



Prespacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase

Olga Musharova^{a,b,1,2}, Sofia Medvedeva^{a,1}, Evgeny Klimuk^b, Noemi Marco Guzman^a, Daria Titova^a, Victor Zgoda^c, Anna Shiriaeva^{a,d}, Ekaterina Semenova^e, Konstantin Severinov^{a,b,e,2}, and Ekaterina Savitskaya^{a,b,3}

^aCenter of Life Sciences, Skolkovo Institute of Science and Technology, Moscow 121205, Russia; ^bInstitute of Molecular Genetics, National Research Centre “Kurchatov Institute,” Moscow 123182, Russia; ^cAdvanced Mass Spectrometry Core Facility, Skolkovo Institute of Science and Technology, Moscow 121205, Russia; ^dResearch Center of Nanobiotechnologies, Peter the Great St. Petersburg Polytechnic University, St. Petersburg 195251, Russia; and ^eWaksman Institute of Microbiology, Rutgers, The State University of New Jersey, Piscataway, NJ 08854

Edited by Luciano A. Marraffini, The Rockefeller University, New York, NY, and approved April 21, 2021 (received for review October 11, 2020)

For Type I CRISPR–Cas systems, a mode of CRISPR adaptation named priming has been described. Priming allows specific and highly efficient acquisition of new spacers from DNA recognized (primed) by the Cascade–crRNA (CRISPR RNA) effector complex. Recognition of the priming protospacer by Cascade–crRNA serves as a signal for engaging the Cas3 nuclease–helicase required for both interference and primed adaptation, suggesting the existence of a primed adaptation complex (PAC) containing the Cas1–Cas2 adaptation integrase and Cas3. To detect this complex in vivo, we here performed chromatin immunoprecipitation with Cas3-specific and Cas1-specific antibodies using cells undergoing primed adaptation. We found that prespacers are bound by both Cas1 (presumably, as part of the Cas1–Cas2 integrase) and Cas3, implying direct physical association of the interference and adaptation machineries as part of PAC.

CRISPR–Cas | CRISPR interference | CRISPR adaptation

CRISPR–Cas systems of adaptive immunity provide prokaryotes with resistance against bacteriophages and plasmids (1–4). They consist of CRISPR DNA arrays and *cas* genes. Functionally, CRISPR defense can be subdivided into the interference and adaptation steps. The interference step involves specific recognition of regions in foreign nucleic acids, named protospacers, based on their complementarity to CRISPR arrays spacers followed by their destruction (5). The CRISPR adaptation step leads to integration of new spacers into the array (6, 7), forming inheritable memory that allows the entire lineage of cells derived from a founder that acquired a particular spacer to do away with genetic invaders carrying matching protospacers (8).

Both interference and adaptation can be subdivided into multiple steps. For interference to occur, the CRISPR array is transcribed from a promoter located in the upstream leader region. The resulting pre-CRISPR RNA (pre-crRNA) is processed into short CRISPR RNAs (crRNAs), each containing a spacer flanked by repeat fragments (9). Individual crRNAs are bound by Cas proteins forming the effector complex, which is capable of recognizing sequences complementary to the spacer part of crRNA (10). Upon protospacer recognition, the target is destroyed either by a protein component of the effector complex or by additional recruitable Cas nucleases (3, 11–14). In a well-studied Type I–E CRISPR–Cas system of *Escherichia coli*, the effector comprises a multisubunit Cascade protein complex bound to a crRNA (11, 12, 15). The complementary interaction of Cascade-bound crRNA with a target protospacer leads to localized protospacer DNA melting and formation of an R-loop complex, where the crRNA spacer is annealed to the protospacer “target” strand, while the opposing “nontarget” strand is displaced and is present in a single-stranded form (16, 17). To avoid potentially suicidal recognition of CRISPR array spacers from which crRNAs originate, target recognition and R-loop complex formation require, in addition to complementarity with the crRNA

spacer, the presence of a three-nucleotide long PAM (protospacer adjacent motif) preceding the protospacer (15, 18, 19). For *E. coli* type I–E system, the consensus PAM sequence is 5′-AAG-3′ on the nontarget strand. Some other trinucleotides also allow target recognition, though with decreased efficiency (15, 20). Below, we will refer to consensus PAM as “PAM^{AG}.” The Cas3 nuclease–helicase is recruited to the R-loop complex and is responsible for target destruction (21–24). Cas3 first introduces a single-stranded break in the nontarget protospacer strand 11 to 15 nucleotides downstream of the PAM on the nontarget strand (25). Next, Cas3 unwinds and cleaves DNA in the 3′-5′ direction from the PAM (26–29). In vitro, Cas3-dependent degradation of DNA at the other side of the protospacer was also detected (16). Bidirectional Cas3-dependent degradation of DNA was also detected in vivo (30). The details of Cas3 “molecular gymnastics” required for such bidirectional destruction of DNA around the R-loop complex are not known.

The main proteins of CRISPR adaptation are Cas1 and Cas2. In vitro, these proteins interact with each other, and the resulting complex is capable of inserting spacer-sized fragments in substrate DNA molecules containing at least one CRISPR repeat and a

Significance

Primed adaptation allows rapid acquisition of protective spacers derived from foreign mobile genetic elements into CRISPR arrays of the host. Primed adaptation requires ongoing CRISPR interference that destroys foreign genetic elements, but the nature of this requirement is unknown. Using the *Escherichia coli* I–E CRISPR–Cas as a model, we show that prespacers, short fragments of foreign DNA on their way to become incorporated into CRISPR arrays as spacers, are associated with both the adaptation integrase Cas1 and the interference nuclease Cas3, implying physical association of the interference and adaptation machineries during priming.

Author contributions: E. Semenova, K.S., and E. Savitskaya designed research; O.M., S.M., E.K., N.M.G., D.T., V.Z., and E. Savitskaya performed research; S.M., V.Z., A.S., E. Semenova, and K.S. analyzed data; and K.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹O.M. and S.M. contributed equally to this work.

²To whom correspondence may be addressed. Email: O.Musharova@skoltech.ru or severik@waksman.rutgers.edu.

³Deceased October 29, 2019.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021291118/-DCSupplemental>.

Published May 25, 2021.

repeat-proximal leader region (31, 32). In the course of spacer integration, the Cas1–Cas2 complex first catalyzes a direct nucleophilic attack by the 3'-OH end of the incoming spacer at a phosphodiester bond between the leader and the first repeat in the top CRISPR strand (32, 33). This reaction proceeds via concurrent cleavage of the leader-repeat junction and covalent joining of one spacer strand to the 5' end of the repeat. Subsequently, the 3'-OH on the second spacer strand attacks the phosphodiester bond at the repeat-spacer junction in the bottom CRISPR strand leading to full integration (32, 33). As a result, an intermediate with the newly incorporated spacer flanked by single-stranded repeat sequences is formed (32, 34). The gaps are filled in by a DNA polymerase, possibly DNA polymerase I (35).

When overexpressed, *E. coli* Cas1 and Cas2 can integrate new spacers into the array in the absence of other Cas proteins (7, 36). During such “naive” adaptation, ~50% of newly acquired spacers are selected from sequences flanked by the 5'-AAG-3' trinucleotide, that is, consensus interference-proficient PAM^{AG}. It thus follows that at least 50% of spacers acquired by Cas1 and Cas2 alone will be defensive during the interference step. The adaptation process must be tightly controlled, activated in the presence of the infecting mobile genetic elements, and directed toward foreign DNA, for otherwise, spacers acquired from host DNA will lead to suicidal self-interference. The details of the activation of CRISPR adaptation upon the entry of foreign DNA into the cell remain elusive. Some data indicate that active replication and/or a small size of phage or plasmid DNA may be responsible for a preferential selection of spacers from these molecules compared to selection of self-targeting spacers from host chromosomes (19). In addition, DNA repair/recombination signals present in host DNA, but lacking in foreign DNA may also increase the bias of the adaptation machinery to the latter (37).

The bias of spacer acquisition machinery toward foreign DNA does not have to be significant, for acquisition of a self-targeting spacer by an infected cell will lead to the demise of such a cell in an act of altruism that would help control the spread of the infectious agent through the population. In contrast, acquisition of interference-proficient spacers from foreign DNA may allow the infected cell to survive, clear the infection, and endow its progeny with inheritable resistance—clearly an advantageous trait.

