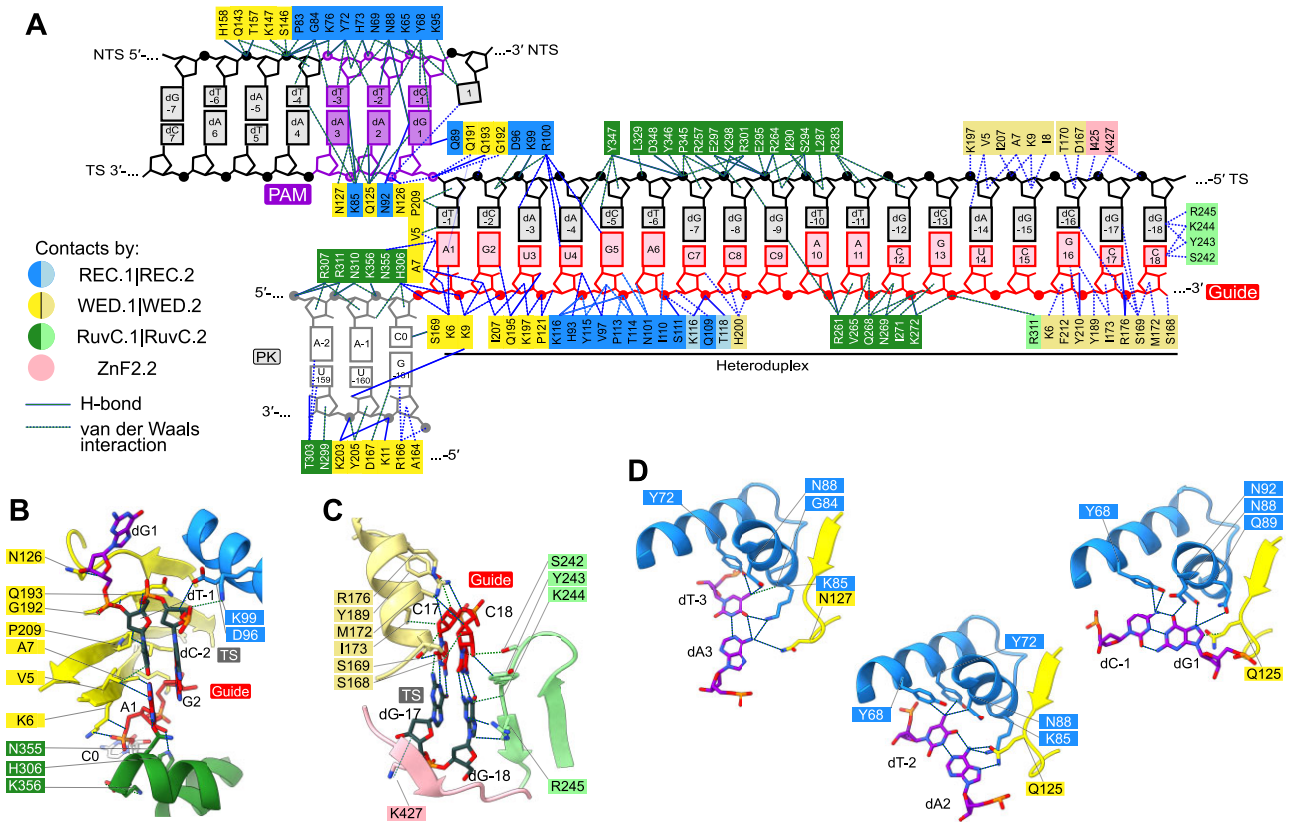
Separation of the TS and NTS DNA strands immediately downstream of the PAM is facilitated by the REC domain residues Y68 and Q89, which insert between the strands and stabilize the last unseparated PAM DNA base pair (Fig. 2A and B). SpCas12f1 forms multiple interactions with the phosphodiester backbone of the DNA–RNA heteroduplex, similar to other Cas12f1 enzymes and the smaller TnpB nuclease (Fig. 2A). The key interaction regions include (i) the PAM-proximal end of the heteroduplex, which is stabilized by the WED domain residues V5, A7, and P209 stacking against the first heteroduplex base pair; (ii) the PAM-proximal region, which is enclosed between the REC and WED domains of the primary subunit; (iii) the 7–9 bp region contacted by the REC/RuvC domains of the primary and the REC/WED domains of the secondary subunits; (iv) the 10–13 bp region contacted by the RuvC domain of the primary subunit; and (v) the PAM-distal region enclosed by the REC, WED, RuvC, and ZnF2 domains of the secondary subunit (Fig. 2A). The maximum length of the heteroduplex (18 bp) is defined by the end-stacking of Sp-Cas12f1.2 RuvC domain residues 242–245 against the 18th DNA:RNA base pair (Fig. 2A and C).

The 5'-TTC-3' PAM-containing region of the target dsDNA is accommodated between the REC and WED domains of the primary subunit (Figs 1B and 2A). All six bases of the 5'-TTC-3' trinucleotide are read out through base-specific H-bond and van der Waals (vdW) contacts (Fig. 2A and D).

The PAM recognition mechanism is partially conserved across the characterized Cas12f1 proteins. For instance, the two consecutive thymine bases (5'-NTTN-3') at PAM positions –3 and –2 are read-out by the SpCas12f1 Y72 residue, equivalent to Y76 in AsCas12f1 and Y146 in Un-Cas12f1, which interacts with the thymine 5-methyl groups (Supplementary Fig. S3C).

Characterization of R-loop formation and stability by SpCas12f1 for different spacer lengths

The obtained atomic structure suggests that SpCas12f1 accommodates and stabilizes an R-loop of 18 nt length similarly to AsCas12f1 [18] but in contrast to UnCas12f1, which accommodates a 20 bp R-loop [17, 16]. To probe whether other R-loop lengths are possible, we employed a previously established magnetic tweezers assay based on DNA supercoiling [31, 32]. For the type I Cascade and the type II Cas9 complexes, it is well-established that R-loop formation is initiated at the PAM at which base pairing between the protospacer and the sgRNA is primed. Subsequently, the R-loop expands in a 1 bp random-walk (forward-and-backward) fashion [47, 31, 48] (Fig. 3A). Once it extends over the entire sgRNA spacer, the R-loop is stabilized by a conformational change of the complex, which subsequently initiates the DNA cleavage. During R-loop formation the DNA becomes unwound. This can be observed in magnetic tweezers experiments by negatively supercoiling a DNA molecule at low force (0.3 pN)



such that its length becomes reduced due to writhe formation (Fig. 3B). DNA unwinding during R-loop formation absorbs a corresponding amount of the supercoiling, which is seen as a sudden increase in length (Fig. 3B) being proportional to the R-loop length [47]. We first investigated R-loop formation by the nuclease-dead variant of SpCas12f1 using an sgRNA with an 18 nt spacer. As expected, R-loop formation at negative supercoiling by the effector complex was seen as an upwards step of the DNA extension (see bottom panel of Fig. 3B). This shifted the characteristic hat-like curve for DNA twisting towards negative supercoiling (Fig. 3C, top right, blue versus black curve). When turning from negative towards positive supercoiling the shift of the supercoiling persisted, indicating that the R-loop became stabilized since it resisted the positive torsion [31, 32]. We next repeated the supercoiling assay for different spacer lengths (Fig. 3C), ranging from 14 to 24 nt. We observed efficient R-loop formation at negative supercoiling (100% of all attempts) for all spacers. When challenging the R-loops with positive supercoiling, the shift of the supercoiling predominantly persisted only for the 18 nt spacer. For other spacer lengths, no shift was observed in most of the attempts, indicating that the formed R-loops were often not stable. We further analyzed the shift of the supercoiling curve (Fig. 3D). For spacers of 14–18 nt, the DNA untwisting increased from ~14 to ~18 nt, while it did not increase further for the longer spacers. Calculating the fraction of stable R-loops (Fig. 3E) revealed that ~90% of the R-loops were stable for the 18 nt spacer, while no or barely any stable R-loops were formed for 14 and 24 nt spacers, respectively.

