

**FULL PAPER**

# Applying machine learning to define best models for screening potential antiviral compound for sars-cov-2 protein target

Herin Setianingsih<sup>a,\*</sup> | Nasywa Zahra Sajida Tsuroyya<sup>a</sup> | Hilyatuz Zahroh<sup>b</sup> | Eka Diyah Putri Lestari<sup>a</sup> | Didik Huswo Utomo<sup>a</sup> | Muhammad Rezki Rayak<sup>c</sup>

<sup>a</sup>Faculty of Medicine, Hang Tuah University, Surabaya, Indonesia

<sup>b</sup>Faculty of Medicine, Yarsi University, Surabaya, Indonesia

<sup>c</sup>Eijkman Institute for Molecular Biology, Jakarta, Indonesia

Protein target identification is a crucial part of drug discovery. This study used a machine learning approach to screen the potential target from marine natural products. A total of 6,314 compounds from 11 marine taxa were collected from CMNPD or the Comprehensive Marine Natural Products Database as drug repurposing candidates for COVID-19. SARS-CoV-2 well identified proteins, including Spike, PLpro, Mpro, Nucleocapsid, ORF9b, ORF3a, and ORF8, are designed as protein targets. The supervised learning classification method that we use consists of three data processing, namely logistic regression (LR), super vector machine (SVM), and random forest (RF). Machine learning is carried out using algorithm found in scikit-learn. We also carried out a deep learning model approach and predict active compounds by applying the algorithm to h2o.ai. Finally, reverse docking approach was also used to get reliable result. The result revealed that compounds from bryozoan, sponge, and bacteria have the best binding affinity score for spike proteins. The best model of machine learning is the LR model. The compilation results of screening predictions from both machine learning and deep learning showed more consistent results and were proven to show more stable bond interactions than compounds that were predicted to have activity in just one of the screening methods.

**\*Corresponding Author:**

Herin Setianingsih

Email: [herinsetianingsih1@gmail.com](mailto:herinsetianingsih1@gmail.com)

Tel.: +81232555576

**KEYWORDS**

Antiviral; SARS-CoV-2; bryozoan; sponge; bacteria; machine learning.

**Introduction**

COVID-19 or Coronavirus Disease has come to the global community's attention, including Indonesia because it has caused respiratory disorders with high rates of transmission and death in more than 100 countries. Coronavirus (CoV) belongs to the family Coronaviridae, which can cause zoonotic diseases with

common flu symptoms to severe and deadly. Coronavirus types that have spread and are known to cause infectious diseases in severe categories include the Middle East Respiratory Syndrome Coronavirus (MERS-CoV), Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV), and most recently COVID-19 disease caused by SARS-CoV 2. The pandemic

is known to stem from a pneumonia outbreak in Wuhan City, China, at the end of 2019 [1].

The problem is further complicated because there is still no effective treatment to treat this infectious disease, and even vaccine has already been successfully developed, treatment for COVID patients caused by the newest variant is still a major problem [2]. In addition, for about a year since the first case, the mutation of this virus is also known to be very fast and has formed several variants, namely alpha (B.1.1.7), beta (B.1.351), and gamma (P.1) to the newest variant with more fast transmission at this time is the delta (B.1.617.2) and lambda (C.37) variant [3]. The development and trial of new antiviral drugs and vaccines take a long time, so efforts to reposition drugs that have been approved by the Food and Drug Administration (FDA) are the best alternatives. However, relying too much on synthetic antiviral drugs is also considered to be less effective in its application, especially in most developing countries. So, looking for alternative antivirals from herbal bioactive compounds and other natural products can be a better solution, particularly bioactive from marine natural products, which have received much attention from international researchers from the early 2000s the most potential sources of durable discovery [4,5]. In addition, Indonesia, as an archipelagic country affected by the pandemic, has a variety of natural marine natural products. Therefore, conducting drug discovery research from bioactive compounds of marine natural products is very suitable.

Several studies have suggested the concept of stabilizing effective interactions proteins in the discovery and development of antivirals [6,7,8]. This *in silico* approach can significantly reduce the time and cost of drug development, especially the toxicity and pharmacokinetic profile of such drugs that are widely known and tested. However, if the sample being screened is a bioactive compound, a new problem will arise. The problem is the total number of bioactive compounds that are very

large for each source of the organism. Thus, a machine learning approach is needed to simplify the screening process and increase the statistical confidence score of the selection of compounds that are predicted to have efficacy.

## Methods

### *Ligand datasets*

Candidates for bioactive compounds, including 6,314 compounds from 11 taxa, were collected from CMNPD or the Comprehensive Marine Natural Products Database (<http://https://www.cmnpd.org/>) [9]. These taxa include green algae, red algae, brown algae, fungi, bacteria, echinoderms, ascidian, mollusks, bryozoans, cnidarians, and sponges. Furthermore, the 3D structures of the drug compounds were collected in SDF format from PubChem (<https://pubchem.ncbi.nlm.nih.gov>) database [10]. The antiviral function scores of each compound were predicted and collected from pass online software (<http://www.way2drug.com/PASSOnline/>) [11].

