Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Elena Kurilovich

PhD Program: Life Sciences

Title of Thesis: The role of genome maintenance proteins in primed CRISPR adaptation by the type I-E CRISPR-Cas system

Supervisor: Prof. Konstantin Severinov

The thesis document includes the following changes in answer to the external review process.

I am grateful to all Jury members for the valuable comments and suggestions, which allowed to significantly improve the quality of the Thesis and raised several important questions.

Professor Francisco J. Martinez Mojica

1. *Define tracrRNA as "trans-activating CRISPR RNA" (list of abbreviations and Page 19).*

- The definition was corrected accordingly.

2. *Add "Cas" to the list of abbreviations and define it as "CRISPR-associated sequence" (see also Page 15).*

- Cas abbreviation was added to the list.

3. *Note that letters size of some figure quotations in the text are different from the rest.*

- Corrected.

4. *Page 12. Check the sentence "Major groups the bacterial and archaeal defense systems that may be divided".*

- The text was changed to: "Major groups the bacterial and archaeal defense systems may be divided into".

5. *Review the sentence "each of them containing from 1 to 100 CRISPR repeats" considering that CRISPR arrays with several hundred repeats have been reported.*

- The text was changed to: "each of them containing from 1 to several hundred CRISPR repeats".

6. *Be consistent through the text when referring to CRISPR-Cas/CRISPR systems (see for example first line on page 16).*

- Corrected.

7. *Page 17. Check the word "biundergoing".*

- The word was corrected to "undergoing".

8. *Page 27. Check the sentence "A similar construct with 5' overhangs are not bound".*

- The text was changed to: "Similar constructs with 5' overhangs are not integrated".

9. *Page 28. Replace "(3'-CTT-5')" with either "(5'-CTT-3')" or "(3'-TTC-5')".*

- "(3'-CTT-5')" was changed to "(3'-TTC-5')"

10. *Pages 28, 95. It is correctly stated that spacer integration starts "with the Cas1-catalyzed nucleophilic attack by the 3'-OH of the prespacer at the phosphodiester bond between the leader and the first repeat in the CRISPR array, leading to formation of a half-site product (Figure 3) [149]". However, Figure 3 shows, and it is claimed in reference 149, that the leader-repeat junction is the second, rather than the first, target site for integration. An appropriate figure/reference (i.e. doi:10.1126/science.aao0679; doi:10.7554/eLife.08716) should be used/cited instead.*

- The figure and the reference were changed.

11. *Page 31. In agreement with the general thought, it is stated that "RecBCD binds blunt or nearly blunt dsDNA ends [160] and starts unwinding and degrading both strands". However, it has been proposed (doi: 10.1128/mmbr.05026-11; doi: 10.1016/bs.adgen.2022.06.001) that RecBCD does not degrade DNA but translocate on the DNA from the DSB and when reaches a Chi site, nicks it. This possibility might be discussed by the candidate.*

- The discussion was added to the text:

"*In vitro*, RecBCD possesses a Mg^{2+}- and ATP-dependent nuclease activity. High concentration of free Mg^{2+} ions is required for the RecBCD nuclease activity (Eggleston & Kowalczykowski, 1993), while ATP chelates Mg^{2+} (J. E. Wilson & Chin, 1991). In conditions of low Mg^{2+}/ATP concentration ratio, RecBCD unwinds both DNA strands without degrading them (Eichler & Lehman, 1977) and makes a site-specific nick at the Chi-containing strand, 4-6 nt to the 3' side from Chi (Ponticelli et al., 1985; Taylor et al., 1985). When Mg^{2+} is in excess, RecBCD nuclease activity is stimulated, leading to degradation of the upper, Chi-containing strand into up to Chi, while the lower strand is cut into longer fragments (Dixon & Kowalczykowski, 1995; Taylor & Smith, 1995b). The products of the Chi-containing strand degradation in these conditions were shown to vary from tens to hundreds nt, while the lower strand is cut into fragments of a kilobase length (Dixon & Kowalczykowski, 1995). As it is hard to estimate the real intracellular Mg^{2+} and ATP concentrations, whether RecBCD degrades both strands before Chi or only unwinds them, is not yet determined."

12. *Page 36. Replace "Streptococcus thermophil**e**s" with "Streptococcus thermophil**u**s" and "P. aeruginosa **PA14**" with "P. aeruginosa **PA14**"*

- Corrected.

13. *Page 38. Replace "1.7.2. Clinical relevance" with "1.7.2. Clinical relevance of P. aeruginosa"*

- Corrected.

14. *Section 4.1. This section, along with related information in other chapters (i.e., Chapter 1), could be removed, as this study is not particularly connected to the main objectives of the thesis. Alternatively, I suggest to either add related materials and methods to Chapter 3 or provide further details in section 4.1 to ensure sufficient understanding of the experiments performed.*

- The related materials and methods were added to Chapter 3.

