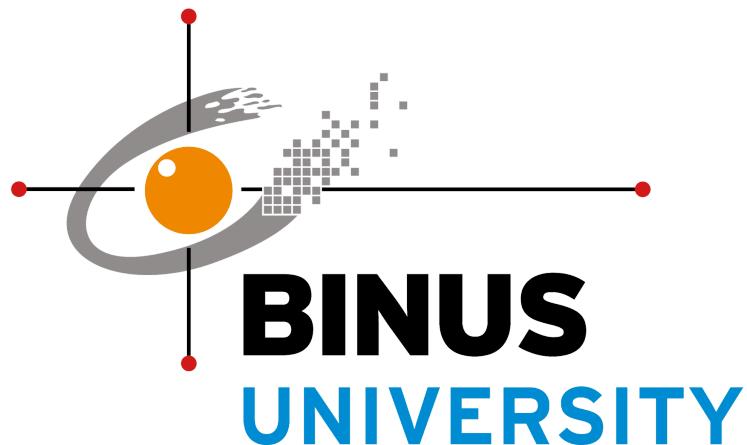


**Protein Modelling and Prediction of Capsaicin as an Anti-Viral Agent  
Against SARS-CoV-2**

**Project Report  
*Computational Biology*  
Dr. Dwiyantari Widyaningrum, S.Si., M.Si.  
D6158**



**LU01  
Group 8**

TISHA JILLIAN	(2602078410)
TARA NIRMALA KUSUMA	(2602084893)
JONATHAN ALVINDO FERNANDI	(2602089143)
VINSON LUCKIANTO	(2602091066)
ARYA MAULANA	(2602209961)

Bina Nusantara University  
2024

## TABLE OF CONTENT

<b>1. Background.....</b>	<b>3</b>
<b>2. Methodology.....</b>	<b>4</b>
2.1 FASTA Sequence.....	4
2.2 3D Protein Modeling.....	7
2.3 Molecular Docking.....	10
<b>3. Result.....</b>	<b>23</b>
3.1 FASTA Format of Amino Acid Sequence.....	23
3.2 Prediction of Protein 3D Structure.....	24
3.2.1 Spike Protein S1 SARS-CoV-2.....	24
3.2.2 RNA-dependent RNA polymerase (RdRp) SARS-CoV-2.....	26
3.3.3 Envelope Protein (E) Sars-CoV-2.....	28
3.3 Molecular Docking Outputs.....	29
3.3.1 Docking Results using AutoDock Vina.....	29
3.3.1 Docking Poses and Interactions Between the Ligand and the Protein Visualization Using PyMOL.....	31
<b>4. Discussion.....</b>	<b>33</b>
4.1 Spike Protein.....	33
4.2 RdRp Protein.....	33
4.3 Envelope Protein.....	33
4.4 Potential Implications.....	33
4.5 Limitations and Future Directions.....	34

## **1. Background**

Proteins are essential molecules in living organisms and they perform crucial functions, serve as structural building blocks, and carry out various processes with cells. A protein's function is linked to its three-dimensional (3D) structure.

Computational techniques are used to bridge the gap between sequences and structures. A 3D protein's structure that was made from predicted amino sequences using a computational method is called protein modelling. This process involves several steps such as identifying and selecting a structure template from the Protein Data Bank (PDB), alignment of the query sequence to achieve high-quality sequence alignment, building a 3D model using different methods for instance, rigid-body assembly, segment matching, spatial restraint, and artificial evolution, the last steps were loop modelling to determine the protein's function and structure.

Protein modelling is a crucial step for making 3D protein structures and later will be used for molecular docking. Molecular docking itself is a method to predict the preferred affinity of ligands within the binding site of a protein. This technique is used in structure-based drug design as it helps predict the binding conformation of small molecule ligands to their binding sites hence facilitating the rational design of drugs.

Protein modelling and molecular docking are crucial tools to analyze the potential of Indonesian herbs as antiviral agents. By understanding the interactions between herbal compounds and the target proteins that are essential for identifying effective antiviral compounds.

SARS-COV-2 also known as Severe Acute Respiratory Syndrome - coronavirus is a strain of coronavirus that causes COVID-19. First identified in Wuhan, China, coronavirus has been a global pandemic that causes a lot of fatalities as it has high mortality rates. The relation between SARS-COV-2 and Indonesian herbs is that several Indonesian plants and their compounds have been identified as potential agents for preventing or reducing SARS-COV-2 infection. These compounds have been found to bind specific therapeutic targets of the virus, such as ACE-2 receptor, spike protein, and protease that could help restrict the virus's entry into human cells and replication. Some of the key compounds identified include emodin and luteolin, hesperidin, curcumin, and trans-cinnamaldehyde; these compounds have been identified through various methods, including computational models and molecular docking.

One of the Indonesian herbs that can be used as an anti-virus is capsaicin, capsaicin as an active component of chilli peppers has been found to exhibit antiviral properties against various viruses, including SARS-COV-2. A key point about the antiviral actions of capsaicin against SARA-COV-2 is capsaicin has been shown to bind strongly to viral 3C like protease by enzyme folding, this binding suggests that capsaicin could be a candidate as a model drug for novel treatment against SARS-COV-2.

## 2. Methodology

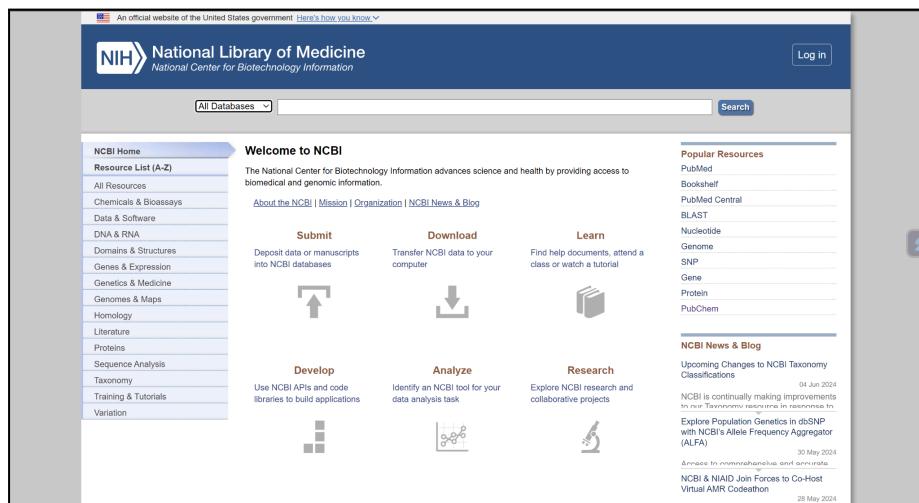
### 2.1 FASTA Sequence

The National Center for Biotechnology Information (NCBI) databases are essential resources for obtaining protein FASTA information. NCBI gets its information from databases, such as GenBank, and RefSeq. The information about protein sequences is submitted by researchers worldwide. NCBI's search system allows users to find specific protein information using various query parameters like gene names, accession numbers, or organism names.

The steps of retrieving protein sequence FASTA format from the NCBI database:

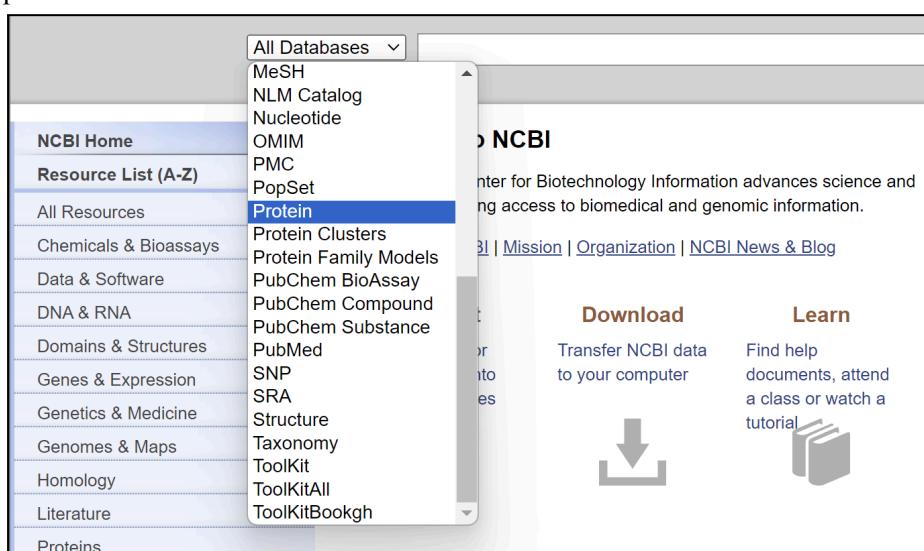
#### Step 1: Access the NCBI Website

To access various databases, open the NCBI homepage by visiting the [National Center for Biotechnology Information](#) website.



#### Step 2: Select Database

On the search bar section, there is a dropdown menu that allows you to select a specific database. Select “Protein” from the dropdown. So the database information given from the search only gives protein information.



### Step 3: Search Query

Type the keyword query, which can be a gene name, protein name, accession number, or protein FASTA format. Click the “Search” button to begin searching

The screenshot shows the National Library of Medicine's Protein search interface. The search term 'RNA-dependent RNA polymerase sars-cov-2' is entered in the search bar. The results page displays a summary of 320 species, 548 PDB entries, and 3,882,327 RefSeq entries. A specific result for 'ORF1ab – ORF1a polyprotein;ORF1ab polyprotein' is highlighted, showing it is processed from ORF1ab, has a Gene ID of 43740578, and is associated with 1 RefSeq protein, 1 RefSeq genome, and 226 PubMed entries. The page also includes sections for RefSeq Proteins, NCI Virus, SARS-CoV-2 proteins, and search details.

### Step 4: Search Results

From the search results, click the title of the protein to view detailed information.

This screenshot shows the detailed view of the protein 'ORF1ab – ORF1a polyprotein;ORF1ab polyprotein'. It includes the gene name, gene ID (43740578), and links to RefSeq proteins, RefSeq genome, and PubMed. A 'Sequence Viewer' button is also present. The page also features a 'Was this helpful?' rating system with thumbs up and thumbs down icons.

## Step 5: FASTA Format

In the detailed information, there is a link to “FASTA”. Click it to access the FASTA sequence format.

GenPept ▾ Send to

### RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]

NCBI Reference Sequence: YP\_009725307.1  
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to: ▾

Locus YP\_009725307 932 aa linear VRL 18-JUL-2020  
Definition RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2].  
Accession YP\_009725307  
Version YP\_009725307.1  
DBLink BioProject: [PRJNA485481](#)  
DBSource RefSeq: accession [YP\\_009724389.1](#)  
Keywords RefSeq.  
Source Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)  
Organism [Severe acute respiratory syndrome coronavirus 2](#)  
Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirinae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus; Severe acute respiratory syndrome-related coronavirus.  
Reference 1 (residues 1 to 932)  
Authors Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y., Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H., Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.  
Title A new coronavirus associated with human respiratory disease in China  
Journal Nature 579 (7798), 265-269 (2020)  
PubMed [32015508](#)

## Step 6: View the FASTA Sequence

There is a protein sequence in FASTA format. The format starts with a “>” symbol with an amino acid sequence.

RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]

NCBI Reference Sequence: YP\_009725307.1  
[GenPept](#) [Identical Proteins](#) [Graphics](#)

>YP\_009725307.1 RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]  
SADAQSFLNRVCVSAARLTPCGTGTSTDVVYRAFDIYNDKVGFAKFLKTNCCRQEKDEDDNLIDSYF  
VVKRHTFSNYQHEETIYNLLKDCAVAKHDFKFIDGDMVPHISRQLTKYTADLVYALRFDEGNCD  
TLKEILVTYNNCCDDYFNKKDWYDFVENPDILRVYANLGERVRQALLTVQFCDAQRNAGIVGVLTLDNQ  
DLNGNWYDFGDFIQTTPGSGVPVDSSYYSLMPILTLTRALTAESHVDTDLTKPYIKWDLKYDFTTEERL  
KLFDRYFKYWDQTYHPNCVNCLDDRCLILHCANFNLFSTVFPPTSFGPLVRKIFVDGVPFVSTGYHFR  
LGVVHNQDVNLHSSRLSFKELLVYAADPAMHAASGNLLDKRTTCFSVAALTNNVAFQTVKPGNFNKDFY  
DFAVSKGFFKEGSSVELKHFFAQDGNAISDYDYYRYNLPTMCDIRQLLFVVEVVDKYFDCYDGGCINA  
NQVIVNNLDKSAGFPFNKGKARLYYDSMSYEDQDALFAYTKRNVIPTITQMNLYKAISAKNRARTVAGV  
SICSTMNRQFHQKLLKSIATRGATVVIGTSKFYGGWHNMLKTVYSDVENPHLMGWDPKCDRAMPNML  
RIMASLVLARKHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGTSSGDATTAYANSVFNICQAV  
TANVNALLSTDGNKIADKYVRNLQHRLYECLYRNRDVTDVFNEFYAYLRKHFSMMILSDDAVVCFNSTY  
ASQGLVASIKNFKSVLYQQNNVFMSEAKCWETDLTKGPHEFCSQHTMLVKQGDDYVYLPPDPSRILGA  
GCFVDDIVKTDTLMIERFVSLAIDAYPLTKHPNQEYADVFLYLYQYIRKLHDLTGHMLDMYSVMLND  
NTSRYWEPEFYEAAMYTPHTVLQ

## 2.2 3D Protein Modeling

Swiss-Model is a valuable tool for protein structure modelling due to its ease of use, comprehensive template library, automated processes, quality assessment features, and integration capabilities. It empowers researchers to generate reliable 3D models of proteins, facilitating a deeper understanding of their biological functions and applications in various fields of life sciences.

The steps of retrieving the 3D protein structure model using the Swiss-Model:

### Step 1: Retrieve the FASTA format of the protein

Obtain the amino acid sequence of the target protein that wants to be modelled. The sequence needs to be in FASTA format, a text-based format that represents protein sequences. The FASTA format can be obtained from NCBI databases.

**spike protein [Severe acute respiratory syndrome coronavirus 2]**

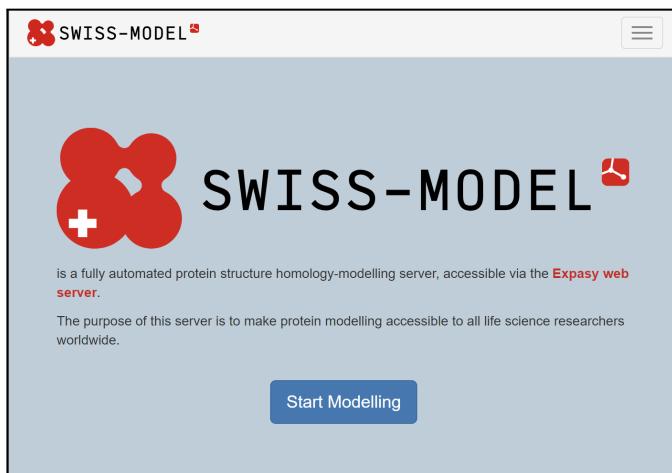
GenBank: QIH45093.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>QIH45093.1 spike protein [Severe acute respiratory syndrome coronavirus 2]
MFVFLVLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFPLFFSNVTWFHAIHV
SGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVEFQFCNDPF
LGVYYHKNNKSWMESEFRVYSSANCTFEVVSQPFLMDLEGKQGNFKNLEREFVFKNIDGYFKIYSKHTPI
NLVRDLPQGFSALEPLVLDLPIGINITRFQTLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYN
ENGTTTDAVDCAALDPSETKCTLKSFTVEKGIVQTTSNFRVOPTESIVRFPNTINLCPCGEVFNAUTRFAKV
YAWNRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRIAPGQTGKIAD
NYYKLPLDFTCVIAWNSNLDSKVGGNNYLYRLFRKSNLKPFERDISTEYIQAGSTPCNGVEGFCVYF
PLQSYGFQPTNGVYQPYRVVVLSELLHAPATVCGPKNSTLNVKNCVNFNFNLGTGVLTESNKFL
PFFQFGRDIADTTDAVRDQPTLEILDTPCSFGGSVITPGTNTSNOAVLYQDVNTEVPVVAIHDQLT
PTWRVYSTGSNVFQTRAGCLIGAEHVNNSYECDIP1GAGICASYQTQTNSPRRARSVASQSIAYTMSLG
AENSVAYSNNSLIAPTNTFISVTTEILPVSMKTKTSVDCTMICGDSTECNSNLLQYGSFCQLNRLALTGI
AVEQDKNTQEVAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIQYGDC
LGDIARDLICAQKFNGLTVPPLLTDEMTIAQTSALLAGITTSGWTFGAGAALQIPFAMQOMAYRFNGIG
VTQNVLYENQKLIQNSAIGKIQDSSLSTASALGKLQDVNNQNAQALNTLVKQLSSNFGAISSSVLNDI
LSRLDKVEAEVQIDRLITGRQLSQLQTVYTQQLTRAEEIRASANLAATKMECVLQSKRVDFCGKYHLM
SFPQSAPHGVFLHVTVYVPAQEKNFTTAPAIHDGKAHFREGVFSNGTHWFVTQRNFYEPQIITTDNT
FVSGNCDVIGIVNNTVYDPLQPELDSFKEELDKYFKNHSTSPDVLGDISGINASVNIQKEIDRLNEVA
KLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMTIMLCCMTCCSCLKGCCSGSCCKFDEDD
SEPVLKGVKLHYT
```

### Step 2: Access Swiss-Model

Open the web browser and go to the Swiss-Model website at [Swiss-Model](#). Click on “Start Modeling” to start a new project.



### Step 3: Input the Sequence

Paste the protein sequence in FASTA format into the provided text box. The project can be named in the project title text box.

The left screenshot shows the 'Start a New Modelling Project' page. The 'Target Sequence(s)' field contains the FASTA sequence: `MHEVFLVLLPLVSSQCVNL TTRQLPAPAYINSFTRGVYYPDVKFRSSVLHS`. Below it, the 'Project Title:' field is set to 'Untitled Project'. The right screenshot shows the same page with the 'Project Title:' field explicitly named 'Spike Protein s1 Sars-CoV-2'.

### Step 4: Template Search

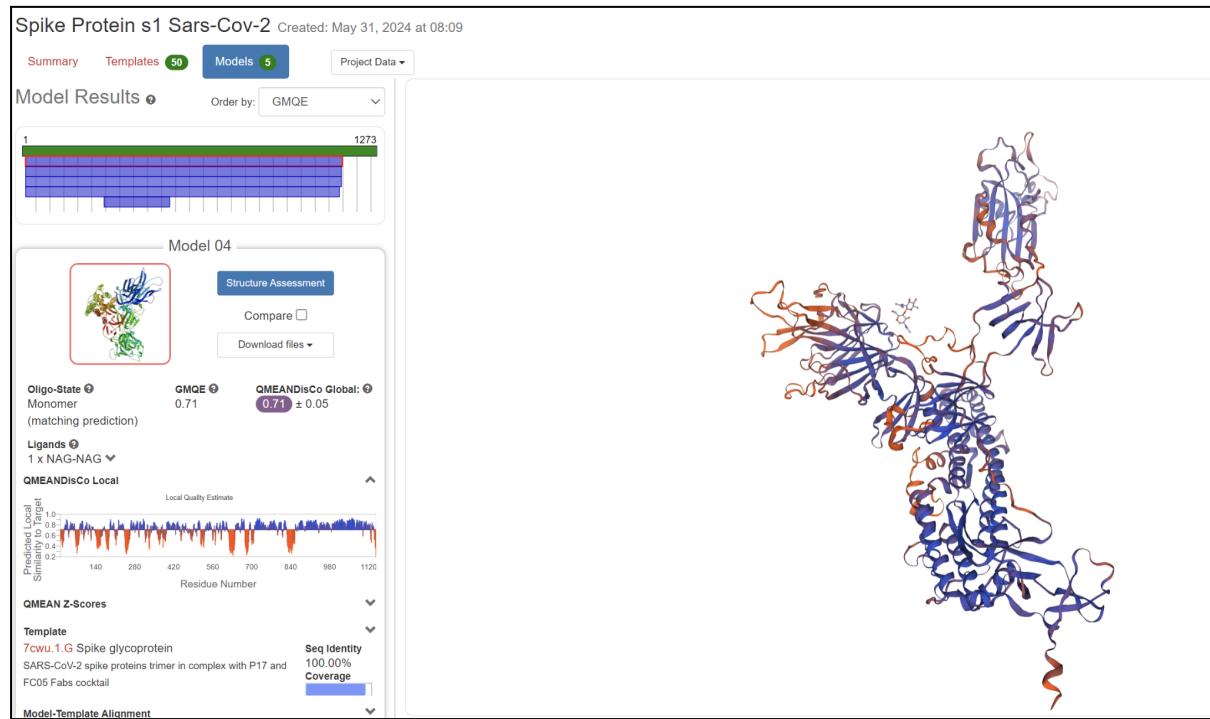
Swiss-Model searches for template structures that match with targeted protein sequences. By comparing the target sequence to sequences of protein structure in the Protein Data Bank (PDB). Swiss-Model selects the best template automatically, but Swiss-Model gives the option to manually choose a different template.

The top screenshot shows the 'Template Results' section for the 'Spike Protein s1 Sars-CoV-2' project. It displays the target sequence: `MHEVFLVLLPLVSSQCVNL TTRQLPAPAYINSFTRGVYYPDVKFRSSVLHS`. Below the sequence, a list of template structures is shown, with the first entry being '8tc0 1.A Spike glycoprotein'.

The bottom screenshot shows the 'Build Models' interface. It includes a 'Clear Selection' button and a message 'No templates selected'.

## Step 5: Model Building

Each selected template, Swiss-Model, will align the sequence and build the 3D model. After the model is built, the Swiss-Model provides information about the 3D protein model structure with metrics such as GMQE (Global Model Quality Estimation) and QMEAN (Qualitative Model Energy Analysis) scores, sequence identity, coverage percentage, and templates used to build the model. Swiss-Model provides built-in tools to inspect the 3D model.



## 2.3 Molecular Docking

### Step 1: Prepare the Structure of the Bioactive Compound

- 1) Open the PubChem website at <https://pubchem.ncbi.nlm.nih.gov>. Then search for the bioactive compound, which is Capsaicin.

The screenshot shows the PubChem homepage with a search bar at the top containing "Capsaicin". Below the search bar is a large search result table. The table has three columns: "Compound", "Gene", and "Taxonomy". The "Compound" column lists various forms of capsaicin, such as Capsaicin, CAPSAICINE, Capsaicin beta-D-Glucopyranoside, Capsaicin-d3, capsicinol, Capsaicin palmitate, Capsaicin-5,7-dene, Capsaicin(E/Z-Mixture), cis-Capsaicin, and 17-Hydroxy Capsaicin. The "Gene" column lists genes related to capsaicin sensitivity, including capsaicin sensitivity related QTL 1, 2, 3, 4, calcin, CaeNaCin, Caprin, Cyclicin-2, and Cyclicin 1. The "Taxonomy" column lists the organisms where these genes are found, such as Lectera capsici, Cohnella capsici, Rhizobium capsici, Diaporthe capsici, Lysobacter capsici, Cercospora capsici, Alternaria capsici, Pseudomonas capsici, Russelliana capsici, and Clavibacter capsici. At the bottom left of the table, it says "118M Compounds". At the bottom right, it says "996 Data Sources".

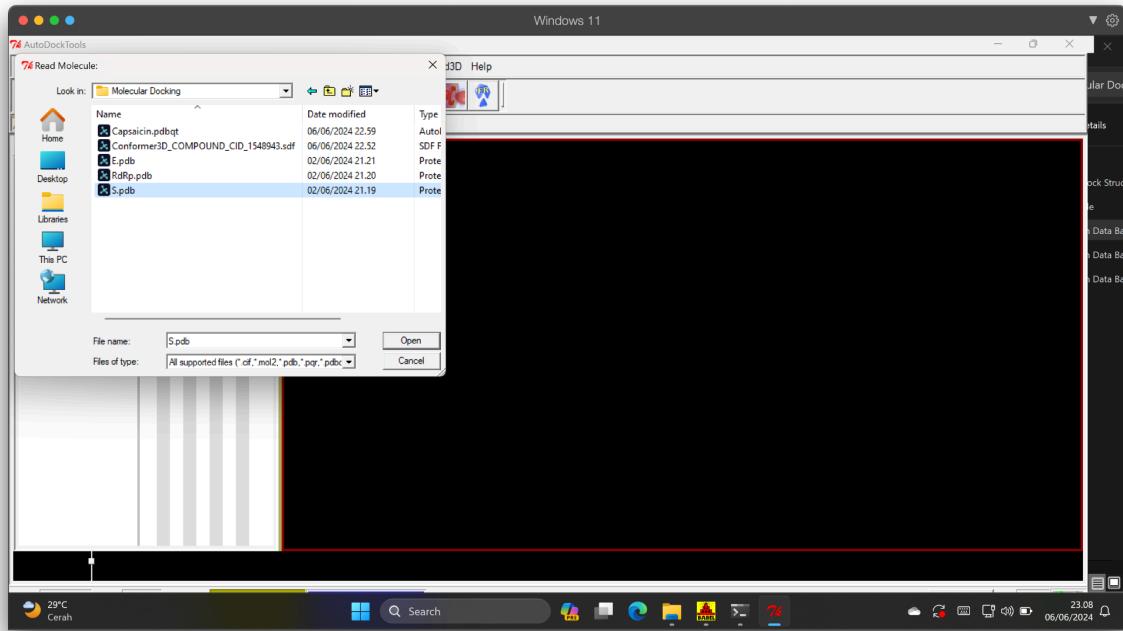
The screenshot shows the detailed view for Capsaicin (Compound CID: 1548943). The page includes a chemical structure diagram, basic properties (MF: C<sub>18</sub>H<sub>22</sub>NO<sub>3</sub>, MW: 305.4g/mol), IUPAC Name, Isomeric SMILES, InChIKey, InChI, Create Date (2005-03-25), and a list of tags from PubChem. Below this, there are links for Summary, Similar Structures Search, Related Records, and PubMed (MeSH Keyword). The URL at the bottom is https://pubchem.ncbi.nlm.nih.gov/compound/1548943.

- 2) Scroll down to the “1.2 3D Conformer” section, then select “Download Coordinates” as an sdf file.

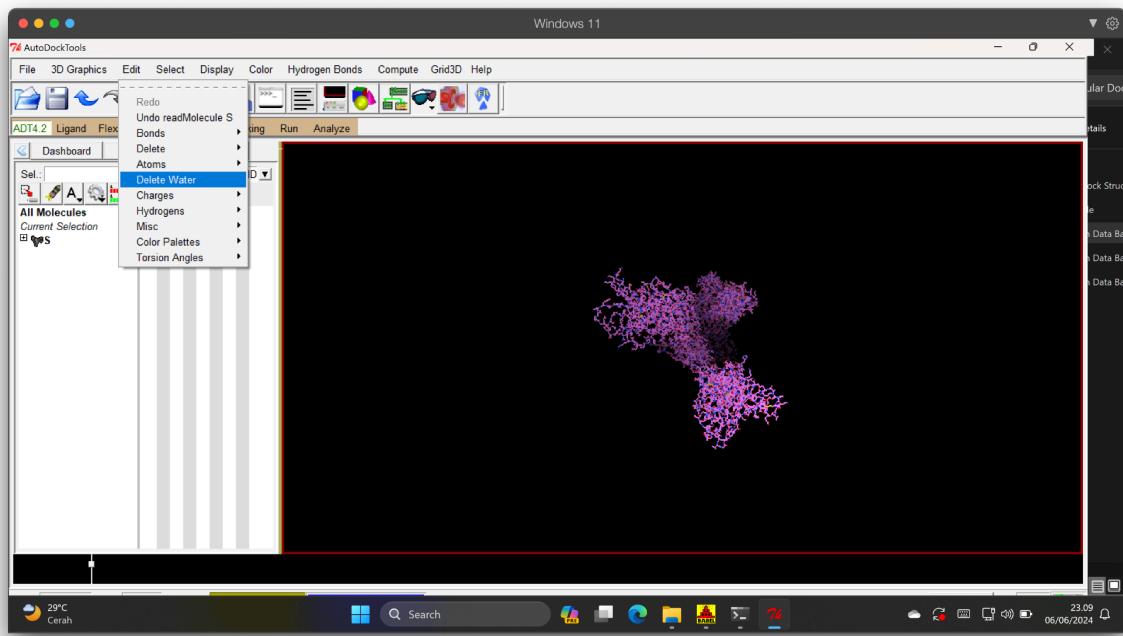
- 3) Open the OpenBabelGUI software. Set the “INPUT FORMAT” as “sdf”, then select the saved sdf file. Set the “OUTPUT FORMAT” as “pdbqt”, then set the location and name for the pdbqt file. Then, click the “CONVERT” button.

## Step 2: Prepare the Structure of the Protein

- 1) Open the AutoDockTools software, then go to File > Read Molecule > select the spike protein pdb file > Open.



- 2) To remove the water molecules in the structure, go to Edit > Delete Water.



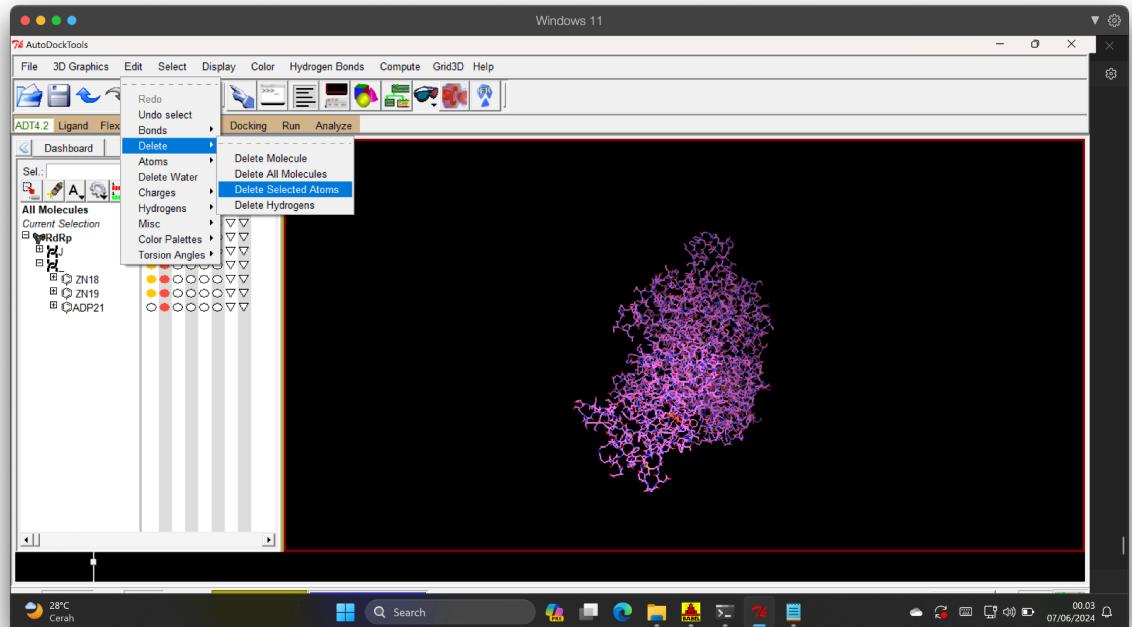
- 3) Check for non-bonded atoms in the structure by opening the pdb file using a text editor, then read the HETATM data thoroughly.
  - a) Spike Protein (S)  
There are no non-bonded atoms in the S structure. Therefore, no further action is required.
  - b) RNA-dependent RNA polymerase (RdRp)

