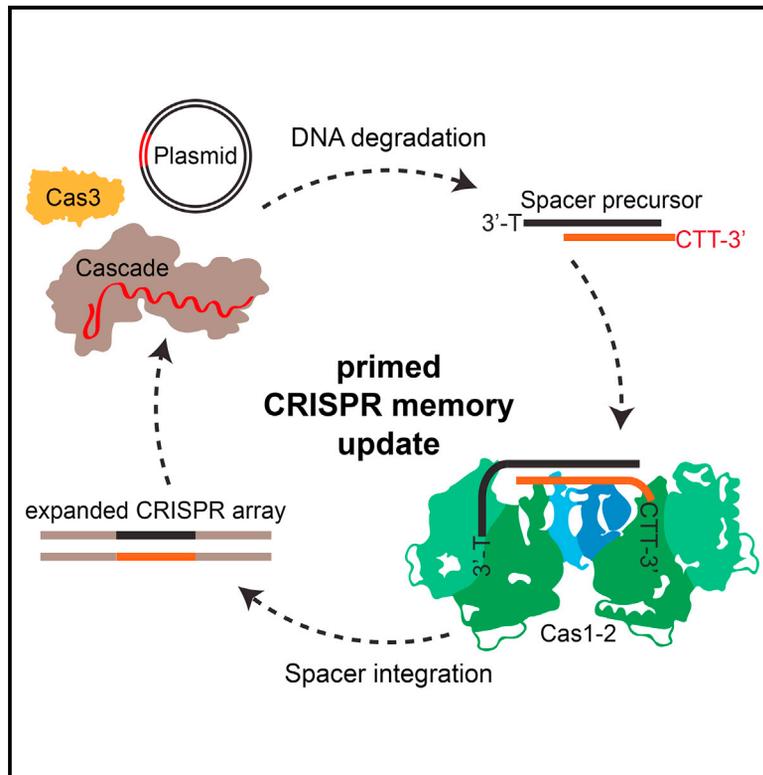


# Molecular Cell

## Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation

### Graphical Abstract



### Authors

Tim Künne, Sebastian N. Kieper, Jasper W. Bannenberg, ..., Martin Depken, Maria Suarez-Diez, Stan J.J. Brouns

### Correspondence

stanbrouns@gmail.com

### In Brief

Künne et al. demonstrate that CRISPR systems cleverly couple target interference to CRISPR memory update. The Cas3 nuclease fragments invader DNA into pieces of near-spacer length enriched for PAM sequences on 3' ends to form ideal spacer precursors.

### Highlights

- CRISPR interference and adaptation are coupled processes
- Moderate direct interference rates stimulate primed spacer acquisition
- Cas1-2 recycle target DNA degradation fragments to form new spacers
- The cleavage specificity of Cas3 contributes to functional PAM selection



# Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation

Tim Künne,<sup>1</sup> Sebastian N. Kieper,<sup>1</sup> Jasper W. Bannenberg,<sup>1</sup> Anne I.M. Vogel,<sup>1,4</sup> Willem R. Miellet,<sup>1</sup> Misha Klein,<sup>2</sup> Martin Depken,<sup>2</sup> Maria Suarez-Diez,<sup>3</sup> and Stan J.J. Brouns<sup>1,2,\*</sup>

<sup>1</sup>Laboratory of Microbiology, Wageningen University, 6708 WE Wageningen, the Netherlands

<sup>2</sup>Kavli Institute of Nanoscience and Department of BioNanoscience, Delft University of Technology, 2629 HZ, Delft, the Netherlands

<sup>3</sup>Laboratory of Systems and Synthetic Biology, Wageningen University, 6708 WE Wageningen, the Netherlands

<sup>4</sup>Present address: Department of Biotechnology, Norwegian University of Science and Technology, 7491 Trondheim, Norway

\*Correspondence: [stanbrouns@gmail.com](mailto:stanbrouns@gmail.com)

<http://dx.doi.org/10.1016/j.molcel.2016.07.011>

## SUMMARY

Prokaryotes use a mechanism called priming to update their CRISPR immunological memory to rapidly counter revisiting, mutated viruses, and plasmids. Here we have determined how new spacers are produced and selected for integration into the CRISPR array during priming. We show that Cas3 couples CRISPR interference to adaptation by producing DNA breakdown products that fuel the spacer integration process in a two-step, PAM-associated manner. The helicase-nuclease Cas3 pre-processes target DNA into fragments of about 30–100 nt enriched for thymine-stretches in their 3' ends. The Cas1-2 complex further processes these fragments and integrates them sequence-specifically into CRISPR repeats by coupling of a 3' cytosine of the fragment. Our results highlight that the selection of PAM-compliant spacers during priming is enhanced by the combined sequence specificities of Cas3 and the Cas1-2 complex, leading to an increased propensity of integrating functional CTT-containing spacers.

## INTRODUCTION

Priming is a mechanism by which immune systems provide an improved immune response to parasite exposure. In vertebrates, priming of adaptive immunity can occur upon first contact of a T or B cell with a specific antigen and causes epigenetic changes as well as cell differentiation into effector T or B cells, producing high levels of antibodies (Bevington et al., 2016). More recently, immune priming has been observed in invertebrates, where it provides increased resistance to previously encountered pathogens (Kurtz and Franz, 2003; Schmid-Hempel, 2005). In plants, priming refers to a state in which the plant can activate its defense responses more rapidly and strongly when challenged by pathogenic microbes, insects, or environmental stress (Conrath et al., 2015). In microbes, priming is a mechanism in which cells can update their immunological memory to provide protection against previously encountered

but slightly changed viruses or conjugative plasmids (Datsenko et al., 2012; Li et al., 2014; Richter et al., 2014; Swarts et al., 2012; Vorontsova et al., 2015). Microbial adaptive immune systems do this by integrating short fragments of invader DNA sequences (called spacers) into clusters of regularly interspaced short palindromic repeats (CRISPR). These spacers are transcribed and processed into small CRISPR RNAs (crRNAs) and guide Cas (CRISPR-associated) surveillance complexes such as Cascade, Cas9, Cpf1, Csm, and Cmr to their DNA or RNA target sequences, resulting in target cleavage and neutralization of the invading threat (Carter and Wiedenheft, 2015; Charpentier et al., 2015; Makarova et al., 2015; Marraffini, 2015; Reeks et al., 2013).

For many years, the acquisition of new spacers was the least understood process in CRISPR-Cas defense, but recent advances have begun to change this (Amitai and Sorek, 2016; Fineran and Charpentier, 2012; Heler et al., 2014; Sternberg et al., 2016). In the type I-E system of *E. coli*, Cas1 and Cas2 form a complex that binds, processes, and integrates DNA fragments into the CRISPR array to form spacers (Arslan et al., 2014; Nuñez et al., 2014, 2015b; Rollie et al., 2015; Wang et al., 2015). Apart from priming, spacers can also be acquired in a naive manner. During naive acquisition, the host acquires spacers from an invading DNA element that has not been cataloged in the CRISPR array yet. This process is dependent on DNA replication of the invading DNA element (Levy et al., 2015) and requires only *cas1* and *cas2* genes (Yosef et al., 2012). In type I CRISPR-Cas systems, primed acquisition makes use of pre-existing spacers that partially match an invading DNA element. Therefore, primed acquisition of spacers is important to rapidly counter invaders that escape immunity by mutating their target site (Cady et al., 2012; Datsenko et al., 2012; Fineran et al., 2014; Semenova et al., 2011; Xue et al., 2015). Priming allows new spacers from such an “escaper” to be rapidly acquired, leading to renewed immunity. Priming is especially advantageous for a host because the process quickly generates a population of bacteria with different spacers against the same virus, efficiently driving the virus extinct (van Houte et al., 2016). In addition to Cas1-2, all remaining Cas proteins are required for priming, including the crRNA effector complex Cascade and the nuclease-helicase Cas3 (Datsenko et al., 2012; Richter et al., 2014). Although the genetic requirements for priming are known, the exact role of these proteins during priming remains

unknown. Several models that explain parts of the priming process have been proposed.

In the Cascade-sliding model, Cascade moves along the DNA until a protospacer-adjacent motif (PAM) is encountered, which marks the DNA for acquisition of a new spacer (Datsenko et al., 2012). A second model was proposed in which a Cas1: Cas2-3 complex translocates away from the primed protospacer marked by the crRNA-effector complex until a new PAM is encountered (Richter et al., 2014). A new spacer is then acquired from this new PAM site. Recently, supporting evidence for this hypothesis has been obtained. Single-molecule studies have suggested that Cascade bound to a priming protospacer recruits Cas1-2, which in turn recruit a nuclease inactive Cas3 (Redding et al., 2015). A complex of Cas1-3 may then translocate along the DNA to select new spacers. Although these models describe the biochemistry and movement of the proteins involved in priming, it has remained unknown how actual DNA fragments from an invading element are obtained to drive the priming process. We have previously put forward a model in which we propose that DNA breakdown products of Cas3 provide the positive feedback needed to fuel the priming process (Swarts et al., 2012). Similar models were proposed for priming in I-B and I-F systems (Li et al., 2014; Vorontsova et al., 2015). In line with that hypothesis, it has recently been suggested that during naive acquisition, spacer precursors are generated during DNA repair at double-stranded breaks (Levy et al., 2015). These breaks are frequently formed at stalled replication forks during DNA replication and are repaired by the RecBCD complex. RecBCD unwinds the DNA strands with its helicase activity, while degrading the subsequent single-stranded stretches using exonuclease activity. The resulting DNA oligomers have been proposed to form precursors for Cas1-2 to produce new spacers. Similar to RecBCD, Cas3 is also a nuclease-helicase that degrades double-stranded DNA (dsDNA) by unwinding, with the difference that Cas3 has been shown to degrade one strand at a time (Gong et al., 2014; Huo et al., 2014; Mulepati and Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012). This leads to the hypothesis that Cas3 also produces substrates for Cas1-2 mediated spacer acquisition during priming.

