BIOLOGY CENTRE CAS
Institute of Parasitology
address: Branišovská 1160/31, 370 05 České Budějovice, Czech Republic
IBAN – CZ24 0800 0000 0000 0606 3942 | SWIFT CODE – GIBACZPX | VAT No.: CZ60077344
phone: +420 387 775 403 | fax: +420 385 310 388 | www.paru.cas.cz | e-mail: paru@paru.cas.cz

POSUDEK OPONENTA NA BAKALÁŘSKOU PRÁCI

Autor práce: Dominika Reichensdörferová

Název práce: Development of DiCre parental lineage of *Babesia divergens*

Školitel práce: RNDr. Daniel Sojka, PhD and RNDr. Marie Jalovecká, PhD.

Oponent práce: RNDr.Alena Zíková, PhD.

Pracoviště oponenta: Biologické centrum AVČR, České Budějovice

It was my pleasure to review thesis by Dominika Reichensdörferová entitled Development of DiCre parental lineage of *Babesia divergens*.

I am a big fan of developing new genetic tools for parasites of medical and veterinary importance and am aware of the difficulty of this task. Even with the development of genome-wide applications of Crispr/Cas9 to study the loss of function of genes at the genome level, there is still a need for a tool to study the function of essential genes. In this case, we need to be able to turn on and off the expression of the gene of interest to determine its functions.

In this work, the student has chosen to initiate the setup of the DiCre system. This system is quite complex as it requires the introduction of the dimerizable Cre recombinase into the genome of the parasite. In addition, the gene of interest must be Lox-flanked to ensure site-specific excision of the target DNA.

The student takes advantage of the fact that this system is already established in related organisms and therefore there is some likelihood that it will also work in *B. divergens*. To establish this system, several conditions must be met: identification of a locus to which the DiCre recombinase will be targeted, generation of vectors that ensure expression of DiCre from the endogenous locus, testing of drugs used for selection, testing of toxicity of rapamycin, development of a method to introduce LoxP sites flanking the gene of interest, optimization of the transfection protocol, elaboration of a proof-of-concept strategy.

The student was able to achieve some of these goals. She designed and assembled a plasmid encoding for both DiCre subunits flanked by *B. divergens* specific UTRs, she sequenced and

therefore mapped the site for introduction of the DiCre cassette using homology recombination, and she attempted to determine concentrations of two drugs for selection of genetically modified *B. divergens* after the transfections.

In my opinion the submitted thesis fulfills all the requirements set by the Faculty of Science and the student deserves the be awarded the title of MSc.

In České Budějovice, 19.5.2022

Alena Zíková, PhD.

Questions and suggestions:

1. There were two drugs tested for their potential as selective markers, G-418 and WR99210. Can a student explain how these drugs work to select drug-resistant genetically modified organisms?

2. In her introduction, the student mentions several different methods of achieving conditional genome editing (e.g., Tet-regulated promoters, ribozymes, TetR aptamters, FKBP-DD, DHFR, but not RNAi). Why was the DiCre method chosen? Is the RNA interference pathway missing in these parasites?

3. What is the maximum parasitemia that can be achieved with *B. divergens* and bovine blood in the in vitro system?

4. From the text it is not clear how parasitemia is determined by flow cytometry. The methodology describes sample preparation well, but lacks a description of data analysis. How exactly is the area under the curve (AUC) associated with parasitemia? If student can give me a one example of data analysis that would be great.

Result 1.

5. *"Figure 8 shows the statistical determination of the IC50 value on 8 DPI after triplicates were averaged, their standard deviation was determined, and the data was* ***normalized (logarithmized)".***
   Normalization is not the same as converting values to the logarithm. Speaking of normalization, do you think it would be useful to normalize your data?

6. How was the concentration scale chosen for WR99210? The concentration gradient starts with a dilution of 2 (20nM, 10nM, 5nM) and then suddenly jumps to a dilution of about 4 (1.25 nM, 0.31 nM, 0.07 (should be 0.08) nM, 0.02). The curve is missing the value that

indicates the middle of the curve (drug's potency). Your NC is your baseline (no effect), while the maximum effect of the drug is well defined by three concentration values. If dilution 2 had been maintained, a middle value would be most likely (0.625 nM) and would increase the credibility of the curve.

7. Figure 8 - it is surprising to me that the curve reaches negative values, it would be helpful to understand how exactly the analysis was done in Graph Pad Prism. There are several ways in which the data can be analyzed.

