Supplementary Materials for

PAM recognition by miniature CRISPR-Cas14 triggers programmable double-stranded DNA cleavage

Tautvydas Karvelis[†*], Greta Bigelyte[†], Joshua K. Young[†], Zhenglin Hou, Rimante Zedaveinyte, Karolina Pociute, Arunas Silanskas, Česlovas Venclovas, Virginijus Siksnys[*]

[†]These authors contributed equally to this work.
[*]Correspondence to: tautvydas.karvelis@bti.vu.lt and siksnys@ibt.lt

This PDF file includes:

Materials and Methods
Figs. S1 to S8
Table S1
References (1-25)

Other Supplementary Materials for this manuscript include the following:

Data S1 (.xlsx)

Materials and Methods

Plasmid-borne Cas14b5 CRISPR-Cas system

The contig containing the CRISPR-Cas14b5 locus (GeneBank: MK005740.1 (30,101-34,986)) was first modified to target the plasmid DNA PAM library described in (16). This was accomplished by reducing the number of repeat:spacer:repeat units to three in the CRISPR array and replacing the native spacer sequences with a 36 bp (average length of repeats in the natural locus) sequence matching a region immediately 3' to the region of randomization. Next, the entire Cas14b5 locus engineered to target the 7N PAM library was synthesized (GeneScript) and cloned into an *E. coli* plasmid DNA (pACYC184, NEB) using PacI and BglII restriction sites. As control for some experiments, the engineered locus was further modified to contain a frameshift mutation in the *cas14b5* gene resulting in a disrupted ORF. This construct was generated by digestion with BsaI followed by klenow fill-in and re-ligation to introduce a 4bp insertion in the *cas14b5* gene. To enhance expression of the CRISPR-Cas14b5 locus, it was sub-cloned into pET-duet1 (MilliporeSigma) modified to contain a single T7 promoter and terminator. Deletions of the *cas1*, *cas2*, *cas4*, and *cas14b5* genes were introduced by first digesting the pET-duet1 containing the modified Cas14b5 locus with EcoNI and EcoRV. Fragments with overlapping ends reconstituting the lacO, T7 promoter, RBS (synthesized as geneblocks (IDT)) and the desired deletion locus (amplified by PCR) were incorporated into the EcoNI and EcoRV digested pET-duet1 backbone using seamless cloning (NEBuilder, NEB) and confirmed by sequencing. The links to the plasmid sequences are provided in Table S1.

Cas14a1 *E. coli* plasmid expression cassette

CRISPR-Cas14a1 *E. coli* lysate experiments to corroborate Cas14b5 dsDNA cleavage activity were performed with the pLBH531_MBP-Cas14a1 plasmid (gift from Jennifer Doudna, Addgene plasmid # 112500).

Cas14a1 sgRNA

Cas14a1 single guide RNAs (sgRNA) were produced by *in vitro* transcription using TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) and purified using GeneJET RNA Purification Kit (Thermo Fisher Scientific). Templates for T7 transcription were generated by PCR using overlapping oligonucleotides, altogether, containing a T7 promoter at the proximal end followed by the sgRNA sequence. Sequences of the sgRNA used in our study is available in Data S1 file.

Detecting Cas14 dsDNA cleavage and PAM recognition

Plasmid DNA targets were cleaved with Cas14-gRNA ribonucleoprotein (RNP) complexes produced from the modified locus (Cas14b5) or by combining *E. coli* lysate containing Cas14 protein with T7 transcribed sgRNA (Cas14a1). Cell lysate was obtained by transforming Cas14 plasmids into DH5α (pACYC plasmids) or ArcticExpress (DE3) (pET-duet1 and pLBH531 plasmids) cells and cultures grown in LB broth (30 ml) supplemented with either chloramphenicol (25 µg/ml) (pACYC plasmids) or ampicillin (100 µg/ml) (pET-duet1 and pLBH531 plasmids). For plasmids with a T7 promoter (pET-duet1 and pLBH531 plasmids), expression was induced with 0.5 mM IPTG when cultures reached an OD_{600} of 0.5 and incubated overnight at 16°C. Cells (from 10 ml) were collected by centrifugation and resuspended in 1 ml of lysis buffer (20 mM phosphate, pH 7.0, 0.5 M NaCl, 5% (v/v) glycerol) supplemented with 10 µl PMSF and lysed by sonication. Cell debris was removed by centrifugation. In the case of

Cas14b5 experiments, 10 µl of the supernatant containing Cas14b5-gRNA RNP was used directly in digestion experiments. For Cas14a1, 20 µl of clarified supernatant was combined with 1 µl of RiboLock RNase Inhibitor (Thermo Fisher Scientific, EO0381) and 2 µg of sgRNA and allowed to complex with the clarified lysate as described below.