```

ATOM 7479 U12 IHE J 929 192.98/ 157.514 112.111 1.00 0.65 C
ATOM 7479 OXT TIR J 929 193.001 136.923 114.942 1.00 0.65 C
TER 7479 THR J 929
HETATM 7473 ZN ZN - 18 174.541 156.302 155.878 1.00 60.89 ZN
HETATM 7474 ZN ZN - 19 187.565 159.932 139.846 1.00 61.60 ZN
HETATM 7475 PB ADP - 21 158.124 127.279 169.264 1.00 70.91 P
HETATM 7476 O1B ADP - 21 158.248 125.359 168.218 1.00 70.91 O
HETATM 7477 O2B ADP - 21 159.109 127.395 170.400 1.00 70.91 O
HETATM 7478 O3B ADP - 21 156.798 126.998 169.697 1.00 70.91 O
HETATM 7479 PA ADP - 21 160.055 125.729 167.967 1.00 70.91 P
HETATM 7480 O1A ADP - 21 160.109 124.480 167.125 1.00 70.91 O
HETATM 7481 O2A ADP - 21 160.574 127.054 167.463 1.00 70.91 O
HETATM 7482 O3A ADP - 21 158.559 125.921 168.528 1.00 70.91 O
HETATM 7483 O5' ADP - 21 160.826 125.420 169.339 1.00 70.91 O
HETATM 7484 C5' ADP - 21 162.096 125.384 169.695 1.00 70.91 C
HETATM 7485 C4' ADP - 21 162.365 125.384 169.695 1.00 70.91 C
HETATM 7486 O4' ADP - 21 162.519 124.285 171.081 1.00 70.91 O
HETATM 7487 C3' ADP - 21 163.767 126.065 171.507 1.00 70.91 C
HETATM 7488 O3' ADP - 21 163.744 127.313 172.200 1.00 70.91 O
HETATM 7489 C2' ADP - 21 164.193 124.947 172.436 1.00 70.91 C
HETATM 7490 O2' ADP - 21 164.006 125.383 173.793 1.00 70.91 O
HETATM 7491 C1' ADP - 21 163.299 123.814 172.192 1.00 70.91 C
HETATM 7492 N9 ADP - 21 163.935 122.672 171.588 1.00 70.91 N
HETATM 7493 C8 ADP - 21 165.245 122.657 171.281 1.00 70.91 C
HETATM 7494 N7 ADP - 21 165.610 121.464 170.749 1.00 70.91 N
HETATM 7495 C5 ADP - 21 164.513 120.687 170.715 1.00 70.91 C
HETATM 7496 C6 ADP - 21 164.198 119.309 170.273 1.00 70.91 C
HETATM 7497 N6 ADP - 21 165.156 118.508 169.749 1.00 70.91 N
HETATM 7498 N2 ADP - 21 162.924 118.880 170.418 1.00 70.91 N
HETATM 7499 C2 ADP - 21 161.965 119.668 170.930 1.00 70.91 C
HETATM 7500 N3 ADP - 21 162.188 120.928 171.351 1.00 70.91 N
HETATM 7501 C4 ADP - 21 163.412 121.484 171.276 1.00 70.91 C
CONNECT 7476 7476 7476 7477 7478
CONNECT 7475 7475 7475
CONNECT 7476 7475 7475
CONNECT 7477 7475
CONNECT 7478 7475
Ln 7706, Col 1 162 of 618.761 characters

```

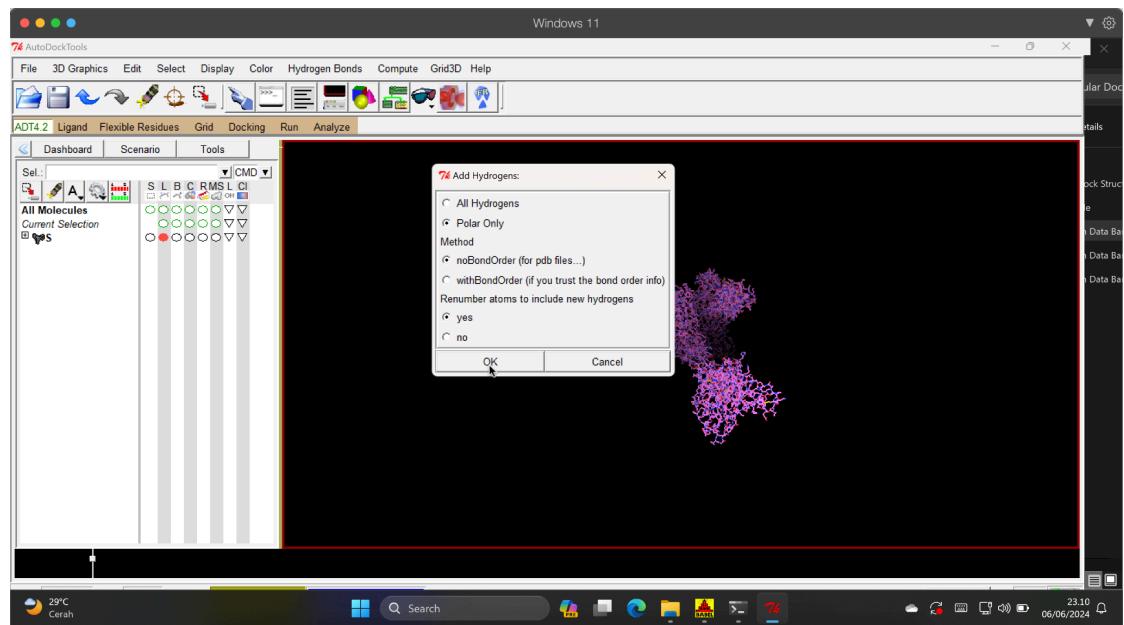
There are 2 non-bonded atoms in the RdRp structure, which are the ZN18 and ZN19 atoms. To remove the non-bonded atoms, click the dropdown button of the molecule, then select both atoms by clicking the “S” button. Go to Edit > Delete > Delete Selected Atoms.



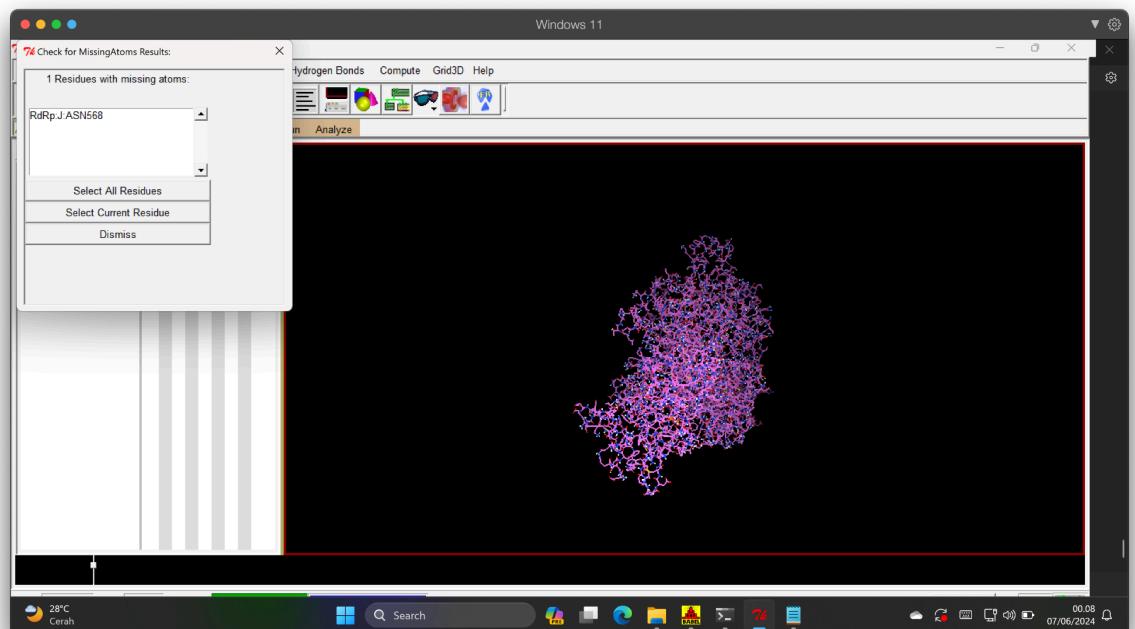
### c) Envelope Protein (E)

There are no non-bonded atoms in the E structure. Therefore, no further action is required.

- 4) To add polar hydrogens to the structure, go to Edit > Hydrogens > Add > Polar Only > OK.



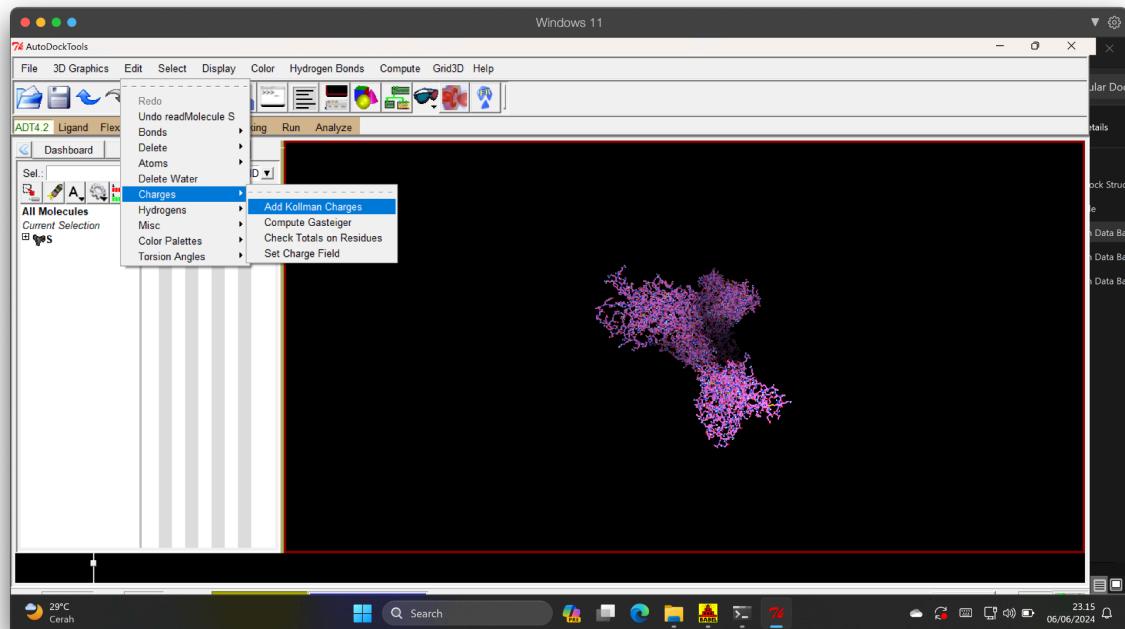
- 5) To check for missing atoms in the structure, go to Edit > Misc > Check for Missing Atoms.
- Spike Protein (S)  
There are no residues with missing atoms in the S structure. Therefore, no further action is required.
  - RNA-dependent RNA polymerase (RdRp)  
There are no residues with missing atoms in the RdRp structure. Therefore, no further action is required.
  - Envelope Protein (E)



There is 1 residue with missing atoms in the E structure. Click “Select All Residues”, then click “Dismiss”. To repair the missing atoms, go to Edit > Misc > Repair

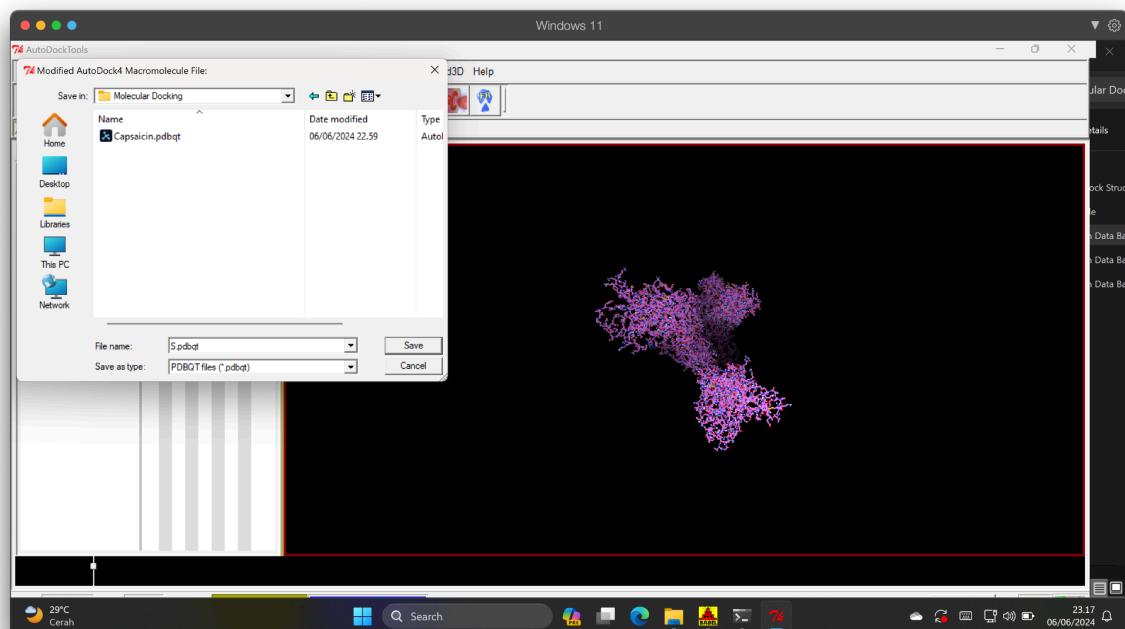
Missing Atoms. Repeat this step until there are no residues with missing atoms in the structure.

- 6) To add Kollman Charges in the structure, go to Edit > Charges > Add Kolman Charges.



To check residues with non-integral charges in the structure, go to Edit > Charges > Check Totals on Residues.

- 7) To save the molecules as a *.pdbqt* file, select Grid > Macromolecule > Choose... > select a molecule > Select Molecule > set the location and name of the *pdbqt* file > Save.



- 8) Repeat these steps for the RdRp and E protein as well.

### Step 3: Find the Size and Center of the Grid Box

- 1) Open the NCBI website at <https://www.ncbi.nlm.nih.gov>. Search for the protein, which is Spike SARS-CoV-2. Click the RefSeq Proteins dropdown button, then select the RefSeq.

a) Spike Protein (S)

Search for “Spike SARS-CoV-2”. Click the RefSeq Proteins dropdown button, then select the RefSeq.

b) RNA-dependent RNA polymerase (RdRp)

Search for “RNA-dependent RNA polymerase SARS-CoV-2”. Click the RefSeq Proteins dropdown button, then select the RefSeq.

c) Envelope Protein (E)

There are no RefSeq Proteins mentioned for “Envelope SARS-CoV-2” in the NCBI website/related literature. Therefore, the E protein will be docked using the blind docking technique to create a grid box that covers the whole protein. Skip to Step 4.

- 2) Find the active site/allosteric site/binding site of the protein.

a) Spike Protein (S)

The screenshot shows the NCBI protein page for Capsaicin | C18H27NO3 | CID. The URL is https://www.ncbi.nlm.nih.gov/protein/YP\_009724390.1. The page displays the amino acid sequence of the Spike protein (S) with various binding sites highlighted. A sidebar on the right lists external resources like Bacterial and Viral Bioinformatics Resource Center and GlyGen glycoinformatics resource. Below the sidebar, a "Recent activity" section shows search history for "surface glycoprotein [Severe acute respiratory syndrome coronavirus 2]", "S [Severe acute respiratory syndrome coronavirus 2]", "Spike SARS-CoV-2 (46582)", "spike s1 sars cov (1365)", and "spike s1 (9678)".

b) RNA-dependent RNA polymerase (RdRp)

The screenshot shows the NCBI protein page for Capsaicin | C18H27NO3 | CID. The URL is https://www.ncbi.nlm.nih.gov/protein/YP\_009725307.1. The page displays the amino acid sequence of the RNA-dependent RNA polymerase (RdRp) with various binding sites highlighted. A sidebar on the right lists external resources like RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus Protein] and RNA-dependent RNA polymerase SARS-CoV-2 (3882672). Below the sidebar, a "Recent activity" section shows search history for "RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]", "S [Severe acute respiratory syndrome coronavirus 2]", and "Spike SARS-CoV-2 (46582)".

- 3) To find the x, y, and z center coordinates of the grid box, open the pdb file of the protein using any text editor, then calculate the average of coordinates from the last active site/allosteric site/binding site.

a) Spike Protein (S)

CenterGridBox

N1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	ATOM	3236	NZ	LYS	G	417	146.04	119.387	211.411	1	0.77	N	Center Grid Box									
2	ATOM	3463	O	GLY	G	446	142.149	94.389	207.584	1	0.52	O	xcenter:	143.572								
3	ATOM	3487	OH	TYR	G	449	141.755	99.606	210.569	1	0.61	O	ycenter:	109.058								
4	ATOM	3527	OH	TYR	G	453	144.516	111.939	210.362	1	0.81	O	zcenter:	213.046								
5	ATOM	3546	C2	LEU	G	455	147.167	113.842	213.296	1	0.77	C										
6	ATOM	3557	CZ	PHE	G	456	148.189	117.768	215.339	1	0.69	C										
7	ATOM	3718	CB	GLY	G	475	134.516	104.002	227.402	1	0.49	C										
8	ATOM	3791	CZ	PHE	G	486	135.457	116.042	227.488	1	0.49	C										
9	ATOM	3799	N2	ASN	G	487	156.059	118.566	224.995	1	0.59	N										
10	ATOM	3817	OH	TYR	G	489	150.123	114.899	221.425	1	0.61	O										
11	ATOM	3852	N2	GLN	G	493	144.972	108.746	212.977	1	0.78	N										
12	ATOM	3874	O	GLY	G	496	138.538	104.044	207.283	1	0.71	O										
13	ATOM	3894	N2	GLN	G	498	139.451	98.921	208.687	1	0.68	N										
14	ATOM	3908	C2	THR	G	500	132.429	96.491	208.088	1	0.65	C										
15	ATOM	3916	N2	ASN	G	501	135.7	103.485	206.415	1	0.72	N										
16	ATOM	3920	O	GLY	G	502	130.582	105.925	205.291	1	0.70	O										
17	ATOM	3943	OH	TYR	G	505	132.893	109.813	211.875	1	0.70	O										
18				Center			143.572	109.058	213.046													
19																						
20																						
21																						
22																						
23																						
24																						
25																						
26																						
27																						
28																						
29																						
30																						
31																						
32																						
33																						
34																						
35																						

Sheet1 Accessibility: Investigate

Average: 155.2254902 Count: 7 Sum: 465.6764706

### b) RNA-dependent RNA polymerase (RdRp)

CenterGridBoxRdRp

N1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	ATOM	4439	OG	SER	J	549	160.199	143.697	116	1	0.91	O	Center Grid Box									
2	ATOM	5450	C2	THR	J	680	168.706	149.374	133.988	1	0.91	C	xcenter:	168.921								
3	ATOM	5456	OG	SER	J	681	168.325	156.239	131.82	1	0.88	O	ycenter:	148.822								
4	ATOM	5462	O	SER	J	682	157.199	153.571	136.671	1	0.88	O	zcenter:	128.068								
5	ATOM	5493	C2	THR	J	687	150.843	150.623	127.862	1	0.93	C										
6	ATOM	5522	N2	ASN	J	691	170.941	146.425	130.905	1	0.92	N										
7	ATOM	6085	OG	SER	J	759	173.215	144.144	129.229	1	0.9	O										
8				Center			168.921	148.822	128.068													
9																						
10																						
11																						
12																						
13																						
14																						
15																						
16																						
17																						
18																						
19																						
20																						
21																						
22																						
23																						
24																						
25																						
26																						
27																						
28																						
29																						
30																						
31																						
32																						
33																						
34																						
35																						

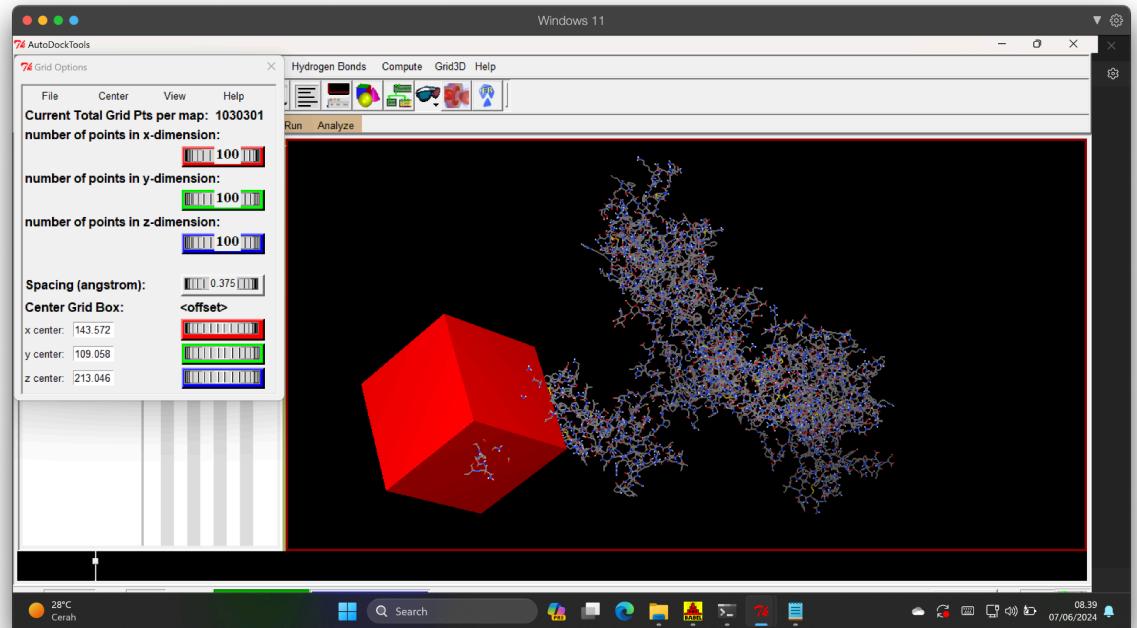
Sheet1 Accessibility: Investigate

Average: 148.6037619 Count: 7 Sum: 445.8112857

- 4) Load the ligand by selecting Grid > Set Map Types > Open Ligand... > select the pdbqt of the Capsaicin structure > Open. Load the macromolecule by selecting Grid > Macromolecule > Open... > select the pdbqt of the protein structure > Open. To visualize the grid box for docking, select Grid > Grid Box.

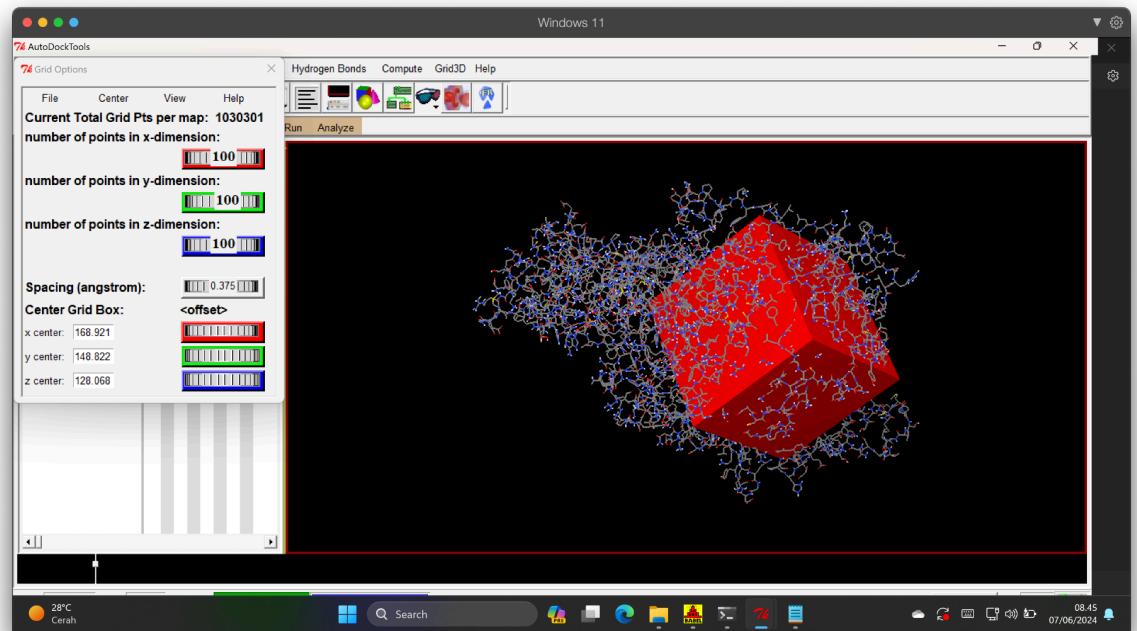
#### a) Spike Protein (S)

Input the calculated centre coordinates of the grid box, then adjust the numbers of points in x, y, and z dimensions to cover the whole active site/allosteric site/binding site.



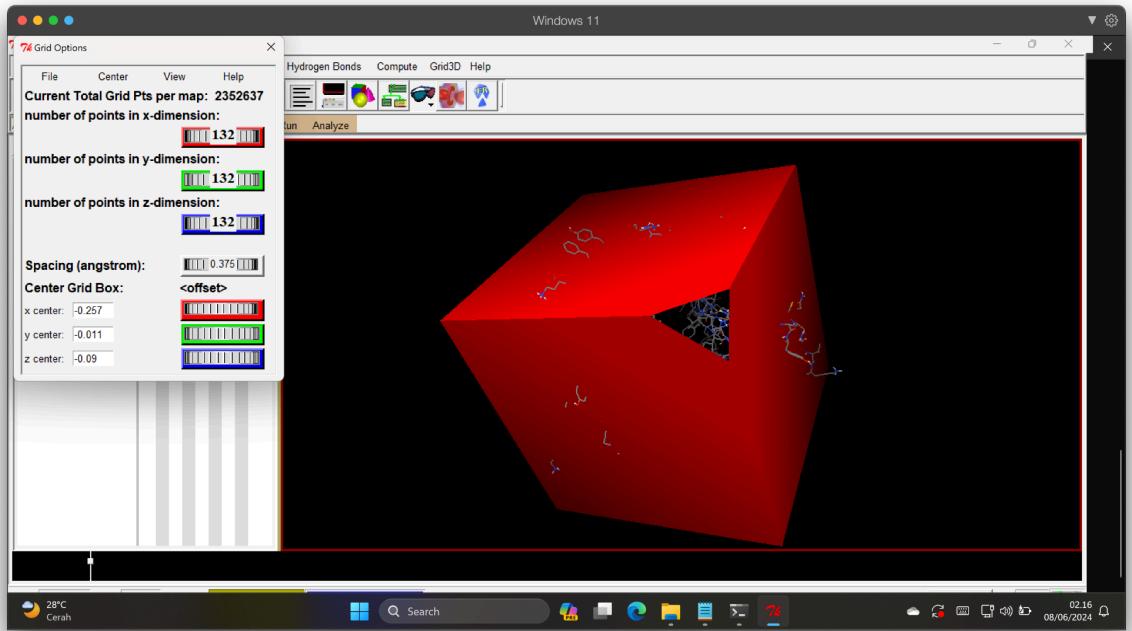
b) RNA-dependent RNA polymerase (RdRp)

Input the calculated center coordinates of the grid box, then adjust the numbers of points in x, y, and z dimensions to cover the whole active site/allosteric site/binding site.



c) Envelope Protein (E)

Adjust the numbers of points in x, y, and z dimensions to cover the whole protein.



#### Step 4: Prepare the Configuration File for Docking

- 1) To create the configuration file for docking, open any text editor, then create a txt file with the format below:

