MUNI SCI National Centre for Biomolecular Research

Review of Doctoral Dissertation

Title: Molecular modeling of structure-dynamics-function relationships in proteins

Candidate: Bartłomiej Surpeta, Adam Mickiewicz University, Poznań, Poland

Bartłomiej Surpeta presents a doctoral thesis summarising his pioneering scientific achievements from applying molecular simulations to understand and design novel quorum quenching (QQ) enzymes.

The QQ enzymes disrupt communication between bacterial cells and prevent gene expression responsible for bacterial virulence factors, biofilm formation, swarming and resistance. These enzymes have potential uses in many fields, especially medicine, where bacterial resistance becomes a serious threat. Since QQ enzymes act extracellularly, it is hypothesized that their action will put smaller selective pressure, and thus, they would escape from common resistance mechanisms. Moreover, QQ enzymes are biodegradable molecules and less harmful to the environment. These properties make them suitable for a wide range of biotechnological, agricultural, and industrial applications. Especially those which can benefit from the disruption of biofilm formation. From this perspective, the studied systems are highly relevant and have high socio-economic potential.

Quorum sensing is triggered by N-acyl homoserine lactones, which are a target of QQ enzymes. To study the degradation of the signalling molecules, the candidate combined several modelling approaches, which range from applied bioinformatics to mainly molecular modelling employing unbiased and biased molecular dynamics simulations. He also studied the reaction mechanisms, which is generally a very challenging task. He employed a hybrid QM/MM approach and combined it with the calculations of the activation free energy barriers utilizing steered molecular dynamics simulations. The entire process comprises several steps, including the entrance of the signalling molecule into the active site, its chemical transformation, and the release of the products to recover the enzyme function. Since a single static structure is not usually enough to describe such a complex process reliably, he studied the ensemble of structures generated by MD simulations. The shift from a single structure, typically represented by an experimental structure, to an ensemble of structures is a novel approach becoming possible only recently due to the increased computational power of supercomputers. Thus, the candidate fulfils the doctoral study's main topic, structure-dynamics-function relationships in proteins. The achievement by the candidate brings significant novelty from the perspective of QQ enzymes and, in general, for molecular modelling of dynamic and complex biomolecular systems, even though not all aspects of the enzymatic function mentioned above were addressed in the case of studied QQ enzymes.

The candidate published the obtained results in three papers in impacted journals (ACS Catalysis, International Journal of Molecular Sciences, Bioinformatics), and one he submitted as a reprint to bioRxiv. The candidate's contribution to all presented papers is significant: in two articles, the candidate is a first author and a shared first author in the other one. In addition, he contributed to two other papers (Briefings in Bioinformatics and MethodsX), which are not included in the thesis.

Altogether, 6 published papers in about five years long doctoral study (according to the candidate's record on LinkedIn) is a very good achievement.

The thesis is written in English except for the Polish abstract and other bibliographical parts. The main text consists of an introduction to structure-dynamics-function relationships in proteins (4 pages), a summary of the doctoral research (13 pages), conclusions and future perspectives (2 pages), and references (5 pages, 78 items). Then, the four papers are included but without supplementary materials, followed by co-author statements.

Even though the thesis and articles were carefully prepared, several errors escaped attention and appeared in the final version. These are an incorrect abbreviation for the hydrogen-deuterium exchange mass spectroscopy (HDX-MD), which should be HDX-MS (page 10); corrupted diacritics in the one author's name for reference 78 (page 34), incorrect spelling of a force constant unit for positional restraints (kcal/mol^{-1} Å^{-2}) in the fourth article, which should be kcal mol^{-1} A^{-2} or kcal/mol/Å^{2}. The statement, "The fastest motions are related to the vibrations of covalent bonds, which occur within femtoseconds to picoseconds and are mostly in the amplitude of one angstrom." (page 12), is incorrect. The amplitude is approximately one hundred times smaller. Perhaps, the accessed time for reference 33 (page 31) should be skipped.

The main thesis text (pages 11 to 29) nicely summarizes the structure-dynamics-function relationships in proteins and aligns it with four published papers. While the length and provided information are sufficient to dive into the problems, providing a pictorial representation of each paper's key achievements would be beneficial. Also, the real medical and industrial applications of the QQ enzymes were not fully clear until the very nice summary in the introductory part of the fourth article. Lastly, despite the main topic being the dynamics of the proteins, the rate-limiting step is the reaction itself. Therefore, additional information about modelling enzymatic reactions and calculating activation barriers would also be beneficial. Altogether, these are only minor comments and do not impact this section's quality.

