Cas4-Cas1 fusions drive efficient PAM selection and control CRISPR adaptation

Almendros, Cristóbal; Nobrega, Franklin L.; McKenzie, Rebecca E.; Brouns, Stan J.J.

DOI
10.1093/nar/gkz217

Publication date
2019

Document Version
Final published version

Published in
Nucleic Acids Research

Published online 2 April 2019

Nucleic Acids Research, 2019, Vol. 47, No. 10 ***5223–5230***
doi: 10.1093/nar/gkz217

Cas4–Cas1 fusions drive efficient PAM selection and control CRISPR adaptation

Cristóbal Almendros[1], Franklin L. Nobrega[1], Rebecca E. McKenzie[1] and Stan J.J. Brouns[1,2,*]

[1]Kavli Institute of Nanoscience, Department of Bionanoscience, Delft University of Technology, Van der Maasweg 9, 2629 HZ Delft, Netherlands and [2]Laboratory of Microbiology, Wageningen University, Stippeneng 4, 6708 WE Wageningen, Netherlands

Received January 29, 2019; Revised March 15, 2019; Editorial Decision March 18, 2019; Accepted March 28, 2019

ABSTRACT

Microbes have the unique ability to acquire immunological memories from mobile genetic invaders to protect themselves from predation. To confer CRISPR resistance, new spacers need to be compatible with a targeting requirement in the invader's DNA called the protospacer adjacent motif (PAM). Many CRISPR systems encode Cas4 proteins to ensure new spacers are integrated that meet this targeting prerequisite. Here we report that a gene fusion between *cas4* and *cas1* from the *Geobacter sulfurreducens* I-U CRISPR–Cas system is capable of introducing functional spacers carrying interference proficient TTN PAM sequences at much higher frequencies than unfused Cas4 adaptation modules. Mutations of Cas4-domain catalytic residues resulted in dramatically decreased naïve and primed spacer acquisition, and a loss of PAM selectivity showing that the Cas4 domain controls Cas1 activity. We propose the fusion gene evolved to drive the acquisition of only PAM-compatible spacers to optimize CRISPR interference.

INTRODUCTION

To counteract the constant threat posed by invading genetic elements, bacterial and archaeal cells developed both innate and adaptive immune systems. While many innate systems have been described (1), so far only one adaptive immune system in prokaryotes has been identified. This system is composed of clustered regularly interspaced short palindromic repeats (CRISPR) and their associated proteins (Cas) (2–4), and can be classified into two major classes, six types and 33 subtypes (5). All DNA targeting CRISPR–Cas systems share a similar mechanism consisting of genomic acquisition of CRISPR memories, formation of guide RNAs from the CRISPR memory bank, and interference with a target sequence (6). Naïve spacer acquisition is the process in which Cas1 and Cas2 proteins integrate a short piece of foreign DNA (prespacer) into the CRISPR array generating a new spacer (7). A sequence next to the CRISPR array, the *leader*, is recognized by Cas1–2 to initiate spacer integration at the first repeat of the CRISPR array (8–10). During the expression step, the new spacer is transcribed generating a new CRISPR RNA (crRNA), which is loaded into a crRNA-effector complex and guides the Cas proteins to the target sequence (11). Finally, the crRNA-effector complex will find and bind the target sequence flanked by the protospacer adjacent motif (PAM) (12), and cleavage the invader DNA sometime with help of accessory nucleases such as Cas3 (7). The products of this degradation can be used by Cas1–2 to efficiently generate new spacers from the same target in a process called primed spacer acquisition (13,14), increasing the diversity of the CRISPR repertoire within the population (7).

Although only Cas1–2 is strictly conserved in all CRISPR systems (5), other genes are often associated with Cas1–2, including reverse transcriptases, *csn2* and *cas4,* implying that the spacer adaptation mechanism varies among systems (5). Recently, the importance of Cas4 protein in spacer acquisition in different CRISPR systems has been reported. In the I-D system Cas4 is not strictly required for acquisition, but it plays a crucial role in the selection of functional PAM sequences and prespacer processing (15). A similar role was observed for two Cas4 proteins that act in coordination in the I-A system of *Pyrococcus furiosus* (16). Moreover, 3′ PAM processing of prespacers and a strong interaction between Cas4 and Cas1 was described in the I-C system of *Bacillus halodurans* (17), suggesting Cas4 as part of the CRISPR adaptation complex.

A fusion between *cas4* and *cas1* genes is found in both Class I (I-U and I-B) and Class II (V-B) CRISPR–Cas systems (5,18) suggesting a strong functional link between the activities of Cas4 and Cas1 proteins. Here, we have selected a fused Cas4/1 protein from the uncharacterized type I-U CRISPR–Cas system and found that the Cas4/1 fu-

*To whom correspondence should be addressed. Tel: +31 15 278 3920; Email: stanbrouns@gmail.com

sion provides a high frequency of PAM-compatible, functional spacers. Mutations in the Cas4 domain drastically decreased spacer acquisition rates and disrupted the PAM selection process. We observed that Cas4-domain activity is not only required during naïve spacer acquisition, but also during primed spacer acquisition.

The Cas4/1 fusion is an example of co-evolutionary refinement of the CRISPR adaptation process to adjust to the PAM requirements of CRISPR interference. A highly efficient PAM selection process reduces the acquisition of non-functional spacers, providing a better chance of survival in hostile environments rich in mobile genetic elements.

MATERIALS AND METHODS

Bacterial strains and growth conditions

Escherichia coli strains DH5 and BL21-AI were grown at 37°C in LB media with shaking or on LB agar (LBA) plates containing 1.5% (w/v) agar. When required, media was supplemented with 50 μg/ml spectinomycin, 25 μg/ml chloramphenicol, 100 μg/ml ampicillin, 1 mM IPTG and 0.2% (w/v) L-arabinose (see Supplementary Table S1 for plasmids and their corresponding selection markers).

Plasmid construction and transformation

Plasmids used in this work are indicated in Supplementary Table S1. All cloning steps were performed in *E. coli* DH5a. Primers described in Supplementary Table S2 were used for PCR amplification of the type I-U CRISPR–Cas acquisition module from *G. sulfurreducens* DSMZ 12127 using the Q5 High-Fidelity Polymerase (New England Biolabs). We made two constructs that harbor the acquisition module: pCas4/1–2LR (*cas4/1*, *cas2*, and *leader*-repeat) for naïve acquisition and pCas4/1–2LRSR (*cas4/1*, *cas2*, and *leader*-repeat-spacer1-repeat) for priming assays. For priming and interference assays we made a construct with the interference complex: *cas3*, *cas8u2*, *cas7* and *cas5*/6 (pCas3-8-7-5/6). PCR amplicons were subsequently cloned into His6x-SUMO tag p13S-S ligation-independent cloning (LIC) vector (http://qb3.berkeley.edu/macrolab/addgene-plasmids/) for the acquisition modules and restriction and ligation cloning for the interference complex. The different mutants and derived plasmids were obtained by PCR-mutagenesis using primers listed in Supplementary Table S2. The *cas4/1* fission construct was constructed by introducing a stop codon at the end of the *cas4* gene, and a new ribosome binding site and start codon directly upstream of the *cas1* gene. The interference complex was cloned into the pACYCDuet-1 vector system (Novagen, EMD Millipore) using conventional restriction-ligation cloning. Target plasmids (pTarget) used in the interference study and in priming assays were obtained by PCR- mutagenesis with the protospacer sequence included in one of the primers (Supplementary Table S2). All plasmids were verified by Sanger sequencing (Macrogen Europe, The Netherlands). Bacterial transformations were either carried out by electroporation (200 Ω, 25 μF, 2.5 kV) using an ECM 630 electroporator (BTX Harvard Apparatus) or using chemically competent cells following the manufacturer's manual (Mix&Go, Zymo research). Transformants were selected on LBA supplemented with appropriate antibiotics.

Spacer acquisition assays

For naïve adaptation assays, *E. coli* BL21-AI strain was transformed with pCas4/1–2LR and mutant derivatives. For priming assays, the same strains were co-transformed with three plasmids: pCas4/1–2LRSR, pCas3-8-7-5/6 and $pTarget_{+1}$. Three colonies of transformants were grown independently in 5 ml of LB supplemented with the appropriate antibiotics at 37°C with shaking. Any assays involving $pTarget_{+1}$ and pNon-Target were performed in the absence of antibiotic selection for these plasmids. After 2.5 h of growth, *cas* genes were induced with IPTG and L-arabinose, and the cultures incubated for additional 24 h. Detection of acquisition and isolation of expanded CRISPR arrays from the cultures were performed with a sensitive, two-step PCR method as described previously (15,19). The first round of PCR uses a mix of degenerate primers with three different 3′ nucleotides, stimulating the amplification of arrays with new spacers (15). Then, the hypothetical expanded CRISPR DNA band was purified with an automated size selection and submitted to a second round of PCR using the degenerated primers and an internal reverse primer (15,19).

Interference assays

Plasmid loss assays were performed in BL21-AI. Three colonies of transformants carrying plasmids pTarget, pCas3-8-7-5/6 and pLRSR (derivated from pCas4/1–2LRSR lacking *cas4/1* and *cas2*) were independently grown in LB at 37°C with shaking. After 2.5 h of growth, *cas* genes were induced with IPTG and L-arabinose, and the cultures were incubated for an additional 24 h. A dilution for each replicate was plated on LBA containing the appropriate antibiotics (without the marker of pTarget). Fifty colonies from each replicate were inoculated onto LBA supplemented with and without the pTarget marker (ampicillin). Only colonies that retained the plasmid will grow with ampicillin. The proportion of ampicillin-resistant colonies after incubation at 37°C for 24 h was calculated with respect to the total number of colonies that grew without ampicillin.

Expanded CRISPR arrays sequencing

PCR amplicons corresponding to the expanded CRISPR arrays were purified using GeneJET PCR Purification kit (Thermo Fisher Scientific). The concentration of DNA in each sample was measured using Invitrogen Qubit fluorometric quantification. Samples were prepared for sequencing with the NEBNext Ultra II DNA Library Prep Kit for Illumina and each library individually barcoded with the NEBNext Multiplex Oligos for Illumina (Index Primers Set1 and Set2). Prepped samples size and concentration were assessed with the Agilent 2200 TapeStation D100 high sensitivity kit so that samples could be pooled with equal molarity. Combined samples were then denatured and diluted as recommended by Illumina and spiked with 15% of the PhiX174 control DNA (Illumina) to artificially increase

the genetic diversity. 600 μl were loaded into the cassette for sequencing on a Nano flowcell (2 × 250 base paired-end) with an Illumina MiSeq. Image analysis, base calling, demultiplexing, and data quality assessments were performed on the MiSeq instrument. FASTAQ files generated by the MiSeq were analysed by pairing and merging the reads using Geneious 9.0.5. Acquired spacers were extracted and analysed as described previously (15).

RESULTS

Cas4 domain boosts naïve acquisition

The *cas4* gene is widely distributed in both Class I and Class II CRISPR–Cas systems, but it is found fused to *cas1* only in three CRISPR subtypes (I-B, I-U and V-B; Figure 1A) (18,20,21). Here, we studied the role of the Cas4/1 fusion protein of the I-U CRISPR–Cas system of *G. sulfurreducens* PCA in naïve spacer acquisition. Genes *cas4/1* and *cas2* from this system, together with a minimal CRISPR array with the full leader (205 bp) and one repeat, were cloned into a vector under T7 promoter (pCas412LR). *E. coli* BL21-AI was transformed with this plasmid to test the ability of the cells to acquire new spacers. Using a sensitive PCR (15), spacer acquisition was detected after 24 hours of induction in the first round of PCR (Figure 1B), demonstrating that the I-U acquisition module does not require any specific host factor from *G. sulfurreducens*. To test the requirement of Cas2 (*Δcas2* mutant) and the importance of the physical link between the Cas4 and Cas1 domains, acquisition assays were performed using different mutant strains. Both the *Δcas2* mutant and the *cas4/1* fission displayed a sharp drop in adaptation frequencies, indicating that both Cas2 and the translationally fused architecture of Cas4 and Cas1 are required for spacer acquisition (Figure 1B). To better understand the importance of each Cas4/1 domain in the integration of new spacers, we made three mutants: one in the Cas1 (metal coordinating residue; E380A) (22) and two in the RecB domain of Cas4 (K102A and D87A) (15,23). As expected, mutation E380A in Cas1 abolished spacer acquisition activity, while mutations in the active site of Cas4 drastically reduced spacer acquisition rates (Figure 1B). After the second round of PCR (see Materials and Methods), spacer acquisition was not detected in the *Δcas2*, *cas4/1* fission and the Cas1 E380A mutants (Figure 1C). Interestingly however, acquisition was detected using Cas4 domain mutants (Figure 1C), suggesting that in the systems with a fused Cas4 and Cas1 architecture, Cas4 activity controls the rate of spacer acquisition.

