Evaluation report on Ph.D. thesis

Evaluator: doc. Ing. Vojtěch Spiwok, Ph.D.

Title: Inter and Intra Domain Interactions in the Motor Domain of EcoR124I: A Computational Study

Applicant: Dhiraj Sinha, M.Sc.

The thesis of *Dhiraj Sinha, M.Sc.* entitled *Inter and Intra Domain Interactions in the Motor Domain of Eco*R124I: *A Computational Study* presents interesting applications of homology modelling, molecular dynamics simulations, quantum mechanics and other techniques on an interesting DNA-modifying system with an important role in prokaryotic defense against foreign nucleic acids. This is a very hot topic owing to recent discoveries and applications of the related CRISPR/Cas9 system. The results of the thesis contributed to publication of three articles published in the time of thesis submission, one article accepted for publication and two manuscripts in preparation. The thesis presents 125 pages of Introduction, Methods, Results and Discussion, together with first pages of articles and submitted manuscripts.

The Introduction presents the topic of prokaryotic immune system with focus on restriction-modification systems and *Eco*R124I. This is followed by Methods that present introduction to biomolecular simulations and quantum chemistry (without going to computational details). Here I would mention a minor stylistic imperfection of referencing to titles of some chapters or sub-chapters in their initial sentences. It is always better to repeat the title of a chapter in the first sentence instead of starting the a chapter with "This is ...". The author is also inconsistent in using italic for mathematical variables.

Chapter 3.1 presents the results of molecular simulations aimed at deciphering the role of inter-subunit contacts in the ATP binding site. The key mutant studied in this chapter is K220A (or referred to as Lys220Ala in the thesis). Maybe I missed some part but I did not find information about the effect of this mutation on activities of *Eco*R124I or at the cellular level. Is this information available in literature?

I like very much the title of Chapter 3.2. It is always good to call a chapter or article "Protein ... does ....", instead of "Study of ...". The author used principal component analysis (PCA) on simulation trajectories in chapters 3.2, 3.3 and 3.4. As far as I understood, trajectories of wild-type and mutants were concatenated to make one "supertrajectory". This "supertrajectory" was analysed by PCA to obtain collective motions. Finally, all trajectories were projected on collective motions to obtain their low-dimensional embeddings. As an alternative I can imagine analysing just one trajectory, for example the wild-type enzyme, by PCA, followed by projection of individual trajectories on resulting collective motions. I would ask the applicant to explain (and confirm) why he used the "supertrajectory" approach and to discuss its advantages and disadvantages.

Chapter 3.5 presents prediction of 3D structure of the *HsdR* subunit. It seems that the alignment in Figure 3.5.1 is not aligned, it presents amino acid sequences simply written without the alignment step. If yes, it is clear from other results that proper sequence alignments were used for building of models. Nevertheless, I believe that the author should provide a set of sequence alignments that were used in modelling of the studied system. The resulting models were evaluated by molecular dynamics simulation. The idea behind evaluation of protein models by molecular dynamics simulations is that a good model is likely to be stable and bad model is likely to be unstable. In principle, a model can be improved (refined) by simulation. I would ask the applicant to present his opinion on this approach.

In conclusion, the thesis presents a series of interesting molecular modelling studies on bacterial complex *Eco*R124I. It is well written and its published results will inspire experimentalists. Minor imperfections such as Figure 3.5.1 or some typesetting errors (missing spaces in and between sentences) do not reduce good quality of this woork. For this reason I recommend that the candidate be awarded the Ph.D. degree.

In Prague, 26 January 2017

Vojtěch Spiwok

Vienna, 27th Jan 2017

Thomas Stockner, Ass-Prof.
Institute of Pharmacology
Medical University of Vienna
Waehringerstrasse 13A
1090 Vienna
Austria

Review of the PhD thesis

Inter and intra domain interactions in the motor subunit of EcoR124I: A computational study

submitted by

M. Sc. Dhiraj Sinha

M.Sc. Dhiraj Sinha focused in the PhD on the bacterial defense system of restriction-methylation enzymes using computational methods and teamed up with experimentalist, which studied the same systems. Mutations were introduced as predicted or the origin of the effects observed experimentally were investigated by simulations. This collaboration was very productive, resulting in 4 peer reviewed and published papers, in two of these as shared first author. Two more manuscripts are in preparation.

The these of M.Sc. Dhiraj Sinha is well written, with a broad introduction on the studied system and the methods applied. Only the large number of missing spaces between words is a bit disturbing, but does not impact on the value of the presented work.