### *SARS-CoV-2 protein target*

As we did in previous research, we studied the genome structure of SARS-CoV-2 and the resulting proteins. The genome size of SARS-CoV-2 ranges from 29,903 bp (NC\_045512.2). SARS-CoV-2 has four structural proteins, 16 non-structural proteins, and 9 accessory proteins [12]. From the screening results, there were 14 protein structures from SARS-CoV-2 that we used as molecular docking repurposing targets. In addition, we used the ACE2 protein structure of humans as a target. The binding affinity results showed that only 8 of the 14 proteins used had good and significant results [13]. These proteins are Spike, PLpro, Mpro, Nucleocapsid, ORF9b, ORF3a, ORF8, and ACE2. Each structure of those proteins was downloaded from the RCSB Database of PDB (<https://www.rcsb.org/>) [14] in format (.pdb) (Table 1).

**TABLE 1** List of protein targets for virtual screening

No.	Macromolecules	PDB ID	Method	Resolution	Seq Length	Positions	Chains	Ref.
1	Spike Protein (prefusion)	6VSB	EM	3.46 Å	1288	1-1208	A/B/C	[12]
2	nsp3/PLpro	6WX4	XRD	1.66 Å	326	1562-1879	D	[16]
3	Mpro/nsp5/3CLpro	7BQY	XRD	1.70 Å	306	3264-3569	A	[15,16]
4	Nucleocapsid Protein	6M3M	XRD	2.70 Å	136	41-174	A/B/C/D	[12]
5	orf9b	6Z4U	XRD	1.95 Å	97	1-97	A/B	[12]
6	orf3a	6XDC	EM	2.90 Å	284	1-275	A/B	[17]
7	orf8	7JTL	XRD	2.04 Å	107	18-121	A/B	[18]
8	ACE2	6M18	EM	2.90 Å	814		B, D	[17]

### Ligands screening using machine deep learning

The feature extraction process is carried out according to data on types of drugs (small molecules/drugs) compounds that have the potential to be selected based on the clinical trial level, clinical trial level. Drugs or compounds with the best clinical test results were used as active compounds (molecules) for machine learning. The active compound was then combined with the decoy compound (fake) from DUDE Docking ([www.dude.docking.org](http://www.dude.docking.org)) [19]. A total of 300 decoys from and DUD-E was then combined with 96 selected active compounds for further feature extraction into the Klekota-Roth .csv bit fingerprint column (with a total of 2048 fingerprints).

The supervised learning classification method that we use consists of three data processing, namely logistic regression (LR), super vector machine (SVM), and random forest (RF). Machine learning was carried out using jupyter notebook with a machine learning algorithm found in scikit-learn (<https://scikit-learn.org/stable/>) with the distribution of training and testing from data with a proportion of 70% training and 30% testing [20]. We also carried out a deep learning model approach and predicted active compounds by applying the algorithm to h2o.ai (<https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html>) with the distribution of training

and testing from data with a proportion of 70% training and 30% testing [21]. The deep learning was applied using a double hidden layer, where the first hidden layer has 400 nodes and the second hidden layer has 100 nodes, with data loops (epochs = 500). Confirmation tests in the form of accuracy, precision, and area under the curve (AUC/ROC) were calculated based on the sensitivity and specificity values, to determine the accuracy of the model.

### Preparation of ligands macromolecules

Protein preparation was performed using Discovery Studio 2021 program by BIOVIA. Water molecules and ligands were removed from the 3D structures. The prepared structures were saved back in PDB format. Furthermore, the preparation process was also carried out on ligands using PyRx 0.9.8 software to lower or minimize the free energy of the ligands as well as convert compounds into AutoDock ligands. The pre-prepared ligands were then saved in PDB format.

### Molecular docking and visualization of protein-ligand interaction

The docking process is carried out using an inverse docking approach. The first group of ligands was 675 compounds that were predicted to have antiviral activity in both machine learning and deep learning

prediction methods. There were also the second group of 665 compounds predicted to have antiviral activity in either machine learning or deep learning prediction method. This process was carried out with the help of the AutoDock Vina 1.1 program from Pyrx 0.9.8 Software [22,23]. After docking was complete, the results were saved in PDB format and then visualized and interpreted using PyMOL v.2.3.2.1, and Discovery Studio 2021 Client.

## Results and discussion

### *Virtual screening using machine learning and deep learning*

Each of the used machine learning methods, LR, SVM, and RF, showed different results. LR appeared as the best model for data testing among others with an accuracy of 0.879 (88%), while SVM and RF resulted in 0.822 (82%) and 0.839 (84%) for the accuracy, respectively. The LR sensitivity was 0.889 (89%), while for SVM and RF were 0.889 (89%) and 0.846 (85%). For AUC, LR also had the best result with 0.915 (91%), while SVM and RF had 0.896 (89%) and 0.834 (83%), respectively. To increase the predictive ability and add a comparison to the machine learning

models, we also used the deep learning method for screening the ligands. The results showed that our deep learning model had very good scores with an accuracy value of 0.87 (87%), a precision of 1.0 (100%), and a specificity of 1.0 (100%). Following these results, we acquired the AUC of 0.904 (90%) with an area under the Precision-Recall Curve of 0.968 (97%).

### *Molecular docking as confirmation test*

Out of the 675 bioactive compounds used in molecular docking, bioactive compounds from bryozoans and sponges showed the most negative binding affinity values for almost all molecular targets, except for ORF9b and ORF3a targets (Table 2). Therefore, we chose two targets with the most potential and most used as targets for antiviral drug discovery, Spike and Mpro [24]. Furthermore, the pipeline confirmation was carried out whether the active compound predicted to be active in both ligand-based screening methods, both machine learning and deep learning (BM), had a better binding affinity than the compound predicted to be active only in one of the screening methods (OM), machine learning (M) or deep learning (D).