Altogether, this provides evidence that SpCas12f1 can accommodate R-loops of up to 18 nt length, which also get most stabilized. This agrees with the atomic structure and indicates that the stacking of RuvC.2 residues 242–245 to the most PAM-distal base pair significantly contributes to R-loop stability as well as DNA cleavage, which revealed a maximum efficiency for 18 nt R-loops [21]. Evaluating the shift of the supercoiling curve of stable R-loops for positive supercoiling revealed shorter R-loops compared to the negative supercoiling, e.g. ~12 bp for the 16 and 20 nt spacers. This indicates that contacts to the R-loop at shorter distances from the PAM are also important for stabilization.

SpCas12f1 binds target dsDNA in a multiple step process

We next inspected more closely the R-loop formation trajectories for nuclease-dead SpCas12f1 with the 18 nt spacer. Following the induction of negative supercoiling (Fig. 4A), the protein complex typically required several minutes to fully unwind the DNA (Fig. 4B). Notably, during the unwinding process, small excursions to a position corresponding to ~8 base pairs were frequently observed, with a short lifetime of a few seconds (Fig. 4B). These transient excursions probably represent intermediate states in the unwinding pathway, suggesting dynamic and reversible interactions within the protein–DNA complex. This agrees with the extensive contacts that the RuvC.1 subunit makes with the TS in the region of –7 to –10 nt (Fig. 2A). Comparing the width of the distributions of the different untwisting states, we observed that the intermedi-

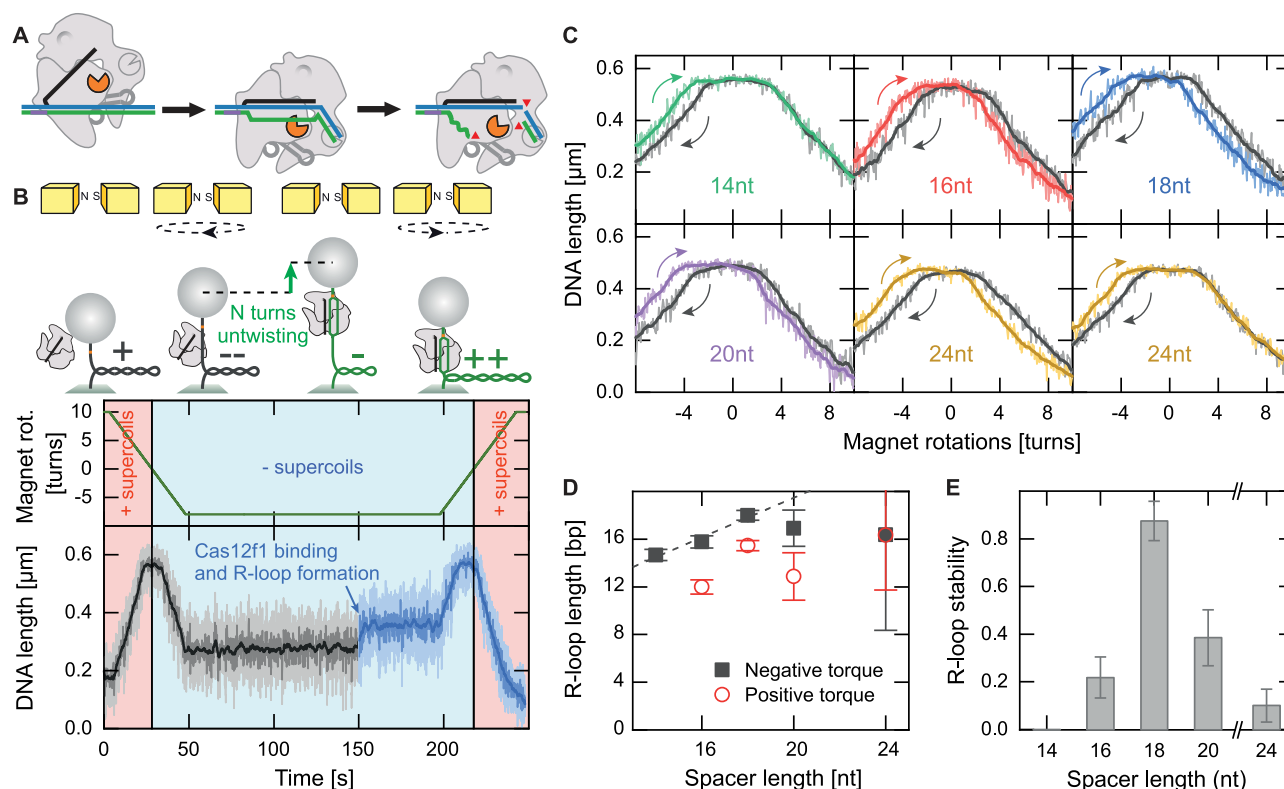


Figure 3. Magnetic tweezers experiment to quantify the extent of R-loop formation by dCas12f1 on supercoiled DNA. **(A)** Scheme of the SpCas12f1-sgRNA effector complex on a dsDNA target in the PAM-bound state (left), after R-loop formation (center), and upon DNA cleavage (right). Shown are sgRNA, TS, NTS and the PAM. The nuclease domains are depicted as small Pac-Man symbols in orange (active domain) and gray (inactive domain). The known DNA cutting sites are indicated by small arrows. **(B)** Representative time trace of a typical R-loop formation experiment by dCas12f1 (at 0.3 pN) with cartoons above illustrating the supercoiling state of the DNA and the R-loop formation in the magnetic tweezers. The experiment starts with positively supercoiled DNA at reduced length due to writhing to prevent R-loop formation. To support R-loop formation, negative turns are applied during which the DNA length initially elongates when reaching zero supercoiling and subsequently reduces when negative supercoiling is obtained. R-loop formation is seen as a sudden increase in DNA length since it absorbs part of the negative supercoiling. Subsequently, positive supercoiling is applied. **(C)** Representative DNA supercoiling curves (at 0.3 pN) measured for Cas12f1 with different spacer lengths of the sgRNA. After R-loop formation, the supercoiling curve (colored curve) is shifted towards negative turns compared to the initial curve (black) due to the absorbed supercoils. Unstable R-loops exhibit a shift only in the negative part of the curve, while for stable R-loops, both sides shift. For all spacer lengths, we see stable and unstable R-loops, as illustrated for the 24 nt spacer, for which both cases are displayed. **(D)** Mean R-loop length as function of the spacer length obtained from the shift of the negative (squares) and the positive (circles) site of the supercoiling curves. Error bars correspond to S.E.M. We analyzed 19, 23, 16, 17, and 14 events for spacer lengths of 14, 16, 18, 20, and 24 nt, respectively. **(E)** Fraction of stable R-loops as a function of the spacer length. Errors represent the standard deviation of a binomial distribution $\sigma = \sqrt{p \cdot (1 - p) / N}$.