To overcome CRISPR resistance, viruses and plasmids accumulate “escaper” mutations in the targeted protospacer or its PAM (36, 38). Given that the acquisition of protective spacers in infected cells is likely to be a rare event and the ease with which escaper mutations accumulate, the complex multistage CRISPR defense could become costly and ineffective (39). To increase the efficiency of CRISPR defense and counter the spread of mobile genetic elements with escaper mutations, CRISPR-Cas systems have evolved a specialized mode of spacer acquisition referred to as “primed adaptation” or “priming” (36, 40–47). Unlike the naive adaptation, in Type I CRISPR-Cas systems, priming requires, in addition to Cas1 and Cas2, a Cascade charged with crRNA recognizing the foreign target and the Cas3 nuclease–helicase. Spacers acquired during priming originate almost exclusively from DNA located *in cis* with the protospacer initially recognized by the effector complex (referred to hereafter as the “priming protospacer” or “PPS”). Furthermore, 90% or more of spacers acquired during priming by the I-E system of *E. coli* originate from protospacers with PAM^{AG} and are therefore capable of efficient interference. Another hallmark of primed adaptation is the following: spacers acquired from DNA located at different sides of the PPS map to opposite DNA strands. The mapping of spacers acquired during naive adaptation shows no strand bias (48). Thus, the strand bias of spacers acquired during priming is probably related to Cas3 nuclease activity; however, exact details are lacking.

The overall yield of spacers acquired during priming is increased when the PPS is imperfectly matched with a Cascade-bound crRNA spacer or when the PAM of the PPS is suboptimal (49).

Thus, escaper protospacers serve as PPS, and priming initiated by inefficient recognition of such protospacers allows cells to quickly update their immunological memory by specific and efficient acquisition of additional interference-proficient spacers from mobile genetic elements that accumulated escaper mutations to earlier acquired spacers.

The exact molecular mechanism of primed adaptation is not fully understood. Clearly, it should involve tight coordination between suboptimal interference against escaper targets and the spacer acquisition process. The DNA fragments produced by Cas3, a nuclease responsible for target degradation during interference, may feed primed adaptation, directly or indirectly, providing a functional link between the interference and adaptation arms of the CRISPR-Cas response. Based on results of *in vitro* experiments, it has been proposed that Cas3-generated degradation products may be used as substrates for the generation of pre-spacers (50)—DNA fragments that can be incorporated by the Cas1–Cas2 complex into arrays. However, no Cas3-generated products were detected in cells undergoing interference only, suggesting that Cas3 may degrade DNA to very short, subspace length products (30). On the other hand, mutations abolishing the Cas3 nuclease activity lead to very little primed adaptation, indicating that priming requires the Cas3 nuclease activity (51). A possible way out from this impasse would be the existence of a “priming complex” that includes both Cas1–Cas2 and Cas3 and is responsible for the generation of pre-spacers by the Cas1–Cas2 complex from DNA along which Cas3 moves. Single-molecule analysis supports the existence of such a complex and even suggests that PPS-bound Cascade may be part of the priming complex (52). Here, we show that both Cas1–Cas2 and Cas3 associate with the same set of pre-spacers in cells undergoing primed adaptation, functionally linking CRISPR interference and adaptation machineries during priming. We also investigate the phenomenon of strand bias of spacer acquisition during priming and show that this bias does not depend on the orientation of PPS.

Results

The Cas1–Cas2 Adaptation Complex and the Cas3 Nuclease–Helicase Associate with Short DNA Fragments in Cells Undergoing Primed Adaptation. To study the association of Cas proteins with DNA during primed adaptation, the previously developed *E. coli* KD263 strain (53) was used. KD263 contains the *cas3* gene under the control of the *lacUV5* promoter and the *cse1-cse2-cas7-cas5-cas6-cas1-cas2* operon under the *araBp8* promoter control. The KD263 cells harbor a single genetically modified CRISPR array with two repeats and a single spacer named G8. KD263 was transformed, in the absence of inducers of *cas* genes expression, with the pG8mut plasmid carrying a protospacer that matches the G8 spacer except for a single position adjacent to fully interference-proficient ATG PAM (36). As is shown elsewhere, the addition of *cas* gene expression inducers to KD263 cultures carrying pG8mut leads to CRISPR interference coupled with a primed adaptation of spacers from pG8mut (54). Thus, the G8mut protospacer functions as PPS, the priming protospacer. There are two versions of the pG8mut plasmid, pG8mut_dir and pG8mut_rev, differing in the orientation of a 209-bp (base pair) fragment carrying the PPS. Primed adaptation occurs in induced KD263 cultures transformed with both versions of pG8mut, but the pattern of acquired spacers differs depending on the orientation of PPS: most spacers acquired from each plasmid map to the nontarget strand of PPS (*SI Appendix, Fig. S1*).

Induced KD263 cells harboring each pG8mut plasmid variant were used for chromatin immunoprecipitation (ChIP) with Cas1- or Cas3-specific polyclonal antibodies. Uninduced cells were used as a control. Induced cells were collected 3 h postinduction when ~20% of CRISPR arrays were expanded by a single spacer-repeat unit (*SI Appendix, Fig. S2*). As additional controls, we also used

KD263 derivatives KD454 (lacks the *cas3* gene) and KD471 (lacks *cas1* and *cas2*). No CRISPR array expansion was observed in induced cultures of these cells as expected (SI Appendix, Fig. S24). Induced KD471 but not KD454 cultures interfered with pG8mut plasmids also as expected (20) (SI Appendix, Fig. S3). The culture aliquots were crosslinked with formaldehyde followed by immunoprecipitation as described previously (55). Protein-bound DNA fragments were purified and subjected to high throughput sequencing using Accel-NGS 1S Plus DNA sequencing protocol for Illumina. For each culture analyzed, “input” DNA that was not subjected to immunoprecipitation was also sequenced.

The length distribution of plasmid DNA fragments associated with Cas1–Cas2 or Cas3 was compared to input DNA and uninduced controls. As can be seen from Fig. 1B, all input samples contained a broad distribution of DNA fragment lengths with a maximum at about 130 nucleotides, consistent with the expected lengths generated by sonication. A similar distribution was observed in immunoprecipitated material from KD471 and KD454 samples. In contrast, specific enrichment with ~30 to 40–nucleotide long plasmid DNA fragments was observed in immunoprecipitated material from induced KD263 cultures harboring either version of pG8mut. The enrichment was detected both in the Cas1 and Cas3 antibody precipitated samples. No enrichment with short fragments was detected in immunoprecipitated material from uninduced KD263 cultures (Fig. 1B) or in chromosomal fragments from induced KD263 (SI Appendix, Fig. S4).

Short DNA Fragments Specifically Associated with Cas1–Cas2 and Cas3 in Cells Undergoing Primed Adaptation Are Plasmid-Derived Prespacers. Reads corresponding to immunoprecipitated 20 to 50–nucleotide long DNA fragments mapping to the pG8mut plasmids were extracted and analyzed. The DNA fragments had a bimodal distribution of lengths with two peaks of 33 and 37 nucleotides. The shorter fragments mapped mainly to the nontarget strand of plasmid DNA, while the longer ones preferentially mapped to the target strand (Fig. 2A). In each case, the longer fragments contained a CTTNN consensus sequence at their 3′ end, while the shorter fragments were enriched with a guanine residue at their 5′ ends (Fig. 2C and SI Appendix, Fig. S5). The length distributions of precipitated plasmid-derived short fragments and their preferred ends were the same in either Cas1- or Cas3-precipitated material (Fig. 2A and B).

Because of the heterogeneity in size of shorter and longer fragments, we grouped the 32 to 34– and 36 to 38–nucleotide long fragments together for further analysis. As can be seen from Fig. 2E, ~33 and ~37 nucleotide fragments mapping to opposing DNA strands tended to colocalize ($R = 0.5$, P value = 10^{-14}), suggesting that most Cas protein-associated material exists as double-stranded fragments with four-nucleotide 3′ overhangs at one end. The overhangs contain the 5′-CTT-3′ sequence complementary to consensus PAM^{AAG}. Thus, Cas1- and Cas3-associated fragments correspond to prespacers that were earlier observed in cells undergoing primed adaptation (30).