Cas14 RNP complexes were used to cleave either the 7 bp randomized PAM plasmid DNA library described previously (16) or a plasmid DNA containing a fixed PAM and gRNA target. Briefly, 10 µl of Cas14-gRNA RNP containing lysate was mixed with 1 µg of PAM library or 1 µg of plasmid containing a single PAM and gRNA target in 100 µl of reaction buffer (10 mM Tris-HCl, pH 7.5 at 37°C, 100 mM NaCl, 1 mM DTT, and 10 mM MgCl2). After a 1 h incubation at 37°C, DNA ends were repaired by adding 1 µl of T4 DNA polymerase (Thermo Fisher Scientific) and 1 µl of 10 mM dNTP mix (Thermo Fisher Scientific) and incubating the reaction for 20 min at 11°C. The reaction was then inactivated by heating it up to 75°C for 10 min and 3'-dA overhangs added by incubating the reaction mixture with 1 µl of DreamTaq polymerase (Thermo Fisher Scientific) and 1 µl of 10 mM dATP (Thermo Fisher Scientific) for 30 min at 72°C. Additionally, RNA was removed by incubation for 15 min at 37°C with 1 µl RNase A (Thermo Fisher Scientific). Following purification with a GenJet PCR Purification column (Thermo Fisher Scientific), the end repaired cleavage products (100 ng) were ligated with a double-stranded DNA adapter containing a 3'-dT overhang (100 ng) for 1 h at 22°C using T4 DNA ligase (Thermo Fisher Scientific). After ligation, cleavage products were PCR amplified appending sequences required for deep sequencing and subjected to Illumina sequencing (16, 17).

Double-stranded DNA target cleavage was evaluated by examining the unique junction generated by target cleavage and adapter ligation in deep sequence reads. This was accomplished by first generating a collection of sequences representing all possible outcomes of dsDNA cleavage and adapter ligation within the target region. For example, cleavage and adapter ligation at just after the 21st position of the target would produce the following sequence (5'-CCGCTCTTCCGATCTGCCGGCGACGTTGGGTCAACT-3') where the adapter and target sequences comprise 5'-CCGCTCTTCCGATCT-3' and 5'-GCCGGCGACGTTGGGTCAACT-3', respectively. The frequency of the resulting sequences was then tabulated using a custom python script and compared to negative controls (experiments setup without Cas14b5 locus or the Cas14a1 sgRNA) to identify target cleavage.

Evidence of PAM recognition was evaluated as described previously (16, 17). Briefly, the sequence of the protospacer adapter ligation exhibiting an elevated frequency in the previous step was used in combination with a 10 bp sequence 5' of the 7 bp PAM region to identify reads that supported dsDNA cleavage. Once identified, the intervening 7 bp PAM sequence was isolated by trimming away the 5' and 3' flanking sequences using a custom python script and the frequency of the extracted PAM sequences normalized to the original PAM library to account for inherent biases using the following formula.

Normalized Frequency = (Treatment Frequency)/((Control Frequency)/(Average Control Frequency))

Following normalization, a position frequency matrix (PFM) (25) was calculated and compared to negative controls (experiments setup without Cas14b5 locus or the Cas14a1 sgRNA) to look for biases in nucleotide composition as a function of PAM position. Biases were considered significant and indicative of PAM recognition if they deviated by more than 2.5-fold from the

negative control. Analyses were limited to the top 10% most frequent PAMs to reduce the impact of background noise resulting from non-specific cleavage coming from other components in the *E. coli* cell lysate mixtures.

Expression and purification of Cas14a1 protein

To further study Cas14 dsDNA cleavage, Cas14a1 protein was expressed in either *E. coli* BL21(DE3) or ArcticExpress (DE3) strains from the pLBH531_MBP-Cas14a1 plasmid. Cas14a1^{D326A} and Cas14a1^{D510A} expression plasmids were engineered from pLBH531 using Phusion Site-Directed Mutagenesis Kit (Thermo Fisher Scientific). For the NLS-tagged Cas14a1 variant, the pRZ47-NLS-Cas14-NLS plasmid was constructed by inserting synthetic NLS-Cas14-NLS construct (Twist Bioscience) into pBAD24-6×His-MBP vector using Gibson assembly kit (NEB). For all, *E. coli* cells were grown in LB broth supplemented with ampicillin (100 μg/ml) at 37°C temperature. After culturing to an OD_{600} of 0.5, temperature was decreased to 16°C and expression induced with 0.5 mM IPTG for Cas14a1, Cas14a1^{D326A} and Cas14a1^{D510A}, or 0.2 % (w/v) arabinose for the NLS-tagged Cas14a1 variant. After 16 h cells were pelleted, re-suspended in loading buffer (20 Tris-HCl, pH 8.0 at 25°C, 1.5 M NaCl, 5 mM 2-mercaptoethanol, 10 mM imidazole, 2 mM PMSF, 5% (v/v) glycerol) and disrupted by sonication. Cell debris was removed by centrifugation. The supernatant was loaded on the Ni^{2+}-charged HiTrap chelating HP column (GE Healthcare) and eluted with a linear gradient of increasing imidazole concentration (from 10 to 500 mM) in 20 Tris-HCl, pH 8.0 at 25°C, 0.5 M NaCl, 5 mM 2-mercaptoethanol buffer. The fractions containing Cas14a1 were pooled and subsequently loaded on HiTrap heparin HP column (GE Healthcare) for elution using a linear gradient of increasing NaCl concentration (from 0.1 to 1.5 M). The next fractions containing Cas14a1 were pooled and the 10×His-MBP-tag was cleaved (or 6xHis-MBP-tag for NLS-tagged variant) by overnight incubation with TEV protease at 4°C. To remove cleaved His-MBP-tag and TEV protease, reaction mixtures were loaded onto a HiTrap heparin HP 5 column (GE Healthcare) for elution using a linear gradient of increasing NaCl concentration (from 0.1 to 1.5 M). Next, the elution from the HiTrap column was loaded on a MBPTrap column (GE Healthcare) and Cas14a1 protein was collected as flow-through. The collected fractions with Cas14a1 were then dialyzed against 20 mM Tris-HCl, pH 8.0 at 25°C, 500 mM NaCl, 2 mM DTT and 50% (v/v) glycerol and stored at -20°C. The sequences of the Cas14a1 protein variants are listed in Data S1 file.

