BMEG 250: Cellular Physiology & Biophysics

Project Report

Table of Contents

A. Background

I. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) 1.

II. Virology 1.

III. Transmission 1.

B. Introduction

I. Problem Statement 1.

II. Significance 2.

III. Objective 2.

C. Methodology

I. Sequence Retrieval 2.

II. Multiple Sequence Alignment 3.

III. Homology Modelling 6.

IV. Residue Retrieval 8.

V. Protein-Protein Docking 9.

VI. Gibbs Free Energy 9.

D. Results

I. Verification 10.

II. Evaluation 10.

III. Discussion 10.

E. Conclusion 10.

F. References 11.

A. Background

I. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2)

Severe Acute Respiratory Syndrome Coronavirus 2, abstracted as SARS-COV-2, is a strain of the coronavirus family which is responsible for the origination and subsequent propagation of the Coronavirus Disease 19, colloquially referred to as COVID-19. The first documented case of COVID-19 was identified in Wuhan, China in November 2019, which resulted in a widespread, global outbreak, that was declared a 'pandemic' by the World Health Organization in March 2020.

III. Virology

SARS-COV-2 is a positive-sense, single-stranded virus which is highly contagious within humans. Each SARS-COV-2 virion comprises of four structural polypeptides: the S (spike), E (envelope), M (membrane), and N (nucleocapsid) proteins [1]. The Ribonucleic Acid (RNA) genome of the coronavirus is contained within the N (nucleocapsid) protein, whereas the 'viral envelope' of the coronavirus is formulated by the S (spike), E (envelope) and M (membrane) proteins [2].

III. Transmission & Infection

SARS-COV-2 is an 'airborne' virus that can be transmitted through close contact, as well as through exposure to infectious respiratory fluids, such as aerosols and respiratory droplets produced in the sinuses. Having entered the respiratory tract, SARS-COV-2 then encounters cells of the respiratory epithelium, trachea, and lungs, as the S (spike) protein recognizes the Angiotensin-Converting Enzyme 2 [3]. Attachment of SARS-COV-2 to the obligate receptor protein is catalyzed by the S1 subunit of the S (spike) protein, whereas the fusion of SARS-COV-2 with the membrane of the host cell is facilitated by the S2 subunit of the S (spike) protein following dynamic, conformational changes. The viral Ribonucleic Acid (RNA) genome of SARS-COV-2, which is encapsulated within the N (nucleocapsid) protein, is deposited within the host cell, where it triggers a signalling cascade for the production of more variants in the body [4].

B. Introduction

I. Problem Statement

The COVID-19 pandemic, propagated by SARS-COV-2, is noted to be amongst the most virulent and morbid medical outbreaks in history, with over 517 million cases and 6.25 million deaths recorded to date [5]. To boost immunization against SARS-COV-2, mRNA and viral vectors were genetically engineered to develop COVID-19 vaccines, which were administered en masse globally. However, SARS-COV-2 mutates at least once every week, which could expedite the emergence of new variants. Although 40% of the mutations hinder the virus' survivability, nonetheless, the high mutability rate of SARS-COV-2, could, significantly undermine the efficacy of the COVID-19 vaccine, and correspondingly, overburden the healthcare system [6]. Therefore, it is critical that the scientific community coordinates its efforts towards enhancing the potency of COVID-19 vaccinations by undertaking a better understanding of the SARS-COV-2 variants.

II. Objective

To classify the transmissibility of SARS-COV-2 by determining the binding affinities of the S (spike) protein for the Original, Delta, and Omicron variants with the ACE-2 obligate receptor.

III. Significance

The following project would provide meaningful insights into the contagiousness of the Original, Delta, and Omicron variants for SARS-COV-2. This would inform decision-making for the development of COVID-19 vaccinations and induce greater immunity against SARS-COV-2. Ultimately, the timely dissemination of COVID-19 inoculations would minimize loss of life, while ensuring that the world emerges sooner from the socio-economic hardships of the pandemic [7].

