دانشگاه علوم پزشکی و خدمات بهداشتی،درمانی کرمانشاه
Kermanshah University Of Medical Sciences

Nucleic Acids Research Journal

Journal club
Tuesday , Feb23 , 2021

PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage 2020

Presenter : Sasan Bouk
Supervisor : Dr. Yadollah Bahrami

Department of Medical Biotechnology
Kermanshah University of Medical Sciences

Faculty of Medicine, Class No.3 [Shahid Nowroznezhad] 12:30 p.m.
سه شنبه ، 5 اسفند 1399 ، ساعت 12:30

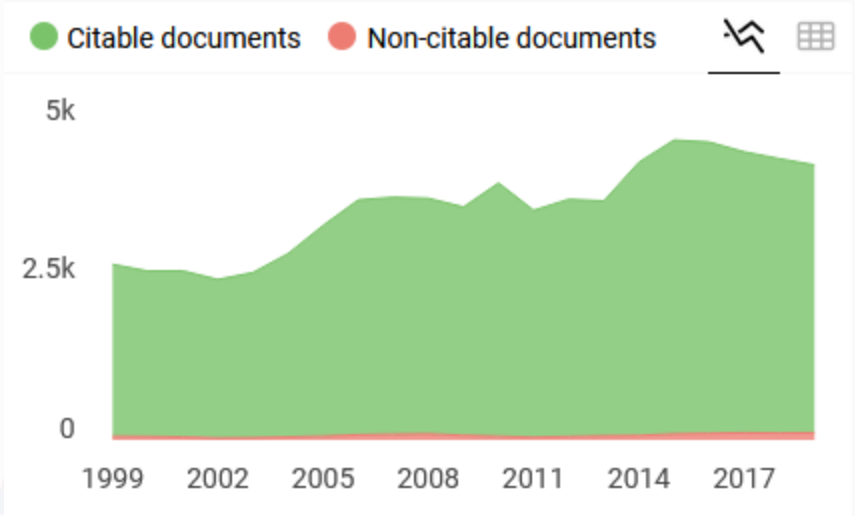
5016–5023 Nucleic Acids Research, 2020, Vol. 48, No. 9 Published online 4 April 2020
doi: 10.1093/nar/gkaa208

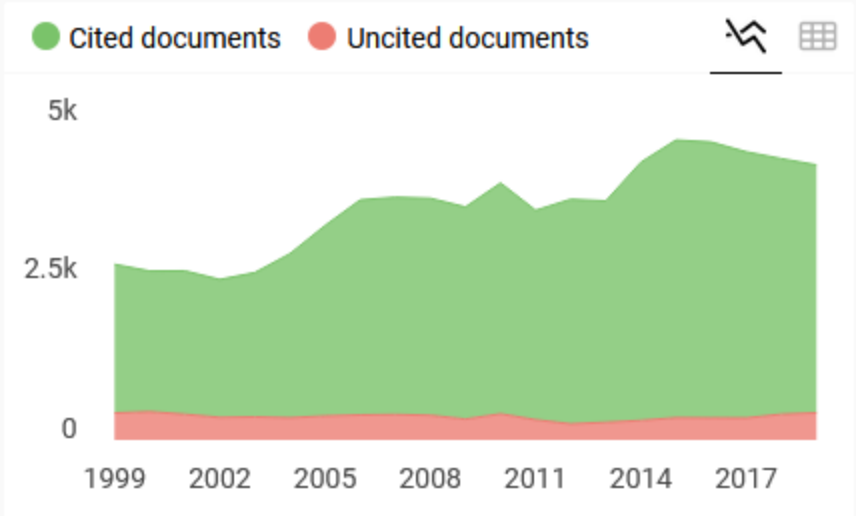
PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage

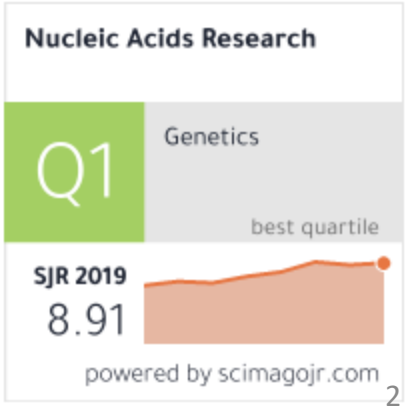
Tautvydas Karvelis[1,*,†], Greta Bigelyte[1,†], Joshua K. Young[2,*,†], Zhenglin Hou[2], Rimante Zedaveinyte[1], Karolina Budre[1], Sushmitha Paulraj[2], Vesna Djukanovic[2], Stephen Gasior[2], Arunas Silanskas[1], Česlovas Venclovas[1] and Virginijus Siksnys[1,*]

[1]Institute of Biotechnology, Vilnius University, Vilnius LT-10257, Lithuania and [2]Corteva Agriscience™, Johnston, IA 50131, USA

Received February 20, 2020; Revised March 17, 2020; Editorial Decision March 19, 2020; Accepted April 02, 2020

