
Homology modeling and mutation prediction of ACE2 from COVID-19

**Purnawan Pontana Putra*, Annisa Fauzana, Khairunnisa Assyifa Salva,
Maya Sofiana, Intan Permata Sari, Henny Lucida**

*Faculty of Pharmacy, Universitas Andalas
Jl. Universitas Andalas, Padang, West Sumatra, Indonesia*

Submitted: 21-12-2020

Reviewed : 15-06-2021

Accepted: 29-06-2021

ABSTRACT

SARS-CoV-2 has become a pandemic in the world. The virus binds to the Angiotensin-Converting Enzyme 2 (ACE2) receptor, which is found in epithelial cells such as in the lungs, to generate the pathology of COVID-19. It is essential to analyze the characteristics of ACE2 in understanding the development of the disease and study potential new drugs. The analysis was carried out using computer simulations to speed up protein analysis that utilized Artificial Intelligence technology, databases, and big data. Homology modeling is a method to exhibit homologous of protein families, hence the model and arrangement of protein sequences modeled are established. This research aims to determine the possibility of mutations in ACE2 by performing the mutation prediction. The result shows reliable homologous modeling with the score of GA341, MPQS, Z-DOPE, and TSVMOD NO35 were 1; 1.28252; -0.47; and 0.793, respectively. Moreover, Gene Ontology (GO) analysis describes that ACE2 has a molecular transport function in cells while there are no mutations found occurred in ACE2 analyzed using SIFT and PROVEAN.

Keywords: ACE2; SARS-CoV-2; homology modeling; mutation prediction

***Corresponding author:**

Purnawan Pontana Putra

Faculty of Pharmacy, Universitas Andalas

Jl. Universitas Andalas, Padang, West Sumatra, Indonesia

Email: purnawanpp@phar.unand.ac.id

INTRODUCTION

SARS-CoV-2 is considered to be a new Betacoronavirus that infects humans (Lu et al., 2020). Phylogenetic analysis of the SARS-CoV-2 genome shows that the virus is closely related (88% similarity) to two coronaviruses that resemble severe acute respiratory syndrome (SARS) from bats collected in 2018 in eastern China (bat-SL-CoVZC45 and bat-SL-CoVZXC21) and are genetically different from SARS-CoV (approximately 79% similarity) and MERS-CoV (Lu et al., 2020). The coronavirus originated from bats (*Rhinolophus affinis*) from Yunnan Province, with 96.2% similarity in genomic sequence identities (Zhou et al., 2020).

The virus that causes COVID-19 enters through the ACE2 receptor pathway (Wan et al., 2020). ACE2 is found in alveolar type II epithelial cells, which are receptors for the SARS-COV2 virus (Zhao et al., 2020). To speed up the analysis of virus proteins, computational methods are used. The computational method is able to simulate a protein, drug, receptor with the help of a computer. The use of computation in the field of drug development is a new trend in drug discovery using Artificial Intelligence (AI), databases, and big data, thus helping to understand the properties of a compound much more quickly and effectively in screening new drug candidates (Idakwo et al., 2018).

The hinge-bending motion at the transmembrane helix does not change the shape of the ACE2 homodimer or Angiotensin-Converting Enzyme 2-Receptor Binding Domain (ACE2-RBD). This sterically flexible shape makes the spike trimer binding change its conformation. These changes indicate that ACE2 affects the opening of glycoproteins to affect the increase in the number of infections (Barros et al., 2021). The N82 residue in ACE2 facilitates contact with the SARS-CoV-2 protein in humans (Luan et al., 2020)

There is a salt bridge on Lys417 of the SARS COV 2 protein and Asp30 of ACE2, three hydrogen bonds on Tyr449, Gln493, and Gln498 of SARS CoV 2, and Asp38, Glu35, and Lys353 of ACE2 (Ali and Vijayan, 2020). D614G mutation increases binding efficiency to ACE2. The trimer in SARS-S exhibits an open conformation, thereby conferring greater contractility on the receptor binding (RBD) that interacts explicitly with ACE2 (Ozono et al., 2021).

Homology modeling is a method based on observations showing that all protein family members have the same fold shape. The use of this method relies on a homologous protein structure, where the protein structure has been determined experimentally (template) and allows the creation of a model starting from the protein sequence to be modeled (target). It is possible to observe the structural changes of each ACE2 variant that interacts with the SARS-COV-2 (Hussain et al., 2020). It can see possible binding stability and essential interactions with viruses (Ali and Vijayan, 2020). The resulting model from homology modeling can be used for ligand docking, macromolecular docking, ligand design, and studying catalytic mechanisms (Webb et al., 2014). It is hoped that the computational method will find out the active side of specific viral proteins so that it will help in the search for drugs and vaccines for COVID-19. With the help of computation, mutations can also be predicted, such as changes in amino acid substitutions, physical properties of proteins, insertions, and deletions of amino acids.