15. *Fig. 15. Replace "Δrec ΔBrecJ" with "ΔrecB ΔrecJ"*

- Corrected.

16. *Section 4.2.1. The heading of this section, "Primed CRISPR adaptation is impaired in ΔrecJ, ΔrecB ΔrecJ and ΔrecB ΔsbcD mutants", does not exactly correspond to its content. Even though figures in this section show data with additional single (recJ, sbcD, sbcB) and double mutants, only results with single recB, recC and recD mutants are outlined in the main text. The results obtained with the other mutants should be mentioned and discussed as well.*

- The missed description of the results was added.

17. *Page 65. "induction of self-targeting in the mutants caused SOS response, as judged by cessation of culture growth (Figure 15)…" Indeed, SOS response could be, and probably is, at least in part responsible for this phenotype. However, without additional experimental prove (use of appropriate controls, such as inactive cas3 mutants, or demonstration of SOS induction with SOS response reporters), the contribution of protein overexpression cannot be dismissed and should be considered as massive protein expression after transcription induction might account for the growth delay, CFU decrease and cell elongation (shown in Fig 17, but not mentioned in the text) observed, independently of interference activity.*

- CRISPR interference and subsequent SOS response are considered to be the main reason for the observed phenotype, as extensive DNA degradation in the priming protospacer (PPS) region was observed in the induced KD403 culture (Shiriaeva et al. 2019) and its mutant derivatives (unpublished). It is worth noting that in this experimental system *cas* genes are expressed not from the plasmid, but from the inducible promotors inside the genome, which means the expression is not so massive in this case. Previously it was demonstrated (Semenova et al. 2016) that induction of *cas* gene expression in KD263 strain, which has the same inducible CRISPR-Cas system but lacks a self-targeting spacer, does not cause growth delay.

18. *"the addition of cas genes inducers as judged by differential staining of live and dead cells (Figure 17)". Refer to Figure 18B as well.*

- The suggested reference was added.

19. *Fig. 23. Define "WT (-Ind)"*

- The definition was added to the text.

20. *Fig. 24. Replace "Fragments mapping to the NT-strand upstream of the PPS and to the T-strand downstream of the PPS are shown in green" with "Fragments mapping to the NT-strand upstream of the PPS and to the T-strand downstream of the PPS are shown in gray".*

- Corrected.

21. *Page 80. Add ")" to the end of the sentence "... presumably generated by random DNA fragmentation"*

- Corrected.

22. *Page 95. Replace "Than the second" with "Then, the second".*

- Corrected.

23. *A schematic figure illustrating the HIPs detection method would be appreciated.*

- A scheme illustrating the method was added.

24. *Page 97. Replace "RecJ of ExoVII" with "RecJ or ExoVII"*

- Corrected.

25. *Page 100. Clarify if the sentence "No preferences in selection of new spacers with specific orientation of the PAM sequence have been reported so far" refers to just naïve adaptation.*

- The clarification was added.

26. *Page 101. Review the sentence "... for plasmids become unstable in Rec mutants"*

- Corrected.

27. *Discussion. Your model of primed adaptation seems to apply to the acquisition of spacers from both the upstream and the downstream regions of the PPS. If so, do you assume that after PPS binding the PAC translocates either in one direction or the other from the PPS? Alternatively, do you think it is possible that upstream prespacers are generated by the PAC and downstream by other proteins/complexes?*

- Currently, the molecular mechanism that could explain the observed bidirectional CRISPR adaptation and CRISPR-Cas-induced DNA degradation, is not established. The inverted gradient pattern of acquired spacers downstream of PPS with a characteristic strand bias, and strong AAG PAM preferences both downstream and upstream of PPS suggest that PAC, not other cellular proteins, is responsible for generating prespacers on both sides. One could suggest that after the initial recognition and cleavage of the NT-strand by Cas3, another Cas3 molecule could be recruited and start translocation on the opposite strand. This hypothesis was supported by the *in vitro* experiments using single-molecule imaging (Redding et al. 2015), where PAC could translocate in either direction during priming.

28. *Bibliography. A few references are incomplete or contain errors, such as:*

- *The month of publication is missing (e.g., pp. 466–472, 2005; 467–469, 1992)*
- *Additional numbers appear before the volume (e.g., Science (80-.); Nat. Struct. Mol. Biol. 2004 119; Nat. Struct. Mol. Biol. 2011 186)*

- *Author's initials are duplicated (e.g., D. C. D. C. Swarts; S. J. J. S. J. J. Brouns; S. J. J. J. Brouns; P. M. P. M. Nussenzweig)*
- *Not abbreviated (e.g., Nature Reviews Microbiology; Methods in Molecular Biology) or incomplete (e.g., Proc. Natl. Acad. Sci.,) journal names*
- *Doi is included (e.g., http://dx.doi.org/10.4161/rna.8.3.15190)*

- The Bibliography was corrected.

