UTILIZING COMPUTATIONAL METHODS TO STUDY THE BIOMOLECULES

WENHAN GUO

Master's Program in Computational Science

APPROVED:

Lin Li, Ph.D., Chair

Jorge A. Muñoz, Ph.D.

Weihong Qiu, Ph.D.

Stephen L. Crites, Ph.D.
Dean of the Graduate School

To the important person in my life.

UTILIZING COMPUTATIONAL METHODS TO STUDY THE BIOMOLECULES

by

Wenhan Guo, M.Med.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Computational Science Program

THE UNIVERSITY OF TEXAS AT EL PASO

August 2022

ACKNOWLEDGEMENTS

In this research, firstly I would like to acknowledge my advisor Dr. Lin Li, from the Department of Physical Sciences, for his patient guidance and help. Dr. Li encourages me a lot, both in the research work and in my attitude to the future. Meanwhile, I would like to express gratitude to my lab colleagues, my boyfriend, and my friends in El Paso, they have inspired me a lot and given me continual support. And I would like to thank the committee members Dr. Weihong Qiu from the Department of Physics of Oregon State University and Dr. Jorge Muñoz from the Department of Physical Sciences for their kind help and advice. Finally, I am also grateful to my family for their wonderful love and understanding. Thank them for supporting me in my research path.

ABSTRACT

COVID-19 is increasingly affecting human health and global economy. Understanding the fundamental mechanisms of Severe Acute Respiratory Syndrome CoronaVirus 2 (SARS-CoV-2) is highly demanded to develop treatments for COVID-19. SARS-CoV and SARS-CoV-2 share 92.06% identity in their N protein RBDs' sequences, which results in very similar structures. However, the SARS-CoV-2 is more easily to spread. Utilizing multi-scale computational approaches, this work studied the fundamental mechanisms of the nucleocapsid (N) proteins of SARS-CoV and SARS-CoV-2, including their stabilities and binding strengths with RNAs at different pH values. Electrostatic potential on the surfaces of N proteins show that both the N proteins of SARS-CoV and SARS-CoV-2 have dominantly positive potential to attract RNAs. The binding forces between SARS-CoV N protein and RNAs at different distances are similar to that of SARS-CoV-2, both in directions and magnitudes. The electric filed lines between N proteins and RNAs are also similar for both SARS-CoV and SARS-CoV-2. The folding energy and binding energy dependence on pH revealed that the best environment for N proteins to perform their functions with RNAs is the weak acidic environment.

Kinesins are microtubule-based motor proteins that play important roles ranging from intracellular transport to cell division. Human kinesin-5/Eg5 is essential for mitotic spindle assembly during cell division. By combining molecular dynamics (MD) simulations with other multi-scale computational approaches, we systematically studied

the interaction between Eg5 and the microtubule. Our results showed that electrostatic potential on the binding interface of the Eg5 motor domain is dominantly positive while electrostatic potential on the binding interface of αβ-tubulin heterodimer is dominantly negative. Detailed electrostatic distributions on the binding interfaces were illustrated in this work. We found that binding forces between the Eg5 motor domains and the αβ-tubulin heterodimer at different distances are consistent with the attractive electrostatic forces in both directions and magnitudes. Electric field lines between Eg5 and the αβ-tubulin heterodimer indicate a strong, attractive force between Eg5 and the αβ-tubulin heterodimer. The folding and binding energy dependence on pH reveals that the Eg5 motor domain performs its functions best with microtubules in the weak acidic environment. The analyses on hydrogen bonds and salt bridges demonstrate that on the binding interfaces of Eg5 and tubulin heterodimer, the salt bridge plays the most significant role in holding the complex structure. The salt bridge residues on the binding interface of Eg5 are mostly positive, while salt bridge residues on the binding interface of tubulin heterodimer are mostly negative. Such salt bridge residue distribution is consistent with the electrostatic potential calculations. On the contrast, the interfaces between α- and β-tubulins are dominated by hydrogen bonds rather than salt bridges. Compared with the salt bridges between Eg5 and α-tubulin interfaces, the salt bridges between Eg5 and β-tubulin have a greater number and higher occupancies. This asymmetric salt bridge distribution may play a significant role in Eg5's directionality. The residues involved in hydrogen bonds and salt bridges are identified in this work, which may be helpful for anticancer drug design.