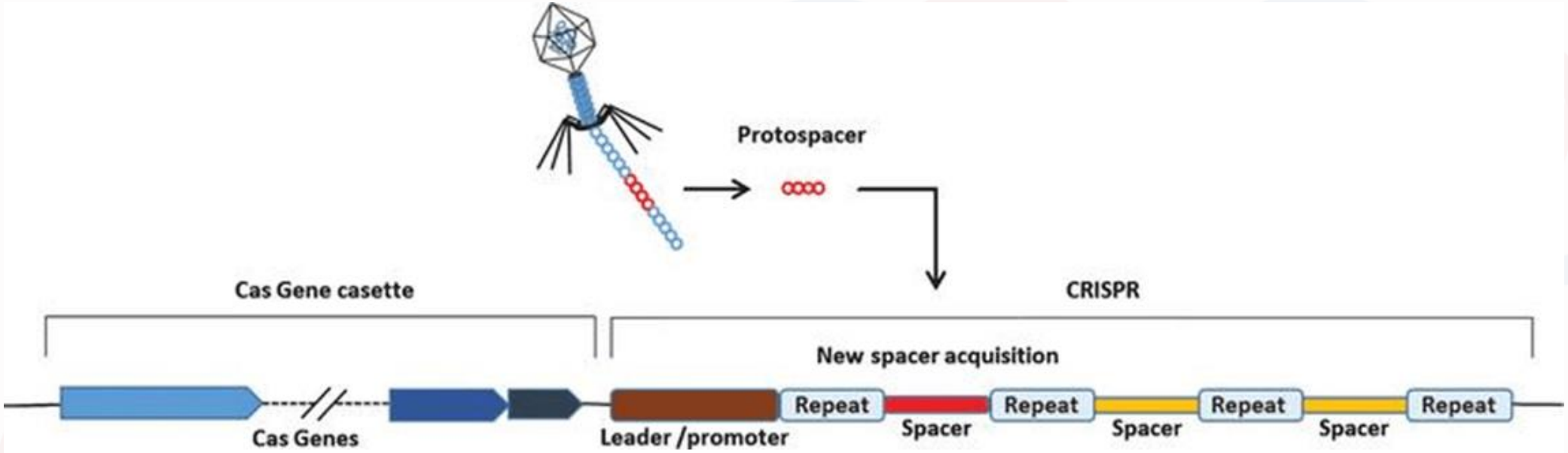






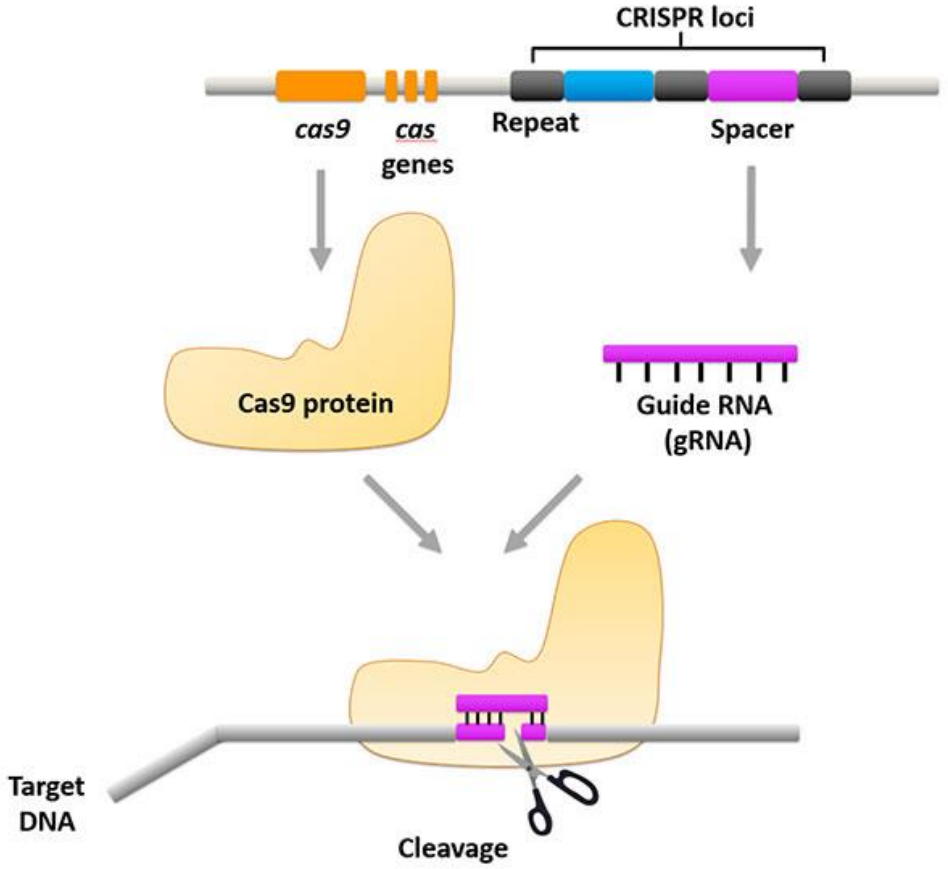
introduction

- CRISPR (clustered regularly interspaced short palindromic repeats)
- family of DNA sequences found in the genomes of prokaryotic organisms
- derived from DNA fragments of bacteriophages that had previously infected the prokaryote
- used to detect and destroy DNA from similar bacteriophages during subsequent infections



- Cas9 (or "CRISPR-associated protein 9")
- an enzyme that uses CRISPR sequences as a guide
- recognize and cleave specific strands of DNA that are complementary to the CRISPR sequence
- RNA harboring the spacer sequence helps

Cas (CRISPR-associated) proteins recognize and cut foreign pathogenic DNA



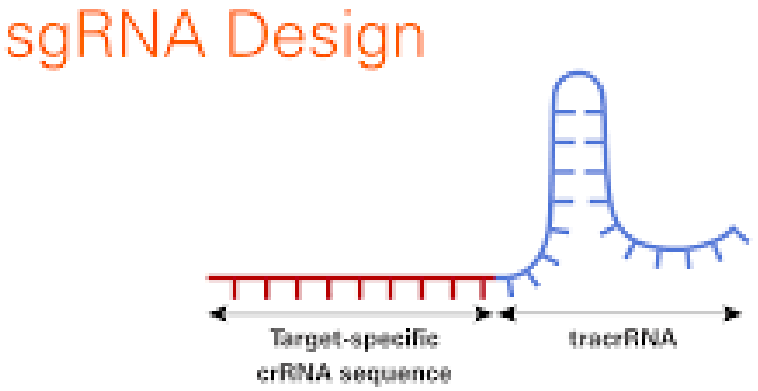
- **The guide RNA** is a specific RNA sequence
- recognizes the target DNA region of interest
- directs the Cas nuclease there for editing
- made up of two parts:

crispr RNA (crRNA), a 17-20 nucleotide sequence complementary to the target DNA, is the *customizable component* that *enables specificity* in every CRISPR and

a **tracr RNA**, which serves as a binding scaffold for the Cas nuclease

- **The sgRNA** is an abbreviation for “single guide RNA“
- is a single RNA molecule that contains both the custom-designed short crRNA sequence fused to the scaffold tracrRNA sequence
- can be synthetically generated or made in vitro or