Professor Edward Bolt

1. Page 16, final paragraph – It is not clear from the way this paragraph is written if the candidate comprehends/understands the molecular mechanism(s) driving evolution of primed adaptation apparently developed by bacteria as they 'learned to fight' escape mutations. There seems to be detail missing. This section would benefit from a slightly longer and more detailed description about what the candidate is thinking.

- This section was extended:

"Genetic elements may acquire point escape mutations in PAM or protospacer sequences to avoid recognition by bacterial CRISPR-Cas systems (Semenova et al., 2011). In turn, bacteria "learned" to fight this through primed CRISPR adaptation, or priming (Datsenko et al., 2012). This type of CRISPR adaptation is coupled with CRISPR interference. Priming was demonstrated for the Type I and Type II CRISPR-Cas systems (Datsenko et al., 2012; Nussenzweig et al., 2019; Swarts et al., 2012). CRISPR adaptation machinery is attracted to target-bound effector-crRNA complex, leading to a robust CRISPR adaptation process, incorporating new spacers from DNA located *in cis* to the targeted protospacer (Datsenko et al., 2012; Savitskaya et al., 2013; Swarts et al., 2012). The primed CRISPR adaptation yield is strongly increased by mutations in PAM or mismatches of the protospacer sequence with crRNA (Datsenko et al., 2012; Fineran et al., 2014). Priming mechanism allows the bacteria to update the memory record, increase the immune diversity in the population, and prevent the propagation of mutated genetic elements, which could otherwise overcome CRISPR immunity. CRISPR diversity appears important at both individual and population levels. The existence of several different spacers against the invader in one cell makes it harder for the foreign DNA element to avoid CRISPR immunity, as it would need to acquire more escape mutations (Levin et al., 2013). In population, escape mutants that managed to avoid CRISPR-Cas system in one cell, may not survive in other cells possessing different spacers, due to the limited ability to acquire escape mutations (van Houte et al., 2016)."

2. Naturally occurring CRISPRi-like functions for some CRISPR Types (ie for natural gene regulation) are absent from the Introduction. They should be included and linked in narrative to the development and use of biotechnology CRISPRi.

- The following text was added to the Chapter 1.2.:

"The most popular gene editing system utilizes Cas9 of *Streptococcus pyogenes* (SpyCas9) and a fused single guide RNAs (sgRNAs), instead of a complex of individual crRNA:tracrRNA (Cong et al., 2013; Jinek et al., 2012). To control gene expression without editing, catalytically inactivated dCas9 (**d**ead Cas9) variant is used. The system is called CRISPRi (CRISPR interference) (Qi et al., 2013), though, in contrast to RNAi which prevents translation, it mainly affects gene transcription. dCas9 can be used not only for gene repression, but also for gene activation if fused to the transcription activation domains (C. Dong et al., 2018; Gasiunas et al., 2012). Programmable RNA-targeting can be performed by Cas13. Given its smallest size among all known CRISPR effectors,

Cas13 is a promising tool for the RNA interference applications in medicine, biotechnology, and research (Jinek et al., 2012). However, the "collateral" activity of Cas13 makes specific RNA targeting problematic, though many successful application cases are already described. Besides Cas13, Cas9 variants able to target RNA are also known (Strutt et al., 2018).
Interestingly, several Type II CRISPR-Cas systems were found to naturally perform gene regulation functions important for bacterial virulence. In *Francisella novicida* strain U112 possessing Type II-B CRISPR-Cas system, the expression of bacterial lipoprotein (BLP) is downregulated by the complex of Cas9, tracrRNA, and a small CRISPR–Cas-associated RNA (scaRNA), which is transcribed from the locus near the CRISPR array (Sampson et al., 2013). At the same time, CRISPR array itself and other Cas proteins are not required for this function. BLP was shown to play a crucial role in bacterial virulence, as it is recognized by the innate immune system of the host. Thus, the repression of BLP expression by CRISPR-Cas system promotes bacterial infection, though the details of the mechanism remain unidentified. Similarly, *blp* expression was shown to be upregulated in Δ*cas9* mutants of *Streptococcus agalactiae* and *Riemerella anatipestifer*, possessing Type II-A and II-C CRISPR-Cas systems, respectively (Ma et al., 2018; Y. Wang et al., 2019). In several other strains, Cas9 was also demonstrated to regulate the adhesion, cytotoxicity and survival during infection in the host (Gao et al., 2019; Saha et al., 2020; Shabbir et al., 2018; Spencer et al., 2019).
…
Type I CRISPR-Cas systems are promising for long-range genome editing in eukaryotic cells (Dolan et al., 2019). Besides, Type I-E CRISPR-Cas system with deleted *cas3* was applied for transcription repression in bacteria (Rath et al., 2015).
The additional role of the Type I-F CRISPR-Cas system in virulence was shown for *Pseudomonas aeruginosa* PA14, where Cas3 targets the mRNA of LasR protein, which is crucial for recognition by the host immunity (R. Li et al., 2016). Type I-C CRISPR-Cas system of *Myxococcus xanthus* was proposed to be involved in the fruiting body development, as the expression of the *M. xanthus cas* genes is tightly regulated by the inter- and intracellular signals and was demonstrated to be activated in the fruiting body, but not in the peripheral cells (P. Viswanathan et al., 2007). In *Salmonella*, which possesses Type I-E CRISPR-Cas system, the deletion of *cas3* reduced the biofilm formation and virulence (Cui et al., 2020), suggesting its role in gene regulation."

