Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Karyna Karneyeva

PhD Program: Life Sciences

Title of Thesis: Exploring type III CRISPR-Cas immunity in *Thermus thermophilus*

Supervisor: Professor Konstantin Severinov

Name of the Reviewer: Dr Robert Fagerlund

I confirm the absence of any conflict of interest	**Date: 18-03-2024**

Reviewer's Report

Exploring type III CRISPR-Cas immunity in *Thermus thermophilus*

Karyna karneyeva

Karyna has investigated the type III-A and type III-B CRISPR Cas systems in *Thermus thermophilus*, a model organism for CRISPR-Cas research. Type III CRISPR-Cas systems provide a sophisticated immune system that relies on targeting RNA for cleavage, in some cases cleavage of single-stranded DNA, and the production of a signalling molecule for activation of accessory proteins. There are several known subtypes of type III CRISPR-Cas systems, with the most abundant and well-characterised being the type III-A and -B systems. The importance of this study is the *in vivo* characterisation of function in the natural host cell. The results chapters were in the logical order of first demonstrating type III immunity against phages and then the requirements of interference. It was demonstrated that both type III systems were active and they used guides from the same CRISPR arrays to carry out immunity. A very interesting aspect of the first results chapter was the genetic analysis of escaper cells that avoid immunity, where mutations ranged from the removal of 3 bases where the CRISPR complex targets to massive deletions from recombination occurring between repeats in different CRISPR arrays. The second chapter undertook a comprehensive investigation into the target sequence requirements for systems, which showed the systems have a tolerance for mismatches and length of complementarity between the guide and target RNAs. The type III-A system was studied further and the crucial role of the HD nuclease domain was demonstrated.

The thesis was written at an excellent standard and the thesis was structured in a logical order. The content and chosen methods were relevant to the topic of the dissertation. Type III CRISPR-Cas immunity is complex and the full understanding of the mechanism of the different subtypes is still being discovered. To that end, this thesis provides significant results towards revealing these mechanisms, and importantly in an *in vivo* context. The methodologies and results comply with international levels and are performed at a world-class standard. This is demonstrated by the high-quality papers that Karyna has contributed to, notably her first-author paper in the Journal of Molecular Biology and second-author in Nucleic Acids Research.

I have listed comments and questions below by chapter. Key points or questions are in bold.

Abstract:

It is stated that *Thermus thermophilus* is a convenient model for Type III CRISPR-Cas research. **What are the pros and cons of this system?**

Chapter 2:

Page 32 The thesis writes that type III CRISPR-Cas investigations have had contradicting results. Highlighting holes in the literature is important, but care is needed when describing past contradictory results with present-day hindsight. For example, early work on type III-A and -B systems did appear contradictory; however, these papers were published before the mechanism of accessory nuclease activation was revealed, which may explain many of these early results. Best to focus on what is known now.

Page 35 Note Csm and Cmr complexes are comprised of 5 and 6 *different* complexes. Total number of subunits per complex is often >10.

Pg 35,37 There are two Palm domains but only one is active and has the GGDD motif.

Chapter 2 would be considerably enhanced if it ended with a summary and direction towards the presented work. Why was this introduction important? What are the key questions this thesis addresses?

Chapter 3:

Appears comprehensive and in a logical order. See note about Acr methods.

Chapter 4:

It was not clear to me that both type III systems recognise the same repeat sequence. I don't believe this is common for co-occuring type III-A and -B systems, and therefore made understanding and interpreting some of these results difficult. **It needs to be stated early that all arrays have the same repeat sequence** (is this the case?) and both systems can use all arrays. If repeats are different, consider a table showing each sequence.

Fig 4.1 Highlight the accessory proteins. An early mention of each one present should be stated. This is an important part of interpreting the results as it wasn't clear which immunity mechanisms were in play. Do both systems have adaptation genes?

Fig 4.3 The analysis of this result on pg 65 **would be strengthened with quantification of the EOP.** Without this, some subtle effects may be missed. For example, are spacers #33 and #26 lower than the control? It appears #36 has approx. >1,000x decrease in EOP – is partial the best term to describe that reduction?