**TABLE 2** Binding affinity values of active compounds predicted to be active by both machine learning and deep learning methods (BM)

No.	Target Protein	Binding Affinity (Kcal/mol)	Source Organism
1	Spike (S)	-8.9	Bryozoan
2	NSP3 (PLpro)	-7.7	Bryozoan
3	NSP5 (Mpro)	-8.3	Sponge
4	Nucleocapsid Protein (N)	-9.5	Sponge
5	ORF9b	-9.6	Fungi
6	ORF3a	-8.6	Fungi
7	ORF8	-9.3	Sponge
8	ACE2 (human)	-9.2	Bryozoan

Based on the docking results, compounds from bryozoan, sponge, and bacteria showed the best binding affinity score for spike proteins. Meanwhile, in terms of the way the compounds were screened, it was noticed that

the BM method had better results except for bacteria. The binding affinity scores of the BM method were almost all better than the OM method, wherein the bryozoan was -8.9 kcal/mol compared to -8.6 kcal/mol, in

sponges -8.8 kcal/mol compared to -8.6 kcal/mol. moles, except for bacteria which both have a score of -8.6 kcal/mol (Table 3). In addition, the OM method shows inconsistent results, where the deep learning (D) sponge

method predicts the activity described by the binding affinity score better than machine learning (K). However, on bacteria and bryozoan machine learning shows better predictions.

**TABLE 3** Binding affinity values between Spike and top five active compounds that predicted to be active by both methods (BM and OM)

Source		Bioactive		Binding Affinity (Kcal/mol)	
BM	OM	BM	OM	BM	OM
Bryozoan	D_ Sponge	7- Ketocholesterol	6-Oxofascaplysin	-8.9	-8.6
Sponge	M_ Bacteria	Xestosaprol M	Amycocyclopiazonic Acid	-8.8	-8.6
Bacteria	M_ Bryozoan	Shewanelline C	Flustramine M	-8.6	-8.6
Sponge	M_ Bacteria	Tedarene A	(16Z)-4,11-Dimethyl-1,8-Diazatetracyclo [6.6.4.02,7.09,14] Octadeca-2(7),3,5,9(14),10,12,16-Heptaene-15,18-Dione	-8.5	-8.5
Sponge	D_ Sponge	Cinanthrenol A	(+)-Frondosin B	-8.5	-8.5

For Mpro, compounds predicted to have the best affinity are originated from distinct profile of organisms, including bryozoans, sponges, and molluscs, to spikes. In addition, there are compounds from sponges, only predicted from deep learning method, that have a binding affinity score much better than the top five compounds from the BM method. However, both methods showed that the compound from the sponge was the most potent compound where the score from the BM method was -8.3 kcal/mol, much higher than the OM method, which was -8.8 kcal/mol. In addition to compounds from sponges, compounds from molluscs also showed that the OM method had a lower binding affinity score than BM, which was -8.2 kcal/mol

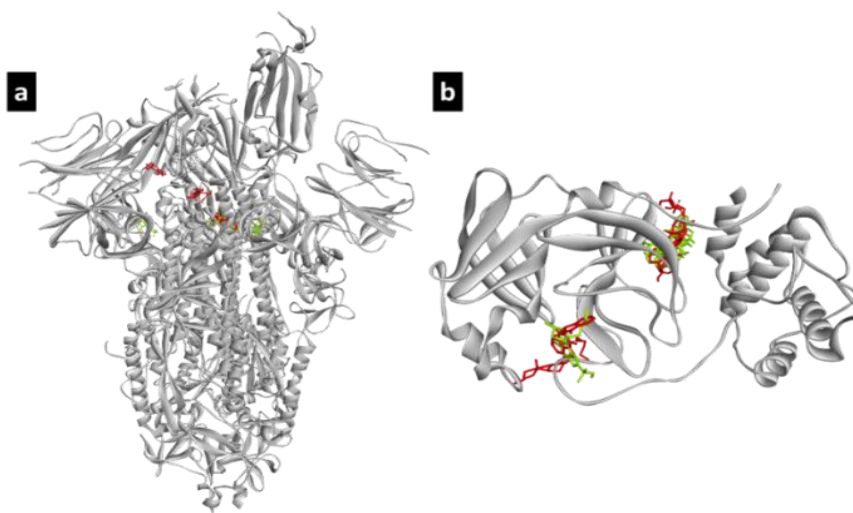
compared to -8 kcal/mol (Table 4). While compounds from bryozoans showed better results where the BM method had a lower score, namely -8.2 kcal/mol compared to -8 kcal/mol. However, the equation that is still visible is the inconsistent results of the OM method, where deep learning (D) shows that compounds from sponges and molluscs are better than machine learning methods. Furthermore, bryozoan shows that the machine learning (M) method is better. These results also support the results of the spike where sponge activity is better predicted by the deep learning (D) method, while the bryozoan is better predicted by the machine learning (M) method.