ate state exhibited a wider distribution (3.7 nt SD) compared to the unbound and fully unwound states (3.2 and 3.3 nt SD), respectively (see Fig. 4B, right). This widening supports a certain variability of the R-loop length in the intermediate state in agreement with the rather extended region of the RuvC domain interaction with the TS backbone.

dsDNA cleavage is preceded by rapid cleavage of one of the strands

We next explored DNA cleavage in our supercoiling experiments by employing WT Cas12f1. After initiation of the experiment by negatively supercoiling the DNA molecule, first a small DNA length increase was observed, indicative of R-loop formation (Fig. 5A and B). The full R-loop state was only short-lived and terminated by a sudden large length increase to the extension of relaxed DNA, which indicates DNA nicking upon which all supercoils are rapidly released [49] (Fig. 5B). For most molecules, nicking was followed by cleavage of the second DNA strand after an extended period seen as disappearance of the magnetic bead (Fig. 5B). Nicking occurred

7.4 ± 0.1 s after R-loop formation, while the second strand cleavage required additional 213 ± 9 s (Fig. 5C). These data indicate that the cleavage of NTS and TS occurs on significantly different time scales.

R-loop formation and the cleavage states may induce strong structural distortions to the DNA, including an overall bend that may change the length of stretched DNA molecules (see “Discussion” section). This in turn would make the individual reaction steps dependent. We thus probed DNA cleavage at negative supercoiling but an elevated force of 3 pN, which yielded a greatly reduced cleavage efficiency and an increased time of ~ 4000 s for the full cleavage reaction (Supplementary Fig. S5A). Since at these forces writhing formation is inhibited, we could not distinguish whether R-loop formation or one of the DNA cleavage steps was predominantly affected by force. To reveal this, we reduced the force for the uncleaved molecules at the end of the cleavage experiment to 0.3 pN. For $38 \pm 17\%$ of the molecules (8 in total), R-loop formation followed by first and second strand cleavage was observed (Supplementary Fig. S5B), suggesting that R-loop formation is highly force sensitive, presumably

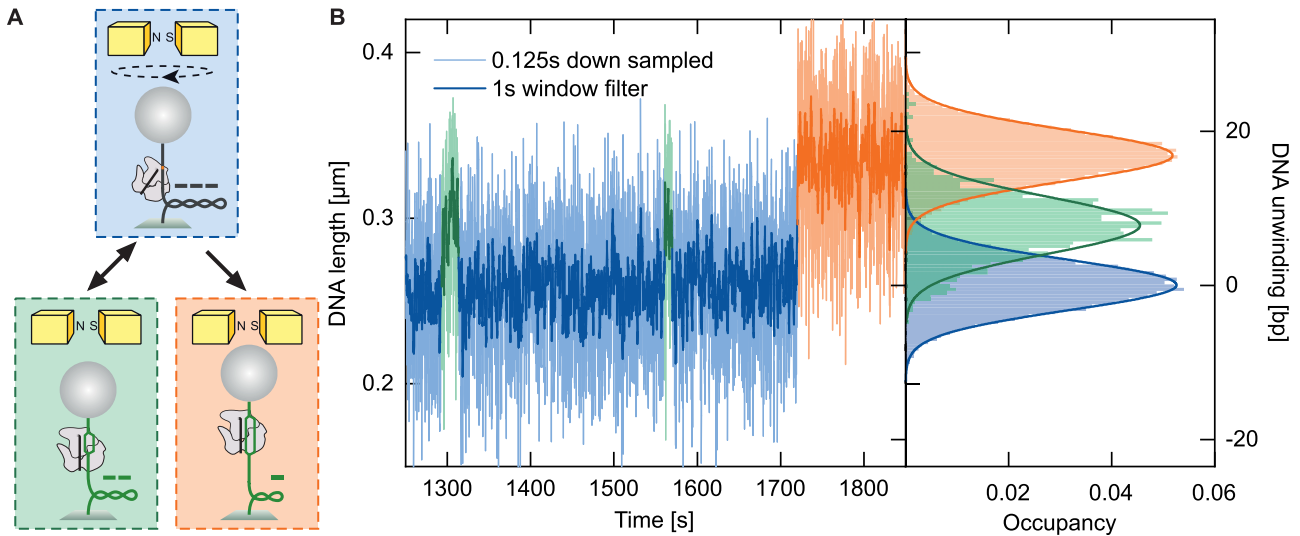


Figure 4. Detection of intermediate R-loop states of dCas12f1. **(A)** Cartoon of the supercoiling experiment depicting the unbound state (top), a reversible intermediate R-loop state (left), and the full R-loop state (right). **(B)** Time trace of the DNA length revealing reversible intermediate R-loops prior to full R-loop formation using dCas12f1 with an 18 nt spacer measured at 0.3 pN and -7 turns. The recorded DNA length was smoothed with a sliding time window of either 0.125 s (light colors) or 1 s Hz (dark colors). Histograms on the right were obtained from correspondingly colored sections of the time trace. Solid lines represent Gaussian fits to the different states. Data were replicated in multiple traces across four different molecules.

due to the associated DNA bending. To probe whether DNA cleavage was additionally affected by force, we carried out R-loop formation experiments at 0.3 pN and -7 turns and increased the force only after R-loop formation and DNA nicking were observed. In this case, we found the mean second strand cleavage time to be moderately increased to ~ 600 s (Supplementary Fig. S5C). This suggests that DNA bending or looping upon second strand cleavage is less pronounced compared to R-loop formation.

Previously, it had been shown that second strand cleavage by the monomeric Cas12a is accompanied by clamping of the PAM distal DNA end [12]. During clamping, the DNA became again torsionally constrained, which can be attributed to tight protein–DNA contacts on both the PAM-proximal and the PAM-distal DNA. We therefore explored whether Cas12f1 also uses clamping during DNA cleavage. To this end, we started to supercoil the DNA as soon as it became nicked (Fig. 5D). In case of clamping, a strong DNA-length reduction would be expected due to writhe formation. Remarkably, following nicking, part of the DNA molecules ($46 \pm 14\%$, for a total of 13 molecules) became supercoilable again, indicative of clamping, albeit for shorter periods (Fig. 5E). Thus, Cas12f1 can bind tightly to the PAM-distal DNA. Notably, brief excursions into an intermediate state were also observed during the supercoiling phase (see binding states measurement in Fig. 5E). When inspecting DNA cleavage occurring during these experiments, we observed that cleavage did not occur during clamping but rather in the torsionally relaxed state of the DNA. This was independent of whether positive or negative supercoiling was applied (Supplementary Fig. S6). This indicates that the clamped state is not the state in which the second strand gets cleaved in contrast to Cas12a.