If the Cas1/Cas3-associated fragments were prespacers, their abundance should correlate with the abundance of corresponding

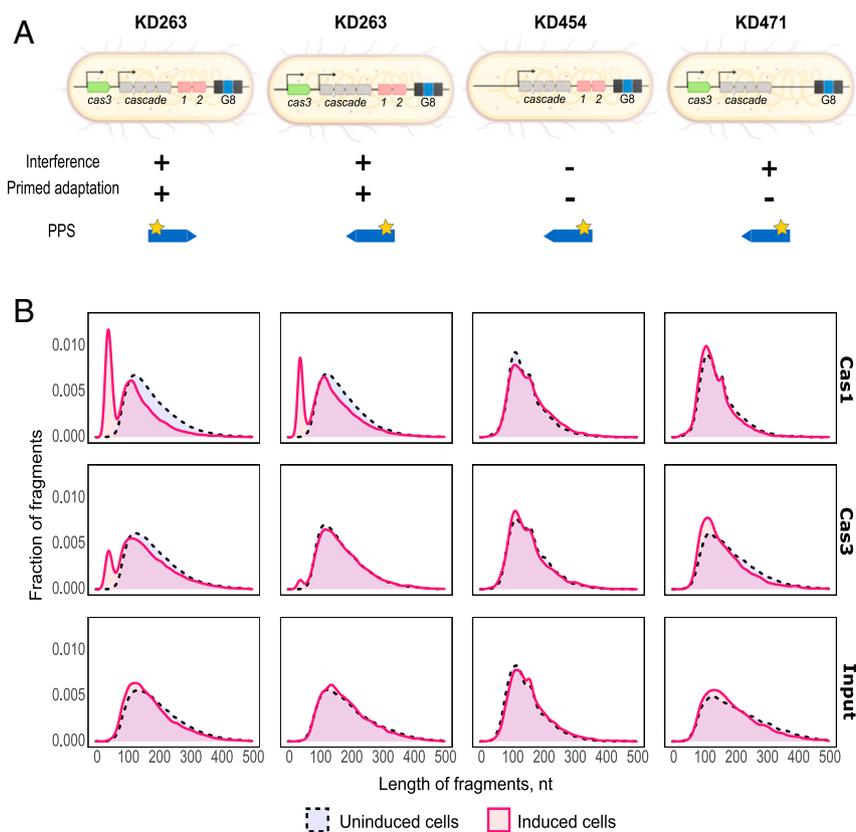


Fig. 1. Spacer-length DNA fragments associate with Cas1 and Cas3 proteins in cells undergoing primed adaptation. (A) Schematic representation of cells of three *E. coli* strains capable of 1) both CRISPR interference and primed adaptation (“wild-type” strain KD263), 2) CRISPR interference only (Δ *cas1,2* strain KD471), or 3) neither (Δ *cas3* strain KD454). Strains were transformed with pG8mut_dir or pG8mut_rev plasmids (KD263) or with pG8mut_rev (KD471 and KD454). The orientation of the PPS in each of these plasmids is shown. The PPS is shown as a blue arrow, and the yellow star shows the position of a mismatch with the G8 crRNA spacer. (B) Length distribution of plasmid-derived DNA fragments recovered after CHIP experiments with (Top) Cas1, (Middle) Cas3 specific antibodies, and (Bottom) input DNA. Red and dotted lines show length distributions of plasmid fragments in induced cells and uninduced cells, respectively. Distributions are normalized such that the total amount of fragments is the same for all samples (y-axes show fractions of the total number of fragments).

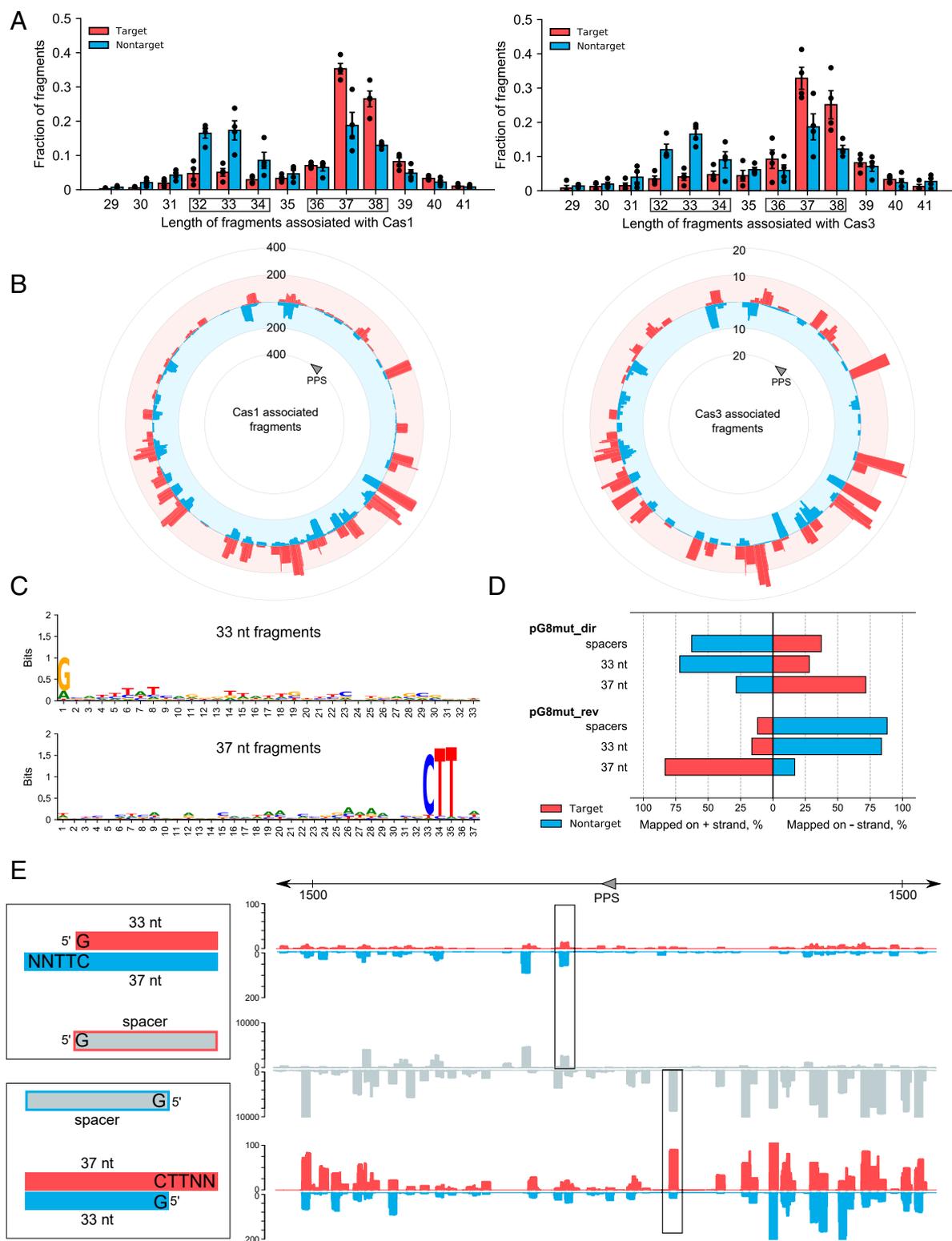


Fig. 2. Short DNA fragments associated with Cas1 and Cas3 in cells undergoing primed adaptation are prespacers. (A) Length distribution of Cas1- (Left) and Cas3- (Right) associated fragments mapping to target (red color) and nontarget (blue) strands of the pG8mut_rev plasmid. The heights of bars show the mean values obtained in four replicates. Dots show values obtained in each individual replicate. Error bars show SEM of four replicates. The fragment lengths selected for further analysis (32 to 34 nt and 36 to 38 nt) are highlighted with black frames. (B) Mapping of short fragments associated with Cas1 (Left) or Cas3 (Right) to both strands of pG8mut_rev. The red color is used to show the mapping of fragments to the target strand; the blue color shows the mapping to the nontarget strand of the plasmid. The position and orientation of the PPS is shown in the inner gray ring with an arrow. The scale shows the obtained sequence coverage of DNA fragments. (C) The structure of Cas1-associated fragments. Logo representation of nucleotide conservation for 33 nt (Top) and 37 nt (Bottom) fragments. (D) The bar plot shows strand bias of acquired spacers and ~33/~37 nt DNA fragments associated with Cas1 for pG8mut_dir and pG8mut_rev plasmids. (E) A linear representation of the pG8mut_rev plasmid and abundance of spacers (gray) and ~33/~37 nt DNA fragments associated with Cas1 for target (red) and nontarget (blue) strands. The structures of Cas1-associated DNA fragments and corresponding spacers are shown on the left side. Two examples of matching ~33 and ~37 nt fragments mapping to opposite strands, and acquired spacers are highlighted with black rectangles.

spacers acquired from the pG8mut plasmids. Indeed, spacers acquired from pG8mut_rev and mapping to its nontarget strand (constituted ~90% of all acquired plasmid-derived spacers) corresponded, both in sequence and abundance, to ~37 nucleotide prespacer fragments of the target strand ($R = 0.69$, P value = 10^{-16}) and ~33 nucleotide prespacer fragments from the nontarget strand ($R = 0.43$, P value = 10^{-10}) (Fig. 2D and E). The same was true for pG8mut_dir (SI Appendix, Fig. S6).

The association of Cas3 and Cas1 with prespacers suggests that Cas3 and Cas1 may interact with each other. The KD263 cells undergoing primed adaptation were processed following the ChIP protocol using Cas1- or Cas3-specific antibodies, and precipitated proteins were eluted and subject to proteolytic digestion followed by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). The MS data presented in SI Appendix, Table S3 and Fig. S8 showed that the Cas3 and the Cas1 proteins are indeed specifically coimmunoprecipitated.