8. The effect of a drug generally depends on the exposure time. Therefore, the IC50 is a function of exposure time. How did you choose the time of 8 days and for what reason? Do you think the IC50 values would be different if you had chosen a different time point? How is the timing relevant to the subsequent transfection protocol?

9. The concepts of IC50 and EC50 are fundamental to biological research. Does the student know the difference between the EC50 and IC50 definitions? Can she also calculate the EC50 values?

10. How did you define that the WR99210 concentration of 5 nM was optimal? The drug seems to work quickly. Would this concentration give you a long enough window for genetic recombination to occur?

Result 2.

11. In contrast to WR99210, the data on G-4218 are not as conclusive. The response is delayed and the maximum inhibitory effect was not reached. From the data presented, it appears that the dose-response curve is incomplete. The data simply do not form an upper or lower plateau, and even if the curve "fits," all IC50 values are meaningless and must have a wide confidence interval. This section of the paper is inconclusive. I want to emphasize that this is perfectly fine, and that it is much better to state this than to give an incorrect IC50 value on which to base the concentration for the selection procedure. How should you have proceeded in this case?

Result 3.

12. This section reads like a method section. No explanation of what segment Bdiv_04560c is, how it was identified, why it was amplified, why it was sequenced. Why is this locus considered a suitable recipient locus?

Result 4.

13. Here I am missing images of DNA gels showing the progress and verification of the assembly of the final plasmid. How was correct assembly verified (restriction digest, pcr)

and/or by sequencing. Which strategy worked in the end? Combining cassette A with PCR4 or with two smaller fragments PCR2 and PCR3?

Discussion

14. *"Unique restriction sites were added to the DiCre plasmid so that the 5'UTR and 3'UTR regions may be replaced for future DiCre cassette editing."*
Why would you like to replace the UTRs region in the future?

15. *"Based on a literature search, we have selected the 6-Cys-E gene (Bdiv_04560c) as the homologous recipient locus of the B. divergens genome for the DiCre cassette incorporation. This gene is highly homologous to the p230p gene (PF3D7_0208900) that has been used as a stable recipient locus for the DiCre cassette transfections of the Babesia related malaria parasite P. falciparum (Knuepfer et al., 2017)."*
What does "highly homologous" mean? If the locus is "highly" homologous, can you adopt plasmids from the Plasmodium genetic toolbox that use this region for incorporation to transfect Babesia cells?

16. *"Even though this mutation changes one amino acid in primary protein structure residue (arginine to serine), no effect on gene expression is expected."*
Why are you interested in a mutation in the primary sequence of amino acids? If I understand it correctly, you are going to insert the cassette into the gene, disrupting its function, so why are you interested in the gene expression of this gene?

17. *"Therefore, WR99210 concentration of 5 nM and G-418 concentration of 3000 µg/ml will be used as optimal for transfection systems of B. divergens."*
I would advise not taking the 3000 ug/ml as a definitive answer. Especially since it is different from the value used for Plasmodium. Also, I think the concentration range must be tested for the transfection procedure anyway. The determined value is only a rough estimate.

18. *"To verify that DiCre recombinase expressing B. divergens lineage can mediate rapamycin-induced excision, wild type (WT) and parental DiCre lineages will be transfected with the green fluorescent protein (GFP) reporter episomal plasmid where the GFP gene will be flanged by loxP sites."*
Is the rapamycin conctrenation for induction known? Has it been determined for B. divergens? Considering that DiCre recombinase operates in the nucleus, has it been demonstrated that it can also act on an episomal plasmid localized in the cytoplasm?

19. When introducing a novel genetic tool, it is always good to perform a proof-of-concept analysis, i.e., to choose a known gene whose effect when knocked out is known and can be easily followed. Would the student suggest such a gene for B. divergens?

Minor points (not to be read at the defense). These are just for the student to learn how to use language properly:

Annotation:

20. *"We have validated and determined the optimal concentration of WR 99210 and G-418 for their use as selection markers"*

    - None of this was accomplished. You would have "validated" if you used the concentration determined for the transfection and the transfection was successful, i.e., you would get properly genetically modified B. divergens. You would have "determined the optimal concentration" if you performed transfections at different concentrations and found that you obtained the best transfection efficiency at that concentration. The correct conclusion from your study is that you have determined the IC50 values for WR99210, and this value is a good starting point to determine the optimal concentration for transfection and validate this drug as a selection marker.

Introduction: please, do not use abbreviations without their explanation e.g. Flp/FRT, DiCre, GlmS, or TetR, or FKBP-DD, DHFR, BSD, hdhfr etc.