Review Article

Detection of CRISPR adaptation

Anna Shiriaeva, Ivan Fedorov, Danylo Vyhovskyi and Konstantin Severinov

1Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow 121205, Russia; 2Waksman Institute, Rutgers, the State University of New Jersey, Piscataway, NJ 08854, U.S.A.; 3Institute of Gene Biology, Russian Academy of Sciences, Moscow 119334, Russia; 4Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow 119334, Russia

Correspondence: Konstantin Severinov (severik@waksman.rutgers.edu)

Prokaryotic adaptive immunity is built when short DNA fragments called spacers are acquired into CRISPR (clustered regularly interspaced short palindromic repeats) arrays. CRISPR adaptation is a multistep process which comprises selection, generation, and incorporation of prespacers into arrays. Once adapted, spacers provide immunity through the recognition of complementary nucleic acid sequences, channeling them for destruction. To prevent deleterious autoimmunity, CRISPR adaptation must therefore be a highly regulated and infrequent process, at least in the absence of genetic invaders. Over the years, ingenious methods to study CRISPR adaptation have been developed. In this paper, we discuss and compare methods that detect CRISPR adaptation and its intermediates in vivo and propose suppressing PCR as a simple modification of a popular assay to monitor spacer acquisition with increased sensitivity.

Introduction

CRISPR-Cas are diverse (two classes, six types [1–3]) prokaryotic adaptive immunity systems that protect cells from phages and other mobile genetic elements (MGEs) [4,5]. They consist of CRISPR arrays and CRISPR-associated cas genes [6,7]. CRISPR arrays are composed of identical or highly similar repeats separated by unique DNA sequences called spacers [6,7]. The total number of spacers in array varies from one to several hundreds [6,8]. The source of the vast majority (≈93%) of spacers remains unknown, they constitute the ‘dark matter’ of CRISPR [9]. Most of the remaining spacers map to MGEs and can be regarded as memories of prior encounters that cells store in CRISPR arrays [9]. Upstream of the CRISPR loci, there is an AT-rich sequence called the ‘leader’ [7]. CRISPR arrays are transcribed from a promoter located in the leader and the primary transcript is processed into CRISPR RNAs (crRNAs) containing a single spacer and flanking sequences derived from repeats [10–17]. Cas proteins together with crRNAs form effector complexes (Cascade complex in the type I-E system of Escherichia coli) that recognize ‘protospacers’ — DNA or, sometimes, RNA sequences, complementary to a crRNA spacer [13,18–20]. Recognition of protospacers in MGEs leads to their destruction [18–20].

CRISPR immunity is built during CRISPR adaptation, a process which entails incorporation of new spacers in the array [4]. New spacers are typically incorporated at the boundary between the leader and the first repeat and, therefore, the chronological order of spacer acquisition matches the inverse order of spacers in the array [4,21,22]. For every acquired spacer, a new copy of repeat is generated [4,21,22]. Two most conserved Cas proteins, Cas1 and Cas2, common to almost all CRISPR-Cas systems, catalyze integration of spacer precursors (prespacers) into arrays [23–25]. Generally, the acquisition of spacers is not specifically targeted to MGEs and thus spacers from cell’s own genome can also be acquired [23,26]. This can result in auto-immune response inhibiting cell growth [27–29]. Not surprisingly, CRISPR adaptation is a tightly controlled process that normally proceeds with very low efficiency and can be difficult to detect both in natural settings and in laboratory experiments. Several methods for the detection of CRISPR adaptation have been developed and helped to shed light on molecular mechanisms governing spacer choice. These methods and their limitations are discussed below.
Selection-based methods of detection of CRISPR adaptation in individual cells or clones

