FACULTY OF
CHEMICAL ENGINEERING
UCT PRAGUE

doc. RNDr. Michal H. Kolář, Ph.D.
Department of Physical Chemistry
UCT Prague, Technická 5, 16628 Prague
kolarh@vscht.cz

PhD thesis review

Thesis title Computational modeling of biomolecular interactions in proteins using MD simulations

Thesis author Zara Zeenat

Supervisor Dr. David Řeha, Ph.D.

The thesis is written in English, so I decided to write this review in English as well.

Zara Zeenat's doctoral thesis summarizes her work on two research projects with rather diverse focuses: the first project deals with an enzyme PETase, which catalyzes degradation of polyethylene terephthalate (PET), while the second project examines a G protein-coupled receptor (GPCR) in complex with its biomolecular ligand. The PETase part represents roughly 40% of the results, the GPCR the remaining 60%. On the other hand, the introduction of the PETase project is more than twice as lengthy as the introduction of the GPCR project. The unifying element of the thesis is the methodology, which encompasses a wide array of computational techniques, with a strong emphasis on molecular dynamics (MD) simulations. These methods are applied to investigate intricate biomolecular systems.

I appreciate Mrs. Zeenat's effort invested into the two research projects. The thesis undoubtedly yielded insights into the association of GPCRs with tetrameric G proteins and the interaction between PETase and its substrate in non-aqueous solvents. However, despite these achievements, I have found specific concerns throughout the thesis, which I describe in detail below. Unfortunately, I am unable to recommend the thesis for further consideration at this time. I believe that with additional work and refinement, there is a potential to produce a high-quality doctoral thesis.

Major concerns, organized in the order they appear in the thesis, are as follows:

Scientific Questions The scientific questions in the Introduction section are too superficial. I would strongly encourage including a dedicated subsection outlining the aims and objectives of the thesis, which is now missing. These questions can also enhance the Conclusion section by providing clear guidance to both the author and readers regarding the anticipated answers.

PETase Methods The methodology employed in the PETase project lacks sufficient detail and requires enhancement. The parametrization of ionic liquids, which is a critical part of the methodology, is unclear. In the thesis and accompanied Paper 1 it is stated that the parametrization originates from GAFF, yet both references 244 in the thesis and 41 in the Paper 1 point to the Lipid 14 AMBER force field. Furthermore, there is mention of validation against experimental thermodynamic data in Ref. 245 [Sprenger et al.], but it is unclear whether this validation was applied to the ionic liquids used in the PETase project.

Consequently, it remains uncertain, whether the published force field for ionic liquids was adopted in the PETase project, or the procedure was repeated for the ionic liquids studied. The essential trick of the published procedure, a factor of 0.8 used to scale electrostatic interactions between the ion pairs, is not mentioned in the thesis nor the Refs 245–247. Moreover, I have noticed methodology differences between validations in Ref. 245, and PETase simulations. For instance, the cut-off radii for PME and short-range electrostatics were 1.4 nm in Ref. 245, but 1.0 nm in the thesis. Since the electrostatics treatment may notably affect the ionic liquids, these differences must be justified in the text, or possibly further validated.

Additionally, there is a lack of information regarding the force fields used for the enzyme and water, which could compromise the reliability of the results. The use of GAFF for protein simulations is discouraged and may lead to incorrect outcomes.

Finally, the rationale for employing a large cubic box (a=12 nm) when dealing with a relatively small and spherical protein like PETase is unclear. My own calculations suggest that a truncated dodecahedron box of approximately 8 nm is sufficient to prevent interactions with periodic images of the enzyme. Was the author aiming for a specific PETase concentration, or was the large box necessary for accommodating a certain number of ion pairs?

Reliance on Single Trajectories The PETase project draws conclusions based on evidence from single trajectories, which is a potentially problematic practice, as discussed by Gapsys and de Groot [eLife 2019]. This approach may lead to incorrect conclusions, and the absence of error analysis renders the results inconclusive. Implementing error analysis, as exemplified by Grossfield et al. [LifeCoMS 2019] even at the level of single trajectories, is crucial to strengthen the student's findings.

