Virtual Screening and In Silico Interactions Studies for Potential Antivirals and Diagnostics against the Spike protein from the Novel Coronavirus SARS-CoV-2

F I Che Abd Aziz¹, F A Ahmad Fuad¹², S Tanbin¹

¹Department of Biotechnology Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia

²Corresponding author’s email: fazia_adyani@iium.edu.my

Abstract. COVID-19 is a newly-emerged respiratory disease that is caused by the SARS-CoV-2, the seventh known Coronaviruses strain that has struck a global pandemic. The sharp increase in the number of positive cases worldwide necessitates highly-sensitive diagnostics kits and effective antiviral drugs to be developed for the populations. One of the antigens that is targeted for antibody neutralisation is the coronavirus Spike protein that consists of the S1 and S2 subunits, which mediated the entry pathway into the host’s cell. Thus, the Spike protein has been suggested as a potential target for Covid-19 diagnostics and drug design. This study aims to evaluate the interactions between the SARS-CoV-2 Spike protein and the known monoclonal antibodies from Coronaviruses and to screen for potential Spike protein inhibitors. Virtual screening was conducted based on two compounds, N-acetyl-D-glucosamine (NAG) and Hesperetin, which is a small molecule that binds to the SARS-CoV-2 Spike protein structure and a natural compound that has prophylactic agents against SARS-CoV-2 infection as it binds to Spike protein, respectively. Protein-protein interaction studies were conducted by using the STRING webservice, prior to performing rigid docking using SWISSDOCK and visualised using UCSF Chimera. Meanwhile, ligand-based screening was conducted through Ultrafast Shape Recognition Virtual Screening Database (USR-VS), and structure-based screening was performed via AutoDock4 software. The toxicity of the compounds was predicted using ProTox-II database. Possible interactions have been observed between the known monoclonal antibodies with the SARS-CoV-2 Spike protein, where M396 monoclonal antibody has shown the strongest interaction with a binding energy of -8.50 kcal/mol. Meanwhile, virtual screening has yielded several compounds that indicate the possibility to inhibit the SARS-CoV-2 Spike protein, where Tamarixetin has shown the strongest binding energy of -7.93 kcal/mol. These findings have potentials to be further evaluated in the future for the development of improved diagnostic kits and potential therapeutic drugs that specifically target the Spike protein of SARS-CoV-2.

1. Introduction
Coronavirus originates from the Coronaviridae family and a subfamily of Coronavirus, which is divided into four genera, α-coronavirus, β-coronavirus, γ-coronavirus and δ-coronavirus [1]. It is a single-stranded, positive sense of RNA genome (+ssRNA) with 5’-cap and 3’-poly-A tail that is approximately 26-32 kilobases in size [2]. As of today, there are seven types of HCoV identified until early 2020, which are 229E, OC43, NL63, HKU1, SARS (Severe Acute Respiratory Syndrome), MERS (Middle East Respiratory Syndrome) and SARS-CoV-2, which is the newly-emerged novel coronavirus that caused the Covid-19 pandemic. The virus was firstly detected in Wuhan, Hubei Province of China from the Huanan local traditional market in December 2019. The infected patients showed similar symptoms to previous diseases known as SARS-CoV and MERS-CoV, which were declared epidemics in the year 2002 and 2012, respectively. The resemblance is due to SARS-CoV-2 carrying similar pathogenic characteristics with SARS-CoV and MERS-CoV with 79% and 50% similarity, accordingly [1].
SARS-CoV-2 contains several structural proteins that are crucial for fusion and assembly of the virus, including the Spike protein, Membrane protein, Envelope protein and Nucleocapsid protein. The Spike protein has been proposed as a potential drug target for disease treatment, as it consists of the S1 and S2 subunits, where S1 contains the receptor binding domain (RBD) that binds with the host cell receptor, while S2 facilitates the fusion and entrance process into the host’s cell targeting the angiotensin-converting enzyme 2 (ACE2) to lower the host’s immune response [3]. A ligand, 2-acetamido-2-deoxy-beta-D-glucopyranose (NAG) is one of the small molecules that bind between the receptor-binding-domain (RBD) of the SARS-CoV-2 Spike protein and the host cell receptor angiotensin-converting-enzyme 2 (ACE2). This ligand was added to stabilize the crystal structure of the SARS-CoV-2 Spike protein, due to the presence of the chemical bond between the carbohydrate moieties and the asparagine side chain [4]. Meanwhile, Hesperetin is a compound that can be found in Citrus sinensis (sweet orange), which is one of the flavonoids that contain flavanones, a group of phenolic natural chemical compound. It is rich in antioxidant as well as anti-microbial function that can inhibit the key proteins that are involved in the coronavirus infective cycle [5]. These compounds will be further analysed using in silico methods against the Spike protein, which is the basic requirement for the development of biomaterials involving substances that have been engineered, which subsequently can be used to directly control the interactions in the course of any therapeutic or diagnostic procedure.