B. Methodology

The following steps were undertaken to compute, obtain, and analyze the SARS-COV-2 lineages:

I. Sequence Retrieval

Sequences for the Original, Delta, and Omicron variants of SARS-COV-2 were extracted from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), an open-source database which stores the three-dimensional configurations of polypeptide molecules. Since a convenient representation was required to smoothly process the nucleotide sequences for SARS-COV-2 variants, the text-based FASTA format was considered appropriate for this purpose.

```
>sp|P0DTC2|SPIKE_SARS2 Spike glycoprotein OS=Severe acute respiratory
          syndrome coronavirus 2 OX=2697049 GN=S PE=1 SV=1

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTK
RFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWME
SEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPI
GINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTV
EKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPT
KLNDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLK
PFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCV
NFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVN
CTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSII
AYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGI
AVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAAR
DLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIA
NQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRL
QSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTA
PAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKY
FKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTI
                MLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
```

Figure 1: *FASTA Sequence of the Original SARS-COV-2 Strain.*

>UNJ26567.1 surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]

MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTT
RFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDVYYHKNNKSWME
SGVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGI
NITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEK
GIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKL
NDLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPF
ERDISTEIYQAGSKPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNF
NFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCT
EVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSRRRARSVASQSIIAY
TMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV
EQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDL
ICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQ
FNSAIGKIQDSLSSTASALGKLQNVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQS
LQTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPA
ICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFK
NHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIML
CCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT

Figure 2: *FASTA Sequence of the Delta SARS-COV-2 Variant.*

>UNK07387.1 surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]

MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRF
DNPVLPFNDGVYFASIEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLDHKNNKSWMESEFRV
YSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVREPEDLPQGFSALEPLVDLPIGIN
ITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKG
IYQTSNFRVQPTESIVRFPNITNLCPFDEVFNATRFASVYAWNRKRISNCVADYSVLYNLAPFFTFKCYGVSPTKLN
DLCFTNVYADSFVIRGDEVRQIAPGQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKVSGNYNYLYRLFRKSNLKPFE
RDISTEIYQAGNKPCNGVAGFNCYFPLRSYSFRPTYGVGHQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFN
FNGLKGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTE
VPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTKSHRRARSVASQSIIAYT
MSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLKRALTGIAVE
QDKNTQEVFAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLI
CAQKFKGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQF
NSAIGKIQDSLSSTASALGKLQDVVNHNAQALNTLVKQLSSKFGAISSVLNDIFSRLDKVEAEVQIDRLITGRLQSL
QTYVTQQLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAI
CHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKN
HTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLC
CMTSCCSCLKGCCSCGSCCKFDEDDSEPLLKGVKLHYT

