Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation

Künne, Tim; Kieper, Sebastian N.; Bannenberg, Jasper W.; Vogel, Anne I M; Miellet, Willem R.; Klein, Misha; Depken, Martin; Suarez-Diez, Maria; Brouns, Stan J J

DOI
10.1016/j.molcel.2016.07.011

Publication date
2016

Document Version
Accepted author manuscript

Published in
Molecular Cell

Citation (APA)
Künne, T., Kieper, S. N., Bannenberg, J. W., Vogel, A. I. M., Miellet, W. R., Klein, M., Depken, M., Suarez-Diez, M., & Brouns, S. J. J. (2016). Cas3-Derived Target DNA Degradation Fragments Fuel Primed CRISPR Adaptation. *Molecular Cell*, *63*(5), 852-864. https://doi.org/10.1016/j.molcel.2016.07.011

Link to formal publication Molecular Cell: http://dx.doi.org/10.1016/j.molcel.2016.07.011

Classification: BIOLOGICAL SCIENCES: Microbiology

Cas3-derived target DNA degradation fragments fuel primed CRISPR adaptation

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

[1]Laboratory of Microbiology, Wageningen University, 6708 WE Wageningen, Netherlands.

[2]Kavli Institute of Nanoscience and Department of BioNanoscience, Delft University of Technology, 2629 HZ, Delft, The Netherlands

[3]Laboratory of Systems and Synthetic Biology, Wageningen University, 6708 WE Wageningen, Netherlands.

[4]Current address: Department of Biotechnology, NTNU, N-7491 Trondheim, Norway

Corresponding author: Brouns, S.J.J. (stanbrouns@gmail.com, +31 15 278 3920)

Keywords: CRISPR-Cas; Priming; Interference; adaptive immunity; Phage resistance; Cascade; Cas1; Cas2; Cas3; Spacer acquisition

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 around 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 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; Nunez et al., 2014; Nunez et al., 2015b; Rollie et al., 2015; Wang et al., 2015). Apart from priming, spacers can also be acquired in a naïve manner. During naïve acquisition the host acquires spacers from an invading DNA element that has not been catalogued 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). Despite knowing the genetic requirements for priming, 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 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). This new site is then used to acquire a new spacer from. 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. While 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 naïve 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 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 (Kunne 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, 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^-) (Figure 1A, B). 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 hours 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 ~3 hpi) and this decrease was accompanied by 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 naïve acquisition had occurred within 48 h 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 (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^+P^+), but that it does not strictly require direct interference as exemplified by the D^-P^+ 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 transformation efficiency to a co-transformed control plasmid (Almendros and Mojica, 2015). While the *bona fide* target plasmid exhibited a relative transformation efficiency that was 512x lower than the control plasmid (1/512), also mutants with up to two effective mutations gave rise to strongly decreased transformation efficiencies (1/16 to 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 (~1), suggesting a dominant role for the position of the mutations in the protospacer. Mutants with 4 or 5 effective mutations transformed as efficient 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 lead to rapid spacer acquisition, while low levels or the absence of direct interference lead to delayed spacer acquisition. This also confirms that late plasmid loss in the D^-P^+ 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^+P^-, D^+P^+, D^-P^+, and D^-P^- (Figure S1). While D^+P^- and D^+P^+ had either 2 or 3 effective mutations, the D^-P^+ mutants had 3 or 4 mutations and the D^-P^- mutants carried 3 or 5 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^+P^- and D^+P^+ 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^-P^- 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^+P^- and D^+P^+ groups (both 95% confidence, Figure S3B). Surprisingly, the analysis also revealed that groups showing direct interference (D^+P^-, D^+P^+) had no mutations at the third position of each segment (significantly lower than expected, 95% confidence), whereas D^-P^+ and D^-P^- 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 if 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 (Kunne et al., 2015). While 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^{-1}), while the mutants cover a range of ratios ranging from very weak binding (>0.008 nM^{-1}) to almost the same levels as the *bona fide* target (<0.1 nM^{-1}). D^-P^- mutants all cluster together with low ratios (<0.02 nM^{-1}), and 5 out of 8 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 (Figure 2B, Figure 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 to 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 higher average activity than D^-P^+ mutants, although there is overlap between the two groups. 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^+ vs 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 this end, we performed Cascade-mediated plasmid degradation assays with Cas3 and plasmids containing the *bona fide* target or M4 target. 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 (Figure S5A, B). Based on 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 from gel 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 around 30-70 nt (Figure 3B, Figure 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.* 32/33 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 see if 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 (Figure 3C,D and Figure 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 (Nunez 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 panel). Interestingly, when Cas1-2 was present in the reaction we observed twice as much nicking of plasmid pCRISPR, suggesting half site integration of DNA fragments into pCRISPR had occurred (Figure 4B, right panel) (Nunez et al., 2015b). The same pCRISPR nicking activity was observed using purified Cas3 degradation products (integration assay 2) indicating 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. Since we did not know the sequence 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) (Nunez 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, while 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 (Nunez et al., 2015b; Rollie et al., 2015), both integration sites show a preference for 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 to 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) 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. These spacer precursors are 30-100 nt long partially double stranded DNA molecules formed by fragmentation of the target DNA. 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 (Nunez et al., 2015a; Rollie et al., 2015; Shipman et al., 2016; Wang et al., 2015). 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 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, 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 (Nunez 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 naïve 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.

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 half site integrations initiating 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 from both the

target and non-target strand of the plasmid could be used for integration. 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 (Ivancic-Bace et al., 2015; Nunez et al., 2016). IHF ensures that the first integration event takes place at the leader-proximal end of the repeat (site 1) and would 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 length. We speculate that trimming of the precursor to 33 nt 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 spacers that end up being part of the first population of bacteria following a priming event.

It was surprising that 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 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 due to 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, 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 naïve spacer acquisition (Levy et al., 2015). It was shown that CRISPR adaptation is linked to double stranded DNA 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 to 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 naïve 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; Kunne, 2014 #298; Wiedenheft et al., 2011; Xue et al., 2015), we uncover that for direct interference pairing of the middle nucleotide in each 5-nucleotide segment of the protospacer is disproportionately important, and may represent a tipping point in the binding of a target. None of the 30 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 (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.

Conclusion

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 IPTG where appropriate. Briefly: Transformation efficiency was assessed by comparing 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, Plasmid loss assay* in **Supplemental Experimental Procedures.**

Protein purification. All proteins were expressed in Bl21-AI cells. Cascade was purified as described earlier (Jore et al., 2011). MBP-Cas3 was purified as described in (Mulepati and Bailey, 2013). The Cas1-2 complex was purified similar to Cascade using affinity chromatography (see *protein purification* in **Supplemental Experimental Procedures**)

EMSA 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. 1) Cas3 plasmid DNA degradation assays were carried out as described above, the reaction products were incubated with Cas1-2 and pWUR869 in buffer R for 60 min. 2) 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**). Sequencing data was deposited at the European Nucleotide Archive under the accession number PRJEB13999. For details on data processing, see *NGS Data analysis* in **Supplemental Experimental Procedures.**

Author contributions

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

Acknowledgements

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 start up 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 dataset.

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 in molecular biology *1311*, 161-170.

Amitai, G., and Sorek, R. (2016). CRISPR-Cas adaptation: insights into the mechanism of action. Nature reviews. Microbiology *14*, 67-76.

Arslan, Z., Hermanns, V., Wurm, R., Wagner, R., and Pul, U. (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., Kunne, 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. Journal of bacteriology *194*, 5728-5738.

Carter, J., and Wiedenheft, B. (2015). SnapShot: CRISPR-RNA-guided adaptive immune systems. Cell *163*, 260-260 e261.

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. Annual review of phytopathology *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. Nature Communications *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., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proceedings of the National Academy of Sciences of the United States of America *111*, E1629-1638.

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. Proceedings of the National Academy of Sciences of the United States of America *111*, 16359-16364.
Han, D., and Krauss, G. (2009). Characterization of the endonuclease SSO2001 from Sulfolobus solfataricus P2. FEBS Letters *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. Proceedings of the National Academy of Sciences of the United States of America *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.
Ivancic-Bace, I., Cass, S.D., Wearne, S.J., and Bolt, E.L. (2015). Different genome stability proteins underpin primed and naive 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, Ü., Wurm, R., and Wagner, R. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature structural & molecular biology *18*, 529-536.
Kunne, T., Swarts, D.C., and Brouns, S.J. (2014). Planting the seed: target recognition of short guide RNAs. Trends Microbiol *22*, 74-83.
Kunne, T., Westra, E.R., and Brouns, S.J. (2015). Electrophoretic Mobility Shift Assay of DNA and CRISPR-Cas Ribonucleoprotein Complexes. Methods in molecular biology *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. Nature reviews. Microbiology *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. The Journal of biological chemistry *288*, 22184-22192.
Mulepati, S., Heroux, 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.
Nunez, 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.
Nunez, 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.
Nunez, 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.
Nunez, 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.
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. The Biochemical journal *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*.
Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V., and Seidel, R. (2015). Directional R-Loop Formation by the CRISPR-Cas Surveillance Complex Cascade Provides Efficient Off-Target Site Rejection. Cell reports *10*, 1534-1543.

Savitskaya, E., Semenova, E., Dedkov, V., Metlitskaya, A., and Severinov, K. (2013). High-throughput analysis of type IE CRISPR/Cas spacer acquisition in E. coli. RNA biology *10*, 0--1.
Schmid-Hempel, P. (2005). Evolutionary ecology of insect immune defenses. Annual review of entomology *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. Proceedings of the National Academy of Sciences *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. Proceedings of the National Academy of Sciences of the United States of America.DOI: 10.1073/pnas.1602639113
Shipman, S.L., Nivala, J., Macklis, J.D., and Church, G.M. (2016). Molecular recordings by directed CRISPR spacer acquisition. Science. DOI: 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 Journal *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. Proceedings of the National Academy of Sciences of the United States of America *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.*, et al.* (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.*, et al.* (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., 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.
Westra, E.R., van Erp, P.B., Kunne, 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.
Wiedenheft, B., van Duijn, E., Bultema, J.B., Waghmare, S.P., Zhou, K., Barendregt, A., Westphal, W., Heck, A.J., Boekema, E.J., and Dickman, M.J. (2011). RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. Proceedings of the National Academy of Sciences *108*, 10092-10097.
Xue, C., Seetharam, A.S., Musharova, O., Severinov, K., SJ, J.B., 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 research *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. Proceedings of the National Academy of Sciences of the United States of America *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.

Figures

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

585 behaviors that were observed: Rapid plasmid loss without spacer integration (D^+P^-), delayed plasmid loss and
586 spacer integration (D^+P^+), strongly delayed plasmid loss and spacer integration (D^-P^+), and no plasmid loss with
587 no spacer integration (D^-P^-). B) Summary of plasmid behavior of all mutants, showing timing of first plasmid loss
588 and time of first observable spacer integration. C) The relative transformation efficiency is plotted for all
589 mutant plasmids (fold change compared to co-transformed non-target plasmid, log2 scale). Bars are color
590 coded based on plasmid behavior classification. Error bars represent the standard error of the mean of
591 triplicate experiments. The positions of mutations are indicated schematically for each mutant (Pos1: Bottom,
592 Pos32: Top). Open ovals represent mutations on positions 6, 12, 18, 24, 30. Closed ovals represent mutations
593 outside of those positions (effective mutations). The amount of effective mutations is indicated above or below
594 the schematic. For a more detailed overview of the mutations, see Figure S1.

595 **Figure 2: EMSA and Cas3 activity assay.** A) Electrophoretic mobility shift assay (EMSA) of the mutant plasmid
596 set. The affinity ratio (Amplitude/K_d) is plotted for each mutant (see Table S3 for more details). Mutants are
597 separated by the previously made plasmid behavior classification. The mean and standard deviation for each
598 group are indicated. The *bona fide* target is abbreviated as WT. B) Cas3 DNA degradation activity assay of
599 mutant plasmid set. The initial Cas3 DNA cleavage rate [%/min] is plotted for each mutant. Mutants are
600 classified according to previously identified plasmid behavior. The mean and standard deviation for each group
601 are indicated. Individual gels for all activity assays can be found in Figure S4.

602 **Figure 3: Next generation sequencing analysis of Cas3 DNA degradation products.** A) Left: Schematic of R-loop
603 formed by binding of Cascade to dsDNA target. Right: Schematic showing the four distinct Cas3 cleavage sites
604 in dsDNA target. B) Length distribution of Cas3 DNA degradation fragments of M4 target. C) Heat map of
605 nucleotide frequencies around cleavage sites. The cleavage site is between position -1 and 1. Positions
606 indicated in black are on the fragments, positions indicated in grey are outside of fragments. D) Heat map of
607 dinucleotide frequencies around cleavage sites. Abundance of dinucleotides was measured in a shifting frame
608 within 4 nucleotides around the cleavage sites.