Here we have tested that hypothesis and prove that plasmid degradation products produced by Cas3 are bound by the Cas1-2 complex, processed into new spacers and integrated into the CRISPR array. The cleavage frequency and cleavage specificity of Cas3 facilitate the production of functional spacer precursor molecules that meet all requirements of new spacers. To achieve this, Cas3 produces fragments that are in the range of the length of a spacer (30–100 nt). Furthermore, the cleavage specificity of Cas3 leads to an enrichment of PAM sequences in the 3' end of these fragments, which enhances the selection of productive spacer precursors by Cas1-2. Our results demonstrate that the DNA degradation fragments produced by Cas3 are the direct link between CRISPR interference and adaptation that make the priming mechanism so robust.

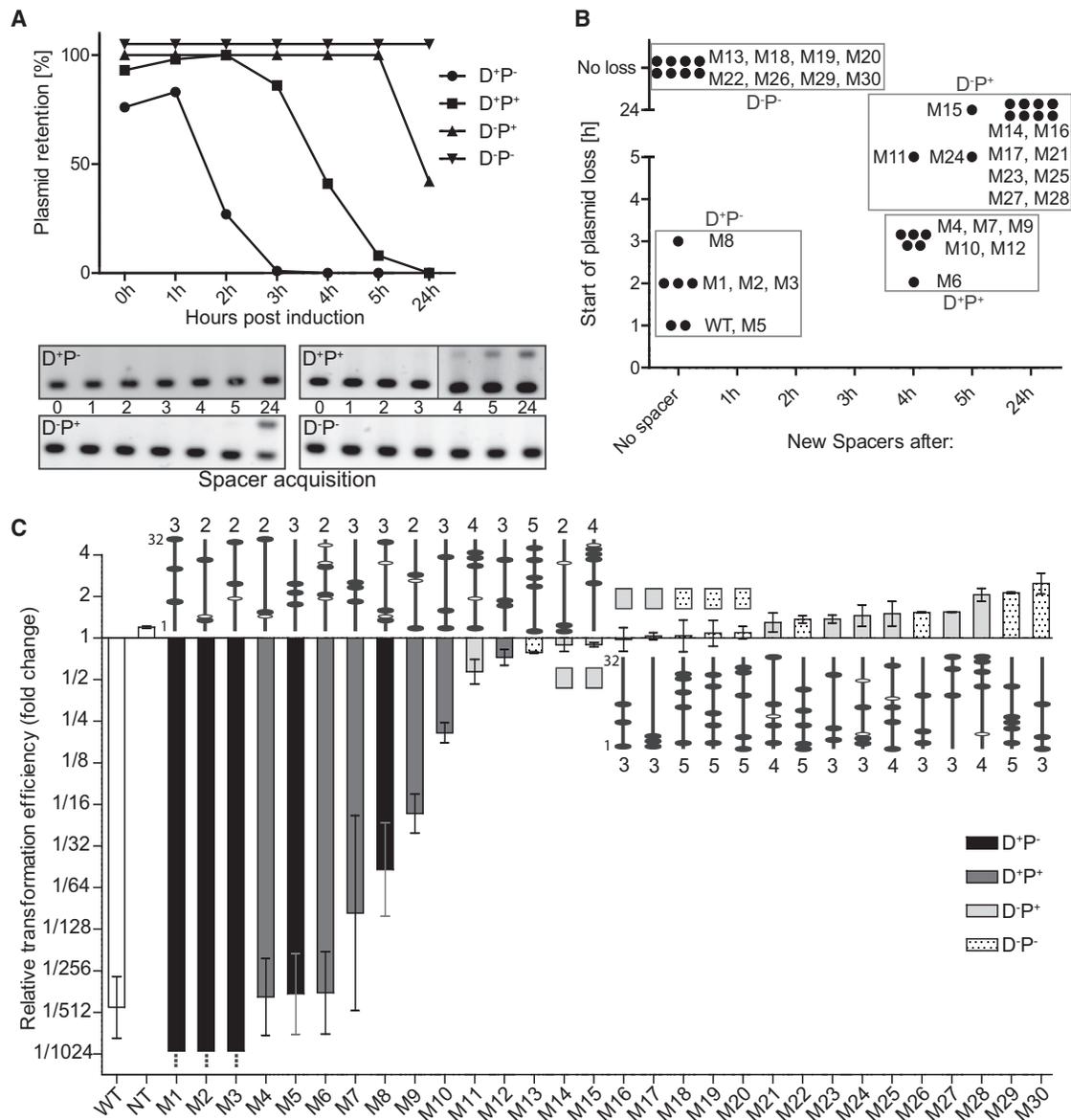
## RESULTS

Previous studies have shown that direct interference in type I CRISPR-Cas systems (i.e. the breakdown of Cascade-flagged

invading DNA by Cas3) is relatively sensitive to mutations in the PAM and seed sequence of the protospacer (Künne et al., 2014; Semenova et al., 2011; Wiedenheft et al., 2011; Xue et al., 2015). Priming on the other hand is an extremely robust process capable of dealing with highly mutated targets with up to 13 mutations. Priming is influenced by a complex combination of the number of mutations in a target, the position of these mutations, and the nucleotide identity of the mutation. Furthermore, the degree of tolerance of mutations in a protospacer during interference and priming depends on the spacer choice (Xue et al., 2015).

### Timing of Plasmid Loss and Spacer Acquisition Reveals Distinct Underlying Processes

In order to find the molecular explanation for why some mutants with equal numbers of mutations show priming while others do not, we performed detailed analysis of a selected set of target mutants obtained previously (Fineran et al., 2014). From the available list, we chose the bona fide target (WT) and 30 mutants carrying an interference permissive PAM (i.e. 5'-CTT-3'). The mutants had between 2 and 5 effective mutations (i.e., mutations outside the kinked positions 6, 12, 18, 24, and 30; Fineran et al., 2014; Jackson et al., 2014; Mulepati et al., 2014; Zhao et al., 2014) (Figure S1). We used *E. coli* strain KD263 with inducible expression of *cas3* and *cascade-cas1-2* genes (Shmakov et al., 2014) to test both direct interference and priming in a plasmid loss setup. Plasmid loss curves of individual mutants (Figure S2) showed four distinct behaviors that led us to classify these target mutants into four groups: mutants capable of only direct interference ( $D^+P^-$ ), mutants capable of direct interference and priming ( $D^+P^+$ ), mutants capable of only priming ( $D^-P^+$ ), and mutants incapable of both direct interference and priming ( $D^-P^-$ ) (Figures 1A and 1B). As expected, rapid plasmid loss was observed for the bona fide target, but also for five mutant targets. These target variants ( $D^+P^-$ ) showed plasmid loss within 2 hr post-induction (hpi), reaching complete loss after 3 hpi (Figure 1B, bottom left cluster), and did not incorporate new spacers. The  $D^+P^+$  group of mutants showed a slower decrease in plasmid abundance (starting at ~3 hpi), and this decrease was accompanied by the incorporation of new spacers 4 hpi (Figure 1B, bottom right cluster). The  $D^-P^+$  group of mutants showed more strongly delayed plasmid loss (>5 hpi), and this loss was preceded or directly accompanied by spacer acquisition (Figure 1B, top right cluster). Therefore, these mutants could not be cleared from the cells by direct interference initially, but after primed spacer acquisition, the plasmid was rapidly lost. No spacer incorporation was observed for  $D^-P^-$  targets, and these variants did not show any plasmid loss within 48 hpi, similar to a non-target plasmid (Figure 1B, top left cluster). This group exemplifies that no naive acquisition had occurred within 48 hr in our experimental setup and that all spacer integration events observed in  $P^+$  groups were due to priming. To validate that spacer acquisition occurred by priming, we sequenced the newly incorporated spacers for a representative set of clones, especially including mutants with late acquisition. We did indeed observe the 9:1 strand bias of new spacers that is typical for priming



**Figure 1. Plasmid Loss and Transformation Assay**

Plasmid loss was assessed by plating cells and scoring for the GFP signal at various time points after induction of cas genes. Individual assays can be seen in Figure S2. The bona fide target is abbreviated as WT.

(A) Example curves and CRISPR PCR of four different types of plasmid behaviors that were observed: rapid plasmid loss without spacer integration (D<sup>+</sup>P<sup>-</sup>), delayed plasmid loss and spacer integration (D<sup>+</sup>P<sup>+</sup>), strongly delayed plasmid loss and spacer integration (D<sup>-</sup>P<sup>+</sup>), and no plasmid loss with no spacer integration (D<sup>-</sup>P<sup>-</sup>).

(B) Summary of plasmid behavior of all mutants, showing timing of first plasmid loss and time of first observable spacer integration.

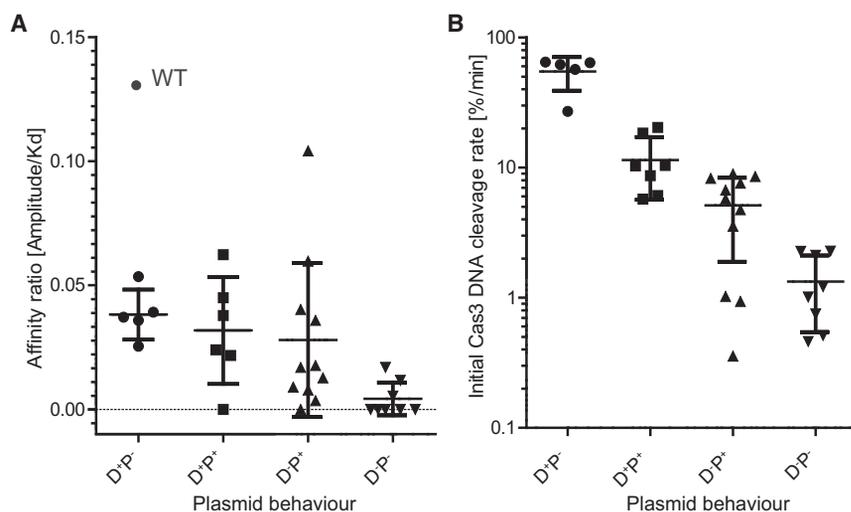
(C) The relative transformation efficiency is plotted for all mutant plasmids (fold change compared with co-transformed non-target plasmid, log<sub>2</sub> scale). Bars are color coded on the basis of plasmid behavior classification. Error bars represent SEM of triplicate experiments. The positions of mutations are indicated schematically for each mutant (position 1, bottom; position 32, top). Open ovals represent mutations on positions 6, 12, 18, 24, and 30. Closed ovals represent mutations outside of those positions (effective mutations). The amount of effective mutations is indicated above or below the schematic.