Figure 3: *FASTA Sequence of the Omicron SARS-COV-2 Variant.*

II. Multiple Sequence Alignment

Sequences for the Original, Delta, and Omicron variants of SARS-COV-2 were subjected to an ‘alignment’, based on local similarity information and domain conservation. T-Coffee, a bioinformatics tool for multiple sequence alignment, was selected for this purpose, as its ideal for processing shorter sequences and yields better performances over progressive alignment methods.

```
SCORE=997
*
 BAD AVG GOOD
*
P0DTC2          :  99
UNJ26567.1      :  99
UNK07387.1      :  99
cons            :  99

P0DTC2          MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIH
UNJ26567.1      MFVFLVLLPLVSSQCVNLRTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIH
UNK07387.1      MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHV--

cons            ********************* **************************************************.

P0DTC2          VSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCND
UNJ26567.1      VSGTNGTTRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCND
UNK07387.1      ISGTNGTKRFDNPVLPFNDGVYFASIEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCND

cons            :*******.****************** ***********************************************

P0DTC2          PFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKH
UNJ26567.1      PFLDVYYHKNNKSWMES--GVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKH
UNK07387.1      PFLD---HKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKH

cons            ***.   ***********  **************************************************

P0DTC2          TPINL--VRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRT
UNJ26567.1      TPINL--VRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRT
UNK07387.1      TPIIVREPEDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRT

cons            ***:  .*************************************************************

P0DTC2          FLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN
UNJ26567.1      FLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFN
UNK07387.1      FLLKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFDEVFN

cons            ************************************************************.****

P0DTC2          ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAP
UNJ26567.1      ATRFASVYAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAP
UNK07387.1      ATRFASVYAWNRKRISNCVADYSVLYNLAPFFTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAP

cons            ***************************** *.* ************************************

P0DTC2          GQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCN
UNJ26567.1      GQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYRYRLFRKSNLKPFERDISTEIYQAGSKPCN
UNK07387.1      GQTGNIADYNYKLPDDFTGCVIAWNSNNLDSKVSGNYNYLYRLFRKSNLKPFERDISTEIYQAGNKPCN

cons            ****:*****************************._***** ************************.._***

P0DTC2          GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTG
UNJ26567.1      GVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTG
UNK07387.1      GVAGFNCYFPLRSYSFRPTYGVGHQPYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLKGTG

cons            ** ********:**.*:** ***:*****************************************._***

P0DTC2          VLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTE
UNJ26567.1      VLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTE
UNK07387.1      VLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTE

cons            ************************************************************._*****
```

```
P0DTC2          VPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVA
UNJ26567.1      VPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSRRRARSVA
UNK07387.1      VPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDIPIGAGICASYQTQTKSHRRARSVA

cons            ************************************:**********************:* *******


P0DTC2          SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYG
UNJ26567.1      SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYG
UNK07387.1      SQSIIAYTMSLGAENSVAYSNNSIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYG

cons            ************************************************************************


P0DTC2          SFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKV
UNJ26567.1      SFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKV
UNK07387.1      SFCTQLKRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKYFGGFNFSQILPDPSKPSKRSFIEDLLFNKV

cons            ******:****************************** ******************************


P0DTC2          TLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ
UNJ26567.1      TLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ
UNK07387.1      TLADAGFIKQYGDCLGDIAARDLICAQKFKGLTVLPPLLTDEMIAQYTSALLAGTITSGWTFGAGAALQ

cons            ******************************:*****************************************


P0DTC2          IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVK
UNJ26567.1      IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQNVVNQNAQALNTLVK
UNK07387.1      IPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNHNAQALNTLVK

cons            ************************************************************:***:**********


P0DTC2          QLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV
UNJ26567.1      QLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV
UNK07387.1      QLSSKFGAISSVLNDIFSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECV

cons            ****:***********:************************************************************


P0DTC2          LGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHW
UNJ26567.1      LGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHW
UNK07387.1      LGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHW

cons            ************************************************************************


P0DTC2          FVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISG
UNJ26567.1      FVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISG
UNK07387.1      FVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISG

cons            ************************************************************************


P0DTC2          INASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSC
UNJ26567.1      INASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSC
UNK07387.1      INASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSC

cons            ************************************************************************


P0DTC2          CSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
UNJ26567.1      CSCLKGCCSCGSCCKFDEDDSEPVLKGVKLHYT
UNK07387.1      CSCLKGCCSCGSCCKFDEDDSEPLLKGVKLHYT

cons            ************************:*********
```

Figure 4: *Multiple Sequence Alignment of Original, Delta and Omicron SARS-COV-2 Variants.*

III. Homology Modelling

Sequences for the Original, Delta, and Omicron variants of SARS-COV-2 were processed over a homology-modelling server to obtain three-dimensional polypeptide configurations. The automated protein-structure developer, SWISS MODEL, which is accessible via the web-based platform, ExPASy, utilizes experimentally determined complexes as its basis to generate reliable structural models. Since a consistent representation was required to capture the structure of the individual nucleic acids, the text-based PDB format was considered appropriate for this purpose. Consequently, to visualize the three-dimensional atomic-resolution images generated for each SARS-COV-2 lineage, an open-source molecular-editing system, PyMol, was utilized to do so.

