Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Alexander Martynov

PhD Program: Life Sciences

Title of Thesis: Using mathematical modeling to understand prokaryotic adaptive immunity

Supervisor: Prof. Konstantin Severinov

Co-Supervisor: Prof. Jaroslav Ispolatov

Chair of PhD defense Jury: Prof. Mikhail Gelfand ***Email**: m.gelfand@skoltech.ru*

Date of Thesis Defense: 25 October 2018

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 14-10-2018**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

- **Brief evaluation of the thesis quality and overall structure of the dissertation.**

This is a mathematical analysis of characteristics of CRISPR systems that can promote their maintenance and optimal performance within the cell, and of kinetics of CRISPR-plasmid ecological interactions. It includes a detailed and well-structured literature review, and two main parts. In the first part, the optimal characteristics of a CRISPR system are modeled, depending on parameters of the system including binding efficiency, rate at which the crRNA abundance decreases in the 5' to 3' direction, and viral mutation rate. Some results are obtained, which appear general and relevant. In the second part, which also included experimental work, the author demonstrates an "Allee-effect"-like dynamics of the plasmid density under CRISPR interference, whereby an unstable equilibrium exists below which the plasmids go extinct but above which they thrive.

- **The relevance of the topic of dissertation work to its actual content**

Although the title is perhaps somewhat too general, it sums up the two projects nicely.

- **The relevance of the methods used in the dissertation**

Mainly old-style ODEs are used, which is great because this way one can actually understand what is going on. There is also an experimental part informing the second model.

- **The scientific significance of the results obtained and their compliance with the international level and current state of the art**

The results are without a doubt interesting and relevant to the currently pressing issue of how CRISPR immunity works and why it works the way it does. To inform the model, up to date knowledge about the structure of the system and estimates for parameter values are used. The proposed model predicts some features of existing CRISPRs, and makes testable predictions for others. Most importantly, it predicts how the optimal number of spacers depends on parameters – some of which can be tweaked experimentally, so the model is testable. The model is comprehensive. Even though its details are nearly certain to be imprecise, as not enough is known about CRISPR interference, it is very useful for informing intuition; e.g., the set of "rules of thumbs" set out in p. 58 can be tested experimentally. While the author doesn't discuss the evolutionary mechanism by which the optimality is maintained (if it is), knowledge of the position of the optimum is undoubtedly useful for further unraveling of these mechanisms.

- **The relevance of the obtained results to applications (if applicable)**

CRISPR system is a big thing in biotechnology, but the conceptual understanding of how and why it has evolved and is maintained by prokaryotes is lagging behind. This work is a potentially important contribution.

- **The quality of publications**

The one publication listed so far is a 1st paper in PLOS Comp Bio, which is a major journal for publication of mathematical models.

The summary of issues to be addressed before/during the thesis defense

The quality of English throughout the text is very uneven. Some parts read like a charm; others are nearly incomprehensible. For example, the very second phrase of the thesis (p. iii, "it is as an anti-viral tool...") makes no sense. Sometimes, the low quality of English impedes understanding.

The second major concern is confounded structure. It appears that large segments of text were copy-pasted from the paper. This is not a problem by itself, but the problem is that many of the links and references were badly confounded along the way. The text includes multiple references to non-existent supplementary materials or "main text" (e.g. p. 48); to equations with non-existent numbers (e.g. p. 48); and to non-existent figures (!) (e.g. Fig. 5 in p. 83). Conversely, none of the figures in chapters 1 and 2 are referenced in the text. The text discussing the experiment shown in fig. 1.5 (even though this figure is not mentioned in it) only occurs in p. 26, while the figure itself is in p. 23. The mutation probability in fig. 3.3 (p. 49) is not explained till p. 53. (And, to add insult to injury, the usual notation for the mutation rate, \mu, is the opposite to that used here, 1-\mu.) All this makes the text very hard to comprehend.

On a more scientific note, the key assumption of the first model is that the involvement of a spacer in an effector complex declines exponentially with spacer age. The author motivates this by pointing out that the 5' crRNAs tend to be younger; and that the 5' crRNAs are expected to be generally more abundant than the 3' crRNAs (although they provide no data on the shape of this dependence). Although this is

outside the scope of this work, it would be interesting to see if the results change a lot when these assumptions are violated. What would change if the decay of crRNA abundance with distance from the promoter is, in fact, not exponential; e.g., if there is a threshold length, up to which all crRNAs are used equiprobably?

In chapter 4, the author suggests that the surviving plasmids have undergone a period of stochastic expansion. As this expansion is "against odds", probability theory tells us that it would have to be fast. Therefore, if the plasmid is fixed in a cell, it has experienced an unusually rapid period of initial expansion. Can this be tested somehow? (Both above questions go beyond the scope of the thesis, but are interesting if this topics are pursued further.)

There are some apparent errors. Fig. 3.3 caption seems to contradict the text under eq. 3.18: should the product be maximized or minimized? Less importantly, for the references to "altruism" (e.g. p. 70) to be valid, it needs to be shown not just that the cell decreases the number of secondary infections, but that it does so at a cost to its own survival. As far as I understood, this is not shown, at least not explicitly.

These criticisms in no way undermine the fact that the author has put a substantial amount of work into the projects, and has produced a solid thesis which I read with much interest.

Provisional Recommendation

X *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*