In this study, protein-protein interactions (PPIs) between the Spike protein (antigen) and the known monoclonal antibodies were analysed, towards the development of antibody-based diagnostics that is designed to detect the presence of SARS-CoV-2 virus that caused the devastating Covid-19. The known monoclonal antibodies chosen are CR3022, M396 and 4C2, which are known antibodies for SARS-CoV-2, SARS-CoV and MERS-CoV, respectively [6][7]. For drug design, ligand-based and structure-based approaches were conducted to determine the potential inhibitors of Spike protein, by using NAG and Hesperetin as the query molecules to search for similar molecules that can bind to Spike protein with high binding affinity. Subsequently, the toxicity of these compounds were predicted using the ProTox-II webserver (https://tox-new.charite.de/protox_II/).

2. Materials and Methods

2.1 Programmes and databases

Table 1 lists the programmes and databases used during the in silico screening process.

| Programme / Database                  | Function                                      |
|--------------------------------------|------------------------------------------------|
| USR-VS                               | To screen for similar compounds or analogues of the query molecules. |
| AutoDock 4                           | To perform docking between the SARS-CoV-2 Spike protein with the analogues. |
| SWISSDOCK                            | To perform docking between the SARS-CoV-2 Spike protein with monoclonal antibodies. |
| USCF Chimera                         | To visualise the docking of monoclonal antibodies onto the SARS-CoV-2 Spike RBD-ACE2 protein. |
| ProToxII                             | To predict the toxicity of the compounds. |
| Protein Data Bank (PDB)              | To obtain the PDB file of the structure of the SARS-CoV-2 Spike protein. |
| ZINC15                               | To obtain the structures of the selected ligands that will be used for docking. |
| PubChem                              | To obtain the SD file of the structure of the SARS-CoV-2 Spike protein and the analogues. |
| STRING                               | To study protein-protein interactions between the Spike protein and the monoclonal antibodies. |

2.2 Protein-protein Interactions

The known monoclonal antibodies specific to Coronaviruses, which are M396 for SARS-CoV and 4C2 for MERS-CoV, as well as CR3022 for SARS-CoV-2 were evaluated by using the STRING webserver.
The STRING database features the largest number of 5090 organisms and 24.6 million proteins, which can be the benchmarked data sources, while providing intuitive and fast viewer for online use [8]. Chain H in M396 antibody, chain B in CR3022 antibody and Chain A in 4C2 antibody were found to have the highest combined score, which was determined prior to rigid molecular docking by using the SWISSDOCK webserver, in order to determine the binding energy between the SARS-CoV-2 Spike-ACE2 protein and the monoclonal antibodies.

### 2.3 Virtual Screening Analyses

The ligand-based method was conducted through USR-VS webserver by using NAG and Hesperetin as the query molecules. This program was able to screen 93.9 million 3D conformers of 23 million molecules using the ZINC database [9]. The similarity score was ranked based on the similarity of the screened compounds with NAG and Hesperetin and the top five compounds from each query molecule were chosen to be docked with the SARS-CoV-2 Spike protein. The structure-based method was performed with the top five compounds, where it was docked using the AutoDock4 software. The docking process was conducted to determine the binding energy between the Spike protein-compounds complexes. Finally, a toxicity test was performed for all selected compounds by using the ProTox-II webserver.