The Spacer Choice Is Not Affected by the Orientation of the PPS. The hallmark of primed adaptation from plasmids is a strand bias: most acquired spacers map to the nontarget strand with respect to PPS (illustrated in Fig. 2D) (36, 49, 54, 56). The effect of PPS orientation on spacer choice (as opposed to the efficiency of spacer selection) was not investigated before, however. To address this question, we compared randomly selected 200,000 spacer sequences extracted from reads of extended CRISPR array amplicons of cells harboring either pG8mut_rev or pG8mut_dir. Within the selected set of spacers, only unique spacers acquired from both plasmids were taken into account; the quantity of individual spacers (i.e., the number of high throughput sequencing reads corresponding to a spacer with a given sequence) was not considered. Furthermore, only unique spacers derived from protospacers with PAM^{AAG} were considered (note that such spacers constitute more than 90% of spacers acquired during priming) (48, 57). Because the pG8mut_rev and pG8mut_dir plasmid backbones are identical except for the 209-bp PPS-containing insert which carries 10 AAG sequences, the remaining 103 AAGs present in either DNA strand are located in the same context in both plasmids but on different (target versus nontarget) strands with respect to PPS (Fig. 3A). With each plasmid, almost all of the “common”

AAG trinucleotides (101 out of 103) are used as PAMs to acquire spacers (Fig. 3A). A similar result was obtained for a pair of plasmids with opposite orientations of PPS in a different backbone (91 out of 102 “common” spacers were shared, SI Appendix, Fig. S7). We therefore conclude that during primed adaptation from plasmids, most protospacers with the PAM^{AAG} are used for generation of prespacers and spacers irrespective of the PPS orientation. The abundances of shared spacers acquired from nontarget and target strands were proportional to a 78:22 ratio (Fig. 3B). Notably, spacers located close to PPS (Fig. 3B, purple color) deviated from the general distribution, possibly because of more active degradation of DNA in this area. Overall, we conclude that the observed strand biases of spacer acquisition that depend on PPS orientation (Fig. 3B) are likely determined not by the kind of prespacers generated by the priming machinery but by their abundance.

Inversion of the Priming Protospacer Orientation Does Not Affect the Strand Bias of Acquired Spacers during Self-Targeting. Prespacers were originally detected in a self-targeting model system where a PPS targeted by crRNA was located in the *E. coli* genome (30). Induction of *cas* gene expression in this system led to primed adaptation with acquired spacers mapping at both sides of the PPS exhibiting a very strong (>98%) strand bias that was inverted at the PPS (30). The results of analysis of primed adaptation from plasmids with opposing PPS orientations made us interested in whether a change of PPS orientation in the self-targeting model system will affect the strand bias of acquired spacers. A self-targeting strain with crRNA recognizing PPS located in the same region of the *E. coli* genome as in the strain used in ref. 30, but on the opposite strand of DNA, was created (Materials and Methods), and spacers were acquired by cultures of both kinds of cells after *cas* gene induction was sequenced and mapped. The results are shown in Fig. 4. As can be seen, spacers derived from DNA to the left-hand side of PPS in Fig. 4A mapped to the top strand (which was either a target or nontarget strand, depending on the orientation of the PPS), while spacers derived from DNA to the right-hand side of PPS mapped to the bottom strand (which, again, was either a nontarget or target strand, depending on the orientation of the PPS). Thus, the strand bias of spacer acquisition

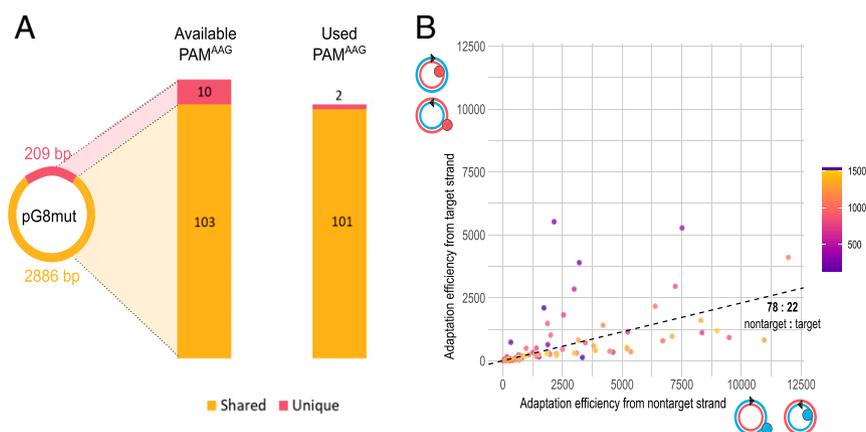


Fig. 3. Statistics of AAG PAM distribution and usage during primed adaptation by pG8mut plasmids with inverted PPS. (A) A schematic representation of a pG8mut plasmid is shown on *Left*. The 209-insert fragment containing the PPS is shown with red color. This fragment is present in different orientations in the two pG8mut variants, pG8mut_dir and pG8mut_rev; the rest of the backbone (shown with orange color) is the same in both plasmids. The first bar plot represents the total number of PAM^{AAG} on both strands of pG8mut plasmids in the PPS-containing insert (red color) and in the plasmid backbone (orange color). The second bar plot shows the number of PAM^{AAG} associated with acquired spacers from both plasmids (orange) and from one of the plasmids (red). (B) Spacers that were acquired from plasmids with each of the PPS orientations are presented on a scatter plot. Abundances of each spacer (“adaptation efficiencies”) when acquired from nontarget (*x*-axis) or target strands (*y*-axis) of each plasmid are shown. Depending on the orientation of the PPS, the top strand of the pG8mut plasmid serves as the target (in case of pG8mut_rev) or nontarget strand (in case of pG8mut_dir). The color gradient shows the distance from the PPS for individual spacers (purple, close to the PPS; yellow, distant from the PPS). The regression line corresponds to a 78:22 ratio (nontarget:target).

is not affected by the orientation of the PPS in the self-targeting model system. In contrast, the overall yield of spacers acquired from both sides was sensitive to PPS orientation: ~60% of spacers were acquired from the nontarget strand, while the remaining ~40% were from the target strand. The estimation of Cas3-mediated strand bias in the I-E CRISPR-Cas system of *E. coli* varies greatly in different works depending on experimental conditions (27, 58), mismatches between crRNA and protospacer (49), and the sequence of the priming protospacer and/or its local context (57). We obtained similar values of strand bias (60:40) with different orientations of PPS in the self-targeting system, despite the opposite local contexts of two protospacers. If Cas3 movement in secondary 3'-5' direction along the target strand is enhanced by Cas1-Cas2 binding (27), then the value of strand bias might be influenced by the expression level of Cas1-Cas2 proteins.

Similarly to the situation in the plasmid system, 54% of acquired spacers were shared between the strains with inverted PPS. The analysis of shared spacers revealed that the relative acquisition efficiency of each individual spacer remained constant when spacers acquired by each strain were compared among themselves (Fig. 4B). It has been reported that the acquisition efficiency of individual spacers negatively correlates with the number of AAG trinucleotides in the spacer sequence (48). Our observation suggests that this PAM-dependent mechanism responsible for differences in acquisition efficiency is the same for spacers originating from target and nontarget strands.

Discussion

The principal result from this work is the demonstration that same set of prespacers generated during primed adaptation by the Type I-E CRISPR-Cas system of *E. coli* is bound by Cas1, presumably as part of the Cas1-Cas2 adaptation complex, and by the Cas3 nuclease-helicase. While prespacer interaction with Cas1-Cas2 is most likely direct (based on the known role of Cas1-Cas2 in spacer acquisition), Cas3 is either physically associated with Cas1-Cas2, which allows to capture Cas1-bound prespacers with Cas3-specific antibodies, or, alternatively, Cas3 alone may generate prespacers and then pass them on to the Cas1-Cas2 adaptation complex (50). We consider the second scenario less likely, since in cells where the adaptation complex is absent or inactivated, prespacers are not detected (see ref. 30 and this work). Thus, if prespacers were generated by Cas3 alone, they must be very unstable. If Cas3 and Cas1-Cas2 associate during priming, as our data suggest, the adaptation complex must either 1) alter the specificity of the Cas3 nuclease to make it generate prespacers or 2) excise prespacers from DNA being unwound by Cas3. It is also

possible that prespacers are excised by a non-CRISPR cellular nuclease(s) from DNA unwound by Cas3 and then handed over to Cas1-Cas2.