For a more detailed overview of the mutations, see Figure S1.

(Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). Taken together, we found that priming is facilitated by slow or delayed direct interference (D<sup>+</sup>P<sup>+</sup>), but that it does not strictly require direct interference as exemplified by the D<sup>-</sup>P<sup>+</sup> group.

### Moderate Direct Interference Activity Facilitates the Priming Process

To verify that rapid plasmid loss indeed results from direct interference, we performed plasmid transformation assays of the target plasmid set into *E. coli* KD263 and compared the



**Figure 2. Electrophoretic Mobility Shift Assay and Cas3 Activity Assay**

All mutants are classified according to previously identified plasmid behavior. The mean and SD for each group are indicated.

(A) Electrophoretic mobility shift assay (EMSA) of the mutant plasmid set. The affinity ratio (amplitude/ $K_d$ ) is plotted for each mutant (see Table S3 for more details). The bona fide target is abbreviated as WT.

(B) Cas3 DNA degradation activity assay of mutant plasmid set. The initial Cas3 DNA cleavage rate (percentage per minute) is plotted for each mutant. Individual gels for all activity assays can be found in Figure S4.

transformation efficiency to a co-transformed control plasmid (Almendros and Mojica, 2015). Although the bona fide target plasmid exhibited a relative transformation efficiency that was 512 times lower than the control plasmid (1/512), also mutants with up to two effective mutations gave rise to strongly decreased transformation efficiencies (1/16–1/512) (Figure 1C). This means that these target variants still triggered an efficient direct interference response. Triple mutants showed a range of relative transformation efficiencies from full direct interference (i.e. 1/512) to no direct interference ( $\sim 1$ ), suggesting a dominant role for the position of the mutations in the protospacer. Mutants with four or five effective mutations transformed as efficiently as the reference plasmid and displayed no direct interference. When we mapped the classification of all the mutants onto the relative transformation efficiency data, the same trend was observed that target variants with the highest direct interference showed no priming. Instead, intermediate levels of direct interference led to rapid spacer acquisition, while low levels or the absence of direct interference led to delayed spacer acquisition. This also confirms that late plasmid loss in the D<sup>-</sup>P<sup>+</sup> group is indeed not caused by direct interference with the original spacer but by primed spacer acquisition followed by direct interference.

### Pairing at the Middle Position of Each Segment Is Important for Direct Interference

The average number of effective mutations in a protospacer increases gradually over the groups D<sup>+</sup>P<sup>-</sup>, D<sup>+</sup>P<sup>+</sup>, D<sup>-</sup>P<sup>+</sup>, and D<sup>-</sup>P<sup>-</sup> (Figure S1). While D<sup>+</sup>P<sup>-</sup> and D<sup>+</sup>P<sup>+</sup> had either two or three effective mutations, the D<sup>-</sup>P<sup>+</sup> mutants had three or four mutations, and the D<sup>-</sup>P<sup>-</sup> mutants carried three or five effective mutations in the protospacer. In order to quantify how significant the shifts in the average number of mutations are, we used empirical bootstrapping to test against the hypothesis that the classification does not depend on the number of mutations. Our analysis showed that the D<sup>+</sup>P<sup>-</sup> and D<sup>+</sup>P<sup>+</sup> groups have significantly fewer mutations than would be expected if the classification did not correlate with the number of mutations (>95% and >68% confidence, respectively), while D<sup>-</sup>P<sup>-</sup> has significantly more mutations (>95% confidence) (Figure S3A). We next looked in detail

at the number of mutations in each segment and the position of mutations in each five-nucleotide segment. As has been observed for the seed sequence (Semenova et al., 2011; Wiedenheft et al., 2011), this showed a significantly lower than average number of mutations in segment 1 for D<sup>+</sup>P<sup>-</sup> and D<sup>+</sup>P<sup>+</sup> groups (both 95% confidence; Figure S3B). Surprisingly, the analysis also revealed that groups showing direct interference (D<sup>+</sup>P<sup>-</sup>, D<sup>+</sup>P<sup>+</sup>) had no mutations at the third position of each segment (significantly lower than expected, 95% confidence), whereas D<sup>-</sup>P<sup>+</sup> and D<sup>-</sup>P<sup>-</sup> groups were enriched for mutations at this position (>68% and >95% confidence, respectively; Figure S3C). This observation therefore suggests that pairing of the middle nucleotide of the segment is somehow important for direct interference. The third nucleotide of each segment could represent a tipping point in the directional pairing of the crRNA to the DNA. This may occur during canonical, PAM-dependent target DNA binding, which leads to R-loop locking, efficient Cas3 recruitment and target DNA degradation (Blosser et al., 2015; Huo et al., 2014; Rutkauskas et al., 2015).

### Cascade-Plasmid Binding Is Required for Interference and Priming

To determine the biochemical basis of priming, we first asked the question what determines whether a mutant target can prime or not, and we hypothesized that the affinity of Cascade for a target plasmid would determine its fate. To test this, we performed plasmid-based mobility shift assays with purified Cascade complexes (Künne et al., 2015). Although the bona fide target and most of the mutant targets were bound to completion at increasing Cascade concentrations, some mutant target plasmids were only partially bound (Table S3), as has been observed before (Hochstrasser et al., 2014). By calculating an affinity ratio (amplitude/ $K_d$ ) and using it as an index for the binding strength, we were able to directly compare the binding properties of all target mutants (Figure 2A). The results show that the bona fide target plasmid had the highest affinity ratio (0.31 nM<sup>-1</sup>), while the mutants cover a range of ratios ranging from very weak binding (>0.008 nM<sup>-1</sup>) to almost the same levels as the bona fide target (<0.1 nM<sup>-1</sup>). D<sup>-</sup>P<sup>-</sup> mutants all cluster together with low

ratios ( $<0.02 \text{ nM}^{-1}$ ), and five of eight show no measurable Cascade binding. This suggests that a minimal level of target plasmid binding by Cascade is required for both direct interference and priming. However, the affinity ratio alone does not predict direct interference and/or priming behavior of a target plasmid.

### Cas3 DNA Cleavage Activity Determines Plasmid Fate

Next, we analyzed if the catalytic rate of target DNA degradation by Cas3 would be related to direct interference and priming. Target DNA degradation is required for direct interference and might be required for priming as well, since all *cas* genes are required for priming in *E. coli* (Datsenko et al., 2012). To test this, we performed Cas3 activity assays with the same panel of target plasmids (Figures 2B and S4). This showed that there is a strong dependence between plasmid fate and Cas3 activity. Mutants capable of only direct interference ( $D^+P^-$ ) display 5–10 times higher activity than priming mutant classes ( $D^+P^+$ ,  $D^-P^+$ ), while stable mutants ( $D^-P^-$ ) show the lowest Cas3 activity. Furthermore,  $D^+P^+$  mutants show a slightly higher average activity than  $D^-P^+$  mutants. The difference between the Cascade affinity and the Cas3 activity plots shows that Cas3 activity is not a simple reflection of Cascade affinity, but is likely influenced by other factors such as conformational differences or the dynamics of Cascade binding. Taken together, there is a link between the Cas3 activity on a target and target plasmid fate. Direct interference requires the highest Cas3 activity, while priming requires a level of target degradation and occurs at a broad range of intermediate or low Cas3 activities. Finally, it is striking that higher Cas3 activities seem to result in faster priming ( $D^+P^+$  versus  $D^-P^+$ ), while very high Cas3 activities ( $D^+P^-$ ) do not lead to priming.

### Cas3 Produces Degradation Fragments of Near Spacer Length

After establishing a connection between plasmid degradation (direct interference) and primed spacer acquisition, we sought to analyze whether the degradation fragments created by Cas3 could serve as spacer precursors. To that end, we performed Cascade-mediated plasmid degradation assays with Cas3 and plasmids containing the bona fide target or M4 mutant. Agarose gel electrophoresis showed that both target plasmids were degraded into similar-sized products smaller than 300 nt. Further biochemical analysis of the products revealed that the products were of double-stranded nature and contained phosphates at their 5' end (Figures S5A and S5B). On the basis of the unidirectional unwinding and single-stranded DNA cleavage mechanism of Cas3 (Gong et al., 2014; Huo et al., 2014; Mulepati and Bailey, 2013; Sinkunas et al., 2013; Westra et al., 2012), we had expected to find single-stranded DNA. However, it appeared that complementary fragments had re-annealed to form duplexes, most likely generating annealed products with both 3' and 5' overhangs.

In order to determine the exact cleavage patterns of target plasmids by Cas3, we isolated DNA cleavage products and sequenced them using the Illumina MiSeq platform. Analysis of the length of the DNA degradation products from the bona fide and M4 target revealed that the majority of fragments from the

target strand had a size of about 30–70 nt (Figures 3B and S6A). The non-target strand displayed a shifted distribution, with most fragments being 60–100 nt long. Instead of cleaving the target DNA randomly, Cas3 produces fragments with a distinct length profile. Furthermore, the length of the main fraction, especially in the target strand, is close to the length of a spacer molecule (i.e., 32 or 33 nucleotides), supporting the idea that these fragments might be used as spacer precursor molecules.