### 3. Results and Discussion

#### 3.1 The Interactions between the SARS-COV-2 Spike protein and the Monoclonal Antibodies

The currently available rapid test kits are antigen and antibody-based detection systems, where these can be used for initial detection of viral infection and after the incubation period of 14 days post-infection of SARS-CoV-2, respectively. The common method to design a diagnostic kit is by using the enzyme-linked immunosorbent assay (ELISA) and some examples of the current rapid test kit developed are Sofia SARS Antigen from QUIDEL company and SARS-CoV-2 LgG Antibody from Beckman Coulter company, apart from real-time reverse transcription–polymerase chain reaction (RT-PCR) diagnostics kits [10]. On the other hand, this study was conducted to determine the known monoclonal antibodies from previous coronaviruses that are able to neutralize the SARS-CoV-2 Spike protein using in silico approach by evaluating the binding energy form between the known monoclonal antibodies with the Spike protein.

Based on the interactive network analysed by the STRING database, it was shown that the top three highest combined score of protein interactions among the antibodies are chain H in M396 antibody, chain B in CR3022 antibody and chain A in 4C2 antibody, as tabulated in Table 2. The combined scores are calculated based on the functional association, which links between the known monoclonal antibodies contributed due to the specific biological function [8]. Thus, the binding energy was generated between the SARS-CoV-2 Spike protein’s RBD-ACE2 with the antibody, which shows that the M396 antibody exhibits the highest binding affinity towards the Spike protein, as shown in Figure 1, indicating the formation of stronger complexes that is ideal for antigen-antibody interactions. Therefore, the M396 antibody may well have potential to be used as the capture antibody in the diagnostics kit to selectively detect the SARS-CoV-2 Spike protein. However, the specificity of M396 antibody can be improved via antibody engineering for future diagnostics development for detection of Covid-19.

| Monoclonal antibody | Chain | Combined Score | Binding Energy (kcal/mol) |
|---------------------|-------|----------------|--------------------------|
| CR3022              | B     | 0.934          | -7.90                    |
| M396                | H     | 0.998          | -8.50                    |
| 4C2                 | A     | 0.903          | -7.67                    |
3.2 Virtual Screening of Compounds against the SARS-CoV-2 Spike Protein

Since there are no clinically proven treatment for Covid-19, only supportive symptomatic treatments such as oxygen therapy, steroids on a per case basis, fluid management and broad-spectrum antibiotics against the co or secondary bacterial infection were provided for patients [11]. Other than that, similar to any viral diseases, Covid-19 can be aided by therapeutic intervention, which can be extended to either direct acting antivirals (DAAs) or immunomodulatory adjuvants and substances [12]. Several drugs have been tested for repurposing to treat Covid-19, which can be divided into two categories; those that targeted the viral replication cycle and the ones that aim to control the symptoms of the diseases. One of the drugs tested is known as chloroquine, which were classically used as anti-malarial medication, meanwhile in Covid-19, it was found to be an effective antiviral agent that can inhibit SARS-CoV-2 infection [13]. Moreover, there are several clinical trials that have reported that the use of chloroquine can alter terminal glycosylation of the angiotensin-converting enzyme 2 (ACE2) receptor by suppressing the SARS-CoV-2 Spike protein binding, which notably reduces viral replication by blocking the virus fusion process [14].

Through virtual screening analyses, two potential inhibitors, which are ZINC 6614489 and Tamarixetin were shown to possess lower binding energy compared to the query molecules; NAG and Hesperetin respectively, as shown in Table 3. Each compound fulfilled the Lipinski’s Rule of Five, which is the rule of thumb in designing oral drugs. It was shown that each compound must consist of no more than 5 hydrogen bond donors, have no more than 10 hydrogen bond acceptors, molecular mass less than 500 Dalton and partition coefficient (log P) must not be greater than 5 [15]. The analyses showed that the binding energy for ZINC 6614489 is close to NAG, while Tamarixetin’s binding energy is lower than Hesperetin, which indicates the possibility of these potential inhibitors to bind strongly to SARS-CoV-2 Spike protein. The dockposes of both ZINC 6614489-Spike protein RBD and Tamarixetin-Spike protein RBD protein-ligand complexes are shown in