Fig 4.4 Are these essential genes? What would happen if non-essential genes were targeted? Knowledge of the accessory genes is important to understand this experiment. In the legend, what do the colours on genes represent?

Pg 67 Why were the three genes selected as "strong Acr candidates"? How were the Acr expression plasmids constructed? I don't see this in the methods. Is expression induced? Are you confident proteins were expressed?

Pg 69 Typo in the figure for pBAD33.

Fig 4.9 and 4.10 Nice work on the RNA sequencing to determine early, middle and late expressed genes. In my experience, BPROM can miss promoter sites and other tools, like that from Silas' group, can be useful. Are all promoter sites identified? I expect some analysis and correlation of promoter sites to the heat map. Is there a promoter in front of genes 4 and 18? Genes coloured lighter may better differentiate some genes.

Pg 74 and in discussion The thesis claims there is no correlation between temporal class and protection efficiency. Can you be certain when only one spacer from late and middle were tested?

Pg 75 Quantification of EOP would be useful for the comparison of WT to the mutant strains. The thesis writes that they have comparable efficiency, but a closer look could reveal other interesting observations.

Is there a difference between WT and the single mutants? Are these systems having an additive effect? Does this add some weight to the reasoning for why there are two systems?

Fig 4.17 Do the regions without mutations in panel C imply they are less important for targeting? How does this correlate with results from chapter 5?

Pg 83 Why is co-transcriptional targeting likely to be primarily directed towards nascent transcripts produced by the RNAP? What evidence is there that the single-subunit polymerase transcribes early genes? The cited paper refers to polymerase from an *E. coli* bacteriophage is expressed early to mediate late mRNA synthesis.

Chapter 5:

Pg 86 typo repeated "downstream of".

Fig 5.2. Are all repeats the same in the arrays? Could slight differences explain why some regions have no recombination? Are there spacer similarities between the "hot spots?

Pg 92 The percentage change in interference efficiency is recorded. Is this from EOP data? I recommend not referring to bases as residues as this could be confused with aminoacids.

Pg 94 The thesis concludes "The results support the fact that target abundance determines the degree of tolerance to mismatches", this is also discussed later. The experimental setup is complicated by the promoter on the bottom strand, which would result in duplex RNA forming and triggering a bacterial response to eliminate this. **In hindsight, what features in your plasmid design would you include to manipulate target expression levels and prevent antisense expression?**

Fig 5.5 CRISPR-Cas expression is typically tightly regulated to avoid fitness costs and autoimmunity. Would Cas levels be different in infected cells? I'm not sure how differences in subunit abundances correlate with stricter target complementarity requirements.

Pg 109. It appears the HD mutant still has some interference activity (about 10-fold reduction). How does this result compare to type III-A systems from other bacteria? Were the accessory genes still present in this setup? How does that impact data interpretation?

Pg 111 and discussion It is proposed that a decrease in plasmid copy number explains the decreased growth rate. **But why is there a drop in copy number?** Rather than an alternative mechanism of type III

function, could it just be that interference is still occurring but at a slower rate that doesn't clear the plasmid?

The discussion on why *Thermo* has two systems was very interesting. Investigation into the role of the accessory genes would be important towards this end. **Could differences in gene regulation also be important, where the systems are turned on by different stimuli?** What is known about the transcription profile of *T. thermophilus* under different stresses?

Pg 120 What is meant by "distinct pathways" in the type III-B system? Is this speculation of non-interference roles? Or it must rely more on the activation of accessory proteins?

General comments:

The convention for labeling CRISPR-Cas systems is to include "type" before the Roman numerals.

Generally referencing literature is done well. It is important to remember to include the reference when stating a fact.

Pg 29, Thesis has "in the presence of antisense transcription". Is this in reference to CRISPR array antisense?

Check the spelling of cOA, in places it is cAO

Ref 22 has first names.

Provisional Recommendation

× *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*