Determinants of spacer choice during naïve and primed CRISPR adaptation by type I-E CRISPR-Cas system of *Escherichia coli*

Ekaterina Savitskaya

Skolkovo Institute of Science and Technology, Moscow, Russia, savitskaya.e@yandex.ru

Sofia Medvedeva

Skolkovo Institute of Science and Technology, Moscow, Russia, sofia.medvedeva@skolkovotech.ru

Olga Musharova

Skolkovo Institute of Science and Technology, Moscow, Russia, olga.musharova@skolkovotech.ru

Konstantin Kuznedelov

Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, NJ USA, kuznedel@waksman.rutgers.edu

Ekaterina Semenova

Waksman Institute, Rutgers, The State University of New Jersey, Piscataway, NJ USA, semenova@waksman.rutgers.edu

Konstantin Severinov

Skolkovo Institute of Science and Technology, Moscow, Russia, k.severinov@skoltech.ru

Prokaryotic CRISPR-Cas immunity systems consist of CRISPR DNA arrays containing variable spacers separated by identical repeats and CRISPR associated (*cas*) genes. 2 classes comprising up to 6 types of CRISPR-Cas systems have been described in diverse eubacteria and archaea (1, 2). Despite this diversity, all CRISPR-Cas systems share common principles of function. Protective short CRISPR RNAs (crRNAs) are produced upon CRISPR array transcription and processing and contain a single spacer flanked by fragments of repeats. A crRNA enters into a complex with *cas* genes products. Resulting ribonucleoprotein effector complex recognizes protospacer, a DNA sequence matching crRNA spacer, and subsequently degrades protospacer-containing DNA either directly or indirectly, by recruiting a Cas executor nuclease, in a process called “CRISPR interference”. For most CRISPR-Cas systems protospacer adjacent motive (PAM) is necessary for CRISPR interference to occur (3). PAMs allow CRISPR effectors to distinguish a non-self protospacer in foreign DNA from self spacer in CRISPR array. New spacers are acquired in CRISPR arrays in a process named “CRISPR adaptation”. While CRISPR interfering complexes are evolutionarily unrelated between CRISPR-Cas systems from different classes and types, the Cas1 and Cas2 proteins required for CRISPR adaptation are conserved in all CRISPR-Cas systems (1).

New spacers are acquired into CRISPR arrays in (at least) three stages: 1) generation of DNA substrates recognized by Cas1-Cas2 proteins; 2) processing of intermediates into fragments capable for integration, and 3) incorporation of such intermediates into array. Protospacers

proceeded by PAM should be selected during adaptation to allow subsequent CRISPR interference.

For type I CRISPR-Cas systems, it was shown that the Cas1 and Cas2 proteins form a stable complex (4). In vitro and in vivo studies showed that this complex is capable for performing the last stage of CRISPR adaptation via a mechanism resembling transposon insertion. In contrast, little is known about the fist two stages of the adaptation process. The situation is complicated by existence of two modes of adaptation: naïve and primed. The less efficient naïve CRISPR adaptation occurs in the absence of CRISPR interference machinery. It was suggested that the Cas1-Cas2 adaptation complex uses byproducts of RecBCD recombination as a source of spacers (5). Much more efficient primed adaptation is guided by CRISPR interference and most spacers come from targeted DNA (6). Thus, one may assume that target DNA degradation products are the substrates for the adaptation complex in the course of primed adaptation. Interestingly, only protospacers containing consensus AAG PAM are selected by the type I-E *E. coli* CRISPR-Cas system during primed adaptation, while this specificity is relaxed during naïve adaptation. Other factors determining adaptation complex specificity were suggested besides PAM, as it was shown that some protospacers are reproducibly selected as spacers more often than others during primed adaptation (7), while spacers with 3'-terminal AA motif are overrepresented in spacer sets obtained during naïve adaptation (8). Here, we studied sequence and structural requirements for spacer selection during CRISPR adaptation by the type I-E CRISPR-Cas system from *E. coli*.