3. The PhD thesis is themed especially on links between genome stability and CRISPR adaptation, but there is no mention in the Literature Review of these links in archaea, which seem to be more prolific users of CRISPR systems than bacteria. This should be rectified, being directly relevant to the PhD thesis body of work, especially archaeal Cas3, Cas1 and Cas2, even though the candidate's experimental work was in bacteria. I suggest incorporating the works of Uri Gophna and Anita Marchfelder, as starting points/examples.

- The beginning of the Chapter 1.6. was modified and supplemented with the suggested information on links between genome stability and CRISPR-Cas systems in archaea:

"It is logical to assume that cellular repair enzymes should become activated during CRISPR-Cas system function, as DNA breaks occur during both CRISPR interference and incorporation of new spacers. On the other hand, the observed domain structure similarity of some of Cas proteins with the DNA repair enzymes initially suggested their possible roles in DNA repair (Makarova et al., 2002). Indeed, *cas1* deletion in *E.coli* was demonstrated to affect DNA damage sensitivity and chromosome segregation (Babu et al., 2011). A negative association between Type II CRISPR-Cas systems and NHEJ was observed in bacterial genomes, which can be explained by the inhibition of NHEJ by Csn2 protein (Bernheim et al., 2017). In contrast, the components of the DSB repair system in Proteobacteria, namely RecBCD, SbcCD, and SbcB, were found to frequently co-occur with CRISPR-Cas systems (Bernheim et al., 2019).

There is also multiple evidence of the interplay between CRISPR-Cas systems and DNA repair in archaea. For instance, in *Sulfolobus islandicus*, possessing Type I-A CRISPR-Cas system, CRISPR-associated factor Csa3a was demonstrated to simultaneously control the expression of the CRISPR adaptation and DNA repair genes and activate the DNA damage response (DDR) (T. Liu et al., 2015, 2017; Z. Liu et al., 2020). One of the activated DNA repair mechanisms induces the transfer of chromosomal DNA between cells so that it can be used for homologous recombination. Such crosstalk is proposed to help evade self-immunity, as self-derived spacers are acquired frequently (~7%) (T. Liu et al., 2017). In naturally polyploid *H.volcani*, autoimmunity was shown to be well tolerated and resolved mainly by recombination between two CRISPR arrays flanking *cas* genes (Stachler et al., 2017). Cas1 of the *Haloferax volcanii* Type I-B CRISPR-Cas system was demonstrated to be crucial for cell growth in the oxidative stress conditions (Wörtz et al., 2022). The authors propose Cas1 to act similarly to Fen1 repair protein, which cleaves the flapped DNA intermediates formed during DNA repair and replication. Moreover, in *H.volcani* the Cas3 translocase/helicase activity was found to participate in DNA repair by restraining HR together with Mre11-Rad50 and promoting MMEJ (Miezner et al., 2023)."

4. Page 24, Figure 1: The figure legend needs to be improved, though the figure itself is acceptable – it requires additional detail explaining every visual aspect of the Figure without the reader having to find explanation in the main text. For example, why are some genes black, others grey and others white? What is CRISPR I specifically, not simply CRISPR? This is E. coli work and as such although some readers may be familiar with the specific paper cited for this Figure (ref 125) many others will not be, or at least would benefit from more explanation.

- The figure legend was improved according to recommendations. Additional related information was also added to the main text.

5. Page 27, Figure 2: The figure should be improved by including useful detail such as N- and C-termini of proteins, why different colors are used, and especially, highlighting where the Cas1 active site residues are located, in relation to the PAM(s). Overall, this Figure like others, requires more detail.

- Figure 2 was replaced with the more detailed one. An additional figure was also added showing the DNA-protein interactions in the active site. The corresponding details were also added to the main text.

6. Figure 3 on page 29 is incorrect and needs to be re-written and replaced. The evidence is that the mechanism of integration has the 1st nucleophilic attack at the Leader/repeat junction not where it is shown from this reference (Ref 149). The candidate needs to read, include, and cite Rollie et al eLife 2015 to correct this.

- The figure was replaced by the one from (Nuñez et al. 2016), and the reference to (Rollie et al. 2015) was added according to the recommendations.

7. Page 30-31 section 1.5. Homologous recombination is not adequately introduced or is pretty-much omitted entirely, including the interplay of DNA recombination and DNA replication, which is critical for generating CRISR immunity. This section will require substantial re-writing and/or defense at the panel.

