

PhD Thesis Acceptance Report
Research Discipline Council of Biological Sciences
Jagiellonian University in Kraków

Candidate's name and surname:Dhanasekaran Balakrishnan.....

PhD Thesis Title:*in silico* modeling of natural and designed biological structures...

Thesis Supervisor:Prof. Dr. Jonathan G.....

Heddle.....

Assistant Supervisor / Second Supervisor/ Co-supervisor (if applicable):.....

Reviewer:Prof. Dr. Maya Topf.....

This thesis explores the intersection of protein engineering and synthetic structural biology, aiming to create novel biomolecules with improved functionalities. Chapter 3 focuses on predicting the 3D structure of the NinH protein, revealing insights into its conformation and function. Chapter 4 investigates the dynamic process of nanodisc fusion with lipid bilayers, offering a foundation for developing efficient cargo delivery systems. The results presented in this chapter are compelling, well-presented, and thoroughly analyzed. The elucidation of the nanodisc-membrane fusion mechanism has the potential to aid in the design of optimized nanodiscs for efficient cargo delivery. However, the research presented in Chapter 3, although enriching our understanding of nucleoid-associated proteins, is not complete. The analysis contains errors, and better analysis and the possible application of additional bioinformatics tools/simulations could further deepen our insights into the calculated models of the NinH protein. Additionally, I found areas for improvement throughout the thesis, including grammatical errors, a lack of detailed explanations of different terms and scores, and several errors in figures and figure legends. Addressing these issues would enhance the clarity and impact of the thesis. The Summary chapter provides a good overview but would benefit from a more detailed discussion of the implications of the findings. It should acknowledge the study's limitations and suggest future research areas beyond one simple example.

Below I detailed my comments on each chapter:

Chapter 1: Introduction

In this chapter, general topics are covered, such as introduction to protein structure, protein structure prediction methods, nucleoid-associated proteins, and biological membranes, membrane fusion, and artificial cells. The theory behind some of the methods used in the thesis are also presented. The introduction generally gives a good cover of the literature. However, it would make more sense to give the introduction to MD and AI before describing the biological topics, as many of the methods (e.g., MD) are referred to already in that part of the chapter (e.g., Figures 1.18, 1.19).

Other minor points for this chapter:

- Page 52: Table 1.1: Comparing conventional methods with AlphaFold2: AlphaFold2 can be used with templates.
- Page 53: "A "force field" is a mathematical formula that includes a set of parameters used to represent the energy of a protein based on its atomic positions, and it can be divided into two

groups: i) bonded and ii) non-bonded.” A forcefield as to be explained better as well as the MD potential energy function. In general, given that it plays a significant part of the thesis, the student should provide more details about MD.

- Page 59: the description of the aims of the research (two main aims). This part sounds very technical. It would be better to have a more detailed description of why the student wants to model these protein/lipid bilayers, what is the motivation, ultimate goal, why it is useful, etc.

Chapter 2: Materials and Methods

This thesis chapter describes the methodologies and computational tools used to analyze protein structures, focusing on the NinH and Fis proteins from *E. coli*, which are the subject of Chapter 3. Specifically, it details the use of the Consurf webserver to identify conserved residues, PSIPRED for secondary structure prediction, Clustal Omega for multiple sequence alignment, and MODELLER for homology modeling. The chapter also includes the use of HADDOCK for protein-protein docking, CPORT for interface residue prediction, and GalaxyRefineComplex for refining predicted protein complexes. Additionally, it mentions the use of deep learning tools like AlphaFold2 and ESMFold for structure prediction. Finally it summarises tools like CHARMM-GUI and GROMACS, and how they were used for MD simulations of protein-lipid interactions described in Chapter 4.

Overall, the chapter presents a comprehensive approach to predicting and analyzing protein structures using a combination of bioinformatics tools and computational methods. Analysis tools should be explained better, e.g., what is RMSD, between which atoms it is calculated in the context of this thesis, and how. Other measures should also be explained in more detail. For example, scores used to assess HADDOCK models or AlphaFold2 models.