Cas4/1 selects TTN spacers from non-chromosomal origin

To determine the source of the new spacers and their PAM we subjected expanded CRISPR array amplicons to deep sequencing. Analysis of newly integrated spacers revealed a predominant spacer length of 36 bp (Figure 2A) resembling the size distribution of the *G. sulfurreducens* spacers (Supplementary Figure S1). Next, we mapped new spacer sequences to the *E. coli* genome and the pCas4/1–2LR plasmid. Approximately 70% of (unique) spacers match the plasmid (Figure 2B), indicating a preference of the acquisition module towards the incorporation of new spacers from plasmid sources. We found that some highly abundant spacers matched regions common to both plasmid and genome (e.g. the *lacI* gene). Mapping of all non-*lacI* spacers onto the genome showed enrichment of new spacers at the origin of DNA replication (*oriC*), and in the region between *terA* and *terC* DNA replication termination sites (Supplementary Figure S2). This suggests that the I-U acquisition module, similar to I-E CRISPR–Cas systems, may use DNA repair fragments that are produced during replication fork stalling as prespacers substrates (24,25). Furthermore, we observed no strand bias for the new spacers (Supplementary Table S3).

To determine the PAM of the acquired spacers, upstream flanking regions of protospacers were extracted and analysed for sequence motifs. We found that 50% of the new unique spacers and >80% of all spacers, matched targets with upstream TTN motifs (Figure 2C), indicating this sequence is highly selected during spacer acquisition and could constitute the PAM in CRISPR interference in I-U CRISPR systems. In summary, the I-U spacer acquisition module is directed to plasmids and is highly efficient in selecting TTN spacers, even during naïve spacer acquisition.

TTN PAMs licence type I-U CRISPR interference

To confirm the TTN PAM provides interference, the full I-U CRISPR–Cas system was cloned into two plasmids under control of the T7 promoter and transformed into BL21-AI cells. pCas3-8-7-5/6 contained copies of *cas3*, *cas8*, *cas7* and *cas5/6*, while pLRSR encoded a minimal CRISPR array with the full leader and a spacer flanked by two repeats (Figure 1A). To test CRISPR interference activity, plasmid loss assays were performed using a third plasmid containing the protospacer (pTarget). Different variants of pTarget were used: a non-target (NT) without protospacer, a target with TTN PAM, and a CRISPR repeat PAM (AGC). The assays were performed in the absence of *cas4/1-cas2* to eliminate possible interference as a consequence of new spacer integration. As expected, plasmid loss was observed only for bacteria containing target plasmids with TTN PAMs (TTA, TTT, TTC and TTG) (Figure 3), demonstrating that the TTN PAM enrichment observed in naïve acquisition assays drives efficient CRISPR interference and target plasmid clearance.

The Cas4 domain processes the PAM sequence

To understand the role of the Cas4 domain in the adaptation process, expanded arrays obtained from the acquisition assays using Cas4 RecB catalytic site D87A and K102A mutants were subjected to deep sequencing (Figure 1C). We found that TTN PAM selection was not retained in any of the two Cas4-domain mutants (Figure 4A), indicating the requirement of Cas4 catalytic activity for PAM processing. Interestingly, in contrast to previous studies (15,16) no differences were found in terms of the origin of new spacers and spacer length, suggesting that target selection and spacer length determination are not dependent on Cas4 activity.