609 **Figure 4: *In vitro* spacer acquisition assays.** A) Illustration of the three types of assays performed. In the oligo
610 assay, pCRISPR is incubated with Cas1-2 and a spacer oligo (BG7415/6), leading to half site integration. In assay
611 1, pTarget and pCRISPR are incubated with Cascade, Cas3 and Cas1-2 for simultaneous degradation of pTarget

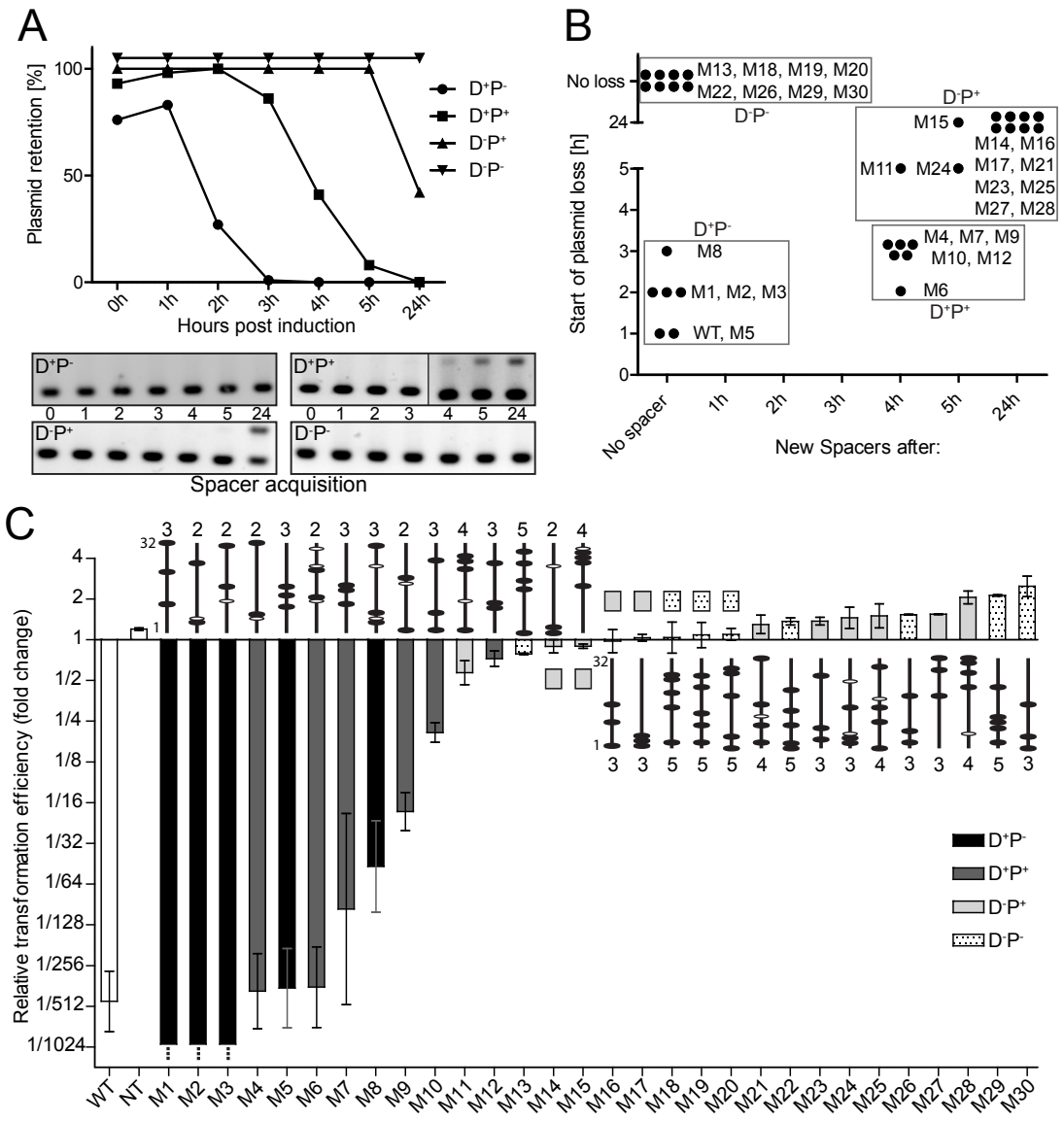
and half site integration into pCRISPR. In assay 2, pTarget is incubated with Cascade and Cas3 and the resulting DNA degradation products are then separately incubated with pCRISPR and Cas1-2. B) Gel electrophoresis of integration assay 1. The *bona fide* target is abbreviated as WT. Left gel, untreated; right gel, Proteinase K treated. Cas1-2 presence causes upwards shift of DNA. Original plasmids are supercoiled (SC), half site integration causes nicking of pCRISPR, resulting in the open circular conformation (OC).

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 based on frequently incorporated spacers (hotspots) *in vivo* (Fineran et al., 2014). B) Gel electrophoresis of half site integration PCR based on 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).

Figure 6: Sequencing analysis of spacer integration. A) Frequencies of exact integration locations for integration at site 1 (grey bars) and site 2 (black bars) as determined by sequencing. X-axis gives the backbone nucleotide to which the spacer is coupled. Frequencies of coupled spacer nucleotides are indicated for the 2 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.

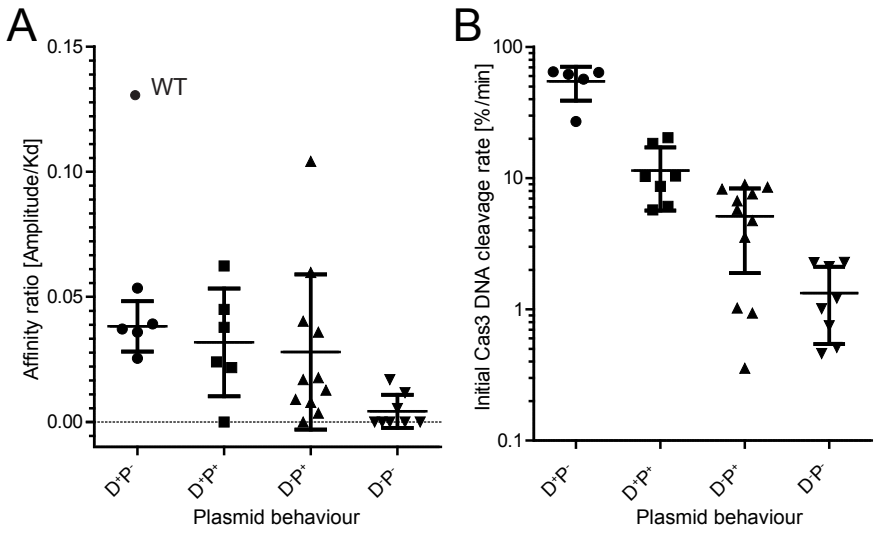
Figure 7: Model of primed spacer acquisition. Cleavage of a targeted plasmid during direct interference by Cascade and Cas3. Cleavage products are near-spacer length and reanneal to form duplexes with 5’ and/or 3’ overhangs. The fragments are enriched for NTT sequences on their 3’ ends. A fraction of the duplexes fulfils 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 (Nunez 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 (Nunez et al., 2015b; Rollie et al., 2015). Integration at the leader-repeat junction occurs first (Nunez et al., 2016), subsequently the PAM derived 3’ cytosine is integrated to assure 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 (Ivancic-Bace et al., 2015).

