Poznań, March 16, 2023

The review of the doctoral thesis "Prediction of ligand binding sites from protein structure" by Mgr. Radoslav Krivák.

The thesis summarizes the author's contributions to the computational methods for structure-based prediction of ligand binding sites on protein surfaces, which have implications for research into protein structure-function relationships and from a more practical perspective for rational drug discovery efforts.

The thesis is based on ten articles published at reputable conferences and journals. Mgr. Krivák was the first author of five publications and presented his significant contributions to the remaining papers, except for the two consortial publications [con19 and con21]. The integration of these two papers into the thesis gives a bit controversial impression, as the author declared minimal contribution, and the full texts are not included in the thesis. In my view, the participation of Mgr. Krivák in the European-wide network by integrating the methods developed in his thesis to PDBe-KB is a badge of the relevance and robustness of these tools, and as such, need not be diminished as a "fractional only" contribution. The candidate's participation in different roles, i.e., as a lead author in specific projects or collaborator in larger ones, demonstrates his ability to deliver impactful contributions as well as contribute to collaborative research and speaks well of his scientific maturity. Overall, there is no doubt that the research topic presented in the thesis is very relevant and timely.

The thesis is well written, giving a cohesive story about the candidate's research in binding site predictions, covering their re-ranking and identification implemented as web services, stand-alone programs, and larger bioinformatics pipelines. The developed methods exhibit very significant performance improvements over other existing tools. It is commendable that Mgr. Krivák had not stopped with publications of his methods but had exerted considerable efforts to turn them into open-sourced, user-friendly tools that have already found many users within the scientific community. Additionally, the tools were clearly well-tested for large-scale applications, making them attractive for integration into PDBe-KB. One drawback of the thesis is its, perhaps too strong, focus on describing the features of the developed methods and tools at the expense of the deeper critical interpretation of the limitations of the principles implemented.

Surprisingly, an entire chapter of the thesis is devoted to the method for search of apo-holo protein pairs covering only one of the publications, to which the candidate has declared lesser contribution. This arrangement seems to overemphasize this research direction over the studies into binding site prediction methods. The content of this chapter could have been easily integrated within the chapter about binding site prediction as a means to obtain a high-quality dataset for benchmarking and training of new methods not biased by the holo-structures inputs, which the candidate proposed in the thesis.

ul. Uniwersytetu Poznańskiego 6,
Collegium Biologicum,
61-614 Poznań
janbre@amu.edu.pl

www.ibmib.amu.edu.pl

In summary, this thesis and included papers undoubtedly prove the ability of the author to produce original and impactful scientific work, clearly fulfilling the requirements for the award of the doctoral degree.

Specific comments and questions

- In the first paragraph on page 4, the candidate introduces the definition of binding sites. Given their connections with surface features only, does it mean that binding sites located inside the protein structures were not considered in this work?

- In chapter 3.2.2., a couple of intriguing observations about the performance of predictive methods, for which I could not find the explanation in the thesis and the attached publications.

  o P2RANK performs markedly better than the PRANK rescoring of the sites found by Fpocket. Why is that? Is that connected with more sites detected by Fpocket than P2RANK or do the nature of the sites differ, affecting their scores?

  o Even more surprising is that the elementary, protrusion-based implementation of P2RANK still performs better than the majority of more complex tools. What might be the reasons behind the superiority of the P2RANK approach?

- What were the objective reasons not to publish the P2Rank-Pept method, as stated on page 23?

- In the introduction to chapter 4, the candidate states that „The generally accepted definition is that a protein in the Apo state does not bind any ligands at all". I would challenge this statement's general validity. I can imagine such situations in very extreme cases of induced-fit recognition mechanisms only. What were the examples that inspired such an interpretation?

Formal aspects of the thesis

- Despite the overall excellent English, the author had not avoided some minor language issues and typos, e.g., page 16 "that our machine learning based approach can predict that the other methods are not able to identify at all", page 18 "prtoein", page 28 "uploaed", etc.
- Perhaps this is field-specific, but the significance of the * in references might not be clear without some annotation in the bibliography section.

Jan Brezovsky, Ph.D., assoc. prof.
Head of the Laboratory of Biomolecular
Interactions and Transport

Signed by Jan Brezovský
(Certificate qualified).).
Created at 2023-03-16
17:23:02 +0100

ul. Uniwersytetu Poznańskiego 6,
Collegium Biologicum,
61-614 Poznań
janbre@amu.edu.pl