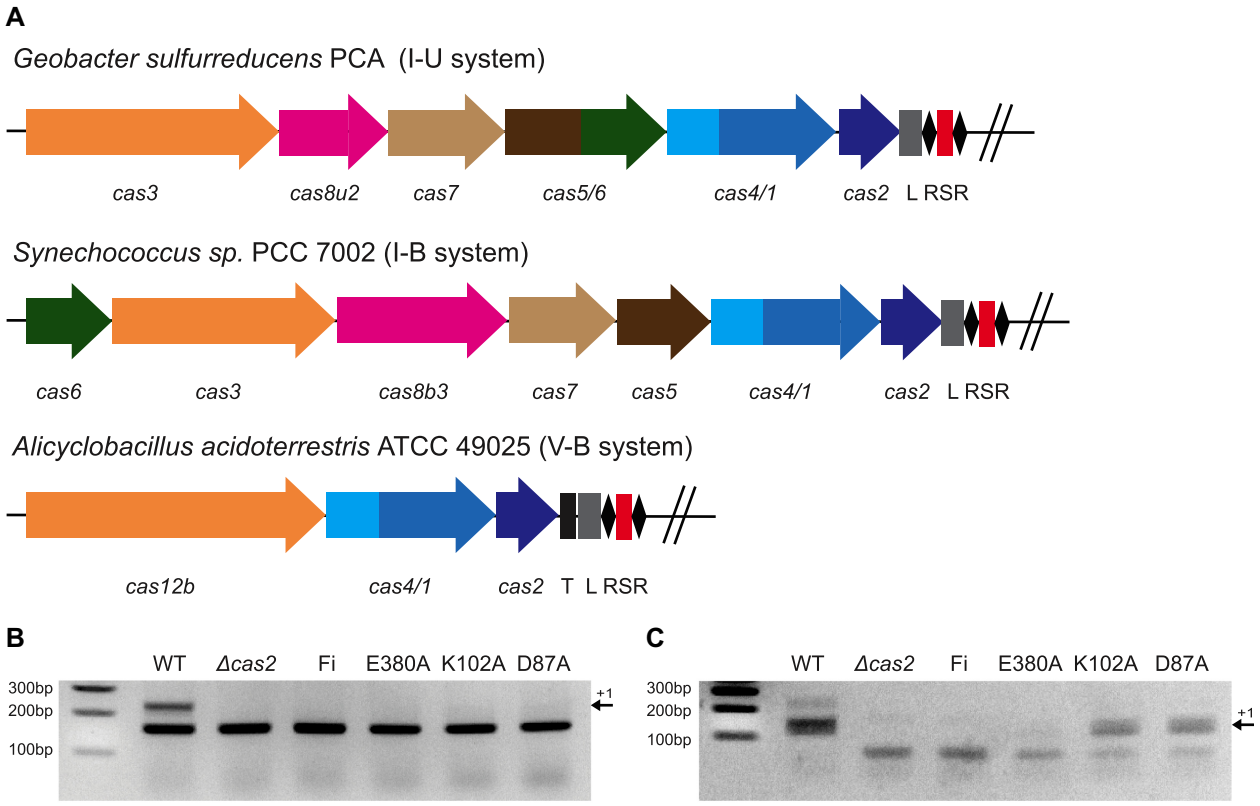

Figure 1. Naïve spacer acquisition in type I-U CRISPR–Cas systems. (**A**) Representation of CRISPR–Cas systems that contain the *cas4/1* fusion genes. Downstream of the *cas* genes is the *leader* sequence (L) in grey followed by the CRISPR array constituted by repeats (R) and spacers (S). In the V-B CRISPR–Cas system the sequence that encodes the tracrRNA (T) is between *cas2* and the *leader*. (**B**) Naïve spacer acquisition assays performed in *E. coli* BL21-AI overexpressing the CRISPR adaptation module (*cas4/1*, *cas2*, *leader* and repeat) of the I-U CRISPR–Cas system (WT). Moreover, spacer acquisition assays were performed using Δ*cas2*, fission of the Cas4 and Cas1 domains (Fi), Cas1 active site mutant (E380A) and Cas4 RecB mutants (D87A and K102A). The PCR products corresponding to the expanded CRISPR array is indicated with a +1 arrow. (**C**) Detection of spacer acquisition after the second round of PCR (see material and methods).

Protospacer mutations stimulate primed acquisition in type I-U

In many CRISPR–Cas systems, primed acquisition dramatically improves the incorporation of PAM-compatible spacers (7). To test if the efficiency of consensus PAM selection could be increased beyond the 50–80% observed during naïve acquisition, we performed priming assays using an experimental set up similar to the interference assays with a few modifications. First, the target plasmid used in the plasmid loss assay with the consensus TTT PAM, was mutated in the *seed* sequence (*i.e.* region of the protospacer flanking the PAM; A>T at position +1; Figure 5A). The *seed* mutation in pTarget$_{+1}$ compromises interference activity and promotes priming (13,26). Second, plasmid pCas4/1–2LRSR, containing *cas4/1*, *cas2* and a minimal array (with *leader*, two repeats and a spacer), was used instead of pLRSR, and combined with pCas3-8-7-5/6. Interestingly, a single mutation of the first position of the protospacer abolishes interference against pTarget$_{+1}$. Moreover, higher levels of spacer acquisition were observed after 24 h under priming conditions in comparison to the control cultures harboring a non-target plasmid (Supplementary Figure S3), suggesting that the presence of pTarget$_{+1}$ in the cell was stimulating the adaptation process. To confirm the heightened level of spacer acquisition was due to priming, expanded CRISPR arrays were subjected to deep sequencing. By mapping the new spacers we found that >50% of the unique spacers and nearly 80% of all new spacers targeted pTarget$_{+1}$ compared to 15% for the non-targeted control (Figure 5B), confirming that spacer acquisition was due to priming. Analysis of the new spacer distribution patterns showed that primed spacer acquisition in the I-U system occurs very evenly over the target plasmid, and does not lead to a strand bias, or location bias in close proximity to the primed protospacer (Figure 5C) (13,27–29).

We observed a high conservation of G (∼57%) and C (∼41%) in the first position of unique spacers as a consequence of using a selective PCR approach with degeneracy (G, C, T) at the 3′ position of the primer (19), in which the variants containing G or C in the 3′ end PCR amplify more efficiently. To correct for this bias, we plotted the density of the TTNS (i.e. TTNG and TTNC) sequences along the pTarget$_{+1}$ (Figure 5C). Rather than a strand or location bias, primed spacer acquisition appears to mostly follow the PAM density of the plasmid.

Next, we analysed the PAM sequence of all new spacers and again observed a TTN PAM in nearly 60% of the unique spacers and 90% of all spacers (Figure 5D). This

