CRISPR adaptation in *Escherichia coli* subtype I-E system

Ruth Kiro*, Moran G. Goren*, Ido Yosef* and Udi Qimron*1

*Department of Clinical Microbiology and Immunology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

**Abstract**

The CRISPRs (clustered regularly interspaced short palindromic repeats) and their associated Cas (CRISPR-associated) proteins protect prokaryotes against horizontally transferred DNA [1-3] and RNA [4]. The CRISPR–Cas system is analogous to the mammalian immune system [5,6], and to eukaryotic RNAi mechanisms [7,8]. Three major types, and ten subtypes, of CRISPR–Cas systems [9] have been classified. These systems are found across ~90% of archaeal genomes and ~50% of bacterial genomes. All types encode a CRISPR array: short repeated sequences called ‘ repeats’ flanking similarly sized DNA sequences called ‘spacers’. The array is usually preceded by a ‘leader’, an AT-rich DNA sequence that drives expression of the CRISPR array and is required for acquiring new spacers into the array [10,11]. The adjacent *cas* genes encode proteins that process the transcript, interfere with foreign nucleic acids and acquire new spacers [12-14]. DNA transcribed from the CRISPR array is processed by Cas proteins into RNA-based spacers flanked by partial repeats [crRNA (CRISPR RNA)]. These crRNAs specifically direct Cas-interfering proteins to target nucleic acids matching the spacers. The short DNA sequences from which the spacers are derived, and are consequently being targeted, are termed ‘protospacers’. The process of protospacer cleavage and insertion into the CRISPR array is called ‘adaptation’. CRISPR adaptation thus enables the system to target new DNA molecules.

In the present short review, we focus on the adaptation process particularly in the *Escherichia coli* subtype I-E CRISPR system, highlighting the most recent findings. For a thorough review on this process, the reader is referred to [15]. Barrangou et al. [1] were the first to experimentally demonstrate CRISPR adaptation. They showed that the CRISPR–Cas system in the Gram-positive bacterium *Streptococcus thermophiles* inserted between one and four new spacers into their CRISPR array following challenges by different phages [1]. The newly acquired spacers were inserted in the leader-proximal end and corresponded to protospacers on either the coding or the template strands of the phage chromosome. Mojica and colleagues analysed numerous CRISPR arrays from sequenced genomes of different *E. coli* strains [16]. They showed that spacers were derived from phage genomic sequences, mobile elements, plasmids, the host chromosome and from unknown sources [16]. Robust experimental systems for thoroughly studying the adaptation system in *E. coli* were only recently established. Two systems identified new spacer adaptation by selecting for a phenotype (plasmid loss or phage resistance) derived from adaptation of an active spacer [17,18]. Two other systems detected adaptation without selection for functionality of the acquired spacers [11,19]. Each of these systems has its advantages and disadvantages as elaborated below.

Swarts et al. [18] used an *E. coli* mutant with an active CRISPR–Cas system due to deletion of the *hns* gene, whose product represses transcription of the *casABCDE* operon and the crRNA [10,21]. They propagated this mutant, harbouring a high-copy plasmid for 1–2 weeks, and then screened for bacteria having lost the plasmid due to CRISPR/Cas activity. This procedure yielded a majority of bacteria cured of the plasmid, all of them carrying at least one functional spacer against the plasmid. This system thus enabled information on newly acquired spacers to be obtained and demonstrated that the *E. coli* adaptation machinery is functional upon derepression. Analysis of the acquired spacers showed that the spacers inserted into the repeats in the leader-proximal end. The majority of the protospacers initiated with a 5′-G base, and had a 5′-AA motif upstream. This suggested that the adaptation machinery selects protospacers having such a motif. However, since the experimental system relies on...
the functionality of the spacers to eliminate the plasmid, this evidence is only circumstantial, as one could argue that the motif is enriched due to its functionality rather than due to selection by the adaptation machinery. The fact that the plasmid carried the lacI gene, having a similar copy in the host chromosome, allowed monitoring of whether self-targeting spacers would be excluded from the array. Indeed, the authors observed a significant underrepresentation of spacers from the lacI gene on the plasmid. This suggested that spacers from genomic origin are counter-selected. However, in this case too, it is hard to assess whether this exclusion occurs during the adaptation step, or as a result of selection against bacteria incorporating the spacers against the chromosome. An intriguing phenomenon was observed in arrays having more than one newly acquired spacer. It was shown that in all of these cases, all spacers were derived from the same strand. The authors propose that the interference machinery produces substrates for the adaptation machinery from the targeted strand and thus, in most cases, multiple adaptation events are derived from the same strand.