in vivo from a DNA template



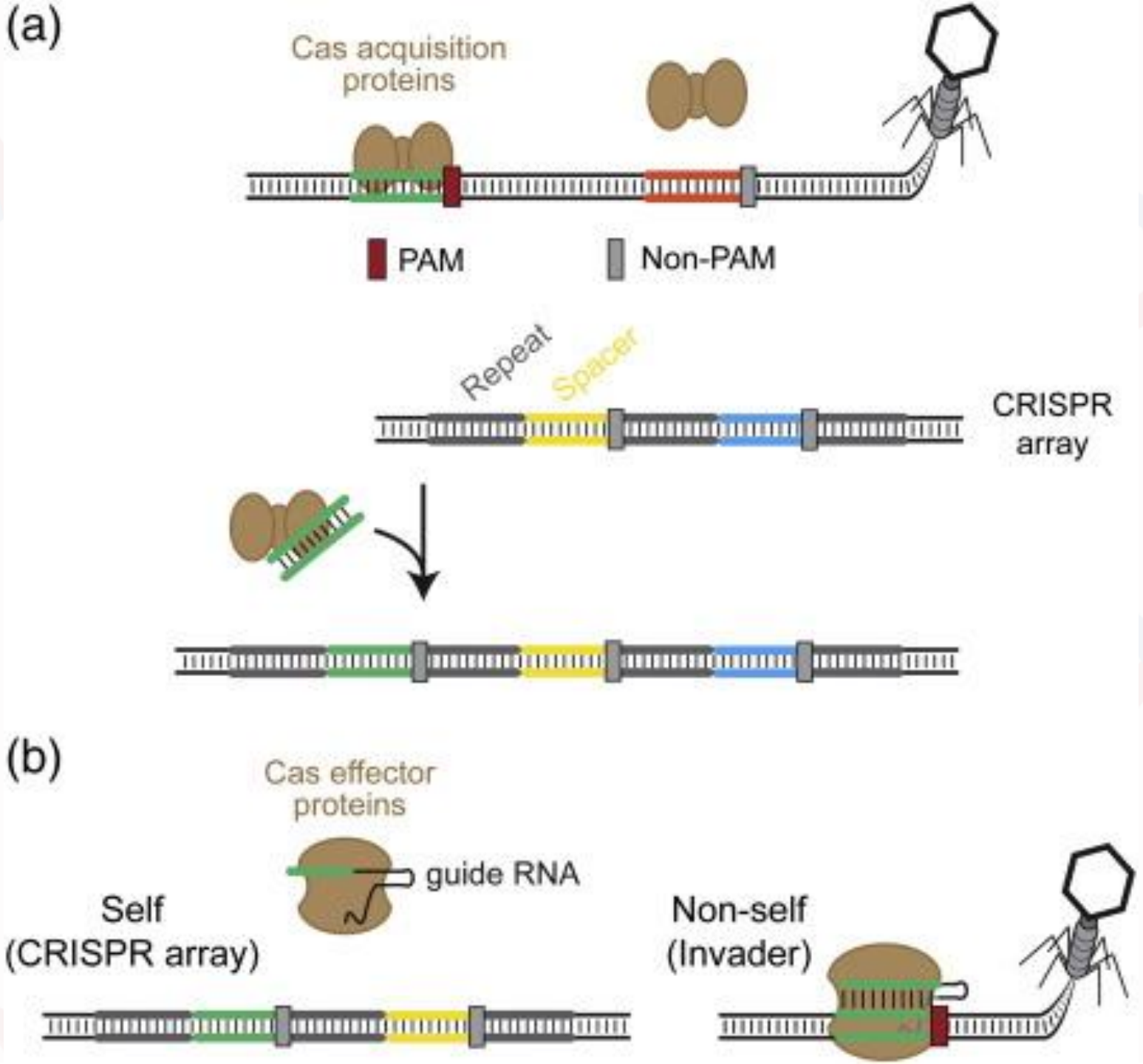
- The protospacer adjacent motif (or **PAM** for short)
- a short DNA sequence (usually 2-6 base pairs in length)
- follows the DNA region targeted for cleavage by the CRISPR system
- is required for a Cas nuclease to cut and is generally found 3-4 nucleotides downstream from the cut site

- The most commonly-used Cas9 from Streptococcus pyogenes recognizes the PAM sequence 5'-NGG-3' (where "N" can be any nucleotide base)

- there are many different Cas endonucleases from different bacterial species, and each recognizes a different PAM

- The very basis of bacteria evading their own endonucleases is the PAM sequence

- Thus, when researchers design a gRNA sequence, they generally exclude the PAM from the guide RNA


(a)
Cas acquisition proteins
PAM
Non-PAM
Repeat
Spacer
CRISPR array
(b)
Cas effector proteins
guide RNA
Self (CRISPR array)
Non-self (Invader)

- Based on the number and composition of proteins involved in nucleic acid interference, CRISPR-Cas systems are categorized into distinct classes 1–2 and types I–VI
- Class 2 systems encode a single effector protein and are further subdivided into types, II, V and VI
- Cas9 (type II) and Cas12 (type V) proteins
- have been shown to cleave invading double-stranded (ds)DNA, single-stranded (ss) DNA and ssRNA
- the size of Cas9 and Cas12 provides constraints on cellular delivery that may limit certain applications, including therapeutics

		# of subtypes	Cas endonuclease	Target	Requires tracrRNA?
Class 1	type I	7	cas3	DNA	no
	type III	4	cas10	DNA/RNA	no
	type IV*	1	--	--	--
Class 2	type II	3	cas9	DNA	yes
	type V	3	cas12	DNA	yes (1 subtype)
	type VI	3	cas13	RNA	no

*putative subtype

- Recently, small CRISPR-associated effector proteins (Cas12f) belonging to the type V-F subtype have been identified through the mining of sequence databases
- More recent classifications have grouped this family into

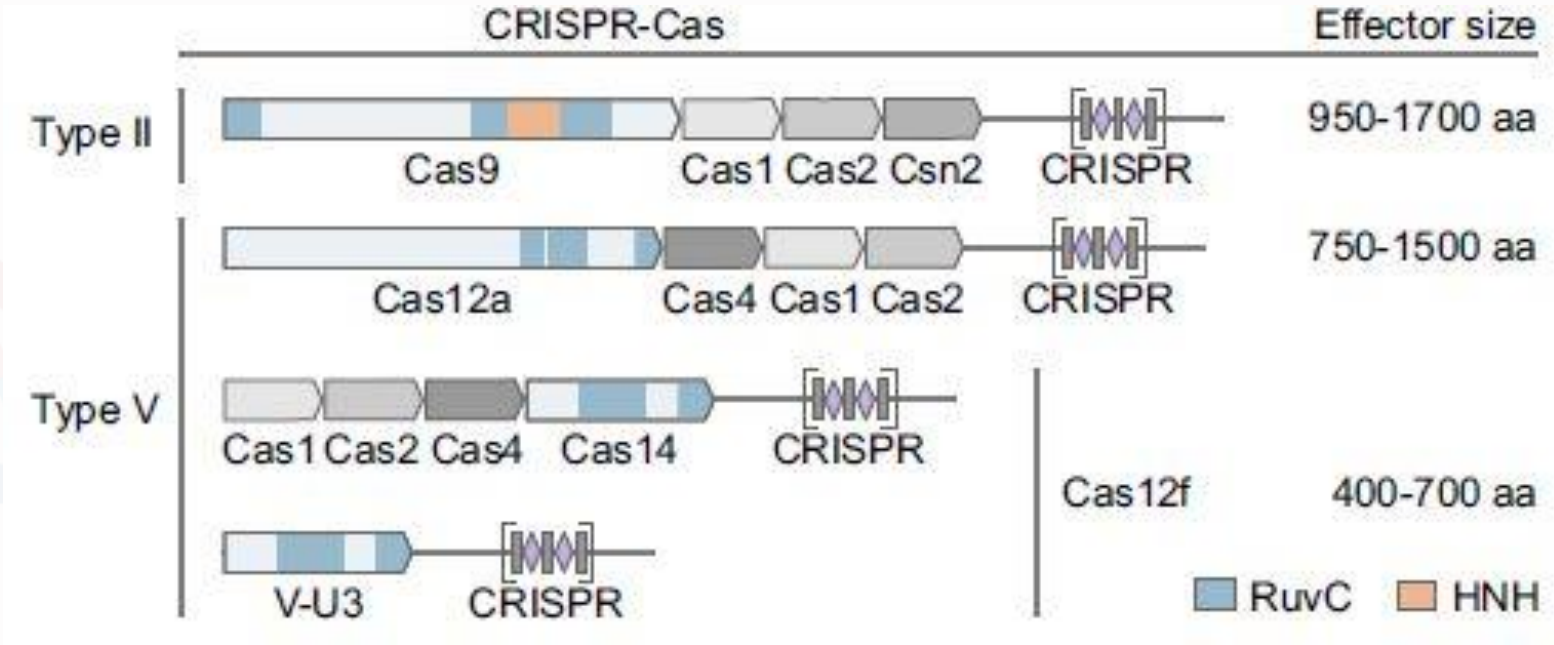
Cas12f1 (Cas14a and type V-U3),

Cas12f2 (Cas14b) and

Cas12f3 (Cas14c, type V-U2 and U4)