A number of points for improvement are described below:

- The student needs to be consistent with the naming of the methods: is it MARTINI or Martini? Same for GROMACS, MODELLER, ESMFold, etc.
- Page 44: bacteriorhodopsin appears in full name for the first time, although above it was abbreviated to bR. It is not in the abbreviation list.
- Page 61: explain Hidden Markov Models better and add a reference.
- Page 62: the link to CPORT is incorrect. This is the correct one: <https://alcazar.science.uu.nl/services/CPORT/>
- Page 64: the student needs to explain the different machine learning tools that he used in Chapter 3. Why is ESMFold not mentioned in the introduction in more detail, where AlphaFold is described? It has a different machine-learning approach (using language models) and therefore this should be explained in the beginning of the thesis (or at least in Chapter 2, but I only saw the explanation in Chapter 3, page 83).
- Page 66: SAXS is mentioned, but there is nothing in the Introduction chapter about SAXS and why it is a helpful method in the context of the thesis. There is a description of various methods and programs used in this context (ATSAS, GNOM, DAMMIF), but what exactly they do and how is missing.
- Page 67: PyMOL is not referenced.

Chapter 3: NinH protein structure prediction and analysis

In this chapter, bioinformatics tools like Modeller 9V21, AlphaFold2, and ESMFold were utilized to predict the three-dimensional structure of the NinH protein, a member of the nucleoid-associated protein family. Through bioinformatics, it was discovered that the NinH protein forms a homodimer and shares functional characteristics with the Fis protein. However, the C-terminal region of the NinH protein remains poorly understood. SAXS experiments supported these findings, but higher resolution data is needed for precise placement of the C-terminal regions. On one hand, these results enhance

understanding of NinH protein structure and function, providing a foundation for future research. On the other hand, I find the results lacking attention to details, and the analysis often too simplistic and the discussion too simplistic and not sufficiently critical. The student concludes that the MODELLER/HADDOCK “manual” model makes sense. I am not sure what “manual” means in this case, but the alphafold model also makes sense and it has not been tested against the SAXS model. The 4th helix in AlphaFold correlates with the SS prediction. I would have liked to see a better discussion about these findings. Overall I think the work presented in this chapter could be improved, possibly with the use of MD simulations (this could also be another chapter).

In this chapter, bioinformatics tools like Modeller 9V21, AlphaFold2, and ESMFold were utilized to predict the three-dimensional structure of the NinH protein, a member of the nucleoid-associated protein family. Through bioinformatics, it was discovered that the NinH protein forms a homodimer and shares functional characteristics with the Fis protein. However, the C-terminal region of the NinH protein remains poorly understood. SAXS experiments supported these findings, but higher-resolution data is needed for precise placement of the C-terminal regions. On one hand, these results enhance understanding of NinH protein structure and function, providing a foundation for future research. On the other hand, I find the results lacking attention to detail, and the analysis is often too simplistic, with the discussion not being sufficiently critical. The student concludes that the MODELLER/HADDOCK “manual” model makes sense. I am not sure what “manual” means in this case, but the AlphaFold model also makes sense and it has not been tested against the SAXS data. The 4th helix in AlphaFold correlates with the SS prediction. I would have liked to see a better discussion about these findings. Overall, I think the work presented in this chapter could be improved, possibly with the use of MD simulations.

Other minor points for this chapter:

- Figure 3.2: It seems that the model that is shown was produced by Phyre2. Why is it not mentioned in Chapter 2? What is the approach behind Phyre2? Also, I don’t really see a structural alignment; I only see the Fis protein. Or is it the model of NinH that is shown in the inset? If it is NinH, why is the protein labeled Fis?
- Figure 3.5: “helixes” should be “helices”; this error appears a number of times.
- Page 73: Why were the top models selected only based on structural similarity to the template and not by a model assessment approach (e.g., DOPE)? ERRAT is used later, but only to describe the quality of the selected model. Was this model actually selected by ERRAT, or if not, how was it selected?
- Figure 3.8: How was the RMSD calculated? Between what regions of the structure? Was it only C α ? Why is RMSD not explained?
- Figure 3.9: There is an error in the figure and the legend. How can His45 be so far from His55 and Arg56?
- Page 77: The last sentence makes no sense. Also, we don’t know if it was modeled correctly; we can just assume that the model makes sense because of the resemblance to Fis proteins.
- Page 79: “The residues discovered across the interface are consistent with the CPORT algorithm prediction of interfacial residues” – CPORT also predicted H45, but I don’t see it in the analysis, possibly because there is an error in Figure 3.9. The student needs to explain this clearly. For some reason, this is explained in the Discussion and Figure 3.22 but not shown in the results.
- Figures 3.14, 3.16: The pLDDT score should be visualized on the structure so it is easy to see the regions that are not well predicted (like the loops).
- Figure 3.15: The comparison between the models is not clear. The orientation of the models is not the same. It would be better if the student superposed the models to indicate the differences and also indicated those differences in the context of the sequence and secondary structure predictions. On Page 83, the student claims that the 4th predicted helix from the