```
receptor = [receptor_filename].pdbqt
ligand = [ligand_filename].pdbqt
```

```
center_x = [x_center]
center_y = [y_center]
center_z = [z_center]
```

```
size_x = [number_of_points_in_x_dimension]
size_y = [number_of_points_in_y_dimension]
size_z = [number_of_points_in_z_dimension]
```

```
energy_range = 4
```

```
exhaustiveness = 8
```

A screenshot of a Windows 11 desktop environment. At the top, there's a taskbar with several open windows: 'S.pdb', 'RdRp.pdb', 'E.pdb', 'Capsaicin.pdb', 'Config5.txt' (which is the active window), and others partially visible. The main area is a terminal window with a dark background. It contains the following Python code:

```
receptor = S.pdbqt
ligand = Capsaicin.pdbqt

center_x = 143.572
center_y = 109.058
center_z = 213.046

size_x = 100
size_y = 100
size_z = 100

energy_range = 4
exhaustiveness = 8
```

The bottom of the screen shows standard Windows system status icons (battery, signal, volume) and the date/time (07/06/2024, 09:15). The bottom-left corner displays the system temperature as 28°C.

- 2) Repeat the step above for the RdRp and E protein as well.

## Step 5: Docking Process

- 1) To do the docking process using the AutoDock Vina program, open a command-line program in the location of the saved pdbqt and txt files. Then, input the command below:  
“[location\_of\_vina\_exe]” --receptor [receptor\_filename].pdbqt --ligand  
[ligand\_filename].pdbqt --config [config\_filename].txt --log [log\_filename].txt --out  
[output\_filename].pdbqt

- 2) To get the docking results, open the generated output log txt file.