An obvious way to detect the acquisition of a new spacer is to amplify the leader-proximal end of CRISPR array with a pair of primers: one matching the leader, and another matching an internal, pre-existing spacer [23,30,31]. Since new spacers are usually incorporated in front of the first, leader-proximal repeat, and result in repeat duplication [4,21–23], detection of PCR-products extended by integral number of spacer-repeat units reveals CRISPR adaptation events. However, since spacer acquisition can be very infrequent, specific selection of adapted cells is required (Figure 1). Examples of such selections include obtaining colonies of BIMs (bacteriophage insensitive mutants) (Figure 1A) [4,21,22,31,32] or PIMs (plasmid interfering mutants) (Figure 1B) [30,31,33]. These methods are cheap and do not require genetic manipulation of cells under study but they are biased towards interference-proficient spacers acquired from MGEs and thus cannot be used to detect spacers that do not lead to interference against MGEs or lead to self-interference due to acquisition of a spacer from cell's own genome (depending on the CRISPR-Cas subtype, when interference is inactivated, such spacers can constitute from 2 to 99% of acquired spacers [23,26,34,35]).

Two powerful experimental systems have been developed to overcome these limitations and increase the sensitivity of the detection of spacer acquisition events [36,37] (Figure 1C,D). Both systems are based on a reporter gene introduced upstream of the leader sequence of a specifically designed miniaturized E. coli CRISPR array. The reporter is transcribed from a promoter located downstream of the array in a direction opposite to the direction of leader-initiated CRISPR array transcription. The resulting mRNA includes a start codon followed by the leader-CRISPR array segment (cloned in reverse orientation) and the sequence of the reporter which does not have a translational start of its own. In cells with unexpanded CRISPR arrays translation of the reporter ORF is prevented due to an in-frame stop codon within the leader. Insertion of an additional 61-bp long unit (33-bp spacer/28-bp repeat) changes the reading frame and allows the synthesis of the reporter leading to either chloramphenicol resistance [36] (Figure 1C) or fluorescence [37] (Figure 1D) of cells that acquired a spacer. Rare chloramphenicol-resistant colonies can be directly screened for CRISPR array expansion by PCR. With the fluorescent protein-based system, live fluorescent microscopy is used to observe and quantify cells that acquired spacers [37]. Though this has not been implemented yet, the use of FACS (fluorescence-activated cell sorting) should allow one to enrich the population of cells with expanded arrays for downstream analysis. With both systems, the acquisition of spacers that carry stop codons located in the reading frame of the reporter remains undetected. Likewise, incorporation of more than one spacer-repeat unit or incorporation of a single non-standard spacer that fails to restore the reporter reading frame will be undetected. Finally, CRISPR-Cas systems where incorporation of a standard spacer-repeat unit does not shift the reading frame (i.e. introduces an insertion whose length is $n \times 3$ bp, where $n$ is an integral number of nucleotides) cannot be studied.

Detection of CRISPR adaptation in cell populations

In early studies of adaptation, the sequences of newly acquired spacers were determined for individual colonies by Sanger sequencing [4,21–23,30–33]. To analyze millions of CRISPR arrays in a single experiment, high-throughput sequencing (HTS) is usually used [38–40]. This allows one to study biases in spacer length, the distribution of corresponding protospacers along different DNA sources and their nucleotide composition [34,35,38–52]. In principle, with sufficient sequencing depth, HTS of total genomic DNA purified from a culture should reveal reads corresponding to expanded arrays [53]. In a model system of E. coli cultures over-producing the Cas1–Cas2 adaptation protein complex and transformed with spacer-sized oligonucleotides, ~350× genomic coverage allowed to confidently detect CRISPR array expansion that occurred in ~10% of cells [54]. Moreover, rarer off-target integration events elsewhere in the genome were also detected [54].

While clearly powerful and unbiased, the shotgun sequencing approach requires high sequencing coverage and provides very low (dozens) numbers of reads corresponding to expanded arrays making it unsuitable for studies aimed at qualitative understanding of spacer selection preferences [54]. Therefore, the common strategy is to prepare PCR amplicons of arrays from cultures undergoing CRISPR adaptation and then subject them for HTS [38–40]. Gel-electrophoresis is used to separate amplicons of initial, unexpanded CRISPR arrays (+0) from those that acquired one (+1), two (+2) or more spacer-repeat units (Figure 2A).

The main problem with PCR-based in-culture methods of detection of CRISPR adaptation is their low sensitivity due to more efficient amplification of shorter (and, in most interesting cases, much more abundant)
Figure 1. Selection-based methods to detect CRISPR adaptation.

