

Prof. Matthias Bochtler

International Institute of Molecular and Cell Biology &
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

June 24th, 2024

Opinion on the PhD thesis by Dhanasekaran Balakrishnan

Dhanasekaran Balakrishnan submitted a dissertation entitled "*In silico* modeling of natural and designed biological structures". It reports computational work in structural biology. Judging this work now is difficult, because structural biology has changed dramatically since the project was undertaken. Some of the computational approaches to protein structure that were perfectly legitimate at the time the work was initiated were badly outdated by the time the work was completed. Machine learning inspired tools that were difficult to run just a few years ago (AlphaFold, AlphaFold Multimer, ESMFold), are now easy to use, thanks to Google Colab, graphics cards with sufficient memory to accommodate large protein language models, pre-prepared docker containers, or even servers as in case of AlphaFold3. Molecular dynamics approaches have changed less, but the difficulty of performing molecular dynamics experiments has also decreased, thanks to better documentation, more tutorials and GUIs that make it much easier to prepare simulations and to analyze the results.

Part 1 of the thesis

The first part of the dissertation focuses on the NinH protein, which belongs to a group of the nucleoid-associated proteins which also include Lrp (leucine responsive regulatory protein), IHF (integration host factor), HU (heat unstable protein), Dps (a DNA-binding protein from starved cells), and Fis (factor for inversion stimulation, a protein that strongly activates ribosomal RNA transcription). According to Dhanasekaran Balakrishnan, NiH is 40% identical to *E. coli* Fis, which has been structurally characterized in complex with DNA (Hancock et al., "Control of DNA minor groove width and Fis protein binding by the purine 2-amino group," NAR 41:6750-60 (2013)). The relatively high similarity between NinH and Fis makes this an exercise in template-based modelling. This was done using Modeler, which generated a number of candidate models. These were filtered, further refined using GalaxyRefine2 and checked using ERRAT, a program that checks the plausibility of the biophysical environment of residues, and PROCHECK, a program that checks bond lengths and angles and the Ramachandran plot. In template-based modelling, problems with the Ramachandran plot are generally not expected since the protein backbone angles are largely copied from the (hopefully correct) template backbone angles. Hence, a positive PROCHECK report is largely self-fulfilling prophecy and not really a meaningful validation of the model.

Subsequently, AlphaFold2 became available and generated a predicted structure that was very different from the Modeler's predictions. I have little doubt that the AlphaFold2 structure is correct,

Prof. Matthias Bochtler

International Institute of Molecular and Cell Biology &
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

especially given the very high predicted Local Distance Difference Test (pLDDT) score of over 95% for almost the entire protein. However, I am surprised by the poor quality of the Modeler model, given the 40% identity between template and the unknown structure. It almost seems that simply "grafting" the NinH sequence onto the Fis backbone without any adjustments might have been superior to the initial modelling exercise. Is this the case?

Since Fis forms dimers, it was expected that NinH would also form dimers. My approach to elucidating the mode of dimerization would be to assume that the dimer interface is conserved between NinH and Fis. I would then try to test this idea by looking at the conservation of the interface residues. Surprisingly, this has not been done. Why? Instead, Dhanasekaran Balakrishnan performed docking with monomers to find potential dimers, using Haddock. It was then shown that the predicted dimer was consistent with experimental small-angle X-ray scattering data for NinH, so everything seemed fine. But then, AlphaFold Multimer became available, and predicted a rather different dimer than the dimer that had been deduced using Haddock. If I understand the thesis correctly, the AlphaFold dimer was actually close to the Fis dimer, as I would have expected. However, this does not necessarily vindicate the AlphaFold Multimer predicted dimer interface, as the experimental structures of the Fis dimer are quite old and therefore must have been part of the AlphaFold Multimer training set. Since small angle X-ray scattering data for the NinH dimer were available, it was obvious to compare the two models with the experimental scattering data to see whether the data could distinguish between the two models. However, this does not appear to have been done. Why not?

Since NinH binds and presumably also bends DNA, I was very surprised that Dhanasekaran Balakrishnan did not attempt to model the DNA complex. From the structure of the NinH dimer (according to Modeler/Haddock or AlphaFold Multimer), it appears that the characteristic DNA binding helix-turn-helix (HTH) motifs are similarly positioned as in the Fis DNA complex. Now that AlphaFold3 is available, can this prediction be confirmed?