- Section 1.5 was extended according to the recommendations. An additional figure was added, illustrating one-ended and two-ended DSB repair in *E. coli*.

8. Page 39, Project Objectives do not all match well with the content of the thesis. For example, Chapter 4 contains work that is not at all mentioned in the list of Objectives. Suggest this list be modified, or some statement added about how the initial objectives were evolved.

- The following text was added to the Objectives:

"In parallel to the main work, an additional project was done, focusing on primed CRISPR adaptation by the Type I-F CRISPR-Cas system in *Pseudomonas aeruginosa* PA14. *P. aeruginosa* is a human pathogen of high clinical relevance, which causes severe infections, commonly treated with antibiotics. The problem of multidrug resistance raised the interest in phage-antibiotic combination therapies. However, around 36% of clinical *P. aeruginosa* isolates were found to possess functional CRISPR-Cas systems. The project aim was to investigate the effects of antibiotics on the CRISPR-Cas immunity development during phage infection. The objectives were:

1. Conduct evolution experiments with *P. aeruginosa* PA14 infected by the DMS3 phage in both rich and minimal media in the presence of different antibiotics
2. Study the impact of antibiotics on the CRISPR-Cas immunity development
3. Check the effects of antibiotics on the fitness of the resistant clones and phage production

"

9. Figure 15 and associated data have some aspects that will need to be modified and/or discussed. Including the assertion that the Figure 15 data shows SOS response, the reproducibility of Figure 15 data, presentation of the data (fonts). Some re-arrangement of text is needed. It will be useful to discuss how the E. coli SOS response can be confirmed using a 'gold standard' molecular method that is not shown in the thesis.

- Figure 15 was improved according to recommendations. Data points in the figure show mean OD 600 values for 3 replicates for each strain. The replicates were well-matched, and the error bars were not shown to avoid overloading the figure. Indeed, the growth curves in Figure 15 do not show the SOS response itself. However, they clearly show the cessation of growth in induced cultures after ~5h of induction (in contrast to uninduced cells), which, together with CFU reduction and cell elongation, is one of the hallmarks of SOS response. Extensive DNA degradation in the PPS region was earlier demonstrated during CRISPR-Cas induction in this model system (Shiriaeva et al. 2019) and the mutant strains (unpublished data), which allows us to speculate that the observed cell elongation and growth cessation reflect DNA damage-induced SOS response.

10. The candidate should discuss to what degree of certainty there is that Primed adaptation outcomes associated with RecBCD and SbcCD are directly caused by molecular processes of RecBCD/SbcCD (i.e. functional coupling, even if not physical coupling) rather than the loss/gain of each changing more global DNA replication physiology (esp. that at Ter sites) that has an indirect 'knock-on' effect on adaptation. The latter may be, for example, by changing DNA substrates available to Cas1-Cas2/Cascade Cas3.

- In the current model system, PPS-derived fragments constitute the absolute majority (~97%) of newly derived spacers. The data on prespacer generation efficiency and their chromosomal distribution presented in the Thesis clearly show that the decrease in CRISPR adaptation efficiency in Δ*recBCD* mutants results from the lowered prespacer generation efficiency in the PPS region. Moreover, our unpublished data on total DNA coverage in induced strains show, that the gap in the PPS region is narrowed in induced Δ*recBCD* strains and even more narrowed in Δ*recB* Δ*sbcD* mutant, compared to *wt*. The results suggest that RecBCD and SbcCD continue Cas3-

induced DNA degradation. Thus, RecBCD and SbcCD could indeed change the pool of evaluable DNA substrates for CRISPR adaptation. Moreover, RecBCD was shown to participate in the processing of 5' ends of prespacers. Overall, these results allow to propose the functional coupling of priming and RecBCD/SbcCD to be the main reason for the observed effects on CRISPR adaptation.
The viability of Δ*recB* Δ*sbcD* strain was shown to be mildly (by ~15%) reduced after 5h of CRISPR-Cas induction, compared to *wt.* Though we suggest that it should not largely influence CRISPR adaptation efficiency, additional experiments (e.g. plasmid complementation, experiments with different model systems) would be needed to exclude the possibility that repair deficiency in this strain affects the outcome (no CRISPR adaptation is detected in Δ*recB* Δ*sbcD* strain).
As regards the influence of these mutations on replication termination, indeed, we observed differences in short (21-270 nt) non-prespacer-like fragments distribution around *dif* chromosome-decatenation site in Δ*recBC* mutants (Shiriaeva et al. 2022). An extensive degradation around *dif* site was previously reported in *ΔrecB* strain due to repair deficiency (Mei et al. 2021 doi: 10.1126/sciadv.abe2846). Accordingly, two symmetrical peaks of short fragments at ~250kb to both sides from *dif* in *recBC* mutants were observed in our work, suggesting these fragments are formed by the unknown nuclease generating the gap. Nevertheless, though this process could potentially interfere with naïve CRISPR adaptation relying on replication and repair, we suggest it is unlikely to interfere with primed CRISPR adaptation in our mutant strains.