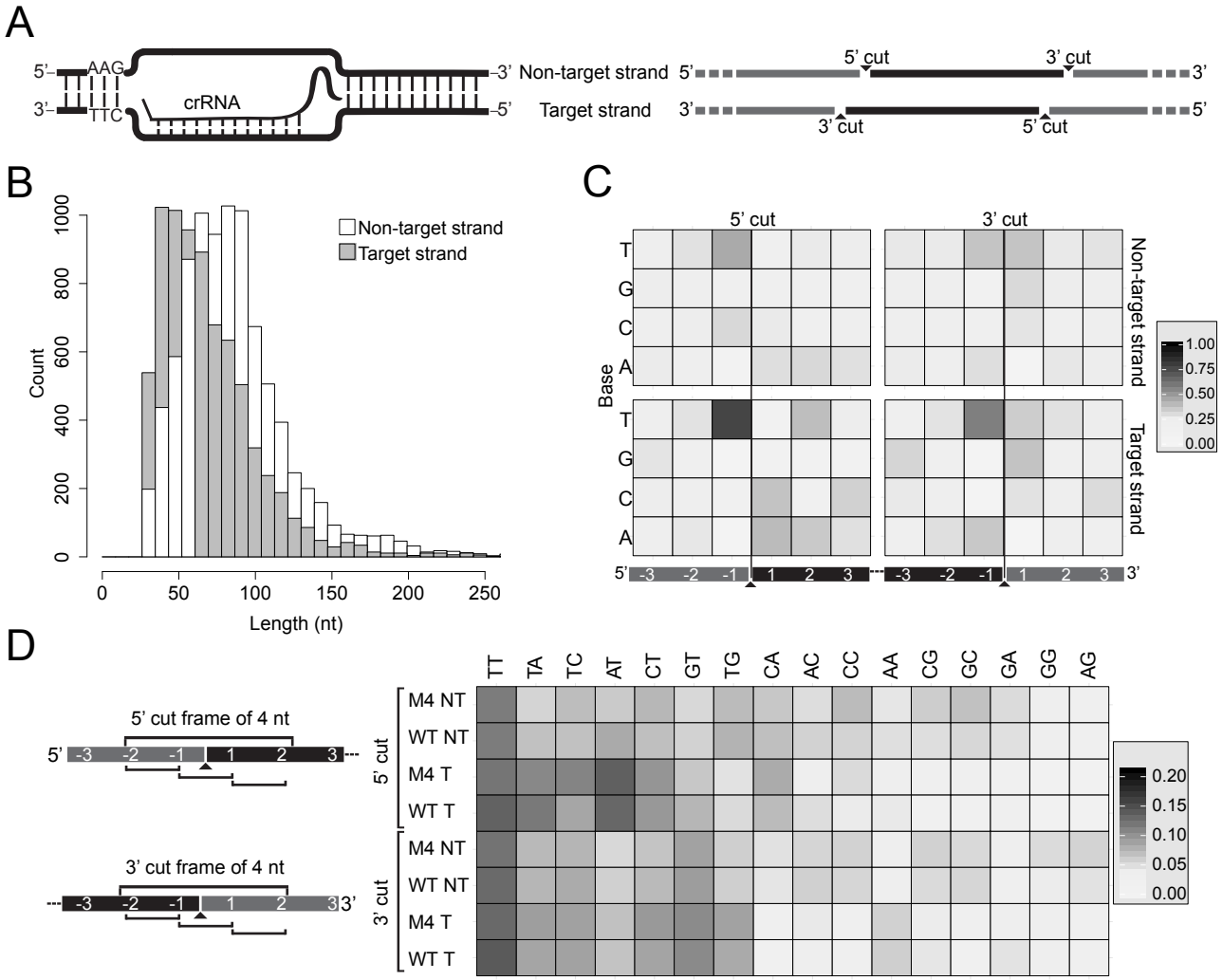


Künne et al
Figure 1

Figure2

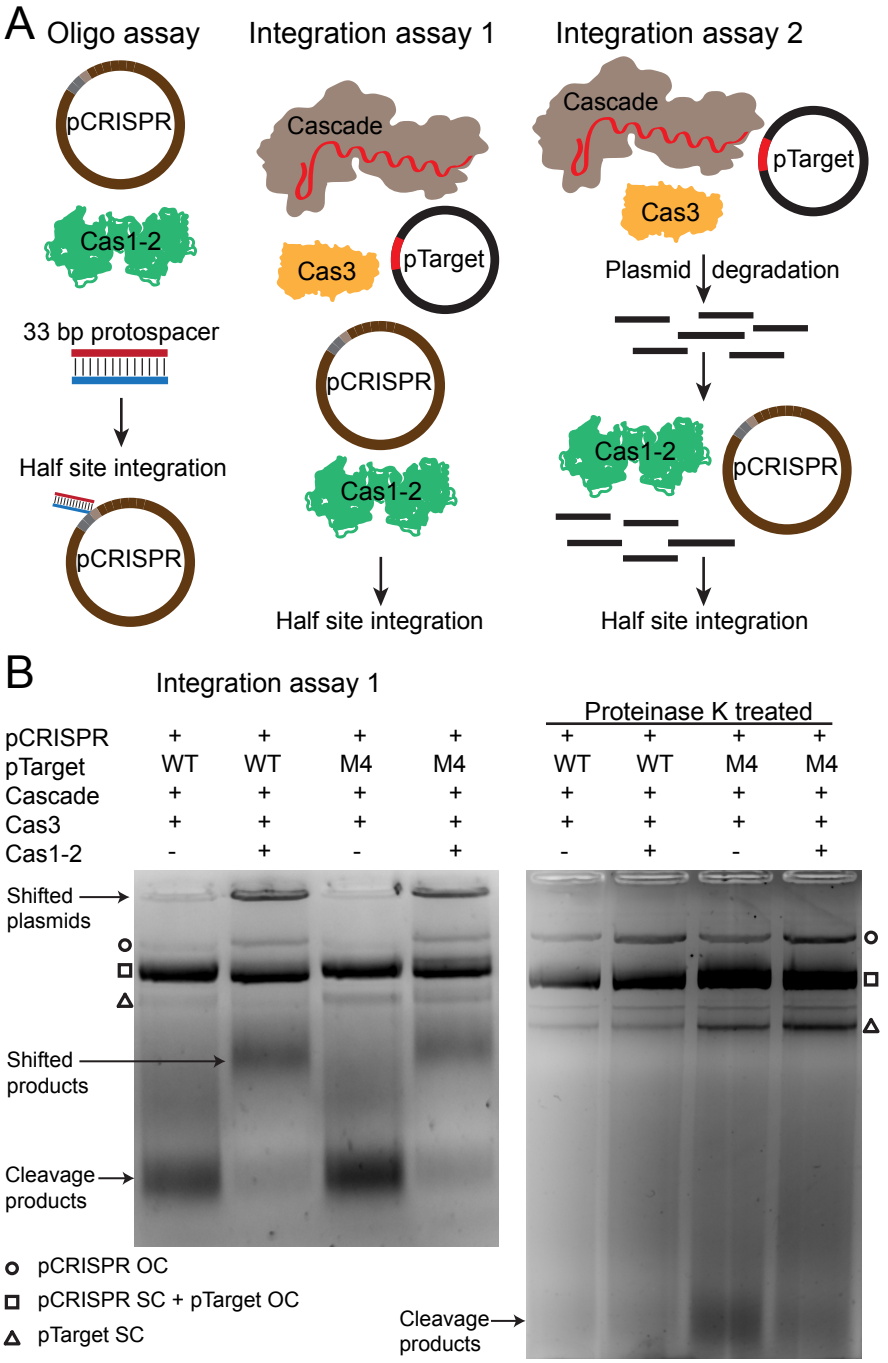


Künne et al
Figure 2



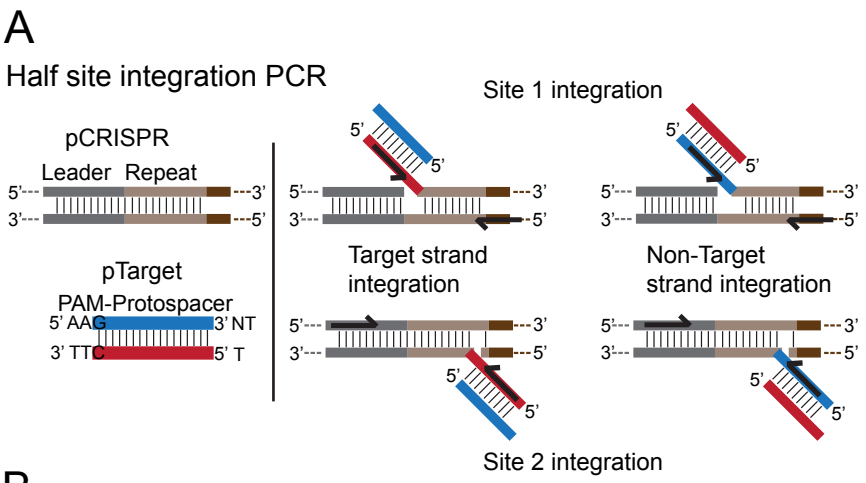
Künne et al
Figure 3

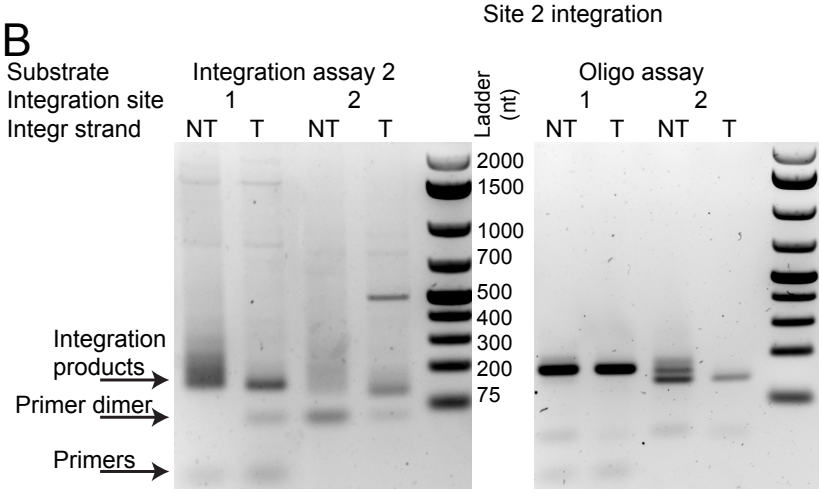
Figure4



Künne et al
Figure 4

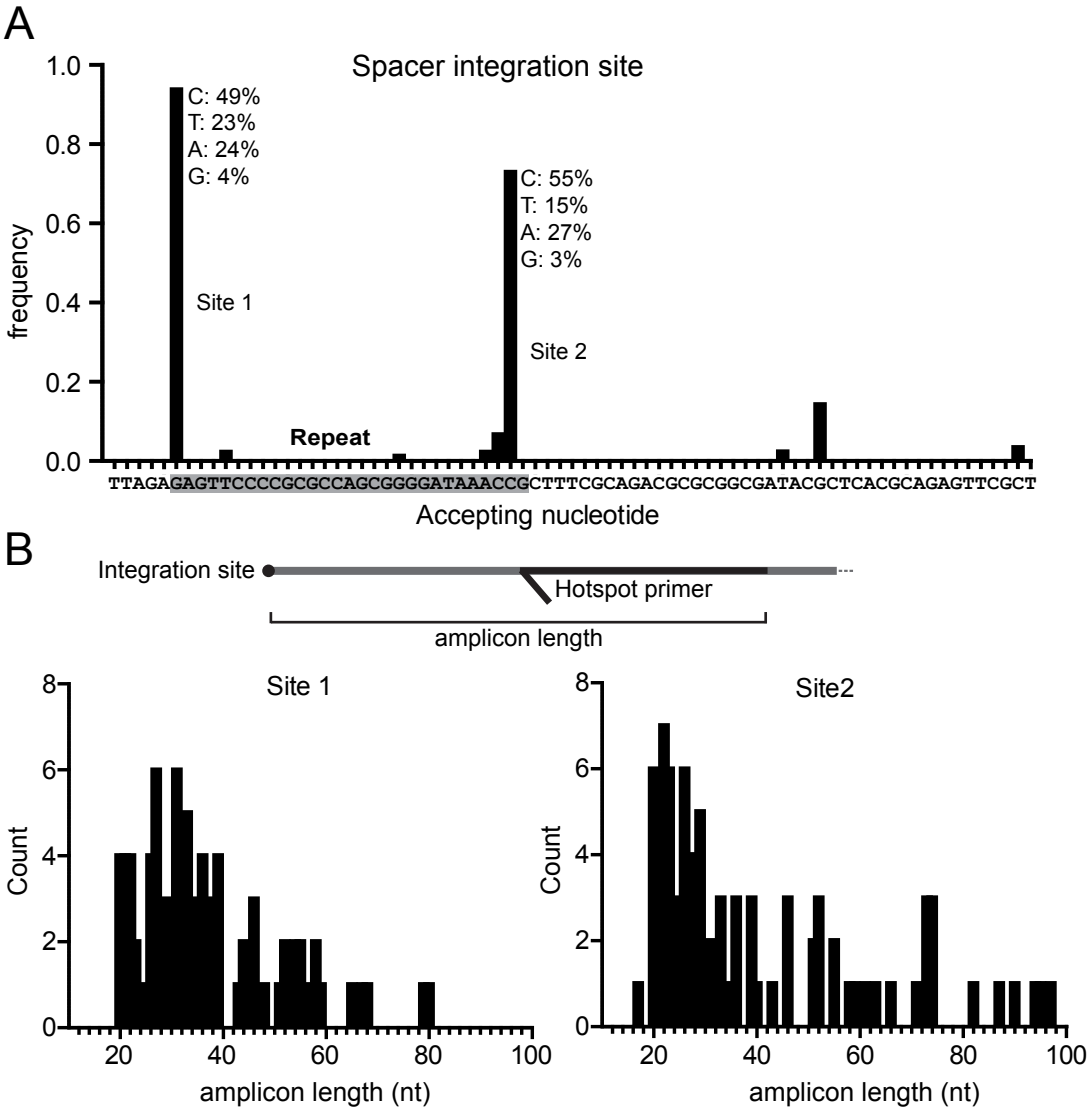
Figure5





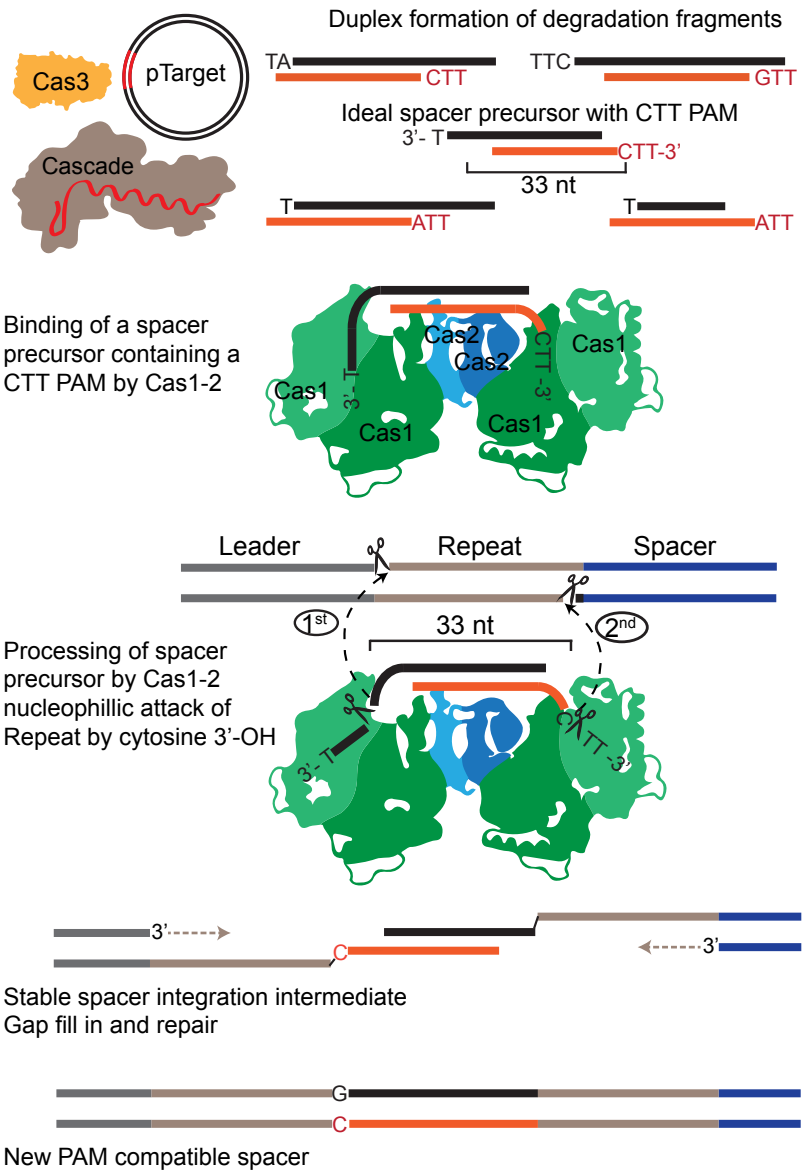
Künne et al
Figure 5

Figure6



Künne et al
Figure 6

Figure7



Künne et al
Figure 7

Supplemental Information

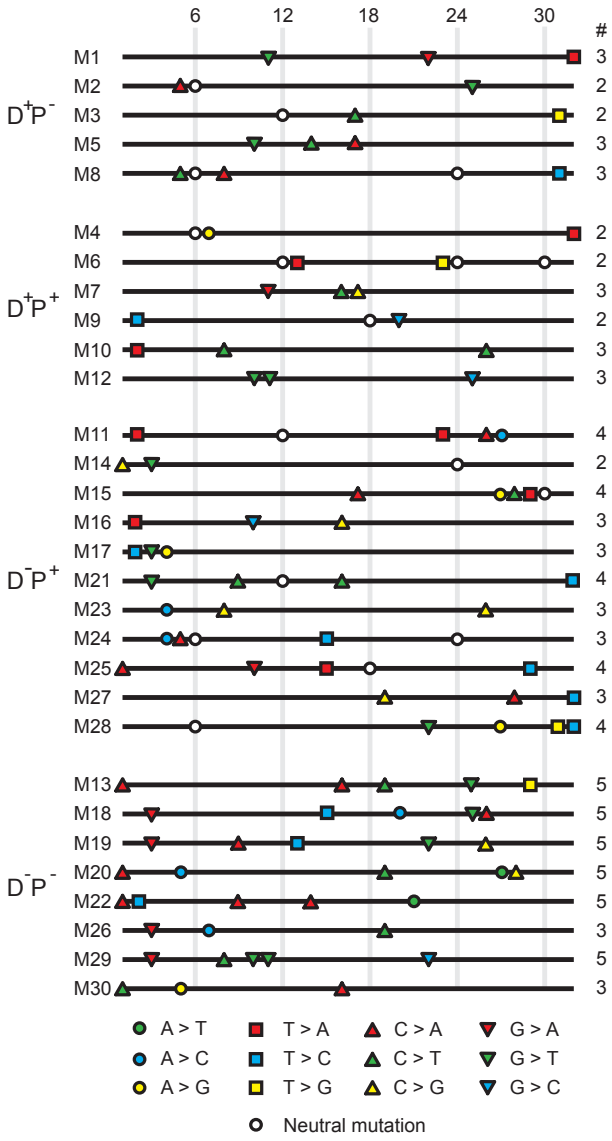
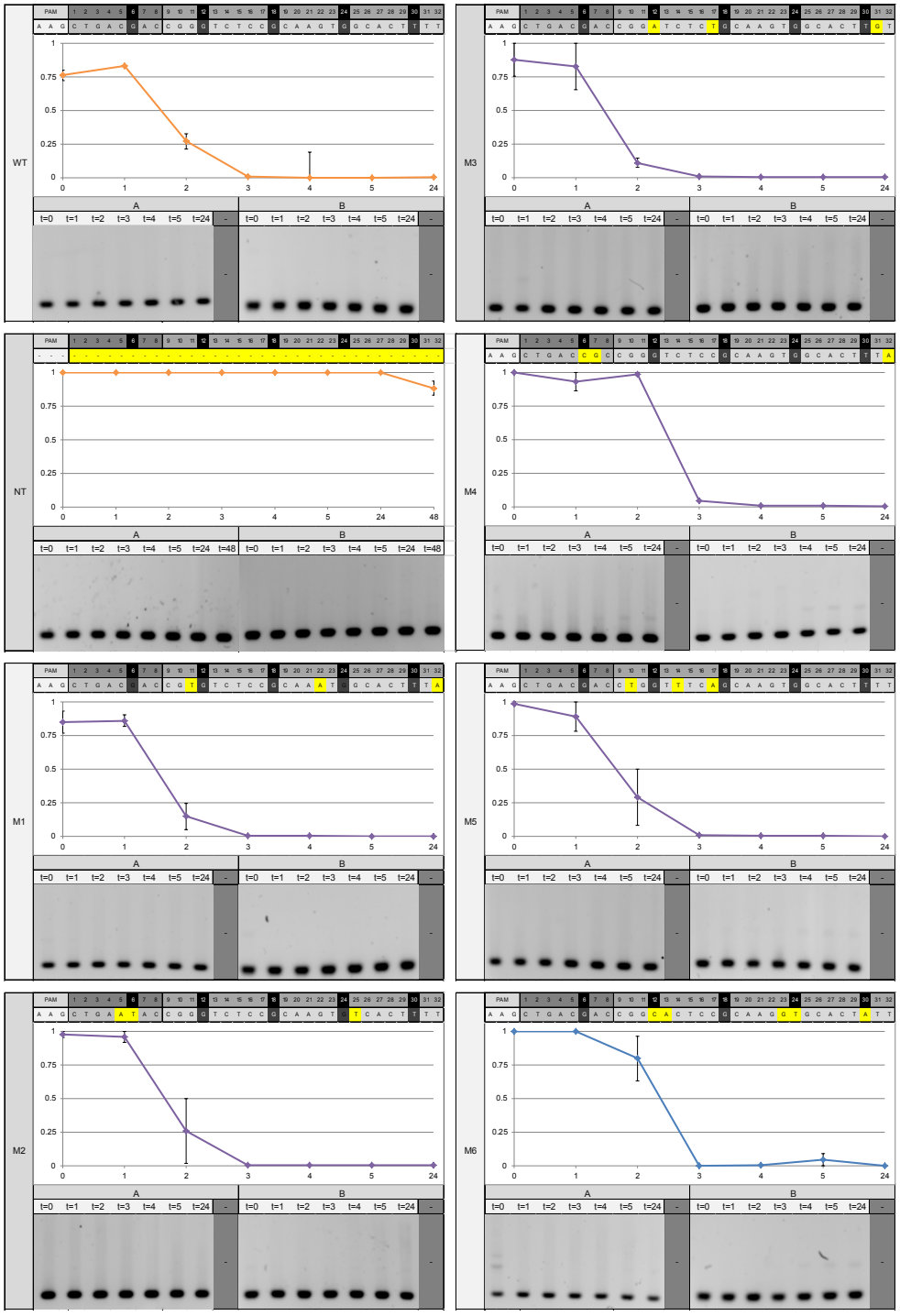