- The majority of Cas12f proteins (from V-U3 and Cas14 families)
- are nearly half the size of the

smallest Cas9 or Cas12 nucleases



- the N-terminal half ofCas12f proteins differs significantly in length accounting for most of the size difference between the two groups

- Due to this and their similarity to transposase associated TnpB proteins,

- hypothesized that they are remnants or intermediates of type V CRISPR–Cas system evolution and are incapable of forming the protein architecture required for dsDNA target recognition and cleavage

- functional characterization of Cas12f proteins are limited to a single protein identified from an uncultured archaeon (Un1), initially named Cas14a1 and renamed here to Un1Cas12f1, which was shown to exclusively target and cleave ssDNA in a PAM independent manner

- Here, using biochemical assays,reported that miniature Cas12f effectors (from Cas14 and type V-U3 families), like most other Cas12 proteins, are able to cleave dsDNA targets if a 5 PAM sequence is present in the vicinity of the guide RNA target

ABSTRACT

- more compact Versions of CRISPR-associated (Cas) nucleases would simplify delivery and extend application
- a collection of 10 exceptionally compact (422–603 amino acids) CRISPR–Cas12f nucleases that recognize and cleave dsDNA in a PAM dependent manner
- Categorized as class 2 type V-F
- previously identified Cas14 family and distantly related type V-U3 Cas proteins found in bacteria
- demonstrate that a 5 T- or C-rich PAM sequence triggers dsDNA target cleavage
- they can protect against invading dsDNA in Escherichia coli and find that some but not all can

MATERIALS AND METHODS

1.Engineering CRISPR-Cas12f systems to target a PAM library

- CRISPR-Cas12f systems were modified to target the 7 bp randomized PAM library
- by replacing the native CRISPR array with three repeat:spacer:repeat units
- a spacer capable of complementing to a sequence (anti-sense strand) immediately 3 of the region of PAM randomization
- (33–39 nt depending on the average spacer length observed
- in the respective Cas12f system)
- then synthesized (GenScript) and cloned into a modified pET-duet1 (MilliporeSigma) or pACYC184 (NEB) plasmid
- For the CRISPR-Cas12f1 system [initially named Cas14a1 and renamed here as

Un1Cas12f1] the pLBH531 MBP-Cas14a1 plasmid was used

2.RNA synthesis

- Un1Cas12f1 (Cas14a1) single guide RNAs (sgRNA) were produced by in vitro transcription using TranscriptAid T7 High Yield Transcription Kit

- and purified using GeneJET RNA Purification Kit

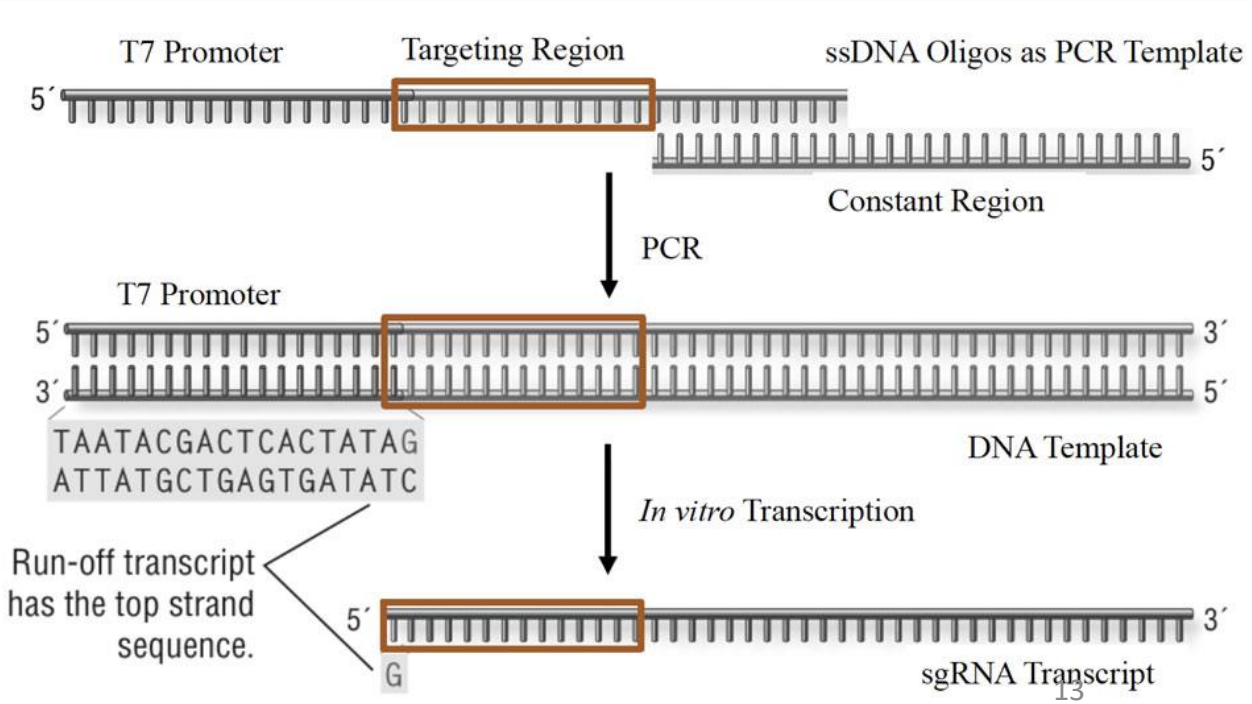
- Templates were generated by PCR using

overlapping oligonucleotides,

altogether, containing a T7 promoter

at the proximal end

followed by the sgRNA sequence



3.Detecting Cas12f dsDNA cleavage and PAM recognition

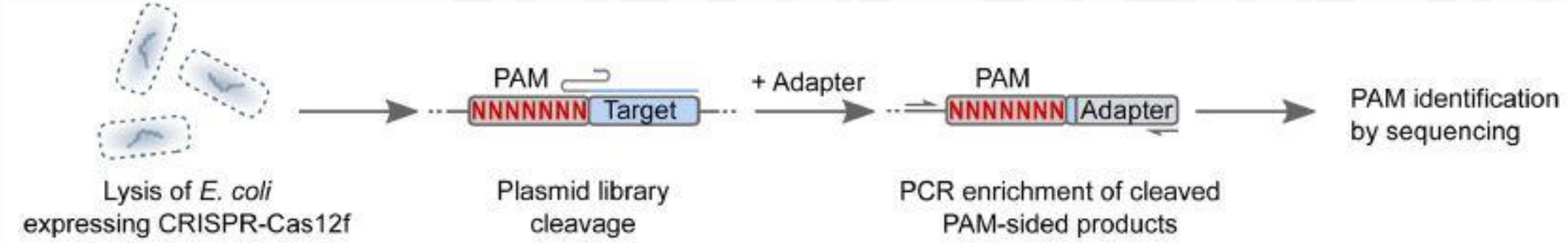
- Plasmid DNA targets were cleaved with Cas12f ribonucleo protein (RNP) complexes produced from the modified locus or by combining Escherichia coli lysate containing Un1Cas12f1 (Cas14a1) protein with T7 transcribed sgRNA (20 nt spacer)