Detection of acquisition of interference-proficient spacers (S⁺) in bacteriophage insensitive mutants (A) or in cells that lost plasmids (B). The structures of unexpanded (UA) CRISPR arrays in cells from an initial culture or of expanded (EA) arrays in cells that either survived phage infection or lost the plasmid are shown at the top. The leader (light peach rectangle), CRISPR repeats (white rectangles), pre-existing spacer S₀ (dark peach rectangle), and newly acquired spacer S⁺ (turquoise rectangle) are shown. Acquisition of spacers is tested by PCR with primers matching the leader (light peach arrow), and the S₀ spacer (dark peach arrow). Amplicons from expanded and unexpanded arrays are shown as brackets below the primers. In A, a colony formed after phage infection is directly tested by PCR and results of agarose gel electrophoresis of amplicons obtained with starting cells (1) and the phage-resistant colony (2) are schematically shown on the right. In B, a liquid culture is inoculated with cells from an antibiotic-resistant (Ab-resistant) plasmid-bearing colony (1) with unexpanded CRISPR array. After growth and plating on non-selective (−Ab) medium, the presence of the plasmid is tested by streaking colonies (2, 3, 4) on plates with and without Ab. Cells from the antibiotic-sensitive colony (2) are further tested by PCR to reveal a spacer acquisition event. M, molecular-weight size marker. Detection of CRISPR adaptation using cat (C) or yfp (D) reporter systems. Insertion of an
inverted leader/single-spacer CRISPR array disrupts the reading frame of *cat* (C) or *yfp* (D) reporter genes. The elements of CRISPR arrays are shown as in A and B. *P*_reporter — a promoter directing the synthesis of reporter mRNA; *P*<sub>CRISPR</sub> — a promoter directing the synthesis of pre-crRNA. The composition of reporter mRNAs is shown below. Each rectangle corresponds to one codon; rbs — a ribosome binding site; AUG — a start codon, UAA — a stop codon located in-frame with the AUG. Translation of mRNAs transcribed from unexpanded CRISPR reporters stops at the UAA codon in the leader sequence resulting in (C) chloramphenicol-sensitivity (Cm-sensitivity) or (D) absence of fluorescence. Insertion of a 61-bp (33-bp spacer/28-bp repeat) unit into the CRISPR array shifts the in-leader UAA stop codon out of the frame and restores the reporter gene reading frame resulting in chloramphenicol-resistance (C) or fluorescence (D) of cells with expanded CRISPR arrays.

Figure 1. Selection-based methods to detect CRISPR adaptation.

In all panels, unexpanded CRISPR arrays are labeled UA; expanded arrays — EA, the leader is shown as a light peach rectangle; repeats — as white rectangles; the pre-existing spacer S<sup>0</sup> — as a dark peach rectangle; the newly acquired spacer S<sup>+</sup> — as a turquoise rectangle. Primers are shown as arrows below CRISPR arrays. A flipped end of a primer arrow indicates the absence of complementarity between the primer’s end and the template. The brackets below primer pairs indicate PCR products (dashed lines correspond to low-efficiency PCR with degenerate primers whose 3′-ends are not complementary to the template). (A) Standard PCR-based spacer acquisition assay. The leader-proximal part of the CRISPR array is amplified with a leader-specific primer (light peach arrow) and a primer annealing to the S<sup>0</sup> spacer (dark peach arrow). The products of PCR are resolved on an agarose gel. M, molecular-weight size marker. The band corresponding to expanded CRISPR arrays (EA) is purified from the gel and subjected to HTS (high-throughput sequencing). (B) Selective amplification of expanded arrays by PCR with a leader-specific primer and a degenerate repeat-specific primer (white arrows with turquoise arrowheads). All nucleotides of the degenerate primer, except for the last one, are annealed to the repeat placing the 3′-end nucleotide opposite the last nucleotide of the leader or the last nucleotide of the acquired spacer. The last position of the degenerate primer includes one of three nucleotides that are not complementary to the leader’s last nucleotide enabling efficient amplification only when the last nucleotide of the acquired spacer is different from the last nucleotide of the leader. (C) Selective amplification of expanded arrays using CAPTURE. Top, CRISPR arrays are amplified and the band corresponding to expanded CRISPR arrays is purified after electrophoresis as in A. Bottom, to further increase the percentage of expanded arrays in sequencing libraries, the products of the first PCR are reamplified with primers (white arrows) annealing to repeats. The expanded arrays are selectively amplified since the products of the first-stage PCR of unexpanded and expanded arrays contain one and two repeats, correspondingly. (D) Selective amplification of expanded arrays using SENECA. An FaqI endonuclease recognition site (maroon rectangle) is introduced immediately following the first CRISPR repeat. FaqI cleaves DNA upstream of its recognition site creating sticky ends in the repeat sequence. A matching adapter is ligated to this sticky end and PCR amplification with adapter-specific (light blue) and repeat-matching (white) primers is performed.