MATERIALS AND METHODS

Materials

The research material is used a sequence of Angiotensin-converting enzyme 2 (ACE2) from humans with the Uniprot® database code (ID Q695T7). It is full-length human ACE2, in the presence of a neutral amino acid transporter BOAT1. The software is used UCSF Chimera 1.15®. It has forcefield tools to perform molecular dynamic analysis such as Parameters for solvation energy (PARSE), AMBER, SWANSON, CHARMM, PEOEPB, and TYL06. In addition, web servers are used, such as NCBI®, BLAST, MolProbity®, Clustal Omega®, Cofactor®, Sorting Intolerant From Tolerant (SIFT®), Protein Variation Effect Analyzer (PROVEAN), Modeler® and Swiss-Model®.

It is utilized because it opens access and user-friendly. The modeler has a script file so that parameters can be adjusted.

Methods

Homology modeling using swiss model

The preparation of the ACE2 enzyme sequence was carried out by searching the Homo sapiens (Human) ACE2 enzyme sequence in the Uniprot® database. Data from Uniprot is then inputted, and a model is made. Furthermore, validation was carried out using Molprobity® and analyzed in 3D using UCSF Chimera.

Homology modeling using modeller

Making enzyme models using the homology modeling method begins with a search for templates with the help of the NCBI® web server www.ncbi.nlm.nih.gov/. An alignment search is carried out using The Basic Local Alignment Search Tool (BLAST) to find templates that are similar to Homo sapiens ACE2 sequence. Then the blastp is input by entering the FASTA Sequence to find out the similarities of the sequences. Obtained five templates that have similarities, then the FASTA sequence is downloaded. After that, Multiple Sequence Alignment is carried out using Clustal Omega <https://www.ebi.ac.uk/Tools/msa/clustalo/> then the Align is downloaded, then analyzed using the Modeler <https://modbase.compbio.ucsf.edu/modweb/> and analyzed the 3D form using UCSF Chimera.

Prediction analysis of ligand binding site, gene ontology and enzyme commission using cofactor

Protein data from the Data Bank with code 6M17 was uploaded at <https://zhanglab.ccmb.med.umich.edu/COFACTOR/> then analyzed the ligand-binding site and gene ontology.

Prediction of Mutations from ACE2

Mutation prediction analysis with interset G8790A using the Sorting Intolerant From Tolerant (SIFT) application <https://sift.bii.a-star.edu.sg/> and also used the PROVEAN application http://provean.jcvi.org/seq_submit.php. Sequence sequences in Fasta format are inputted into SIFT Sequences and PROVEAN.

RESULTS AND DISCUSSION

Homology modeling using swiss model

Setup the ACE2 sequence by inputting the Uniprot ID code file format (Q695T7). Obtained amino acid sequence which is used to find the template in making ACE2 model. The sequence of the ACE2 sequence can be seen in [Figure 1](#). This method is used to generate a previously unknown protein structure by "fitting" its sequence (target) into a known structure (template), with a certain degree of sequence homology (at least 30%) between the target and template (Sensoy et al., 2017). It is accurate for making structural models of proteins (Skariyachan and Garka, 2018).

```

>sp|Q695T7|S6A19_HUMAN Sodium-dependent neutral amino acid transporter B(0)AT1 OS=Homo
sapiens OX=9606 GN=SLC6A19 PE=1 SV=1
MVRVLVLPNPGLDARIPSLAELETIEQEEASSRPKWDNKAQYMLTCLGFCVGLGNVWRFYPY
LCQSHGGGAFMIPFLILLVLEGIPLLYLEFAIGQLRRGSLGVWSSIHPPALKGLGLASML
TSFMVGLYYNTIISWIMWYLFNSFQEPWPWSDCPLNENQTYVDECARSSPVDYFWYRET
LNISTSISDSGSIQWWMLLCLACAWSVLYMCTIRGIETTGKAVYITSTLPYVVLTIPLIR
GLTLKGATNGIVFLFTPNVTELAQPDTWLDAGAQQVFFSFLAFGLISFSSYNSVHNNCE
KDSVIVSIINGFTSVYVAIVVYVIGFRATQRYDDCFSTNLTLLINGFDLPEGNVQTQENF
VDMQQR CNASDPAAAYQLVFQTC DINAF LSEA VEGTGLAFIVFTEAITKMPLSPLWSVLF
FIMLFCLGLSSMFGNMEGVVPLQDLRVIPPKWPKEVLTGLICLGTFLIGFIFTLNSGQY
WLSLLDSYAGSIPLLIIFAFCEMFSVVYVYGVDRFNKDIEFMIGHKPNIFWQVTWRVVSPL
LMLIIFLFFFVVEVSQELTYSIWDPGYEEFPKSKISYPNWVYVVVVVIVAGVPSLTIPGY
AIYKLRNHCQKPGDHQGLVSTLSTASMGDLKY