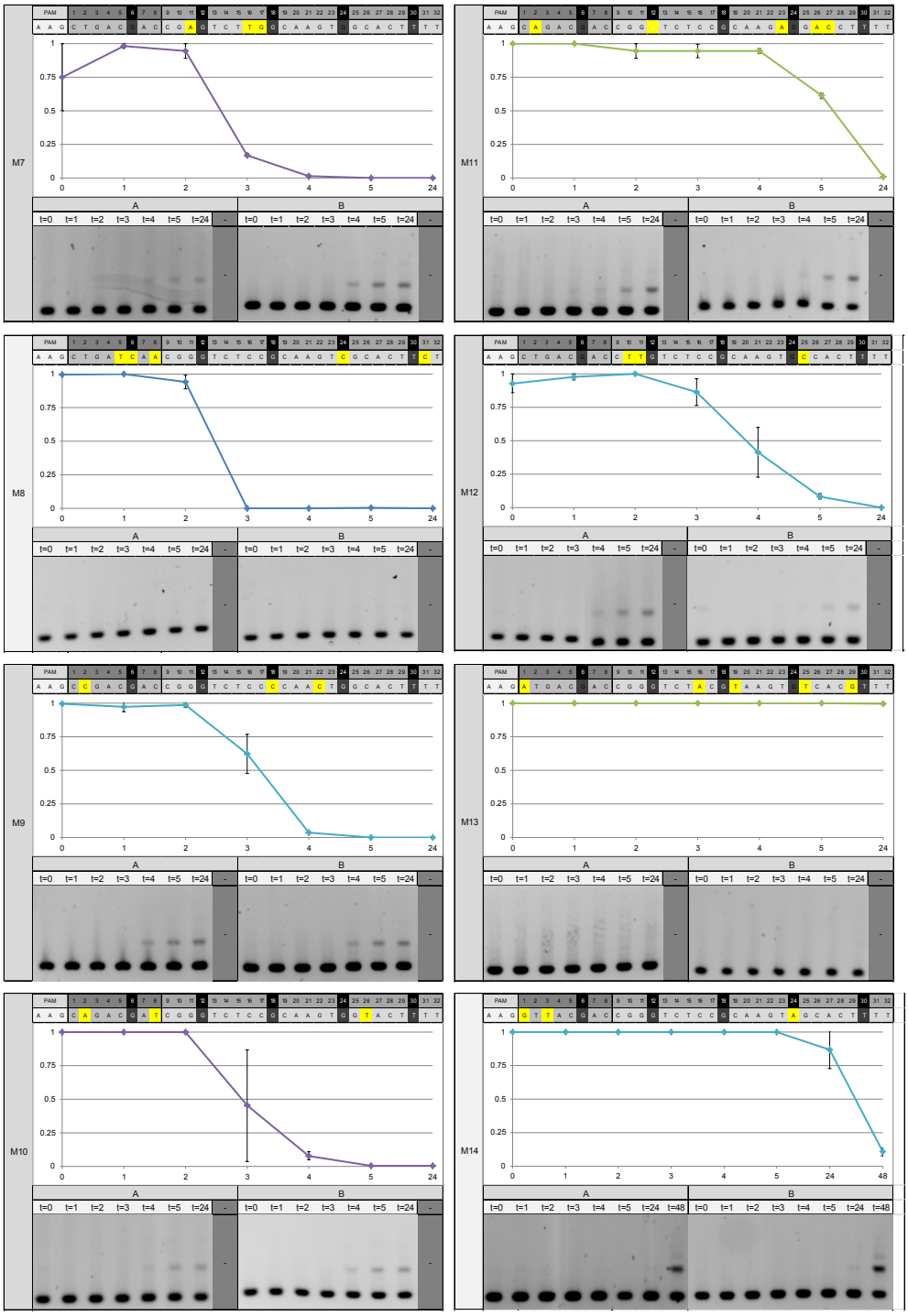
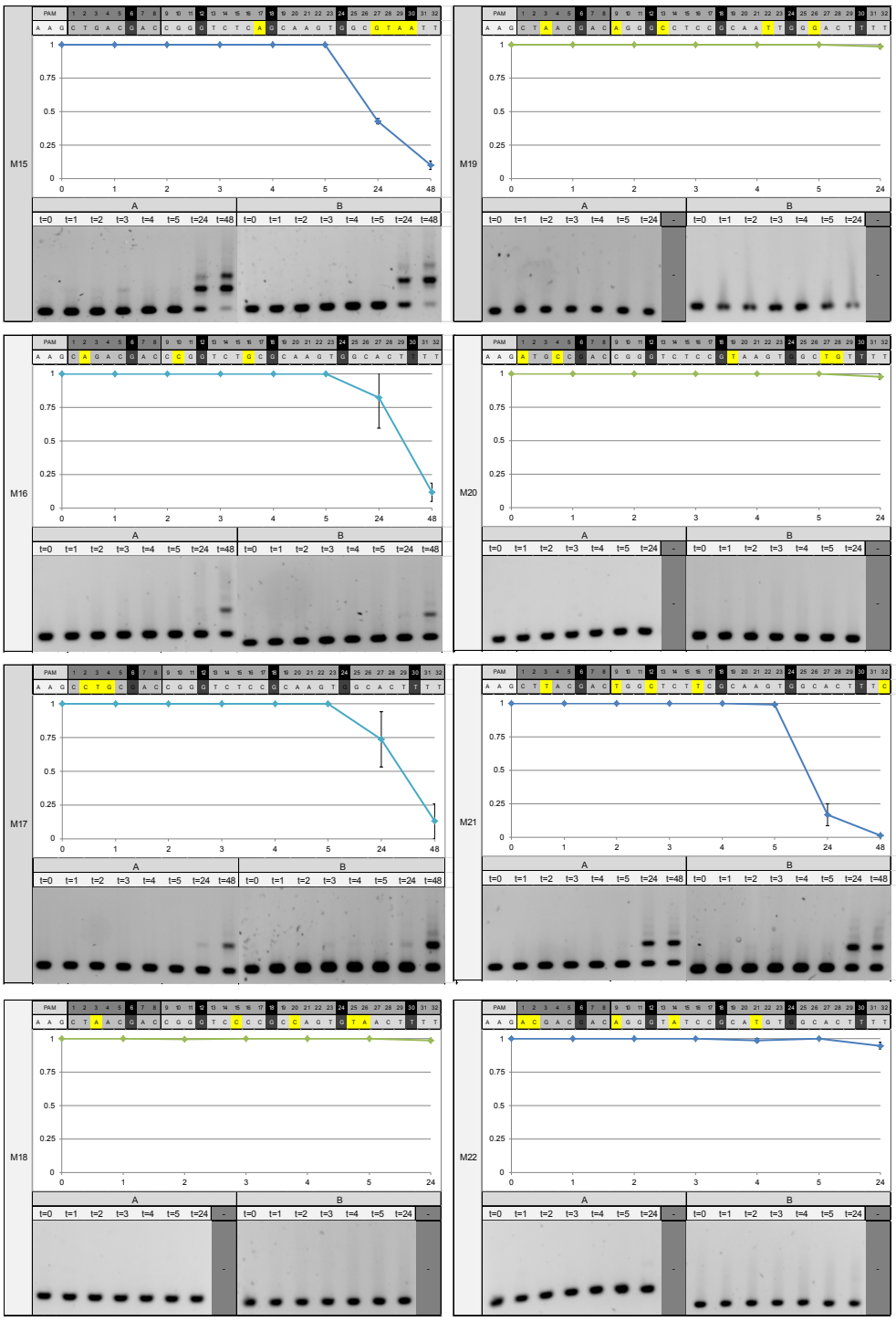


Figure S1: Overview of Protospacer8 mutants. 30 mutants of protospacer8 containing either 3 or 5 total mutations were used throughout the study. Mutations on positions 6, 12, 18, 24, 30 (empty circles) are not participating in base-pairing and are therefore not considered as effective mutations. Types of mutations are indicated by colored symbols. Mutants are separated into categories based on their behavior in plasmid loss assays (see also Figure 1B).


WT
M3
NT
M4
M1
M5
M2
M6
PAM
A
B
t=0 t=1 t=2 t=3 t=4 t=5 t=24 t=48


M7
M11
M8
M12
M9
M13
M10
M14
PAM
A
B
t=0 t=1 t=2 t=3 t=4 t=5 t=24 -
t=0 t=1 t=2 t=3 t=4 t=5 t=24 t=48 -


M15
M19
M16
M20
M17
M21
M18
M22
A
B
t=0 t=1 t=2 t=3 t=4 t=5 t=24 t=48

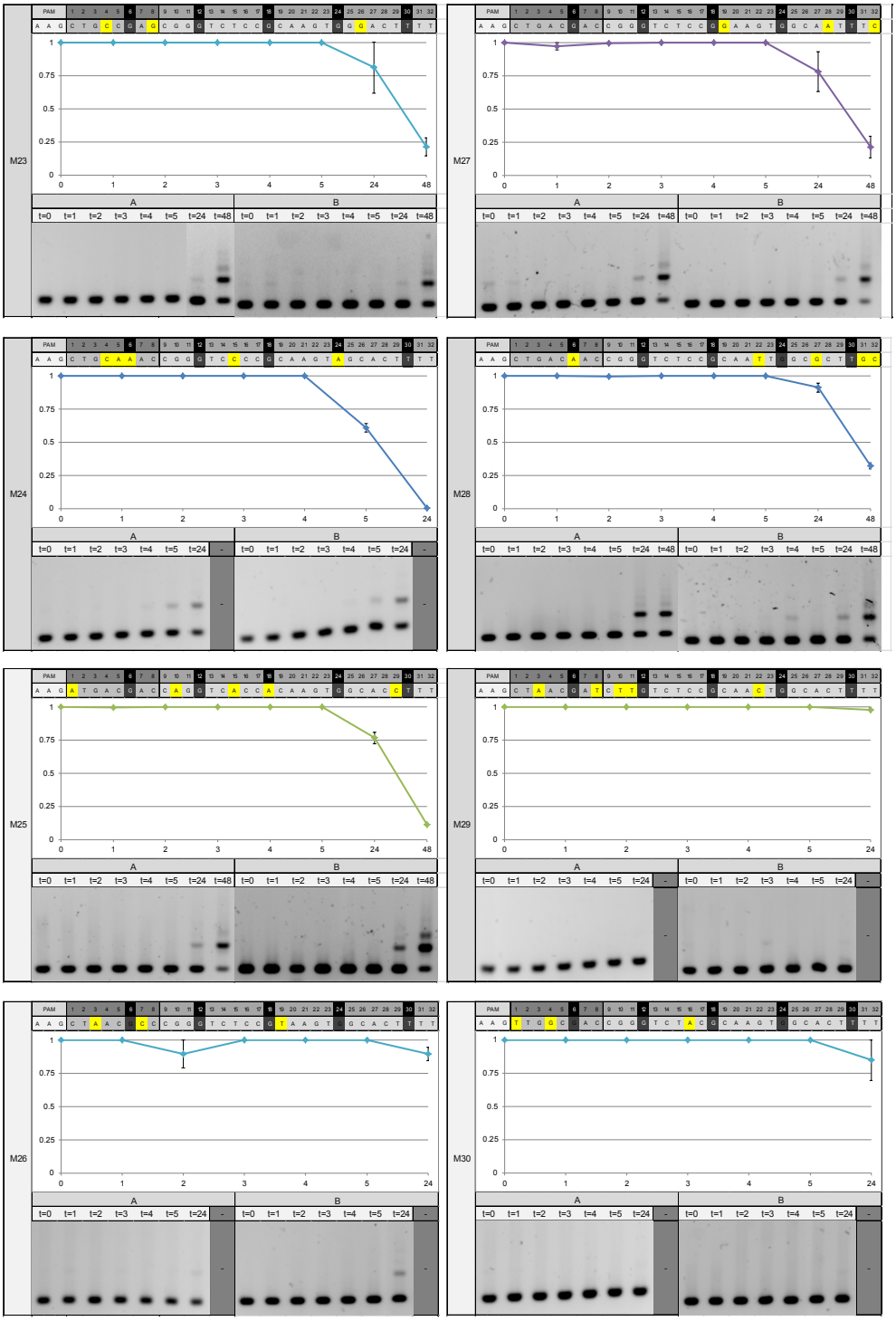
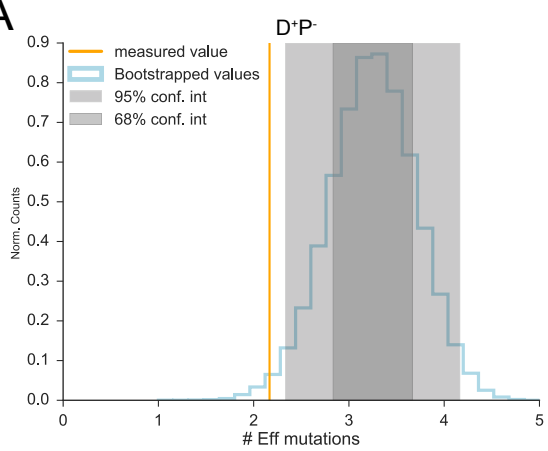
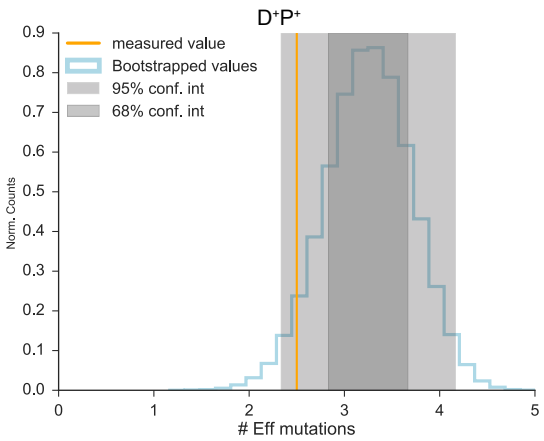
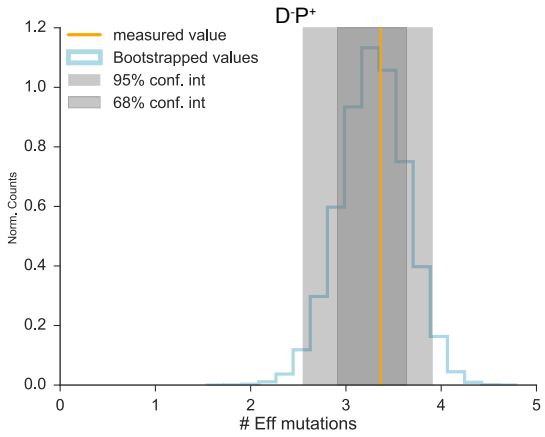

M23
M27
M24
M28
M25
M29
M26
M30
A
B
t=0 t=1 t=2 t=3 t=4 t=5 t=24 t=48

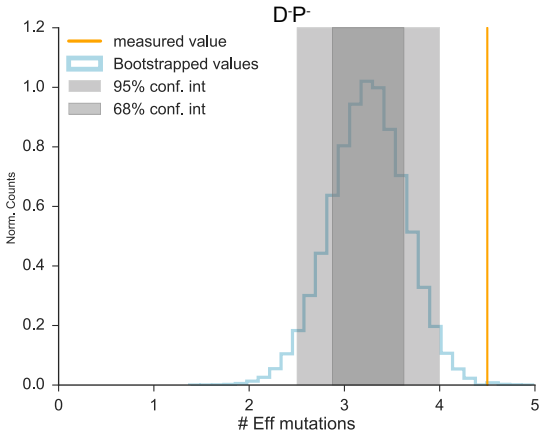
Figure S2. Related to Figure 1: Individual plasmid loss assays. Panels for each plasmid mutant with (top to bottom): Sequence with indicated mutations, plasmid loss curves from 0 h to 24 h or 48 h, duplicate of CRISPR PCR showing spacer acquisition. The bottom bands in the PCR gels represent the unextended array, higher bands represent the array with an extra spacer. Error bars in plasmid loss graphs represent the standard deviation of replicate experiments. The *bona fide* target is abbreviated as WT.

A

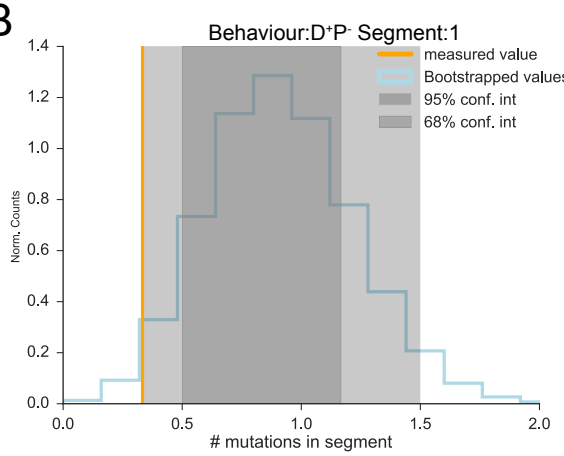

D+P-
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
Eff mutations

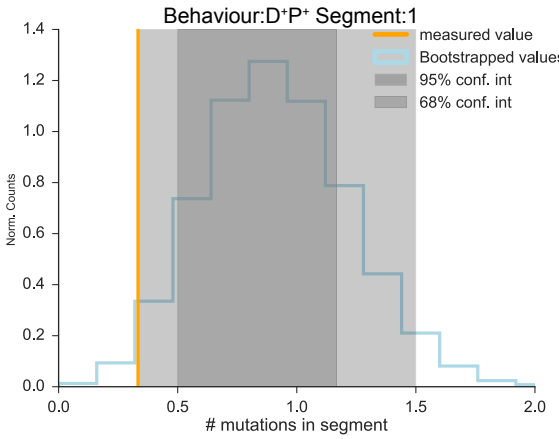

D+P+
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
Eff mutations

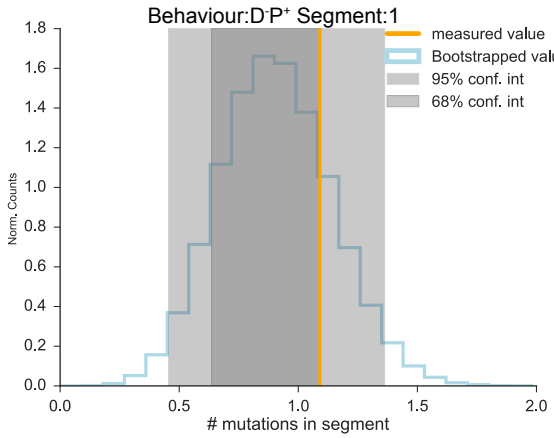

D-P+
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
Eff mutations

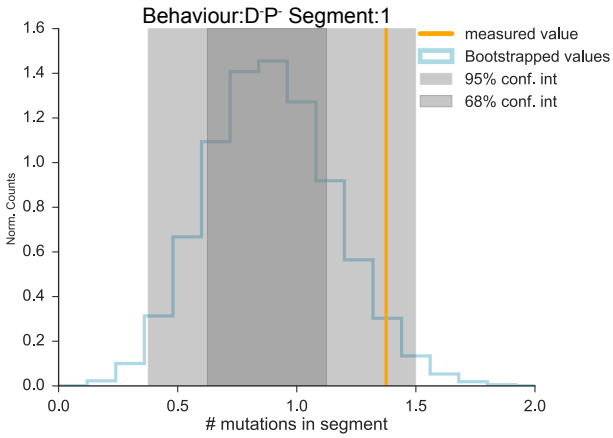

D-P-
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
Eff mutations

B

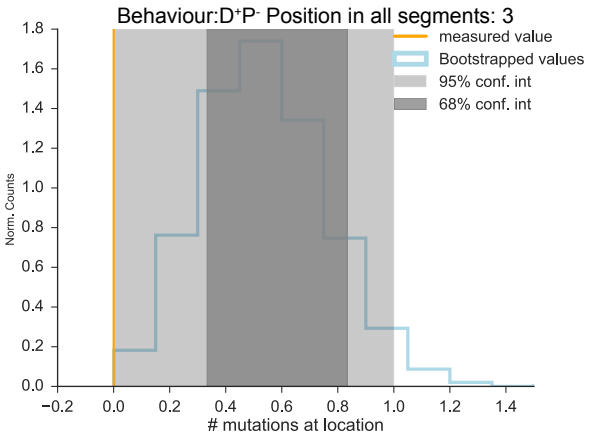

Behaviour:D+P- Segment:1
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations in segment

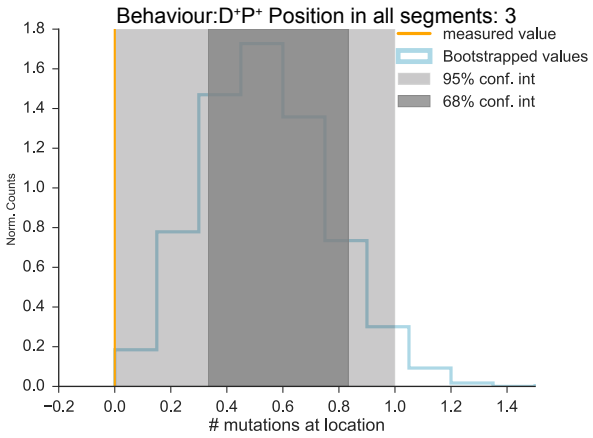

Behaviour:D+P+ Segment:1
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations in segment