- First, E. coli DH5a or ArcticExpress (DE3) cells were transformed with CRISPR-Cas12f encoding plasmids

- the obtained **supernatant** containing RNPs was used **directly** in the digestion experiments

- For Un1Cas12f1, 20 micro l of clarified supernatant was combined with 1 micro l of RiboLock RNase Inhibitor (Thermo Fisher Scientific) and 2 micro g of sgRNA and allowed to complex with the clarified lysate

- Cas12f RNP complexes were used to cleave either the 7 bp randomized PAM library or a plasmid containing a fixed PAM and gRNA target
- DNA ends were repaired by adding T4 DNA polymerase
- 3’-dA overhangs added by incubating the reaction mixture with DreamTaq polymerase
- the end repaired cleavage products were ligated with a double-stranded DNA adapter containing a 3’-dT overhang using T4 DNA ligase
- products were PCR amplified appending sequences required for deep sequencing
- target cleavage was evaluated by examining the unique junction generated by target cleavage and adapter ligation in deep sequence reads
- 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 sequence was then tabulated and compared to negative controls (experiments setup without functional Cas12f complexes) to identify target cleavage.

- Evidence of PAM recognition was evaluated

- 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

- 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 >2.5-fold from the negative control

4.Expression and purification of an Un1Cas12f1 (Cas14a1) protein

- Un1Cas12f1 (Cas14a1) protein was expressed in E. coli BL21(DE3) strain from the pLBH531 MBP-Cas14a1 plasmid
- The **supernatant** was loaded on the **Ni2+-charged HiTrap chelating HP column** and **eluted** with a **linear gradient of increasing imidazole** concentration
- The **fractions** containing Un1Cas12f1 variants were pooled and subsequently loaded on **HiTrap heparin HP column** for **elution using a linear gradient of increasing NaCl** concentration
- The **fractions** containing the **protein of interest** were pooled and the **10×His-MBP-tag was cleaved** by overnight **incubation with TEV protease**
- the **elution** from the **HiTrap heparin colum**n was **loaded on a MBPTrap column** and the Un1Cas12f1 proteins were collected in the **flow-through**
- The **collected fractions** with Un1Cas12f1 were then **dialyzed**
- **stored at –20◦C**

5.Un1Cas12f1 (Cas14a1)–sgRNA complex assembly for invitro DNA cleavage

- Un1Cas12f1 (Cas14a1) ribonucleoprotein (RNP) complexes were assembled by mixing Un1Cas12f1 protein with sgRNA at 1:1 molar ratio followed by incubation

6.DNA substrate generation

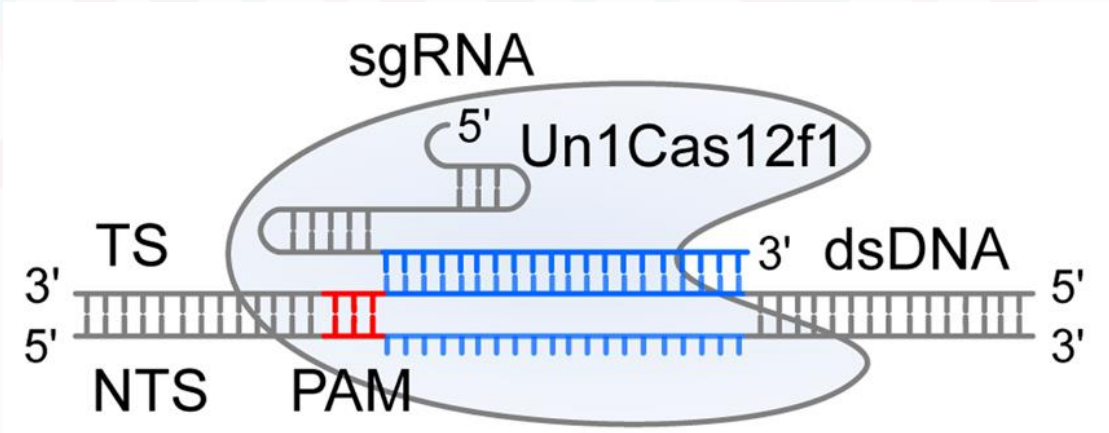
- Plasmid DNA substrates were generated by cloning oligoduplexes assembled into pUC18 plasmid
- To generate radiolabeled DNA substrates, the 5’-ends of oligonucleotides were radiolabeled using T4 PNK and [gama-33P]ATP

7.DNA substrate cleavage assay

- initiated by mixing plasmid DNA with Un1Cas12f1 (Cas14a1) RNP complex

- The final reaction mixture typically contained plasmid DNA, Un1Cas12f1 RNP complex in Tris–HCl, NaCl, DTT and MgCl2 reaction buffer

- Aliquots were removed at timed intervals and mixed with 3× loading dye solution (0.01% Bromophenol Blue and EDTA in 50% (v/v) glycerol) and reaction products were analyzed by agarose gel electrophoresis and ethidium bromide staining.



8.M13 cleavage assay

- The reaction mixture contained M13 ssDNA, ssDNA or dsDNA activator and Un1Cas12f1 RNP
- The reaction was initiated by addition of Un1Cas12f1 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

9.Plasmid interference assay

- were performed in E. coli Arctic Express (DE3) strain bearing Cas12f systems (plasmids encoding CRISPR-Cas12f systems)

- For Un1Cas12f1 (Cas14a1), E. coli BL21 (DE3) strain was transformed with pGB53 plasmid, which was engineered from the pLBH545 Tet-Cas14a1 Locus plasmid by removing tracrRNA and CRISPR array and adding sgRNA encoding sequence with T7 promoter, HDV ribozyme and terminator sequences