```

Figure 1. Sequence of ACE2 sufferers of COVID-19

ACE2's 3D structure prediction uses the Swiss-Model webserver to obtain accurate modeling. In the Homology Modeling study, several steps must be taken, (1) It is identifying the target sequence using BLAST, (2) Perform Sequence Alignment, (3) Correct the alignment order, (4) Identify the backbone, (5) Loop modeling (6) Sidechain modeling using rotamer data, (7) Optimization of the model using energy minimization data, (8) Stereochemical model validation using the Ramachandran plot ([Gromiha et al., 2018](#)).

A template with swiss-model analysis, namely PDB 6M17 and 6M18. These two proteins are specific structures for identifying SARS-CoV-2 using the human ACE2 standard ([Yan et al., 2020](#)), which has 100 percent similarity. The tyrosinase enzyme modeling using the Swiss-Model® webserver was carried out automatically (automated mode).

Sequence alignment or better known as sequence alignment, is intended to determine the similarity between the target amino acid sequence and the template amino acid sequence. Sequence alignment between the ACE2 target and the template (PDB ID 6M17 and 6M18) was carried out using the assistance of the Swiss-Model® web server. The results of the alignment of the sequences with the Swiss-Model can be seen in [Figure 2](#) and [Figure 3](#).

The modeling of the ACE2 enzyme using the Swiss-Model is done by inputting data so that a 3D protein form is obtained. The results are obtained automatically by transferring atomic coordinates and aligning them with the target template. SWISS-MODEL relies on the OpenStructure computational structural biology framework ([Biasini et al., 2013](#)) and the database from Promod3 ([Waterhouse et al., 2018](#)). Swiss-Model® ACE2 3D model prediction assessment generated by using the QMEAN rating function. The QMEAN (Qualitative Model Energy Analysis) assessment function is a function that describes the geometric structure of proteins ([Benkert et al., 2008](#)). From the research data, the QMEAN score data obtained from the modeling is -4.50 for the 6M17 model and -4.92 for the 6M18, where the score shows the results of the modeling quality with low quality. The QMEAN score value ≤ -4.0 indicates that the model made has low quality.

Model Evaluation and Optimization The evaluation carried out on the ACE2 model from the results of homology modeling includes evaluating the stereochemical properties and spatial properties of the model. Evaluation of the stereochemical properties of the model was carried out using the Swiss Model web server by analyzing the Ramachandran and Molprobitry® plots to determine the acquisition of the clash score and molprobitry scores from the model.

Model_01	MVRLVLPNPGLDARIPSLAELETIEQEEASSRPKWDNKAQYMLTCLGFCVGLGNVWRFPYLCQSHGGGAF	70
6m17.1.A	-VRLVLPNPGLDARIPSLAELETIEQEEASSRPKWDNKAQYMLTCLGFCVGLGNVWRFPYLCQSHGGGAF	90
Model_01	MIPFLILLVLEGIPLLYLEFAIGQRRLRRGSLGVWSSIHHPALKGLGLASMLTSPFMVGLYNTIISWIMWYL	140
6m17.1.A	MIPFLILLVLEGIPLLYLEFAIGQRRLRRGSLGVWSSIHHPALKGLGLASMLTSPFMVGLYNTIISWIMWYL	160
Model_01	FNSFQEPLEPWSDCPLNENQGTGVDECARSSPVDYFWYRETINISTSISSDGSIQWMLLCLACAWSVLYM	210
6m17.1.A	FNSFQEPLEPWSDCPLNENQGTGVDECARSSPVDYFWYRETINISTSISSDGSIQWMLLCLACAWSVLYM	230
Model_01	CTIRGIETTGAKEYITSTLPYVVLTIPLIRGLTLKLGATNGIVFLFTPNVTELAQPDITWLDAGAQQVFFSFS	280
6m17.1.A	CTIRGIETTGAKEYITSTLPYVVLTIPLIRGLTLKLGATNGIVFLFTPNVTELAQPDITWLDAGAQQVFFSFS	300
Model_01	LAFGGLISFSSYNSVHNNCEKDSVIVSIINGFTSVYVAIVVYSVIGFRATQRYDDCFSTNILTLLINGFDL	350
6m17.1.A	LAFGGLISFSSYNSVHNNCEKDSVIVSIINGFTSVYVAIVVYSVIGFRATQRYDDCFSTNILTLLINGFDL	370
Model_01	PEGNVTOENFVDMQQRNASDPAAYQLVFTQCDINAFLEAVEGTGLAFIVFTEAITKMPLSPLWSVLF	420
6m17.1.A	PEGNVTOENFVDMQQRNASDPAAYQLVFTQCDINAFLEAVEGTGLAFIVFTEAITKMPLSPLWSVLF	440
Model_01	FIMLFCLGLSSMFGNMEGVVVPLQDLRVIPKPKPEVLTGLICLGTFLIGFIFTLNSGQYWSLLDSYAG	490
6m17.1.A	FIMLFCLGLSSMFGNMEGVVVPLQDLRVIPKPKPEVLTGLICLGTFLIGFIFTLNSGQYWSLLDSYAG	510
Model_01	SIPLLIIAFCEMFSVVVYGVDRFNKDIIEFMIGHKPNIFWQVTRVVSPLLMLIIFLFFVVEVSQELTY	560
6m17.1.A	SIPLLIIAFCEMFSVVVYGVDRFNKDIIEFMIGHKPNIFWQVTRVVSPLLMLIIFLFFVVEVSQELTY	580
Model_01	SIWDPGYEYEFKSKKISYPNWWVVVVIVAGVPSLTIPGYAIYKLIIRNHCQKPGDHQGLVSTLSTASMNG	630
6m17.1.A	SIWDPGYEYEFKSKKISYPNWWVVVVIVAGVPSLTIPGYAIYKLIIRNHCQKPGDHQGLVSTLSTASMNG	650
Model_01	DLKY	634
6m17.1.A	DLKY	654