Cas14-sgRNA complex assembly

Cas14a1 ribonucleoprotein (RNP) complexes (1 μM) were assembled by mixing Cas14a1 protein with sgRNA at 1:1 molar ratio followed by incubation in a complex assembly buffer (10 mM Tris-HCl, pH 7.5 at 37°C, 100 mM NaCl, 1 mM EDTA, 1 mM DTT) at 37°C for 30 min.

DNA substrate generation

Plasmid DNA substrates were generated by cloning oligoduplexes assembled after annealing complementary oligonucleotides (Metabion) into pUC18 plasmid over HindIII (Thermo Fisher Scientific) and EcoRI (Thermo Fisher Scientific) sites. The sequences of the inserts are listed in Data S1 file.

To generate radiolabeled DNA substrates, the 5'-ends of oligonucleotides were radiolabeled using T4 PNK (Thermo Fisher Scientific) and [γ-33P]ATP (PerkinElmer). Duplexes were made by annealing two oligonucleotides with complementary sequences at 95°C following slow

cooling to room temperature. A radioactive label was introduced at the 5'-end of individual DNA strands before annealing with unlabeled strands. The sequences of the oligoduplexes are provided in Data S1 file.

DNA substrate cleavage assay

Plasmid DNA cleavage reactions were initiated by mixing plasmid DNA with Cas14a1 RNP complex. The final reaction mixture typically contained 3 nM plasmid DNA, 100 nM Cas14a1 RNP complex in 2.5 mM Tris-HCl, pH 7.5 at 37°C, 25 mM NaCl, 0.25 mM DTT and 10 mM $MgCl_2$ reaction buffer. Aliquots were removed at timed intervals (30 min if not indicated differently) and mixed with 3× loading dye solution (0.01% Bromophenol Blue and 75 mM EDTA in 50% (v/v) glycerol) and reaction products were analyzed by agarose gel electrophoresis and ethidium bromide staining.

Reactions with oligoduplexes were typically carried out by mixing labelled oligoduplex with Cas14a1 RNP complex and incubating at 46°C. The final reaction mixture contained 1 nM labeled duplex, 100 nM Cas14a1 RNP complex, 5 mM Tris-HCl, pH 7.5 at 37°C, 50 mM NaCl, 0.5 mM DTT and 5 mM $MgCl_2$ in a 100 µl reaction volume. Aliquots of 6 µl were removed from the reaction mixture at timed intervals (0, 1, 2, 5, 10, 15 and 30 min), quenched with 10 µl of loading dye (95% (v/v) formamide, 0.01% Bromophenol Blue and 25 mM EDTA) and subjected to denaturing gel electrophoresis (20% polyacrylamide containing 8.5 M urea in 0.5× TBE buffer). Gels were dried and visualized by phosphorimaging.

M13 cleavage assay

M13 ssDNA cleavage reactions were initiated by mixing M13 ssDNA (New England Biolabs) and DNA activator with Cas14a1 RNP complex at 46°C. Cleavage assays were conducted in 2.5 mM Tris-HCl, pH 7.5 at 37°C, 25 mM NaCl, 0.25 mM DTT and 10 mM $MgCl_2$. The final reaction mixture contained 5nM M13 ssDNA, 100 nM ssDNA or dsDNA activator and 100 nM Cas14a1 RNP. The reaction was initiated by addition of Cas14a1 RNP complex and was quenched at timed intervals (0, 5, 15, 30, 60 and 90 min) by mixing with 3× loading dye solution (0.01% Bromophenol Blue and 75 mM EDTA in 50% (v/v) glycerol). Products were separated on an agarose gel and stained with SYBR Gold (Thermo Fisher Scientific). The sequences of the activators are listed in Data S1 file.

In vivo genome editing

HEK293 (ATCC) cells were cultured in DMEM (Gibco) with 10% FBS (Gibco) and penicillin/streptomycin (Gibco) at 37°C in 5% CO_2. A day prior to transfection cells were seeded in 96-well plates at 3.6×10^4 density. NLS-tagged Cas14a1 RNP complex was assembled by mixing 20 pmol of purified NLS-Cas14a1-NLS protein with 20 pmol of sgRNA in 25 µl Opti-MEM (Gibco) and incubated at room temperature for 30 min. After complex assembly 25 µl of Opti-MEM, containing 1.2 µl Lipofectamine 3000 (Thermo Fisher Scientific) was added and the mixture was incubated for additional 15 min at room temperature before transfection of the cells. Genomic DNA was extracted 72 h after transfection using QuickExtract DNA Extraction Solution (Lucigen) and regions surrounding Cas14a1 target sites amplified using Phusion high fidelity PCR mix (NEB) tailing on sequences needed for Illumina deep sequencing bridge-amplification and sample deconvolution. Libraries were single-read deep sequenced on a MiSeq Personal Sequencer (Illumina, USA) with a 25% (v/v) spike of PhiX control v3 (Illumina, USA) and sequences post-processed and deconvoluted per the manufacture's instruction. Only reads

with an insertion or deletion (indel) mutation within a 12 bp window centered on the expected Cas14a1 cut-site were counted as modified. Total reads were calculated by summing the number of reads matching the expected wildtype amplicon and indel reads. Statically relevance over controls (transfections assembled without the sgRNA) was calculated using a one-tailed t-test at a 0.05 significance level.