**TABLE 4** Binding affinity values between Mpro and top 5 active compounds that predicted to be active by both methods (BM and OM)

Source of organism		Compound name		Binding Affinity (Kcal/mol)	
BM	OM	BM	OM	BM	OM
Sponge	D_	Xestosaprol M	Meridine	-8.3	-8.8
Bryozoan	D_	7-Ketocholesterol	Kuanoniamine A	-8.2	-8.2
Bryozoan	M_	(3S,7R,8S,9S,10R,13R,14S,17R)-7-methoxy-10,13-dimethyl-17-[(E,2R)-6-methylhept-3-en-2-yl]-2,3,4,7,8,9,11,12,14,15,16,17-dodecahydro-1H-cyclopenta[a]phenanthren-3-ol	Flustramine M	-8	-8
Mollusca	D_	(22Z)-3 $\alpha$ ,24 $\zeta$ ,25-trihydroxycholesta-5,22-diene	6-Oxofascaplysin	-8	-7.8
Bacteria	D_	Shewanelline C	Fronodosin B	-7.9	-7.8
	Sponge				

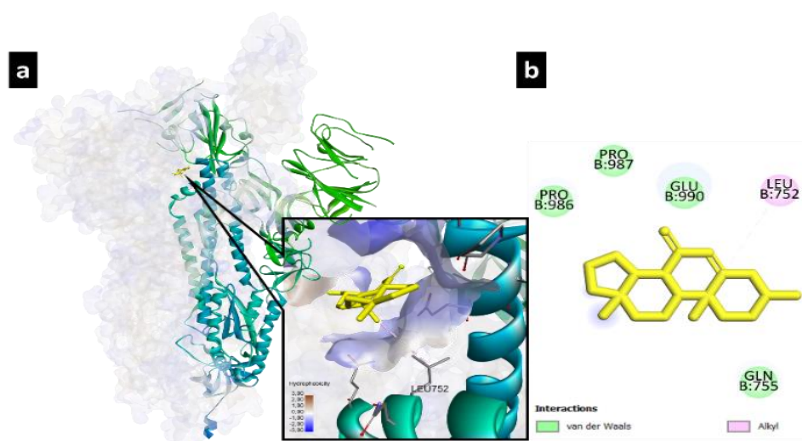
Even though some exceptions showed that the binding affinity score of the OM method is lower, in general, the results of each deep learning and machine learning have different tendencies for different sources of organisms or can also change according to groups of compounds (Figure 1). Therefore, we recommend using the BM method because it is more consistent than the OM method. However, in Mpro, the distribution of interacting ligands was divided into two groups, with lower binding affinity found in the interactions that occur in the core protein. While the inhibitor site in the previous study

was shown in the outer region, where the N3 inhibitor formed a bond [15,25]. Based on these results, we also visualized and analyzed the BM and OM ligands administered with the inhibitor site, namely Shewanelline C (Figure 6) and Flustramine M (Figure 7). Based on the results of this second group, it can be seen that although the binding affinity formed between the OM ligands is lower than the BM ligands, the OM ligand in the form of Flustramine M does not form hydrogen bonds so it can be predicted that the interactions formed will be less strong and less stable.

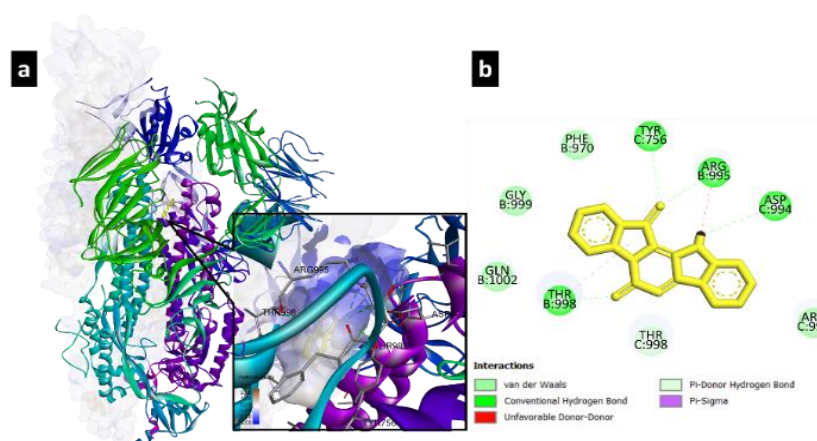
**FIGURE 1** Binding sites of top 5 ligands from the BM (Red) and OM (Green) methods against (a) spike and (b) Mpro

The visualization results also support the conclusion that the BM method is more accurate in predicting the activity of compounds as ligands when used as antivirals. This can be seen from the interaction that is formed, where the results of OM on the spike target form an unfavorable bond which can make the ligand-receptor interaction less stable (Figure 3b). Meanwhile, in Mpro protein, the ligand interactions of the BM method only form one type of hydrogen bond (Figure 4b) as the OM ligands form hydrophobic interactions (Figure 5b). Therefore the binding affinity of the OM ligand