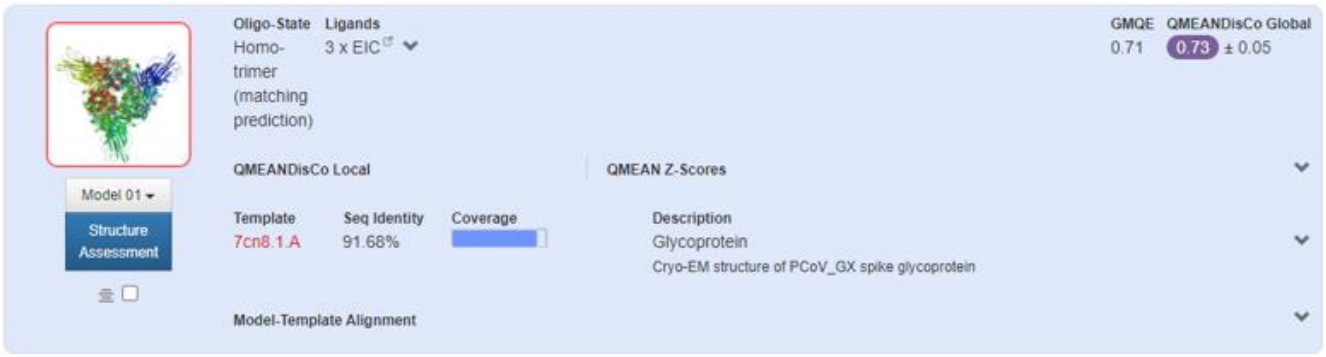

Figure 5.1: *QMEAN Distance Constraint Score for the Original SARS-COV-2 Strain.*

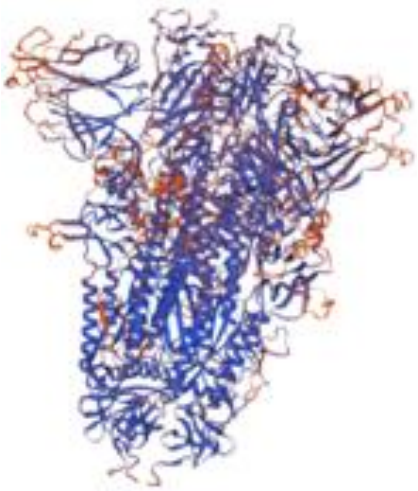

Figure 5.2: *Three-Dimensional Homology Model of the Original SARS-COV-2 Strain.*