TABLE OF CONTENTS

LIST OF FIGURES

PROJECT 1: ELECTROSTATIC FEATURES FOR NUCLEOCAPSID PROTEINS OF SARS-COV AND SARS-COV-2

1.1. INTRODUCTION

Severe Acute Respiratory Syndrome Corona Virus 2 (SARS-CoV-2) is currently affecting human health and global economy seriously. A similar situation happened in 2003 with SARS-CoV, which also belongs to Coronavirus family. SARS-CoV and SARS-CoV-2 share 92.06% identity in their N protein RBDs' sequences [1,2]. Both SARS-CoV and SARS-CoV-2's genomes encode nonstructural replicase polyproteins and structural proteins [1], including the nucleocapsid phosphoprotein (N protein). The main function of N protein is to link envelopes to the +RNA. The N protein of SARS had shown to play a crucial role in regulating viral RNA synthesis in replication and transcription [3]. Understanding the fundamental mechanisms of how N proteins Receptor Binding Domains (RBDs) of SARS-CoV and SARS-CoV-2 bind RNAs is highly demanded for developing new antiviral drugs and vaccines [4].

Some groups studied the N proteins of SARS-CoV and SARS-CoV-2 using experimental methods. Sisi Kang et. al [3] utilized chemical experiments and X-ray analysis to obtain the structure of N protein of SARS-CoV-2, which helped revealing potential drug targeting sites. Veverka V's laboratory [5] performed NMR-based titration experiments, combined with computational model, to build the complex model of the Nucleocapsid N-Terminal RNA binding Domains (N-NTD) with RNA. Unfortunately, only a few groups have conducted research on the structure and function of N proteins of SARS-CoV. Peter Kuhn's team [6] is one of them who characterized the structures of the N-NTD of SARS-CoV. Compared with experimental studies,

some effort has been also made to investigate SARS-CoV and SARS-CoV-2 using computational approaches. Most of these computational studies focused on the spike (S) proteins of the SARS-CoV and SARS-CoV-2 [7,8], including discoveries of potential drug targets for SARS-CoV-2 [9,10,11], few works focused on N proteins. Some studies calculated the electrostatic potential on N protein surfaces in coronavirus [3,5,6]. Electrostatic features of N proteins help us understanding different mechanisms of RNA recognition and assembly. Other calculations between N proteins and RNAs explore more fundamental principles for their binding mechanisms.

Due to the relatively high cost of experiments and the rapid development of computational algorithms [12,13], computational methods are now widely used to study biology phenomena, including biomolecular structures [14,15], biomolecular interactions [16,17,18], pH dependence of protein-protein/DNA/RNA interactions [19,20], etc. Using such state of art computing techniques, a lot of efforts have been contributed to study viruses [7,21,22]. In this work, several computational approaches are used to study nucleocapsid proteins of SARS-CoV and SARS-CoV-2, including DelPhi [23], DelPhiForce [24,25], DelPhiPKa [26,27]. The electrostatic features are critical in analyzing the interactions between the N protein and RNA. Thus, the electrostatic potential, electric field lines and electrostatic forces were analyzed based on the structures of N proteins of SARS-CoV/SARS-CoV-2 RBDs and RNAs. It was found that SARS-CoV and SARS-CoV-2 have similar electrostatic potential distributions on their binding surfaces, which demonstrated that the net charges play a significant role to attract the RNAs. In addition, DelPhiPKa was implemented to calculate the binding energy pH dependence. Such method has been proved successful and reliable [20,28,29,30]. The pH effects on the binding energies for N proteins' RBDs interacting with RNAs and folding energies of N proteins was

analyzed, which demonstrated the optimal pH for N proteins' folding and binding with RNAs. Such details assist us to understand how the N proteins' RBDs recognize RNAs. These findings pave the way for research on future coronavirus-caused diseases. No experimental studies have been conducted to reveal the differences between the biology functions of SARS-CoV and SARS-CoV-2. Therefore, this work of comparing the N proteins of SARS-CoV and SARS-CoV-2 can also be useful for future experimental design.

1.2. METHODS

1.2.1 Structure Preparation

The complex structure of SARS-CoV-2 with the Double Strand RNA (dsRNA) was obtained from Protein Data Bank (pdb ID: 7ACS [5]). The SARS-CoV structure was obtained from Protein Data Bank (pdb ID: 2OFZ [6]), which does not include the dsRNA structure. Therefore, the complex structure of dsRNA combined with SARS-CoV N protein was modeled by aligning the SARS-CoV structure to SARS-CoV-2 based on the template of 7ACS using Chimera [31]. This study is mainly focused on the electrostatic features of Nucleocapsid N-Terminal RNA binding Domains (N-NTDs) of SARS-CoV and SARS-CoV-2. In the SARS-CoV N protein structure, the N and C terminals are not determined [6]. Figure 1.1 shows the complex structures of SARS-CoV-2 N protein RBD binding with RNA, which is determined by NMR experiments [5]. The NMR structures demonstrate none of the N or C terminals of SARS-CoV-2 binds to RNAs, therefore the N and C terminals are extremely flexible. Due to this experimental evidence, N and C terminals of SARS-CoV-2 were deleted in this work. After the deletion, we obtained the same length of N proteins for SARS-CoV-2 and SARS-CoV.