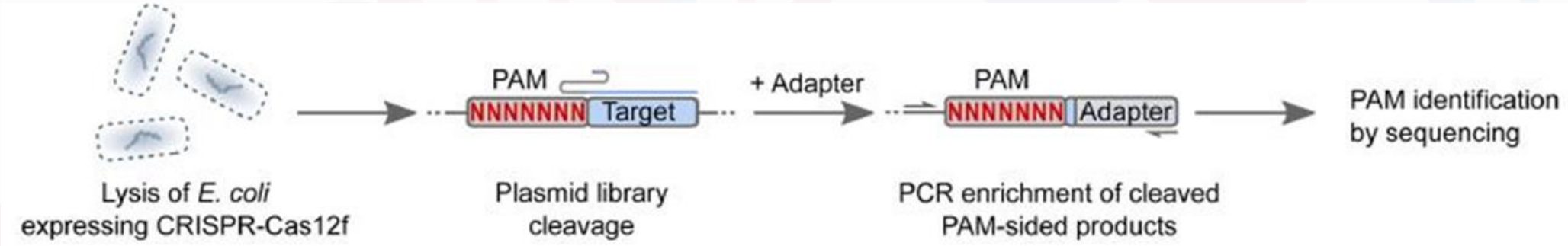
- The cells were grown and electroporated with 100 ng of low copy number pSC101 target plasmids

- The co-transformed cells were further diluted by serial 10× fold dilutions and grown at 37◦C for 16–20 h on plates containing inductor and antibiotics

RESULTS

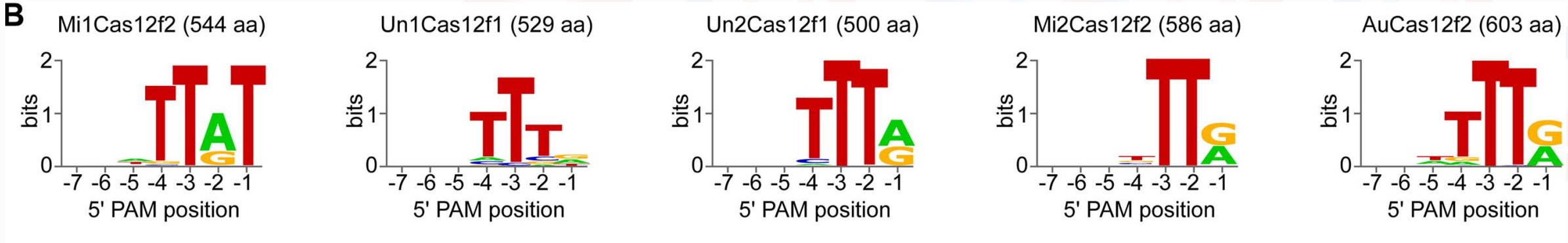
1.DNA targeting requirements by Cas12f proteins

- To our knowledge, just a single Cas12f protein, Un1Cas12f1 (Cas14a1), it was shown to only cleave ssDNA targets in a PAM independent manner
- To establish DNA cleavage requirements, we initially tested a Cas12f2 protein from Micrarchaeota archaeon (Mi1) previously named Cas14b4
- A slight increase (2.5-fold) in the number of adapter ligated sequences was recovered after the 21st protospacer position 3’ of the randomized PAM
- Analysis of these fragments showed the recovery of a T-rich sequence (5’-TTAT-3’) immediately 5’ of the gRNA target only in the Mi1Cas12f2 treated sample



- To confirm the PAM sequence and the dsDNA cleavage position, a plasmid was constructed containing a target adjacent to the identified 5’- TTAT-3’ PAM sequence and subjected to cell lysate cleavage experiments.
- To increase Mi1Cas12f2 concentration in the cell lysate, a higher copy number DNA expression plasmid equipped with an inducible T7 promoter was also utilized Sequencing of the target plasmid cleavage products confirmed cleavage at the 21st position

- Four additional Cas12f proteins consisted of **two Cas12f1** proteins (**Un1** and **Un2**) and **two Cas12f2** proteins from Micrarchaeota archaeon (**Mi2**) and Aureabacteria bacterium (**Au**)
- Similarly to Mi1Cas12f2, a cleavage signal distal to the PAM region was detected for all proteins
- For **Cas12f2** nucleases, this occurred just after the **21st** protospacer position (a second cleavage site was also observed after the 23rd position)
- While for **Cas12f1** proteins the signal was found after the **24th** position
- Analysis of the sequences supporting cleavage yielded 5’ T rich PAM recognition like that recovered for Mi1Cas12f2



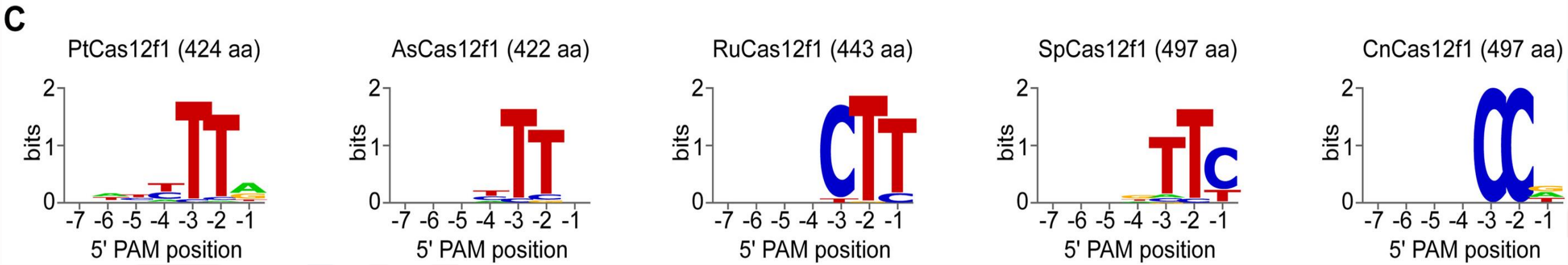
to further explore the **DNA cleavage requirements** of **miniature CRISPR-Cas** effectors, we **sought to evaluate** the **dsDNA cleavage activity** of **proteins from** an **uncharacterized** putative group of **class 2** CRISPR–Cas systems, **type V-U3**, **which lack** the proteins involved in **adaptation**

- **First**, PSI-BLAST searches were performed to identify a group of CRISPR-associated proteins primarily from bacteria, in particular, lineages of Clostridia and Bacilli, that belonged to the type V-U3 family
- They contained a conserved **C-terminal** tri-split **RuvC domain** similar to other Cas12 nucleases and a **short variable N-terminal** sequence as observed for the Cas12f proteins from Cas14 family
- Next, five members were selected for functional characterization
- they were chosen to represent the uncovered diversity and ranged in size between 422–497 amino acids

- all produced cleavage around the 24th position 3’ of the PAM region

- For nucleases from Parageobacillus thermoglucosidasius (Pt) and Acidibacillus sulfuroxidans (As), secondary cleavage signals (>5% of all reads) were also recovered either before or after the 24th position

- Like Cas12f proteins from Cas14 family, type V-U3 nucleases cleaved the dsDNA library in a 5’ PAM-dependent manner and altogether expanded miniature Cas nucleases PAM diversity to encompass not only Trich but also C-rich motifs