SpCas12f1 cleaves NTS and TS in a sequential manner

The single-molecule cleavage experiments revealed different time scales for the cleavage of the first and the second strand.

To attribute whether the NTS at either of the two sites or the TS gets cleaved first, we determined the dsDNA cleavage kinetics using fluorescently labeled DNA strands and analyzed the reaction products with denaturing PAGE. Labeling the TS at the PAM proximal end and the NTS at the PAM-distal end revealed the expected TS cut at ~ 22 nt from the PAM and an NTS cut at site 2 at ~ 24 nt from the PAM (Fig. 6A). When labeling the NTS at both the PAM-proximal and the PAM-distal end, one obtains fragments corresponding to cuts at site 1 at ~ 13 nt from the PAM and at site 2 at ~ 24 nt from the PAM, respectively (Fig. 6B). Interestingly, for either label, one obtains only the fragment of the closest cut position. Longer intermediates from cutting at the more distal site are absent. This indicates that cutting of both NTS sites occurs for our time resolution practically instantaneously. The mean cleavage times in these experiments were 3.3 ± 0.2 min for NTS and 7.3 ± 0.3 min for TS cutting (Supplementary Fig. S7). Also, in the subsequent experiments, TS cutting appeared to be delayed compared to NTS cutting, suggesting that the fast nicking (~ 7 s) in the single-molecule experiments was due to NTS cutting, while the slow second strand cleavage (~ 3 min) was due to TS cutting. The NTS cutting appears to be much slower in the ensemble experiments, since one does not resolve the preceding R-loop formation step, which is again in the minute range (see earlier).

To further support a sequential NTS and TS cleavage mechanism, we investigated how blockage of the individual cut sites affects the DNA cleavage using a phosphorothioate-modified DNA backbone at these positions [11] (see Fig. 6C for the different substrates). Blocking NTS site 1, within error, did not affect the cleavage efficiency nor the cleavage position at NTS site 2 nor the TS (Supplementary Fig. S8A). Using a PAM-proximal NTS label surprisingly revealed that site 1 cleavage was not blocked but rather shifted to 17 nt from the PAM, i.e. to the end of the blocked region (Supplementary Fig. S8B). When further expanding the block up to 21 nt from the PAM, the NTS cleavage position shifted accordingly, while TS cleavage was unaf-

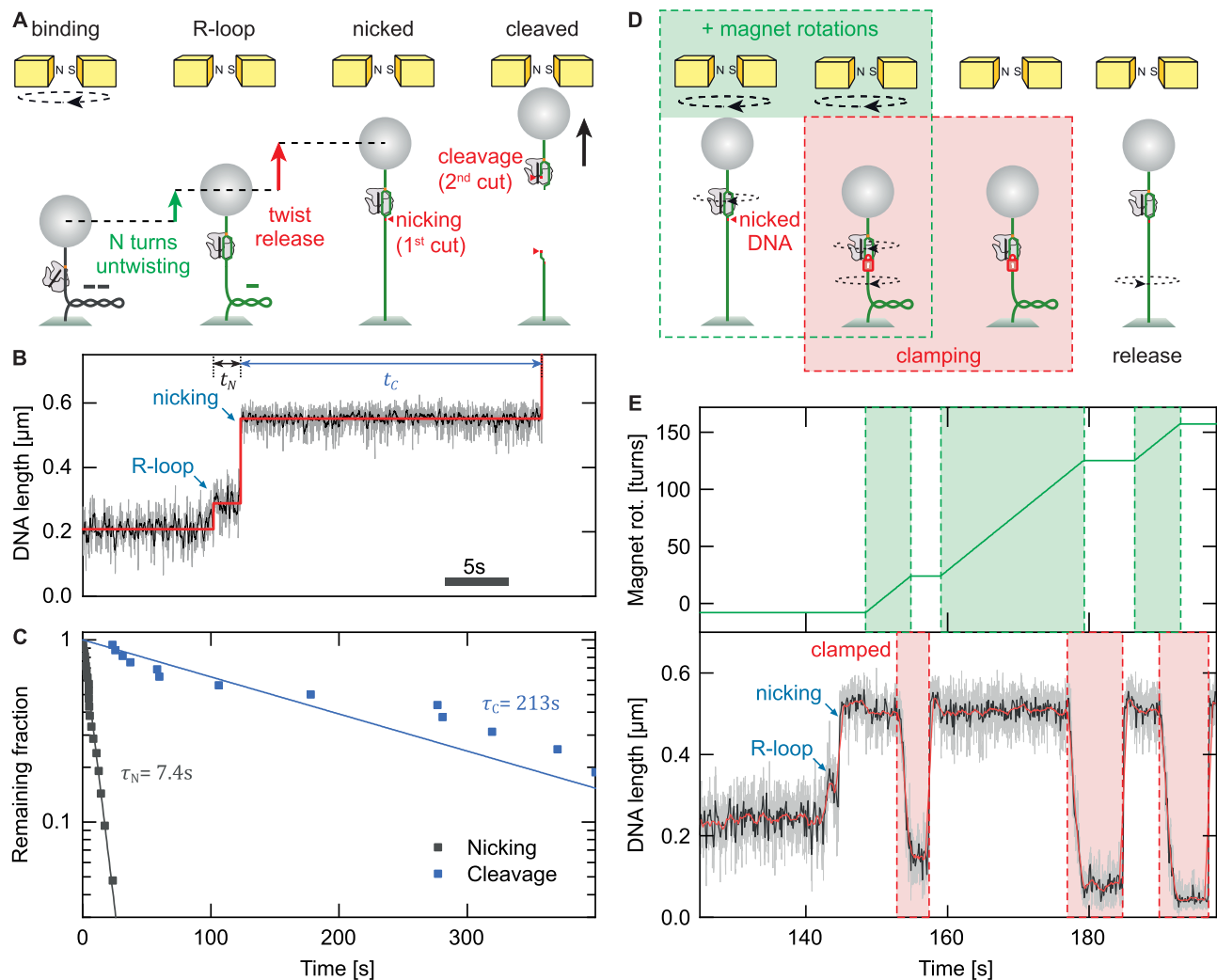


Figure 5. Following R-loop formation as well as first and second strand cleavage in single-molecule magnetic tweezers experiments. **(A)** Cartoon of the different successive states and the associated DNA length increases starting with negatively supercoiled DNA in absence of an R-loop, which is followed by R-loop formation, complete release of twist caused by DNA nicking (first strand cleavage), and subsequent bead loss due to second strand cleavage by SpCas12f1. **(B)** Representative time trace of an R-loop formation-DNA cleavage experiment. Nicking occurs rather quickly after R-loop formation with time t_N , while second strand cleavage occurs considerably slower after time t_C . **(C)** Survival plot of unnicked (black squares) and nicked (blue squares) molecules as function of time providing single exponential decays for the nicking and cleavage. Mean values for t_N and t_C given in the plot were obtained from exponential fits to the data. A total of 21 events were analyzed, which are shown as individual points in the plot. **(D)** Testing clamping of the PAM-distal DNA end by Cas12f1 after first-strand cleavage using DNA supercoiling. After DNA nicking by Cas12f1, positive supercoiling is applied, which is released via the nick. In the case of DNA clamping, the DNA length would suddenly start to decrease until the clamping is released. **(E)** Time trace of a clamping experiment with periods of DNA clamping by Cas12f1 shown on a colored background. After DNA nicking, the application of positive supercoiling does initially not change the length. Only if the nick is topologically trapped by the Cas12f1 the extension decreases again due to DNA writhing. When leaving the DNA in the positively supercoiled state, the termination of this clamping is seen as a sudden DNA length increase. Clamping can occur repeatedly. Clamping events were measured on four different DNA molecules, yielding one or more clamping events per molecule.