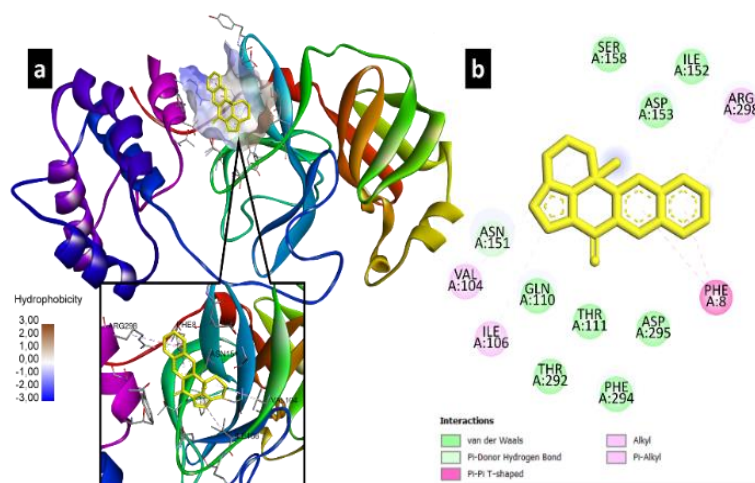
is lower than that of the BM ligand. However, the four interactions formed in the BM complex will increase the stability of the bond between Xetosaprol M and Mpro. According to the visualization results on the hydrophobicity diagram, all ligands form a type of hydrophobic interaction with amino acids that are both hydrophobic and hydrophilic. Surface hydrophobicity is shown based on the colour of the protein surface with brown indicating hydrophobic, white (intermediate), and hydrophilic as blue color. The more hydrophobic the resulting interaction, the more stable the conformation of the complex.



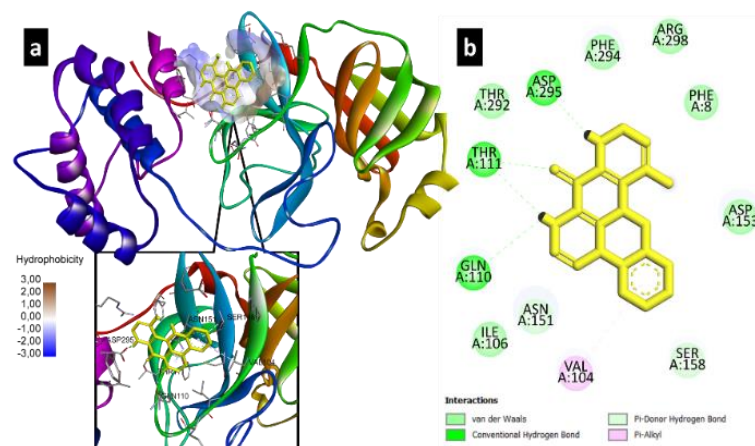
**FIGURE 2** Binding sites of top 5 ligands from the BM (Red) and OM (Green) methods against (a) spike, (b) Mpro



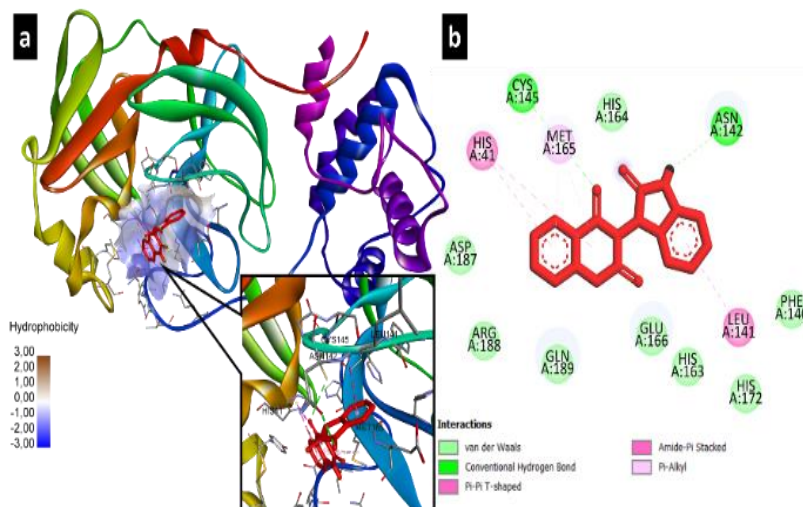
**FIGURE 3** The most potent OM ligand interactions are compounds 6-Oxofascaplysin to spike protein in terms of (a) hydrophobicity diagram and (b) 2D diagram of non-bond



**FIGURE 4** The most potent BM ligand interactions are compounds Xestosaprol M to Mpro protein in terms of (a) hydrophobicity diagram and (b) 2D diagram of non-bond

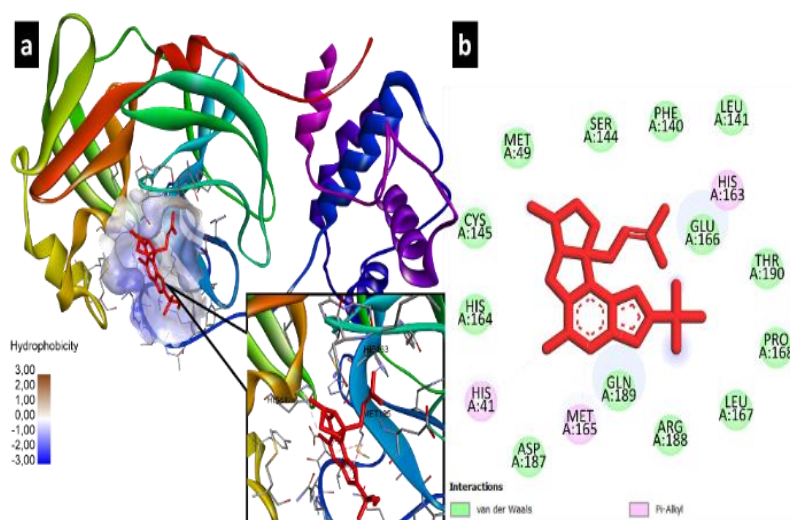


**FIGURE 5** The most potent OM ligand interactions are compounds Meridine to Mpro protein in terms of (a) hydrophobicity diagram and (b) 2D diagram of non-bond