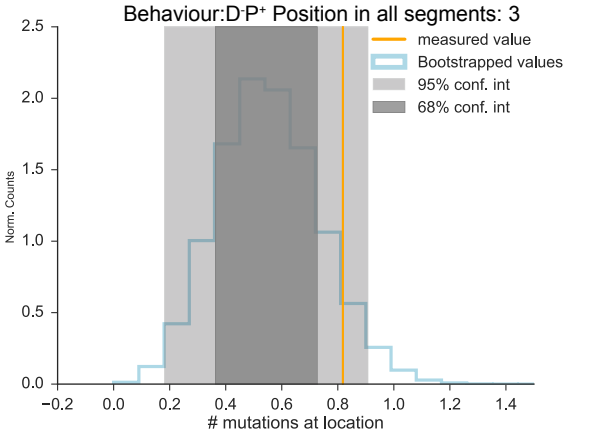

Behaviour:D-P+ Segment:1
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations in segment

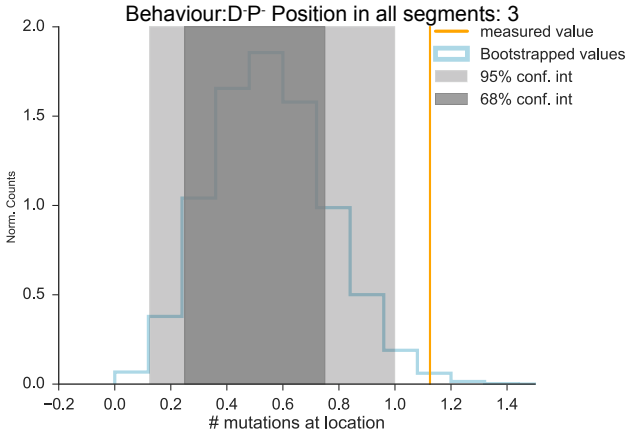

Behaviour:D-P- Segment:1
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations in segment

C


Behaviour:D+P- Position in all segments: 3
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations at location


Behaviour:D+P+ Position in all segments: 3
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations at location


Behaviour:D-P+ Position in all segments: 3
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations at location


Behaviour:D-P- Position in all segments: 3
measured value
Bootstrapped values
95% conf. int
68% conf. int
Norm. Counts
mutations at location

D

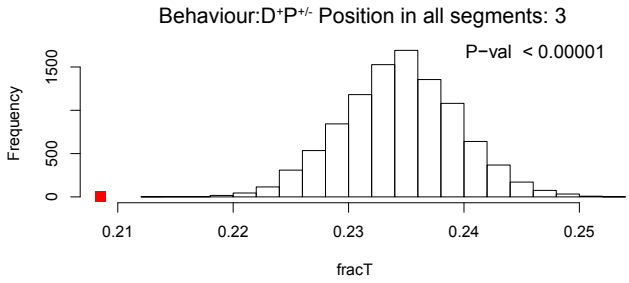

Behaviour:D+P+/- Position in all segments: 3
P−val < 0.00001
Frequency
fracT

Figure S3: Statistical pattern analysis of 30 mutants set. Three properties were analyzed separately for each group of plasmid behavior. The average of each behavioral group is indicated by the yellow vertical line. To test if the plasmid behavior depends on a certain property, for each property a distribution was made based on empirical bootstrapping of the whole set of 30 mutants (blue line). The 95% and 68% confidence intervals of each distribution are indicated by the light and dark grey boxes respectively. A) Average number of effective mutations. B) Average number of mutations in segment 1. C) Average number of mutations on position 3 within all segments combined. D) Average number of mutations on position 3 within all segments combined but the analysis was performed on a previously published large dataset (Fineran et al., 2014). From this dataset, mutants with 3 mutations (all canonical PAM) were analyzed. The average of the direct interference group is indicated by the red square.

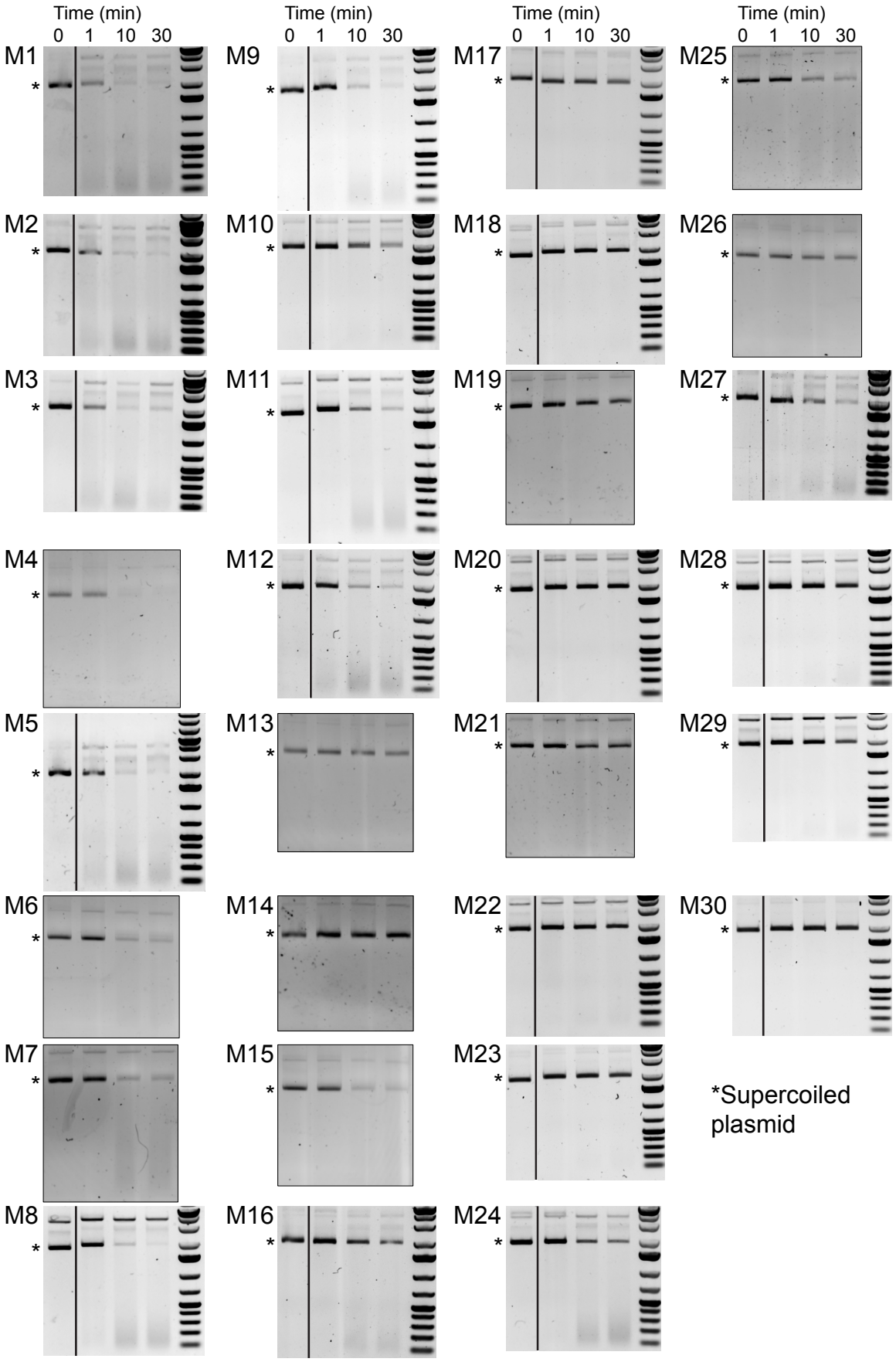
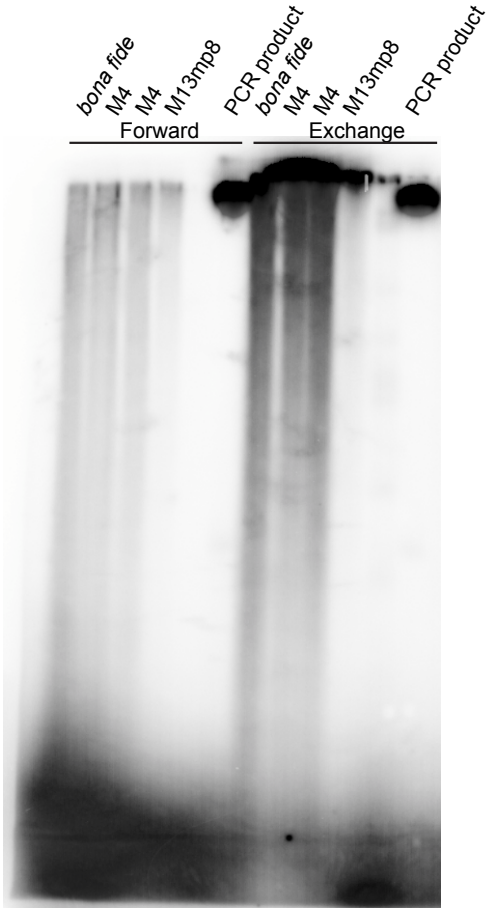

Time (min)
0 1 10 30
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14
M15
M16
M17
M18
M19
M20
M21
M22
M23
M24
M25
M26
M27
M28
M29
M30
*Supercoiled plasmid

Figure S4. Related to Figure 2B. Representative gels of Cas3 activity assays. Individual gels for each mutant showing Cas3 plasmid degradation reactions at time points 0, 1, 10, 30 minutes. Vertical black lines indicate removal of 3 gel lanes with irrelevant samples. Supercoiled plasmid is indicated with an asterisk, gel lanes above are linearized and nicked plasmids, which are not considered in quantification.

A


bona fide
M4
M4
M13mp8
PCR product
bona fide
M4
M4
M13mp8
PCR product
Forward
Exchange

B

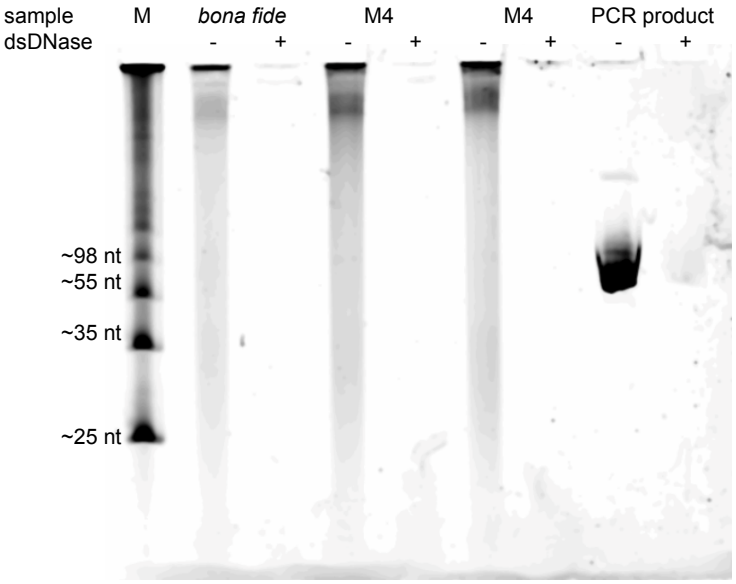

sample
M
bona fide
M4
M4
PCR product
dsDNase
-
+
-
+
-
+
-
+
~98 nt
~55 nt
~35 nt
~25 nt

Figure S5. Related to Figure 3: Biochemical analysis of Cas3 DNA degradation fragments. A) ^{32}P PNK labeling of degradation fragments from *bona fide* target plasmid, M4 target plasmid and m13mp8 single stranded plasmid. Forward reaction can only label non-phosphorylated 5'ends, exchange reaction can label both phosphorylated and non-phosphorylated 5' ends. Non-phosphorylated PCR product for reference. B) dsDNase incubation with degradation fragments of *bona fide* target plasmid and M4 target plasmid. dsDNase is a double stranded DNA specific endonuclease with no activity on single stranded DNA.