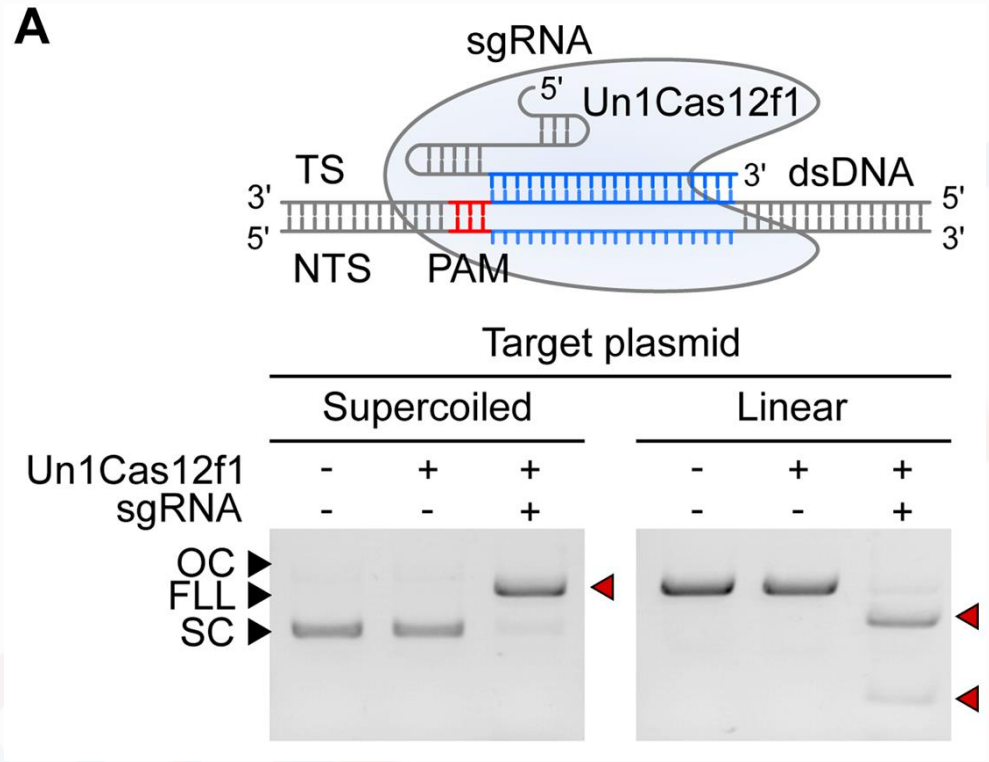
2.Biochemical characterization of Un1Cas12f1 (Cas14a1) mediated dsDNA cleavage

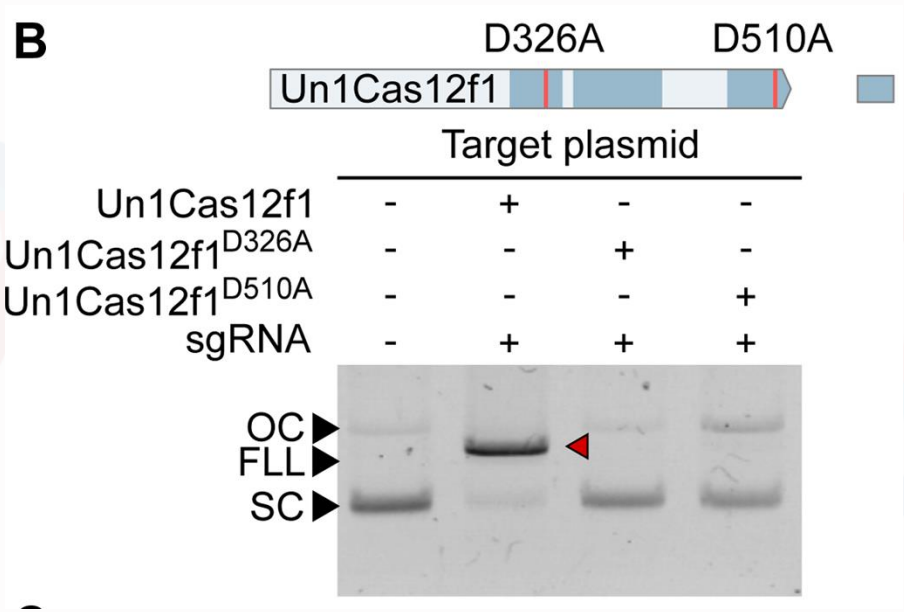
- Since a sgRNA solution was available for Un1Cas12f1 (Cas14a1)
- First, evaluated Un1Cas12f1 dsDNA cleavage activity to decipher optimal reaction conditions
- Experiments revealed that Un1Cas12f1 ribonucleoprotein (RNP) complex is a Mg2+-dependent endonuclease that functions best in low salt concentrations (5–25 mM) and is active in a wide temperature range ∼37–50◦C, with a temperature optimum of ∼46◦C
- sgRNA spacers of around 20 nt supported the most robust dsDNA cleavage activity

- Under optimized reaction conditions, supercoiled (SC) plasmid DNA containing a target sequence flanked by an Un1Cas12f1 PAM (5’-TTTA-3’) was completely converted to a linear form (FLL) indicating the formation of a dsDNA break

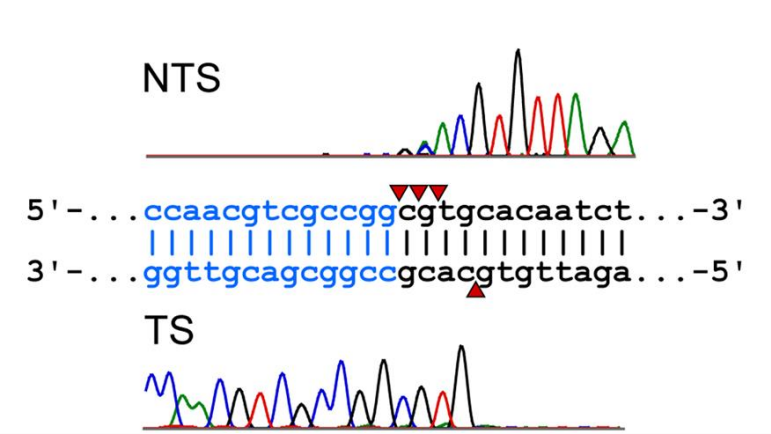
- Additionally,cleavage of linear DNA yielded DNA fragments of expected size, further validating Un1Cas12f1 mediated dsDNA break formation

