Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Maria Sokolova

PhD Program: Life Sciences

Title of Thesis: Functional and Structural Analysis of a Non-Canonical RNA Polymerase Encoded by Giant Bacteriophage AR9

Supervisor: Prof. Konstantin Severinov

Chair of PhD defense Jury: Prof. Yuri Kotelevtsev ***Email**: y.kotelevtsev@skoltech.ru*

Date of Thesis Defense: 15 June 2018

Name of the Reviewer: Professor Finn Werner

I confirm the absence of any conflict of interest	**Signature:**  **Date: 19/05/2018**

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

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

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

Report on Maria Sokolova's PhD thesis

Overall assessment

This thesis is eloquently written and presents a biochemical research project of the *highest caliber*. The introduction chapter describes many important aspects of the structure, function and evolution of multisubunit RNA polymerases, and of phage transcription systems and their regulation. The first results chapter describe the production and purification of the AR9 RNAP from infected Bacillus cultures, and a detailed functional analysis using in vitro transcription, DNA-binding, DNaseI- and Potassium permanganate foot printing. The second results chapter describes the production of recombinant AR9 nvRNAP using an *E. coli* expression system, and several (mostly) successful crystallization and structure determination attempts (using core RNAP, and holo RNAP-promoter complexes), and the early stages of a cryoEM approach for structure determination. Each chapter ends with a short conclusion section, and the thesis is rounded up with a discussion chapter putting all discoveries in context with the state of the art according to the current literature. All chapters are accompanied by thoroughly annotated, very clear figures that are informative and well presented, very *easy on the eye*. Referencing is solid throughout.

This thesis is the result of a very thorough and labour-intensive research project, the thesis is written in an authoritative and convincing style, and the author is clearly in command of her field. In summary, this combined effort merits the award of a PhD degree without any concerns or hesitation.

Research highlights

This thesis reports the following discoveries that are of high interest to the field of transcription. The AR9 nvRNAP is unusual in several interesting ways: the (i) DPBB catalytic subunits do apparently not require any assembly platform subunits to form a catalytically active RNAP, (ii) the subunit responsible for transcription initiation bears no striking homology to known sigma factors and therefor may provide a 'new' solution to an old problem (i.e. mechanisms of transcription initiation), (iii) this enzyme specifically recognizes Uracil-containing promoter templates (and will not utilize normal, T-containing templates) and (iv) single-stranded promoter DNA. These findings imply that the promoter recognition and nucleic acid strand handling (in particular RNA-DNA separation) of the AR9 nvRNAP is different from canonical msRNAP. This is all very exciting, and I cannot wait to see the high-resolution X-ray and cryoEM structures of the AR9 nvRNAP initiation complex!

Publications

The candidate's research project described in this thesis has resulted in three excellent publications:

1. A non-canonical multisubunit RNA polymerase encoded by the AR9 phage recognizes the template strand of its uracil-containing promoters. **Sokolova M**, Borukhov S, Lavysh D, Artamonova T, Khodorkovskii M, **Severinov** K.; Nucleic Acids Res. 2017 Jun 2;45(10):5958-5967.
2. Transcription Profiling of *Bacillus subtilis* Cells Infected with AR9, a Giant Phage Encoding Two Multisubunit RNA Polymerases. Lavysh D, **Sokolova M**, Slashcheva M, Förstner KU, **Severinov** K.; MBio. 2017 Feb 14;8(1).
3. The genome of AR9, a giant transducing Bacillus phage encoding two multisubunit RNA polymerases. Lavysh D, **Sokolova M**, Minakhin L, Yakunina M, Artamonova T, Kozyavkin S, Makarova KS, Koonin EV, **Severinov** K.; Virology. 2016 Aug;495:185-96.

These are all very respectable, peer reviewed journals, and in particular the first author NAR article will support the candidate's standing in the field and enhance her opportunities for securing future appointments as postdoctoral researcher.

Key research questions to be discussed during the viva

1. In your opinion, was the ancestral DPBB RNAP a single or a multisubunit enzyme? Are the phage RNAPs including AR9 and PhiKZ *derivatives* of cellular RNAP, or *progenitors*? Which methods could one develop and apply to gain insights into this question?
2. I would have liked to see complete (protein) sequence alignments of AR9 RNAP subunits with sequences of msRNAPs from all domains of life and including viral sequences such as PhiKZ RNAP, annotated with all functional elements of canonical msRNAP. Are there 'obvious' extensions or additions that could account for the absence of alpha-like assembly platform subunits?
3. The AR9 RNAP subunits correspond to split beta/beta prime subunits. Are the split sites the same as found in archaeal (and some chloroplast) Rpo1 and -2 RNAP subunits? What are the evolutionary implications for either case?
4. I'm intrigued by the ability of AR9 RNAP to utilize ssDNA promoter templates. This works, as does the RNA-DNA strand separation. Why do you think that the AR9 RNAP has evolved to cope with this situation? (I don't buy the argument that the 72% GC genome melts spontaneously as reason) What is known about AR9 replication, are there single stranded stages that are 'long lived' and relevant as transcription templates?
5. Several phages shut down host transcription using *early* expressed genes that specifically inhibit the host RNAP (e.g. the T7 gp2 protein). How does AR9 inhibit host transcription? Or does it not need to inhibit the host RNAP interference, since its genome contains U instead of T residues?
6. You prepare a E. coli produced recombinant AR9 RNAP for structural analysis. Have you ascertained that no important post translational modifications (PTMs) are missing in this type of prep? AR9 is unusual (U rather than T), and missing PTMs could thwart or compromise your structure.
7. You provide a comparative analysis between AR9 and Sso RNAP by discarding a significant amount of the Sso RNAP structure (all non-homologous polypeptides). How does your structural model of the AR9 RNAP compare with the archaeal polD DNAP structure, which is much closer to AR9 (than Sso RNAP) with respect to size and subunit complexity?

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*