Figure 6.1: *QMEAN Distance Constraint Score for the Delta SARS-COV-2 Variant.*

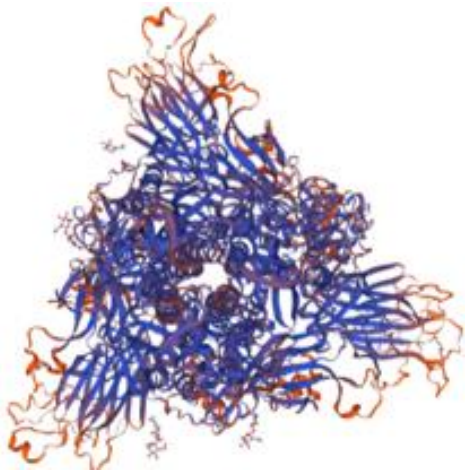

Figure 6.2: *Three-Dimensional Homology Model of the Delta SARS-COV-2 Variant.*

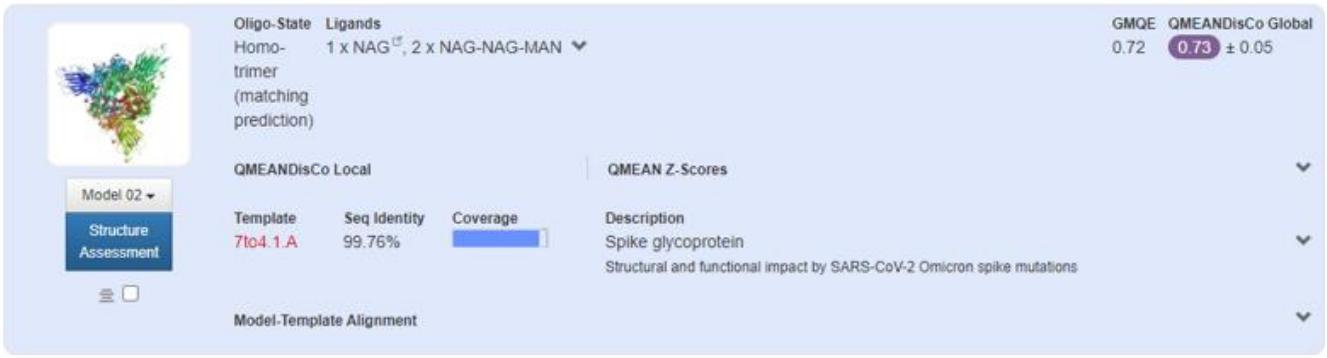

Figure 7.1: *QMEAN Distance Constraint Score for the Omicron SARS-COV-2 Variant.*

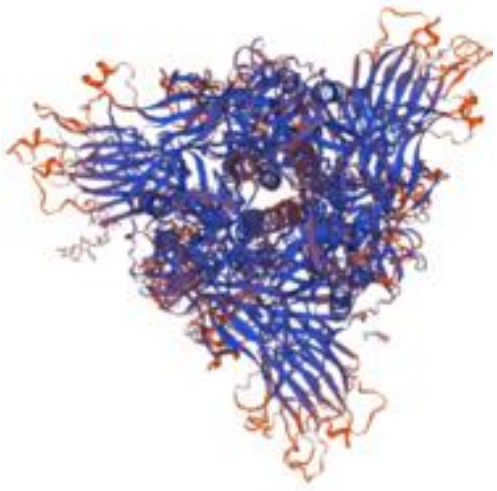

Figure 7.2: *Three-Dimensional Homology Model of the Omicron SARS-COV-2 Variant.*

IV. Residue Retrieval

Residues for the S (spike) protein pertaining to the Original SARS-COV-2 strain was acquired from UNIPROTKB, an open-source database that provides access to protein sequence information. Meanwhile, residues for the S (spike) protein related to the Delta, and Omicron variants of SARS-COV-2 were derived from PUBMED, a search-engine based medical databank operated by the National Center for Biotechnology Information (NCBI). Therefore, to allow for polypeptide-based docking of the SARS-COV-2 lineages with the ACE-2 receptor, the residues for the S (spike) protein were demarcated separately with a different color for easier identification in PyMol.

Figure 8: *Residues of the ACE-2 Receptor (Red) for the SARS-COV-2 Protein Complex.*

V. Protein-Protein Docking

Macromolecular docking was performed to procure the quaternary structure for the biological complex following ligand-receptor interaction. The Original, Delta, and Omicron variants of SARS-COV-2 were selected as the ligand, respectively, whereas the ACE-2 input molecule was chosen as the receptor. HDOCK, a web-based server that implements a hybrid algorithm with template-based modeling and ab-initio docking, was used to obtain the macromolecular structure.