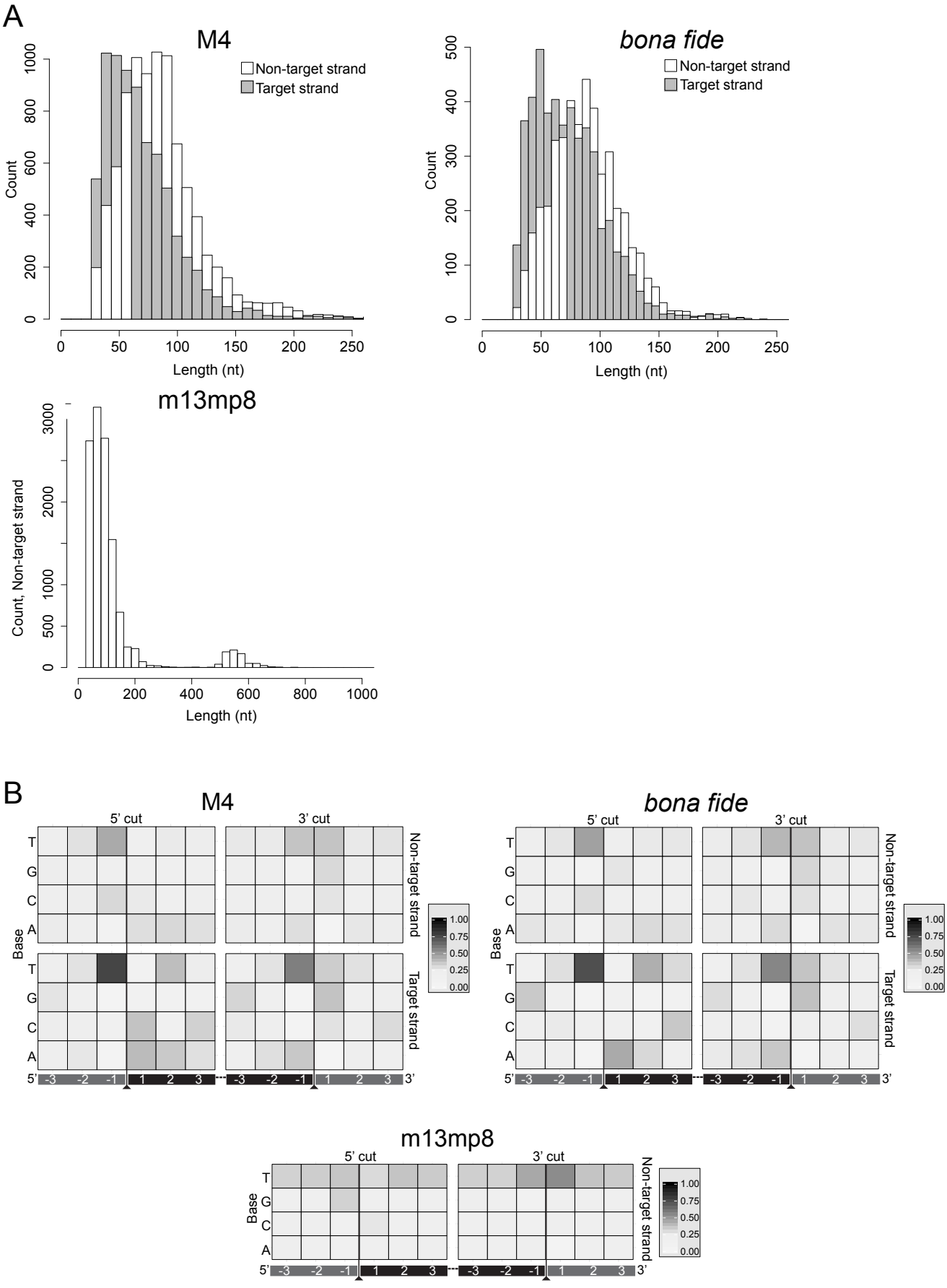

A
M4
Non-target strand
Target strand
Count
Length (nt)
bona fide
Non-target strand
Target strand
Count
Length (nt)
m13mp8
Count, Non-target strand
Length (nt)
B
M4
5' cut
3' cut
Non-target strand
Target strand
Base
T
G
C
A
1.00
0.75
0.50
0.25
0.00
5'
-3
-2
-1
1
2
3
3'
bona fide
5' cut
3' cut
Non-target strand
Target strand
Base
m13mp8
5' cut
3' cut
Non-target strand
Base

Figure S6. Related to Figure 3: Next generation sequencing analysis of Cas3 DNA degradation products. A) Length distribution bar charts for Cas3 DNA degradation products of *bona fide* target plasmid, M4 target plasmid and m13mp8 single stranded plasmid. B) Heat maps of nucleotide frequencies around cleavage sites for *bona fide* target plasmid, M4 target plasmid and m13mp8 single stranded plasmid. 5' and 3' cut sites are displayed separately for both target and non-target strand. The cleavage site is between position -1 and 1. Positions indicated in black are on the fragments, positions indicated in grey are outside of fragments.

Supplemental Tables

Table S1: Oligo nucleotides used in this study

Name	Sequence	Description
BG4556	ATCCCGGGATGACCTGGCTTCCCCTT	Cas1 fw (SmaI)
BG4557	AGTGAGCTCTCAAACAGGTAAAAAAGACACC	Cas2 rv (SacI)
BG5301	AAGGTTGGTGGGTTGTTTTTATGG	CRISPR leader forward primer
BG5302	GGATCGTCACCCTCAGCAGCG	M13_g8 spacer reverse primer
BG6170	CACTCTTTCCCTACACGACGCTCTTCCGATCTG CCTAA	NGS PE 5'Adapter 3
BG6174	CACTCTTTCCCTACACGACGCTCTTCCGATCTG ATCTG	NGS PE 5'Adapter 7
BG6176	CACTCTTTCCCTACACGACGCTCTTCCGATCTC TGATC	NGS PE 5'Adapter 9
BG6179	AATGATACGGCGACCACCGAGATCTACACTCT TTCCCTACACGACGC	NGS PE 5'Adapter extension primer
BG6180	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC TTTTTTTTTTTTTTTTTTTTTTTTTTTTVN	NGS PE 3' Tail primer 1
BG6183	CAAGCAGAAGACGGCATACGAGATGCCTAAGT GACTGGAGTTCAGACGTGTG	NGS PE 3' Tail primer 2.3
BG6187	CAAGCAGAAGACGGCATACGAGATGATCTGGT GACTGGAGTTCAGACGTGTG	NGS PE 3' Tail primer 2.7
BG6189	CAAGCAGAAGACGGCATACGAGATCTGATCGT GACTGGAGTTCAGACGTGTG	NGS PE 3' Tail primer 2.9
BG6713	GCTCTGCTGAAGCCAGTT	Reverse S437 hot spot pBR322
BG6714	GATCCTCTAGAGTCGACCT	Reverse S429 hot spot bb
BG6715	GCTAGTTGAACGGATCCAT	Reverse S416 hot spot GFP
BG7213	CGCTGCTGCGAAATTTGAAC	pWUR477 single repeat fw
BG7214	AACTCTGCGTGAGCGTATCG	pWUR477 single repeat rv
BG7215	ATCCGTTCAACTAGCAGACC	GFP hotspot nested forward
BG7216	GGTCTGCTAGTTGAACGGAT	GFP hotspot nested reverse
BG7415	CAATTTACTACTCGTTCTGGTGTTTCTCGTCAG GG	Protospacer 35 forward
BG7416	ACGAGAAACACCAGAACGAGTAGTAAATTGG GCTT	Protospacer 35 reverse
BG7522	CTGCGCTAGTAGACGAGTC	pWUR477 behind array reverse

Table S2: Plasmids used in this study

Plasmid	Description (positions of all mutations)	Name in paper	source
pWUR835	pGFP-UV Amp	-	(Fineran et al., 2014)
pWUR836	pGFP-UV Km protospacer8 WT	pTarget *bona fide*	(Fineran et al., 2014)
pWUR837	pGFP-UV Km protospacer8 mutant pos. 1, 3, 24	pTarget M14	(Fineran et al., 2014)
pWUR838	pGFP-UV Km protospacer8 mutant pos. 10, 11, 25	pTarget M12	(Fineran et al., 2014)
pWUR839	pGFP-UV Km protospacer8 mutant pos. 1, 4, 16	pTarget M30	(Fineran et al., 2014)
pWUR840	pGFP-UV Km protospacer8 mutant pos. 2, 3, 4	pTarget M17	(Fineran et al., 2014)
pWUR841	pGFP-UV Km protospacer8 mutant pos. 3, 7, 19	pTarget M26	(Fineran et al., 2014)
pWUR842	pGFP-UV Km protospacer8 mutant pos. 4, 8, 26	pTarget M23	(Fineran et al., 2014)
pWUR843	pGFP-UV Km protospacer8 mutant pos. 2, 10, 16	pTarget M16	(Fineran et al., 2014)
pWUR844	pGFP-UV Km protospacer8 mutant pos. 2, 18, 22	pTarget M9	(Fineran et al., 2014)
pWUR845	pGFP-UV Km protospacer8 mutant pos. 10, 14, 17	pTarget M5	(Fineran et al., 2014)
pWUR846	pGFP-UV Km protospacer8 mutant pos. 11, 16, 17	pTarget M7	(Fineran et al., 2014)
pWUR847	pGFP-UV Km protospacer8 mutant pos. 11, 22, 32	pTarget M1	(Fineran et al., 2014)
pWUR848	pGFP-UV Km protospacer8 mutant pos. 5, 6, 25	pTarget M2	(Fineran et al., 2014)
pWUR850	pGFP-UV Km protospacer8 mutant pos. 2, 8, 26	pTarget M10	(Fineran et al., 2014)
pWUR851	pGFP-UV Km protospacer8 mutant pos. 19, 27, 32	pTarget M27	(Fineran et al., 2014)
pWUR852	pGFP-UV Km protospacer8 mutant pos. 12, 17, 31	pTarget M3	(Fineran et al., 2014)
pWUR853	pGFP-UV Km protospacer8 mutant pos. 6, 7, 32	pTarget M4	(Fineran et al., 2014)
pWUR854	pGFP-UV Km protospacer8 mutant pos. 1, 10, 15, 18, 29	pTarget M25	(Fineran et al., 2014)
pWUR855	pGFP-UV Km protospacer8 mutant pos. 1, 16, 19, 25, 29	pTarget M13	(Fineran et al., 2014)
pWUR856	pGFP-UV Km protospacer8 mutant pos. 1, 4, 19, 27, 28	pTarget M20	(Fineran et al., 2014)
pWUR857	pGFP-UV Km protospacer8 mutant pos. 2, 12, 23, 26, 27	pTarget M11	(Fineran et al., 2014)
pWUR859	pGFP-UV Km protospacer8 mutant pos. 3, 8, 10, 11, 22	pTarget M29	(Fineran et al., 2014)
pWUR860	pGFP-UV Km protospacer8 mutant pos. 3, 15, 20, 25, 26	pTarget M18	(Fineran et al., 2014)
pWUR859	pGFP-UV Km protospacer8 mutant pos. 3, 9, 13, 22, 26	pTarget M19	(Fineran et al., 2014)
pWUR860	pGFP-UV Km protospacer8 mutant pos. 5, 6, 8, 24, 31	pTarget M8	(Fineran et al., 2014)
pWUR861	pGFP-UV Km protospacer8 mutant pos. 4, 5, 6, 15, 24	pTarget M24	(Fineran et al., 2014)
pWUR862	pGFP-UV Km protospacer8 mutant pos. 1, 2, 9, 14, 21	pTarget M22	(Fineran et al., 2014)

pWUR863	pGFP-UV Km protospacer8 mutant pos. 6, 22, 27, 31, 32	pTarget M28	(Fineran et al., 2014)
pWUR864	pGFP-UV Km protospacer8 mutant pos. 12, 13, 23, 24, 30	pTarget M6	(Fineran et al., 2014)
pWUR866	pGFP-UV Km protospacer8 mutant pos. 3, 9, 12, 16, 32	pTarget M21	(Fineran et al., 2014)
pWUR867	pGFP-UV Km protospacer8 mutant pos. 17, 27, 28, 29, 30	pTarget M15	(Fineran et al., 2014)
pWUR868	pGFP-UV Km non-target	pTarget NT	(Fineran et al., 2014)
pWUR748	pMAT11-MBP-Cas3		(Mulepati and Bailey, 2013)
pWUR868	pACYC poly spacer8 CRISPR array		This study
pWUR514	*cse2* with Strep-tag II (N-term)-*cas7-cas5-cas6e* in pET52b		(Jore et al., 2011)
pWUR408	*cse1* in pRSF-1b, no tags		(Brouns et al., 2008)
pWUR477	pACYC with artificial CRISPR array		(Brouns et al., 2008)
pWUR872	pWUR477 with only one repeat	pCRISPR	This study
pWUR871	Cas1-Cas2 operon with Strep-tag II (N-term) in pET52b		This study

Table S3. Related to Figure2A: EMSA data from regression analysis

Plasmid	Amplitude	Kd (nM)	Amplitude/Kd
bona fide (WT)	1.0 ± 0.01	7.6 ± 0.8	1.31E-01
M1	0.85 ± 0.01	23.6 ± 2.0	3.59E-02
M2	0.92 ± 0.04	23.6 ± 4.6	3.92E-02
M3	0.99 ± 0.02	18.5 ± 2.7	5.35E-02
M4	1.02 ± 0.04	16.4 ± 3.34	6.23E-02
M5	0.87 ± 0.03	34.3 ± 5.3	2.54E-02
M6	0.0	--	0.00E+00
M7	0.69 ± 0.01	31.6 ± 2.7	2.17E-02
M8	0.65 ± 0.01	17.4 ± 2.0	3.71E-02
M9	0.94 ± 0.03	24.8 ± 4.7	3.78E-02
M10	1.05 ± 0.05	23.4 ± 5.3	4.50E-02
M11	0.39 ± 0.02	22.1 ± 6.0	1.77E-02
M12	0.0	--	0.00E+00
M13	0.0	--	0.00E+00
M14	1.2 ± 0.13	360 ± 79.4	3.46E-03
M15	0.46 ± 0.01	4.4 ± 0.4	1.04E-01
M16	0.78 ± 0.02	46.3 ± 6.7	1.69E-02
M17	1.19 ± 0.02	152.6 ± 10.0	7.79E-03
M18	0.0	--	0.00E+00
M19	0.0	--	0.00E+00
M20	0.0	--	0.00E+00
M21	0.0	--	0.00E+00
M22	0.94 ± 0.01	55.9 ± 2.7	1.69E-02
M23	0.69 ± 0.02	54.1 ± 5.3	1.27E-02
M24	0.9 ± 0.03	22.4 ± 4.0	4.03E-02
M25	0.31 ± 0.01	34.6 ± 6.0	9.02E-03
M26	0.93 ± 0.03	79.4 ± 8.7	1.17E-02
M27	0.74 ± 0.02	20.7 ± 2.7	3.59E-02
M28	1.04 ± 0.04	17.4 ± 3.3	5.97E-02
M29	0.4 ± 0.02	74.2 ± 18.0	5.40E-03
M30	0.0	--	0.00E+00