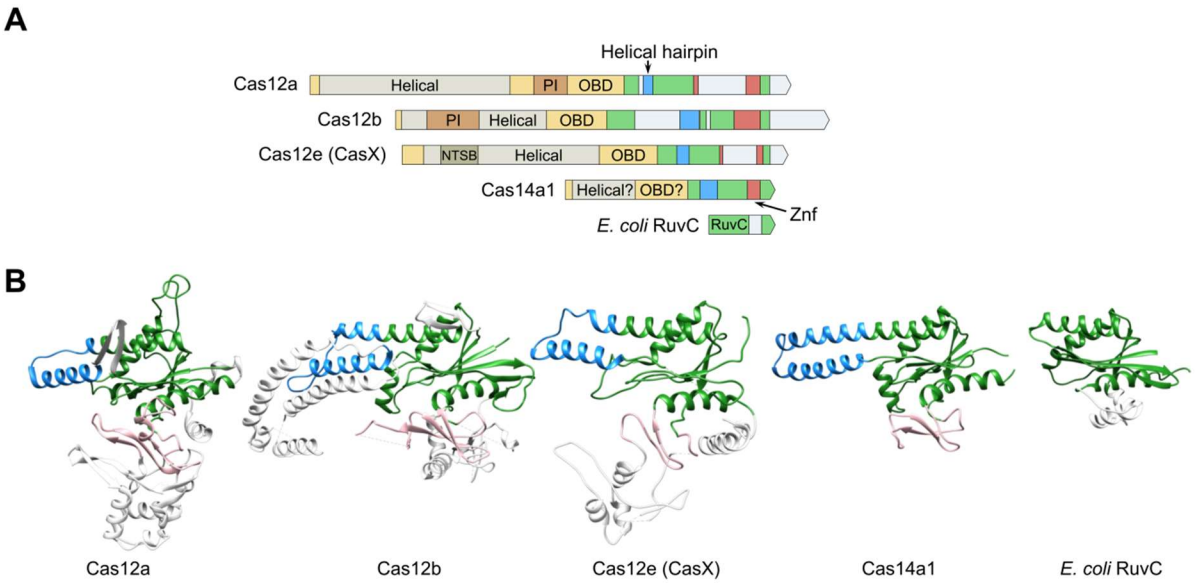

Fig. S1.

Protein architecture and C-terminal RuvC domain structural comparisons between Cas12 dsDNA effectors and Cas14.

Comparison includes Cas12a (PDB id: 5xut), Cas12b (PDB id: 5wti), Cas12e (CasX) (PDB id: 6ny2), Cas14a1 (model), and *E. coli* RuvC (PDB id: 1hjr). (**A**) Organization of protein domains and relative size comparison between known type V dsDNA effectors and Cas14. (**B**) C-terminal RuvC domain of known type V dsDNA effectors, Cas14, and *E. coli* RuvC. The common RuvC core is colored green in all the structures. Helical hairpin common to Cas proteins, but absent from the *E. coli* RuvC is shown in blue. Common zinc finger motif (CasX and Cas14a1) or a zinc finger-like motif (Cas12a and Cas12b) is shown in pink. Additional structure-specific motifs are shown in grey.