```

Windows 11
S.pdb RdRp.pdb E.pdb ConfigS.txt ConfigRdRp.txt ConfigE.txt Log5.txt
File Edit View
#####
# If you used AutoDock Vina in your work, please cite: #
# O. Trott, A. J. Olson, #
# AutoDock Vina: improving the speed and accuracy of docking #
# with a new scoring function, efficient optimization and #
# multithreading, Journal of Computational Chemistry 31 (2010) #
# 455-461 #
# DOI 10.1002/jcc.21334 #
# Please see http://vina.scripps.edu for more information. #
#####

WARNING: The search space volume > 27000 Angstrom^3 (See FAQ)
Detected 4 CPUs
Reading input ... done.
Setting up the scoring function ... done.
Analyzing the binding site ... done.
Using random seed: 170680464
Performing search ... done.
Refining results ... done.

mode | affinity | dist from best mode
|(kcal/mol)| rmsd l.b.| rmsd u.b.
-----
1   -5.4    0.000   0.000
2   -5.3    3.024   4.597
3   -5.3    33.254  37.006
4   -5.2    13.628  15.554
5   -4.5    54.405  56.968
6   -4.5    12.170  14.671
7   -4.4    23.313  25.436
8   -4.4    32.232  34.132
9   -4.3    25.263  27.657