Table S4. Related to Figure 3: NGS data processing and mapping

Sample name	Total number of reads	Reads mapping to NT strand	Reads mapping to NT strand (%)	Reads mapping to T strand	Reads mapping to T strand (%)
bona fide (WT)	215218	57217	26.6	158001	73.4
M4	101327	23334	23	77993	77
M13mp8	46205	46109	>0.99	96	<0.01

Table S5. Related to Figure 3: NGS data processing for cleavage sites

	Non-target strand (NT)			Target strand (T)		
Sample name	# Distinct Fragments	# Distinct Start	# Distinct End	# Distinct Fragments	# Distinct Start	# Distinct End
bona fide (WT)	8777	1381	1479	7448	1318	1151
M4	4432	971	1076	4784	1029	920
M13mp8	12243	3737	2620			

Supplemental Experimental Procedures

Bacterial Strains and Growth Conditions. *Escherichia coli* strain KD263 was obtained from (Shmakov et al., 2014). *E. coli* strains were grown at 37 °C in Luria Broth (LB; 5 g/L NaCl, 5 g/L yeast extract, and 10 g/L tryptone) at 180 rpm or on LB-agar plates containing 1.5% (wt/vol) agar. When required, medium was supplemented with the following: ampicillin (Amp; 100 μg/mL), chloramphenicol (Cm; 34 μg/mL), or kanamycin (Km; 50 μg/mL). Bacterial growth was measured at 600 nm (OD600).

Molecular Biology and DNA Sequencing. All oligonucleotides are listed in Table S1. All plasmids are listed in Table S2. All strains and plasmids were confirmed by PCR and sequencing (GATC-Biotech). Plasmids were prepared using GeneJET Plasmid Miniprep Kits (Thermo Scientific). DNA from PCR and agarose gels was purified using the DNA Clean and Concentrator and Gel DNA Recovery Kit (Zymo Research). The library of pGFPuv sp8 mutants was available from a previous study (Fineran et al., 2014). pMAT MBP-Cas3 was a kind gift from Scott Bailey lab (Mulepati and Bailey, 2013).

Transformation assay. Transformation assays were carried out in *E. coli* KD263. Cells were grown to OD600 ~0.4, induced with 0.2% L-arabinose and 0.5 mM IPTG and allowed to grow for 1h. Cells were then made chemically competent for heat shock transformation using the $RuCl_2$ method. Cells were co-transformed with 10 ng target plasmid (pWUR836-868, Kan^R) and 10 ng control plasmid (pWUR835, Amp^R) simultaneously (Almendros and Mojica, 2015). Dilutions of transformants were then plated on LBA plates with Amp and LBA plates with Kan. The transformation efficiency of mutated target plasmids was normalized against the transformation efficiency of the control plasmid.

Plasmid loss assay. *E. coli* KD263 cells were transformed with the target plasmids (pWUR836-868) by heat shock. Individual colonies were picked in triplicate and grown overnight in 5 ml LB supplemented with 2% glucose to repress *cas* gene expression. The next day, cultures were transferred 1:100 into induced medium (0.2% L-Arabinose, 0.5 mM IPTG) and plasmid loss was monitored. Samples were taken every hour until 5h, and then again at 24h and 48h. Dilutions were plated on non-selective plates and plasmid loss was counted based on loss of fluorescence using a Syngene G-box imager. Plasmid-free colonies were screened for spacer integration by colony PCR using DreamTaq Green DNA polymerase (Thermo Scientific). Acquisition of spacers was detected by PCR using primers BG5301 and BG5302. PCR products were visualized on 2% agarose gels and stained with SYBR-safe (Invitrogen). PCR products were sequenced using Sanger sequencing at GATC (Konstantz, Germany) using primer BG5301.

EMSA assays. Purified Cascade complex with spacer8 crRNA was incubated with plasmid at a range of molar ratios (1:1-100:1, Cascade:DNA) in buffer A (20 mM HEPES pH7.5, 75 mM NaCl, 1 mM DTT) for 30 min. Reactions were run on 1% native agarose gels for 18h at 22 mA in 8 mM sodium-borate buffer. Gels were post stained with SYBR Safe (Invitrogen). Shifted (Cascade bound DNA) and unshifted (free DNA) bands were quantified using the GeneTools software (Syngene) and total Cascade concentration (X) was plotted against the fraction of bound DNA (Y). The curves were fitted with the following formula: $Y = (\text{amplitude} * X)/(K_d + X)$ (van Erp et al., 2015). The amplitude is the maximum fraction of bound DNA. Since the amplitude is not always 1, we cannot directly compare K_d values, instead the 'affinity ratio' was calculated as: $\text{amplitude}/K_d$ (i.e. normalizing the K_d against the variable amplitude).

Cas3 DNA degradation assays. Cas3 DNA degradation activity was routinely tested by incubating 500 nM Cas3 with 4 nM M13mp8 single stranded circular DNA in buffer R (5 mM HEPES, pH8, 60 mM KCl) supplemented with 100 μM Ni^{2+} at 37 °C for 1 h. Plasmid-based assays were performed by incubating 70 nM Cas3 with 70 nM Cascade, 3.5 nM plasmid DNA in buffer R (+ 10 μM $CoCl_2$, 10 mM $MgCl_2$, 2 mM ATP) at 37 °C for 10-60 minutes unless indicated otherwise. For quantifying Cas3 activity, assays were run at normal conditions and samples were taken at 0 min, 1 min, 10 min and 30 min. Samples were immediately quenched with 6x DNA loading dye (Thermo scientific) on ice. Samples were run on agarose gels and supercoiled plasmid bands were quantified using the GeneTools software (Syngene). The DNA degradation was plotted (X: time [min]; Y: Intact Plasmid [%]) and the initial activity of Cas3 [%/min] calculated from the initial slope of the curve.

Protein purification. The Cas1-2 operon was PCR amplified with primers BG4556/7 and cloned into pET52b (SmaI/SacI) to make pWUR871. The Cas1-2 complex was purified using the N-terminal StrepII tag on Cas1. Briefly, cells were grown to an OD_{600} of 0.4, cooled on ice for 30 minutes and induced with 0.5 mM IPTG and 0.2% l-arabinose. Protein was expressed at 20 °C overnight. Cells were collected by centrifugation and lysed in buffer L (20mM HEPES pH 7.5, 75mM NaCl, 1mM DTT, 5% glycerol, 0.1% Triton X100) using a Stansted pressure cell homogenizer. The lysate was cleared by centrifugation and filtration. The cleared lysate was

incubated with Strep-tactin beads (IBA) for 30 minutes at 4 °C and loaded into a gravity column. The column was washed with buffer A (20mM HEPES pH 7.5, 300mM NaCl, 1mM DTT, 5% glycerol) and the proteins eluted in buffer B (20mM HEPES pH 7.5, 75mM NaCl, 1mM DTT, 5% glycerol, 2.5 mM biotin). The presence and purity of the Cas1-2 complex was checked via Tris-tricing SDS PAGE (10-20%). The final complex was snap frozen in liquid nitrogen and stored at -80 °C.

Degradation product analysis. To test if Cas3 produces single- or double-stranded DNA products, the reaction products of the plasmid based assay were incubated with dsDNase (Thermo Scientific) according to manufacturer's protocol. dsDNase exclusively degrades double-stranded DNA. Products were run on a 5% denaturing PAGE gel and visualized using Sybr-Gold (Thermo Scientific). To determine the phosphorylation state of the degradation products, the products were ^{32}P labelled with T4 PNK (Thermo) using the forward and exchange reaction according to the manufacturer's protocol. Labelled DNA was run on an 8% PAGE gel and visualized using a phosphor imaging screen (GE healthcare) and a Personal molecular imager (Bio-Rad).

Statistical testing against the null hypothesis. To establish the confidence with which the null hypothesis can be disregarded, we construct randomized mock behavioral groups by repeatedly (10^5 times, resulting in an accuracy in the significance intervals of about $1/\sqrt{10^5} \approx 0.3\%$) drawing a random selection (allowing repetitions) of sequences from the complete set of 31 protospacers (including the *bona fide* spacer). The average property of interest is then calculated for the generated mock behavioral groups, giving histograms showing the distribution over the mock sets. The above procedure is performed for the total number of effective mismatches, and the number of mutations within segment 1, and the number of mutations on position 3 within all segments combined.

***In vitro* acquisition assay**. Component concentrations for assay 1 and 2 were as follows: 70 nM Cascade, 70 nM Cas3, 300 nM Cas1-2, 3.5 nM target plasmid, 5 nM pWUR869 (pCRISPR). Reaction products of both assays were run on a 1.8% TAE-agarose gel. To verify half-site integration of spacers in the CRISPR array as described in (Nunez et al., 2015), nicked pWUR869 was isolated from gel and analyzed by PCR. PCR was performed with forward primer BG5301 (site2) or BG7522 (site1) and reverse primers BG7415/6 (control) or BG6713-15 (3 hotspots) or BG7215/6 (fw/rv of hotspot3). These primers match spacers that are frequently incorporated *in vivo* (Fineran et al., 2014). To verify and analyze integration, PCR products were cloned into a pGEMT-easy vector (Promega) and individual clones were sequenced.

NGS library construction. Degradation products were gel purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research), cutting out DNA up to ~500bp. DNA was then poly-A tailed with TdT (Invitrogen) according to manufacturer's protocol (approximately 100 nt tails). Tailed DNA was purified using the DNA Clean and Concentrator Kit (Zymo Research). Subsequently, tailed products were 5' phosphorylated with T4-PNK (Thermo Scientific). Next, the DNA was heated to 95°C to separate DNA strands and a barcoded ssDNA adapter (BG6170/4/6) was ligated to the 5' end of the products. Unincorporated adapters were removed using the DNA Clean and Concentrator Kit (Zymo Research). PCR amplification was performed with BG6179 and BG6180. A second round of PCR amplification was performed with BG6179 and BG6183/7/9 (barcoded). PCR products were purified and sent to the Imagif, Centre for Molecular Genetics, Centre National de la Recherche Scientifique, France for sequencing (paired-end, 2x250nt). Based on the procedure outlined above, a fraction of degradation fragments smaller than 50 nucleotides was purified with lower yields during the initial agarose gel extraction, and could be less populated in the size distribution shown in Fig 3B/S6A.

NGS Data analysis. Samples were de-multiplexed using their barcodes. All pair-end reads were mapped to their originating sequences (pWUR836/853, m13mp8) using BLAST and allowing for up to one mismatch. Reads for which both ends could not be aligned to the reference sequence were discarded. For the cleavage sites, distinct start/end positions were analyzed independently (see Table S4 and Table S5 for details). For the duplets a sliding window around the cut point was used. For the duplets the following positions were considered: (-2,-1), (-1,1) and (1,2) . In this notation the cut point is between -1 and 1, positive positions are inside the considered fragment and negative positions are outside. Enrichment analysis was performed using a hypergeometric probability distribution to model the background probability density associated to the originating sequence. R packages stats (R-Development-Core-Team, 2008) and ggplot2 (Wickham, 2009) were used for these computations and to generate corresponding graphics.

Supplemental 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 in molecular biology *1311*, 161-170.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.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.

Fineran, P.C., Gerritzen, M.J., Suarez-Diez, M., Kunne, T., Boekhorst, J., van Hijum, S.A., Staals, R.H., and Brouns, S.J. (2014). Degenerate target sites mediate rapid primed CRISPR adaptation. Proceedings of the National Academy of Sciences of the United States of America *111*, E1629-1638.

Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R., Waghmare, S.P., Wiedenheft, B., Pul, Ü., Wurm, R., and Wagner, R. (2011). Structural basis for CRISPR RNA-guided DNA recognition by Cascade. Nature structural & molecular biology *18*, 529-536.

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. The Journal of biological chemistry *288*, 22184-22192.

Nunez, J.K., Lee, A.S., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193-198.

R-Development-Core-Team (2008). R: A language and environment for statistical computing (Vienna).

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.

van Erp, P.B., Jackson, R.N., Carter, J., Golden, S.M., Bailey, S., and Wiedenheft, B. (2015). Mechanism of CRISPR-RNA guided recognition of DNA targets in Escherichia coli. Nucleic Acids Res *43*, 8381-8391.

Wickham, H. (2009). Ggplot2 : elegant graphics for data analysis (New York: Springer).