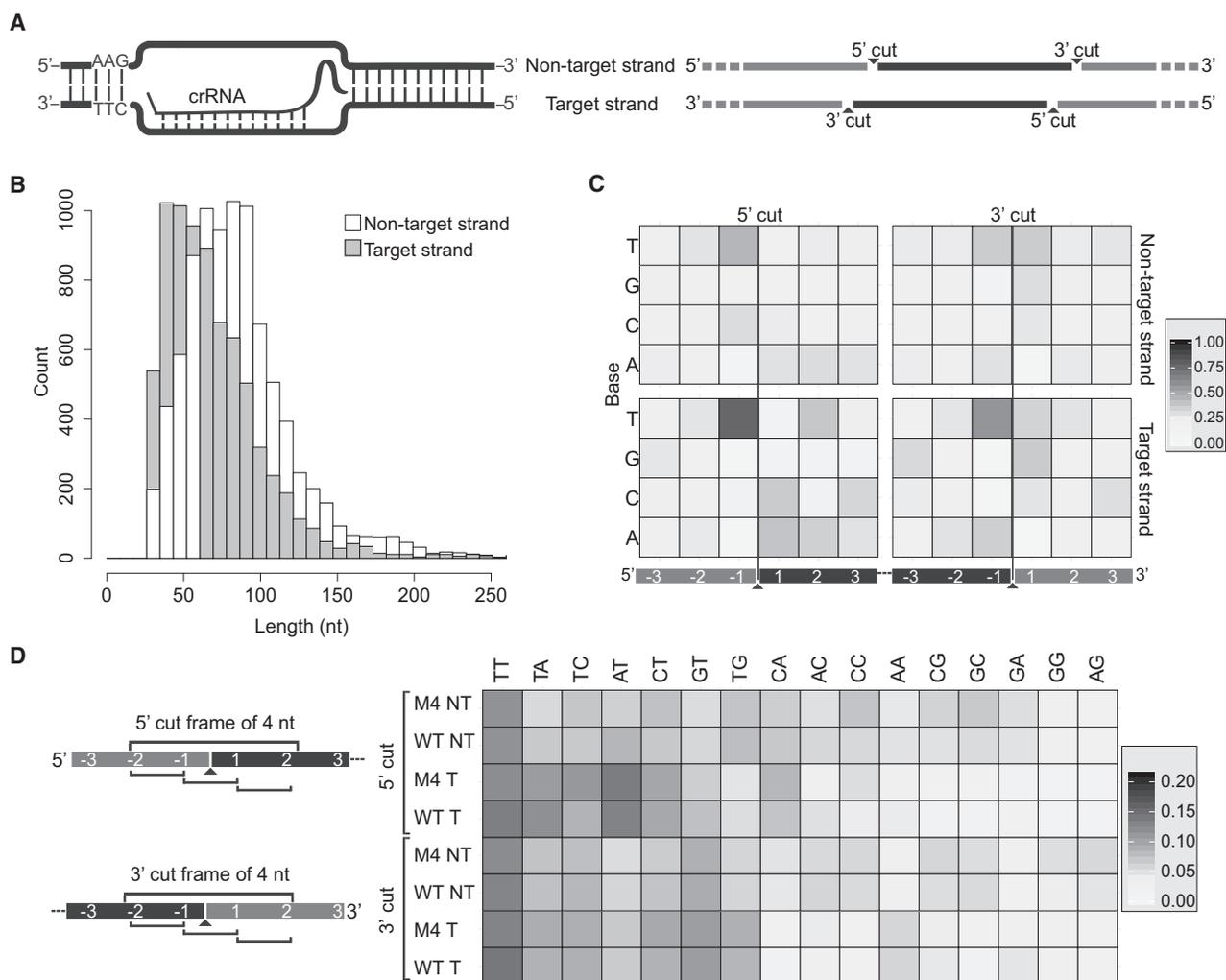
### Cas3 Cleavage Is Sequence Specific for Thymine Stretches

In order to determine whether Cas3 cleaves the target DNA in a sequence-specific manner, we analyzed the region encompassing the cleavage site. This revealed a preference for Cas3 to cleave in thymine-rich sequences for both the bona fide and the M4 target, preferably cleaving 3' of a T nucleotide (Figures 3C, 3D, and S6B). The same pattern was also observed for single-stranded m13mp8 DNA cleaved in the absence of Cascade, indicating that T-dependent cleavage specificity is an inherent feature of the HD domain of Cas3. The cleavage specificity of Cas3 leaves one or multiple T nucleotides on the 3' ends of DNA degradation products. This enriches the 3' ends of the fragments for NTT sequences, including the PAM sequence CTT. A considerable proportion of degradation fragments therefore satisfies the requirement of Cas1-2 for having CTT sequences in the 3' ends of spacer precursors in order for these to be correctly integrated into the CRISPR array (Shipman et al., 2016; Wang et al., 2015). Interestingly, C/T-associated cleavage has previously been shown for *Streptococcus thermophilus* Cas3 cleaving oligo nucleotides (Sinkunas et al., 2013), suggesting that this cleavage specificity may be common for HD-domains of Cas3 proteins.

### Cas1-2 Integrate Cas3-Derived Degradation Fragments

To find out if Cas3 degradation products can indeed serve as spacer precursors, we reconstituted spacer integration in vitro using purified Cas proteins. Two types of spacer integration assays were performed (Figure 4A): the first assay used all Cas proteins simultaneously (Cascade, Cas3, Cas1-2) to degrade a target plasmid and integrate the resulting fragments into a plasmid carrying a leader and single CRISPR repeat (pCRISPR). The second assay used DNA degradation products from a separate Cascade-Cas3 reaction. These products were incubated with Cas1-2 and pCRISPR, as described (Nuñez et al., 2015b). We noticed a pronounced Cas1-2-dependent shift of the degradation fragments in the gel, suggesting the fragments are bound by Cas1-2 (Figure 4B, left). Interestingly, when Cas1-2 was present in the reaction we observed twice as much nicking of plasmid pCRISPR, suggesting that half-site integration of DNA fragments into pCRISPR had occurred (Figure 4B, right) (Nuñez et al., 2015b). The same pCRISPR nicking activity was observed using purified Cas3 degradation products (integration assay 2), indicating that the integration reaction was not dependent on Cascade or Cas3.

To verify that spacer half-site integration had taken place and not just pCRISPR nicking, we gel-isolated the nicked pCRISPR band for PCR analysis. Because we did not know the sequence



**Figure 3. Next-Generation Sequencing Analysis of Cas3 DNA Degradation Products**

(A) Left: schematic of R-loop formed by binding of Cascade to dsDNA target. Right: schematic showing the four distinct Cas3 cleavage sites in dsDNA target.

(B) Length distribution of Cas3 DNA degradation fragments of M4 target.

(C) Heatmap of nucleotide frequencies around cleavage sites. The cleavage site is between positions –1 and 1. Positions indicated in black are on the fragments; positions indicated in gray are outside of fragments.

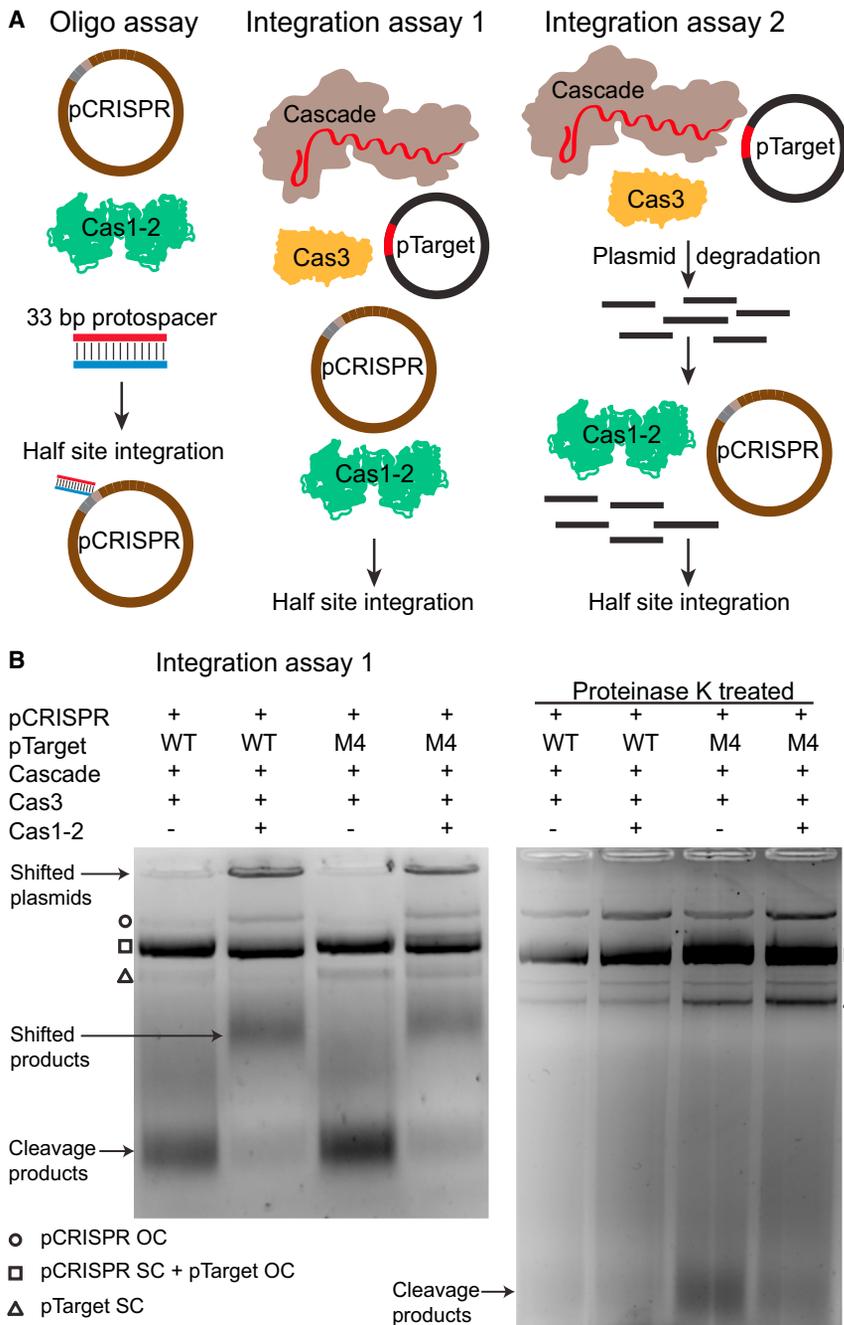
(D) Heatmap of dinucleotide frequencies around cleavage sites. Abundance of dinucleotides was measured in a shifting frame within four nucleotides around the cleavage sites.