- AlphaFold model doesn't make sense, yet the SS prediction in Figure 3.1 shows a 4th helix in the same region.
- Figure 3.17: Both AlphaFold2 and ESMFold predicted the 4th helix, as mentioned. Yet, the student is not explaining that it makes sense, and the position of this helix may be different in the model due to the flexibility of the loop connecting it.
- Figure 3.18: Why is the model colored from N- to C-terminal in the dimers? Each monomer has to be colored separately.
- Page 88, Discussion: The second sentence should be reworded; it is too long.
- Page 89, last sentence: Not sure what this sentence means: "with residues present in the C-terminal region more likely to be involved in the NinH."
- Page 93, Summary: Given that the model is inconclusive, I would have liked to see more discussion about it. The last sentence, "Overall, these findings shed light on the structure and function of the NinH protein and lay down the foundation for future research," is vague. What future research? What light does the model shed on function?

Chapter 4: Nanodiscs as a delivery vehicle

This final chapter details the investigation of calcium-mediated membrane fusion between a bacteriorhodopsin (bR)-nanodisc and a lipid bilayer via coarse-grained simulations. The use of bR embedded in a PS nanodisc enriched with PC and stabilized by MSPs is a sound choice.

The chapter effectively demonstrates the potential of this system as a "molecular shuttle," capable of transferring membrane proteins to preformed GUVs or SUVs. The method's applicability to a wide range of prokaryotic and eukaryotic proteins, due to the negatively charged membranes, broadens its utility. The potential success of this approach could enable the development of nanodisc systems for precise control of protein orientation, useful for equipping protocells with essential membrane proteins. The chapter also acknowledges and addresses potential drawbacks, such as cross-fusion of recipient membranes, proposing viable solutions like microfluidic setups and DNA nanotechnology for precise positioning. This foresight enhances the work's credibility and practical relevance.

Overall, this chapter provides a compelling and detailed exploration of nanodisc-lipid bilayer fusion, showcasing innovative approaches and practical solutions. However, I still found a few mistakes in sentences, figure legends, and general explanations of the figures.

Below are my specific comments:

- Page 94: bR-nanodiscs - again bR appears without explanation (which comes further in the text).
- Page 100: VMD is not referenced. Was it used in all visualizations?
- Figure 4.7: MSP chain A has to be indicated in the figure.
- Page 102: The second sentence has to be rephrased.
- Figure 4.8: The legend has to be rewritten.
- Figure 4.10 and page 102: How was the RMSD calculated exactly? Which atoms were used?
- Figure 4.12: The legend has to be rephrased. Also, the transported nanodisc lipids should have a different color as green was used to show bR in previous images.

I, hereby, declare that the reviewed PhD thesis by **Dhanasekaran Balakrishnan** meets the criteria pursuant to art. 187 of Act of 20 July 2018 The Law on Higher Education and Science (Journal of Laws of 2018, item 1668, as amended) and request that the Research Discipline Council of Biological Sciences of the Jagiellonian University in Kraków accepts **Dhanasekaran Balakrishnan** for further stages of doctoral proceedings in the field of exact and biological sciences, in the discipline of biological sciences

YES/NO

I, hereby, request that the thesis is accepted with distinctions

YES/NO

Justification of the request (if YES is selected)

18/06/2024

.....
date


.....
Reviewer's signature

INFORMATION FOR THE REVIEWER:

A digital copy should be sent to:
nauki.biologiczne@uj.edu.pl

A duly signed original should be sent to:

**Rada Dyscypliny Nauki biologiczne
Dziekanat Wydziału Biologii
Uniwersytet Jagielloński w Krakowie
ul. Gronostajowa 7
30-387 Kraków**