Figure 2. PCR-based methods of studying CRISPR adaptation in bacterial cultures.

In all panels, unexpanded CRISPR arrays are labeled UA; expanded arrays — EA, the leader is shown as a light peach rectangle; repeats — as white rectangles; the pre-existing spacer S<sup>0</sup> — as a dark peach rectangle; the newly acquired spacer S<sup>+</sup> — as a turquoise rectangle. Primers are shown as arrows below CRISPR arrays. A flipped end of a primer arrow indicates the absence of complementarity between the primer’s end and the template. The brackets below primer pairs indicate PCR products (dashed lines correspond to low-efficiency PCR with degenerate primers whose 3′-ends are not complementary to the template). (A) Standard PCR-based spacer acquisition assay. The leader-proximal part of the CRISPR array is amplified with a leader-specific primer (light peach arrow) and a primer annealing to the S<sup>0</sup> spacer (dark peach arrow). The products of PCR are resolved on an agarose gel. M, molecular-weight size marker. The band corresponding to expanded CRISPR arrays (EA) is purified from the gel and subjected to HTS (high-throughput sequencing). (B) Selective amplification of expanded arrays by PCR with a leader-specific primer and a degenerate repeat-specific primer (white arrows with turquoise arrowheads). All nucleotides of the degenerate primer, except for the last one, are annealed to the repeat placing the 3′-end nucleotide opposite the last nucleotide of the leader or the last nucleotide of the acquired spacer. The last position of the degenerate primer includes one of three nucleotides that are not complementary to the leader’s last nucleotide enabling efficient amplification only when the last nucleotide of the acquired spacer is different from the last nucleotide of the leader. (C) Selective amplification of expanded arrays using CAPTURE. Top, CRISPR arrays are amplified and the band corresponding to expanded CRISPR arrays is purified after electrophoresis as in A. Bottom, to further increase the percentage of expanded arrays in sequencing libraries, the products of the first PCR are reamplified with primers (white arrows) annealing to repeats. The expanded arrays are selectively amplified since the products of the first-stage PCR of unexpanded and expanded arrays contain one and two repeats, correspondingly. (D) Selective amplification of expanded arrays using SENECA. An FaqI endonuclease recognition site (maroon rectangle) is introduced immediately following the first CRISPR repeat. FaqI cleaves DNA upstream of its recognition site creating sticky ends in the repeat sequence. A matching adapter is ligated to this sticky end and PCR amplification with adapter-specific (light blue) and repeat-matching (white) primers is performed.
unexpanded CRISPR arrays [55]. In the case of *E. coli* type I-E system, the standard method allows one to reliably detect expanded arrays amplicons only in cultures which contain, in our experience, at least 5% of adapted cells. Several modifications aimed to increase the sensitivity have been developed. The simplest one relies on amplification with a leader-specific primer and a primer matching a newly acquired spacer whose sequence is known [40]. After calibration to the amount of PCR product amplified from a region outside of CRISPR array and reflecting the total number of DNA molecules in the sample, this method can be used to determine the efficiency of adaptation by qPCR [56]. The obvious drawback of this method is that it requires prior knowledge about acquired spacer(s) and thus cannot be applied to study spacer acquisition in systems with unknown adaptation preferences. However, it is very powerful when studying acquisition from spacer-sized oligonucleotides transformed into cells [51,54,57].

Another modification uses a leader-specific primer and ‘degenerate’ primers matching the repeat sequence and containing one extra 3’-end nucleotide (Figure 2B) [42]. The additional position contains, in equal proportions, three nucleotides except for the one complementary to the last nucleotide of the leader. While amplification products are only expected if (i) a spacer has been acquired and (ii) its last nucleotide is different from the last nucleotide of the leader, in practice amplicons from unexpanded arrays are still observed [42]. The method was reported to detect as little as 0.01% of cells with expanded arrays [42]. However, by design, up to ~25% of acquired spacers remain undetected. Moreover, the method is effective only when applied to engineered miniaturized CRISPR arrays reduced to just one repeat, since multiple amplification products from unexpanded arrays with multiple repeats are produced which can’t be distinguished from amplicons from expanded arrays.

Reamplification of gel-purified amplicons of expanded arrays allows one to increase the sensitivity of detection of CRISPR adaptation in cell cultures [58]. Products of standard amplification are separated by gel-electrophoresis and purified. The reamplification step is repeated until a fragment of expected length becomes clearly visible on the gel (Figure 2C). The use of automated BluePippin system (agarose gel electrophoresis with automated elution for size selection) allows one to improve the quality of separation, reduce contamination from unexpanded arrays amplicons, and increase the reproducibility of analysis. Even when amplicons of expanded arrays are invisible after first electrophoretic separation, DNA extracted from the corresponding position of the gel can be used for reamplification. Depending on the set of primers designed for reamplification (‘internal’, ‘degenerate’, or ‘repeat-specific’), this method is reported to detect, correspondingly, 1, 0.01, and 0.1% of cells with expanded arrays within *E. coli* cultures. ‘Internal’ primers can be either the same as the ones used during initial amplification, or a leader-specific ‘nested’ reamplification primer annealing closer to the array can be used to increase specificity and avoid amplification of non-CRISPR DNA. In either case, amplification of unexpanded arrays co-purified with expanded ones is not suppressed. ‘Degenerate’ primers selectively suppress unexpanded array reamplification as described above. Reamplification with ‘repeat-specific’ primers relies on the fact that amplicons corresponding to expanded arrays have two repeats after the first PCR step (Figure 2C). Thus, amplification with primers matching the halves of the repeat sequence yields PCR product only for expanded arrays.

The SENECA [59] pipeline selectively amplifies expanded CRISPR arrays. At the heart of the method is the construction of a plasmid-borne CRISPR array with an FaqI endonuclease recognition site immediately downstream of a miniaturized CRISPR ‘array’ consisting of a single repeat preceded by the leader sequence (Figure 2D). Unlike most Type II restriction endonucleases, FaqI, a Type II-S enzyme, cleaves DNA outside of its recognition site generating a sticky end. The CRISPR array used in SENECA is designed such that the recognition of the FaqI site leads to cleavage in the upstream repeat. An Illumina adapter with a sticky end complementary to that generated by FaqI is ligated and PCR with a pair of primers, one complementary to repeat and another — to adapter, selectively amplifies expanded arrays, since the initial repeat sequence is lost after the FaqI treatment and thus amplicons from unexpanded arrays are not amplified. While the published SENECA protocol is based on the use of FaqI, other Type II-S restriction endonucleases could conceivably be used in lieu of FaqI.