To determine DNA fragment structure that allows CRISPR adaptation, transformation of *E. coli* cells overexpressing *cas1-cas2* genes with different synthetic DNA oligonucleotides was used. Upon transformation and outgrowth, newly acquired spacers were PCR amplified and subjected to high throughput sequencing and analysis. In this way we confirmed that spacer-length (33 bp) double-stranded oligonucleotides give rise to new spacers as was previously shown by the Church group (9). We also observed that longer double-stranded DNA fragments could be used as a source of new spacers. Characterization of other possible substrates such as single-stranded or partially matched DNA fragments was performed. Such analysis allowed to identify preferred substrates of the Cas1-Cas2 adaptation complex generated in the presence or in the absence of CRISPR interference.

To find out sequence determinants of spacer selection, we analyzed new spacers acquired in cells undergoing primed adaptation in the presence of different targets (plasmids and phages) or during naïve adaptation. PCR amplified extended CRISPR arrays containing newly acquired spacers were sequenced using Illumina platform followed by new spacer sequence analysis. We

were unable to reveal features that could distinguish “hot” (very common) from “cold” (rare) spacers, however, clear underrepresentation of internal AAG trinucleotide was observed for “hot” spacers in datasets from primed adaptation experiments suggesting that the presence of internal AAG PAM sequences destroys precursors of some spacers before their incorporation into CRISPR array. To get more insight into sequence requirements for spacer acquisition we performed two series of experiments. First, primed adaptation was induced in the presence of target plasmid DNA with “hot” protospacer carrying blocks of randomized nucleotides. Second, cells expressing *cas1-cas2* genes and capable of naïve adaptation were transformed with double-stranded fragments corresponding to “hot” spacer carrying blocks of randomized nucleotides. Results of comparative bioinformatics analysis of “hot” spacer variants selected for acquisition during primed or naïve adaptation will be presented.

Overall, our results shed light on the mechanisms of CRISPR adaptation and allow partial reconstruction of protospacer-containing DNA fragment processing on its way to CRISPR array.

1. K.S.Makarova, Y.I. Wolf, O.S.Alkhnbashi, F.Costa, S.A.Shah, S.J.Saunders, R. Barrangou, S.J.J.Brouns, E.Charpentier, D.H. Haft, et al. (2015). An updated evolutionary classification of CRISPR–Cas systems, *Nat. Rev. Microbiol.* **13**: 722-736.

2. S. Shmakov, O.O. Abudayyeh, K.S. Makarova, Y.I.Wolf, J.S. Gootenberg, E. Semenova, L. Minakhin, J. Joung, S. Konermann, K. Severinov, et al. (2015). Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. *Mol. Cell* **60**: 385–397.

3. F.J.Mojica, C. Diez-Villansenor, J. Garcia-Martinez, and C.Almendros (2009). Short motif sequences determin the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.

4. J.K. Nuñez, L.B. Harrington, P.J. Kranzusch, A.N. Engelman, and J.A. Doudna (2015). Foreign DNA capture during CRISPR–Cas adaptive immunity. *Nature* **527**: 535-538.

5. A.Levy, M.G. Goren, I. Yosef, O. Auster, M. Manor, G. Amitai, R. Edgar, U. Qimron, and R.Sorek (2015). CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **520**: 505–510.

6. K.A.Datsenko, K.Pougach, A. Tikhonov, B.L. Wanner, K. Severinov, and E.Semenova (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat. Commun.* **3:** 945.

7. O.Musharova, E.I.Klimuk, K.A.Datsenko, A.Z.Metlitskaya, M.Logacheva, E.Semenova,

K.Severinov, and E. Savitskaya (2017) Spacer-length DNA intermediates are associated with Cas1 in cells undergoing primed CRISPR adaptation. *Nucleic Acids Res.* In print.

8. I.Yosef, D.Shitrit, M.G.Goren, D. Burstein, T.Pupko, and U.Qimron (2013). DNA motifs determining the efficiency of adaptation into the Escherichia coli CRISPR array. *Proc. Natl. Acad. Scie. USA* **110**: 14396–14401.

9. S.L .Shipman, J. Nivala, J.D. Macklis,and G.M.Church (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**:aaf1175.