

Reviewer's report on the habilitation thesis

“New methodologies of Machine-Learning modeling of complex chemical systems: mixtures, reactions and ligand-protein complexes”

by Pavlo Polishchuk, Ph.D., M.Sc.

Dr. Polishchuk's habilitation thesis is written in the form of a monograph of more than 100 pages of professional text. The work is complex, well-structured, and accompanied by numerous illustrations, which help to understand individual topics more easily. Overall, the thesis reads very well. The author cites 230 sources, 24 of which are his own publications, which clearly demonstrates how great his contribution to knowledge in the field of chemoinformatics is. Moreover, Dr. Polishchuk and his team are very active in providing ready-to-use computational tools for particular methods, and the thesis contains 12 links to GitHub repositories with computational packages and auxiliary scripts, and even other links to repositories with new test and benchmark datasets. Some of the tools are directly available online from the Elixir CZ resources. If I should point out any imperfections, I can only mention the occasional typos, some unfortunately in important abbreviations and names of molecules, and also one broken link to a repository of 3D models, which I was able to easily find anyway.

The contribution of the author to the field of chemoinformatics is incredibly wide and can be divided into five main areas that are rigorously described in the thesis. First of all, he contributed to the refinement of the simplex representation of molecular structures. This new representation can be used, among others, to model properties of chemical reactions, which is, in my eyes, very appealing for the whole field of systems biology. Secondly, he laid the foundations of the interpretation of machine learning or nonlinear models in general. Proposing novel metrics, he concluded that descriptors are more important than machine learning methods for high-quality interpretation of models. His paradigm makes all models interpretable and solves the long-lasting belief that there is a trade-off between model predictivity and interpretability, and excludes the existence of “black boxes” in QSAR models. Additionally, he proposed a new representation of 3D pharmacophores, including their new scoring technique, and he contributed to the pilot studies implementing multi-instance learning approaches. Finally, his main contribution to the field is the “Chemically reasonable mutations (CRem)” approach, useful for de novo design, i.e., model-driven generation, of new chemical structures with promising predicted properties. CRem has already been included in benchmarking studies of other authors and has become an integral part of third-party projects.

Not only has the author described new methods, approaches, and metrics theoretically, but in the last chapter of the thesis, he has demonstrated their practical use. I really appreciate this as the author was involved in drug design and medicinal chemistry projects, which were related to the development of various agents. This chapter provides an additional view on how particular techniques can be applied to infer biological knowledge, which is invaluable for readers of the thesis.

Questions for the habilitation thesis defense

I propose the following questions to be discussed during the habilitation thesis defense:

1. Could you please briefly summarize the advantages of alignment-free comparison of pharmacophores mentioned in Chapter 3? Is it expected that a consensus model outperforming individual pharmacophore models in both recall and precision could be achieved in the future?
2. Regarding the ever-increasing computing capacity of hardware resources, do you expect that 3D multi-conformer models will generally replace 2D models? If yes, how fast could the replacement be, considering other limitations of 3D models?

Overall Evaluation

In my opinion, Dr. Pavlo Polishchuk represents a very strong candidate, and his contribution to the field of modelling chemical systems is indisputable. His habilitation thesis “New methodologies of Machine-Learning modeling of complex chemical systems: mixtures, reactions and ligand-protein complexes” **fulfills all requirements expected of a habilitation thesis in the field**. Finally, **I recommend the thesis for further progress in the habilitation procedure**.

In Brno, 19th October, 2025

doc. Mgr. Ing. Karel Sedlář, Ph.D.
Head of the Bioinformatics and Systems Biology Lab
Chair of Biomedical Engineering and Bioinformatics

Department of Biomedical Engineering
Faculty of Electrical Engineering and Communication
Brno University of Technology
sedlar@vut.cz