**FIGURE 6** The most potent BM ligand interactions are compounds Shewanelline C to Mpro protein in terms of (a) hydrophobicity diagram and (b) 2D diagram of non-bond





**FIGURE 7** The most potent OM ligand interactions are compounds Flustramine M to Mpro protein in terms of (a) hydrophobicity diagram and (b) 2D diagram of non-bond

Based on the analysis of the amino acid residues that make up the binding site on each target protein, it showed that the binding site of the spike protein is quite large because it consists of 4 functional domains so that the distribution of the ligand positions varies (Figure 1a). However, in general, ligands that are predicted to have antiviral activity will bind between the B and C domains of the spike (Figures 2 and 3). Meanwhile, the Mpro protein showed a smaller binding site where, there were two groups, namely the inhibitor site which was the same as the N3 inhibitor

[15,25] and the alternative active site that we found, based on the molecular docking results, with a lower binding affinity value than the N3 inhibitor site. The inhibitor site residues that were found to interact with each of the most potent compounds were Met165, Phe140, His164, Glu166, Asp187, Arg188, and Gln189. In addition, amino acid residues Asn151, Val104, Asp153, Thr292, and Phe294 were also found to form interactions in the two most potent ligands which were predicted to be alternative active sites of Mpro (Table 5).

**TABLE 5** Amino acids that form binding sites in the most potent complexes for each target and screening method

Protein Target	Compound	Hydrogen Bond Position	Hydrophobic Interaction	van der Waals	Unfavorable Bond
BM_Spike	7-Ketocholesterol		B:LEU752	B:GLN755, B:PRO986, B:PRO987, B:GLU990	
OM_Spike	6-Oxofascaplysin	C:ASP994, B:ARG995, B:THR998, C:TYR756	B:THR998	B:PHE970, B:GLY999, B:GLN1002	B:ARG995
BM_Mpro	Xestosaprol M	<b>A:ASN151</b>	A:PHE8, A:ARG298, A:ILE106, <b>A:VAL104</b>	A:GLN110, A:THR111, A:ASN151, A:ILE152, <b>A:ASP153</b> , A:SER158,	

OM_Mpro	Meridine	A:GLN110, A:THR111, A:ASP295, <b>A:ASN151</b> , A:SER158	<b>A:VAL104</b>	<b>A:THR292</b> , <b>A:PHE294</b> , A:ASP295 A:PHE8, A:ILE106, <b>A:ASP153</b> , <b>A:THR292</b> , <b>A:PHE294</b> , A:ARG298 <b>A:PHE140</b> , A:HIS163, <b>A:HIS164</b> , <b>A:GLU166</b> , A:HIS172, <b>A:ASP187</b> , <b>A:ARG188</b> , <b>A:GLN189</b> A:MET49, <b>A:PHE140</b> , A:LEU141, A:SER144, A:CYS145, A:HIS164, <b>A:GLU166</b> , A:LEU167, A:PRO168, <b>A:ASP187</b> , <b>A:ARG188</b> , <b>A:GLN189</b> , A:THR190
BM_Mpro	Shewanelline C	A:ASN142, A:CYS145	A:HIS41, A:LEU141, ASN142, <b>A:MET165</b>	
OM_Mpro	Flustramine M		<b>A:MET165</b> , A:HIS41, A:HIS163	

In addition, several studies have also tested the efficacy of 7-Ketocholesterol as a treatment for the corona virus [26-30]. These results once again show that the BM method produces better predictions. For 6-Oxofascaplysin, the study is limited, but some reported that this compound has a cytotoxic activity against NFF cells and antiproliferative against human immune cells [31]. Meridine has been widely studied for the treatment of several diseases, such as liver [32], antitumor [33], and also antifungal [34,35]. While the efficacy of Xestosaprol M, Shewanelline C, and Flustramine M has not been studied.

## Conclusion

The best model of machine learning is the LR model. The compilation results of screening predictions from both machine learning and deep learning show more consistent results and are proven to show more stable bond

interactions than compounds that are predicted to have activity in just one of the screening methods.

## Acknowledgments

The authors would like to thank Inbio for helping this project and the LPPM of Hang Tuah University for administrating this project.

## Funding

This research was funded by KemendikbudRistek.

## Authors' Contribution

Herin, Nasywa: data collection; Eka, Hilya, and Didik: Article formulation; and Rezki: Analyzed the data.

## Conflict of Interest

The authors declare that they have no conflict of interest in this study.

### Orcid:

Herin Setianingsih\*:

<https://orcid.org/0000-0002-6247-483X>

Nasywa Zahra Sajida Tsurayya:

<https://orcid.org/0009-0004-7184-5788>

Hilyatuz Zahroh:

<https://orcid.org/0000-0002-7781-768X>

Eka Diyah Putri Lestari:

<https://orcid.org/0000-0002-1028-3871>

Didik Huswo Utomo:

<https://orcid.org/0000-0002-1555-9507>

Muhammad Rezki Rayak:

<https://orcid.org/0000-0002-4016-4072>

## References

- [1] M.A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, COVID-19 infection: Origin; transmission; and characteristics of human coronaviruses, *Journal of Advanced Research*, **2020**, *24*, 91–98. [Crossref], [Google Scholar], [Publisher]
- [2] M. Mohammadi, M. Shayestehpour, H. Mirzaei, The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines, *Brazilian Journal of Infectious Disease*, **2021**, *25*, 101606 [Crossref], [Google Scholar], [Publisher]
- [3] B. Roy B, J.K. Dhillon, N. Habib, B. Pugazhandhi B, Global variants of COVID-19: Current understanding, *Journal of Biomedical Sciences*, **2021**, *8*, 8-11. [Crossref], [Publisher]
- [4] T.F. Molinski, D.S. Dalisay, S.L. Lievens, J. P. Saludes, Drug development from marine natural products, *Nature Reviews Drug Discovery*, **2009**, *8*, 69-85. [Crossref], [Google Scholar], [Publisher]
- [5] A.W. Indrayani, I.M. Jawi, I.G.A. Artini, N.W. Sucindra, S. Martodihardjo, S. Radiono, J. Jumina, I.G.M.N. Budiana, D.A. Arimurni, M.D.P. Wahyudi, L. Chabib, M. Mustofa, Acute toxicity

profile and Sun Protection Factor (SPF) nanoemulgel combination of C-phenylcalix[4]resorcinaryl octacinnamate, C-methylcalix[4]resorcinaryl octabenzoate, and quercetin in vitro and in vivo, *Bali Medical Journal*, **2020**, *9*, 246–252. [Crossref], [Google Scholar], [Publisher]

[6] E. Sijbesma, K.K. Hallenbeck, S. Leysen, P.J. de Vink, L. Sko'ra, W. Jahnke, L. Brunsveld, M. R. Arkin, C. Ottmann; Site-directed fragment-based screening for the discovery of protein-protein interaction stabilizers, *Journal of the American Chemical Society*, **2019**, *141*, 3524–3531. [Crossref], [Google Scholar], [Publisher]

[7] P.I. Nandana, H. Rasyid, Prihantono, I. Yustisia, L. Hakim, Molecular docking studies of Brucein D as a potential inhibitor of the Bcl-2 anti-apoptotic protein, *Bali Medical Journal*, **2023**, *12*, 2148–2152. [Crossref], [Google Scholar], [Publisher]

[8] J. Bosch; PPI inhibitor and stabilizer development in human diseases, *Drug Discovery Today: Technology*, **2017**, *24*, 3–9. [Crossref], [Google Scholar], [Publisher]

[9] C. Lyu, T. Chen, B. Qiang, N. Liu, H. Wang, L. Zhang, Z. Liu, CMNPD: a comprehensive marine natural products database towards facilitating drug discovery from the ocean, *Nucleic Acids Research*, **2021**, *49*, D509-15. [Crossref], [Google Scholar], [Publisher]

[10] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B.A. Shoemaker, P.A. Thiessen, B. Yu, L. Zaslavsky, PubChem 2023 update, *Nucleic Acids Research*, **2023**, *51*, D1373-80. [Crossref], [Google Scholar], [Publisher]

[11] D.A. Filimonov, A.A. Lagunin, T.A. Glorizova, A.V. Rudik, D.S. Druzhilovskii, P.V. Pogodin, V.V. Poroikov, Prediction of the biological activity spectra of organic compounds using the PASS online web resource, *Chemistry of Heterocyclic Compounds*, **2014**, *50*, 444-57. [Crossref], [Google Scholar], [Publisher]

[12] D.E. Gordon, G.M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K.M. White, M.J. O'Meara, Veronica V. Rezelj, J.Z. Guo, D.L. Swaney, T.A.