See also Figure S6.

of the integrated fragments, we selected three primer pairs that would amplify frequently incorporated spacers from the plasmid *in vivo* (Fineran et al., 2014). Two of the three tested primers gave a PCR product of the expected size, and we chose one of the primers for more detailed analysis. It has previously been shown that the first half-site integration may occur at the boundary of the leader and repeat in the sense strand (i.e. site 1) or at the penultimate base of the repeat in the antisense strand (i.e. site 2) (Nuñez et al., 2015b; Rollie et al., 2015). Furthermore, fragments can be integrated in two different orientations. We performed PCR amplification reactions to test for all four different situations (Figure 5A). This showed that integration of Cas3-derived degradation products occurs sequence specifically at both site 1 and site 2 and in both orientations (Figure 5B).

### Integration of Fragments in the Repeat Is Nucleotide and Position Specific

In order to obtain more insight into the accuracy of integration, we sequenced 48 clones for each of the four primer sets. The results confirm that fragments from the target and non-target strands are integrated at both site 1 and site 2 of the repeat. Integration is very specific to the correct positions in the repeat. At site 1, 94% of the integrated fragments were coupled correctly to the first nucleotide of the sense strand of the repeat, whereas at site 2, 73% of integrated fragments were coupled correctly to the penultimate nucleotide of the antisense strand of the repeat, replacing the last nucleotide of the repeat in the process (Figure 6A). In line with previous findings (Nuñez et al., 2015b; Rollie et al., 2015), both integration sites show a preference for



### Figure 4. In Vitro Spacer Acquisition Assays

(A) Illustration of the three types of assays performed. In the oligo assay, pCRISPR is incubated with Cas1-2 and a spacer oligo (BG7415/6), leading to half-site integration. In assay 1, pTarget and pCRISPR are incubated with Cascade, Cas3, and Cas1-2 for simultaneous degradation of pTarget and half-site integration into pCRISPR. In assay 2, pTarget is incubated with Cascade and Cas3, and the resulting DNA degradation products are then separately incubated with pCRISPR and Cas1-2. (B) Gel electrophoresis of integration assay 1. The bona fide target is abbreviated as WT. Left gel, untreated; right gel, proteinase K treated. Cas1-2 presence causes upward shift of DNA. Original plasmids are supercoiled (SC); half-site integration causes nicking of pCRISPR, resulting in the open circular conformation (OC).

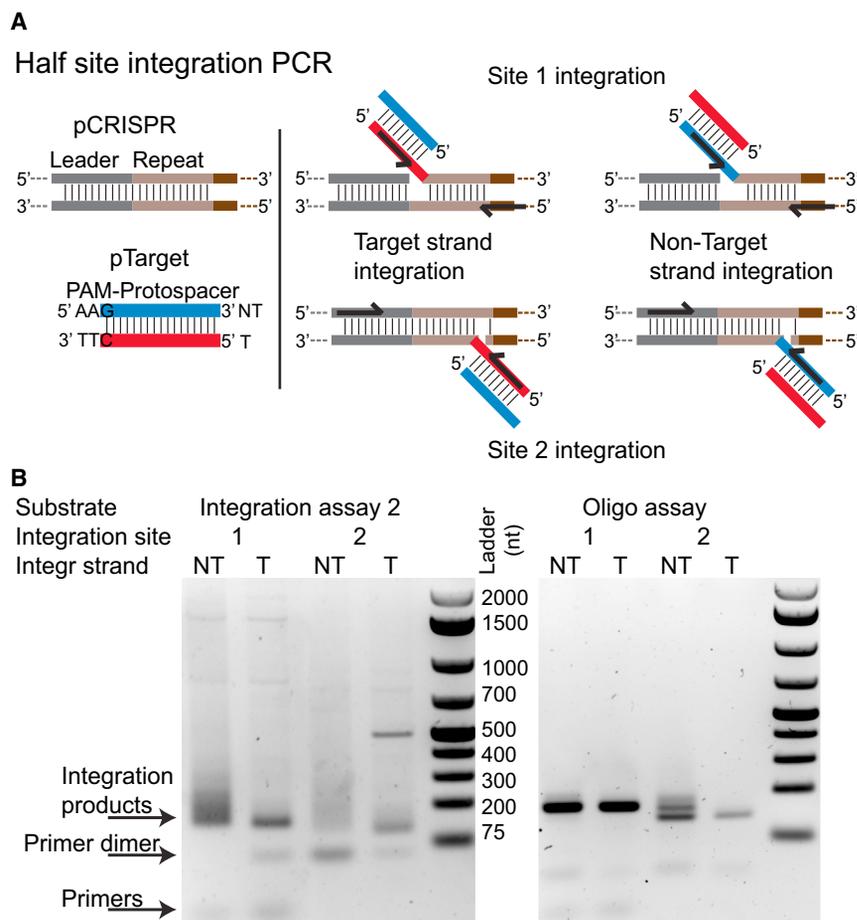
Cas3, which leaves thymines in the 3'-end of DNA fragments, enriching the fragment ends for CTT, and (2) Cas1-2, which prefer CTT-carrying substrates and process and couple the 3' cytosine specifically to both integration sites of the repeat.

### DISCUSSION

A remaining gap in our understanding of type I CRISPR-Cas mechanisms is how new spacers are selected and processed before being incorporated into the CRISPR array. In this work we demonstrate that Cas3 produces spacer precursors for primed adaptation of the CRISPR array. Cas3 DNA degradation fragments fulfill all criteria for spacer precursors that can be deduced from recent studies of the Cas1-2 complex (Figure 7). Ideal spacer precursors in *E. coli* are partially double-stranded duplexes of at least 35 nucleotides containing splayed single-stranded 3' ends with a CTT PAM sequence on one of the 3' overhangs (Nuñez et al., 2015a; Rollie et al., 2015; Shipman et al., 2016; Wang et al., 2015).

coupling incoming C nucleotides: 49% and 55% for site 1 and site 2, respectively (Figure 6A). Considering that Cas3 DNA degradation fragments have T nucleotides on their 3' ends, this suggests that precursors have been pre-processed by Cas1-2 before integration, as has been demonstrated for artificial substrates (Wang et al., 2015). The majority of the integration amplicons had a length of only 20–40 nucleotides (Figure 6B), indicating that the integration reaction prefers short to long substrates. Altogether, we show that the integration of PAM-containing spacers in the repeat during priming is enhanced by the combined sequence specificities of two Cas enzymes: (1)

We have shown that Cas3 DNA degradation products are mainly double-stranded in vitro. This is most likely due to re-annealing of the single-stranded products that are produced by the nuclease-helicase activity of Cas3. It is possible that in vivo, other proteins are involved in the formation of duplexes after degradation. In fact, it has been shown that Cas1 from *Sulfolobus solfataricus* can facilitate the annealing of oligonucleotides (Han and Krauss, 2009). These re-annealed duplexes likely contain a mix of 3' and 5' overhangs, because the two DNA strands of the target are degraded independently. This also results in slightly shorter fragments for the target strand. Despite these differences in



**Figure 5. Half-Site Integration PCR**

(A) Illustration of the half-site integration PCR. Primer sets are chosen to show integration into site 1 (leader-proximal repeat end) and site 2 (leader-distal repeat end) and to see both possible orientations of the integrated spacer. Primer sequences were chosen on the basis of frequently incorporated spacers (hot spots) in vivo (Fineran et al., 2014).

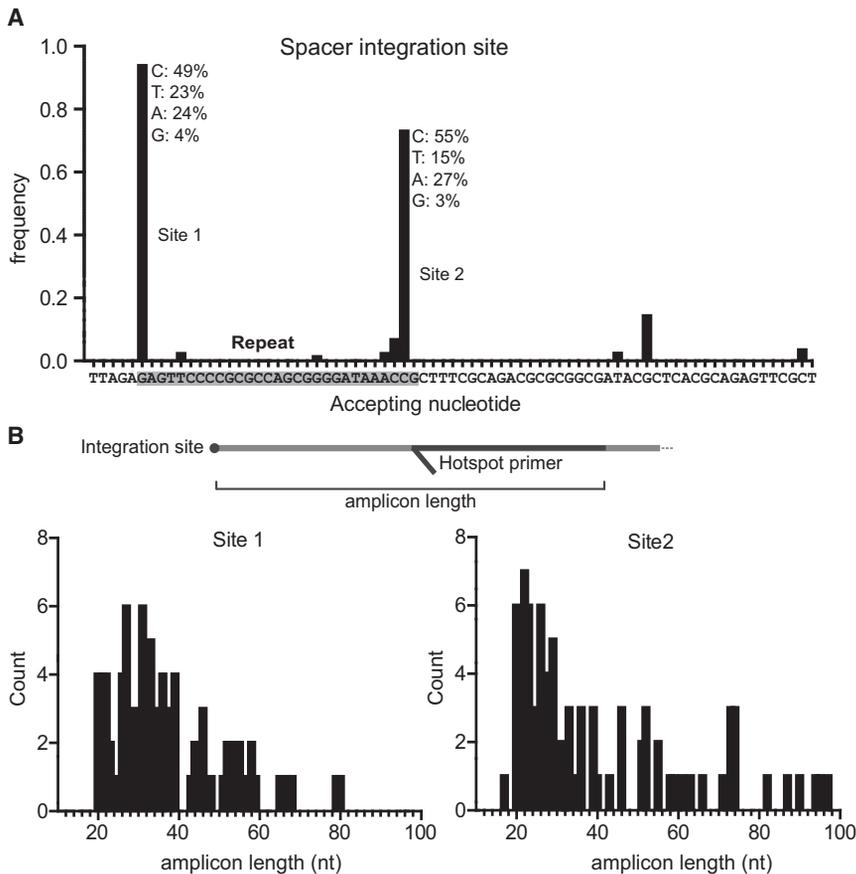
(B) Gel electrophoresis of half-site integration PCR on the basis of integration assay 2 (left) and oligo assay (right). PCR products representing integrations are indicated with an arrow. PCR products were specific to reactions containing all components. Lower running PCR products are primer dimers (verified by sequencing).

We found that the integration reaction is very precise for the two correct integration sites in the repeat (site 1 and site 2), and we observed that the integrated fragments most often were the result of a 3' cytosine coupling reaction. In vivo, however, only the integration of a CTT-containing fragment at site 2 would lead to a functional spacer targeting a protospacer with PAM (Figure 7), while CTT integration at site 1 would result in “flipped” spacers (Shmakov et al., 2014). Using a selective PCR strategy, we detected primed spacer acquisition events at both integration sites, and we identified that DNA fragments are integrated

fragment size, both strands are cleaved by Cas3 with the same specificity, enriching the 3' ends of the fragments for stretches of thymines. Contrary to the CTT requirements for spacer integration, it is known that Cascade tolerates five different PAM sequences (i.e. CTT, CTA, CCT, CTC, and CAT) for direct interference (Fineran et al., 2014; Leenay et al., 2016). However, the vast majority of new spacers (97%) resulting from primed acquisition carry CTT PAM sequences (Shmakov et al., 2014). This further supports the idea that spacer precursors with CTT-ends are selected non-randomly by the Cas1-2 complex from pools of Cas3 breakdown fragments and further trimmed to a 3' C (Wang et al., 2015). These are then coupled to the repeat by nucleophilic attack of the 3'-OH (Nuñez et al., 2014; Rollie et al., 2015). The T-dependent target DNA cleavage specificity of Cas3 further enhances the production of precursors that fit the requirements of new spacers by creating a pool of DNA fragments with the correct size and correct 3' ends. The interference phase of CRISPR immunity is therefore effectively coupled to the adaptation phase, providing positive feedback about the presence of an invader.

It was previously reported that a dinucleotide motif (AA) at the 3' end of a spacer increases the efficiency of naive spacer acquisition (Yosef et al., 2013). We did not observe this motif at the expected distance from the end in the Cas3 DNA degradation fragments, suggesting that Cas3 does not take the AA motif into account when generating spacer precursors.

in both orientations. In type I-E CRISPR-Cas systems, primed spacer acquisitions display a typical 9:1 strand bias for the acquisition of spacers targeting the same strand of DNA as the spacer causing priming (Datsenko et al., 2012; Swarts et al., 2012). This suggests that in vivo, other factors might be involved in further increasing the accuracy of functional spacer integration. This includes the formation of supercomplexes between various Cas proteins (i.e. Cascade, Cas3, Cas1-2) (Plagens et al., 2012; Redding et al., 2015; Richter et al., 2014) and the involvement of non-Cas host proteins such as PriA, RecG, and IHF (Ivančić-Baće et al., 2015; Nuñez et al., 2016). IHF ensures that the first integration event takes place at the leader-proximal end of the repeat (site 1) and might be involved in ensuring that the PAM cytosine gets integrated at the leader-distal end (site 2). Supercomplex formation during precursor generation may lead to the selection of fragments from the target strand containing a CTT PAM at the 3' end. Although the length of the observed integration amplicons is centered around 20–40 nt, we also find amplicons of up to 100 nt. In vivo, *E. coli* integrates fragments of 33 nt in length. We speculate that trimming of the precursor to 33 nt in length occurs after half-site integration and before formation of the stable integration intermediate (Figure 7). Despite the mechanisms that lower erroneous integration of new spacers, it is likely that natural selection of functional spacers in vivo also plays a role in the



**Figure 6. Sequencing Analysis of Spacer Integration**

(A) Frequencies of exact integration locations for integration at site 1 (gray bars) and site 2 (black bars) as determined by sequencing. The x axis gives the backbone nucleotide to which the spacer is coupled. Frequencies of coupled spacer nucleotides are indicated for the two canonical insertion locations.

(B) Top: schematic of integrated fragment and method of length determination. Bottom: length of the integration amplicon for site 1 and site 2.

the authors further propose that priming is usually not observed with fully matching protospacers because these targets are degraded too rapidly.

### Cut-Paste Spacer Acquisition

We have shown that priming reuses target DNA breakdown products as precursors for new spacers, providing support for a cut-and-paste mechanism of spacer selection (Wang et al., 2015). Compatible models have recently been proposed for naive spacer acquisition (Levy et al., 2015). It was shown that CRISPR adaptation is linked to dsDNA breaks that form at stalled DNA replication forks. Invading genetic elements often go through a phase of active DNA replication when they enter a host cell, and a replication-dependent

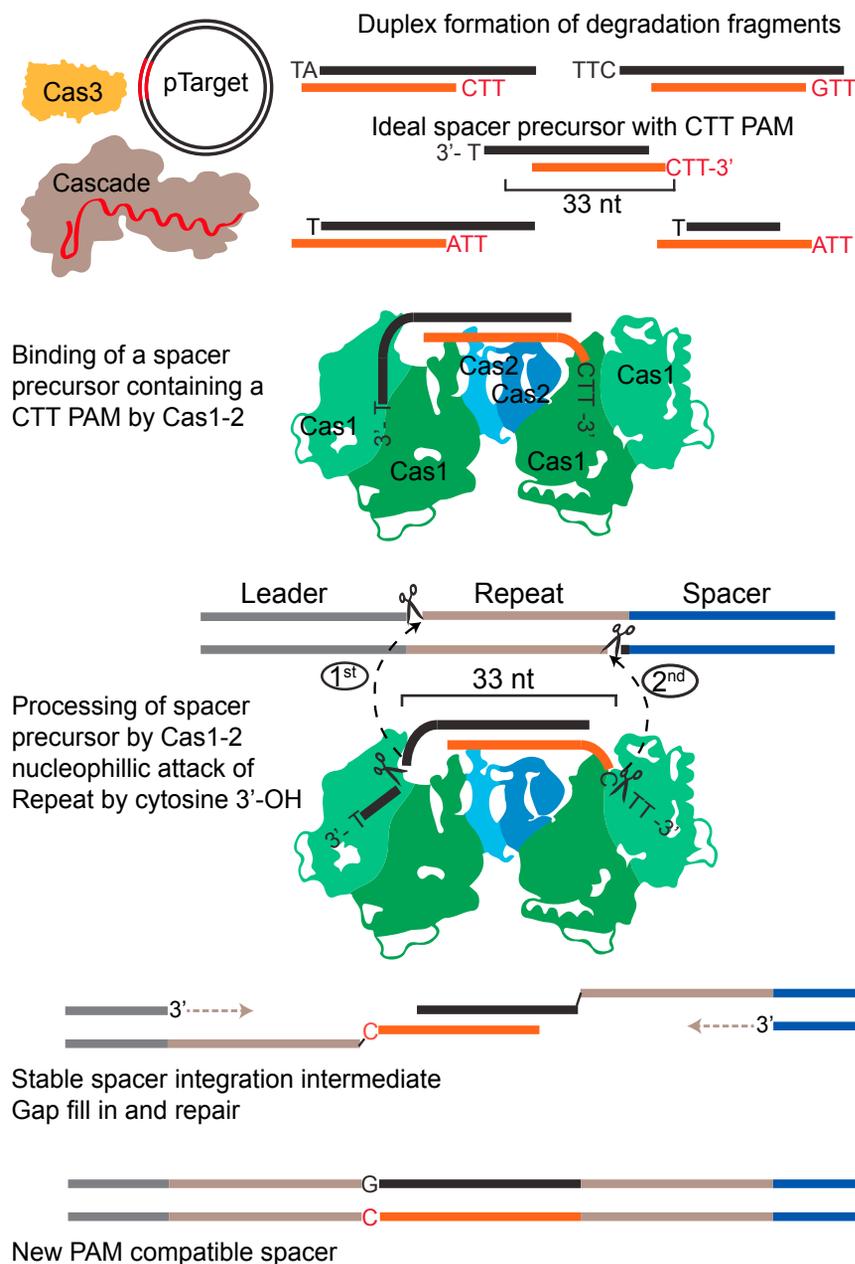
mechanism therefore helps the host primarily select spacers from the invading element. The RecBCD complex is key in this process as it repairs double-stranded breaks by first chewing back the ends of the DNA, creating fragments of tens to thousands of nucleotides (Amitai and Sorek, 2016). These fragments are thought to reanneal and serve as precursors for new spacers. Other studies have shown the direct involvement of crRNA-effector complexes in spacer selection. In the type I-F CRISPR-Cas system of *Pseudomonas aeruginosa*, the Csy complex is required for naive spacer acquisition (Vorontsova et al., 2015). Also, Cas9 in type II systems has a direct role in spacer acquisition (Heler et al., 2015; Wei et al., 2015). Both systems incorporate spacers very specifically from canonical PAM sites, suggesting that the Csy complex and Cas9 are directly involved in PAM recognition during spacer sampling.

spacers that end up being part of the first population of bacteria following a priming event. It was surprising that the bona fide target and several  $D^+P^-$  mutants did not show priming despite providing Cas3 degradation products. Furthermore, the degradation fragments of the bona fide target ( $D^+P^-$ ) were very similar to the fragments of the M4 target ( $D^+P^+$ ), which cannot explain the difference in priming behavior. We propose that these targets are degraded and cured from the cell too rapidly, giving the acquisition machinery insufficient time to generate new spacers. However, a low level of spacer integration might be taking place at undetectable levels even for the bona fide target, as has been observed previously (Swarts et al., 2012; Xue et al., 2015). In this case, cells with additional spacers do not have a selective growth advantage over cells without new spacers, as the plasmid is already effectively cleared from cells without new spacers. Mutant targets with intermediate levels of direct interference, however, are replicated and subject to interference over a longer time period, thereby providing more precursors, more time for spacer acquisition to occur, and therefore a greater selective growth advantage. Low levels of direct interference lead to a slow priming response because of the scarcity of spacer precursor molecules. While this paper was under review, another study showed that perfectly matching protospacers with canonical PAMs can indeed stimulate priming and that plasmid targeting is the stimulating factor (Semenova et al., 2016). In line with our findings,

mechanism therefore helps the host primarily select spacers from the invading element. The RecBCD complex is key in this process as it repairs double-stranded breaks by first chewing back the ends of the DNA, creating fragments of tens to thousands of nucleotides (Amitai and Sorek, 2016). These fragments are thought to reanneal and serve as precursors for new spacers. Other studies have shown the direct involvement of crRNA-effector complexes in spacer selection. In the type I-F CRISPR-Cas system of *Pseudomonas aeruginosa*, the Csy complex is required for naive spacer acquisition (Vorontsova et al., 2015). Also, Cas9 in type II systems has a direct role in spacer acquisition (Heler et al., 2015; Wei et al., 2015). Both systems incorporate spacers very specifically from canonical PAM sites, suggesting that the Csy complex and Cas9 are directly involved in PAM recognition during spacer sampling.