Figure 1: The interactions between the M396 antibody with the RBD-ACE2 region of the SARS-CoV-2 Spike protein.
Figure 2. The first potential inhibitor, ZINC 6614489 is a compound similar to NAG that possesses the binding energy of -4.92 kcal/mol, which is very close to the binding energy of NAG. Meanwhile, Tamarixetin is an analogue of Hesperetin and exhibited the binding energy of -7.93 kcal/mol, which is lower compared to Hesperetin (-7.43 kcal/mol). The well-established and potent inhibitor of Spike protein, Chloroquine was shown to have a binding energy of -7.53 kcal/mol, which means that Tamarixetin has higher potential to inhibit the SARS-CoV-2 Spike protein, compared to both Hesperetin and Chloroquine. Meanwhile, the toxicity prediction value (LD₅₀) is required for drug design evaluation in order to determine whether the compounds are safe for human use. The toxicity classes are categorised into six classes, where the lowest is Class 1, indicating that the compound is fatal if swallowed when the LD₅₀ value is less and equal to 5, while the highest is Class 6 that is non-toxic when the LD₅₀ value is above 5000. In this study, the toxicity prediction value (LD₅₀) for each compound falls under the category of class 5 within the range of (2000 < LD₅₀ ≤ 5000), which indicated that the compounds may be harmful for consumption. Therefore, further analysis is required to find the right amount of dose of the compounds that have potentials as drugs that are safe to be consumed. Hence in this study, virtual screening has successfully indicated that ZINC 6614489 and Tamarixetin could be the potential inhibitors of the SARS-CoV-2 Spike protein that can be further developed in the drug design pipeline for Covid-19 treatment.

Table 3: Potential inhibitors (ZINC 6614489 and Tamarixetin) obtained from virtual screening analyses along with the query molecules, NAG and Hesperetin, respectively. Also shown in the table is a repurposed drug that is currently used to treat Covid-19, Chloroquine.

| Compound Name | Chemical Structure | Binding Energy (kcal/mol) | ProTox-II Prediction (LD₅₀, mg/kg) | H-bond donor (<5) | H-bond acceptor (<10) | MW (<500Da) | XlogP (<5) |
|---------------|-------------------|---------------------------|-----------------------------------|------------------|----------------------|-------------|-----------|
| Chloroquine   | ![Chemical Structure](attachment:chloroquine.png) | -7.53 | - | 1 | 2 | 319.87 | 4.63 |
| NAG           | ![Chemical Structure](attachment:nag.png) | -5.17 | - | 5 | 6 | 221.21 | -1.72 |
| ZINC 6614489  | ![Chemical Structure](attachment:zinc.png) | -4.92 | 5000 | 5 | 7 | 221.209 | -3.03 |
| Hesperetin    | ![Chemical Structure](attachment:hesperetin.png) | -7.43 | - | 3 | 6 | 302.28 | 2.60 |

C₈H₁₇N₂O₄
C₁₆H₁₄O₆
C₁₈H₂₆CIN₄
C₈H₁₇N₂O₄
C₆H₁₄O₆
Tamarixetin

\[ \text{C}_{16}\text{H}_{12}\text{O}_{7} \]

-7.93

| 5000 | 4 | 7 | 316.265 | 1.99 |

Figure 2: a) The docked ZINC 6614489 ligand with SARS-CoV-2 Spike-RBD. b) The type of interactions formed during the docking process with the interacting amino acid residues. c) The docked Tamarixetin ligand with SARS-CoV-2 Spike-RBD. d) The type of interactions formed during the docking process with the interacting amino acid residues.

4. Conclusion
Protein-protein interactions between the SARS-CoV-2 Spike protein with M396 antibody has shown the strongest binding energy of -8.50 kcal/mol, suggesting that this antibody can be further improved in future antibody engineering and design studies, to be developed as a specific probe for rapid detection kit. Meanwhile, the virtual screening analyses, which were conducted to search for potential inhibitors of SARS-CoV-2 Spike protein had shown that Tamarixetin has the strongest binding energy compared to other screened compounds. Therefore, it is suggested that this compound has potentials to be designed as a future therapeutic drug. Overall, this study has computationally shown potential monoclonal antibody-Spike protein interactions, as well as potential candidates for rational drug design against the SARS-CoV-2 Spike protein, which have paved the path towards the control of Covid-19 transmission and future medication.

Acknowledgement
The authors would like to acknowledge the Department of Biotechnology Engineering, Kulliyyah of Engineering, IIUM for the support to conduct this study.