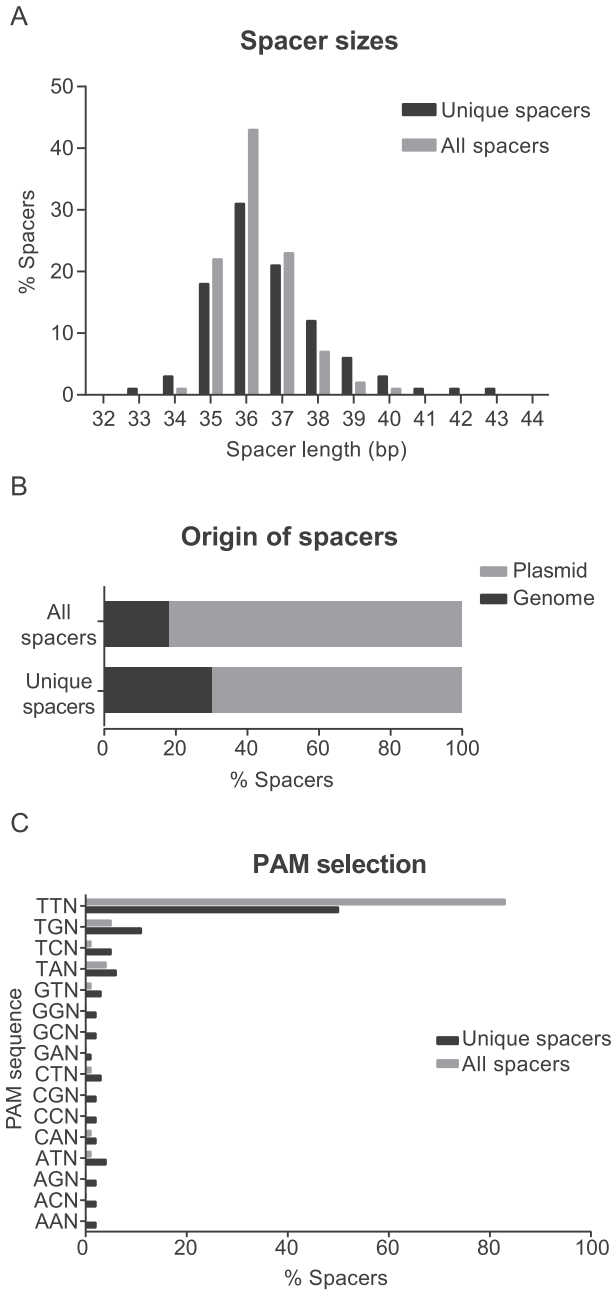

Figure 2. Analysis of newly integrated spacers. (**A**) Spacer size distribution of all and unique spacers. The most predominant size is 36 bp. (**B**) Origin of newly acquired spacers using the percentage of spacers that matches each replicon. (**C**) PAM selection of all and unique upstream protospacers sequences represented as percentage.

shows that acquisition of spacers with a correct PAM remains very high during priming in I-U.

To test the importance of the Cas4 domain in priming adaptation we performed assays using Cas4 K102A and D87A mutants. In both mutants, a strong reduction of acquisition was observed compared to the WT (Supplementary Figure S4), demonstrating that the activity of the Cas4 domain is important for both naïve and primed spacer acquisition events.

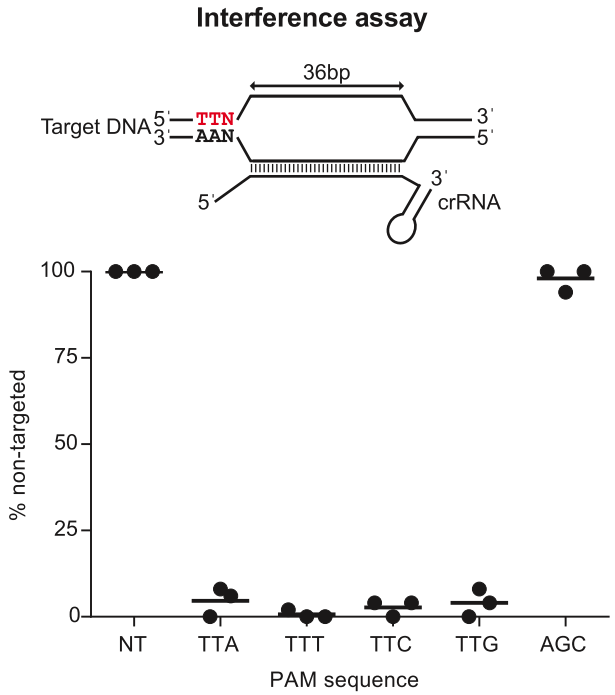

Figure 3. A TTN PAM drives CRISPR interference in type I-U. Plasmids that contain the protospacer with TTN PAM are eliminated from the cell. However, the plasmids with AGC PAM and without protospacer (non-target, NT) do not license interference. Data are represented by dots ($n = 3$) and the mean by a line. The inset shows a crRNA paired to the protospacer and the position of the PAM.

DISCUSSION

The selection of PAM-functional spacers into a CRISPR array is crucial for the recognition and cleavage of target DNA sequences (7,30). Here, we have experimentally investigated a type I-U CRISPR–Cas system for the first time and showed that Cas4/1 fusion proteins, which can be found in type I-U, V-B and I-B CRISPR–Cas systems, catalyze the integration of PAM-compatible spacers with the highest frequency observed to date. PAM-functional spacers were integrated into the CRISPR array from 50% up to 80% compared to ∼20% for unfused *cas4* systems such as type I-D (15). It is important to note that these numbers are obtained in the absence of CRISPR interference, which does not select for spacers with a functional PAM. Recent publications in type I-A systems have shown that the percentage of PAM-functional spacers increases when the full CRISPR–Cas system is present, probably due this selective advantage (16,31). Nevertheless, single mutations in the Cas4 domain proved that unlike some unfused Cas4 systems (e.g. I-D) (15), the spacer adaptation process is in fact reduced, suggesting that the activity of the Cas4 domain strongly controls overall spacer acquisition activity of Cas1. We did not detect spacer integration with the *cas4/1* fission construct indicating that either this control is coordinated through tight linkage between the Cas4 and Cas1 domains, or that the unfused Cas4 and Cas1 proteins are not stable on their own. Either way, the tight linkage between Cas4 and Cas1 may ensure that only PAM-processed prespacers are integrated into the CRISPR array. More-