Spike protein S1 (PRO_0000449647)

With	#Exp.	IntAct
ACE2 [Q9BYF1] from Homo sapiens.	2	EBI-25490323,EBI-7730807

Figure 9: *Interaction of the SARS-COV-2 Spike Protein with the ACE-2 Receptor.*

VI. Gibbs Free Energy Calculation

Docking scores, which represent the Gibbs Free Energy, were obtained for the binding residues of the ligand, the respective lineages of SARS-COV-2, and the receptor, the ACE-2 input molecule. The ligand-receptor affinity is derived from the spontaneity of binding, which is higher for a *negative* Gibbs Free Energy value in relation to a *positive* Gibbs Free Energy value [8]. Hence, a high spontaneity of binding is indicative of a more stable interaction for the Original, Delta, and Omicron variants of SARS-COV-2 and the Angiotensin-Converting Enzyme 2 obligate receptor [9]. Thus, the SARS-COV-2 lineage with a more stable interaction would be hypothesized to have a relatively higher binding affinity with the ACE-2 receptor, and can, thus, be considered to be more infectious compared to other SARS-COV-2 variants which are responsible for COVID-19.

Table 1: *Gibbs Free Energy of Binding for Original, Delta, and Omicron SARS-COV-2 Variants.*

SARS-COV-2 Variant	Docking Score
Original	*-70.90*
Delta	*-115.27*
Omicron	*-152.39*

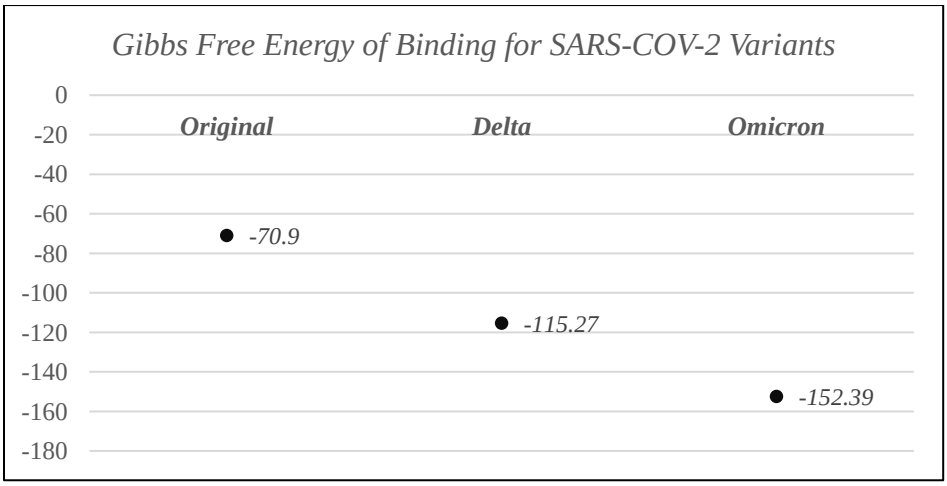

Figure 10: *Graphical Representation of Gibbs Free Energy of Binding for SARS-COV-2.*

D. Results

I. Verification

Data for Gibbs Free Energy corresponding to the Original, Delta, and Omicron Variants for SARS-COV-2 was, subsequently, extracted from wet-laboratory experiments. The values obtained substantiated the results for the computational stimulation, thus, corroborating the hypothesis for the investigation that pertains to the binding affinities of SARS-COV-2 with the ACE-2 receptor.

II. Evaluation

The Docking score for the SARS-COV-2 and ACE-2 ligand-receptor complex was acquired from the open-source HDOCK server, and corresponds to the Gibbs Free Energy of Binding, in kcal/mol (*Table 1*). The binding affinity of the Original SARS-COV-2 strain is the least 'negative', at *-70.90 kcal/mol*, followed by the Delta variant, at *-115.27 kcal/mol*. Meanwhile the Omicron variant is the most 'negative', with a value of *-152.39 kcal/mol* compared to the other lineages.

III. Discussion

As denoted earlier, a more 'negative' Gibbs Free Energy of Binding correlates to a stronger binding affinity between the SARS-COV-2 variant and the ACE-2 receptor protein. Therefore, the Omicron variant can be considered to have a higher spontaneity of binding, and correspondingly, a more stable interaction with the ACE-2 receptor protein due to its more 'negative' value for the Gibbs Free Energy, followed by the Delta variant and the Original SARS-COV-2 strain. Further, due to its greater binding affinity within the ligand-receptor complex, the Omicron variant is considered to have a higher infectivity, compared to the Delta and Original SARS-COV-2 lineages.