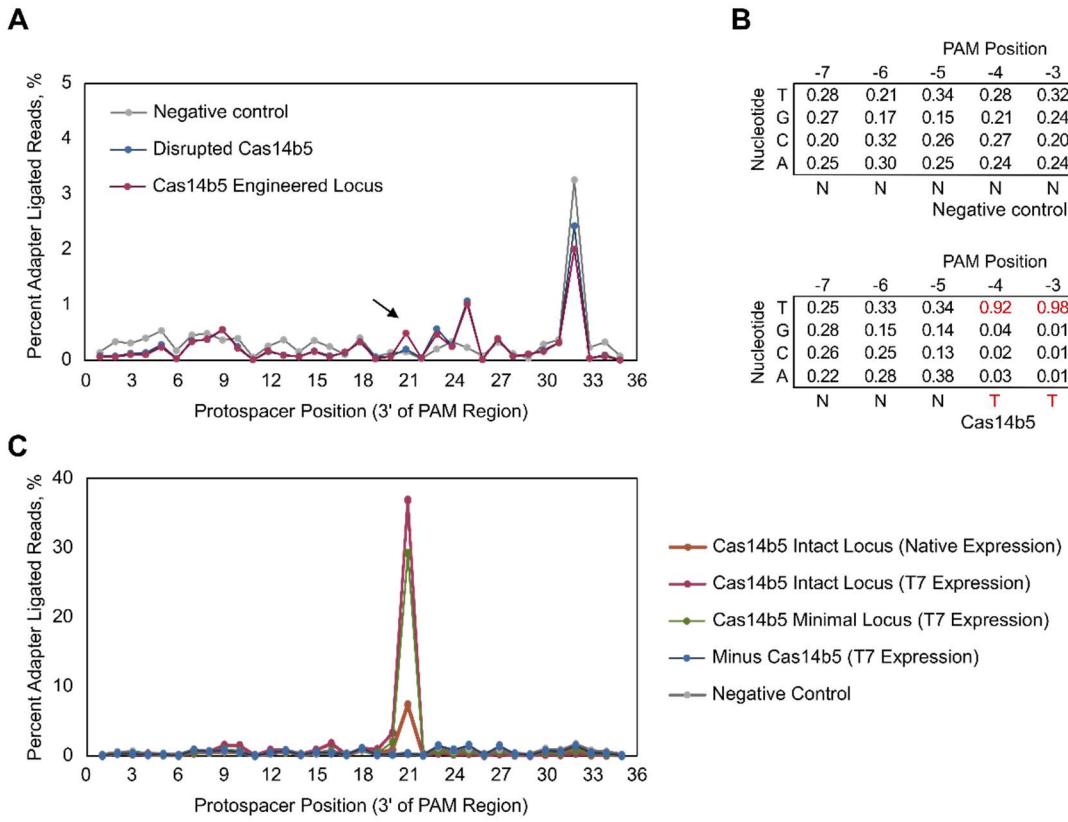

Fig. S2.

Cas14b5 dsDNA recognition and cleavage.

(**A**) Relative to the negative controls, the Cas14b5 locus engineered to target a PAM library produced a spike in the recovery of protospacer fragments ligated to an adapter just after the 21st position 3' of the PAM region. (**B**) PAM sequences that supported cleavage generated a position frequency matrix (PFM) exhibiting preferences for T and A bps 5' of the gRNA target. As a reference, a PFM at the same position in the lysate only control was also calculated. (**C**) dsDNA plasmids containing a PAM and gRNA target showed an even greater enrichment in the recovery of adapters ligated just after the 21st position, especially, for reactions where expression was enhanced with a T7 promoter. Experiments deleting *cas1*, *cas2*, and *cas4* genes (Cas14b5 Minimal Locus) and the *cas14b5* gene itself (Minus Cas14b5) confirmed that Cas14b5 was the only protein required for the observed dsDNA target recognition and cleavage.

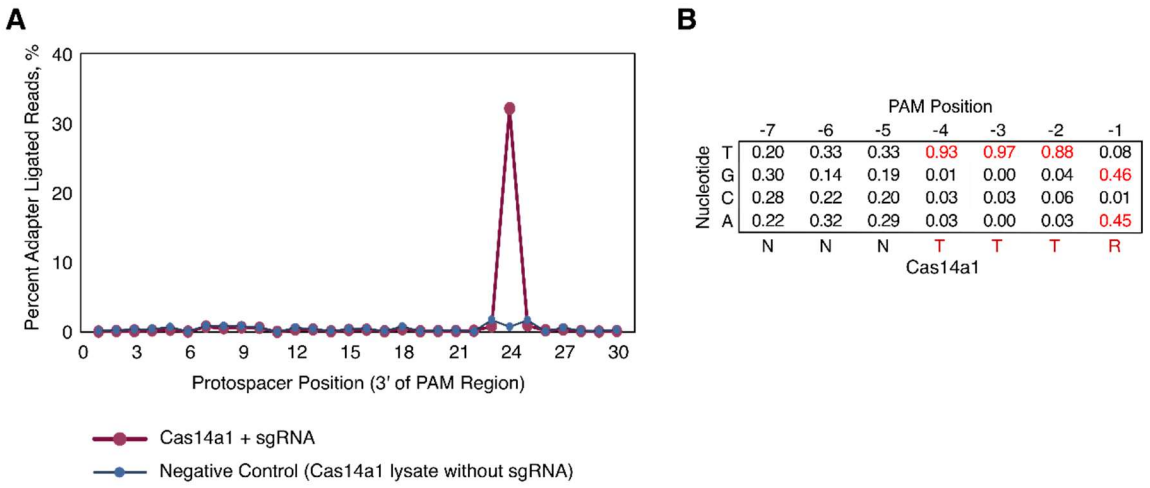

Fig. S3.

Cas14a1 dsDNA recognition and cleavage.

(**A**) *E. coli* lysate containing Cas14a1 and sgRNA targeting the PAM library produced an enrichment in the recovery of protospacer adapter ligated fragments just after the 24st position 3' of the PAM region relative to the negative control. (**B**) Displayed as a PFM, PAM sequences that supported cleavage showed a 5' T-rich PAM.