The initial SENECA protocol was applied for analysis of spacers acquired by the type III CRISPR-Cas system of *Fusicatenibacter saccharivorans* heterologously expressed in *E. coli* [59]. While ~700-fold enrichment for expanded arrays in a sequencing library was reported, the sensitivity of SENECA in terms of detecting the percentage of cells with expanded arrays was not determined. To benchmark SENECA against other methods to detect CRISPR adaptation, we constructed an *E. coli* strain with a spacer containing FaqI recognition site, S_FaqI, incorporated into the genomic type I-E CRISPR array (Figure 3A and Supplementary Figure S1A). Except for S_FaqI, the strain is isogenic to KD263, a strain with *cas* genes under control of inducible promoters and a miniaturized CRISPR array with a single spacer (Supplementary Fig. S1A and Supplementary Table S1). KD263 has
been extensively used to study various aspects of CRISPR adaptation [41,49,56,60–64] and adaptation efficiencies of up to 50–80% have been reported at conditions of priming [41,62]. To compare the sensitivity of standard PCR detection and SENECA, DNA prepared from cells with unexpanded CRISPR array containing the SFaqI spacer was mixed in different proportions with DNA prepared from cells containing an additional spacer-repeat unit (Figure 3B). With the standard protocol, expanded arrays were clearly detected when 10% of DNA had an additional spacer. With SENECA, an expected amplification product was still seen when DNA with expanded array constituted 0.4% of the total, a ∼25-fold increase in sensitivity (Figure 3B). To test if SENECA introduces biases in amplification of individual spacers we analyzed spacers acquired during type I-E primed adaptation by standard protocol or SENECA (Supplementary Fig. S1). No changes in spacer selection patterns were detected between the two methods (Supplementary Fig. S1C).

Suppressing PCR
As yet another alternative, we sought to increase the efficiency of rare adaptation events detection by selective amplification of longer fragments with suppressing PCR (supPCR) that relies on amplification of fragments flanked by ∼30-nt inverted terminal repeats (ITR) [66,67]. DNA fragments containing ITR form pan-like structures due to intra-molecular hybridization (Figure 4A). Since shorter fragments are more likely to form pan-like structures, annealing of a primer identical with ITR and used at the second, reamplification stage is inhibited [68]. To determine whether supPCR can increase the sensitivity of spacer acquisition detection, we used standard leader and first-spacer specific primers extended with ITR. As can be seen from Figure 4B and Supplementary Figure S2, preferential amplification of expanded CRISPR arrays at the second PCR stage with an ITR primer increased the sensitivity of detection of adaptation events by ∼10-fold.

The ratio of DNA template to the ITR primer used during the second stage is a critical parameter for successful supPCR, as suppression relies on competition between intra-molecular hybridization of ITRs and
annealing of the ITR primer for target product amplification (Supplementary Fig. S2A) [68]. This depends on the length of a spacer-repeat unit incorporated in the extended array and has to be determined experimentally. Because the same ITR can be used for amplification of CRISPR arrays from different systems, other parameters, such as annealing temperature and the length of suppression primer remain the same for different amplification reactions. We applied supPCR to analyze spacers acquired during primed adaptation by the E. coli type I-E system and detected similar patterns of acquired spacers for the standard protocol and supPCR (Supplementary Fig. S3). In addition to the E. coli type I-E system, we have successfully used supPCR to reveal spacer acquisition in the Thermus thermophilus Type III CRISPR arrays (our unpublished observations).

Detection of spacer precursors in vivo
The methods described above detect the final result of CRISPR adaptation, spacers integrated into the CRISPR array. New spacers are derived from ‘prespacers’, short pieces of DNA that are generated inside the cell by various processes and then channeled into CRISPR arrays by Cas1 and Cas2 proteins [24,25,34,50–52,69–75]. The knowledge about prespacer structure is essential for the understanding of CRISPR adaptation preferences and the ability to control it. In vitro, the Cas1–Cas2 complex preferentially uses double-stranded DNA fragments with blunt ends or 3′-end overhangs as substrates for integration [24,70,76–78]. Shipman et al. were the