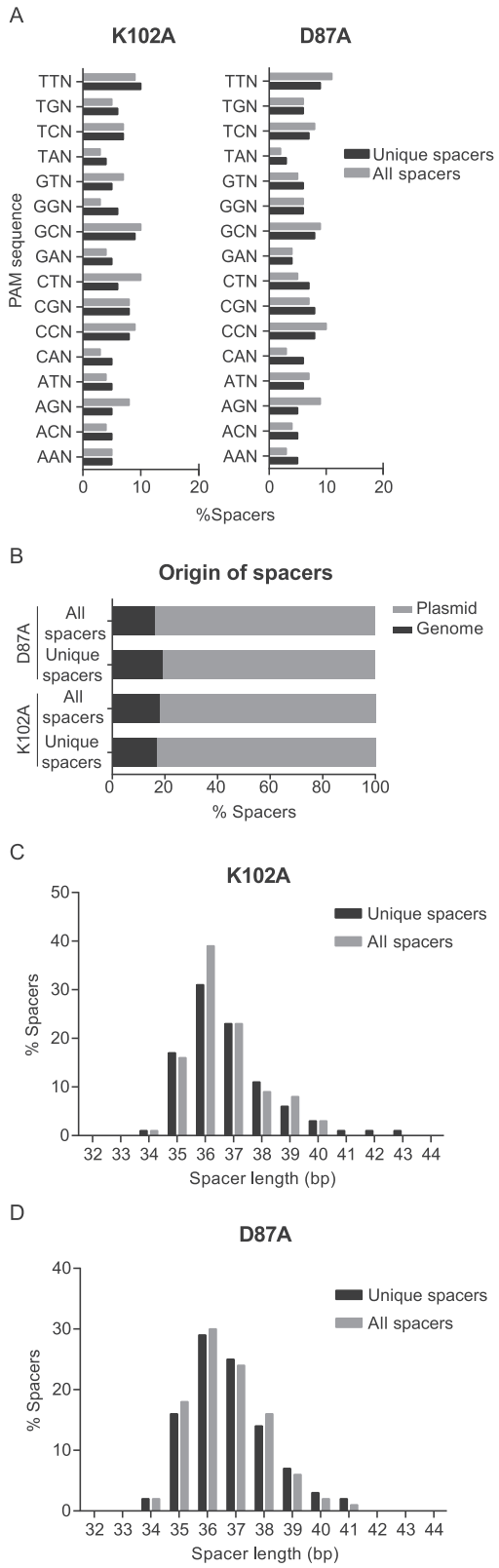

Figure 4. Analysis of spacers integrated by Cas4 RecB mutants. (**A**) PAM selection of all and unique upstream protospacers sequences represented as percentage obtained with Cas4 RecB mutants. (**B**) Origin of newly acquired spacers using the percentage of spacers that matches each replicon obtained with K102A and D87A mutants. (**C** and **D**) Distribution of integrated spacer lengths in the K102A and D87A mutants respectively.

over, contrary to systems with unfused Cas4, we observed no change in spacer length when the Cas4 domain was inactivated (15–17,32). This result suggests a different mechanism of prespacer processing by the fusion protein, probably due to structural differences of the adaptation complex. While in the type I-C system, one unit of Cas4 interacts with two units of Cas1 (17), in the Cas4/1 fusion protein both domains are present at equimolar quantities.

Recently multiple groups have begun to uncover the role of unfused Cas4 systems in CRISPR–Cas systems in relation to Cas1 and have shown that the interaction between the proteins is essential for prespacer processing and PAM recognition (15–17,32). It is thus possible that Cas4/1 evolved as a fused protein to ensure the presence of Cas4 in the acquisition complex and optimize the Cas4–Cas1 interaction for efficient spacer acquisition and PAM selection. However, the way that the fusion protein recognizes and processes the PAM remains unknown. Further experiments should be performed in order to understand why Cas4 domains are required in equimolar ratios with Cas1, while in type I-C systems unfused Cas4 proteins are present only once for every dimer of Cas1 (17). Possibly excess copies of Cas4 in Cas4–Cas1 fusions are not active during PAM recognition, PAM processing and spacer integration.

A phylogenomics study of Cas4 has recently identified that Cas4/1 proteins from I-U and V-B are closely related (20), suggesting the possibility that the adaptation modules have been shared between these two distinct systems by horizontal gene transfer. We provide additional rationale for this functional exchange, as the PAM supporting CRISPR interference in I-U and V-B systems is both TTN (33). Despite this potential horizontal exchange between Class I and II CRISPR–Cas systems, fused Cas4/1 proteins are only found in a subset of type I CRISPR–Cas systems (20), suggesting that the fusion event between *cas4* and *cas1* genes may have occurred relatively recently. Other reasons why these efficient Cas4/1 fusions have not disseminated further in type I systems could be because it might be beneficial for a host to integrate spacers with a PAM that does not drive strong CRISPR interference. Consequently, this will stimulate the acquisition of new spacers by priming, increasing the diversity of spacers stored in the CRISPR array (7).

We have demonstrated the importance of the Cas4 domain in primed adaptation, consistent with previous studies with unfused Cas4 in the type I-B system from *Haloarcula hispanica* (28). Processing of Cas3 DNA degradation products by the Cas4 domain may be required before integration into the CRISPR array by Cas1. Curiously, we found that the type I-U system has a pattern of primed adaptation that is different to other type I systems (34), with no apparent spacer orientation or location bias relative to the primed protospacer. Through the highly accurate PAM selectivity of the Cas4/1 fusion, the type I-U seems to have evolved into a defense system that selects PAM-compatible spacers equally well during naïve and primed spacer acquisition.

Cas4/1 is not the only case of a fusion protein in CRISPR systems. Fusions between reverse transcriptase and Cas1 in type III (35), DnaQ exonuclease and Cas2 in type I-E (36), Cas2 and Cas3 in type I-F (37,38), and Cas3 and Cas8 in type I-E (39) are other examples where fusions of Cas proteins have occurred. It is worth noting that many of these fu-

