

HABILITATION THESIS REVIEWER'S REPORT

Masaryk University

Applicant

Dr. Stanislav Mazurenko

Habilitation thesis

Advanced Data Modelling for Protein Engineering

Reviewer

Prof. Dr. A.D.J. van Dijk

Reviewer's home unit, institution

Biosystems Data Analysis, Swammerdam Institute for Life Sciences, University of Amsterdam, The Netherlands

The candidate, Dr. Stanislav Mazurenko, has broad experience in the field of computational biology. He has been involved as supervisor of multiple doctoral theses, as well as being consultant in the case of a few others. The research in these projects contributes both to development of fundamental methodology for bioinformatics and systems biology, as well as to applications to address specific biotechnological and biomedical issues. Based on this, in my opinion the applicant definitely is well prepared for the promotion from assistant professor to associate professor. One aspect that I could not completely judge based on the provided information is his contribution to education. I had access to a list of courses to which Dr. Mazurenko somehow contributes but the nature of that contribution was not clear to me. I am not sure how important educational contribution is in the habilitation process so I will ignore it in the rest of this report.

Focusing on the habilitation thesis, the methods described in it involve computational methods for the analysis of complex biological data. There is a particular focus on protein engineering, and various methods are described aimed at enabling protein engineers to plan experiments, analyse their data, and interpret the results of data analysis. The thesis distinguishes two main types of computational approaches: (1) low-parameter modelling, primarily based on physical principles; and (2) machine learning-based approaches, which do not start from physical principles but focuses on identifying predictive patterns in the data.

In the introduction, these ideas are well explained, and a number of protein-related tasks are described. Chapter 2 describes the low parameter modelling in a clear way and also gives an overview of relevant gaps and limitations that were addressed in the work of the applicant. Compared to chapter 2, I felt that chapter 3, on the second type of approaches, viz. machine learning, manages a bit less well to give a good overview. This might be unavoidable given that the field of machine learning even on a specific application domain such as protein engineering is vast. Nevertheless I would have appreciated a bit more attempt to give some structure to the types of tasks which are relevant. On a positive note though, the methodology that is presented in this part of the thesis is clearly highly creative and addresses various research frontiers. Also, after the description of the contributions, there is a quite useful discussion on limitations and needs for improvements in this type of work.

Overall I consider that the thesis (including the various research publications which are part of it) definitely warrants the promotion from assistant professor to associate professor.

Reviewer's questions for the habilitation thesis defence (number of questions up to the reviewer)

One main question/comment relates to how the difference between “low parameter modelling” and machine learning is described. For example, in section 1.3 it is claimed that for machine learning, a part of the data is initially put aside and used only to evaluate the generalisability of the fitted model, and that this would not apply to low parameter modeling. The same thought is repeated on p.19 (“To make sure that only meaningful patterns are captured, only part of the data is typically used for fitting the parameters, and the remaining part is used for evaluation of generalisability, in contrast to low parameter modelling, where the entire dataset is used for fitting the model. “). I agree that the need for an independent test set is regularly neglected in low parameter modeling, but I would argue that that the principle discussed here is as relevant in that situation as it is for machine learning. Obviously with the more flexible machine learning models risks of overfitting are larger but to get a proper estimate of prediction performance, also for low parameter modelling one needs to consider either using an independent test set, or use cross-validation, or some variation on this principle. If the focus is more on model comparison, which is often the case for low parameter modelling, one might also consider approaches that penalize a standard goodness-of-fit measure for the number of parameters used by the model (e.g. such as AIC). In any case, I would argue that for “low parameter modelling” approaches generalisability is also relevant. I wonder if the applicant would agree with this, and if not, why would this not be relevant for low parameter modelling? If indeed agreed, does this mean that some of the statements above are too strongly worded w.r.t. the difference between machine learning and low parameter modelling?

A second question relates to the ideas put forward on how to combine the two modelling approaches (low parameter modelling vs machine learning). Currently the thesis states that the two paradigms can be combined in different ways, e.g. at the level of the data, or side-by-side to provide two different perspectives on the task at hand. It is stated that while some evidence is starting to appear that combined these approaches may be able to compensate for each other's weaknesses, this hypothesis is yet to be validated in large-scale experiments. I wonder how the applicant envisions contributing with his research to address this important question. Could some of the specific problems that are addressed in this thesis, be somehow particularly relevant as test-cases for such investigation on how to combine low-parameter modelling and machine learning? If not, what other directions could be taken to make progress in this area?

Conclusion

The habilitation thesis entitled [**Advanced Data Modelling for Protein Engineering**] by [Stanislav Mazurenko] fulfils requirements expected of a habilitation thesis in the field of [Protein Engineering].

Date: 23 July 2026

Signature: