INSTYTUT FIZYKI
POLSKIEJ AKADEMII NAUK

Aleja Lotników 32/46, 02-668 Warsaw, Poland , tel.: (+48) 22 843 70 01 | fax: (+48) 22 843 09 26 | e-mail: director@ifpan.edu.pl www.ifpan.edu.pl NIP: 525-000-92-75, Regon: 000326061

# Review

of the PhD thesis by Mateusz Piotr Marcisz, MSc

***Development of novel computational approaches for glycosaminoglycans***

submitted for the degree of Doctor of Philosophy

University of Gdańsk, Poland

The thesis of Mr. Mateusz Marcisz is devoted to glycosaminoglycans and their interactions with proteins - one of the important issues in interdisciplinary science including biophysics, chemistry, biomedical etc. The dissertation consists of 4 chapters and Conclusion.

Chapter 1 starts with describing structure and function of glycosaminoglycans (GAGs) which are linear anionic periodic polysaccharides that play an important role in different biologically relevant processes. The role of water in GAG-containing systems, in particular in GAG-protein complexes was highlighted.

Since this thesis intensively studied GAG-protein complexes, it would have been logical to include at least a brief introduction to the proteins, but there is no such part. On page 13, the author wrote: "It was shown that LMW HP (Heparin) is a potent drug in the fight with coronavirus and could decrease mortality through anti-viral, anti-inflammatory and anti-coagulant activities". However, to my best knowledge, this molecule has the potential for treating Covid-19, but is not a drug (see https://doi.org/10.3389/fphar.2023.1159363s for a recent review).

In Chapter 2, the author summarized various experimental techniques for studying GAG-containing systems, in particular, the X-ray diffraction was well described. In many cases, the combination of these methods with molecular dynamics simulations has provided detailed information about GAG-protein interactions. Recent references on GAG interactions with the SARS-CoV-2 spike protein are also presented, showing that he is very knowledgeable about the literature.

Molecular docking with various search algorithms and scoring functions was described in detail. Particular attention has been paid to the repulsive scaling replica exchange molecular dynamics

(RS-REMD) method, which is probably the best method for studying GAG-containing systems. Theoretical methods for estimating the binding free energy, such as MM/P(G)BSA, LIE were presented. A good review of the literature on various implicit and explicit water models and their use in GAG simulation was presented. The results obtained using all-atom and coarse-grained molecular dynamics simulations for GAG-containing complexes were summarized and analyzed. Overall, the candidate has demonstrated a good knowledge of the scientific literature in his field.

Page 18: "transmission electron miscopy (TEM) " should be replaced with " transmission electron microscopy (TEM)".

Page 32: What MMFF stands for?

Chapter 3 outlines the objectives of the dissertation.

Chapter 4 is a collection of 7 publications used in the dissertation. I will review the main findings of each publication and comment or criticize if any.

*Publication D1*

In this publication, the effectiveness of RS-REMD for docking glycosaminoglycans to proteins was investigated. This method was applied to a set of 21 protein-GAG complexes with experimentally resolved structures and GAG lengths ranging from dp5 to dp7. Of the 21 systems, RS-REMD successfully predicted binding sites in 19 cases. The RS-REMD method has been shown to fail when binding sites are located in enzymatic pockets, which is likely due to the fact that REMD-RS simulations include replicas with larger van der Waals radii in each subsequent replica, resulting in larger volumes of the ligand and receptor atoms than those in the unmodified force field. Thus, although RS-REMD is better than Autodock3, it does not have a 100% success rate, and one of the goals of this thesis is to improve the original version of RS-REMD.

Figure D1 was not cited in the text.

*Publication D2*

Together with collaborators the author proposed an improved RS-REMD protocol, which replaces the implicit water model by an explicit water model for studying protein-GAG complexes. As expected, application of this protocol to three complexes (FGF1-HP dp6, FGF2-HP dp6, and ATIII-HP dp8) showed that explicit solvent significantly improves docking quality by providing a more realistic description of solvent-solute interactions. It is important to note that the new protocol does not increase computation time, since after production run there is no need to refine the results, as in the original method with an implicit water model. Thanks to this advantage, the new protocol can be more widely used in the future not only for systems containing GAGs, but also for other systems.

*Publication D3*

Autodock3 has been successfully used to study the binding of short GAGs with a degree of polymerization less than 7 to proteins. However, Autodock3 fails in case of sufficiently long GAGs, which motivates the author to develop a new GAG model to overcome this difficulty. In

the new model, the GAG fragment directly in contact with the protein is described by an all-atom (AA) model, and the remaining part is treated using coarse-grained (CG) models. Using MM/GBSA to study the binding affinities of the AA and hybrid AA/CG models, the difference in binding free energies of the two complexes was shown to be about 5.6%, confirming the reliability of the new model.

Since the author did not discuss the contribution of the entropy term, I wonder how big the difference is between the AA and AA/CG models in terms of $T\Delta S$. In addition, Autodock3 fails to dock long GAGs to proteins, but can one use protein-protein docking server like HDOCK etc. ? Figure D3 was not cited in the text.

From a methodological point of view, this publication is the most impressive for me because the author proposed a hybrid AA/CG model for GAGs.

*Publication D4*

The binding of several GAGs to APRIL protein (A proliferation-inducing ligand) was examined using *in vitro* and *in silico* experiments. Understanding this issue is important because binding of APRIL to GAGs is thought to promote its oligomerization and mediate its function in cell signaling. APRIL was found to have direct contact with GAGs not only at its N-terminus, but also with some lysine and arginine residues at the C-terminus and in loops near the N-terminus. Molecular simulations showed that APRIL interacts with HP more strongly than with CS-E (chondroitinsulfate-E), while the binding affinity for CS-C (chondroitinsulfate-C) is very weak. Both experimental and computational studies have confirmed the importance of the N-terminus of APRIL in its binding to GAG, as truncated APRIL has lower binding affinity compared to the full-length variant due to the removal of some charged residues. Moreover, HS (heparansulfate) binding to the TACI (transmembrane activator and calcium modulator and cyclophilin ligand interactor) protein was not observed in either simulations or experiments.

Because GAGs interact with proteins that are involved in pathways of a number of diseases, including cancer, it was stated that the results obtained in this work could be useful for drug discovery. I wonder if some GAGs are FDA approved to treat any diseases?

Figure D4: It would be easier for readers if N- and C-terminal tags were added.

*Publication D5*

As mentioned above in the dissertation, standard docking methods face difficulties in docking GAGs longer than dp6-10 to proteins. Therefore, in this work, an improved all-atom RS-REMD method was used to study the binding of HP dp24 and HP dp48 to APRIL bound to its receptors TACI and BCMA (B-cell maturation antigen). HP was found to bind more strongly to the APRIL-BCMA complex than to the APRIL-TACI complex. Interestingly, the HP-BCMA and HP-TACI interaction is enhanced by the binding of APRIL to its receptors, presumably through long-range electrostatic interactions. All-atom RS-REMD has been shown to be superior to Autodock3 and is a useful tool for studying the binding of long GAGs to proteins.

*Publication D6*

The purpose of this work is to examine the effectiveness of different solvent models in molecular dynamics simulations of GAGs. This issue is important not only because water molecules at protein-GAG interfaces are much more abundant than at protein-protein interfaces, but also because the simulation results depend on the choice of solvent model. To perform benchmark tests, MD simulations were carried out for HP, HS and CS decamers with five implicit solvent models (IGB = 1, 2, 5, 7, 8) and 6 explicit solvent models (TIP3P, SPC/E, TIP4P, TIP4PEw, ORS, TYPE5P). The GLYCAM06 force field was used to describe GAGs. The results obtained for various quantities characterizing GAGs were compared with existing experimental data to justify the validity of each model. As expected, explicit solvent models outperform implicit ones. However, none of the explicit models can fully capture all the characteristics being studied. The most popular TIP3P force field in GAG simulations fails to provide satisfactory results for GAG chain curvature and end-to-end distance, while only the TIP5P and OPC models match experimental data for these quantities. Thus, the main message of this work is that one must be very careful when choosing a solvent model for modeling GAGs.

It has not been mentioned that the GLYCAM06 force field was used to characterize GAGs.

*Publication D7*

This is a review article “Modeling glycosaminoglycan–protein complexes” published in the prestigious journal Current Opinion in Structural Biology. I'm not sure if a review can be used for a dissertation, but the fact that Mr. Mateusz Marcisz is the first author of this paper demonstrates his maturity in science. I believe this review is very useful for researchers working on GAGs and related systems.

To summarize I have read the thesis by Mateusz Marcisz with great interest. It is comprehensive, systematic, logical and well-presented study of important biochemical systems. It was written in rather good English. One should be aware that all parts of this thesis have been already published in 7 papers in highly ranked international journals. Mr. Mateusz Marcisz demonstrates that he has good skill in molecular docking, molecular dynamics simulations, and data analysis. The author has shown a good orientation in the scientific literature related to his research.

Despite some minor, but as always necessary criticism, I think that this is a solid thesis, fully consistent with the Polish standards known to me. Therefore, I recommend proceeding with further steps of the PhD degree procedure for Mr. Mateusz Marcisz .

Warsaw, September 29, 2023

Mai Suan Li, Professor of Physics

Email: masli@ifpan.edu.pl

Tel. +48-22-843 66 01 (ext. 3457)

URL: http://info.ifpan.edu.pl/~masli/index.html