Since I have many questions, I attached them separately to this review alongside additional comments.

In conclusion, Bartłomiej Surpeta demonstrates that he can independently and successfully pursue modern academic research employing molecular modelling of biomolecular systems. The published work represents an original solution to scientific problems. The doctoral thesis of Bartłomiej Surpeta meets the requirements for doctoral theses in biological sciences. Therefore, I unequivocally recommend this excellent scientific work for the defence with high distinction.

In Brno, 4th August 2023

.....

RNDr. Petr Kulhánek, Ph.D.

National Centre for Biomolecular Research
Masaryk University
Kamenice 5
625 00 Brno
Czech Republic

Additional comments and questions:

The thesis contains a nice **introduction**, but some interesting points related to the QQ enzymes and employed simulation methodology were not thoroughly discussed.

1. It is unclear if N-acyl homoserine lactones are the only one found kind of signalling molecules involved in quorum sensing in the bacterial world. Can you comment on the uniqueness of these molecules and, thus generality of their degradation from the perspective of quorum quenching?
2. Since the studied QQ enzymes are of natural origin, it is not evident from the thesis text what is their native role. Do bacteria produce them to self-regulate the quorum-sensing process? Or are they produced by bacteria to suppress rival bacteria?
3. Follow-up question. Since natural enzymes can degrade the N-acyl homoserine lactones, are those selected during the doctoral study the best ones? What would be the key selection criteria to pick the best candidate?
4. Why did the candidate select steered MD for the free energy calculations? Was it simpler than other methods such as metadynamics, umbrella sampling, or constrained dynamics (blue moon sampling)?

The **first paper** concerns three QQ enzymes: ecPGA, aPGA, and paPvdQ. Although the paper underwent a review, I will criticize some parts. Firstly, it was extremely difficult for me to decipher the definition of the collective variables because two similar shades of blue were used for nitrogen and carbon atoms (Figure 2, page 6361). Perhaps normal structural formulas and reaction schemes would be more appropriate.

5. What is pKa of the terminal amino group in Ser1β? Is it possible to assume that the amino group of the catalytic serine is neutral in the initial state?
6. The calculated barriers (Figures 5 and 6, pages 6364 and 6366) are significantly underestimated compared to the experimental data. When I read the article and supporting materials, I thought of the following possible causes. Can the candidate critically evaluate each of them with supporting data perhaps not provided in the paper?
    a. The steered MD calculates the non-equilibrium work applied along the collective variable. Since the CV for the TI-formation does not include all reacting atoms (the distance between the serine oxygen and lactone carbonyl carbon atoms is not included), the calculated work could be lower. Thus, the free energy barrier could be lower than it should be.
    b. The initial state for TI formation (RC=1.1 Å) does not seem to represent a free energy minimum. Thus, this state is higher in the free energy, and consequently, the barrier is lower.

The **second paper** is a very nice review of protein dynamics and its impact on protein design and engineering.

7. With the overall knowledge provided in the review, do you think that substrate (bulky head with long aliphatic chain) transport in the case of the studied QQ enzymes can be done reliably on trajectories of the enzymes without the substrate? Would the steered MD on the substrate/enzyme complexes be a better alternative to characterize the entrances to the active site?

The candidate did not contribute to the code presented in the **third paper**, according to the records on Git Hub and the co-author statement. Nevertheless, he participated in the testing and writing the tutorial, which are unequivocally important contributions.

8. With the amount of data produced by MD simulations on GPU accelerators, do you think the analysis by TransportTools would benefit from parallelization? If yes, how difficult would it be to implement such a feature?

The **fourth paper** is a preprint about the rational engineering of the binding pocket of ecPGA.

9. Was the manuscript sent to a regular journal? If yes, what is its current status?

10. With recent advances in protein structure prediction employing artificial intelligence (AlphaFold, ESM), would it be beneficial in your case to employ the inverse folding technique to generate suitable mutants for further screening of reactivity?

11. In the concluding remarks to the fourth paper (page 26), you wrote that steered QM/MM MDs are performed on additional substrates and native as well as mutated enzymes. What is the current status of these simulations?