ected (Supplementary Fig. S8C). We next tested a block of NTS site 2. Instead of blocking, we observed again a shift of the cut site to ~ 17 nt from the PAM coinciding with the start of the block (Supplementary Fig. S8D). Taken together, these data suggest that the NTS can be cut at any position between position 13 and 24 and that multiple and possibly random cuts are made in order to excise this NTS region. This agrees with a rather flexible conformation of the NTS such that different positions can pass the catalytic center. The observed boundaries may therefore represent the steric constraints for the NTS conformations.

To test whether NTS cleavage is a requirement for TS cleavage, we blocked the entire cut region of the NTS from position 8–28. This led to a strong attenuation of both NTS

cleavage and TS cleavage (Supplementary Fig. S9) and the loss of a dedicated cleavage site in the NTS. The NTS phosphorothioate backbone was not fully resistant to cleavage by Cas12f1, as frequently observed [50, 51]. In turn, this can explain the residual TS cleavage. An additional bulge in the blocked region allowing for an increased flexibility of the NTS revealed a strongly attenuated NTS and TS cleavage (Supplementary Fig. S9B). Using a substrate with a nick in the long blocking region of the NTS restored, however, the full TS cleavage efficiency (Supplementary Fig. S9C). Thus, NTS cleavage tightly controls the TS cleavage (Fig. 6D), which provides evidence that TS cleavage requires a cut NTS, at least at some position. Blocking the TS cut site did not affect the NTS cleavage efficiency nor position (Supplementary Fig. S9D).

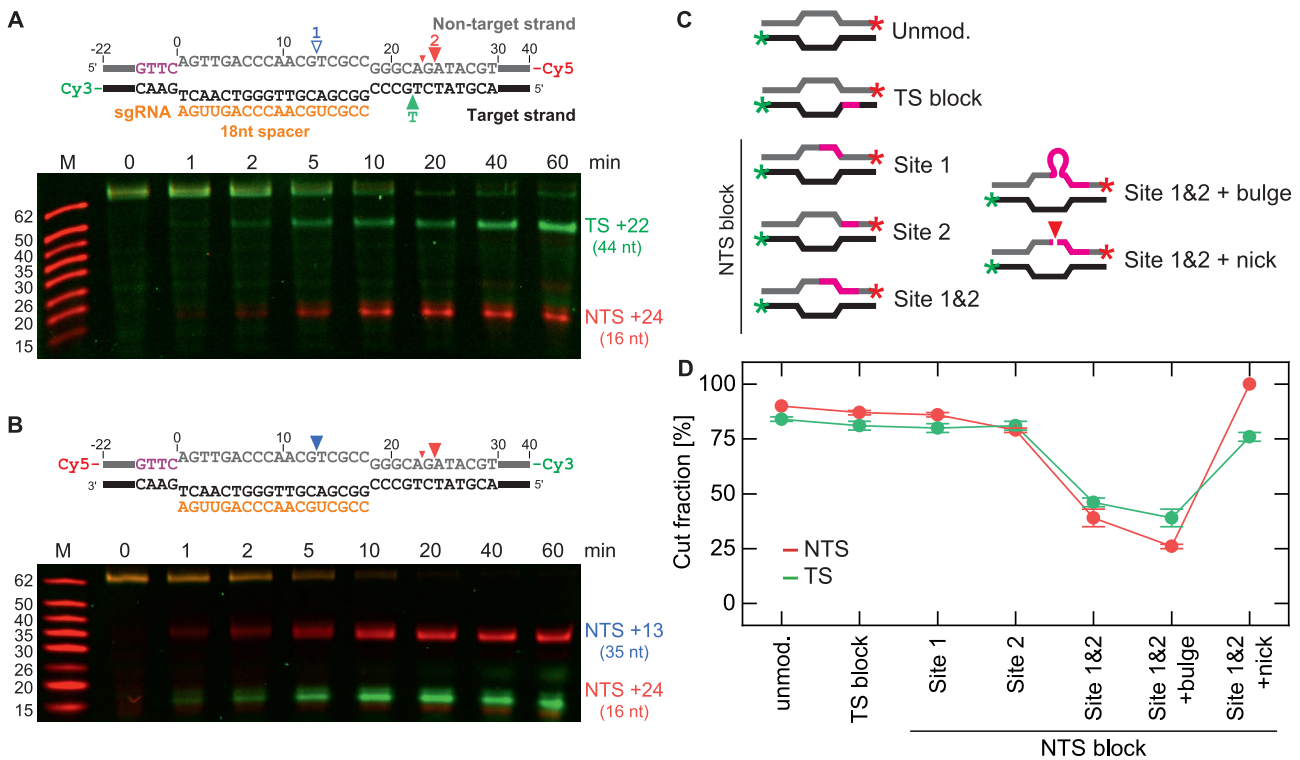


Figure 6. Kinetics of TS and NTS cleavage by SpCas12f1 probed with denaturing PAGE. **(A)** (Top) Schematics of the DNA target with 3'-end fluorescent labels on both strands. Detected cuts at the TS and site 2 of the NTS are shown by solid triangles, respectively. The expected position for an exclusive NTS cleave at site 1 is marked with an open triangle. (Bottom) PAGE gel probing the kinetics of the strand cleavage by SpCas12f1:sgRNA (18 nt spacer). The cleavage times are given on the top of each lane. M is a Cy3-labeled ssDNA size marker. Numbers on the right show the main detected cleavage position on each strand with respect to the PAM as well as, in brackets, the length of the main product. **(B)** PAGE gel analysis of a DNA target with two terminal labels of the NTS (see cartoon). The gel is labeled in analogy to A. Main site 1 and 2 cleavage positions are given on the right as well as triangles in the cartoon. **(C)** Cartoons of the different target DNAs with 3'-end labels of TS and NTS employed in this study. Phosphorothioate modifications are shown in pink. **(D)** Cut fractions of NTS and TS after 60 min for the different target DNAs. NTS and TS cleavage become only significantly inhibited once both NTS cut sites are blocked by phosphorothioate modifications, while a nick can restore TS cleavage despite blockage of both NTS sites. Errors represent the SD of triplicate measurements.