### Mutations in the Protospacer

In this study, we have focused on the effect of mutations in the protospacer on direct interference and priming, while maintaining the dominant interference permissive PAM CTT. Apart from underscoring the importance of the number of mutations and existence of a seed sequence (Semenova et al., 2011; Künne, et al., 2014; Wiedenheft et al., 2011; Xue et al., 2015), we uncover that for direct interference, pairing of the middle nucleotide in each five-nucleotide segment of the protospacer is disproportionately important and may represent a tipping point in the binding of a



target. None of the mutants showing direct interference carried mutations at these middle positions. Also, in a previously obtained list of approximately 3,300 triple mutants showing direct interference (Fineran et al., 2014), mutations at this position were underrepresented (Figure S3D). This suggests that pairing at the middle position of each segment may be important for continuation of the directional zipping process. This process starts at the PAM and leads to the formation of a canonical locked R-loop, which is required for Cas3 recruitment and target DNA degradation (Blosser et al., 2015; Redding et al., 2015; Rutkauskas et al., 2015; Szczelkun et al., 2014). We stress that we have used variants with CTT PAMs only, which can be engaged by Cascade in the canonical PAM-dependent binding mode

### Figure 7. Model of Primed Spacer Acquisition

Cleavage of a targeted plasmid during direct interference by Cascade and Cas3. Cleavage products are near spacer length and reanneal to form duplexes with 5' and/or 3' overhangs. The fragments are enriched for NTT sequences on their 3' ends. A fraction of the duplexes fulfills spacer precursor requirements: 3' overhangs, CTT at one 3' end, and a 33 nt distance between the C and the opposite 3' overhang. Cas1-2 binds spacer precursors with a preference for ideal duplexes as described above (Nuñez et al., 2015a; Wang et al., 2015). The precursor is processed by Cas1-2 to a length of 33 nt with 3' cytosine. In parallel to processing, 3' ends of the precursor perform a Cas1-2 catalyzed nucleophilic attack on the two integration sites of the repeat (Nuñez et al., 2015b; Rollie et al., 2015). Integration at the leader-repeat junction occurs first (Nuñez et al., 2016); subsequently, the PAM-derived 3' cytosine is integrated to ensure correct orientation and production of a functional spacer. A stable spacer integration intermediate is formed (Arslan et al., 2014). The gaps are filled in and repaired by the endogenous DNA repair systems, including DNA polymerase I (Ivančić-Baće et al., 2015).

(Blosser et al., 2015; Hayes et al., 2016; Redding et al., 2015; Rutkauskas et al., 2015) and can also trigger priming. It has become clear, however, that targets with mutations in the PAM display a broad spectrum of distinct characteristics depending on the chosen PAM, including a range of efficiencies of direct interference (Westra et al., 2013) and the reluctance to trigger efficient Cas3 target DNA degradation (Blosser et al., 2015; Hochstrasser et al., 2014; Mulepati and Bailey, 2013; Redding et al., 2015; Rutkauskas et al., 2015; Xue et al., 2015). In many cases, these PAMs still support the priming process (Datsenko et al., 2012; Fineran et al., 2014; Xue et al., 2015). Targets with highly disfavored PAMs (Hayes et al., 2016) are likely engaged in the

non-canonical PAM-independent binding mode (Blosser et al., 2015) and may require recruitment and translocation events of Cas1-2 and Cas3 proteins to initiate the target degradation needed to acquire new spacers.

### Conclusions

The findings presented here showcase the intricate PAM interplay of all Cas proteins in type I systems to update the CRISPR memory when receiving positive feedback about the presence of an invader. The robustness of priming is achieved by three components that co-evolved to work with PAM sequences: Cas3, producing spacer precursors enriched for correct PAM ends, Cas1-2 selecting PAM-compliant spacer precursors, and

Cascade efficiently recognizing targets with PAMs. This process stimulates the buildup of multiple spacers against an invader, preventing the formation of escape mutants (Datsenko et al., 2012; Richter et al., 2014; Swarts et al., 2012). When the original spacer triggers sufficiently strong interference, priming acquisition does not frequently occur. This prevents the unnecessary buildup of spacers and keeps the CRISPR array from getting too long. Any subsequent reduction in effectivity of the immune response by further mutations of the invader will in turn allow priming acquisition, restoring immunity.

## EXPERIMENTAL PROCEDURES

### Transformation and Plasmid Loss Assay

Both assays were carried out in *E. coli* KD263 cells, which have inducible cas gene expression. Expression was induced with 0.2% L-arabinose and 0.5 mM isopropyl  $\beta$ -D-1-thiogalactopyranoside where appropriate. Briefly, transformation efficiency was assessed by comparing colony-forming units (CFUs) of target plasmid transformations to CFUs of a control plasmid. Plasmid loss was assessed by loss of fluorescence in colonies, and spacer acquisition was determined by PCR of the CRISPR array. For details, see “Transformation Assay” and “Plasmid Loss Assay” in [Supplemental Experimental Procedures](#).

### Protein Purification

All proteins were expressed in BI21-AI cells. Cascade was purified as described previously (Jore et al., 2011). MBP-Cas3 was purified as described by Mulepati and Bailey (2013). The Cas1-2 complex was purified similarly to Cascade using affinity chromatography (see “Protein Purification” in [Supplemental Experimental Procedures](#)).

### Electrophoretic Mobility Shift Assays

Purified Cascade complex was incubated with plasmid at a range of molar ratios (1:1–100:1, Cascade:DNA). After electrophoresis, protein-bound and unbound DNA was quantified and the affinity calculated. For details, see “EMSA Assays” in [Supplemental Experimental Procedures](#).

### Cas3 DNA Degradation Assays

Cas3 DNA degradation activity was routinely tested by incubating 500 nM Cas3 with 4 nM M13mp8 single-stranded circular DNA. Plasmid-based assays were performed by incubating 70 nM Cas3 with 70 nM Cascade, 3.5 nM plasmid DNA. For details and activity quantification, see “Cas3 DNA Degradation Assays” in [Supplemental Experimental Procedures](#).

### Statistical Testing

We used a version of the empirical bootstrap method (Dekking, 2005) to test our data against the null hypothesis that observed behaviors ( $D^*P^*$ ) do not correlate with a particular sequence property. For details, see “Statistical Testing against the Null Hypothesis” in [Supplemental Experimental Procedures](#).

### In Vitro Acquisition Assays

Two types of assays were performed. First, Cas3 plasmid DNA degradation assays were carried out as described above, and the reaction products were incubated with Cas1-2 and pWUR869 in buffer R for 60 min. Second, target plasmid, Cascade, Cas3, Cas1-2, and pWUR869 were incubated in buffer R for 60 min. For details, see [Figure 5A](#) and “In Vitro Acquisition Assay” in [Supplemental Experimental Procedures](#).

### Next-Generation Sequencing

Plasmid degradation assays were performed as previously described. Three different targets were chosen: bona fide target plasmid (pWUR836) or M4 target plasmid (pWUR853) with 0.13 mM ATP and the m13mp8 assay as described above. Degradation fragments were processed for Illumina MiSeq sequencing (see “NGS Library Construction” in [Supplemental Experimental](#)

[Procedures](#)). For details on data processing, see “NGS Data Analysis” in [Supplemental Experimental Procedures](#).

## ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is ENA: PRJEB13999.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2016.07.011>.

## AUTHOR CONTRIBUTIONS

T.K. and S.J.J.B. designed research. T.K., S.N.K., J.W.B., A.I.M.V., and W.R.M. performed research. T.K., M.S.-D., S.N.K., M.K., M.D., and S.J.J.B. analyzed data. T.K. and S.J.J.B. wrote the paper with input from all authors.

## ACKNOWLEDGMENTS

This work was financially supported by an LS6 ERC starting grant (639707) and NWO VIDI grant to S.J.J.B. (864.11.005). S.J.J.B. and M.D. were supported by a TU Delft startup grant. This work was supported by the Netherlands Organization for Scientific Research (NWO/OCW), as part of the Frontiers in Nanoscience program. We thank Sebastian Hornung for depositing the DNA sequencing data set.