The above observation, in which a pre-existing spacer positively directs adaptation from the same strand on which it resides was termed ‘priming’ and was shown directly in a study by Datsenko et al. [17]. This study used an experimental system to monitor adaptation events among bacterial colonies expressing different combinations of Cas proteins infected or uninfected with M13 phage. The study provided direct evidence that either a functional or non-functional pre-existing spacer can prime the adaptation machinery to selectively acquire additional spacers from the same DNA strand on which the pre-existing spacer resides. As opposed to the hypothesis of Swarts et al. [18], suggesting that digestion products of the targeted DNA are used for the adaptation, Datsenko et al. [17] proposed that the adaptation machinery slides on the DNA through a Cas3–Cascade (CRISPR-associated complex for antiviral defence)-guiding mechanism and therefore the specific DNA strand to which the complex is loaded on is primed. Nevertheless, the same group analysed ~200,000 spacers obtained by high-throughput sequencing and showed that the ‘sliding’ mechanism probably does not occur during spacer adaptation [23]. The authors hypothesized that, if sliding occurs, then spacers closest to the priming spacer should be acquired most often. However, this was not the case. Clearly, further experiments should be carried out to elucidate the mechanism of the priming phenomenon.

In another experimental system to study adaptation, Yosef et al. [11] used a PCR-based assay to detect adaptation and characterize new spacers. The PCR was carried out on DNA from a bacterial culture overexpressing Cas1 and Cas2, and provided a simple and robust tool to assess whether adaptation takes place under different conditions. Using this assay, Yosef et al. [11] showed that Cas1 and Cas2 are the only Cas proteins that are essential for the adaptation process, and, particularly, that the DNase activity of Cas1 is essential for this process. Moreover, they showed that a small (60 bp) fragment of the leader sequence adjacent to the array is essential for adaptation. In addition, they demonstrated that the first repeat is duplicated upon adaptation, and that a single repeat is necessary and sufficient for the adaptation process. Lastly, analysis of the spacers obtained showed that the majority of the spacers initiated with a 5′-G, and had a 5′-AA motif upstream to the protospacer. Because, in this experimental system, selection for functional spacer was not applied, these results demonstrated that this 5′-AA motif is selected in the adaptation step, and then plays a significant role also in the interference step.

On the basis of the above results, van der Oost and colleagues proposed a model for adaptation that fits well with the above observations [23]. In this model, a double-stranded protospacer DNA is cleaved probably by the adaptation machinery, and, concomitantly, two nicks are generated in the first base of the repeat and in its 28th base, in the opposite strand (Figure 1). The protospacer ends are then joined to each nick, and gap filling, followed by ligation, completes the process. This model thus explains why the 29th base is considered part of the protospacer rather than of the repeat [17,18,24]. It also explains why the first repeat is identical with the new repeat.

The proposed model, however, does not address whether additional motifs, beside the PAM (protospacer-associated motif), affect the efficiency of spacer adaptation. Existence of such motifs was postulated following an extensive analysis of spacers from E. coli that showed clear preference for adaptation of certain protospacers over others, all encoding a PAM [22]. Our laboratory recently established a PCR-based assay that measures the efficiency of spacer adaptation from distinct protospacers. In a thorough analysis of ~200000 spacers, we found that the terminal dinucleotide, 5′-AA, affects the efficiency of adaptation of certain spacers [20]. This dinucleotide motif, which we termed AAM (‘adaptation-affecting motif’) was significantly overrepresented in spacers acquired frequently compared with those acquired rarely. Interestingly, another group analysed spacers acquired from S. thermophilus, and found a bias in spacer sampling [25]. This bias could be explained by the presence of novel motifs that are yet to be identified in S. thermophilus. As a whole, these findings shed more light on the mechanism of adaptation.