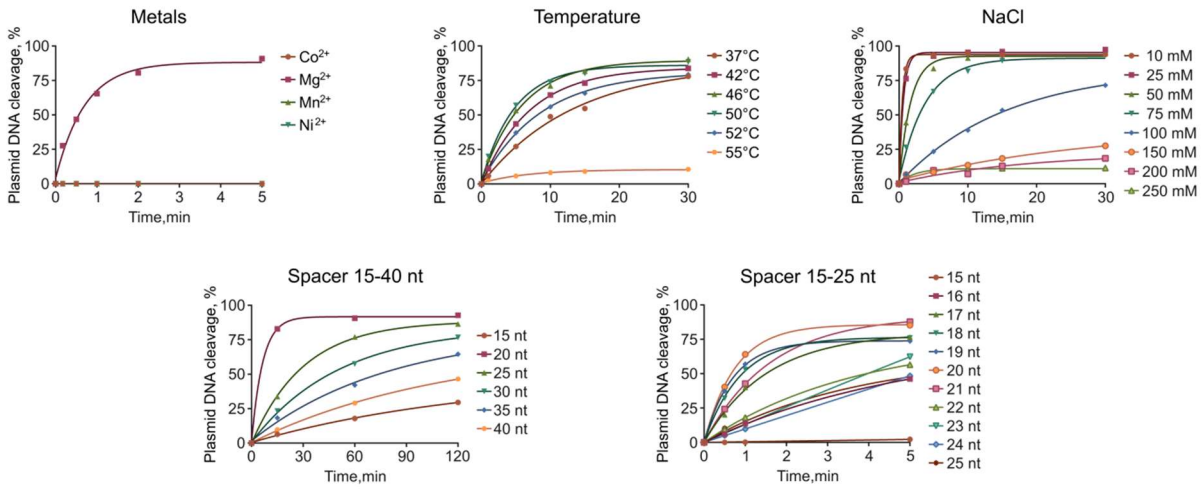

Fig. S4.

Optimization of reaction conditions for Cas14a1 RNP mediated plasmid DNA cleavage. Cas14a1 RNP plasmid DNA cleavage was assayed by independently varying divalent metal ions, temperature, NaCl concentration, and sgRNA spacer length.

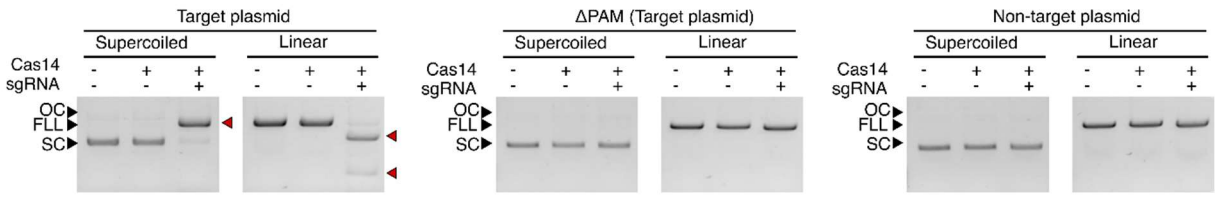

Fig. S5.

Cas14a1 RNP complex is a PAM-dependent dsDNA endonuclease.

Purified Cas14a1 RNP complexes cleaved plasmid DNA targets in a PAM-dependent manner (left panel) requiring both PAM (center panel) and sgRNA recognition (right panel).

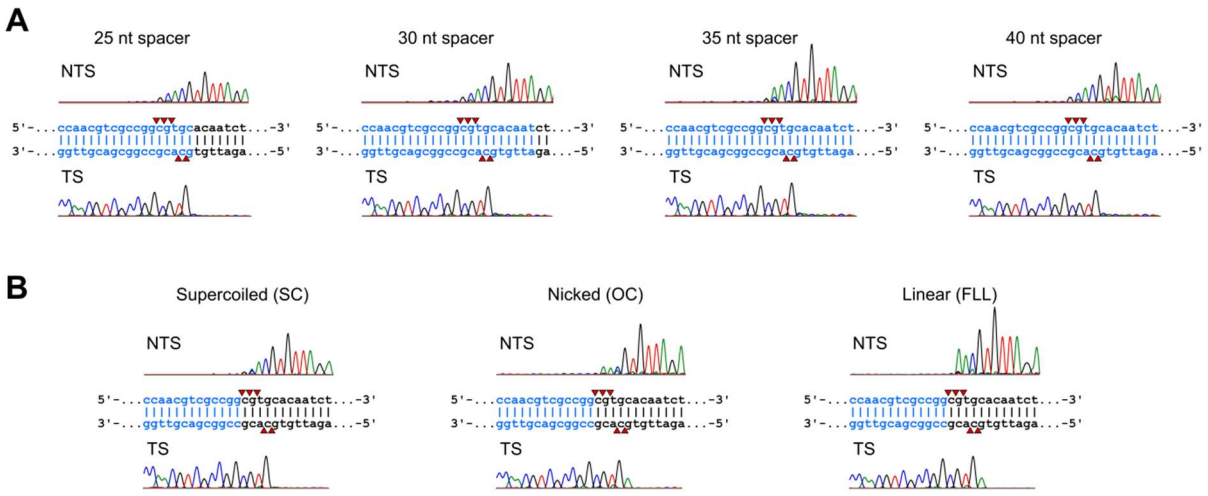

Fig. S6.