11. The Literature review and RecBCD Chapter data need to cite and briefly comment on prior work linking RecBCD and CRISPR, including data showing the RecBCD nuclease activity is not required for adaptation (Radovcic et al 2018, Nucleic Acids Research) and questions re: Chi sites and CRISPR (Subramaniam and Smith, 2022 Adv. In Genetics).

- Chapter 1.6 "Interactions of repair enzymes and CRISPR-Cas systems" was extended and now includes more discussion of the data on links between RecBCD, Chi, and CRISPR.

12. The candidate should discuss how it can be reconciled that ssDNA fragments formed by the Cas3 translocase-nuclease can be returned into flayed duplexes suitable for DNA capture and integration by Cas1-Cas2. Or, alternative ideas about this.

- The following text was added to Chapter 1.4.2.:

"Though branched dsDNA fragments were initially demonstrated to be the preferred substrates for binding and incorporation into CRISPR array (Nuñez, Harrington, et al., 2015), subsequent single-molecule experiments showed that Cas1-Cas2 can capture various substrates, including ssDNA. Particularly, Cas1-Cas2 binds 3′-TTC-5′ -containing ssDNA fragments and then anneals them to the complementary strands, leading to the formation of a dsDNA prespacer precursor (S. Kim et al., 2020). Together with PAC formation, such mechanism allows Cas1-Cas2 to rapidly bind ssDNA fragments directly transferred from Cas3 and ensures high efficiency of primed CRISPR adaptation."

13. Throughout: Cas3 is called a helicase. But its major function is ssDNA translocation and a 'threshing' mechanism to degrade ssDNA. This is not strictly 'helicase' at all, even though it can (in vitro behave as such). I suggest changing helicase to DNA translocase throughout.

- The suggested change was introduced throughout the text.

14. page 13 top line: 'systemc'

- Corrected.

15. page 14 line 2: insert word endonuclease?

- Corrected.

16. Page 15: definition of Cas does not need the word genes.

- Corrected. Cas is now defined as "CRISPR-associated sequence", as was suggested above.

17. Page 40, Section 3.1 Bacterial strains – there is no mention of how strains were validated/verified other than (presumably) that the P1 had generated the appropriate antibiotic resistance. Were they sequenced? Were they phenotyped to ensure an absence of suppressor mutations?

- The following information was added:

"The deletions in obtained colonies were confirmed by PCR with primers annealing to the gene of kanamycin resistance and those annealing to the regions flanking the mutation site. The genotypes of mutant KD403 derivatives, except those containing Δ*xseA* and Δ*recQ* mutations, were additionally verified by sequencing of genomic DNA with Illumina MiniSeq and analyzing the mutations by the breseq pipeline (Deatherage & Barrick, 2014)."

Professor Peter Fineran

1. The abstract was well written and a nice succinct summary of the main findings of the thesis. If a suggestion were to be made, it would be about including some description of the work presented in Chapter 4 around the role of bacteriostatic antibiotics affecting CRISPR-Cas adaptation.

- The suggested information was included in the abstract:

"A project focusing on primed CRISPR adaptation in *Pseudomonas aeruginosa* was undertaken in parallel with the main work. The results show that bacteriostatic antibiotics promote the evolution of CRISPR-Cas immunity by delaying the production of mature phage particles. These findings can be generalized to other conditions that slow down the speed of phage development. The work shows that, in addition to defective phages and nucleases that cleave phage genomes, the time of phage development inside the infected cell is a key determinant of the acquisition of CRISPR-Cas immunity against the phage. The results further support the evidence that the longer the DNA substrates are present in the cell, the higher is the chance for a CRISPR-Cas system to acquire new spacers from them."

2. Chapter 1 Introduction: The introduction was well written with only a few grammatical or typographical issues. It provided an excellent introduction into the topic. Other defence systems were discussed, but fairly briefly to set the broader context, without overwhelming the reader. Coverage of CRISPR-Cas was very good. The complex topic of DNA repair and genome maintenance was well covered. However, it is my opinion that – due to the complexity – this section would benefit from a few more figures/models of how these work and even showing some of the (hypothetical) models for how they are involved in CRISPR-Cas adaptation.

- Two more figures on DSB repair and the interactions of repair and CRISPR adaptation were added to the Introduction.

3. Chapter 2 Project objectives: This brief 1 page provided an appropriate overview of the I-E work in the thesis. However, similar to the abstract, I thought that the work in Chapter 4 should have been mentioned here in the objectives.

- The suggestion was addressed above.

4. Chapter 4.1: As mentioned above, it would have been good to provide more context for this chapter in abstract, introduction and objectives sections, since it is slightly tangential to the main goals of the project on host factors for I-E primed adaptation. This chapter provided further support for the evidence that increasing the time that DNA substrates are available (by slowing growth in this case) can increase the probability of spacer acquisition. It would have been good for the contribution of the candidate to be more explicit – e.g. state precisely which figures were contributed to.