The presented studies are based on the use of molecular dynamics simulations. The interdomain communication between the helicase domains and the endonuclease domains of EcoR124I as well as ATP are the topic of first four results chapters: 3.1 Interdomain communication in the endonuclease/motor subdomain of the type I restriction – modification enzyme EcoR124I; 3.2 The helical domain in the motor subunit of EcoR124I participates in ATPase activity and dsDNA translation; 3.3 The role of motif III and its extended region in positioning the two helicase domains in the motor subunit of the restriction-modification system EoR124I; and 3.4 Alternative conformation of 180 loop. Chapter 3.5 studies the isolated specificity subdomain of the pentameric complex. The overall conformation of the complex, specific interactions and structural changes have been analyzed and the effect of mutations evaluated. These results are discussed in the frame of the function of the restriction-methylation enzymes and the experimental data. These studies make extensive use of principal component analysis to identify the dominant motions.

During the course of this PhD, M.Sc. Dhiraj Sinha was able to push forward our knowledge on function of the EcoR124I defenses system. Given the growing thread posed by bacteria resistance against antibiotics and the associated health risks, an understanding of bacterial defense against their natural enemies could potentially lead to new therapeutic strategies in the future. M.Sc. Dhiraj Sinha contributed to our understanding of domain motions and its link to ATP binding as well as mutations and could also link these changes to the function of the enzyme complex.

The use of molecular dynamics simulations to study the function of large protein complexes remains a challenge. M.Sc. Dhiraj Sinha demonstrated to be able to successfully carry them out. Further discussion of the methodological aspects relevant for the project as well as the pitfalls and their solutions would be welcome.

- In the introduction many practical and methodological aspects of simulations were discussed. The force field and potential energy interaction were introduced. The function the studied EcoR124I complex and of all other proteins depends on $\Delta G = \Delta H - T\Delta S$. Entropy was not directly described, but listed to play a outstanding role in chapter 3.5. A discussion of entropy, how its introduced and treated in simulations should be presented.
- Chapter 3.1 describes the introduction of a restraining force, which needed to be set to 5000 kJ/mol nm^2, a very high number. Such strong forces do often lead to distortions and Figure 3.1.2 does indicate those in the 180 loop. A discussion should be given on the topic of how the quality control of simulations should be carried out. The trajectories of the 3.1 simulations could be used for this purpose.
- Principal component analysis (PCA) are a very useful method to identify significant motions from all the motions along all degrees of freedom. The method does also have several pitfalls. A discussion of how PCA works, which are the requirement for its application and which are its limitations should be given.

.

M.Sc. Dhiraj Sinha has published four publications, two of them as first author. Two more manuscript are in preparation. This underlines the outstanding achievement. I can therefore assess the PhD thesis submitted by M.Sc. Dhiraj Sinha as very good and recommend M.Sc. Dhiraj Sinha for the admission to the PhD defense.

Best regards

Dr. Thomas Stockner

NOVA SOUTHEASTERN UNIVERSITY

To Whom It May Concern:

Subject: Dr. Venk's Comments and Suggestions on Dhiraj Sinha (Ph.D. Candidate's Thesis Titled: Inter and intra domain interactions in the motor subunit of EcoR124I: A computational study.

Mr. Dhiraj Sinha (Ph.D. candidate) was 1) 1st author of 12 total authors in "(J. Mol. Model. (2014) 20:2334, Inter domain communication in the….".

2) Coauthor on "Molecular Dynamics Comparison of *E. coli* WrbA….. (J. Mol. Model (2014) 20:2400".

3) Coauthor on "Quantum Calculations indicate effective electron transfer between FMN… The J. Phys. Chem B (2016) 120: 4867-4877".

4) 2nd Author on "The helical domain of the EcoR124I motor subunit that participates in ATPase activity and dsDNA translocation. PeerJ (2017).

5) First Author (Unpublished/Submitted): The role of motif III and its extended region in positioning the two helicase domains in the motor subunit of the restriction–modification system EcoR124I

All the information from these peer reviewed articles has been compiled nicely in to a Ph.D. thesis.