Part 2 of the thesis

The second part of the dissertation addresses a question that is rather unrelated to the first part of the thesis and requires a different methodology. In the second part of the thesis, Dhanasekaran Balakrishnan investigates nanodiscs as a delivery vehicle for membrane proteins to lipid bilayers. "Loaded" nanodiscs contain a membrane protein of interest embedded in a lipid bilayer. The cylinder of membrane protein and surrounding lipids is wrapped by a "double belt" of a membrane scaffold protein (MSP), which prevents unfavorable exposure of the hydrophobic parts of the lipids at the edge of the disk. It is known that membrane proteins can be transferred from nanodiscs to lipid bilayers. However, the detailed pathway of this transfer is not fully clear. Dhanasekaran Balakrishnan addresses the pathway of the fusion process using the example of bacteriorhodopsin in a phosphatidylserine and phosphatidylcholine nanodisc with a pure phosphatidylserine bilayer as the target membrane.

Prof. Matthias Bochtler

International Institute of Molecular and Cell Biology &
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

processes as complex as membrane fusion can occur rapidly on the time scales relevant to the human experimenter, but they occur very slowly on molecular timelines. Therefore, coarse-grained representations are necessary to make the problem computationally tractable. The individual components were first equilibrated separately and then together, prior to the production run. However, it was found that spontaneous fusion did not occur unless the components overlapped from the outset. If overlap was enforced, fusion proceeded largely as expected, with an increasing apposition of the lipid head groups of the nanodisc and the target bilayer. Very reasonably, the juxtaposition of the negative charges of the phosphate head groups of the lipids required cationic bridges to avoid large electrostatic penalties. In the simulations, Ca^{2+} ions were used. However, it remains unclear to me whether any cations, or just divalent cations, or only Ca^{2+} are suitable. From the work, three key conclusions are drawn: (a) the membranes must be able to achieve sufficient dehydration to enable stalk formation between them. (b) membrane scaffold proteins must be rearranged or removed. (c) trapped reverse micelles must be destabilised. For me, as an outsider to the field of membrane fusion processes, it is not clear how novel these conclusions are. If they are novel, how do these conclusions fit experimental evidence?

In the Discussion, Dhanasekaran Balakrishnan states that “the insights (on nanodisc delivery) can be used to design more effective nanodisc-based delivery systems for large membrane proteins”. To me, it is not obvious how this can be done. Since the size of the membrane proteins was not varied in the experiments, why the emphasis on large membrane proteins? Given that it has been some time since the publication of the work, has there been any attempt to actually use the insights. And if so, have the insights from the computational work been validated?

Conclusions

In summary, this dissertation was written under unusual circumstances, in many ways. The supervisor, Prof. Heddle, moved to the UK and led the group mostly remotely for at least part of the PhD. More importantly, the dissertation was written at a time when the tools for *in silico* structure prediction had changed dramatically, to an extent that few in the community expected. I can imagine how uncomfortable it must have been to operate in this environment of rapidly shifting goalposts. When AlphaFold2 became available, much of the early work on NinH became obsolete, and the same is now true for the entire NinH work with the advent of AlphaFold3. Fortunately, the molecular dynamics part of the thesis is not affected by this problem. Also, in fairness, the work should be judged by the standards of the time when it was done, not retrospectively from today’s perspective.

Overall, it seems to me that the topic of the thesis was too broad. As a consequence, the Introduction is too general. Unfortunately, many contradictions in the computational work could not be resolved. Moreover, key questions about the nucleic acid interactions of NinH could not/were not addressed, which limits the biological insights from the work. In the part dealing with nanodiscs as delivery

Prof. Matthias Bochtler

International Institute of Molecular and Cell Biology &
Trojdena 4, 02-109 Warsaw, Poland
Institute of Biochemistry and Biophysics
Pawinskiego 5a, 02-106 Warsaw, Poland
Tel: 0048 22 5970732
Fax: 0048 22 5970715
e-mail: mbochtler@iimcb.gov.pl

vehicles, it remained unclear to what extent the key conclusions were novel (from a computational perspective) and, if so, to what extent they were supported by/consistent with experimental evidence. However, I acknowledge that Dhanasekaran Balakrishnan has clearly learnt a lot about modelling, including the use of new protein language model-based structure prediction tools. The combined expertise in molecular modeling/structure prediction on the one hand and molecular dynamics on the other hand is certainly going to be useful, at least for the near future. In the longer term, I suspect that the kind of work reported in the thesis will become so easy that one will need to either become part of the methods development community or combine the theoretical work with experimental validation to have an impact.

Dhanasekaran Balakrishnan's work has passed peer review and is published in three publications that are authored or co-authored by Dhanasekaran Balakrishnan. For one of these papers, from 2019, in Nanomedicine, he is the first author. I therefore recommend to proceed with the next stages towards the award of a PhD.

With best regards



Matthias Bochtler