- The work on CRISPR-Cas immunity development in *P. aeruginosa* is now mentioned and described in all the main sections. The Figures which the author contributed to are now stated in the beginning of the corresponding Chapter.

5. Chapter 4.3 Host nucleases generate pre-spacers for CRISPR adaptation: The candidate investigated the effects of recBCD, recJ, sbcD and sbcB on primed adaptation. This revealed that RecJ and SbcB influenced the frequency (abundance) of spacer acquisition. In contrast, RecBCD and SbcD influenced the distribution of spacers. Significant effects on PAM selection were also detected in these high-throughput sequencing assays. RecJ affected the frequency of flipped spacers and those containing internal AAG motifs. Further work using a combination of deep sequencing and in vitro assays led to a convincing model for the role of the host nucleases and helicases during pre-spacer formation. I would have liked to see a schematic of the model to help aid the reader in understanding the proposed mechanism. I actually think that bringing Figure 36 earlier into the thesis and using it to help frame the questions and explain the data as it arises could be useful.

- A scheme of the proposed primed adaptation mechanism highlighting the questions addressed in the Thesis was added to the Objectives.

6. Chapter 4.4 Detection of half integrated pre-spacers: A newly developed assay in the lab enabled either half-site integration to be sequenced and quantified in the various mutant backgrounds. This provided data that supported the model proposed in 4.3. It would be useful to provide a schematic outlining the principle of this methods since it has not been published previously.

- A scheme for the HIPs preparation procedure was added accordingly.

7. Chapter 5 Discussion: The candidate clearly summarized the results of the thesis and provided a nice graphical representation of the proposed mechanism. It would have been nice to have some zoomed in regions in this figure showing some of the mechanism in a little more detail.

- Two more figures were added to the Discussion, which show the proposed roles of RecBCD, RecJ, and ExoVII in CRISPR adaptation in more detail.

8. Referencing: I would prefer to see full author lists in the references. This helps work out which groups/teams have contributed to the work discussed.

- The referencing was corrected.

9. It would be good to hear from the candidate what they view as the limitations of the study (this was not covered particularly in the thesis).

- The biggest issue in studying the role of the cellular repair proteins in CRISPR adaptation is their importance for cell survival, which limits the mutations that can be introduced into the genome. This further complicates the investigation of the redundant pathways. The *ΔrecBΔsbcD* mutant had a modestly increased level of dead cell accumulation after CRISPR induction (~15% of the population). The use of the self-targeting system in our experiments allowed to avoid the problems regarding plasmid instability in mutants lacking DNA repair genes. On the other hand, the extensive chromosomal DNA degradation and subsequent SOS response induced in our model systems could affect the results. Moreover, the size of the DNA target should also affect the role of cellular enzymes involved in priming. Thus, other model systems should be used to further support the conclusions. The experiments with self-targeting strains targeting other protospacers in the genome would be also beneficial to prove the observed effects.

10. It would also be interesting to hear their opinion about the next big questions that the work has raised and what approaches they would take to tackle these.

- While we explored the involvement of the cellular repair enzymes in the 5' prespacer end processing, the mechanisms of the 3' ends trimming remain unclear. Other cellular nucleases which we did not include in the study should be checked. Besides, the role of different RecBCD catalytic activities should be checked using point mutations lacking the nuclease, helicase, and Chi-recognizing activities, to figure out its role in primed CRISPR adaptation in more detail. It would be also interesting to investigate the role of repair proteins for other types of CRISPR-Cas systems using the same methods.

Professor Yuri Kotelevtsev

No comments to address.

Professor Mikhail Gelfand

1. Three paragraphs of Chapters 4 "Results" are based on three published papers. The organization of this chapter is somewhat paradoxical: the first paragraph is not directly related to the main project, although is loosely linked to it; moreover, its results are not reflected in the conclusions. Probably, in a more natural order, it would have been the last one.

- The order of the chapters was changed according to the recommendation. The work on CRISPR-Cas immunity development in *P. aeruginosa* is now mentioned and described in all the main sections.

2. The results of the study are interesting, as they demonstrate and partially dissect the mechanism of repair machinery on CRISPR-Cas adaptation process. They are convincingly integrated in the Discussion and summarized in the Conclusions. The latter should have been more transparent — for instance, what is the intended sense difference between "involvement <…> of RecBCD and SbcCD pathways in spacer acquisition" (1) and "2. RecBCD and SbcCD affect the pattern of acquired spacers"?

- The conclusion on the effects of RecBCD and SbcCD on patterns of acquired spacers was changed:

"2. RecBCD and SbcCD affect the pattern of acquired spacers, suggesting the enzymes could interfere with the processivity of PAC."

While the affected patterns of acquired spacers rather deal with the altered PAC processivity, the effects on CRISPR adaptation efficiency could also reflect the involvement of the corresponding enzymes in other steps of CRISPR adaptation, which was addressed in further experiments.