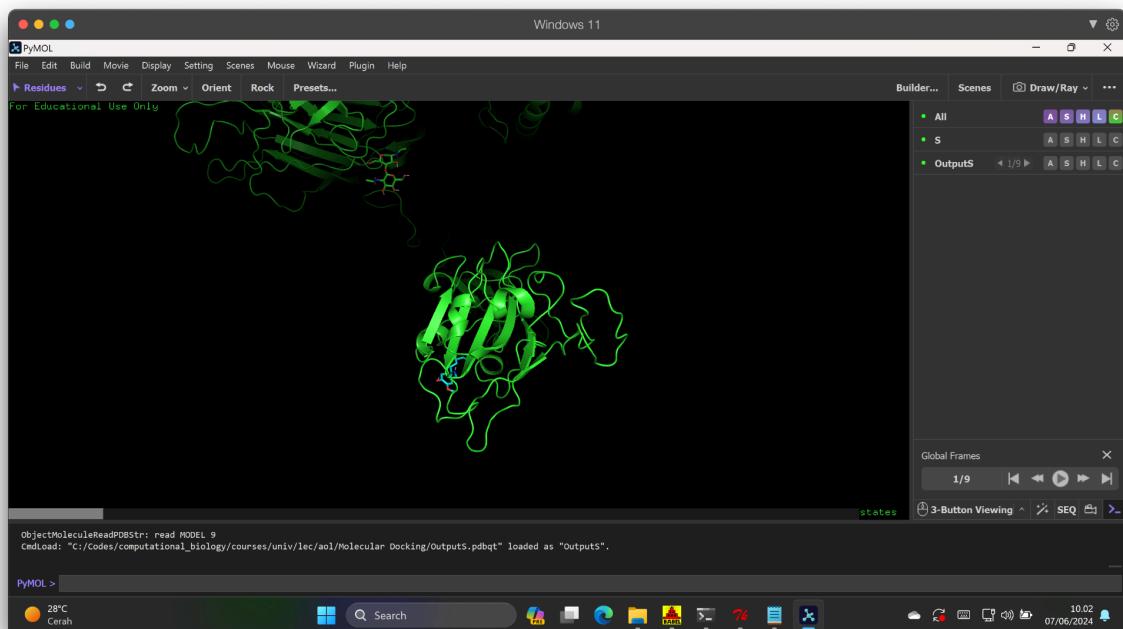
Writing output ... done.
Ln 35, Col 40 481 of 1632 characters
31°C Sebagian cerah
Search 100% Windows (CRLF) 22.20 07/06/2024

```

3) Repeat the steps above for the RdRp and E protein as well.

### Step 6: Docking Visualization

- 1) Open the PyMol software. To open the receptor structure, go to File > Open... > select the receptor pdbqt file. To open the output of the docking, go to File > Open... > select the output pdbqt file.



2) Repeat the step above for the RdRp and E protein as well.

### 3. Result

#### 3.1 FASTA Format of Amino Acid Sequence

##### 3.1.1 S1 Subunit of the Spike Protein

**spike protein [Severe acute respiratory syndrome coronavirus 2]**

>QIH45093.1 spike protein [Severe acute respiratory syndrome coronavirus 2]  
MFVFLVLLPLVSSQCVCNLTRTQLPPAYTNSFTRGVYYPDVKFRSSVLHSTQDLFLPFFSNVTWFHAIHVGSGTNGTKRFDNPVLPFNDGVYFASTEKSNIIRGWIFGTTLDSKTQSLLIVNNATNVVIKVCEFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLMDLEGKQGNFKNLERVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLLALHRSYLTGDSSSGWTAGAAAYVGYLQPRTFLKYNEGTITDAVDCALDPLSETKCTLKSFTVEKGIVQTSNFRVQPTESIVRFPNITNLCPGEVFNFATRFASVYAWNKRKRISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRIQIAPGQTGKIADYNKLPDDFTGCVIAWNSNNLDSKVGGNNYNYLRLFRKSNLKPFERDISTEIQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIA DTTDAVRDPQTLEILDITPCSFGGVSITPGNTNTSNQAVLYQDVNCTEVPAIHADQLPTWR VYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQNSPRRARSVASQSIIAYTMSLGAENSVAYSNNIAPTNFTISVTTEILPVSMKTSVDCTMYICGDSTECNSNLLQYGSFCTQLNRALTGIAVEQDKNTQEVAQVKQIYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIARADLICAQKFNGLTVLPPLLTDEMAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVQLSSNFGAISSVLDILSRDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEI RASANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAHGVVFLHVTYVPAQEKNFTTAP AICHDGKAHFREGVFSNGTHWFVTQRNFYEPQIITDNTFVSGNCVVIGIVNNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCSCLKGCCSCCKFDEDDSEPVLKGVKLHYT

##### 3.1.2 RNA-dependent RNA polymerase (RdRp)

**RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]**

>YP\_009725307.1 RNA-dependent RNA polymerase [Severe acute respiratory syndrome coronavirus 2]

SADAQSFLNRVCVSAARLTPCGTGTSTDVYRAFDIYNDKVAGFAKFLKTNCRCFQEKDDEDNLIDSYFVVKRHTFSNYQHEETIYNLLKDCPAVAKHDFFKFRIDGDMVPHISRQRLTKYTMA DLVYALRHFDENGCDTLKEILVTYNC CDDDYFNKKDWYDFVENPDILRVYANLGERVRQALL KTVQFC DAMRNAGIVGVLTLDNQDLNGNWYDFGDFIQTPGSGVPVVDSYYSLMPILTTR ALTAESHVDTDLTPYIKWDLLKYDFTEERLKLFDRYFKYWDQTYHPNCVNCLDDRCILHCA NFNVLFSTVFPPTSGGPLVRKIFVDGVVPVSTGYHFRELGVVHNQDVNLHSSRLSFKELLVA ADPAMHAASGNLLDKRTTCSVAALTNNVAFQTVKPGNFNKDFYDFAVSKGFFKEGSSVEL KHFFFAQDGNAIASDYDYYRYNLPTMCDIRQLFVVEVVDKYFDCYDGGCINANQVIVNNL DKSAGFPFNWKWGKARLYYDSMSYEDQDALFAYTKRNVPIITQMLKYAISAKNRARTVAGV SICSTMNRQFHQKLLKSIATRGATVVI GTSKFYGGWHNMLKTVYSDVENPHLMGWDPK CDRAMPNMLRIMASLVLARKHTTCCSLSHRFYRLANECAQVLSEMVMCGGSLYVKPGGTSS GDATTAYANSVFNICQAVTANVNALLSTDGNKIADKYVRNLQHRLYECLYRN RDVDTDFVNEYAYLRKHFSMMILSDDAVVCFNSTYASQGLVASIKNFKSVLYYQNNVFMSEAKCWTETDLTKGPHEFCSQHTMLVKQGDDYVYLPYPDPSRILGAGCFVDDIVKTDGTLMIERFVSLAIDAYPL

TKHPNQEYADVFLYLQYIRKLHDELTGHMLDMYSVMLNDNTSRYWEPEFYEAMYTPHTV  
LQ

### 3.1.3 Envelope protein (E)

envelope protein [Severe acute respiratory syndrome coronavirus 2]

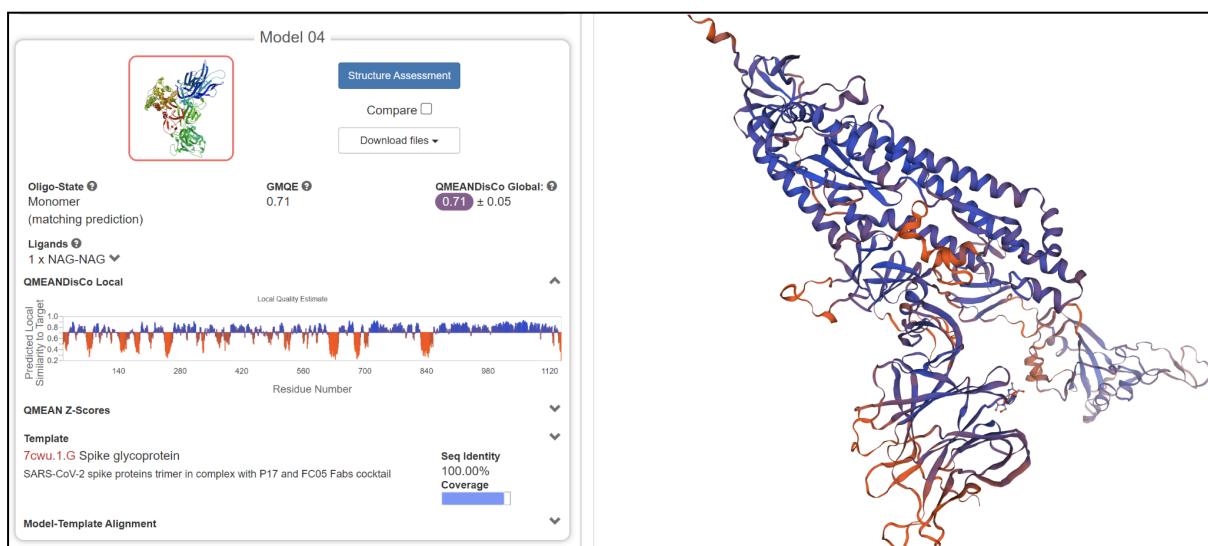
>YP\_009724392.1 envelope protein [Severe acute respiratory syndrome coronavirus 2]

MYSFVSEETGTLIVNSVLLFLAFVVFLVTLAILTALRLCAYCCNIVNVSLVKPSFYVYSRVKN  
LNSSRVPDLLV

## 3.2 Prediction of Protein 3D Structure

Swiss-model produces several models from existing templates. From the model results, they can be sorted based on GMQE, GMEAN, oligo, and others.

### 3.2.1 Spike Protein S1 SARS-CoV-2



#### 1. Oligo State: Monomer

Number and arrangement of subunits (monomer) that together form the functional protein complex.

#### 2. GMQE: 0.71

The quality estimation score shows 0.71, meaning the expected accuracy of the modelled protein structure. A higher score indicates higher reliability, and the GMQE score ranges from 0 to 1. There may still be some areas that could be improved

#### 3. QMEANDisCo Global: $0.71 \pm 0.05$

Evaluates various structural features with  $\pm 0.05$  that indicates the true quality of the model could be expected to fall within the range of 0.66 to 0.76. A smaller QMEAN (e.g.,  $\pm 0.05$ ) value indicates that the QMEAN score is precise and model quality assessment is reliable. A larger value would indicate more variability and less confidence in the exact quality score.

#### 4. Ligands: 1 x NAG-NAG

Ligands are molecules that bind to a protein at specific sites, often playing a critical role in the protein's function, such as substrate binding, regulation, or structural stabilization. "Ligands 1 x

"NAG-NAG" signifies that the protein structure includes a single complex of two N-acetylglucosamine molecules, highlighting an important aspect of the protein's interaction or structural features.

5. Template: 7cwu.1.G Spike glycoprotein

This refers to the Protein Data Bank (PDB) identifier for a specific protein structure. Each structure in the PDB is given a unique four-character alphanumeric code. In this case, "7CWU" identifies a particular protein structure. Proteins often consist of multiple chains, and each chain within a PDB entry is designated by an identifier. Here, "1.G" refers to the specific chain (G) in the structure (7CWU). This is a structural protein commonly found in coronaviruses, such as SARS-CoV-2, responsible for the COVID-19 pandemic.

6. Seq Identity: 100.00%

It measures the degree of similarity between the amino acid sequence of the target protein and the template protein. Models generated using such templates are highly reliable and accurately reflect the native structure and function of the target protein because the score is 100.00%.

7. Coverage

Templates with high coverage ensure that a significant portion of the target protein's sequence is accurately represented in the model, enhancing its reliability for various structural and functional analyses.

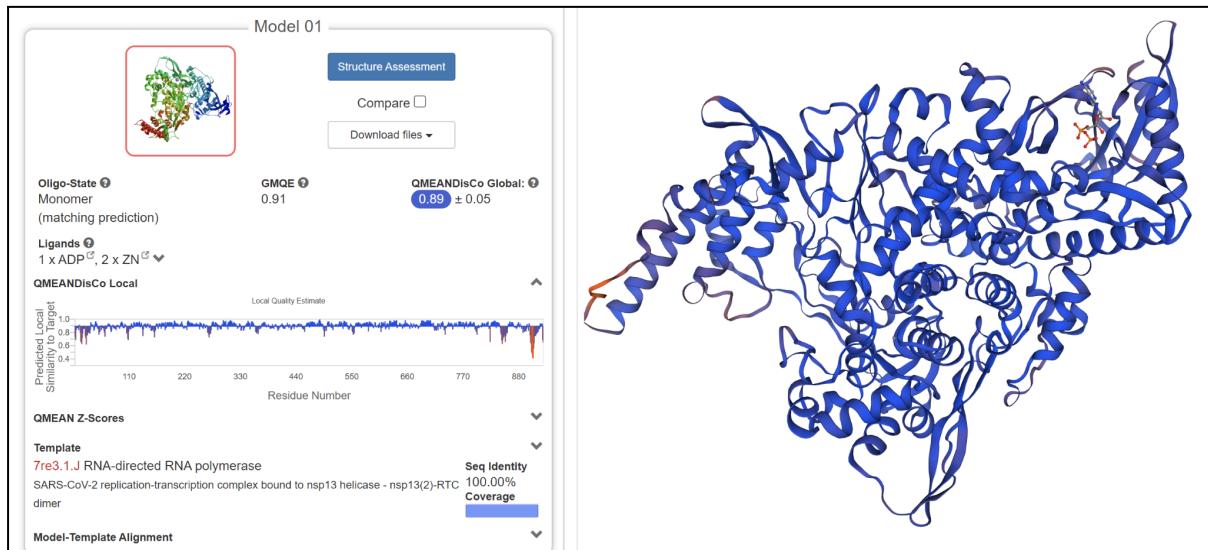
The provided information outlines key aspects of a modelled protein structure, including its oligo state, quality estimation scores (GMQE and QMEANDisCo Global), ligands present, the template used for modelling, sequence identity with the template, and coverage.

The protein is characterized as a monomer, indicating that it functions as a single unit rather than as part of a larger complex. Both the GMQE score of 0.71 and the QMEANDisCo Global score of  $0.71 \pm 0.05$  suggest a moderate to high level of confidence in the accuracy of the modelled structure. The presence of ligands, specifically 1 x NAG-NAG, highlights potential functional or structural interactions within the protein.

The modelling process utilized a highly similar template, as indicated by the 100.00% sequence identity, likely contributing to the reliability and accuracy of the model. Furthermore, the coverage of the template ensured that a significant portion of the target protein's sequence was accurately represented in the model.

In conclusion, the modelled protein structure exhibits promising characteristics, including a monomeric oligo state, reliable quality estimation scores, the presence of ligands, and a high sequence identity with the template. These factors collectively enhance confidence in the structural accuracy and functional relevance of the model, making it a valuable resource for further structural and functional analyses.

### 3.2.2 RNA-dependent RNA polymerase (RdRp) SARS-CoV-2



#### 1. Oligo State: Monomer

Number and arrangement of subunits (monomer) that together form the functional protein complex.

#### 2. GMQE: 0.91

The quality estimation score shows 0.91, meaning the expected accuracy of the modelled protein structure. A higher score indicates higher reliability, and the GMQE score ranges from 0 to 1. There may still be some areas that could be improved.

#### 3. QMEANDisCo Global: $0.89 \pm 0.05$

Evaluates various structural features with  $\pm 0.05$  which indicates the true quality of the model could be expected to fall within the range of 0.94 to 0.84. A smaller QMEAN (e.g.,  $\pm 0.05$ ) value indicates that the QMEAN score is precise and model quality assessment is reliable. A larger value would indicate more variability and less confidence in the exact quality score.

#### 4. Ligands: 1 x ADP, 2 x ZN

Ligands are molecules that bind to a protein at specific sites, often playing a critical role in the protein's function, such as substrate binding, regulation, or structural stabilization. ADP is a nucleotide composed of adenine, ribose, and two phosphate groups. It is a critical molecule in cellular metabolism, serving as a precursor to ATP (adenosine triphosphate) and playing a role in energy transfer within cells. "2 x ZN" indicates that there are two zinc ions associated with the protein structure. Zinc ions often coordinate with specific amino acid residues in the protein to form metal-binding sites, which can be crucial for protein stability and function.

#### 5. Template: 7re3.1.J RNA-directed RNA polymerase

The entry "7RE3.1.J" refers to a specific protein structure found in the Protein Data Bank (PDB). In this case, the "7RE3" is the PDB identifier, and "1.J" designates a specific chain within that structure.

#### 6. Seq Identity: 100.00%

It measures the degree of similarity between the amino acid sequence of the target protein and the template protein. Models generated using such templates are highly reliable and accurately reflect the native structure and function of the target protein because the score is 100.00%.

## 7. Coverage

Templates with high coverage ensure that a significant portion of the target protein's sequence is accurately represented in the model, enhancing its reliability for various structural and functional analyses.

The provided information presents crucial details about a modelled protein structure, including its oligo state, quality estimation scores (GMQE and QMEANDisCo Global), ligands, the template used for modelling, sequence identity with the template, and coverage.