Figure 2. Model-Template Alignment 6M17

Model-Template Alignment		
Model_01	MVRLVLPNPGLDARIPSLAELETIEQEEASSRPKWDNKAQYMLTCLGFCVGLGNVWRFPYLCQSHGGGAFMIPFL	75
6m18.1.A	-VRLVLPNPGLDARIPSLAELETIEQEEASSRPKWDNKAQYMLTCLGFCVGLGNVWRFPYLCQSHGGGAFMIPFL	95
Model_01	ILLVLEGIPLLYLEFAIGQRRLRRGSLGVWSSIHHPALKGLGLASMLTSPFMVGLYNTIISWIMWYLFNSFQEPLEP	150
6m18.1.A	ILLVLEGIPLLYLEFAIGQRRLRRGSLGVWSSIHHPALKGLGLASMLTSPFMVGLYNTIISWIMWYLFNSFQEPLEP	170
Model_01	SDCPLNENQGTGVDECARSSPVDYFWYRETINISTSISSDGSIQWMLLCLACAWSVLYMCTIRGIETTGAKEYI	225
6m18.1.A	SDCPLNENQGTGVDECARSSPVDYFWYRETINISTSISSDGSIQWMLLCLACAWSVLYMCTIRGIETTGAKEYI	245
Model_01	TSTLPYVVLTIPLIRGLTLKLGATNGIVFLFTPNVTELAQPDITWLDAGAQQVFFSFLAFGGLISFSSYNSVHNNCE	300
6m18.1.A	TSTLPYVVLTIPLIRGLTLKLGATNGIVFLFTPNVTELAQPDITWLDAGAQQVFFSFLAFGGLISFSSYNSVHNNCE	320
Model_01	KDSVIVSIINGFTSVYVAIVVYSVIGFRATQRYDDCFSTNILTLLINGFDLPEGNVTOENFVDMQQRNASDPAAY	375
6m18.1.A	KDSVIVSIINGFTSVYVAIVVYSVIGFRATQRYDDCFSTNILTLLINGFDLPEGNVTOENFVDMQQRNASDPAAY	395
Model_01	AQLVFTQCDINAFLEAVEGTGLAFIVFTEAITKMPLSPLWSVLFIMLFCLGLSSMFGNMEGVVVPLQDLRVIP	450
6m18.1.A	AQLVFTQCDINAFLEAVEGTGLAFIVFTEAITKMPLSPLWSVLFIMLFCLGLSSMFGNMEGVVVPLQDLRVIP	470
Model_01	PKWPEVLTGLICLGTFLIGFIFTLNSGQYWSLLDSYAGSIPLLIIAFCEMFSVVVYGVDRFNKDIIEFMIGHK	525
6m18.1.A	PKWPEVLTGLICLGTFLIGFIFTLNSGQYWSLLDSYAGSIPLLIIAFCEMFSVVVYGVDRFNKDIIEFMIGHK	545
Model_01	PNIFWQVTRVVSPLLMLIIFLFFVVEVSQELTYSIWDPGYEYEFKSKKISYPNWWVVVVIVAGVPSLTIPGY	600
6m18.1.A	PNIFWQVTRVVSPLLMLIIFLFFVVEVSQELTYSIWDPGYEYEFKSKKISYPNWWVVVVIVAGVPSLTIPGY	620
Model_01	AIYKLIIRNHCQKPGDHQGLVSTLSTASMNGDLKY	634
6m18.1.A	AIYKLIIRNHCQKPGDHQGLVSTLSTASMNGDLKY	654