Run-off sequencing of Cas14a1 cleaved plasmid DNA.
Plasmid DNA cleavage resulted in a double-stranded break centered around positions 20-24 bp 3' of the PAM and the cleavage pattern was independent of spacer length (**A**) and plasmid topology (**B**). NTS and TS represent non-target strand and target strand, respectively. SC, OC, and FLL stand for supercoiled, open-circle, and full-length linearization, respectively.

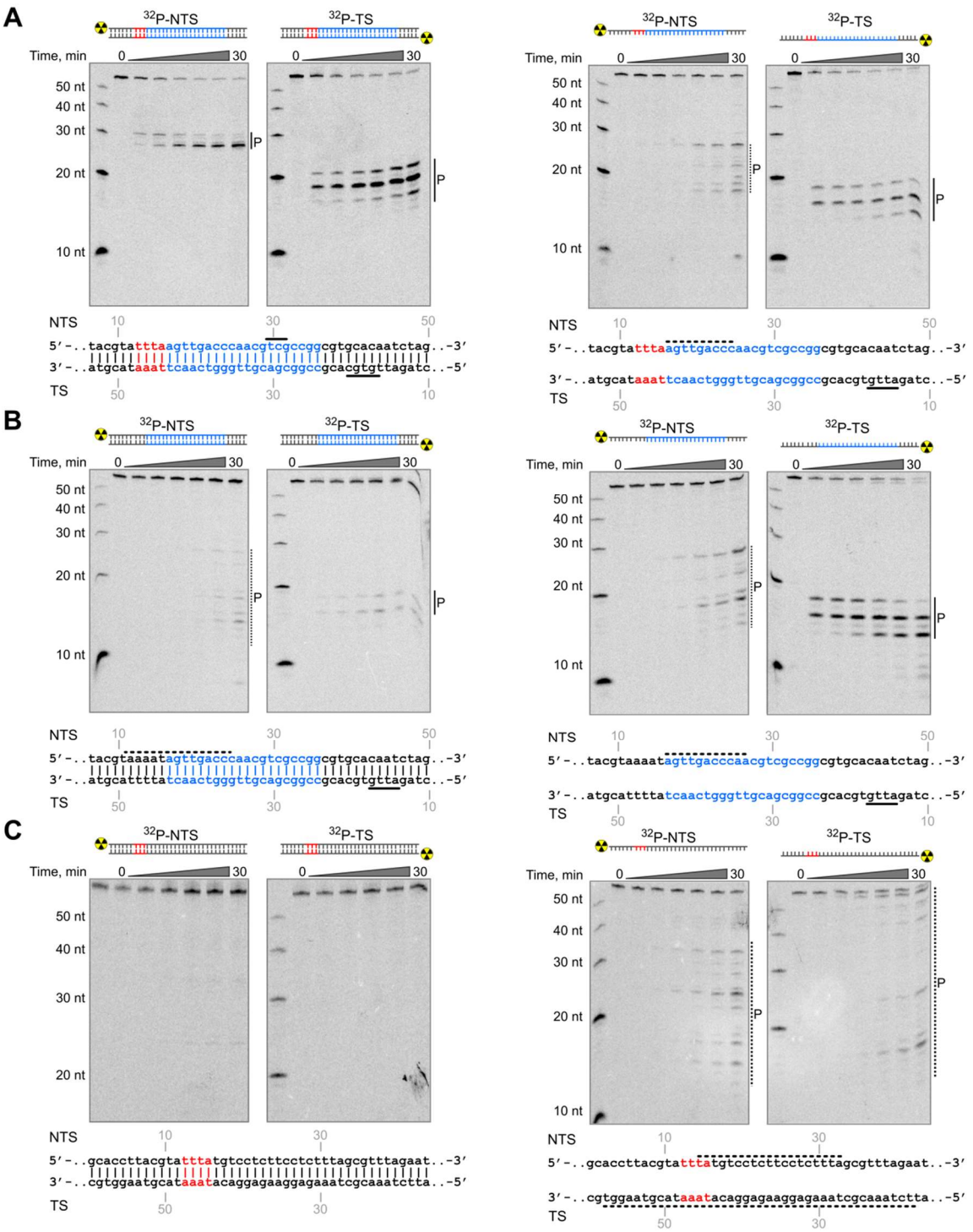

Fig. S7.

Oligoduplex cleavage by Cas14a1 RNP complex.

(**A**) Purified Cas14a1 RNP complexes cleaved radiolabeled dsDNA oligoduplexes containing a sgRNA target in a PAM-dependent manner generating a staggered cleavage pattern. dsDNA substrates without PAM (**B**) or target sequence (**C**) were not cleaved by the Cas14a1 RNP complex. ssDNA substrates complementary to the sgRNA spacer sequence were also cleaved albeit in the PAM-independent manner (**A** and **B**).

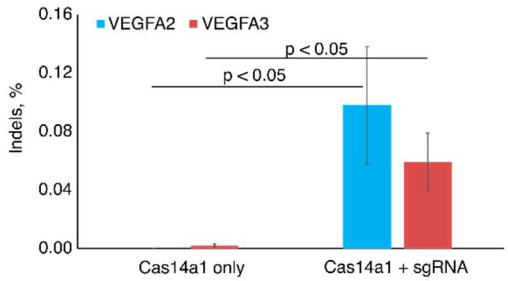

Fig. S8.