Received: May 20, 2016

Revised: July 1, 2016

Accepted: July 15, 2016

Published: August 18, 2016

## REFERENCES

- Almendros, C., and Mojica, F.J. (2015). Exploring CRISPR interference by transformation with plasmid mixtures: identification of target interference motifs in *Escherichia coli*. *Methods Mol. Biol.* *1311*, 161–170.
- Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. *Nat. Rev. Microbiol.* *14*, 67–76.
- Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, Ü. (2014). Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* *42*, 7884–7893.
- Bevington, S.L., Cauchy, P., Piper, J., Bertrand, E., Lalli, N., Jarvis, R.C., Gilding, L.N., Ott, S., Bonifer, C., and Cockerill, P.N. (2016). Inducible chromatin priming is associated with the establishment of immunological memory in T cells. *EMBO J.* *35*, 515–535.
- Blosser, T.R., Loeff, L., Westra, E.R., Vlot, M., Künne, T., Sobota, M., Dekker, C., Brouns, S.J., and Joo, C. (2015). Two distinct DNA binding modes guide dual roles of a CRISPR-Cas protein complex. *Mol. Cell* *58*, 60–70.
- Cady, K.C., Bondy-Denomy, J., Heussler, G.E., Davidson, A.R., and O’Toole, G.A. (2012). The CRISPR/Cas adaptive immune system of *Pseudomonas aeruginosa* mediates resistance to naturally occurring and engineered phages. *J. Bacteriol.* *194*, 5728–5738.
- Carter, J., and Wiedenheft, B. (2015). SnapShot: CRISPR-RNA-guided adaptive immune systems. *Cell* *163*, 260–260.e1.
- Charpentier, E., Richter, H., van der Oost, J., and White, M.F. (2015). Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* *39*, 428–441.
- Conrath, U., Beckers, G.J., Langenbach, C.J., and Jaskiewicz, M.R. (2015). Priming for enhanced defense. *Annu. Rev. Phytopathol.* *53*, 97–119.

- 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.
- Dekking, M. (2005). *A Modern Introduction to Probability and Statistics: Understanding Why and How* (London: Springer).
- Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* 434, 202–209.
- Fineran, P.C., Gerritzen, M.J., Suárez-Díez, M., Künne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U S A* 111, E1629–E1638.
- Gong, B., Shin, M., Sun, J., Jung, C.H., Bolt, E.L., van der Oost, J., and Kim, J.S. (2014). Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci. U S A* 111, 16359–16364.
- Han, D., and Krauss, G. (2009). Characterization of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* 583, 771–776.
- Hayes, R.P., Xiao, Y., Ding, F., van Erp, P.B., Rajashankar, K., Bailey, S., Wiedenheft, B., and Ke, A. (2016). Structural basis for promiscuous PAM recognition in type I-E Cascade from *E. coli*. *Nature* 530, 499–503.
- Heler, R., Marraffini, L.A., and Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* 93, 1–9.
- Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* 519, 199–202.
- Hochstrasser, M.L., Taylor, D.W., Bhat, P., Guegler, C.K., Sternberg, S.H., Nogales, E., and Doudna, J.A. (2014). CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U S A* 111, 6618–6623.
- Huo, Y., Nam, K.H., Ding, F., Lee, H., Wu, L., Xiao, Y., Farchione, M.D., Jr., Zhou, S., Rajashankar, K., Kurinov, I., et al. (2014). Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* 21, 771–777.
- Ivancić-Baće, 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.
- Jackson, R.N., Golden, S.M., van Erp, P.B., Carter, J., Westra, E.R., Brouns, S.J., van der Oost, J., Terwilliger, T.C., Read, R.J., and Wiedenheft, B. (2014). Structural biology. Crystal structure of the CRISPR RNA-guided surveillance complex from *Escherichia coli*. *Science* 345, 1473–1479.
- Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, U., Wurm, R., Wagner, R., et al. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* 18, 529–536.
- Künne, T., Swarts, D.C., and Brouns, S.J. (2014). Planting the seed: target recognition of short guide RNAs. *Trends Microbiol.* 22, 74–83.
- Künne, T., Westra, E.R., and Brouns, S.J. (2015). Electrophoretic mobility shift assay of DNA and CRISPR-Cas ribonucleoprotein complexes. *Methods Mol. Biol.* 1311, 171–184.
- Kurtz, J., and Franz, K. (2003). Innate defence: evidence for memory in invertebrate immunity. *Nature* 425, 37–38.
- Leenay, R.T., Maksimchuk, K.R., Slotkowski, R.A., Agrawal, R.N., Gomaa, A.A., Briner, A.E., Barrangou, R., and Beisel, C.L. (2016). Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell* 62, 137–147.
- 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.
- 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.
- Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736.
- Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. *Nature* 526, 55–61.
- Mulepati, S., and Bailey, S. (2013). In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* 288, 22184–22192.
- Mulepati, S., Héroux, A., and Bailey, S. (2014). Structural biology. Crystal structure of a CRISPR RNA-guided surveillance complex bound to a ssDNA target. *Science* 345, 1479–1484.
- Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* 21, 528–534.
- Nuñez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N., and Doudna, J.A. (2015a). Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* 527, 535–538.
- Nuñez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* 519, 193–198.
- Nuñez, J.K., Bai, L., Harrington, L.B., Hinder, T.L., and Doudna, J.A. (2016). CRISPR immunological memory requires a host factor for specificity. *Mol. Cell* 62, 824–833.
- Plagens, A., Tjaden, B., Hagemann, A., Randau, L., and Hensel, R. (2012). Characterization of the CRISPR/Cas subtype I-A system of the hyperthermophilic crenarchaeon *Thermoproteus tenax*. *J. Bacteriol.* 194, 2491–2500.
- Redding, S., Sternberg, S.H., Marshall, M., Gibb, B., Bhat, P., Guegler, C.K., Wiedenheft, B., Doudna, J.A., and Greene, E.C. (2015). Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-Cas system. *Cell* 163, 854–865.
- Reeks, J., Naismith, J.H., and White, M.F. (2013). CRISPR interference: a structural perspective. *Biochem. J.* 453, 155–166.
- Richter, C., Dy, R.L., McKenzie, R.E., Watson, B.N., Taylor, C., Chang, J.T., McNeil, M.B., Staals, R.H., 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.
- Rollie, C., Schneider, S., Brinkmann, A.S., Bolt, E.L., and White, M.F. (2015). Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *eLife* 4, 4.
- Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksny, V., and Seidel, R. (2015). Directional R-loop formation by the CRISPR-Cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* 10, 1534–1543.
- Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type I-E CRISPR/Cas spacer acquisition in *E. coli*. *RNA Biol.* 10, 716–725.
- Schmid-Hempel, P. (2005). Evolutionary ecology of insect immune defenses. *Annu. Rev. Entomol.* 50, 529–551.
- Semenova, E., Jore, M.M., Datsenko, K.A., Semenova, A., Westra, E.R., Wanner, B., van der Oost, J., Brouns, S.J., and Severinov, K. (2011). Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U S A* 108, 10098–10103.
- Semenova, E., Savitskaya, E., Musharova, O., Strotskaya, A., Vorontsova, D., Datsenko, K.A., Logacheva, M.D., and Severinov, K. (2016). Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U S A* 113, 7626–7631.
- Shipman, S.L., Nivala, J., Macklis, J.D., and Church, G.M. (2016). Molecular recordings by directed CRISPR spacer acquisition. *Science*, Published online June 9, 2016. <http://dx.doi.org/10.1126/science.aaf1175>.

- Shmakov, S., Savitskaya, E., Semenova, E., Logacheva, M.D., Datsenko, K.A., and Severinov, K. (2014). Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* *42*, 5907–5916.
- Sinkunas, T., Gasiunas, G., Waghmare, S.P., Dickman, M.J., Barrangou, R., Horvath, P., and Siksnys, V. (2013). In vitro reconstitution of Cascade-mediated CRISPR immunity in *Streptococcus thermophilus*. *EMBO J.* *32*, 385–394.
- Sternberg, S.H., Richter, H., Charpentier, E., and Qimron, U. (2016). Adaptation in CRISPR-Cas systems. *Mol. Cell* *61*, 797–808.
- Swarts, D.C., Mosterd, C., van Passel, M.W., and Brouns, S.J. (2012). CRISPR interference directs strand specific spacer acquisition. *PLoS ONE* *7*, e35888.
- Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U S A* *111*, 9798–9803.
- van Houte, S., Ekroth, A.K., Broniewski, J.M., Chabas, H., Ashby, B., Bondy-Denomy, J., Gandon, S., Boots, M., Paterson, S., Buckling, A., and Westra, E.R. (2016). The diversity-generating benefits of a prokaryotic adaptive immune system. *Nature* *532*, 385–388.
- Vorontsova, D., Datsenko, K.A., Medvedeva, S., Bondy-Denomy, J., Savitskaya, E.E., Pougach, K., Logacheva, M., Wiedenheft, B., Davidson, A.R., Severinov, K., and Semenova, E. (2015). Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res.* *43*, 10848–10860.
- Wang, J., Li, J., Zhao, H., Sheng, G., Wang, M., Yin, M., and Wang, Y. (2015). Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* *163*, 840–853.
- Wei, Y., Terns, R.M., and Terns, M.P. (2015). Cas9 function and host genome sampling in type II-A CRISPR-Cas adaptation. *Genes Dev.* *29*, 356–361.
- Westra, E.R., van Erp, P.B., Künne, T., Wong, S.P., Staals, R.H., Seegers, C.L., 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.
- Westra, E.R., Semenova, E., Datsenko, K.A., Jackson, R.N., Wiedenheft, B., Severinov, K., and Brouns, S.J. (2013). Type I-E CRISPR-cas systems discriminate target from non-target DNA through base pairing-independent PAM recognition. *PLoS Genet.* *9*, e1003742.
- Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., Dickman, M.J., and Doudna, J.A. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. U S A* *108*, 10092–10097.
- Xue, C., Seetharam, A.S., Musharova, O., Severinov, K., Brouns, S.J., Severin, A.J., and Sashital, D.G. (2015). CRISPR interference and priming varies with individual spacer sequences. *Nucleic Acids Res.* *43*, 10831–10847.
- 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.
- Yosef, I., Shitrit, D., Goren, M.G., Burstein, D., Pupko, T., and Qimron, U. (2013). DNA motifs determining the efficiency of adaptation into the *Escherichia coli* CRISPR array. *Proc. Natl. Acad. Sci. USA* *110*, 14396–14401.
- Zhao, H., Sheng, G., Wang, J., Wang, M., Bunkoczi, G., Gong, W., Wei, Z., and Wang, Y. (2014). Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* *515*, 147–150.