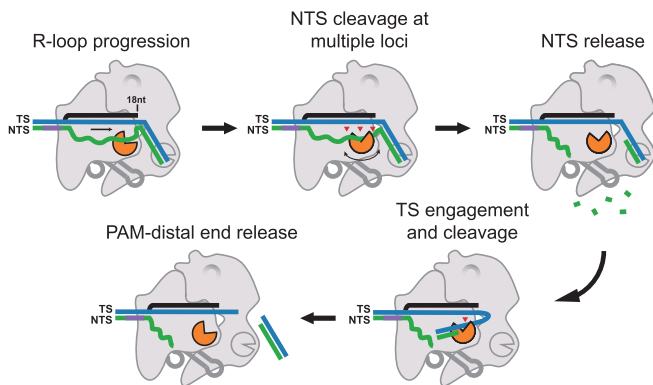


Figure 7. Model for the sequential NTS and TS cleavage by Cas12f1. Upon R-loop formation, the NTS is rather flexibly bound to the protein surface and can easily enter the catalytic center (filled packman symbol), where it gets cleavage at multiple positions, including the PAM distal DNA duplex. This is due to a random repositioning of the flexible NTS as well as an additional unwinding of the PAM-distal DNA duplex when binding to the protein. After release of the cut fragments, the catalytic center becomes accessible for the TS, for which the PAM distal DNA duplex loops back to get cut. In this configuration, the duplex end may get clamped. Subsequently, the PAM-distal DNA duplex dissociates from the protein.

Rather TS cutting was shifted to 27 nt from the PAM, i.e. just behind the block region, and cleavage was highly efficient, demonstrating that TS cutting can occur even further inside the PAM-distal DNA duplex.

We repeated the cleavage experiments with blocks of the NTS site 1 and/or site 2 using spacers of 14 and 16 nt (Supplementary Figs S10 and S11). These experiments supported the results obtained with the 18 nt spacer. Generally, the efficiency of NTS and TS cleavage decreased with decreasing spacer length. As before, TS cleavage was only hampered if the full NTS cleavage region was protected (Supplementary Fig. S12).

Roles of RuvC.1 and RuvC.2 catalytic centers in DNA cleavage

The open conformation of the RuvC catalytic centers in both SpCas12f1 subunits prompted us to examine the potential involvement of SpCas12f1.2 subunit catalytic center in DNA cleavage (Supplementary Fig. S13A). We tested this using a combination of two inactive SpCas12f1 mutants: RuvC^{Mut} (nuclease-dead mutant D228A) and PAM^{Mut} (triple mutant N88A + Q89A + N92A incapable of PAM recognition, Supplementary Fig. S13B). We reasoned that if the RuvC.2 domain of the secondary subunit SpCas12f1.2 is indeed involved

in DNA cleavage, the heterodimer of the inactive RuvC^{Mut} and PAM^{Mut} mutants, where the RuvC^{Mut} acts as the primary subunit capable of PAM recognition and the PAM^{Mut} mutant as the secondary subunit capable of catalysis, would regain DNA cleavage activity (Supplementary Fig. S13B). However, the mixture of RuvC^{Mut} and PAM^{Mut} mutants showed no TS nor NTS DNA strand cleavage activity on DNA substrates containing both the intact and pre-cleaved NTS strands (Supplementary Fig. S13C), suggesting that both DNA strands are instead cleaved by the RuvC domain of the primary SpCas12f1 subunit, as previously demonstrated for another family member UnCas12f1 [16].

Discussion

In this study, we solved the structure of the WT SpCas12f1:RNA:DNA ternary complex and dissected the main steps of its DNA cleavage mechanism. The overall structure of the SpCas12f1 ternary complex was found to be similar to other Cas12f1 effectors complexes such as AsCas12f1 [18], UnCas12f1 [16, 17], and CnCas12f1 [19]. The core part of the sgRNA structure is also conserved between different Cas12f1 family members and rRNA of TnpB, the evolutionary ancestor of Cas12 nucleases. However, the additional structural elements of the sgRNA deviate significantly between different Cas12f1 enzymes, highlighting various complex-specific adaptations. Importantly, some of the engineered SpCas12f1 complexes possessing truncated sgRNAs retained significant cleavage activity *in vitro* and were effective in indel generation *in vivo*, implying significant flexibility in the sgRNA structure that supports site-specific DNA cleavage.

Interestingly, the DNA:RNA heteroduplex observed in the SpCas12f1 structure had a length of 18 bp with the base pairs at either end being stabilized by specific amino-acid residues (Fig. 2B and C; Supplementary Fig. S14). The same heteroduplex length has been found in the structure of AsCas12f, while UnCas12f1 possessed a 20 bp R-loop. DNA cleavage experiments revealed that 18 nt RNA is required for the optimal DNA cleavage by Cas12f1 [21] in agreement with the structure. By directly probing the lengths of formed R-loops using magnetic tweezers, we could show here that the heteroduplex length of 18 nt is not exceeded even when using longer spacers and negative supercoiling (Fig. 3D). Furthermore, the 18 bp spacer supports the highest R-loop stability, suggesting that for shorter spacers the protein contacts to the PAM-distal heteroduplex end are perturbed, while for longer spacers the unpaired RNA bases cause a steric clash with the surrounding protein. During R-loop expansion, an intermediate is observed that coincides with major contacts in the 7–9 bp region of the RuvC domain of the primary subunit to the TS. In contrast to the type I surveillance complex Cascade [48], the free energy landscape of R-loop formation of SpCas12f1 is thus probably not uniformly biased but rather exhibits a pronounced energy minimum in the 7–9 bp region.

A pronounced difference to other Cas12 effectors is that three rather than two cut sites were observed for Cas12f1 effectors, out of which two are on the NTS (at 13 and 24 nt from the PAM for SpCas12f1) and one is on the TS (at 22 nt for SpCas12f1) [23, 20]. A tempting hypothesis based on the accessible catalytic centers observed in the structure on both subunits was that they both carry out DNA cleavage, with the SpCas12f1.1 and SpCas12f1.2 catalytic centers being responsible for the reactions on the NTS and the TS strands,

respectively. However, heterodimers of RuvC^{Mut} and PAM^{Mut} mutants were found to be non-active. This supports the usage of just a single catalytic center by Cas12f1 family proteins, which is located in subunit 1 based on previous reports [16, 17]. When inspecting the DNA cleavage more closely in magnetic tweezers, we observed that the time scale for cleaving the first strand is ~10 s, while it is ~100 s for the second strand. This coincided with a faster NTS and a slower TS cleavage in ensemble cleavage experiments. Further considering that blocking of cleavage throughout both NTS sites using phosphorothioate backbone modifications strongly attenuated TS cleavage and other controls, we could conclude a sequential strand cleavage mechanism, where NTS cutting is a prerequisite for TS cutting. We suggest that the order of the strand cleavage is sterically determined. The NTS is directly guided into the catalytic center of subunit 1 and thus blocks the access of the TS to the catalytic center until it becomes cleaved. It would be highly interesting to resolve in real time the engagement of the TS to the catalytic center and how the presence of the NTS influences this process.

When blocking the cut sites separately with phosphorothioate modifications, we found a remarkable flexibility of the cut positions, which were typically shifted to the closest unmodified base pair. TS cutting could be shifted this way from 22 nt to 27 nt. NTS cutting could be flexibly distributed between 13 and 24 nt (observed e.g. at 17 and 21 nt). We think that this is due to the cutting of the NTS at random positions due to a high flexibility of the NTS, which allows to place it at different positions into the catalytic center (Fig. 7). Presumably, the PAM-distal DNA duplex exits the complex in the cleft formed by the RuvC domain of subunit 1 and the ZnF1/ZnF2 domains of subunit 2, leading to an overall kinking of the DNA upon complex binding and a bulging of the NTS (Fig. 7). Such a kinking is supported by the observed inhibition of complex binding under elevated force (Supplementary Fig. S5A). The kinking may also support cutting of the NTS into the DNA duplex up to 24 nt.