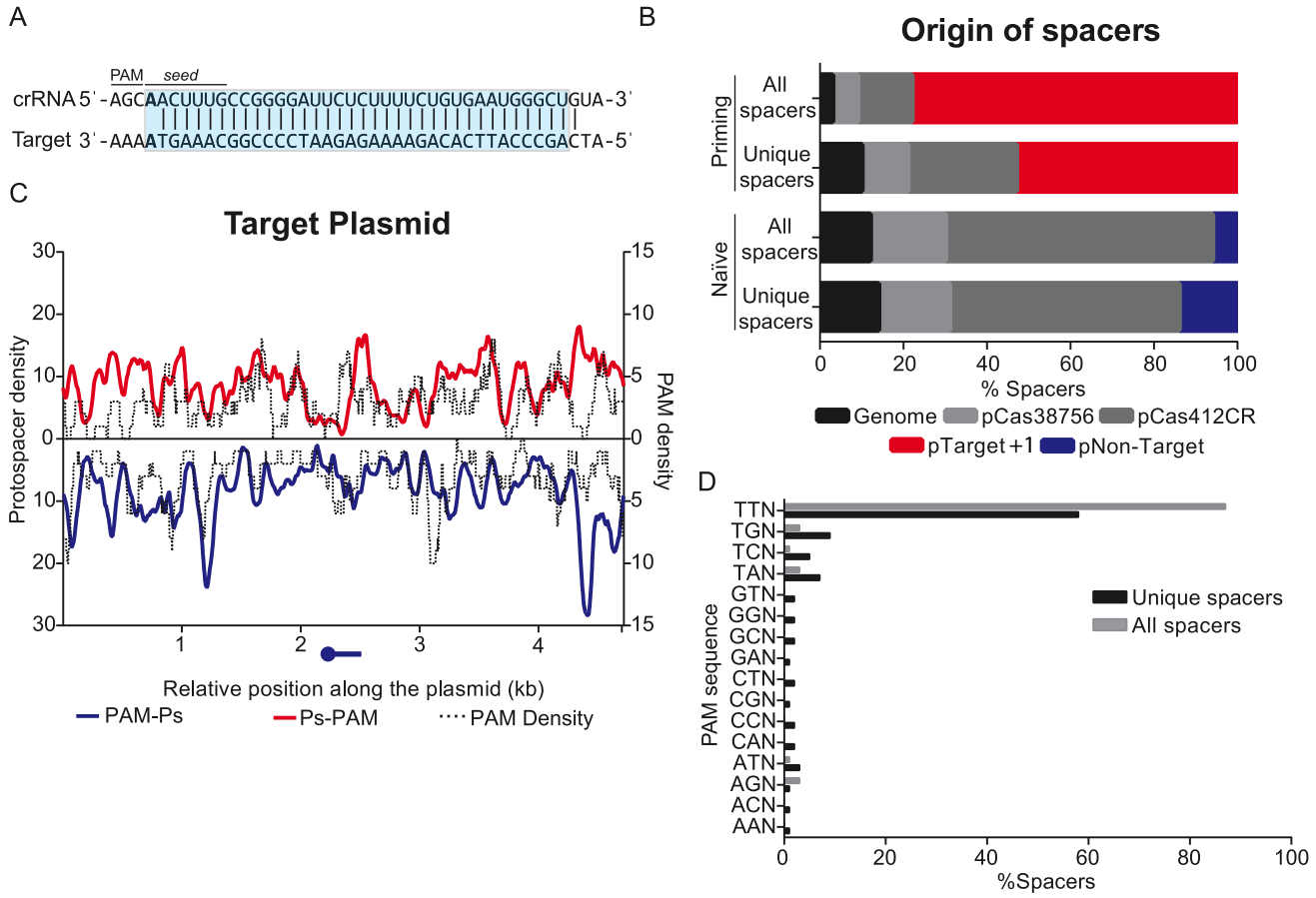

Figure 5. Analysis of spacers integrated by primed acquisition. (**A**) Representation of the imperfect match at position +1 of the protospacer with the crRNA. (**B**) Origin of newly acquired spacers in naïve and priming acquisition assays. The bar representing the percentage of spacers from the target and non-target plasmids are highlighted in red and blue, respectively. (**C**) Representation of the unique protospacer (PS) density using the coverage of the sequences along the plasmid, a sliding 100 nt binning window was applied. Protospacers are represented depending on their orientation (PAM-PS in blue and PS-PAM in red). The PAM density across the plasmid is plotted using a sliding 100 nt binning window and is represented by a black dotted line. The position of the priming target sequence is indicated with a blue line with a dot representing the PAM orientation. (**D**) PAM selection frequencies of all and unique protospacer sequences from all replicons during priming.

sions are involved in the CRISPR adaptation step, suggesting that this step can be optimized to increase the chance of prokaryotes to survive from invasion by mobile genetic elements.

DATA AVAILABILITY

Data has been deposited in the Sequence Read Archive under accession number PRJNA518078: https://www.ncbi.nlm.nih.gov/sra/PRJNA518078.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Carsten J. Blom and Hubertus J.E. Beaumont to provide us the *G. sulfurreducens* strain and Jochem N.A. Vink for assistance with data analysis.
Author contributions: C.A., F.L.N. and S.J.J.B. designed the experiments. C.A., F.L.N. and R.E.M. performed the experiments. C.A. and S.J.J.B. analysed the data. C.A. and S.J.J.B. wrote the paper with input from all authors.

FUNDING

European Research Council (ERC) [Stg grant 638707 to S.J.J.B.]; Start-up grant from Delft University of Technology; Vici grant from the Netherlands Organization for Scientific Research (NWO); Veni grant [016.Veni.181.092 to F.L.N.] from NWO; NWO Frontier of Nanoscience grant (Nanofront) (to R.E.M.). Funding for open access charge: ERC.
Conflict of interest statement. None declared.

REFERENCES

1. van Houte,S., Buckling,A. and Westra,E.R. (2016) Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.*, **80**, 745–763.
2. Jansen,R., van Embden,J.D.A., Gaastra,W. and Schouls,L.M. (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.*, **43**, 1565–1575.

3. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
4. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D.A. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
5. Koonin,E.V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
6. van der Oost,J., Jore,M.M., Westra,E.R., Lundgren,M. and Brouns,S.J.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.*, **34**, 401–407.
7. Jackson,S.A., McKenzie,R.E., Fagerlund,R.D., Kieper,S.N., Fineran,P.C. and Brouns,S.J.J. (2017) CRISPR–Cas: adapting to change. *Science*, **356**, eaal5056.
8. Nuñez,J.K., Lee,A.S.Y., Engelman,A. and Doudna,J.A. (2015) Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature*, **519**, 193–198.
9. Díez-Villaseñor,C., Guzmán,N.M., Almendros,C., García-Martínez,J. and Mojica,F.J.M. (2013) CRISPR-spacer integration reporter plasmids reveal distinct genuine acquisition specificities among CRISPR–Cas I-E variants of Escherichia coli. *RNA Biol.*, **10**, 792–802.
10. Yosef,I., Goren,M.G. and Qimron,U. (2012) Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.*, **40**, 5569–5576.
11. Brouns,S.J.J., Jore,M.M., Lundgren,M., Westra,E.R., Slijkhuis,R.J.H., Snijders,A.P.L.L., Dickman,M.J., Makarova,K.S., Koonin,E.V. and van der Oost,J. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, **321**, 960–964.
12. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
13. Datsenko,K.A., Pougach,K., Tikhonov,A., Wanner,B.L., Severinov,K. and Semenova,E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.*, **3**, 945.
14. Künne,T., Kieper,S.N., Bannenberg,J.W., Vogel,A.I.M., Miellet,W.R., Klein,M., Depken,M., Suarez-Diez,M. and Brouns,S.J.J. (2016) Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell*, **63**, 852–864.
15. Kieper,S.N., Almendros,C., Behler,J., McKenzie,R.E., Nobrega,F.L., Haagsma,A.C., Vink,J.N.A., Hess,W.R. and Brouns,S.J.J. (2018) Cas4 facilitates PAM-compatible spacer selection during CRISPR adaptation. *Cell Rep.*, **22**, 3377–3384.
16. Shiimori,M., Garrett,S.C., Graveley,B.R. and Terns,M.P. (2018) Cas4 nucleases define the PAM, length, and orientation of DNA fragments integrated at CRISPR loci. *Mol. Cell*, **70**, 814–824.e6.
17. Lee,H., Zhou,Y., Taylor,D.W. and Sashital,D.G. (2018) Cas4-dependent prespacer processing ensures high-fidelity programming of CRISPR arrays. *Mol. Cell*, **70**, 48–59.
18. Hein,S., Scholz,I., Voß,B. and Hess,W.R. (2013) Adaptation and modification of three CRISPR loci in two closely related cyanobacteria. *RNA Biol.*, **10**, 852–864.
19. McKenzie,R.E, Almendros,C., Vink,J.N.A. and Brouns,S.J.J. (2019). Using CAPTURE to detect spacer acquisition in native CRISPR arrays. *Nat. Protoc.*, **14**, 976–990.
20. Hudaiberdiev,S., Shmakov,S., Wolf,Y.I., Terns,M.P., Makarova,K.S. and Koonin,E. V. (2017) Phylogenomics of Cas4 family nucleases. *BMC Evol. Biol.*, **17**, 232.
21. Makarova,K.S., Wolf,Y.I., Alkhnbashi,O.S., Costa,F., Shah,S.A., Saunders,S.J., Barrangou,R., Brouns,S.J.J., Charpentier,E., Haft,D.H. *et al.* (2015) An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **13**, 722–736.
22. Wright,A. V, Liu,J.-J., Knott,G.J., Doxzen,K.W., Nogales,E. and Doudna,J.A. (2017) Structures of the CRISPR genome integration complex. *Science*, **357**, 1113–1118.
23. Zhang,J., Kasciukovic,T. and White,M.F. (2012) The CRISPR associated protein Cas4 is a 5′ to 3′ DNA exonuclease with an iron-sulfur cluster. *PLoS One*, **7**, e47232.
24. Levy,A., Goren,M.G., Yosef,I., Auster,O., Manor,M., Amitai,G., Edgar,R., Qimron,U. and Sorek,R. (2015) CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature*, **520**, 505–510.
25. Ivančić-Bace,I., Cass,S.D., Wearne,S.J. and Bolt,E.L. (2015) Different genome stability proteins underpin primed and naïve adaptation in E. Coli CRISPR–Cas immunity. *Nucleic Acids Res.*, **43**, 10821–10830.
26. Künne,T., Zhu,Y., da Silva,F., Konstantinides,N., McKenzie,R.E., Jackson,R.N. and Brouns,S.J. (2018) Role of nucleotide identity in effective CRISPR target escape mutations. *Nucleic Acids Res.*, **46**, 10395–10404.
27. Richter,C., Dy,R.L., McKenzie,R.E., Watson,B.N.J., Taylor,C., Chang,J.T., McNeil,M.B., Staals,R.H.J. and Fineran,P.C. (2014) Priming in the type I-F CRISPR–Cas system triggers strand-independent spacer acquisition, bi-directionally from the primed protospacer. *Nucleic Acids Res.*, **42**, 8516–8526.
28. Li,M., Wang,R., Zhao,D. and Xiang,H. (2014) Adaptation of the Haloarcula hispanica CRISPR–Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.*, **42**, 2483–2492.
29. Rao,C., Chin,D. and Ensminger,A.W. (2017) Priming in a permissive type I-C CRISPR–Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA*, **23**, 1525–1538.
30. McGinn,J. and Marraffini,L.A. (2019) Molecular mechanisms of CRISPR–Cas spacer acquisition. *Nat. Rev. Microbiol.*, **17**, 7–12.
31. Zhang,Z., Pan,S., Liu,T., Li,Y. and Peng,N. (2018) Cas4 nucleases can effect specific integration of CRISPR spacers. bioRxiv doi: https://doi.org/10.1101/486738, 04 December 2018, preprint: not peer reviewed.
32. Rollie,C., Graham,S., Rouillon,C. and White,M.F. (2018) Prespacer processing and specific integration in a Type I-A CRISPR system. *Nucleic Acids Res.*, **46**, 1007–1020.
33. Yang,H., Gao,P., Rajashankar,K.R. and Patel,D.J. (2016) PAM-Dependent Target DNA recognition and cleavage by C2c1 CRISPR–Cas endonuclease. *Cell*, **167**, 1814–1828.
34. Nicholson,T.J., Jackson,S.A., Croft,B.I., Staals,R.H.J., Fineran,P.C. and Brown,C.M. (2018) Bioinformatic evidence of widespread priming in type I and II CRISPR–Cas systems. *RNA Biol.*, **18**, 1–11.
35. Silas,S., Mohr,G., Sidote,D.J., Markham,L.M., Sanchez-Amat,A., Bhaya,D., Lambowitz,A.M. and Fire,A.Z. (2016) Direct CRISPR spacer acquisition from RNA by a natural reverse transcriptase-Cas1 fusion protein. *Science*, **351**, aad4234.
36. Drabavicius,G., Sinkunas,T., Silanskas,A., Gasiunas,G., Venclovas,Č. and Siksnys,V. (2018) DnaQ exonuclease-like domain of Cas2 promotes spacer integration in a type I-E CRISPR-Cas system. *EMBO Rep.*, **19**, e45543.
37. Rollins,M.F., Chowdhury,S., Carter,J., Golden,S.M., Wilkinson,R.A., Bondy-Denomy,J., Lander,G.C. and Wiedenheft,B. (2017) Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 201616395.
38. Fagerlund,R.D., Wilkinson,M.E., Klykov,O., Barendregt,A., Pearce,F.G., Kieper,S.N., Maxwell,H.W.R., Capolupo,A., Heck,A.J.R., Krause,K.L. *et al.* (2017) Spacer capture and integration by a type I-F Cas1–Cas2-3 CRISPR adaptation complex. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, E5122–E5128.
39. Westra,E.R., van Erp,P.B.G., Künne,T., Wong,S.P., Staals,R.H.J., Seegers,C.L.C., Bollen,S., Jore,M.M., Semenova,E., Severinov,K. *et al.* (2012) CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell*, **46**, 595–605.