Whichever the actual molecular mechanism, our data provide support for the existence of a primed acquisition complex (PAC) that includes Cas1-Cas2 and Cas3. The existence of a PAC was first suggested based on elegant single-molecule analysis of *E. coli* Cas3 behavior in the absence or presence of Cas1-Cas2 (27). While only a single Cas3 molecule was bound to Cascade and moved in one direction from the protospacer in the absence of Cas1-Cas2, the addition of Cas1-Cas2-enabled reiterative binding of several Cas3 molecules to stationary Cascade and Cas3 movement in both directions away from the protospacer. Unfortunately, in these experiments, the location of the Cas1-Cas2 complex during Cas3 movement was not followed, so it remains to be determined whether Cascade remains within the *E. coli* PAC. We reexamined our published data obtained using an *E. coli* self-targeting system (30) and observed high levels of accumulation of 37 to 39 nucleotide (nt) fragments mapping to the target strand of the priming protospacer and extending five to six nt downstream in the PAM-distal side in cells undergoing either primed adaptation or self-interference (SI Appendix, Fig. S9). This result suggests that PPS-bound Cascade may be removed from the target DNA by the action of Cas3 or other cellular nucleases, which would be consistent with a PAC consisting of only Cas3 and the Cas1-Cas2 complex associated with target DNA. On the other hand, the existence of PAC consisting of Cas3, Cas1-Cas2, and PPS-bound Cascade was suggested by in vitro studies in the type I-E system of *Thermobifida fusca* (52) in which concerted unidirectional movement of labeled Cas1-Cas2 and Cascade was observed in the presence of unlabeled Cas3. However, only the interaction between Cas1-Cas2 and Cascade was demonstrated in vivo in this case (52). The existence of a PAC consisting of Cas3 and the adaptation complex is also likely in other Type I CRISPR-Cas systems. In the I-F subtype, Cas3 is a default component of the priming complex because it is naturally fused to Cas2 (40, 59). In the subtype I-B system of *Pyrococcus furiosus*, the deletion of Cas3 decreased the efficiency of both primed and naive adaptation processes, which implies a functional interaction between Cas3 and the adaptation complex (41).

Our second important result is the demonstration that during priming, the choice of protospacers from which spacers are acquired qualitatively does not depend on PPS (and, therefore, bound Cascade) orientation. The PPS orientation only determines the abundance of prespacers (and acquired spacers) mapping to each DNA strand, that is, the strand bias, which is the hallmark of

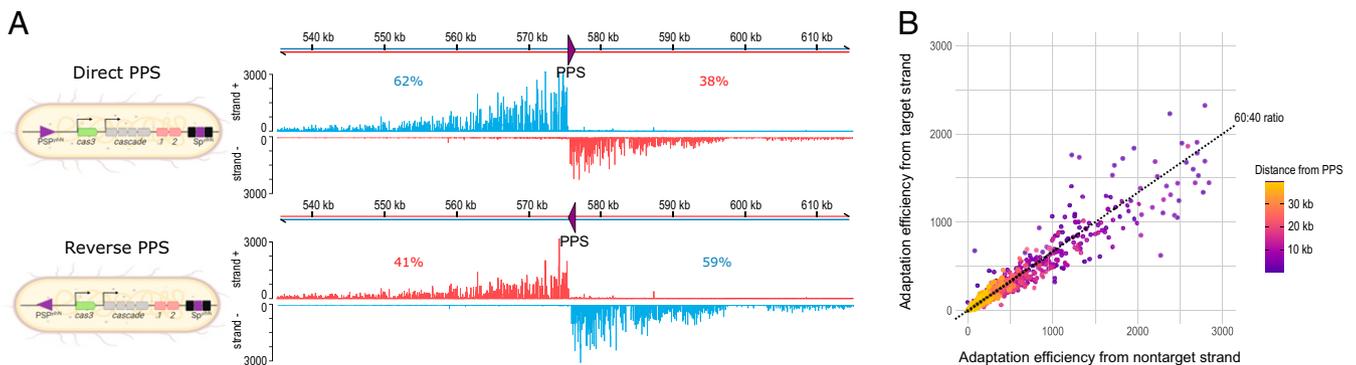


Fig. 4. Inversion of the PPS orientation does not invert the gradient of spacer acquisition. (A) The histograms show the abundance of spacers acquired from target (red) and nontarget (blue) strands around the PPS in two self-targeting cells with inverted PPS orientations (schematically shown on *Left*). The percentage of spacers acquired from target and nontarget strands is indicated above the histograms with blue or red fonts, respectively. Priming protospacers and their orientations are shown as arrows. (B) Spacers that were acquired by cells with both orientations of PPS are presented on a scatter plot. Abundances of each spacer ("adaptation efficiencies") when acquired from nontarget (*x*-axis) or target strands (*y*-axis) are shown. The color gradient shows the distance from the PPS for individual spacers. The regression line corresponds to a 60:40 ratio.

primed adaptation. To explain this finding, we propose a model shown in Fig. 5. According to the model and in agreement with earlier *in vitro* studies (27, 52), the orientation of Cascade bound to PPS determines the timing, rate, and/or efficiency of Cas3-induced degradation of DNA at both sides of PPS *in vivo*. The initial Cas3 cleavage in the nontarget strand of PPS generates a 3' end (50). Next, Cas3 moves along a nicked strand and directs Cas1–Cas2 to the unwound target strand to seek 3'-TTC-5' (AAG PAM complement) sequences that are processed by cellular nucleases (30, 60) with the formation of ~33 nt fragments associated with 5'-AAG-3' PAM and complementary ~37 nt fragments containing single-stranded 3'-NN TTC-5' extensions. Complementary ~33/~37 fragments form partially double-stranded DNA prespacers that stay bound to Cas1/Cas2 and are delivered to the CRISPR array. Prespacers are also associated with Cas3, likely through the Cas1–Cas3 interaction (see above).

Eventually, a Cas3 molecule rebinds the PPS-bound Cascade and proceeds in the 3'-5' direction along the target strand leading to a generation of prespacers composed of ~33 and ~37 nt fragments mapping to the opposite strands compared to fragments generated by the initial process. Because degradation (or, more strictly, prespacer generation) proceeds faster and/or more frequently at one side of the PPS-bound Cascade (in the PAM-proximal direction, counterclockwise in the model shown in Fig. 5A, left side), a strand bias in the abundance of prespacers and corresponding acquired spacers is created. The strand bias thus appears to be due to two processes. First, it arises because

Cas3/the priming complex moves unidirectionally, and only the 5'-CTT-3' sequences located on the strand displaced by the Cas3 helicase are recognized (Fig. 2). Second, the faster/more efficient degradation of DNA at one side of PPS in the context of short circular plasmids leaves less substrate for prespacer generation on the PAM-distal side of PPS. The reversal of the PPS orientation changes the strand bias, but the choice of protospacers is not affected, as it is determined solely by the presence of PAM^{AAG} sequences in degraded DNA. As a result, the sets of unique prespacers generated by the priming process and of acquired spacers are the same irrespective of PPS orientation. In Fig. 5, this is illustrated by two hypothetical prespacers that either become highly abundant or scarce when the orientation of PPS in the plasmid is changed. Somewhere along this process, Cas3 or other cellular nucleases generate cuts in the target strand downstream and upstream of PPS, effectively removing the effector from the target DNA (Fig. 5).

In contrast to primed adaptation from plasmids, during self-targeting of genomic DNA, the overall strand bias of acquired spacers is marginal since spacers acquired from DNA at each side of PPS have opposing strand biases (30, 57). While more active processive degradation of DNA initiated in the direction of PPS PAM on short circular plasmids removes DNA available for degradation/prespacer generation in the PAM-distal direction (see above), degradation of DNA at both sides of PPS in the self-targeting model proceeds independently, and while degradation in the PPS PAM-proximal direction likely starts earlier,