Figure 3. Model-template alignment 6M18

The Ramachandran plot is a two-dimensional graph depicting amino acid residues in the enzyme structure, where the angles ϕ (phi) as the x-axis and ψ (psi) as the y-axis are divided into four quadrants, the angular values of the spectrum from each axis range from -180° to $+180^\circ$ (Choudhuri, 2014). It is plot visualizes the do-and-not areas in the plane of dihedral angles. The quality of the homolog is poor if many forbidden regions are found (Wiltgen, 2018). Visualization of the results of the Ramachandran plot can be seen in Figure 4.

The evaluation of contacts between atoms is based on the clashscore assessment. Evaluation results can be seen in Table 1. Clashscore is the steric number that overlaps per 1000 atoms. The best

clashscore value is <100. It has met the standard value in homologous modeling. The molprobability score is a combination of clashscore, percentage forbidden regions, and the percentage of bad rotamer, which reflects the value of the crystallographic resolution appropriate for the model. Judging from the Molprobability score, the data model has met the standard because it is <84.

Table 1. Evaluation of homology modeling using the swiss model

Model	Clashscore	Plot Ramachandran		Skor MolProbit
		Favored Regions	Outlier	
6M17	0.94	93.37%	0.50%	1.21
6M18	10.32	91.19%	1.11%	2.43

Both models do not meet the standard value of the standard value (> 98%). The two models' outlier value does not enter the standard criteria, namely <0.05%. Low-quality result of Swiss Model because steric hindrance between the C β side chain and the central chain atoms.

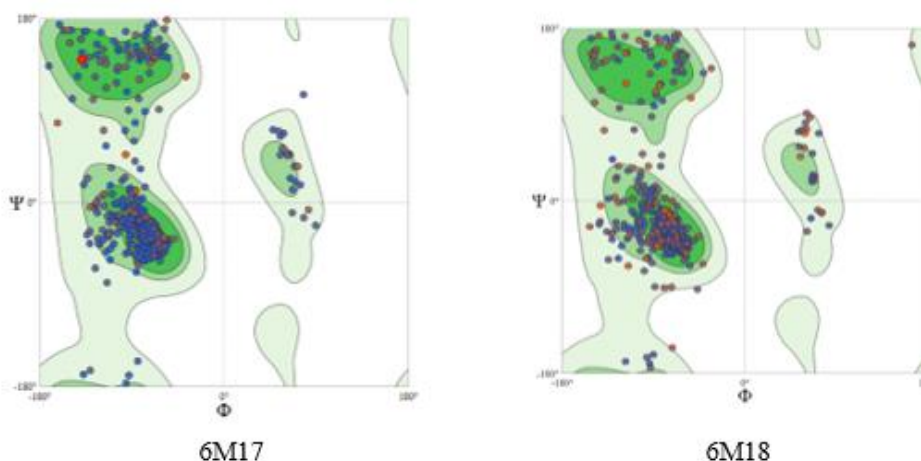


Figure 4. Ramachandran plot of ACE2 using the Swiss-model

Homology modeling using modelers

The search for templates was carried out using the assistance of the NCBI® BLAST web server to find the protein structure that has the highest sequence similarity to the target amino acid sequence, which has been known experimentally in 3D structure. One of the tools that NCBI has that is commonly used to find similar sequences in the BLAST (Basic Local Alignment Search Tool), which looks for similarities in templates with the highest percentage of identity.

Table 2. The results of the sequence analysis using BLAST

Accession	Query Cover	E-Value	Percent Identity	Description
6M17_A	99%	0	100%	Chain A. ACE2 (Human)
4XPH_A	84%	2.00E-79	43.65%	Chain A. Transporter (Drosophila melanogaster)
4XPT_A	84%	2.00E-79	43.51%	Chain A. Dopamine (Drosophila melanogaster)
4XNX_A	84%	2.00E-78	43.32%	Chain A. transporter (Drosophila melanogaster)
4XP5_A	84%	3.00E-78	43.32%	Chain A. Dopamine Transporter, Isoform B (Drosophila melanogaster)