3. Nothing is said about reports at conferences, so this needs to be addressed during the defense.

- The list of conferences was added.

Professor Ekaterina Khrameeva

1. Regarding the section 3.2.2: - It is unclear how reads containing two or more repeats were searched for. Was some specific software used for this task, or a custom script? I guess it is important to make sure that all such cases were accounted for, therefore these details matter.

- Custom scripts were used for all the raw data analysis, including the selection of reads with two or more CRISPR repeats. This is now mentioned in the text.

2. About non-uniquely mapped spacers: did you check that they originate from rRNA operons? I could not find this analysis in the thesis.

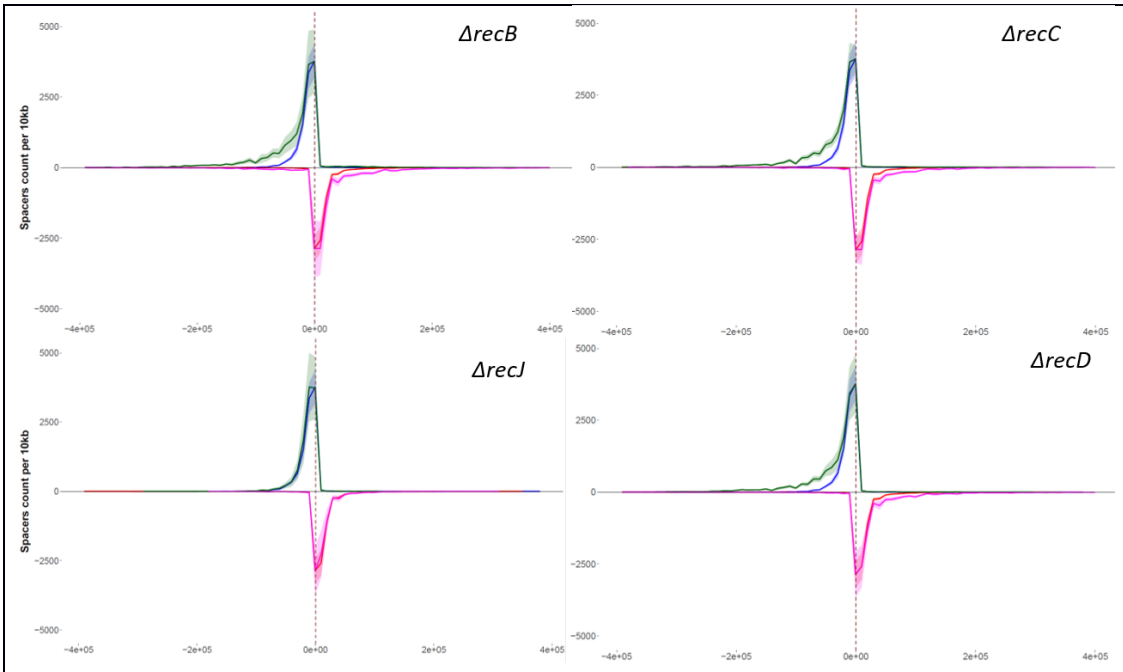
- Indeed, the largest peaks of non-unique new spacers were found to correspond to rRNA operons. This analysis was done in (Shiriaeva et al.2022) but was not included in the current Thesis.

3. About the normalization to the total spacer counts: did you also normalize for the library size (total number of reads)? Is it important to normalize for the coverage difference between samples?

- For the analysis of acquired spacers, only the normalization for the total spacer counts was done, as the goal was to compare the patterns of acquired spacers, not the efficiency of CRISPR adaptation. In this case, the coverage difference seems not relevant. Also, to directly compare the curves, maximal values for mutants and wild-type were adjusted to make them the same, which is indicated in the text.

4. It is ok to average results across two independent experiments but their high similarity has to be demonstrated first, to ensure reproducibility of results. This analysis of replicates consistency should be added to the thesis.

The results of the comparative analysis of CRISPR adaptation patterns in the same strains were reproduced in several experiments performed by different people in our lab and were highly consistent between the experiments. This is now indicated in the text. The data obtained in (Shiriaeva et al. 2022) from 6 biological repeats better illustrates the consistency. The data presented in the figure below was processed in the same way as described in the Thesis, with the difference that the shaded area represents the range of values in 6 biological repeats.



5. *Page 74: "at the border of statistical significance" is more conventionally formulated as "marginally significant". Consider this replacement.*

- The change was introduced.

6. *In Figure 27, Welch's t-test is used to assess the statistical significance of differences. However, this test requires normality of the data as one of its assumptions. Was it tested? I could not find it in the thesis. Either check the normality here and in all similar figures in the thesis or consider replacing with another test not requiring normality (Wilcoxon test).*

- The normality of the data was checked by the Shapiro-Wilk test with a 95% degree of confidence.