- Tummino, R. Hüttenhain, R.M. Kaake, A.L. Richards, B. Tutuncuoglu, H.e Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B.J. Polacco, H. Braberg, Jacqueline M. Fabius, ...N.J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing, *Nature*, **2020**, *583*, 459–68. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [13] H. Setianingsih, E. Yatmasari, M.F. Ilmawan, H. Zahroh, E.D.P. Lestari, Targeting multiple SARS-CoV-2 and human proteins: in silico approach for COVID-19 drug repurposing, *International Journal of Pharmaceutical Research*, **2020**, *12*, 3184-3192. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [14] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The protein data bank, *Nucleic acids research*, **2000**, *1*, 28(1), 235-42. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [15] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, et.al., Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature*, **2020**, *582*, 289-293. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [16] L. Enjuanes, I. Solá, S. Zúñiga, PharmaMar reports positive results for Aplidin ® against coronavirus HCoV-229E, *PharmaMar*, **2020**, 1-2. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [17] S. Drożdżał, J. Rosik, K. Lechowicz, F. Machaj, K. Kotfis, S. Ghavami, M.J. Tos., FDA approved drugs with pharmacotherapeutic potential for SARS- CoV-2 (COVID-19) therapy, *Drug Resistance Updates*, **2020**, *53*, 100719 [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [18] C.N. Cavasotto, J.I. Di Filippo, In silico Drug Repurposing for COVID-19: Target SARS CoV-2 proteins through docking and consensus ranking, *Molecular Informatics*, **2020**, *40*, 2000115. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [19] M.M. Mysinger, M. Carchia, J.J. Irwin, B.K. Shoichet, Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking, *Journal of medicinal chemistry*, **2012**, *55*, 14, 6582-94. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [20] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, Scikit-learn: Machine learning in Python, *The Journal of Machine Learning Research*, **2011**, *12*, 2825-30. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [21] A. Candel, E. LeDell, Deep learning with H<sub>2</sub>O, 2022. [[Google Scholar](#)], [[Publisher](#)]
- [22] J. Eberhardt, D. Santos-Martins, A.F. Tillack, S. Forli, AutoDock Vina 1.2.0: New Docking Methods, expanded force field, and python bindings, *Journal of Chemical Information and Modeling*, **2021**, *61*, 3891-3898. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [23] 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*, **2010**, *31*, 455-461. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [24] D. Tworowski, A. Gorohovski, S. Mukherjee, G. Carmi, E. Levy, et.al., COVID19 Drug Repository: text-mining the literature in search of putative COVID19 therapeutics, *Nucleic Acids Research*, **2020**, *49*, 1113–1121. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [25] S. Keretsu, S.P. Bhujbal, S.J. Cho, Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation, *Scientific Reports*, **2020**, *10*, 17716. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [26] I. Ghzaïel, K. Sassi, A. Zarrouk, T. Nury, M. Ksila, et. al., 7-Ketocholesterol: Effects on viral infections and hypothetical contribution in COVID-19, *The Journal of Steroid Biochemistry and Molecular Biology*, **2021**, *212*, 105939. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [27] Y.M. Mupangati, Mustain, H. Setyawan, I. Riwanto, C.H. Soejono, M.H. Gasem, Factors related to mortality in geriatric patients after elective surgery: a retrospective cohort study, *Bali Medical Journal*, **2023**, *12*, 495–500. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [28] F. Fadhillah, N. Vanawati, Comparison of glucose reduction in urine using benedict

- method heated by methylated flame with 100°C waterbath, *Indonesian Journal of Medical Laboratory Science and Technology*, **2019**, *1*, 44–51. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [29] M.B. Rahmawati, Y. Andhitara, R. Ardhini, A. Kurnianto, A.C. Pertiwi, R.P. Sinaga, A. Larastiti, COVID-19 with ischemic stroke manifestations, *Bali Medical Journal*, **2021**, *10* 412–415. [[Crossref](#)], [[Publisher](#)]
- [30] A. Marcello, A. Civra, R.M. Bonotto, L.N. Alves, S. Rajasekharan, C. Giacobone, C. Caccia, R. Cavalli, M. Adami, P. Brambilla, D. Lembo, G. Poli, V. Leoni, The cholesterol metabolite 27-hydroxycholesterol inhibits SARS-CoV-2 and is markedly decreased in COVID-19 patients, *Redox Biology*, **2020**, *36*, 101682. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [31] S. Khokhar, Y. Feng, M.R. Campitelli, M.G. Ekins, J.N.A. Hooper, K.D. Beattie, M.C. Sadowski, C.C. Nelson, R.A. Davis, Isolation, structure determination and cytotoxicity studies of tryptophan alkaloids from an Australian marine sponge *Hyrtios* sp., *Bioorganic & Medicinal Chemical Letters*, **2014**, *24*, 3329–3332. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [32] R.Z. Zhao, D. Yuan, S.J. Liu, Y.J. Chen, L.J. Liu, Y. Zhao, Liver targeting effect of vinegar-baked *Radix-Bupleuri* on rhein in rats, *Journal of Ethnopharmacology*, **2010**, *132*, 421–428. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [33] E. Delfourne, R. Kiss, L.L. Corre, F. Dujols, J. Bastide, F. Collignon, B. Lesur, A. Frydman, F. Darro, Synthesis and in vitro antitumor activity of ring C and D-substituted phenanthrolin-7-one derivatives, analogues of the marine pyridoacridine alkaloids ascididemin and meridine, *Bioorganic & Medicinal Chemistry*, **2004**, *12*, 3987–3894. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [34] P.J. McCarthy, T.P. Pitts, G.P. Gunawardana, M. Kelly-Borges, S.A. Pomponi, Antifungal activity of meridine, a natural product from the marine sponge *Corticium* sp, *Journal of Natural Products*, **1992**, *55*, 1664–1668. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]
- [35] D.D. Wulansari, A. Kirtishanti, D.D. Wulandari, A.R.P. Santoso, E.A. Christy, L. Hiskiati, D. Winarni, Preliminary study of immunomodulatory effects of fermented garlic honey on non-specific immune responses, *Bali Medical Journal*, **2023**, *12*, 3370–3375. [[Crossref](#)], [[Google Scholar](#)], [[Publisher](#)]

**How to cite this article:** Herin Setianingsih, Nasywa Zahra Sajida Tsuroyya, Hilyatuz Zahroh, Eka Diyah Putri Lestari, Didik Huswo Utomo, Muhammad Rezki Rayak, Applying machine learning to define best models for screening potential antiviral compound for sars-cov-2 protein target. *Journal of Medicinal and Pharmaceutical Chemistry Research*, 2024, 6(8), 1099-1111. **Link:** [https://jmpcr.samipubco.com/article\\_192865.html](https://jmpcr.samipubco.com/article_192865.html)