One NTS and the TS cut occur in the PAM-distal DNA duplex end. This process still remains somewhat enigmatic. A very recent study on AsCas12f1, which also suggested a sequential strand cleavage mechanism, found that at reduced Mg²⁺ concentrations, NTS cutting starts rapidly at position 16 on the unwound NTS and then proceeds significantly slower into the duplex end [20]. Based on single-molecule FRET measurements, the authors concluded an exonuclease activity that unwinds and degrades the PAM-distal duplex end. It is probably difficult to clearly distinguish between a unidirectional exonuclease activity and an endonuclease activity occurring close to the NTS end. Using phosphorothioate protected strands, we demonstrated here that the NTS in the DNA duplex does not have to be unwound/degraded until 24 nt to obtain TS cleavage (Supplementary Fig. S8D) and that the TS can become cut at 27 nt despite an NTS cutting of only 24 nt (Supplementary Fig. S9D). This suggests that the RuvC catalytic center can act on duplex DNA to cut the TS. During this process the duplex gets probably transiently unwound [20], as also seen for other ssDNA-binding proteins [52]. The interactions of the catalytic center with the NTS of the DNA duplex may cause the observed clamping events (Fig. 5E) since in this case the PAM distal duplex would become torsionally coupled to Cas12f1. For TS cleavage, we hypothesize that the TS is looping back to allow its correct orientation in the catalytic center (Fig. 7). Such a rather large

conformational repositioning of the DNA end should be associated with a strong force dependence, which is only partially seen (Supplementary Fig. S5C). Potentially, conformational changes of the protein support the TS cleavage.

In summary, our study establishes a detailed and highly consistent link between structure and mechanism of Cas12f1 effector complexes and sheds light on the complex cutting mechanism of the NTS. To improve the overall editing rates of Cas12f1-based genome editing, we think that engineering efforts focused on optimizing the initial NTS engagement would be highly interesting. Furthermore, the observed staggered cleavage pattern may leverage a facilitated targeted insertion of exogenous DNA fragments into specific genomic sites.

Acknowledgements

Author contributions: J.M.M., M.B., T.K., V.S., G.S., and R.S. conceptualized the study; J.M.M., M.B., S.H.H., G.B.S., A.S., T.K., and G.S. performed the investigation; J.M.M., M.B., S.H.H., D.J.K., P.I., G.B.S., A.S., T.K., V.S., G.S., and R.S. analyzed the data; and J.M.M., T.K., G.S., and R.S. wrote the original draft receiving revisions and edits from the other coauthors. All authors read and approved the final manuscript.

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

V.S. is a Chairman of and has financial interest in Caszyme.

Funding

Research Council of Lithuania (LMTLT) [S-MIP-19-32 to G.S.]; “Mission-driven Implementation of Science and Innovation Programmes” [No. 02-002-P-0001], funded by the Economic Revitalization and Resilience Enhancement Plan “New Generation Lithuania” to T.K. European Research Council (GA 724863) and Deutsche Forschungsgemeinschaft (SPP2141, SE 1646/8-1) to R.S. J.M.M. was supported by a postdoctoral fellowship of the Alexander von Humboldt Foundation. Open access publishing was funded by Leipzig University supported by the German Research Foundation within the program Open Access Publication Funding.

Data availability

Source data are provided with the manuscript: <https://doi.org/10.5281/zenodo.15097003>. The EM density and model coordinates have been deposited in the Electron Microscopy Data Bank and Protein Data Bank under the accession codes EMD-52745/PDB:9i8y. Sequencing data have been deposited on the NCBI Sequence Read Archive under BioProject ID PRJNA1242595.

References

- Barrangou R, Fremaux C, Deveau H *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 2007;315:1709–12. <https://doi.org/10.1126/science.1138140>
- Brouns SJJ, Jore MM, Lundgren M *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 2008;321:960–4. <https://doi.org/10.1126/science.1159689>
- Garneau JE, Dupuis MÈ, Villion M *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 2010;468:67–71. <https://doi.org/10.1038/nature09523>
- Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;346:1258096. <https://doi.org/10.1126/science.1258096>
- Gasiunas G, Siksnys V. RNA-dependent DNA endonuclease Cas9 of the CRISPR system: holy grail of genome editing? *Trends Microbiol* 2013;21:562–7. <https://doi.org/10.1016/j.tim.2013.09.001>
- Hsu PD, Scott DA, Weinstein JA *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;31:827–32. <https://doi.org/10.1038/nbt.2647>
- Harrington LB, Burstein D, Chen JS *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* 2018;362:839–42. <https://doi.org/10.1126/science.aav4294>
- Karvelis T, Druteika G, Bigelyte G *et al.* Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* 2021;599:692–6. <https://doi.org/10.1038/s41586-021-04058-1>
- Sasnauskas G, Tamulaitiene G, Druteika G *et al.* TnpB structure reveals minimal functional core of Cas12 nuclease family. *Nature* 2023;616:384–9. <https://doi.org/10.1038/s41586-023-05826-x>
- Zetsche B, Heidenreich M, Mohanraju P *et al.* Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 2017;35:31–4. <https://doi.org/10.1038/nbt.3737>
- Swarts DC, Jinek M. Mechanistic insights into the *cis*- and *trans*-acting DNase activities of Cas12a. *Mol Cell* 2019;73:589–600. <https://doi.org/10.1016/j.molcel.2018.11.021>
- Naqvi MM, Lee L, Montaguth OET *et al.* CRISPR-Cas12a-mediated DNA clamping triggers target-strand cleavage. *Nat Chem Biol* 2022;18:1014–22. <https://doi.org/10.1038/s41589-022-01082-8>
- Yan WX, Hunnewell P, Alfonse LE *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* 2019;363:88–91. <https://doi.org/10.1126/science.aav7271>
- Karvelis T, Bigelyte G, Young JK *et al.* PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res* 2020;48:5016–23. <https://doi.org/10.1093/nar/gkaa208>
- Wu Z, Zhang Y, Yu H *et al.* Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat Chem Biol* 2021;17:1132–8. <https://doi.org/10.1038/s41589-021-00868-6>
- Takeda SN, Nakagawa R, Okazaki S *et al.* Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol Cell* 2021;81:558–70. <https://doi.org/10.1016/j.molcel.2020.11.035>
- Xiao R, Li Z, Wang S *et al.* Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res* 2021;49:4120–8. <https://doi.org/10.1093/nar/gkab179>
- Hino T, Omura SN, Nakagawa R *et al.* An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* 2023;186:4920–35. <https://doi.org/10.1016/j.cell.2023.08.031>
- Su M, Li F, Wang Y *et al.* Molecular basis and engineering of miniature Cas12f with C-rich PAM specificity. *Nat Chem Biol* 2024;20:180–9. <https://doi.org/10.1038/s41589-023-01420-4>
- Song X, Chen Z, Sun W *et al.* CRISPR-AsCas12f1 couples out-of-protospacer DNA unwinding with exonuclease activity in the sequential target cleavage. *Nucleic Acids Res* 2024;52:14030–42. <https://doi.org/10.1093/nar/gkac989>
- Bigelyte G, Young JK, Karvelis T *et al.* Miniature type V-F CRISPR-Cas nucleases enable targeted DNA modification in cells. *Nat Commun* 2021;12:6191. <https://doi.org/10.1038/s41467-021-26469-4>
- Kim DY, Lee JM, Moon SB *et al.* Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by

- adeno-associated virus. *Nat Biotechnol* 2022;40:94–102. <https://doi.org/10.1038/s41587-021-01009-z>
23. Xu X, Chemparathy A, Zeng L *et al.* Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Mol Cell* 2021;81:4333–45. <https://doi.org/10.1016/j.molcel.2021.08.008>
 24. Wang Y, Wang Y, Pan D *et al.* Guide RNA engineering enables efficient CRISPR editing with a miniature *Syntrophomonas palmitatica* Cas12f1 nuclease. *Cell Rep* 2022;40:111418. <https://doi.org/10.1016/j.celrep.2022.111418>
 25. Kong X, Zhang H, Li G *et al.* Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat Commun* 2023;14:1–13. <https://doi.org/10.1038/s41467-023-37829-7>
 26. Gosse C, Croquette V. Magnetic tweezers: micromanipulation and force measurement at the molecular level. *Biophys J* 2002;82:3314–29. [https://doi.org/10.1016/S0006-3495\(02\)75672-5](https://doi.org/10.1016/S0006-3495(02)75672-5)
 27. De Vlaminck I, Henighan T, van Loenhout MTJ *et al.* Highly parallel magnetic tweezers by targeted DNA tethering. *Nano Lett* 2011;11:5489–93. <https://doi.org/10.1021/nl203299e>
 28. Klaue D, Seidel R. Torsional stiffness of single superparamagnetic microspheres in an external magnetic field. *Phys Rev Lett* 2009;102:028302. <https://doi.org/10.1103/PhysRevLett.102.028302>
 29. Wörle E, Jakob L, Schmidbauer A *et al.* Decoupling the bridge helix of Cas12a results in a reduced trimming activity, increased mismatch sensitivity and impaired conformational transitions. *Nucleic Acids Res* 2021;49:5278–93. <https://doi.org/10.1093/nar/gkab286>
 30. Rutkauskas M, Sinkunas T, Songailiene I *et al.* Directional R-loop formation by the CRISPR-Cas surveillance complex Cascade provides efficient off-target site rejection. *Cell Rep* 2015;10:1534–43. <https://doi.org/10.1016/j.celrep.2015.01.067>
 31. Rutkauskas M, Songailiene I, Irmisch P *et al.* A quantitative model for the dynamics of target recognition and off-target rejection by the CRISPR-Cas Cascade complex. *Nat Commun* 2022;13:7460. <https://doi.org/10.1038/s41467-022-35116-5>
 32. Aldag P, Rutkauskas M, Madariaga-Marcos J *et al.* Dynamic interplay between target search and recognition for a type I CRISPR-Cas system. *Nat Commun* 2023;14:3654. <https://doi.org/10.1038/s41467-023-38790-1>
 33. Bokori-Brown M, Martin TG, Naylor CE *et al.* Cryo-EM structure of lysenin pore elucidates membrane insertion by an aerolysin family protein. *Nat Commun* 2016;7:11293. <https://doi.org/10.1038/ncomms11293>
 34. Punjani A, Rubinstein JL, Fleet DJ *et al.* cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 2017;14:290–6. <https://doi.org/10.1038/nmeth.4169>
 35. Punjani A, Zhang H, Fleet DJ. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* 2020;17:1214–21. <https://doi.org/10.1038/s41592-020-00990-8>
 36. Abramson J, Adler J, Dunger J *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 2024;630:493–500. <https://doi.org/10.1038/s41586-024-07487-w>
 37. Pettersen EF, Goddard TD, Huang CC *et al.* UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci* 2021;30:70–82. <https://doi.org/10.1002/pro.3943>
 38. Emsley P, Lohkamp B, Scott WG *et al.* Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 2010;66:486–501. <https://doi.org/10.1107/S0907444910007493>
 39. Liebschner D, Afonine PV, Baker ML *et al.* Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* 2019;75:861–77. <https://doi.org/10.1107/S2059798319011471>
 40. Olechnovic K, Venclovas C. VoroContacts: a tool for the analysis of interatomic contacts in macromolecular structures. *Bioinformatics* 2021;37:4873–5. <https://doi.org/10.1093/bioinformatics/btab448>
 41. Aldag P, Welzel F, Jakob L *et al.* Probing the stability of the SpCas9-DNA complex after cleavage. *Nucleic Acids Res* 2021;49:12411–21. <https://doi.org/10.1093/nar/gkab1072>
 42. Luzzietti N, Knappe S, Richter I *et al.* Nicking enzyme-based internal labeling of DNA at multiple loci. *Nat Protoc* 2012;7:643–53. <https://doi.org/10.1038/nprot.2012.008>
 43. Daldrop P, Brutzer H, Huhle A *et al.* Extending the range for force calibration in magnetic tweezers. *Biophys J* 2015;108:2550–61. <https://doi.org/10.1016/j.bpj.2015.04.011>
 44. Virtanen P, Gommers R, Oliphant TE *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>
 45. Clement K, Rees H, Canver MC *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* 2019;37:224–6. <https://doi.org/10.1038/s41587-019-0032-3>
 46. Zhang H, Li Z, Xiao R *et al.* Mechanisms for target recognition and cleavage by the Cas12i RNA-guided endonuclease. *Nat Struct Mol Biol* 2020;27:1069–76. <https://doi.org/10.1038/s41594-020-0499-0>
 47. Szczelkun MD, Tikhomirova MS, Sinkunas T *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci USA* 2014;111:9798–803. <https://doi.org/10.1073/pnas.1402597111>
 48. Kauert DJ, Madariaga-Marcos J, Rutkauskas M *et al.* The energy landscape for R-loop formation by the CRISPR-Cas Cascade complex. *Nat Struct Mol Biol* 2023;30:1040–7. <https://doi.org/10.1038/s41594-023-01019-2>
 49. Crut A, Koster DA, Seidel R *et al.* Fast dynamics of supercoiled DNA revealed by single-molecule experiments. *Proc Natl Acad Sci USA* 2007;104:11957–62. <https://doi.org/10.1073/pnas.0700333104>
 50. Potter BV, Eckstein F. Cleavage of phosphorothioate-substituted DNA by restriction endonucleases. 1984;259:14243–8.
 51. Liu G, Ou HY, Wang T *et al.* Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet* 2010;6:e1001253. <https://doi.org/10.1371/journal.pgen.1001253>
 52. Kemmerich FE, Daldrop P, Pinto C *et al.* Force regulated dynamics of RPA on a DNA fork. *Nucleic Acids Res* 2016;44:5837–48. <https://doi.org/10.1093/nar/gkw187>