E. Conclusion

The investigation demonstrated that in-silico protein sequencing via homology modelling is a viable method in bioinformatics to obtain valuable information pertaining to the binding affinities of macromolecular biological structures. As observed, the dry-laboratory experimentation yielded accurate results for the spontaneity of binding for the Original, Delta and Omicron SARS-COV-2 variants with the ACE-2 receptor. To ensure that the insights acquired were replicated, the controlled trial was stimulated again, this time within a wet-laboratory environment. Further, as mentioned previously, the experimental values extracted for the Gibbs Free Energy of Binding were similar for the two different scenarios, which adds to the robustness of the investigation.

Nonetheless, the insights drawn from the experimentation cannot necessarily be abstracted to all subsequent lineages of SARS-COV-2 since the investigation was only concerned with the Original, Delta and Omicron variants. Hence, future iterations of the experimentation would focus upon performing the trails for other, dominant lineages of SAR-COV-2. Patterns can then be derived to model the behavior of downstream mutations for SARS-COV-2. However, given the time-consuming nature of conducting the dry-laboratory simulations, it is incumbent upon the scientific community to refine the hybrid algorithm for polypeptide-based docking. Once this is accomplished, three-dimensional protein configurations can then be synthesized, extracted, and evaluated, remotely, thus providing researchers with rapid access to computational biology tools.

F. References

1. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses (April 2020). The Species Severe Acute Respiratory Syndrome-Related Coronavirus: Classifying 2019-Ncov and Naming it SARS-Cov-2. *Nature Microbiology*. 5 (4): 536–544. doi:10.1038/s41564-020-0695-z.

2. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. (May 2020). Analysis of Therapeutic Targets for SARS-Cov-2 and Discovery of Potential Drugs by Computational Methods. *Acta Pharmaceutica Sinica B*. 10 (5): 766–788. doi:10.1016/j.apsb.2020.02.008.

3. Centers for Disease Control and Prevention (April 2020). *How Coronavirus Spreads*. Wayback Machine.

4. Zhao P, Praissman JL, Grant OC, Cai Y, Xiao T, Rosenbalm KE, et al. (October 2020). Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host & Microbe*. 28 (4): 586–601.e6. doi:10.1016/j.chom.2020.08.004.

5. Aronson J.K. (March 2020). Coronaviruses: A General Introduction. *Centre for Evidence-Based Medicine, Nuffield Department of Primary Care Health Sciences, University of Oxford*.

6. Bordoli, L., Kiefer, F., Arnold, K. (December 2008). Protein Structure Homology Modeling Using SWISS-MODEL. *Nat Protoc* 4, 1–13. https://doi.org/10.1038/nprot.2008.197

7. Zoumpourlis V, Goulielmaki M, Rizos E, Baliou S, Spandidos DA (October 2020). The COVID-19 Pandemic as a Scientific and Social Challenge in the 21st Century. *Molecular Medicine Reports*. 22 (4): 3035–3048. doi:10.3892/mmr.2020.11393.

8. Popovic, M., & Minceva, M. (2020). Thermodynamic Insight into Viral Infections 2: Empirical Formulas, Molecular Compositions and Thermodynamic Properties Of SARS, MERS and SARS-Cov-2 (COVID-19) Viruses. *Heliyon*, *6*(9), e04943. https://doi.org/10.1016/j.heliyon.2020.e04943

9. de Andrade, J., Gonçalves, P. F., & Netz, P. A. (2020). *How Does the Novel Coronavirus Interact with the Human ACE2 Enzyme? A Thermodynamic Answer*. https://doi.org/10.26434/chemrxiv.12361238