EcoR124I is a very complicated yet very nice model system to look at structure/function of multimeric proteins. This protein consists of three types of host specificity determinant subunits (Hsd's) 1) monomeric HsdS: DNA specificity/recognition subunit that recognizes a general sequence of (GAANNNNNNRTCG) 2) homo-dimeric HsdM: methyl transferase subunits that modifies the host DNA that uses universal SAM as methyl group donor, 3) homo-dimeric HsdR subunits: (restriction subunits) that has N-terminal endonuclease activity domain, middle helicase 1 and 2 domains (with ATPase/Translocation activity) and C-terminal helical domain (recognizes hsdM). The interactions (of HsdR) i.e. the C-terminal helical domain with DNA translocation, cleavage and ATPase activities of helicase domain is described. Though the thesis looks mainly at the HsdR, it describes less on the HsdM in relation to HsdR. The HsdR and HsdS interactions are fairly well described. The overall dynamics of the three subunits (HsdR, HsdM and HsdS) is not predicted in depth. However, the study focuses mainly on the HsdR especially, on the helicase 2 domain interactions with C-terminal helical domain.

With D796A Mutant: restriction/ability to cleave DNA activity was affected. In panel E: Why the open circle DNA increases with time? As oppose to decrease in OC with WT?

Health Professions Division
College of Medical Sciences
3200 South University Drive • Fort Lauderdale, Florida 33328-2018
(954) 262-1301 • Fax: (954) 262-1802

The linear DNA (L) increase with time is consistent with WT. Also, why supercoiled (SC) in D796A sort of decreases very slowly/remains the same for a prolonged period compared to WT, where you see gradual decrease in SC and complete loss at time (1800s)? Can you explain this better?

The DNA translocation activity of D796A (measured by Fluorescence Intensity) is low compared to WT. However, the relative ATPase activity is retained fairly well compared to WT.

Did you do Kcat/Km (i.e. catalytic efficiency calculations) to see what is happening with D796A mutant?

Also, did you try docking sulfatoATP or non-hydrolysable ATP to see what happens with WT and D796A mutant proteins?

The projections of the first Eigenvector (in the last 20 ns) of WT and D796A looks about the same, However, K527A, Y736A, and the double mutant exhibited -ve (~-7.5) values. Can you explain the dynamics?

Have you tried in silico deletion of C-terminal helical part on one subunit and a another subunit with intact C-terminal helical part (sort of hybrid dimer) and ask the question what happens to DNA translocation, Cleavage and ATPase activity? Does that reduce the overall activity by half? Also did you try doing that in vitro (in other words hybrid dimer (one with complete HsdR + the other with HsdR without helical domain?)? is that even possible?

Motif III of helicase I domain of EcoR124I changes: leads to altered ATPase activity and cleavage activity. Crystal structure: G436, is located too far? (~7 Å) from the γ phosphate of ATP. However, computational analysis suggests G436 residue in motif III plays a role in the positioning of the two helicase domains towards each other. G436A destroys the interactions of helicase I and helicase II domain. Can you explain why that is? How many fold? What is the predicted kcat/km? Simple methyl versus Valine (isopropyl?) side chain?

Overall Strengths: Dhiraj thesis comprises main elements of Abstract, Introduction, Materials and Methods, Results and Discussion. The abstract, Introduction and Methods sections are written very nicely. Dhiraj's Method section is very strong and he explains it part by part very strongly.

Weakness: The results and discussion section could have been split into separate sections. Also, try to have the figure first and then start explaining the figure, rather than the reverse. Some of the Figures used in introduction (the quality/clarity) of the figures could be better (e.g. fig 1.8). In too many places the words are coupled together especially in the results and discussion section. For e.g. theDNA. Give space where appropriate. With nomenclature: be consistent with "EcoR124I" (avoid EcoR4I or in many places EcoR1241).

NOVA SOUTHEASTERN UNIVERSITY

When you say hemi methylated: can you give Figure or %? And the location perhaps?

In the figure "Domain Representation of HsdR……" Where is domain IV?

What is the %homology between EcoKI and EcoR124I? Can you describe more on the looping? Is it like Eukaryotic Intron Looping with consensus "AG" in the exon intron junctions of Eukaryotes. Are there any consensus in the looping regions?

Page 72: "In order to Cover" not coverage…

Page 75: "The In practice" delete "The"

Nomenclature "FMN": FMN is technically not a nucleotide. It is structurally Flavin phosphate. Why are we still calling Flavin Mono Nucleotide? For intuitive docking we want to think of isoalloxazine type ring (a pteridine type nitrogenous base) with ribosyl-phosphate attached to it.

Overall this is a very good Ph.D. thesis work on computational biology with strengths on molecular dynamics, QM, MM etc. I give 90% marks.

Yours Sincerely,

K. V. Venkatachalam

Jan/29/2017

K.V. Venkatachalam, Ph.D.

Professor, Department of Biochemistry

Health Professions Division
College of Medical Sciences
3200 South University Drive • Fort Lauderdale, Florida 33328-2018
(954) 262-1301 • Fax: (954) 262-1802