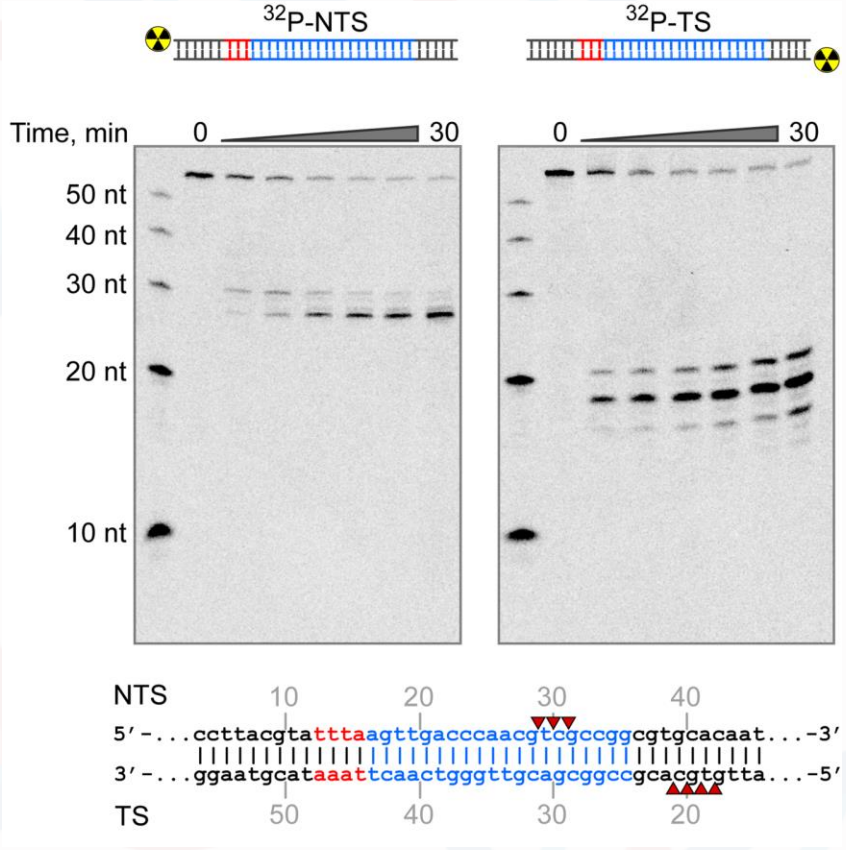
- Finally, alanine substitution of conserved RuvC active site residues abolished cutting activity, confirming that the RuvC domain is essential for the observed dsDNA cleavage activity





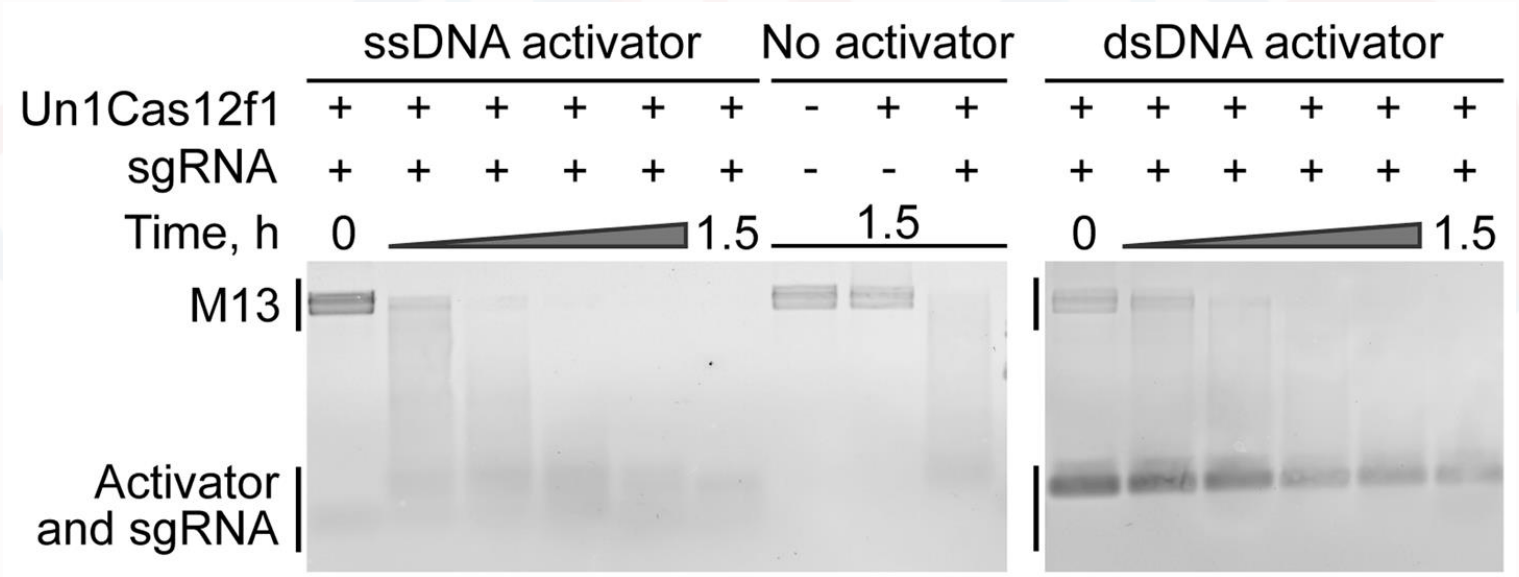
- Using run-off sequencing, we observed that Un1Cas12f1 makes 5’ staggered overhanging DNA cut sites
- Cleavage predominantly occurred centered around positions 20–24 bp in respect to PAM sequence and was independent of spacer length





- The cleavage pattern of Un1Cas12f1 was
- also assessed on synthetic double stranded
- oligodeoxynucleotides.
- As illustrated, a 5’ staggered cut pattern, albeit with less strictly defined cleavage positions than observed with larger DNA fragments was seen.

- Next we investigated if the non-specific ssDNA degradation activity of Un1Cas12f1 could be induced not just by ssDNA targets but also by dsDNA targets
- First, the ability of Un1Cas12f1 to indiscriminately degrade single stranded M13 DNA in the presence of a ssDNA target without a PAM was confirmed
- Then, a dsDNA target containing a 5’ PAM and sgRNA target for Un1Cas12f1 was also tested for its ability to trigger non-selective ssDNA degradation
- the trans-acting ss-DNase activity of Un1Cas12f1 was activated by both ss-DNA and dsDNA targets, similar to observations made for Cas12a
- Additionally, in the absence of a target, the Un1Cas12f1 RNPcomplex non-selectively degraded single stranded M13 DNA



3.Cas12f mediated plasmid DNA interference in E. coli

- next tested if CRISPR-Cas12f systems (from Cas14 and type V-U3 protein effector families) can be programmed to target and cleave invading dsDNA in a heterologous E.coli host

- First, an E. coli plasmid DNA interference assay was adopted using a minimal Cas12f CRISPR locus modified to target the incoming low copy number plasmid DNA.

- To assess transformation efficiency, each experiment was serially diluted by 10×and compared with controls (experiments performed with a plasmid that does not contain a target site)

- none of the Cas12f nucleases from Cas14 protein family interfered with plasmid DNA transformation as evidenced by similar recovery of resistant colonies compared to controls

- In contrast, type V-U3 effectors from A. sulfuroxidans (As) and Syntrophomonas palmitatica (Sp) both produced notable levels of plasmid interference and slight interference was also observable for nucleases from P. thermoglucosidasius (Pt) and Ruminococcus. (Ru)