Supplementary Information

IDENTIFCATION OF LEAD INHIBITORS FOR 3CLPRO OF SARS-COV-2 TARGET USING MACHINE LEARNING BASED VIRTUAL SCREENING, ADMET ANALYSIS, MOLECULAR DOCKING AND MOLECULAR DYNAMICS SIMULATIONS

Sandeep Poudel Chhetri¹, Vishal Singh Bhandari², Rajesh Maharjan¹, Tika Ram Lamichhane^{1,*}

¹ Central Department of Physics, Tribhuvan University, Kathmandu 44600, Nepal

² Central Department of Chemistry, Tribhuvan University, Kathmandu 44600, Nepal

* Correspondence: Email: tika.lamichhane@cdp.tu.edu.np

Figure-S1: Molecular Structure of Azvudine, Ensitrelvir, Simnotrelvir, and Nirmatrelvir.

Table-S1: General performance of 20 different classifiers.

Figure-S2: Confusion matrix for (a) NuSVC (b) ExtraTreesClassifier (c) LGBM Classifier (d) Voting Classifier.

Table-S2: Physicochemical properties, pharmacokinetics, druglikeness, and medicinal chemistry of the molecules.

Figure-S3: Bioavailability radar diagram for Azvudine and Nirmatrelvir as references and candidates M1, M2, and M3.

Table-S3: General performance of 20 different regressors.

Figure-S4: Boiled Egg graph of known inhibitors and potential inhibitors.

Table-S4: Molecular docking results of three candidate compounds with 3CLpro.

Figure-S5: Experimental versus predicted pIC50 for training and testing set.

Table-S5: Active residues of 3CLpro and their interaction with best docking pose of candidate compounds.

Figure-S6: Redocking of Z219104216 in 3CLpro.

Table-S6: Average binding free energies of 3CLpro-M1 complex using the MM-PBSA method (all values are in kcal/mol).

Figure-S7: RMSF plot for ap and M1 binding forms of 3CLpro during 200ns MD simulation.

Table-S7: The binding free energies contributed by active amino acid residues of 3CLpro and M1 (LIG).

Figure-S8: Secondary structure evaluation of apo and M1 binding forms of 3CLpro during 200ns MD simulation.

Figure-S9: Heatmap of the binding free energy contribution by active residues and ligand during last 20ns MD simulations.



Figure-S1: Molecular Structure of Azvudine, Ensitrelvir, Simnotrelvir, and Nirmatrelvir.

Accuracy	ROC AUC
0.89	0.90
0.89	0.89
0.88	0.88
0.88	0.88
0.87	0.88
0.87	0.87
0.86	0.86
	Accuracy 0.89 0.89 0.88 0.88 0.87 0.87 0.87 0.86

Table-S1: General performance of 20 different classifiers.

Perceptron	0.86	0.86
DecisionTreeClassifier	0.86	0.86
ExtraTreeClassifier	0.85	0.85
CalibratedClassifierCV	0.84	0.84
Logisticregression	0.84	0.84
PassiveAggresiveClassifier	0.84	0.84
AdaBoostClassifier	0.83	0.83
SGDClassifier	0.83	0.83
LinearDiscriminantanalysis	0.82	0.82
RidgeClassifier	0.82	0.82
LinearSVC	0.82	0.82
BernoulliNB	0.81	0.81
KNeighboursClassifier	0.82	0.81



Figure-S2: Confusion matrix for (a) NuSVC (b) ExtraTreesClassifier (c) LGBM Classifier (d) Voting Classifier.