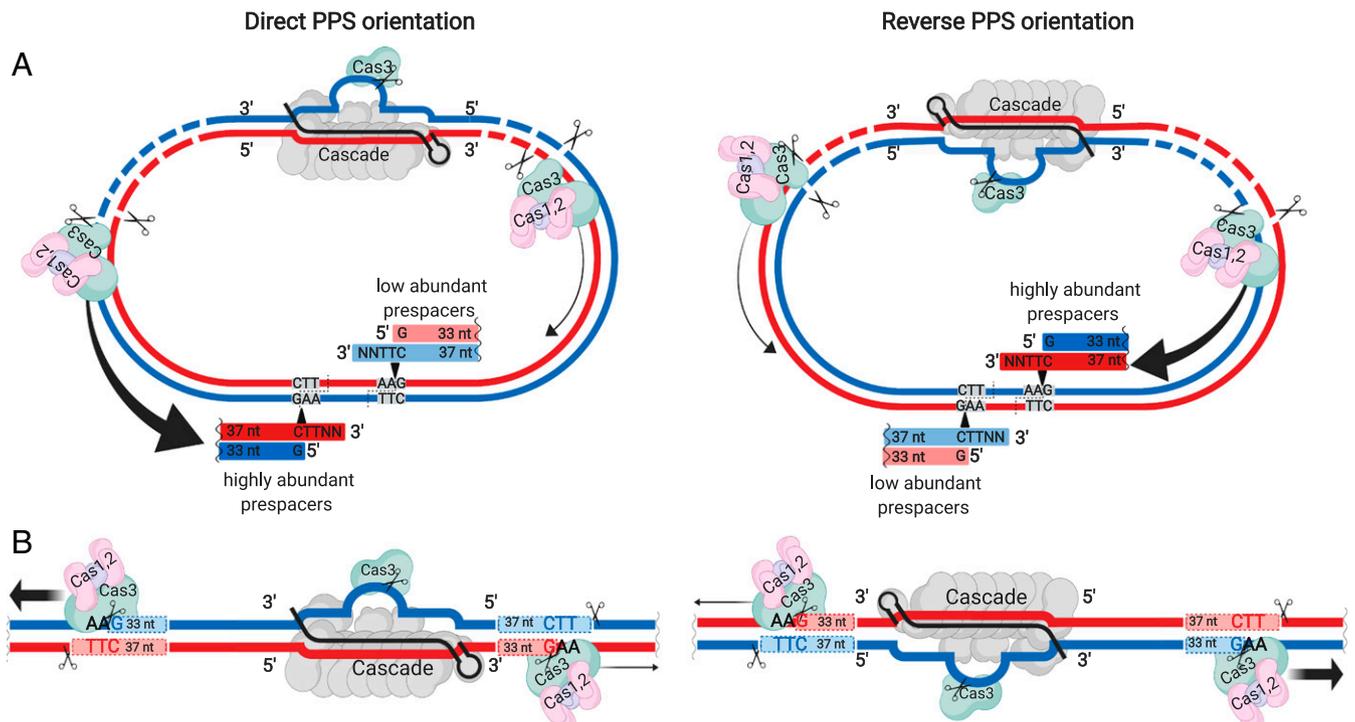


Fig. 5. A model of prespacer selection and strand bias generated during priming. (A) Two plasmids with opposite orientations of the PPS are schematically shown. In each plasmid, the target strand is shown with red color, and the nontarget strand is blue. The first priming complex associates with Cascade and proceeds, after a Cas3-generated cut in the displaced nontarget strand of the R-loop complex, in the 3'-5' direction of the nontarget strand (thick arrow), recognizing 3'-TTC-5' sequences on the opposite strand and generating ~33/~37 nt prespacers. The second priming complex moves in the 3'-5' direction of the target strand (thin arrow). The prespacer and spacer choice is not affected by the orientation of PPS (Left and Right); however, prespacers are excised with a different efficiency depending on which strand serves as a target or nontarget strand. Prespacers generated during the movement along the nontarget strand are more abundant (shown in bright color) compared to prespacers generated during the movement along the target strand (shown in pale color). (B) Two opposite orientations of the PPS located on a linear DNA (like in self-targeting cells) are shown. The color scheme is like in A. The priming complex associates with the 3' end of the nontarget strand after the initial cleavage in the R-loop and translocates (thick arrow) in the 3'-5' direction, recognizing CTT sequences on the target strand and excising prespacers. Another priming complex associates with the target strand and also proceeds in the 3'-5' direction (thin arrow). Unlike the situation with PPS-containing plasmids (A), the unwinding/degradation of DNA in PAM-proximal and PAM-distal directions of PPS are independent because the size of the chromosome greatly exceeds the processivity of the complex.

the process in the opposite direction eventually catches up, generating opposing strong strand biases, which, when all acquired spacers are considered, practically cancel each other out. While one could naively expect that a change in the orientation of PPS should lead to a reversal of the strand bias in this system, this is not observed (Fig. 5B). Thus, the orientation of spacers selected at each side of PPS does not depend on the PPS orientation but on which DNA strand the 3' end associated with the Cas3 helicase domain is located.

Materials and Methods

Strains and Plasmids. The *E. coli* strains used in this study are listed in *SI Appendix, Table S1*. The *E. coli* strain NMG24 was also constructed by Red recombinase-mediated gene replacement integrating the inverted PPS sequence into the BW25113 genome within the *yihN* gene. The presence of inverted PPS was confirmed by the sequencing of an appropriate PCR amplicon, and it was next transferred to KD403 by P1 transduction (61). The kanamycin resistance gene was next removed using the pCP20 plasmid by the site-directed recombination (62).

The plasmid pCas3 carrying the *cas3* gene was constructed by cloning the *E. coli cas3* gene amplified with *Cas3_F*–*Cas3_R* primers downstream of the inducible T7 RNA polymerase promoter in the linearized pYM390 vector (a generous gift of Alfred Anton, University of York, York, United Kingdom) by Gibson Assembly (New England Biolabs). The resulting Cas3 protein is fused to N-terminal 6⁺His-SUMO (small ubiquitin-like modifier) tag. Plasmids pG8mut_dir and pG8mut_rev carrying a 209-bp fragment of the M13 phage (genome positions 1,311 to 1,519) and containing the G8 protospacer with escaper mutation C1T at the first position of the protospacer have been described earlier (54, 55). Plasmids pB_Dir and pB_Rev are derivatives of the pBBR6 plasmid. The PPS-containing fragment in direct or reverse orientation was amplified from, respectively, pG8mut_dir and pG8mut_rev with B6_F_dir and B6_R_dir or B6_F_rev and B6_R_rev primers. Amplicons were then used to perform site-directed PCR mutagenesis of the pBBR6 plasmid. The parental vector plasmid without the insertion was digested with *DpnI*, and the final product was electroplated into *E. coli* BW25113. Selected clones were subjected to Sanger sequencing in order to verify the presence of desired insert.

Cas3 Purification. The *E. coli* BL21 Star (DE3) strain was used for expression of Cas3. Cells transformed with pCas3 were grown in Luria broth (LB) containing 30 μ g/mL kanamycin at 37 °C at 200 rpm until OD₆₀₀ reached 0.2 and was then transferred to 16 °C. After 20 min of growth, 0.2 mM isopropyl β -D-1-thiogalactopyranoside was added, and cells were cultivated overnight. Cells were harvested by centrifugation for 20 min at 5,000 \times *g* at 4 °C, and the pellet was immediately resuspended in Buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 10 mM imidazole, and 10% glycerol) containing 1 mg/mL lysozyme, 1 tablet/100 mL Complete Protease Inhibitor Mixture (Roche), and 20 μ l/100 mL deoxyribonuclease I and ribonuclease A (RNase A) (both Thermo Fisher). The mixture was incubated on ice for 30 min and then sonicated using the Vibra-Cell VCX 130 machine (Sonic) at 100% power for 10 min. The suspension was centrifuged at 16,000 \times *g* for 30 min at 4 °C. The supernatant was loaded onto a 1-mL His-Trap FF column (GE Healthcare) pre-equilibrated with Buffer A. After loading the sample, the column was washed with 20 mL of ice-cold Buffer B (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 10% glycerol). Bound proteins were eluted with a linear (0 to 100%) gradient of Buffer C (20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 250 mM imidazole, and 10% glycerol). Eluted fractions were analyzed by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis. Fractions containing Cas3 were concentrated by centrifugation at 4,000 \times *g* for 30 min at 4 °C in Amicon Ultra-15 100K cutoff Centrifugal Filter Units (Merck Millipore). The concentrated sample was loaded onto a HiLoad 16/600 Superdex 200 size exclusion column (GE Healthcare) pre-equilibrated in Buffer D (20 mM Tris-HCl, pH 8.0, 200 mM NaCl, and 1 mM dithiothreitol [DTT]). Fractions containing Cas3 were pooled and desalted into Buffer P (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 1 mM DTT) in the presence of SUMO Protease overnight at 4 °C. The resulting mixture of proteins was loaded onto a 1 mL His-Trap FF column (GE Healthcare) pre-equilibrated with Buffer B. Cas3 without the SUMO tag was collected in the flowthrough and concentrated as described above followed by additional purification on a HiLoad 16/600 Superdex 200 size exclusion column (GE Healthcare) pre-equilibrated in Buffer D. Fractions containing the Cas3 protein were concentrated by centrifugation at 4,000 \times *g* for 30 min at 4 °C in Amicon Ultra-15 100K cutoff Centrifugal Filter Units (Merck Millipore), flash frozen in liquid nitrogen, and stored at –80 °C.

Antibody Generation and Purification. Recombinant Cas3 protein was used for rabbit immunization. Antisera were tested by Western blotting against Cas3 preparation used for immunization. For antibody purification, Cas3 was immobilized on cyanogen bromide-activated Sepharose 4B (Sigma-Aldrich) following the manufacturer's instructions. Briefly, activated beads were mixed with recombinant Cas3 (5 to 10 mg/1 mg dry beads) in coupling buffer (100 mM NaHCO₃, pH 8.3 to 8.5, and 500 mM NaCl) and incubated overnight at 4 °C in a tumbler. After washing thrice with the coupling buffer beads were incubated in 15 mL 100 mM Tris-HCl, pH 8.0 for 2 h at room temperature in a tumbler. Beads were transferred into an empty XK 16/20 (GE Healthcare) column attached to an ÄKTA start chromatography system (GE Healthcare), washed three times with ice-cold 100 mM Tris-HCl, pH 8.0, and 500 mM NaCl, and followed by a wash with ice-cold 100 mM sodium acetate and 500 mM NaCl, pH 4.0. The beads were finally equilibrated for 30 min at a low flow rate with ice-cold phosphate-buffered saline (PBS; 2.7 mM KCl, 137 mM NaCl, 1.76 mM potassium phosphate, and 10 mM sodium phosphate, pH 7.4). Washed beads were incubated with antiserum (5 mL beads with 5 mL antiserum) overnight at 4 °C in a tumbler. The beads were washed with PBS, and bound antibodies were eluted with ice-cold 200 mM glycine pH 2.8. A total of 1 mL aliquots were collected into tubes containing 27 μ l 3 M Tris-HCl, pH 8.8, and 100 μ l 3 M KCl. Aliquots with antibodies were pooled and dialyzed against the PBS buffer overnight at 4 °C. The concentration of anti-Cas3 antibodies was adjusted to 1 mg/mL by centrifugation at 4,000 \times *g* for 30 min at 4 °C in Amicon Ultra-15 100K cutoff Centrifugal Filter Units (Merck Millipore). Antibodies were stored at 4 °C with 0.2% Na₃N.

The anti-Cas1 antibody was described earlier at ref. 55.

CRISPR System Induction. *E. coli* KD263 were transformed with pG8mut_dir, pG8mut_rev, pB_Dir, or pB_Rev plasmids. Transformants were selected on LB agar plates containing 100 μ g/mL ampicillin for pT7Blue-based plasmids or 15 μ g/mL gentamicin for pB-based plasmids. Individual transformants were grown in liquid medium and induced as described (54). Aliquots of induced and uninduced cultures were processed for ChIP 3 h postinduction.

Spacer Acquisition Analysis. Cells collected from aliquots of induced and uninduced cultures were subjected to PCR with primers ECLDR_F and M13G8_R annealing at the leader region of the CRISPR array and to the G8 spacer, respectively. The results were analyzed by agarose gel electrophoresis. To analyze newly acquired spacers, the PCR products corresponding to the expanded CRISPR array were gel purified with a QIAquick Gel Extraction Kit (QIAGEN) and sequenced using a Miseq Illumina in pair-end 150-bp long reads mode according to manufacturer's protocols. Analysis of reads and spacer mapping was performed as described earlier (63). To compare spacer choices between different experiments, the Pearson coefficient, which is a measure of the linear dependence between two variables, was used. A Pearson coefficient of 1 indicates total positive linear correlation, 0, no linear correlation, and –1, total negative linear correlation.

ChIP Sequencing. ChIP was performed as described earlier (55). Briefly, formaldehyde was added to 20 mL induced or uninduced cultures to a final concentration of 1%, and reactions were allowed to proceed for 20 min at room temperature. The reaction was quenched by the addition of glycine (0.5 M final concentration) and incubation for additional 5 min. The cells were pelleted by centrifugation and washed three times with 1 \times tri-buffered saline (10 mM Tris-HCl, pH 7.5, and 150 mM NaCl). A total of 1 mL lysis buffer (10 mM Tris-HCl, pH 8.0, 20% sucrose, 50 mM NaCl, 10 mM ethylenediaminetetraacetic acid (EDTA), 20 mg/mL lysozyme, and 0.1 mg/mL RNase A) was added, and the samples were incubated at 37 °C for 30 min. After the addition of 4 mL IP buffer (50 mM Hepes-KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, and 1 mM phenylmethylsulfonyl fluoride) the samples were sonicated on a Vibra-Cell VCX 130 machine (Sonic) at 80% power for 5 min yielding DNA fragments with an average length of 100 to 200 bp. This and on subsequent steps were performed on ice. After centrifugation, 800 μ l of supernatant was preincubated with 20 μ l Protein A/G Sepharose beads (Thermo Scientific) to pull down proteins unspecifically interacting with the resin, and unbound material was combined with 30 μ l bovine serum albumin-blocked Protein A/G Sepharose and 2.5 mg/mL anti-Cas1 or anti-Cas3 antibody and incubated overnight on a rotary platform. The standard steps of washing with the IP buffer, high salt IP buffer (the IP buffer supplied with 500 mM NaCl), wash buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet P-40, and 0.5% sodium deoxycholate), tris-EDTA buffer (10 mM Tris-HCl, pH 8.0, and 1 mM EDTA), and elution in elution buffer (10 mM Tris-HCl, pH 7.5, 10 mM EDTA, and 1% SDS) were performed as described (64). Immunoprecipitated samples and sheared DNA samples before IP (input) were

de-crosslinked in 0.5× elution buffer containing 0.8 mg/mL Proteinase K at 42 °C for 2 h followed by incubation at 65 °C for 6 h. DNA was ethanol precipitated in the presence of added glycogen, dissolved in 20 µl of MilliQ water and subjected to next generation sequencing library preparation using Accel-NGS 15 Plus DNA Library Kit (Swift Biosciences). Resulting libraries were sequenced by MiniSeq Illumina in pair-end 150-bp long reads mode, according to the manufacturer's protocols. Three biological replicates or every immunoprecipitation experiment were performed.

ChIP Sequencing Data Analysis. Adapter sequences and low-quality (Phred score < 15) sequences were removed from reads with Trimmomatic (65). Reads were mapped separately to the KD263 genome or plasmid genomes by Bowtie 2 (66) in a local alignment mode, with minimum and maximum fragment lengths set to 20 and 1,500, respectively. Lengths of fragments between mapped paired-end reads were extracted from the same file (TLEN field). For clarity, distributions of fragment lengths were smoothed (loess smoothing with span = 0.05) and visualized with R.

For analysis of short fragments, we assembled paired reads with PEAR (67), leaving only concatenated sequences with lengths <50 nt. Assembled reads were mapped to plasmid sequences by Bowtie 2 in the "local" mode. The precise lengths of short fragments were extracted from the CIGAR field of the resulting sam file with an R script. Fragments with characteristic lengths of 32 to 34 nt and 36 to 38 were selected and analyzed separately. Logo

representations of sequence specificity logos were created for nonunique 33 and 37 nt fragments with the python package logomaker (68).

Proteomic Analysis of Immune Precipitates. The ChIP procedure was performed using KD263 cells containing the pG8mut_dir plasmid as described above up to the elution stage. In addition to Cas1 and Cas3 specific polyclonal antibodies, similarly prepared rabbit antibodies specific to a small terminase subunit of *Thermus thermophilus* bacteriophage G18 were used as a control. Eluates were subjected to proteomic analysis as described (69). Raw data were processed using the MaxQuant software (70) (version 1.6.10.43) with the built-in search engine Andromeda (71). iBAQ (intensity-based absolute quantification) values were calculated by MaxQuant as the (raw) intensities divided by the number of theoretical peptides. The iBAQ* normalized meaning was calculated for each protein as $iBAQ^* = IP_iBAQmean_IP - iBAQmean_Control$.

Data Availability. Illumina sequencing result data have been deposited in the Sequence Read Archive (PRJNA723343) (72). All other study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. This study was supported by the Russian Science Foundation (Grant 19-74-00118 to O.M.), the Russian Science Foundation (Grant 19-74-20130 to S.M.), NIH National Institute of General Medical Sciences (Grant RO1 10407 to K.S.), and the Russian Science Foundation (Grant 19-14-00323 to K.S.).

- Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- L. A. Marraffini, E. J. Sontheimer, CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845 (2008).
- S. J. J. Brouns *et al.*, Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- P. M. Nussenzweig, L. A. Marraffini, Molecular mechanisms of CRISPR-cas immunity in bacteria. *Annu. Rev. Genet.* **54**, 93–120 (2020).
- J. van der Oost, M. M. Jore, E. R. Westra, M. Lundgren, S. J. J. Brouns, CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407 (2009).
- K. S. Makarova, Y. I. Wolf, E. V. Koonin, Classification and nomenclature of CRISPR-cas systems: Where from here? *CRISPR J.* **1**, 325–336 (2018).
- I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in *Escherichia coli*. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
- J. McGinn, L. A. Marraffini, Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).
- E. Charpentier, H. Richter, J. van der Oost, M. F. White, Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**, 428–441 (2015).
- H. Nishimasu, O. Nureki, Structures and mechanisms of CRISPR RNA-guided effector nucleases. *Curr. Opin. Struct. Biol.* **43**, 68–78 (2017).
- M. M. Jore *et al.*, Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat. Struct. Mol. Biol.* **18**, 529–536 (2011).
- B. Wiedenheft *et al.*, Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**, 486–489 (2011).
- M. D. Szczelkun *et al.*, Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9798–9803 (2014).
- S. Nimkar, B. Anand, Cas3/IC mediated target DNA recognition and cleavage during CRISPR interference are independent of the composition and architecture of Cascade surveillance complex. *Nucleic Acids Res.* **48**, 2486–2501 (2020).
- H. Zhao *et al.*, Crystal structure of the RNA-guided immune surveillance Cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014).
- Y. Xiao *et al.*, Structure basis for directional R-loop formation and substrate handover mechanisms in type I CRISPR-cas system. *Cell* **170**, 48–60.e11 (2017).
- C. Xue, Y. Zhu, X. Zhang, Y.-K. Shin, D. G. Sashital, Real-time observation of target search by the CRISPR surveillance complex cascade. *Cell Rep.* **21**, 3717–3727 (2017).
- L. A. Cooper, A. M. Stringer, J. T. Wade, Determining the specificity of cascade binding, interference, and primed adaptation *in vivo* in the *Escherichia coli* type I-E CRISPR-cas system. *mBio* **9**, e02100-17 (2018).
- J. L. Weissman, A. Stoltzfus, E. R. Westra, P. L. F. Johnson, Avoidance of self during CRISPR immunization. *Trends Microbiol.* **28**, 543–553 (2020).
- O. Musharova *et al.*, Systematic analysis of Type I-E *Escherichia coli* CRISPR-Cas PAM sequences ability to promote interference and primed adaptation. *Mol. Microbiol.* **111**, 1558–1570 (2019).
- M. Rutkauskas *et al.*, Directional R-loop formation by the CRISPR-cas surveillance complex cascade provides efficient off-target site rejection. *Cell Rep.* **10**, 1534–1543 (2015).
- B. Gong *et al.*, Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16359–16364 (2014).
- M. L. Hochstrasser *et al.*, CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6618–6623 (2014).
- Y. Xiao, M. Luo, A. E. Dolan, M. Liao, A. Ke, Structure basis for RNA-guided DNA degradation by cascade and Cas3. *Science* **361**, eaat0839 (2018).
- S. Mulepati, S. Bailey, In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* **288**, 22184–22192 (2013).
- L. Loeff, S. J. J. Brouns, C. Joo, Repetitive DNA reeling by the cascade-cas3 complex in nucleotide unwinding steps. *Mol. Cell* **70**, 385–394.e3 (2018).
- S. Redding *et al.*, Surveillance and processing of foreign DNA by the *Escherichia coli* CRISPR-cas system. *Cell* **163**, 854–865 (2015).
- Y. Huo *et al.*, Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* **21**, 771–777 (2014).
- E. R. Westra *et al.*, CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol. Cell* **46**, 595–605 (2012).
- A. A. Shiriaeva *et al.*, Detection of spacer precursors formed in vivo during primed CRISPR adaptation. *Nat. Commun.* **10**, 4603 (2019).
- J. K. Nuñez, L. B. Harrington, P. J. Kranzusch, A. N. Engelman, J. A. Doudna, Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535–538 (2015).
- J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. A. Doudna, Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015).
- J. K. Nuñez, L. B. Harrington, P. J. Kranzusch, A. N. Engelman, J. A. Doudna, Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **534**, 513–514 (2016).
- Z. Arslan, V. Hermans, R. Wurm, R. Wagner, Ü. Pul, Detection and characterization of spacer integration intermediates in type I-E CRISPR-Cas system. *Nucleic Acids Res.* **42**, 7884–7893 (2014).
- I. Ivančić-Baćić, S. D. Cass, S. J. Wearne, E. L. Bolt, Different genome stability proteins underpin primed and naïve adaptation in *E. coli* CRISPR-Cas immunity. *Nucleic Acids Res.* **43**, 10821–10830 (2015).
- K. A. Datsenko *et al.*, Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3**, 945 (2012).
- A. Levy *et al.*, CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**, 505–510 (2015).
- E. Semenova *et al.*, Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 10098–10103 (2011).
- B. R. Levin, S. Moineau, M. Bushman, R. Barrangou, The population and evolutionary dynamics of phage and bacteria with CRISPR-mediated immunity. *PLoS Genet.* **9**, e1003312 (2013).
- C. Richter *et al.*, 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 (2014).
- S. Garrett *et al.*, Primed CRISPR DNA uptake in *Pyrococcus furiosus*. *Nucleic Acids Res.* **48**, 6120–6135 (2020).
- M. Li, R. Wang, D. Zhao, H. Xiang, Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Res.* **42**, 2483–2492 (2014).
- C. Rao, D. Chin, A. W. Ensminger, Priming in a permissive type IC CRISPR-Cas system reveals distinct dynamics of spacer acquisition and loss. *RNA* **23**, 1525–1538 (2017).
- T. J. Nicholson *et al.*, Bioinformatic evidence of widespread priming in type I and II CRISPR-Cas systems. *RNA Biol.* **16**, 566–576 (2019).
- D. C. Swarts, C. Moser, M. W. J. van Passel, S. J. J. Brouns, CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7**, e35888 (2012).
- D. Vorontsova *et al.*, Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res.* **43**, 10848–10860 (2015).
- P. M. Nussenzweig, J. McGinn, L. A. Marraffini, Cas3 cleavage of viral genomes primes the acquisition of new immunological memories. *Cell Host Microbe* **26**, 515–526.e6 (2019).

Musharova *et al.*

Pre-spacers formed during primed adaptation associate with the Cas1–Cas2 adaptation complex and the Cas3 interference nuclease–helicase

48. O. Musharova *et al.*, Avoidance of trinucleotide corresponding to consensus protospacer adjacent motif controls the efficiency of prespacer selection during primed adaptation. *mBio* **9**, e02169-18 (2018).
49. P. C. Fineran *et al.*, Degenerate target sites mediate rapid primed CRISPR adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1629–E1638 (2014).
50. T. Künne *et al.*, Cas3-Derived target DNA degradation fragments fuel primed CRISPR adaptation. *Mol. Cell* **63**, 852–864 (2016).
51. S. Kim *et al.*, Selective loading and processing of prespacers for precise CRISPR adaptation. *Nature* **579**, 141–145 (2020).
52. K. E. Dillard *et al.*, Assembly and translocation of a CRISPR-cas primed acquisition complex. *Cell* **175**, 934–946.e15 (2018).
53. S. Shmakov *et al.*, Pervasive generation of oppositely oriented spacers during CRISPR adaptation. *Nucleic Acids Res.* **42**, 5907–5916 (2014).
54. E. Semenova *et al.*, Highly efficient primed spacer acquisition from targets destroyed by the *Escherichia coli* type I-E CRISPR-Cas interfering complex. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7626–7631 (2016).
55. O. Musharova *et al.*, Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* **45**, 3297–3307 (2017).
56. R. H. J. Staals *et al.*, Interference-driven spacer acquisition is dominant over naive and primed adaptation in a native CRISPR-Cas system. *Nat. Commun.* **7**, 12853 (2016).
57. A. M. Stringer, L. A. Cooper, S. Kadaba, S. Shrestha, Characterization of primed adaptation in the *Escherichia coli* type IE CRISPR-cas system. *bioRxiv* [Preprint] 10.1101/2020.02.10.942821 (Accessed 12 February 2020).
58. H. Morisaka *et al.*, CRISPR-Cas3 induces broad and unidirectional genome editing in human cells. *Nat. Commun.* **10**, 5302 (2019).
59. M. F. Rollins *et al.*, Cas1 and the Csy complex are opposing regulators of Cas2/3 nuclease activity. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5113–E5121 (2017).
60. E. Kurilovich *et al.*, Genome maintenance proteins modulate autoimmunity mediated primed adaptation by the *Escherichia coli* type I-E CRISPR-cas system. *Genes (Basel)* **10**, 872 (2019).
61. L. C. Thomason, N. Costantino, D. L. Court, *E. coli* genome manipulation by P1 transduction. *Curr. Protoc. Mol. Biol.* **Chapter 1**, Unit 1.17 (2007).
62. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645 (2000).
63. E. Semenova *et al.*, The Cas6e ribonuclease is not required for interference and adaptation by the *E. coli* type I-E CRISPR-Cas system. *Nucleic Acids Res.* **43**, 6049–6061 (2015).
64. C. Kahramanoglou *et al.*, Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli*. *Nucleic Acids Res.* **39**, 2073–2091 (2011).
65. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
66. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
67. J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: A fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
68. A. Tareen, J. B. Kinney, Logomaker: Beautiful sequence logos in Python. *Bioinformatics* **36**, 2272–2274 (2020).
69. I. Laptsev *et al.*, METTL15 interacts with the assembly intermediate of murine mitochondrial small ribosomal subunit to form m4C840 12S rRNA residue. *Nucleic Acids Res.* **48**, 8022–8034 (2020).
70. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
71. J. Cox *et al.*, Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
72. O. Musharova *et al.*, Sequencing of fragments associated with Cas1 and Cas3 during priming in I-E system of *E. coli*. National Center for Biotechnology Information Sequence Read Archive (NCBI SRA). <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA723343>. Deposited 20 April 2021.