Human chromosomal genome editing frequencies using Cas14a1 RNP.
Cas14a1 RNP complexes targeting VEGFA2 and 3 loci were delivered into HEK293 cells. Insertion or deletion (indel) mutation frequencies were assayed using targeted deep sequencing. Cas14a1-sgRNA lipofections consistently resulted in the recovery of indel mutations at frequencies greater than those recovered in negative controls. P-value calculated using a one-sided t-test at a 0.05 confidence interval.

Table S1. Plasmid map links.

Plasmid name	Description	Link
PV424	Cas14b5 intact locus (native expression)	https://benchling.com/s/seq-uDaLUdexDYQQSz7QDMXF
PV477	Disrupted Cas14b5 (native expression)	https://benchling.com/s/seq-y7Fh5ryHUNNX7R3sDYMa
R-652	Cas14b5 intact locus (T7 expression)	https://benchling.com/s/seq-DNGdXS8DCheZt5Nv6kMu
R-656	Cas14b5 minimal locus (T7 expression)	https://benchling.com/s/seq-kUL4zwpBZDDJIwIwwLBT
R-658	Minus Cas14b5 (T7 expression)	https://benchling.com/s/seq-A2E4WYBe0vLNbHiWQ2py
pLBH531	10×His-MBP-Cas14a1 expression	https://www.addgene.org/112500/
pGB49	10×His-MBP-Cas14a1 D326A expression	https://benchling.com/s/seq-qpUDPxM6IBhTJFXpNVUJ
pGB50	10×His-MBP-Cas14a1 D510A expression	https://benchling.com/s/seq-W1miOvRPGZ44fgZnV0Nn
pRZ47	6×His-NLS-Cas14a1-NLS expression	https://benchling.com/s/seq-JODc8yzEKzAcomWc9djA
pTZ57	7N PAM plasmid library	https://benchling.com/s/seq-nu2IvfXbn7smVQ7T6MYi
pGB33	Cas14b5 target plasmid	https://benchling.com/s/seq-iYcV6jflHOPbUxMdCGs9
pGB40	Cas14a1 target plasmid	https://benchling.com/s/seq-XGplLg5diY1G7BwBDHXn
pGB41	Cas14a1 ΔPAM (target plasmid)	https://benchling.com/s/seq-QS7mA2qo4e3JRlBvKsXb
pGB42	Cas14a1 non-target plasmid	https://benchling.com/s/seq-rXjk6jlGlmPUSbb2GeT9

Data S1 (separate file).
DNA, RNA and protein sequences used in this study.

References

1. Koonin,E. V., Makarova,K.S. and Wolf,Y.I. (2017) Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.*, **71**, 233–261.
2. Koonin,E. V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
3. 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.
4. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.*, **109**, 2579–2586.
5. 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.
6. Ma,E., Harrington,L.B., O'Connell,M.R., Zhou,K. and Doudna,J.A. (2015) Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell*, **60**, 398–407.
7. Zhang,Y., Rajan,R., Seifert,H.S., Mondragón,A. and Sontheimer,E.J. (2015) DNase H Activity of Neisseria meningitidis Cas9. *Mol. Cell*, **60**, 242–255.
8. Chen,J.S., Ma,E., Harrington,L.B., Da Costa,M., Tian,X., Palefsky,J.M. and Doudna,J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, **360**, 436–439.
9. 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.
10. 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.
11. Knott,G.J. and Doudna,J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*, **361**, 866–869.
12. Klompe,S.E. and Sternberg,S.H. (2018) Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *Cris. J.*, **1**, 141–158.
13. Wu,Z., Yang,H. and Colosi,P. (2010) Effect of genome size on AAV vector packaging. *Mol. Ther.*, **18**, 80–6.
14. Lino,C.A., Harper,J.C., Carney,J.P. and Timlin,J.A. (2018) Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv.*, **25**, 1234–1257.
15. 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.
16. Karvelis,T., Gasiunas,G., Young,J., Bigelyte,G., Silanskas,A., Cigan,M. and Siksnys,V. (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.*, **16**, 253.
17. Karvelis,T., Young,J.K. and Siksnys,V. (2019) A pipeline for characterization of novel Cas9 orthologs. In *Methods in Enzymology*. Elsevier Inc., Vol. 616, pp. 219–240.
18. 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.
19. 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.e4.
20. Strohkendl,I., Saifuddin,F.A., Rybarski,J.R., Finkelstein,I.J. and Russell,R. (2018) Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol. Cell*, **71**, 816-824.e3.
21. Fonfara,I., Richter,H., Bratovič,M., Le Rhun,A. and Charpentier,E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517–521.
22. Stella,S., Alcón,P. and Montoya,G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*, **546**, 559–563.
23. Gootenberg,J.S., Abudayyeh,O.O., Lee,J.W., Essletzbichler,P., Dy,A.J., Joung,J., Verdine,V., Donghia,N., Daringer,N.M., Freije,C.A., *et al.* (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, **356**, 438–442.
24. Gootenberg,J.S., Abudayyeh,O.O., Kellner,M.J., Joung,J., Collins,J.J. and Zhang,F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, **360**, 439–444.
25. Stormo,G.D. (2013) Modeling the specificity of protein-DNA interactions. *Quant. Biol.*, **1**, 115–130.