The results of the sequence analysis using BLAST can be seen in Table 2, selected 5 sequences that have a similarity > 43%—followed by multiple sequence alignment using Clustal Omega. The percentage of similarity in identity is shown by the Query Cover, which is input into BLAST. A good sequence is one that has a similarity sequence > 30%. The Three-dimensional space (3D) form of modeling can be seen in Figure 5. It was obtained three 3D models. Modeler software brought one model with the protein data bank code 6M17. The Swiss model software received two 3-dimensional forms, namely the protein data bank code 6M17 and 6M18

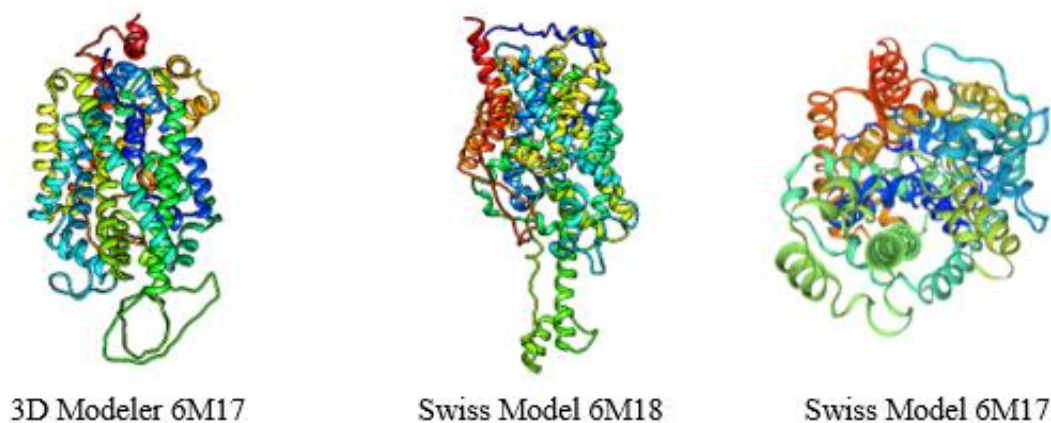


Figure 5. Model 3D ACE2 Homologous

The results of the thread sequence and the model evaluation values using a modeler can be seen in Table 3. Model evaluation of the model is carried out, including the assessment of E-Value, GA341, MPQS, z-DOPE, and TSVMOD NO35. The results obtained are included in the criteria (reliable).

Table 3. Evaluation of homology modeling using modeller

Original Database ID	<u>E-Value</u>	<u>GA341</u>	<u>MPQS</u>	<u>Z-DOPE</u>	<u>TSVMOD NO35</u>
6M17_A Chain A	0	1	1.28252	-0.47	0.793

E-Value is a parameter used to compare the size form's suitability between the design and the actual shape of the score database. The standard score is <0.0001. GA341 is a value that can be trusted from statistical modeling with a standard value of ≥ 0.7 . The MPQS is a composite score consisting of the template sequence identity, z-Dope value, and GA341. Z-Dope is a value that describes the folds in the template model with criteria that meet the standards, namely <0 (Webb and Sali, 2016). TSVMOD NO35 is an estimated value overlap with the standard value, namely $\geq 40\%$. From the evaluation of the data obtained, the homologous modeling above fulfills the standard criteria.

The ACE2 variant exhibits a similar binding affinity for the SARS-CoV-2 spike protein as observed in the structure of the wild-type ACE2 and SARS-CoV-2 spike protein complex. However, the ACE2 alleles, rs73635825 (S19P) and rs143936283 (E329G), show marked variations in their intermolecular interactions with viral spike proteins (Hussain et al., 2020). It has the strongest binding interaction. The earliest isolates of SARS-CoV-2 were surprisingly well adapted to human ACE2, potentially explaining its rapid transmission (Piplani et al., 2020). Twenty residues interact with proteins and can bind to ACE2, of which five residues (Val445, Thr478, Gly485, Phe490, and

Ser494). The interaction between ACE2 and the tertiary structure of the protein is different from that of ACE2 and the RBD protein monomer (Sakkiah et al., 2021).

Analysis of ligand binding site prediction, gene ontology (GO) using cofactor

Analysis of the Predicted Gene ontology and predicted binding site using the Cofactor can be seen in Table 4. Cofactor is an application that can predict the structure and function of proteins. Cofactor can categorize proteins such as Gene Ontology, Enzyme Commission, and Ligand-Binding sites from analogs and template homologs (Zhang et al., 2017).

Table 4. Predicted gene ontology

	Molecular Function (MF)	Biological Process (BP)	Cellular Component (CC)
Predicted GO 6M17	Transporter Activity	Transport	Cell Part
Cscore	0.99	0.97	1
GO Term	GO:0005215	GO:0006810	GO:0044464

Gene ontology (GO) is a framework and concept for describing genes' function in all organisms. Gene ontology facilitates computational interpretation of biological systems. GO Term is a unique code to facilitate and classify gene ontologies (Dessimoz and Škunca, 2017). In terms of gene ontology, ACE2 has biological and molecular functions as a transporter activity. ACE2 is involved in the movement of molecules and ions from outside to into the cell or vice versa. It is essential to know the ligand and binding site in analyzing protein because this is the place where the drug will attach later. This study predicts ligand and binding sites, the results of which can be seen in Table 5.

