CEITEC
MASARYK UNIVERSITY

Brno 6th December 2024

To the Department of Gene Expression, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University in Poznań, Poland

Opponent's Report on M.Sc. Nishita Mandal's Doctoral Thesis entitled „Insights into Molecular Mechanism behind Rare Transport Processes in Enzymes with Buried Active Sites."

The thesis is based on three papers, two of which have already been published in well recognized peer-reviewed journals – Bioinformatics and Journal of Chemical Information and Modeling. The last paper is deposited on biorXiv and is expected to go through a review process soon. The candidate is the first author on two of those papers.

The thesis starts with a brief introduction of enzymes with highlighted importance of tunnels leading to the active sites buried deep in the structure. Importance of tunnel dynamics is explained and dynamical opening and closing of new tunnels and gates is explained together with consequences regulating the active site access. Molecular dynamics simulations are highlighted as ideal tool to study such changes in the protein structure, when enhanced sampling techniques are employed. Subsequently, haloalkane dehalogenase LinB is introduced and the candidate outline her work.

Second chapter summarizes the candidate results with three sections each capturing content of one of her papers. The first paper is mainly co-authored by colleagues from Laboratory of Biomolecular Interaction and Transport and the candidate is on the fifth position. Her contribution is mainly testing of the developed tool to LinB protein and its mutants, using which he found that mutations at p1 and p3 tunnels in LinB, were crucial for understanding the behavior and biological functions of tunnels. The manuscript describes TransportTools software (python library), which was created to analyze and unify large datasets (ensambles) of tunnels from simulations. The tool integrates tunnel data from CAVER with ligand tracking data from AQUADUCT, allowing efficient comparisons between protein variants. The results highlighted the ability of TransportTools to differentiate between main and auxiliary tunnels via clustering, as well as to track water transport through them. The tool was made freely available enabligh for all researchers to handle ale amounts of simulations and use them for detailed

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

UNIVERSITAS MASARYKIANA BRUNENSIS

comparative analysis. These features makes it versatile for studying membrane transport, enzyme catalysis, and tunnel dynamics across different simulation methods.

The second section is focused on Gaussian accelerated molecular dynamics (GaMD) simulations used to explore the tunnel network in enzyme haloalkane dehalogenase LinB. Three version of the protein were investigated LinB-Wt, LinB-Closed and LinB-Open. Every system was simulated with five different replicas each 5 μs long. For comparison, the same systems were simulated for the same time also using standard molecular dynamics simulations. TransportTools were employed to analyze the tunnels.

New tunnel was discovered. It was rarely observed in classical molecular dynamics, while it was readily found in GaMD. The agreement between the two methods supports the finding and suggest that it is not an artifact and higher occurrence in GaMD can be attributed to its enhanced sampling. The tunnel was named the side tunnel based on its emergence at the side helix region. Importantly, the residue at the mouth of this tunnel was experimentally found to significantly affect the enzyme activity. In particular, A189F mutation increased the activity by 21.4% for 1-chlorohexane and 26.2% for 1-bromocyclohexane. The transport of small molecules needed for enzymatic reaction was evaluated through the tunnels using CaverDock software. The side tunnel had lowest energy barrier for transport of water compared to other tunnels. In addition, the role of L176 residue was investigated as it is involved in the side tunnel and it is mutated in LinB-Closed and LinB-Open. It was found that mutants have additional hydrogen bond which need to break fo tunnel opening making it more difficult for both mutants.

Finally, the previously known auxiliary tunnel, p3, is mutated in the LinB-Open. The de novo engineered mutation in this tunnel resulted in the highest frequency of tunnel opening with respect to other LinB variants. Comparative analysis of p3 opening percentages in all studied versions of LinB revealed increased sampling of this transient tunnel in GaMD compared to classical molecular dynamics highlighting again the advantageous enhanced sampling og GaMD.

In the third study, two coarse grained models were employed to study the dynamics of LinB. The employed models Martini 3 and SIRAH are briefly introduced in the section including the employed structural restrains that were used to stabilize the proper tertiary structure. However, the selection of restrains parameters is not described or discussed. It is also unclear why simulations were limited to only 5 μs. Apart from LinB, at least three proteins from each enzyme commission class were investigated each for 2 μs. Despite the shortness of the simulations, it was possible to identify functional known and transient tunnels.

However, many other tunnels were also found, probably due the coarse grained nature of the beads. It would be interesting to know if these new tunnels would remain after the backmaping to all-atom structures. It was found that elastic restrains overstabilized the structures compared to the Go approach/model and SIRAH as was probably anticipated. Go model was identified as faster and exploring broader conformational space compared to other studied models, making it ideal for studying large protein systems.

In general, the thesis is well written. I noticed a few typos and imprecise expressions. The flow of the text could be improved, and the descriptions of the figures could be more detailed - for example, in Figure 3 it would be good to explicitly mention that proteins with secondary structure are shown in cartoon style and are coloured grey. If allowed, it would be beneficial to use one of the many AI tools now available that can correct these in seconds.

Regarding the organisation of the thesis, I was a bit surprised to see the description of the Caverdock tool, Principal Component Analysis (PCA), Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) in the results section. I was expecting them in the methods section, but I am probably biased by the common dissertation format in the Czech Republic.

I would like to know if Figure 2B is the illustrative plot or if it is based on real data. If it is real data, could you please clarify how it was obtained and whether the graph might be system dependent? In other words, how do you expect the graph to look for different systems?

Could the candidate compare how much the sampling of tunnel dynamics is improved using Gaussian Accelerated Molecular Dynamics (GaMD) compared to Classical Molecular Dynamics (cMD)? Is it necessary to reweight the protein configurations for tunnel analysis by CaverDock after running GaMD (page 22)?

On page 23 it is stated: "The first two principal components captured at least 80% variance suggesting that they are enough to understand the system." I would welcome more explanation on how the percentage of captured variance is related to the system understanding.

Have the candidate tested other methods to generate proteins tunnels? For example Alphafold or RosettaFold where different models or less propapble structured would be considered?

Overall, the candidate has demonstrated the quality of her research by being the first author on two provided publications. 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
Professor and Group leader
CEITEC and Faculty of Science
Masaryk University, Czech Republic
Phone: +420 549 496 846
Email: robert.vacha@muni.cz
www: vacha.ceitec.cz