



The Board of Doctoral Study

Faculty of Biology
Adam Mickiewicz University, Poznań

Prague, 31. 7. 2023

Review of the doctoral thesis “Ligand transport pathways in proteins” by Carlos Eduardo Sequeiros Borja

General notes

The presented doctoral thesis deals with the world of protein tunnels and the transport of small molecules within them, exploring the dynamic aspects of these essential biological processes. Throughout the research, novel methodologies, including the development of TransportTools, were designed to efficiently analyze and interpret large-scale molecular dynamics (MD) data. The goal of the research presented in the thesis was to gain deeper insights into the behavior of transport within proteins, paving the way for a more comprehensive understanding of their functional significance. The text of the thesis is split into two main parts: method development and application of the developed methodology.

The first chapter sets the stage by providing an overview of the field of tunnel analysis, providing a historical context and evolution of tunnel analysis, and tracing its origins from the early study of static protein surfaces to the current dynamic view, where a multitude of protein conformations are analyzed. The chapter also focuses on the classification of various cavities present in proteins shedding light on their diversity, ranging from pockets and groves to channels, and on methods to identify them. The chapter's main part introduces TransportTools, a methodology and software tool that facilitates tunnel analysis in proteins. TransportTools combines the outputs of CAVER and AQUADUCT, streamlining the

comparative analysis of tunnel geometries and their usage during MD simulations. The method reduces subjectivity and enhances reproducibility, providing researchers with a more robust framework to study tunnels. To speed up the computation, it presents a divide-and-conquer approach to effectively analyze long MD simulations by slicing them into manageable pieces.

Chapter two focuses on the application of the methodology described in section 1 on two use cases: i) selective transport in plant ABCG transporters and ii) water transport in enzymes. First, the TransportTools methodology was used to study the ATP-binding cassette transporter ABCG46 from *Medicago truncatula*. MD simulations and subsequent tunnel analyses revealed the structural determinants for the selective transport of specific compounds through the protein's transmembrane domain. Mutational studies and umbrella sampling simulations shed light on the impact of mutations on tunnel behavior. Second, TransportTools were used to study water transport through tunnels in enzymes. This was carried out by investigating water transport in three enzymes. It has been discovered that water molecules could traverse tunnels narrower than their own radius, revealing the significance of narrow tunnels in enzyme function.

The thesis is well structured and is, in general, relatively easy to follow. However, what hinders reading of the thesis are the fairly frequent occurrences of odd wordings and formulations, and typos (see the Text comments section at the end of the review). As a computer scientist with a background in mathematics, I would welcome a bit more formalism in describing the problem and also in describing the performance improvements. For example, it would be helpful to specify what the runtime of previous and new (divide and conquer) solutions was in terms of algorithmic time complexity. Then, a claim such as "the time required for the analysis of potential tunnels using a very small probe with a radius of merely 0.7 Å was reduced considerably by the usage of the divide-and-conquer method." would stand on a firmer footing.

The amount of computer science work related to the thesis is either substantial or minor, depending on how much the candidate contributed to the development of TransportTools. This is not obvious as the candidate is not the first or corresponding author on the respective publication, and his contribution is not visible on the project GitHub repo either (see the Comments section). If his contribution mainly rests in the development of the divide and conquer approach, then the amount of software work is rather minor. However, this would be anyway balanced by the amount of experimental work (as in in-silico experimental) demonstrated in section 2. This part also proves the candidate's ability to cooperate with domain specialists, who are the main consumers of the developed software.

The application of the tool to the study of water transport through tunnels in enzymes even challenged the established single-file transport paradigm by showing that water molecules could traverse tunnels narrower than their own radius. I find this extremely interesting. Sadly, this is the only publication that has not been peer-reviewed, and since I am not an expert on MD simulations, it's hard for me to accept this fact without any reservations. Based on my scant exposure to MD, its results are extremely sensitive to parametrization, so I am wondering if the outcome cannot be attributed to the experimental settings. I did not see this being discussed in the publication.

The contributions are backed up by 4 publications, one being a bioRxiv preprint, and two of the three remaining journal publications were published in high-impact journals. However, it should be noted that the candidate is also the main author or co-author of four other impacted journal publications. This, together with the fact that the work of the candidate has already been cited about 70 times (based on Google Scholar) indicates that the candidate is starting to establish himself as a valuable member of the bioinformatics community.

Overall Evaluation

The thesis shows that the candidate has deep knowledge of the scientific domain and is capable of independent creative scientific work, and, despite the critical comments, I recommend the thesis for defense. I believe the PhD candidate can proceed to the final stages towards the award of the doctoral degree.

Comments/Questions

1. What is the contribution of the candidate towards the development of TransportTools? All the commits in the commit history on GitHub were carried out by the group leader, Dr. Brezovsky.
2. I do not fully agree with the statement "*For proteins with buried active sites, ligands are required to traverse internal protein cavities that connect the active site with the external environment*" as buried ligands can be the effect of conformational change of the site itself (cryptic binding sites).
3. On pg18 the candidate states that "*I realized that another limitation of the current approach to evaluating long MD simulations is the computational resources required to perform this task with CAVER*" – how is the runtime of CAVER distributed. What takes the longest time? Why does splitting the computation into parts result in speedup?
4. On pg19 the candidate states that "*this approach allows for a more refined search of tunnels, granting visibility to previously important undetected narrow tunnels*". How exactly? I find this quite interesting that by splitting one can achieve results which

could not be obtained before. This raises the question of equivalency of the results obtained with and without splitting. Is it possible that some tunnels won't be obtained when the splitting procedure is applied?

5. One of the most interesting results (or at least I find it interesting) is the one showing that tunnels used by water can be quite narrow. Could the candidate elaborate on the exact mechanism?
6. In the work on ABCG transporter, the candidate gave structural explanation of a phenomena observed in nature (mutation of residue leading to the change in selectivity of the transport). However, the problem with these sorts of experiments is that one knows the results in advance. Does the author consider in the future using mutation studies to predict the effect and only then validating it by experiment? I would find that a true unbiased validation of them methodology.

Text comments

- pg10 - *"Is at the active site that the chemical reactions occur, hence,..."* – does not make sense
- pg10 - *"I decided to study the importance and methods to study protein cavities and pathways"*
- pg10 - *"This dynamical view brought together, not only new insights and more information but also..."* – extra comma
- pg13 - *"...identify "manually" the existing ..."* -> "manually" identify
- pg15 - *"considerably scarce compared"* - considerably scarser
- pg16 - *"well accepted for the community"* -> by the community
- pg16 - *"alleviated in some grade by another research project"* – to some extent by
- Fig3 - *"The workflow starts by employing a set of molecular dynamics trajectories as inputs, then a) the tunnels (colored surfaces) are calculated with CAVER, b) and the transport of small molecules (red and white spheres) with AQUA-DUCT."* – the last sentence is missing a verb
- ...

Typos

- pg 12 - *"by rejected by the majority"* -> be
- pg14 - *"relays"* -> relies



doc. RNDr. David Hoksza, Ph.D.