---

**Figure 4. SupPCR increases the sensitivity of detection of CRISPR adaptation.**

Schematic representation of the supPCR assay for detection of CRISPR adaptation is presented in A. Both strands of expanded (EA) and unexpanded (UA) CRISPR arrays are shown. The first-stage PCR (I) employs one primer matching the CRISPR array leader (beige arrow), and another one matching the leader-proximal pre-existing spacer S₀ (maroon arrow). Both primers carry at their 5′ ends identical 30-nt sequences (shown in lilac) that become ITRs after amplification. The second PCR stage (II) is carried using a primer whose sequence matches that of the ITR. Amplicons corresponding to expanded arrays are less likely to undergo intra-molecular hybridization, and, therefore, more effectively amplified with the ITR primer. Primers are shown as arrows below or above CRISPR arrays. The brackets below CRISPR arrays indicate regions amplified by PCR (dashed line corresponds to low-efficiency PCR on the UA template). (B) Genomic DNA purified from an E. coli strain with one additional spacer (+1) was mixed in indicated proportions with DNA from a strain with an unexpanded (+0) array. The mixtures were used as templates for supPCR. Products of ‘standard’ amplification (Figure 2A), as well as products of the first (I) and second (II) stages of supPCR separated by agarose gel electrophoresis are shown.
first to test prespacer requirements for integration in vivo [57]. They demonstrated that double-stranded synthetic oligonucleotides electroporated into E. coli cells expressing type I-E cas1 and cas2 can be trimmed to the size of a canonical E. coli spacer (33 bp) and integrated into the CRISPR array (Figure 5A) [57]. The presence of consensus PAM sequence (AAG/CTT) increased the acquisition efficiency of prespacers ~5-fold and ensured specific orientation of integration with the PAM-derived G/C immediately following the first repeat [57].

We developed a method for strand-specific HTS of short DNA fragments generated in vivo that we called FragSeq and that allowed us to detect prespacers in E. coli cells undergoing primed CRISPR adaptation (Figure 5B) [79]. The method uses protocols that avoid the loss of short fragments during purification of DNA from cell lysates (phenol–chloroform DNA purification) and size selection of spacer-sized fragments prior to
library preparation. Our experience shows that even if as little as 0.1% of genomic DNA is randomly fragmented during DNA purification and size selection, it creates background noise that complicates detection of specific signal from short DNA fragments present in the cell. To avoid this complication the FragSeq experiment should be designed in such a way that a specific signal from a particular region of the genome/plasmid from which prespacers are expected is compared with a control region non-specifically fragmented during downstream handling. In our case, we achieve this using self-targeting E. coli cells, in which a crRNA directs the interference machinery to recognize a protospacer positioned within bacterial genome. Upon induction of self-targeting cells stop dividing but remain alive for at least several hours. To ensure that fragments with different end structures are detected, single-strand specific sequencing is required to correctly determine the length of each prespacer strand, meaning that adapters need to be ligated to both single- and double-stranded ends of prespacer fragments at the library construction stage. We used a combination of thermostable 5’ App DNA/RNA ligase (NEB) which is a K97A variant of RNA ligase from Methanobacterium thermoautotrophicum [80] and T4 RNA ligase 1 (NEB). However, we observed a strong bias by the 5’ App DNA/RNA ligase to ligate adapters to DNA fragments carrying a CGN motif on their 3’ ends, which skews the libraries. It is thus possible that the less biased T4 RNA ligase 1 alone may be better suited for library construction. Alternatively, commercial kits for single-strand specific sequencing such as Accel-NGS 1S Plus DNA Library Kit (Swift Biosciences) can be used. But it should be considered that these kits include a step of addition of a low-complexity tail onto 3’ ends, which can complicate the mapping of prespacer ends.

In our work, we used FragSeq to analyze prespacers formed during primed adaptation [79]. The hallmark of primed adaptation by the type I-E CRISPR-Cas system is the acquisition of spacers with the sequence of the nontranscribed strand originating from the nontarget strand of the DNA degraded by CRISPR interference machinery [30,31]. In line with this, we detected 32–34-nt fragments (starting with G or AG on their 5’ ends) that originated from the nontarget strand upstream of the priming protospacer (PPS) due to incision within the AAG motif [79]. Longer 37–38-nt fragments originated from the target strand and had a CTTNN motif on their 3’ ends [79]. Using the approach of Shipman et al. we tested complementary pairs of the most abundant fragments and concluded that prespacers with a blunt PAM-distal end, a 33–34-bp double-stranded region and a 4–3-nt 3’-end overhang on the PAM-derived end are integrated into the CRISPR array with the highest efficiency [79]. It thus appears that the asymmetry in type I-E prespacer structure contributes to specific orientation of prespacer integration since the PAM-distal 3′ end is processed (becomes blunt) earlier than the PAM-derived end and, therefore, has higher chances to be engaged into integration at the leader-repeat boundary.