To demonstrate the issue, I performed MD simulations of the PETase (PDB 5XJH) in pure water, namely four independent trajectories differing in the initial set of randomly generated velocities. Each trajectory was 150 ns long alike to the thesis and underwent a similar equilibration setup. Using CHARMM36m force field for

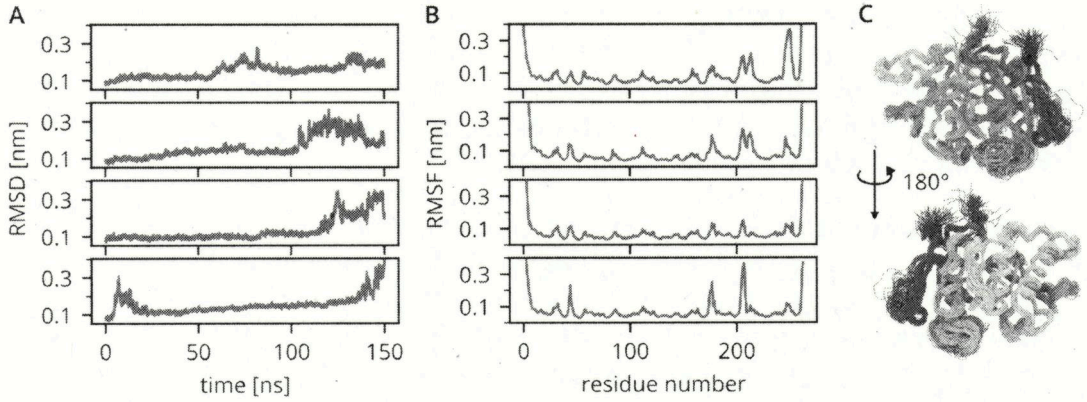

Figure 1: Structural characterization of PETase. Four independent trajectories generated by MD simulations using CHARMM36m force field and explicit TIP3P water are shown. A) Root-mean-square deviation (RMSD) of C_α atoms as a function of simulation time. B) Root-mean-square fluctuations of C_α atoms. C) 100 snapshots equally separated in time from the trajectory 1 and superimposed by C_α atoms. Color coded from the N-terminus in red through white to C-terminus in blue. The two views are rotated around the vertical axis by 180°.

the protein and TIP3P water, I observed notable variations among the trajectories in both RMSD and RMSF (Fig. 1). At the molecular level, the variations are located on the protein termini and around residues 180, and 205. In the magnitude, the variations are comparable to the variations among ionic liquids or their concentrations, as presented in the thesis. Hence, a more rigorous statistical analysis of the results is needed in the thesis to better support the conclusions. My statement holds no matter whether the results have been already published in a peer-reviewed journal or not.

Conformational Ensemble Representation In Figure 10, multiple PETase structures are depicted. From the accompanying text, it appears that these snapshots represent the final states from MD simulations in various solvents. A similar practice seems to be employed in the GPCR project (Figures 20–24). If true, I have doubts about how these data effectively represent the conformational ensembles of the enzyme or the GPCR–G protein complex.

Additional comments:

Methods Section The methodology section comprises three main parts detailing homology modeling, MD simulations, and analysis. The homology modeling is adequately described, one can reproduce the workflow if needed. It provides readers

with insight into the complexity of studying a system involving a conformationally diverse transmembrane protein (GPCR) and ligand (G protein). This feeling is implicit and it is pity that was not expressed more literally. The MD subsection is extended. It could have been more concise by referencing appropriate textbooks and focusing on system-specific setup details.

Formatting and Proofreading The overall formatting of the thesis requires improvement. I recommend including page numbers. I found a large number of typographic errors, and grammar issues with a varying level of seriousness. Dr. Bondar is misspelled in the Acknowledgment section, chemical formulas in text lack proper subscript/superscript formatting, there are typographic errors in Figures 2 and 21A, both *Lewis* and *lewis* are used inconsistently, formatting of Eq. 6 is wrong, to name just a few of my concerns. Careful proofreading should certainly improve the overall quality of the thesis.

References The use of references should be reviewed and improved. For instance, the references have ambiguous formatting in the text including styles like "[41–43]," and "[180][181, 182]." Reference 212 on page 46 is wrong. The algorithms of temperature and pressure coupling in section 2.2.7 are not cited, although some of the citations appear later in the text.

I hope this feedback sounds constructive in improving the thesis and possibly offer a learning opportunity to Mrs. Zeenat. I remain available if Mrs. Zeenat would like to discuss the feedback in more detail or seek guidance on how to address these concerns.

Overall statement in Czech: **Nedoporučuji k obhajobě**.

Michal H. Kolář

Prague, September 29, 2023