References
[1] Fani M, Teimoori A, and Ghafari S 2020. Comparison of the COVID-2019 (SARS-CoV-2)
pathogenesis with SARS-CoV and MERS-CoV infections. Future Virology
https://doi.org/10.2217/fvl-2020-0050.
[2] Weiss S and Navas-Martin S 2005. Coronavirus Pathogenesis and the Emerging Pathogen
Severe Acute Respiratory Syndrome Coronavirus. Microbiology and Molecular Biology Reviews, vol 69 pp 635–664. https://doi.org/10.1128/mmbr.69.4.635-664.2005.
[3] Bharath B, Damle H, Ganju S, and Damle L 2020. In silico screening of known small molecules
to bind ACE2 specific RBD on Spike glycoprotein of SARS-CoV-2 for repurposing against
COVID-19. F1000Research, https://doi.org/10.12688/f1000research.24143.1
[4] Wlodawer A, Dauter Z, Shabalin I and et al 2020. Ligand-centered assessment of SARS-CoV-2
drug target models in the Protein Data Bank FEBS J vol 287 pp 3703-3718.
[5] Bellavite P, and Donzelli A 2020. Hesperidin and SARS-CoV-2: New light on the healthy
function of citrus fruits. Antioxidants vol 9 pp 1–18.
[6] Jiang S, Hillyer C, and Du L 2020. Neutralizing Antibodies against SARS-CoV-2 and Other
Human Coronavirus. Trends Immunol., vol 41 pp 355.
https://doi.org/10.1016/j.it.2020.03.007.
[7] Choudhry H, Bakhrehab M, Abdulael W and et al 2019. “Middle East respiratory syndrome:
Pathogenesis and therapeutic developments,” Future Virol, vol 14 pp 237–246,
doi:10.2217/fvl-2018-0201.
[8] Szklarczyk D, Gable A, Lyon D and et al 2019. STRING v11: Protein-protein association
networks with increased coverage, supporting functional discovery in genome-wide
experimental datasets. Nucleic Acids Research, vol 47 pp 607–613.
https://doi.org/10.1093/nar/gky1131
[9] Schreyer A M and Blundell T 2012. USRCAT: Real-time ultrafast shape recognition with
pharmacophoric constraints. Journal of Cheminformatics, vol 4 pp 1–12.
https://doi.org/10.1186/1757-899X-4-27.
[10] Khalaf K, Papp N, Chou J, Hana D, Mackiewicz A, and Kaczmarek M 2020. SARS-CoV-2:
Pathogenesis, and Advancements in Diagnostics and Treatment. In Frontiers in Immunology,
vol 11 (Frontiers Media S.A) p 570927. https://doi.org/10.3389/fimmu.2020.570927
[11] Libster R, Pérez M, Wappner D and et al 2021. Early High-Titer Plasma Therapy to Prevent
Severe Covid-19 in Older Adults. New England Journal of Medicine, vol 384 pp 610–618.
https://doi.org/10.1056/nejmoa2033700.
[12] Nitulescu G, Paunescu H, Moschos S and et al 2020. Comprehensive analysis of drugs to treat
SARS-CoV-2 infection: Mechanistic insights into current COVID-19 therapies (Review). In
International Journal of Molecular Medicine vol 46 (Spandidos Publications) pp 467–488.
https://doi.org/10.3892/ijmm.2020.4608.
[13] Singh B, Ryan H, Kredo T, Chaplin M, and Fletcher T 2020. Chloroquine or
hydroxychloroquine for prevention and treatment of COVID-19. Cochrane Database
Syst Rev 2021. https://doi.org/10.1002/14651858.CD013587.pub2.
[14] Bolarin J, Oluwatoyosi M, Orege J and et al 2021. Therapeutic drugs for SARS-CoV-2
treatment: Current state and perspective. In International Immunopharmacology vol 90
(Elsevier B.V) p 107228. https://doi.org/10.1016/j.intimp.2020.107228.
[15] Banegas-Luna A J, Cerón-Carasso J P and Pérez-Sánchez H 2018. A review of ligand based
virtual screening web tools and screening algorithms in large molecular databases in the age
of big data. Future Medicinal Chemistry, vol 10 pp 2641–2658. https://doi.org/10.4155/fmc
2018-0076.