A unique and rather ingenious system for detecting adaptation events was recently developed by Mojica and colleagues [19]. This system does not require functional spacers or interference proteins and can detect adaptation events even under low expression of Cas1 and Cas2. The assay selects for antibiotic resistance that is gained regardless of the spacer’s functionality. A recombinant plasmid was uniquely designed to confer antibiotic resistance upon adaptation of a new spacer into a CRISPR array. This enabled the selection of clones having acquired new spacers. Using this system, the authors showed that variations of the PAMs occur in different E. coli strains harbouring the subtype I-E CRISPR-Cas systems. Interestingly, the study showed that spacers are adapted from plasmids encoding the cas genes more frequently than from other plasmids residing in the bacterial cell. The authors speculate that the proximity of the newly
transcribed Cas1 and Cas2 to the plasmid results in this biased adaptation, but this has to be studied further. The authors also demonstrated that different combinations of Cas1 and Cas2 proteins of certain E. coli strains with leader sequences of other E. coli strains results in spacer adaptation having the PAM in the reverse orientation. This indicates that both the Cas proteins and the leader sequence are involved in determining the orientation of adaptation.

A fundamental question for any immune system is how it avoids targeting self components. Stern et al. [26] showed that, in many cases in which a self spacer was acquired, the entire system became inactive. Our results show that, indeed, chromosome-derived spacers are acquired at least 200-fold less frequently than plasmid-derived spacers, in the absence of interference, suggesting that an active mechanism prevents adaptation from the chromosome [11]. It is thus possible that the CRISPR–Cas system reduces targeting its own chromosome by avoiding the adaptation of self spacers. This discrimination against self sampling of spacers is probably not an artefact resulting from self killing of ‘self samplers’ due to damage to the chromosome resulting from dsDNA breaks by Cas1 and Cas2, as no severe toxicity that would account for such a 200-fold discrimination is detected. The discrimination against adaptation from genomic DNA

Figure 1 | Hypothetical adaptation model for subtype I-E CRISPR/Cas system of E. coli
(A) I. Spacer selection. Both Cas1 and Cas2 are essential for cleaving a spacer at the sites indicated by the arrowheads. Additional uncharacterized proteins marked with ? may also play a role in the adaptation process. As explained in the text, the PAM, found upstream of the spacer, and the AAM, found in the last two nucleotides of the spacer enhance the efficiency of spacer adaptation. II. Spacer insertion into the array. A segment of 60 bp of the leader, not encoding the TATA box found upstream this sequence, is sufficient for adaptation. The CRISPR array is nicked in the first base of the repeat and in its 28th base, in the opposite strand, as indicated by the arrowheads. The pink base pair demonstrates that usage of a functional variant of a repeat will duplicate this same variant. III. Final adaptation step. The protospacer ends are ligated to each nick, and gap filling, followed by ligation, completes the process. (B) Schematic representation of the priming process. The first acquired protospacer increases the probability of additional adaptation of subsequent protospacers from the same stand. The crRNA, Cas1, Cas2, Cascade and Cas3 are all essential for this process. Small arrows represent acquired spacers.
perhaps depends on the unique methylation pattern of the chromosome or other post-replication changes, unique structural organization or unique protective DNA sequences present in the chromosome. The mechanism by which this selective adaptation takes place is yet obscure, but it is definitely a cardinal question that awaits resolution in elucidating the adaptation process.

Another cardinal mechanism that is yet elusive is the mechanism by which the size of the acquired spacer is determined. Most spacers are 33 bp long. Approximately 3% of the spacers are 34 bp long, and less than 2% are longer than that [11,16]. The fact that ~95% of the acquired spacers are 33 bp long suggests that there is a mechanism that determines this specific size. The length of the resident spacer does not dictate the length of the incoming spacer, as adaptation of spacers with regular lengths is also observed when a resident spacer is lacking (e.g. in a single-repeat array); moreover, adaptation of a spacer into an array with a resident spacer longer than 33 bp mostly results in a 33 bp spacer [11]. One possible mechanism for determining the size of the spacer is the presence of two DNA motifs at a fixed interval that are recognized by the adaptation machinery. Another possible mechanism is that the recognition of the PAM binds the adaptation machinery to that site and, using another site at a distance of 33 bp. These, and other, important questions are only few of those exciting scientific enigmas that await deciphering in this fascinating field of CRISPR–Cas immunity, and, in particular, in CRISPR adaptation.

Funding

This work was supported by a Marie Curie International Reintegration Grant [grant number PIRG-GA-2009-256340] and by the Leo and Elsa Abrahamson Fund of the Sackler School of Medicine at Tel Aviv University.

References


Received 24 June 2013 doi:10.1042/BST20130109