Cs Score is a value in the range 0-1, the closer to the value 1, the more similar to the physical-chemical properties of the protein with the database. The potential Ligand is Leucine with predicted binding site residues number 49, 51, 52, 53, 125, 129, 277, 278, 280, 283, 431, 435. Cofactor was predicted ligand and binding site by collecting binding site data and mapping it from the query. The ligands originating from the template are placed in the query structure using the query matrix and the template binding site. It was calculated with Monte Carlo then compute the ligand pose to figure the motion and rotation of the atoms. Grouping ligands and their superposition obtain the final ligand into queries with a cut off of 8Å (Zhang et al., 2017).

Table 5. Protein template with similar binding site

PDB Hit	Cs Score	Ligand Name	Predicted binding site residue
3gjcB	0.28	Leucine	49,51,52,53,125,129,277,278,280,283,431,435
3f4jA	0.15	Sodium Cation	49,50,51,53,54,278,310,314
3gwwA	0.12	Fluoxetine	52,53,57,61,129,132,277,395,396,483,486
3gwuA	0.12	Sertraline	56,57,129,277,395,396,482,486,491
2qjuA	0.12	Desipramine	52,56,57,62,277,394,395

Prediction of Mutations from ACE2

Prediction of mutations using SIFT obtained ACE2 had not mutations. The mutations that were chosen to be compared were interset G8790A. G8790A is a gene that has a mutation in ACE2 hypertensive patients (Li, 2012). It was able to predict amino acid substitution that affects protein function based on sequence homology and amino acid physical properties (Sim et al., 2012). Amino acid position was obtained from 1 to 634. Mutation analysis was also carried out using the PROVEAN

application; there were no mutations from the results. PROVEAN is a software that can predict amino acid substitution or indel that affects the biological function of a protein. PROVEAN was able to filter sequence variants on proteins (Choi and Chan, 2015). The subsequent significant study is molecular dynamics simulation to explain the interaction between RBD and ACE2 using NAMD and GROMACS. Free energy analysis should be carried out using the Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA) method to understand the Thermodynamics of Biomolecular Systems.

CONCLUSION

Data Bank Protein Homologous Modeling shows the ID ACE2 protein data bank, namely 6M17 and 6M18. Modeling using a Modeler is included in the Reliable criteria where the modeling results can be accepted. From the Gene Ontology analysis, ACE2 has a function as a transport molecule in cells. The result of ACE2 ligand prediction is Leucine with Predicted Binding Site Residue, namely 49,51,52,53,125,129,277,278,280,283,431,435. From the prediction of mutations using SIFT and PROVEAN, there were no mutations in ACE2.

ACKNOWLEDGMENT

The author would like to thank the Universitas Andalas for supporting and finance this research. With Contract Number: T/18/UN.16.17/PT.01.03/KO-RDP/2020.

REFERENCES

- Ali, A., & Vijayan, R. (2020). Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-71188-3>
- Barros, E. P., Casalino, L., Gaieb, Z., Dommer, A. C., Wang, Y., Fallon, L., Raguette, L., Belfon, K., Simmerling, C., & Amaro, R. E. (2021). The flexibility of ACE2 in the context of SARS-CoV-2 infection. *Biophysical Journal*, 120(6), 1072–1084. <https://doi.org/10.1016/j.bpj.2020.10.036>
- Benkert, P., Tosatto, S. C. E., & Schomburg, D. (2008). QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function and Genetics*, 71(1), 261–277. <https://doi.org/10.1002/prot.21715>
- Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johner, N., Schenk, A. D., Philippsen, A., & Schwede, T. (2013). OpenStructure: An integrated software framework for computational structural biology. *Acta Crystallographica Section D: Biological Crystallography*, 69(5), 701–709. <https://doi.org/10.1107/S09074444913007051>
- Choi, Y., & Chan, A. P. (2015). Provean web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31(16), 2745–2747. <https://doi.org/10.1093/bioinformatics/btv195>
- Choudhuri, S. (2014). Additional bioinformatic analyses involving protein sequences. *Bioinformatics for Beginners*, 183–207. <https://doi.org/10.1016/b978-0-12-410471-6.00008-6>
- Dessimoz, C., & Škunca, N. (2017). *The Gene Ontology Handbook* (Vol. 1446). Humana Press. <https://doi.org/10.1007/978-1-4939-3743-1>
- Gromiha, M. M., Nagarajan, R., & Selvaraj, S. (2018). Protein structural bioinformatics: An overview. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–3, 445–459. <https://doi.org/10.1016/B978-0-12-809633-8.20278-1>
- Hussain, M., Jabeen, N., Raza, F., Shabbir, S., Baig, A. A., Amanullah, A., & Aziz, B. (2020). Structural variations in human ACE2 may influence its binding with SARS-CoV-2 spike protein. *Journal of Medical Virology*, 92(9), 1580–1586. <https://doi.org/10.1002/jmv.25832>
- Idakwo, G., Luttrell, J., Chen, M., Hong, H., Zhou, Z., Gong, P., & Zhang, C. (2018). A review on machine learning methods for in silico toxicity prediction. *Journal of Environmental Science and Health - Part C Environmental Carcinogenesis and Ecotoxicology Reviews*, 36(4), 169–191.

