CEITEC
MASARYK UNIVERSITY

Brno 25th July 2023

To the Department of Gene Expression, Faculty of Biology, Adam Mickiewicz University in Poznań, Poland

Opponent's Report on Carlos Eduardo Sequeiros Borja's Doctoral Thesis entitled „Ligand transport pathways in proteins."

The thesis is based on four papers, three of which have already been published in peer-reviewed journals. The last paper is deposited on biorXiv and is expected to go through a review process. The thesis starts with a short introduction to protein function and the importance of protein tunnels. The main body of the thesis is divided into two chapters, followed by conclusions, references, and attached papers.

The first main chapter is focused on the methodology of tunnel calculations. It begins with the history of tunnel evaluation together with the necessary definitions. Subsequently, two main approaches for tunnel calculations are explained and the differences between them are highlighted. This introduction explains the need for a new tool, which was developed by the candidate. This tool, called TransportTools, combines both approaches in a single program and is the subject of the first paper (Annex 1).

TransportTools is a Python library designed to facilitate the analysis of transport tunnels and ligand migration events in proteins. The library integrates outputs from molecular dynamics simulations and analysis tools of CAVER and AQUA-DUCT. The program analyzes ensembles of tunnels and cluster them together. Ligand transport events are then assigned to the clusters. The performance of TransportTools was tested on three datasets of 50 simulations and processed within a reasonable time frame of one day on a standard workstation. The tool inherits the limitations of CAVER and AQUA-DUCT. The article provides a few use cases that demonstrate the applicability of TransportTools in analyzing ligand transport. Overall, TransportTools provides a freely available tool for the analysis of transport pathways.

The second paper of the first chapter (Annex 2) is focused on effective use of the tool on the simulated trajectory of a protein. The authors proposed a divide-and-conquer approach, which is based on slicing the MD trajectory into smaller parts and analyzing the tunnels in each part separately. This approach significantly

CEITEC – Central European Institute of Technology
Masaryk University
Kamenice 753/5, 625 00 Brno, Czech Republic
www.ceitec.eu

UNIVERSITAS MASARYKIANA BRUNENSIS

reduced the runtime and hardware resources required for tunnel analysis. The TransportTools library was used to merge the results from the smaller parts of the trajectory and provided an overall view of the tunnel network for the complete trajectory. The approach is validated using a 100 ns MD simulation of haloalkane dehalogenase DhaA, which showed equivalent outcomes to the analysis of the full trajectory. The proposed approach makes the tunnel analysis of long MD trajectories more accessible to researchers with limited computational resources.

I appreciate the provided tutorial/guided example, which goes step by step through the tool usage and would be very helpful for scientists starting with this tool. However, I would appreciate additional information for non-experts on how many parts the trajectory can be split into, or even how to recognize that the trajectory has been split too much, and the proposed approach is failing.

I was a bit confused by the employed protocol for the simulation: "Next, another 1 ns of MD in the NPT ensemble with the same setting was performed without any restraints. Final equilibration step constituted a 200 ns simulation with the same settings as the previous step. This was followed by 100 ns of unrestrained production simulation in the canonical NVT ensemble, with constant temperature 310 K and saving frequency of 10 ps." Why was the production run performed in the NVT ensemble instead of the NPT ensemble?

The second chapter of the thesis focuses on two applications of the developed tools, which highlight the need to move beyond the analysis of single structures to the dynamic nature of proteins.

The first application (Annex 3) is to a plant protein called MtABCG46, which belongs to a family of ABCG transporters. The researchers started with an in silico generated structure using the AlphaFold2 program. Molecular dynamics was subsequently employed to obtain a thermodynamic ensemble of structures. The access paths to the central cavity/active site were analyzed. One pathway/tunnel provided the most direct access to the central cavity from the inside of the cell. This tunnel was only transient and unusually narrow. After the amino acid sequence analysis of various ABCG transporters, residue F562 was identified as potentially important for the ligand selectivity of MtABCG46. Various mutants at this position were used to experimentally verify the selectivity effect of F562. The authors then used computer simulations to investigate the structural changes in these mutations. They found that the mutations disrupted non-covalent interactions between protein helices, resulting in the changes in their bending and twisting. Such changes altered the availability of the access tunnels ,further supporting the transport through the identified narrow tunnel. The free energy to open the protein was further calculated and mutants showed different free energy barriers. The

study provides insight into the structure and function of MtABCG46 and its role in the transport of plant compounds.

In this study, I missed information on how the barriers to open MtABCG46 protein and its mutants were calculated from the mutants and what is the estimated error of the free energy profiles. Such information could enhance the significance of the obtained barrier increases of 1.2, 9.4, and 11.6 kcal/mol for the F562L, F562Y, and F562A mutants, respectively.

The last paper of the thesis (Annex 4) is aimed at water transport through tunnels in α/β-hydrolases. Three different hydrolase enzymes (haloalkane dehalogenase, epoxide hydrolase, and lipase) were simulated using molecular dynamics and analyzed with a focus on water in tunnels. Water molecules were observed to traverse narrow tunnels with sub-angstrom bottlenecks. Such finding contradicts the commonly accepted view that the radius of the tunnel needs to be grater then the radius of the water molecules (about 1.4 Å). Analysis of transport events in such narrow tunnels showed a significantly increased number of hydrogen bonds formed between the water molecules and the protein, likely compensating for the steric energy penalty. Such narrow tunnels accounted for approximately 20% of the total water transport observed, highlighting the need to move beyond the standard geometric limits of functional tunnels. In addition, the differences between and the E470G mutant of the human epoxide hydrolase enzyme showed the mutation effect on the water transport tunnel and suggested the relation of water transport with a higher incidence of ischemic stroke. Despite the lack of sufficient data to support extrapolation outside the α/β- hydrolase fold, the study indicates the need to rethink the classical limits of functional tunnels in enzymes and suggests that narrow tunnels may play a significant role in the transport.

I would like to know how was the lengths and number of individual MD simulations of hydrolases selected? In other words, was the efficiency of running multiple simulations in parallel rather than one long simulation tested?

The thesis is straightforward to read. However, it would be even easier if the author would avoid very long and complicated sentences. There are numerous freely available tools that could be used to improve texts now, so it is a pity that they were not employed. I believe that the thesis would also benefit from a more detailed description of the methods/tools used for tunnel calculations and simulations, but I may be biased by the common format of Czech theses.

Overall, the candidate has demonstrated the quality of his research by being the first author on three provided publications, on one of which he is in a shared

CEITEC
MASARYK UNIVERSITY

position due to the collaborative nature of the work. The work clearly represents a substantial original contribution to the field of protein research and has advanced our understanding of the protein tunnels. Therefore, my overall assessment of the thesis is positive. I recommend the thesis for the defense and the award of a distinction and the candidate for the final stages towards the award of the doctoral degree.

Robert Vácha
Group leader/Professor
CEITEC and Faculty of Science
Masaryk University, Czech Republic
Phone: +420 549 496 846
Email: robert.vacha@muni.cz
www: vacha.ceitec.cz