Earlier, Musharova et al. [63] performed primer extension assay on total genomic DNA purified from E. coli cells undergoing primed adaptation to test if there are any specific cleavage sites near two protospacers corresponding to frequently acquired spacers. The reaction with a primer annealed to the nontarget strand downstream of the protospacer sequence revealed two products, one corresponding to DNA cleaved within the AAG PAM and another — at the PAM-distal boundary (Figure 5C) [63]. This observation is fully consistent with the detection of 32–34-nt fragments by FragSeq. Though 37–38-nt CTT-containing fragments were revealed by FragSeq, no primer extension products were detected for the target strand. The possible explanation for this discrepancy could be that while at least the PAM-distal 3′ end of a prespacer is produced via endonucleolytic cleavage of the nontarget strand, the PAM-derived 3′ end is produced due to exonucleolytic cleavage of the target strand upstream of the protospacer resulting in loss of the sequence complementary to the primer used in Musharova et al. (Figure 5C).

**Perspectives**

- **Importance of the field.** CRISPR adaptation is a complex multistep process that remains the least understood stage of adaptive prokaryotic immunity. While most practical applications of CRISPR are built around the interference stage, understanding of adaptation, in addition to providing fundamental insights into the process, will allow one to efficiently create cells with specific resistance profiles and record time-resolved ‘memories’ of specific events. These developments will not be possible without highly sensitive methods to reveal rare spacer acquisition events.
• A summary of the current thinking. Despite considerable effort, in the absence of selection, the best available methods can detect adaptation when cells with expanded arrays constitute at least 0.01% of the population. These levels are likely several orders of magnitude above those observed in natural settings, where high levels of CRISPR adaptation, at least in the absence of viral infection, should be highly unfavorable because of subsequent deleterious autoimmunity. Indeed, most of the currently used methods are used to study artificial systems where adaptation is activated by overexpression of cas genes responsible.

• Future directions. Since some of the available methods, such as SENECA, the use of degenerate primers, and supPCR are orthogonal to each other, combining them may help to increase sensitivity somewhat. In the absence of further developments, very deep sequencing will probably become the method of choice to study adaptation, especially in organisms for which no methods of genetic manipulation are available. Development and application of methods like FragSeq to study short nucleic acids inside the cells should allow complementary studies of spacers. Of most interest will be comparative analyses of spacers generated by different CRISPR-Cas systems types and establishment of relationship between spacer generation and DNA maintenance processes in the cell. Selection of spacers is virtually unstudied. New sensitive and high-throughput methods will have to be developed to uncover the details of this earliest stage of generation of adaptive immunity.

Abbreviations
Ab, antibiotic; BIMs, bacteriophage insensitive mutants; CAPTURE, CRISPR adaptation PCR technique using reamplification and electrophoresis; Cm, chloramphenicol; CRISPR, clustered regularly interspaced short palindromic repeats; EA, expanded CRISPR arrays; FACS, fluorescence-activated cell sorting; HTS, high-throughput sequencing; ITR, inverted terminal repeats; MGEs, mobile genetic elements; PAM, protospacer adjacent motif; PIMs, plasmid interfering mutants; PPS, priming protospacer; PS, protospacer; rbs, ribosome binding site; S, spacer; SENECA, selective amplification of expanded CRISPR arrays; UA, unexpanded CRISPR arrays.

Author Contributions
K.S. conceived the article. A.S., K.S., I.F., and D.V wrote the manuscript. A.S., I.F., and D.V prepared the figures. D.V. performed SENECA experiments and data analysis. I.F. performed supPCR experiments and data analysis.

Funding
This work was supported by grant 075-15-2019-1661 from the Ministry of Science and Higher Education of the Russian Federation and NIH RO1 grant GM104071 to KS.

Acknowledgements
We dedicate this paper to the memory of Dr. Ekaterina Savitskaya who played a central role in CRISPR research in our laboratory.

Competing Interests
The authors declare that there are no competing interests associated with the manuscript.

References


© 2020 The Author(s). This is an open access article published by Portland Press Limited on behalf of the Biochemical Society and distributed under the Creative Commons Attribution License 4.0 (CC BY-NC-ND).