- <https://doi.org/10.1080/10590501.2018.1537118>
- Li, Y. Y. (2012). Lack of association of ACE2 G8790A gene mutation with essential hypertension in the Chinese Population: A meta-analysis involving 5260 subjects. *Frontiers in Physiology*, 3 SEP. <https://doi.org/10.3389/fphys.2012.00364>
- Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, 395(10224), 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
- Luan, J., Lu, Y., Jin, X., & Zhang, L. (2020). Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochemical and Biophysical Research Communications*, 526(1), 165–169. <https://doi.org/10.1016/j.bbrc.2020.03.047>
- Ozono, S., Zhang, Y., Ode, H., Sano, K., Tan, T. S., Imai, K., Miyoshi, K., Kishigami, S., Ueno, T., Iwatani, Y., Suzuki, T., & Tokunaga, K. (2021). SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-21118-2>
- Piplani, S., Singh, P. K., Winkler, D. A., & Petrovsky, N. (2020). *In silico comparison of spike protein-ACE2 binding affinities across species; significance for the possible origin of the SARS-CoV-2 virus*. 2.
- Sakkiah, S., Guo, W., Pan, B., Ji, Z., Yavas, G., Azevedo, M., Hawes, J., Patterson, T. A., & Hong, H. (2021). Elucidating interactions between SARS-CoV-2 trimeric spike protein and ACE2 using homology modeling and molecular dynamics simulations. *Frontiers in Chemistry*, 8. <https://doi.org/10.3389/fchem.2020.622632>
- Sensoy, O., Almeida, J. G., Shabbir, J., Moreira, I. S., & Morra, G. (2017). Computational studies of G protein-coupled receptor complexes: Structure and dynamics. *Methods in Cell Biology*, 142, 205–245. <https://doi.org/10.1016/bs.mcb.2017.07.011>
- Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40(W1). <https://doi.org/10.1093/nar/gks539>
- Skariyachan, S., & Garka, S. (2018). Exploring the binding potential of carbon nanotubes and fullerene towards major drug targets of multidrug resistant bacterial pathogens and their utility as novel therapeutic agents. *Fullerenes, Graphenes and Nanotubes: A Pharmaceutical Approach*, 1–29. <https://doi.org/10.1016/B978-0-12-813691-1.00001-4>
- Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by the novel Coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS Coronavirus. *Journal of Virology*, 94(7). <https://doi.org/10.1128/jvi.00127-20>
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). Swiss-model: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/nar/gky427>
- Webb, B., Eswar, N., Fan, H., Khuri, N., Pieper, U., Dong, G. Q., Sali, A., Francisco, S., & Francisco, S. (2014). Author's personal copy comparative modeling of drug target proteins ☆. In *Chemistry, Molecular Sciences and Chemical Engineering*. Elsevier Inc. <https://doi.org/10.1016/B978-0-12-409547-2.11133-3>
- Webb, Benjamin, & Sali, A. (2016). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, 2016, 5.6.1-5.6.37. <https://doi.org/10.1002/cpbi.3>
- Wiltgen, M. (2018). Algorithms for structure comparison and analysis: Homology modelling of proteins. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 1–

- 3, 38–61. <https://doi.org/10.1016/B978-0-12-809633-8.20484-6>
- Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, 367(6485), 1444–1448. <https://doi.org/10.1126/science.abb2762>
- Zhang, C., Freddolino, P. L., & Zhang, Y. (2017). Cofactor: Improved protein function prediction by combining structure, sequence and protein-protein interaction information. *Nucleic Acids Research*, 45(W1), W291–W299. <https://doi.org/10.1093/nar/gkx366>
- Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., & Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV. *BioRxiv*, 2020.01.26.919985. <https://doi.org/10.1101/2020.01.26.919985>
- Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. Di, Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., ... Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>