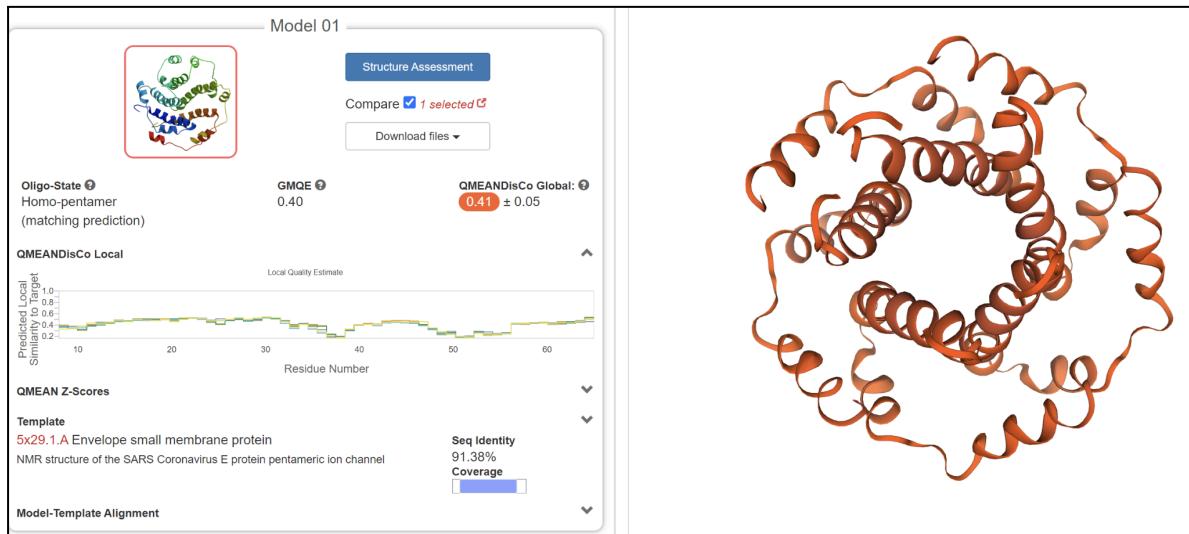
The protein is identified as a monomer, indicating that it functions as a single unit rather than as part of a larger complex. The high GMQE score of 0.91 suggests a high level of confidence in the accuracy of the modelled structure, while the QMEANDisCo Global score of  $0.89 \pm 0.05$  indicates a precise assessment of the model's quality within a narrow range.

The presence of ligands, specifically 1 x ADP and 2 x ZN, suggests potential functional roles for the protein, with ADP likely involved in cellular metabolism and zinc ions contributing to protein stability and function.

The modelling process utilized a highly similar template, as indicated by the 100.00% sequence identity, enhancing the reliability and accuracy of the model. Additionally, the coverage of the template ensured that a significant portion of the target protein's sequence was accurately represented in the model.

In conclusion, the modelled protein structure exhibits promising characteristics, including a monomeric oligo state, high-quality estimation scores, the presence of functionally relevant ligands, and a high sequence identity with the template. These factors collectively enhance confidence in the structural accuracy and functional relevance of the model, making it a valuable resource for further structural and functional analyses.

### 3.3.3 Envelope Protein (E) Sars-Cov-2



1. Oligo State: Homo-pentamer  
Number and arrangement of subunits (Monomer) that together form the functional protein complex.
2. GMQE: 0.40  
The quality estimation score shows 0.40, meaning the expected accuracy of the modelled protein structure. A higher score indicates higher reliability, and the GMQE score ranges from 0 to 1. There may still be some areas that could be improved.
3. QMEANDisCo Global:  $0.41 \pm 0.05$   
Evaluates various structural features with  $\pm 0.05$  that indicates the true quality of the model could be expected to fall within the range of 0.36 to 0.46. A smaller QMEAN (e.g.,  $\pm 0.05$ ) value indicates that the QMEAN score is precise and model quality assessment is reliable. A larger value would indicate more variability and less confidence in the exact quality score.
4. Template: 5x29.1.A Envelope small membrane protein  
The entry "5x29.1.A" corresponds to a specific protein structure found in the Protein Data Bank (PDB), where "5x29" is the PDB identifier and "1.A" designates a particular chain within that structure.
5. Seq Identity: 91.38%  
It measures the degree of similarity between the amino acid sequence of the target protein and the template protein. Models generated using such templates are highly reliable and accurately reflect the native structure and function of the target protein because the score is 91.38%.
6. Coverage  
Templates with high coverage ensure that a significant portion of the target protein's sequence is accurately represented in the model, enhancing its reliability for various structural and functional analyses.

The provided information outlines important details regarding a modelled protein structure, including its oligo state, quality estimation scores (GMQE and QMEANDisCo Global), the template used for modelling, sequence identity with the template, and coverage.

The protein is identified as a homo-pentamer, indicating that it forms a functional protein complex composed of five monomeric subunits. The GMQE score of 0.40 suggests moderate confidence in the accuracy of the modelled structure, while the QMEANDisCo Global score of  $0.41 \pm 0.05$  indicates a moderate level of precision in the assessment of the model's quality.

The modelling process utilized a template with a relatively high sequence identity of 91.38%, enhancing the reliability and accuracy of the model. Additionally, the coverage of the template ensured that a significant portion of the target protein's sequence was accurately represented in the model.

In conclusion, the modelled protein structure exhibits characteristics of a homo-pentameric complex, with moderate-quality estimation scores and a relatively high sequence identity with the template. While there may be areas for improvement, the model provides a valuable starting point for further structural and functional analyses of the protein complex.

### 3.3 Molecular Docking Outputs

#### 3.3.1 Docking Results using AutoDock Vina

- Spike Protein (S)

mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-5.4	0.000	0.000
2	-5.3	3.024	4.597
3	-5.3	33.254	37.006
4	-5.2	13.628	15.554
5	-4.5	54.405	56.968
6	-4.5	12.170	14.671
7	-4.4	23.313	25.436
8	-4.4	32.232	34.132
9	-4.3	25.263	27.657

- RNA-dependent RNA polymerase (RdRp)

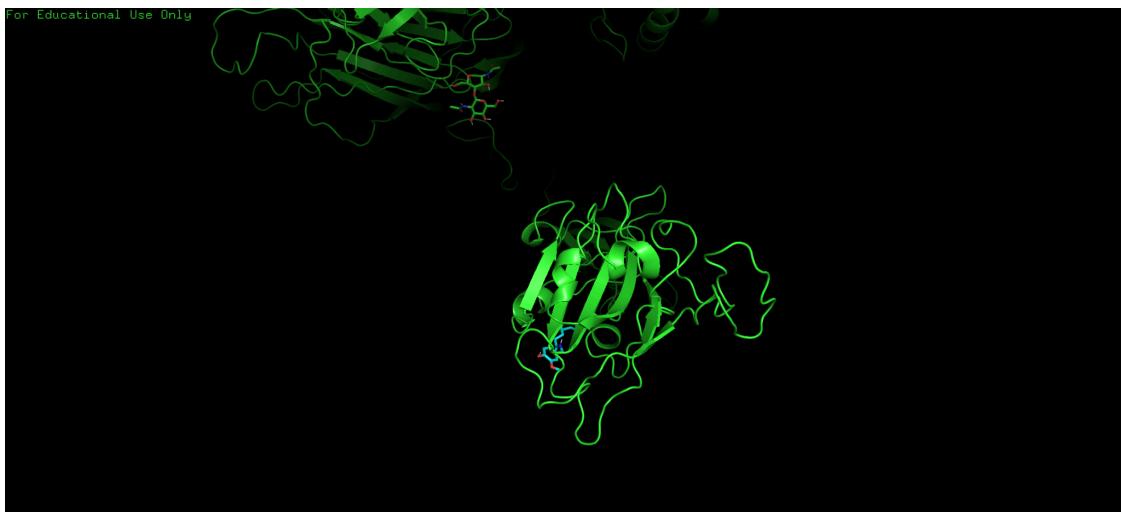
mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-6.0	0.000	0.000
2	-5.8	56.606	58.968
3	-5.5	56.582	59.215
4	-5.5	29.000	31.687
5	-5.4	42.798	44.743
6	-5.4	20.373	23.349
7	-5.3	19.810	21.986
8	-5.2	29.565	32.689
9	-5.2	20.069	23.970

- Envelope Protein (E)

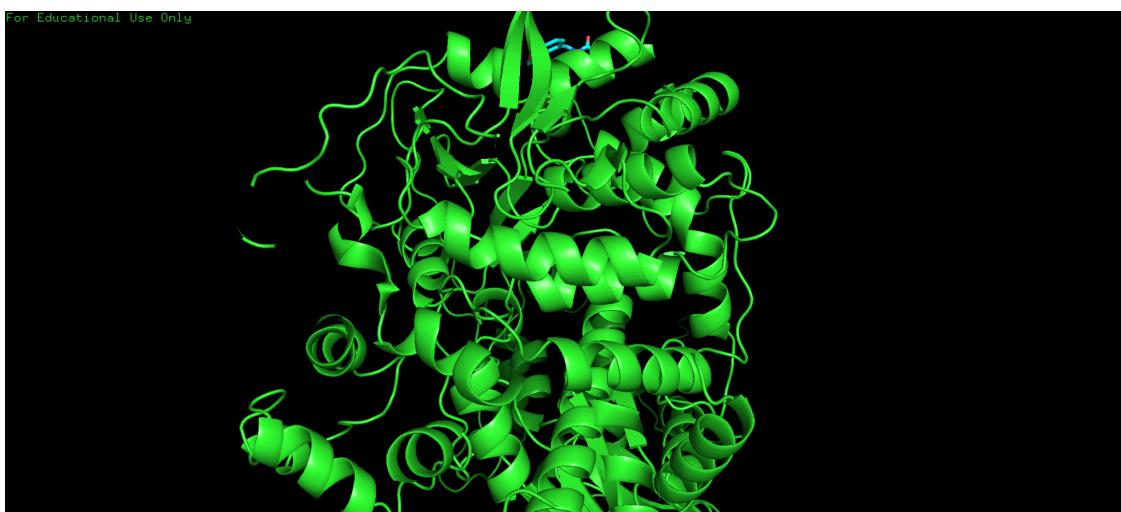
mode	affinity (kcal/mol)	dist from best mode	
		rmsd l.b.	rmsd u.b.
1	-5.4	0.000	0.000
2	-5.2	15.786	17.078
3	-5.1	14.303	16.428
4	-4.8	19.333	21.936
5	-4.8	6.648	9.863
6	-4.8	17.120	18.481
7	-4.7	7.011	10.437
8	-4.6	17.010	19.495
9	-4.6	15.709	18.005

### 3.3.1 Docking Poses and Interactions Between the Ligand and the Protein Visualization Using PyMOL

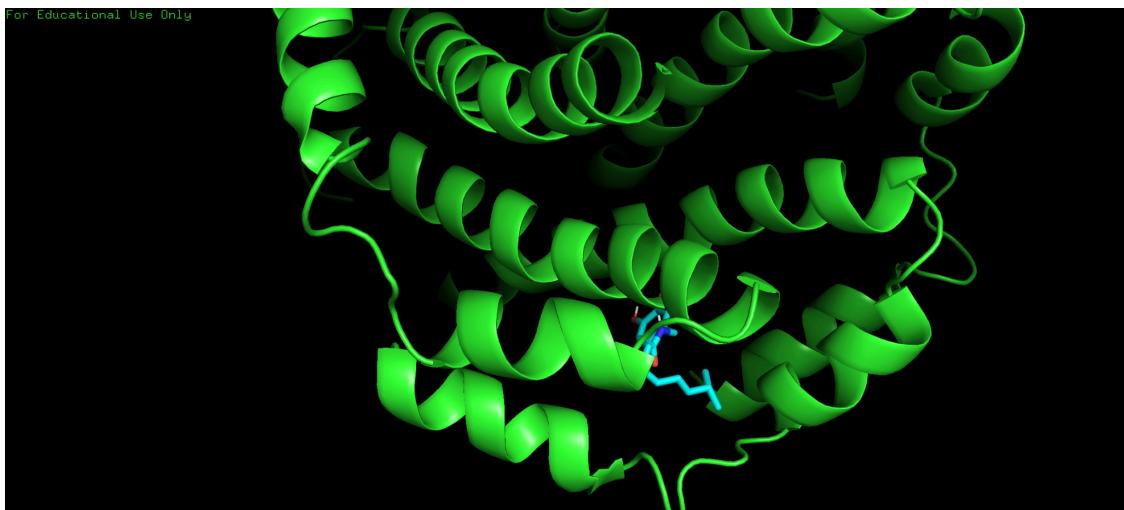
- Spike Protein (S)



- RNA-dependent RNA polymerase (RdRp)



- Envelope Protein (E)



## 4. Discussion

### 4.1 Spike Protein

The S1 Spike Protein model has a GMQE and Global QMEANDisCo score of  $0.71 \pm 0.05$ , indicating moderate reliability. While the model is generally considered reliable, there are some regions of the protein structures that could be refined. The identification of NAG-NAG ligands in the model reveals important docking sites that could be used to target the protein using antiviral compounds, which is critical for understanding how disrupting these interactions could affect protein function and stability. With a 100% sequence identity with the template and high coverage, the modelled protein closely represents the native structure of the Spike protein, giving credibility for further protein function investigations.

Recent updates have unveiled the results of analyzing binding affinity shedding light on the interactions between the S1 subunit and capsaicin. According to docking studies, capsaicin demonstrates a binding affinity of -5.4 kcal/mol suggesting potential inhibitory effects on the Spike protein that might impede viral entry into host cells.

### 4.2 RdRp Protein

RdRp's model has a GMQE and Global QMEANDisCo score of  $0.91 \pm 0.05$ , indicating high confidence in the model and reflecting the reliability and accuracy of the model. The docked ligands and cofactors found in the structure are significant in understanding the protein's function: ADP and Zn ions; here, ADP is involved in energy transfer, and Zn ions play a role in the protein's structural stability. As the sequence identity is 100% (the highest confidence we can achieve), sequence identity and coverage indicate the model is likely closely representing the actual protein structure, ensuring the model's validity for further functionality and inhibitory potential investigation.

The latest findings from docking indicate a binding affinity of -6.0 kcal/mol for capsaicin signifying an interaction with the RdRp protein. This implies that capsaicin could potentially thwart the replication mechanism of SARS CoV 2 thus curbing propagation.

### 4.3 Envelope Protein

The Envelope Protein model has a GMQE of 0.4 and a Global QMEANDisCo score of  $0.41 \pm 0.05$ , indicating lower confidence compared to the Spike and RdRp models. While the scores are lower, suggesting moderate reliability, the model is still useful and presents opportunities for further refinement. Given that the Envelope Protein is key to the homo-pentameric form of the virus's envelope, understanding its structure is important for targeting the protein with inhibitory compounds. With a 91.38% sequence identity, the model provides valuable insights into the Envelope protein.

Fresh insights from docking experiments have disclosed a binding affinity of -5.4 kcal/mol for capsaicin, with the Envelope protein. This interaction has the potential to disrupt either the assembly or functionality of the envelope thereby impeding its capacity to infect host cells.

### 4.4 Potential Implications

The consistency of the modelled structure with what is published in the literature will bolster the credibility of the study results. As an example of the interaction, other studies have reported that the

Spike protein interacts with the NAG ligand which supports our current findings. By comparing our model to the existing literature, discrepancies can suggest unique structural features or new binding sites that might be suitable for antiviral drug design. We have modelled potential targets for antiviral compounds sourced from Indonesian herbal plants by first identifying binding sites and ligands. For example, herbal compounds interacting with the Spike protein binding site could prevent the virus from entering host cells. High-confidence models, especially for RdRp and Spike protein, would be able to guide the development of new antiviral drugs. With this information, compounds can be created to block these binding sites and inhibit replication and spread.

#### **4.5 Limitations and Future Directions**

Modelling provides data, however, experimental validation is necessary to test these computational predictions. For some models, there was a confidence score of only moderate certainty, which is an indication to concentrate on these areas for further investigation. In future studies, the validation of these models, which may include techniques such as X-ray crystallography or cryo-electron microscopy, will be needed. Additionally, further potential modelling errors could be compounded by different herbal compounds and additional model refinement.