Physiochemical		Pharmacokinetics		Druglikeness	
Properties				C C	
Molecular	286.22	GI absorption	Low	Lipinski	Yes;0
weight (g/mol)					violations
Num. heavy	20	BBB period	No	Ghose	No;1
atoms					
Num. arom.	6	P-gp substrate	No	Veber	No;1
heavy atoms					
Fraction Csp3	0.56	CYP1A2 inhibitor	No	Egan	No;1
Num. rotatable	3	CYP2C9 inhibitor	No	Muegge	No;1
bonds					
Num. H-bond	9	CYP2C9 inhibitor	No	Bioavailability	0.55
acceptors				score	
Num. H-bond	3	CYP2D6 inhibitors	No		
donors					
Molar	59.60	CYP3A4 inhibitor	No		
Refractivity					
$TPSA(\dot{A}^2)$	160.35	Log K _p (cm/s)	-8.62		
Molecular	499.53	GI absorption	High	Lipinski	Yes
weight (g/mol)					
Num. heavy	35	BBB period	No	Ghose	No;1
atoms					
Num. arom.	0	P-gp substrate	Yes	Veber	No;1
heavy atoms					
Fraction Csp3	0.78	CYP1A2 inhibitor	No	Egan	Yes
Num. rotatable	11	CYP2C9 inhibitor	No	Muegge	Yes
bonds					
Num. H-bond	8	CYP2C9 inhibitor	No	Bioavailability	0.55
acceptors				score	
Num. H-bond	3	CYP2D6 inhibitors	No		
donors					
Molar	125.68	CYP3A4 inhibitor	Yes		
Refractivity					
TPSA(Å ²)	131.40	Log K _p (cm/s)	-7.81		
Molecular	359.38	GI absorption	High	Lipinski	Yes
weight (g/mol)					
Num. heavy	26	BBB period	No	Ghose	No;1
atoms					
Num. arom.	12	P-gp substrate	Yes	Veber	Yes
heavy atoms					
Fraction Csp3	0.44	CYP1A2 inhibitor	No	Egan	Yes
Num rotatable	4	CYP2C9 inhibitor	No	Muegge	Yes
Num. Iotatable	•	0 1 1 2 0 / minomor			
	Physiochemical Properties Molecular weight (g/mol) Num. heavy atoms Num. arom. heavy atoms Fraction Csp3 Num. rotatable bonds Num. H-bond acceptors Num. H-bond donors Molar Refractivity TPSA(Å ²) Molecular weight (g/mol) Num. heavy atoms Num. arom. heavy atoms Fraction Csp3 Num. H-bond acceptors Num. H-bond acceptors Num. rotatable bonds Num. rotatable bonds Num. rotatable acceptors Num. H-bond acceptors Num. Feature H B H H H H H H H H H H H H H H H H H	Physiochemical PropertiesImage: Section of the sect	Physiochemical PropertiesPharmacokineticsMolecular286.22GI absorptionweight (g/mol)BBB periodNum. heavy20BBB periodatoms6P-gp substrateheavy atoms6P-gp substratefraction Csp30.56CYP1A2 inhibitorNum. arom.6P-gp substrateheavy atoms3CYP2C9 inhibitorbonds3CYP2C9 inhibitornum. rotatable3CYP2C9 inhibitoracceptors77Molar59.60CYP3A4 inhibitordonors6Jup SectorMolar59.60CYP3A4 inhibitorRefractivity160.35Log Kp (cm/s)Molecular499.53GI absorptionweight (g/mol)7160.35Num. heavy35BBB periodatoms711Num. arom.0P-gp substrateheavy atoms711Fraction Csp30.78CYP2C9 inhibitorNum. rotatable11CYP2C9 inhibitorbonds7125.68CYP3A4 inhibitorNum. H-bond3CYP2D6 inhibitorsdonors125.68CYP3A4 inhibitorRefractivity7131.40Log Kp (cm/s)Molecular359.38GI absorptionweight (g/mol)712Num. heavy26BBB periodatoms712Num. arom.12P-gp substrateheavy atoms7 <td>Physiochemical PropertiesPharmacokineticsPharmacokineticsProperties286.22GI absorptionLowweight (g/mol)BBB periodNonum. heavy20BBB periodNoatoms0P-gp substrateNoheavy atomsCYP1A2 inhibitorNobeavy atomsCYP2C9 inhibitorNobonds0.56CYP1A2 inhibitorNoNum. rotatable3CYP2C9 inhibitorNobonds0CYP2C9 inhibitorNoacceptors0CYP2D6 inhibitorsNoMolar59.60CYP3A4 inhibitorNoRefractivity160.35Log Kp (cm/s)-8.62Molecular499.53GI absorptionHighweight (g/mol)0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNoatoms0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNonum. neavy35BBB periodNoatoms0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNoNum. rotatable11CYP2C9 inhibitorNoNum. H-bond3CYP2D6 inhibitorsNoNum. H-bond3CYP2C9 inhibitorNoMolar125.68CYP2C9 inhibitorNoMolar125.68CYP2A1A1 inhibitorYesRefractivity1Log Kp (cm/s)-7.81Molacular359.38GI abs</td> <td>Physiochemical PropertiesPharmacokineticsDruglikenessMolecular weight (g/mol)286.22GI absorptionLowLipinskiNum. heavy atoms20BBB periodNoGhoseNum. arom. heavy atoms6P-gp substrateNoVeberheavy atoms7CYP1A2 inhibitorNoEganFraction Csp30.56CYP1A2 inhibitorNoBioavailabilityNum. rotatable3CYP2C9 inhibitorNoBioavailabilityscoreptors78CYP2D6 inhibitorsNoBioavailabilityMolar59.60CYP3A4 inhibitorNoKoseScoreMolecular499.53GI absorptionHighLipinskiweight (g/mol)160.35Log K_p (cm/s)-8.62-Num. heavy35BBB periodNoGhoseatoms0P-gp substrateYesVeberheavy atoms710CYP2C9 inhibitorNoWum. neavy35BBB periodNoMueggebonds0P-gp substrateYesVeberheavy atoms0P-gp substrateNoMueggeNum. H-bond8CYP2C9 inhibitorNoMueggebonds11CYP2C9 inhibitorNoMueggebonds11CYP2C9 inhibitorNoMueggeNum. H-bond3CYP2D6 inhibitorsNoMueggeNum. H-bond3CYP2C9 inhibitorNoMuegge</td>	Physiochemical PropertiesPharmacokineticsPharmacokineticsProperties286.22GI absorptionLowweight (g/mol)BBB periodNonum. heavy20BBB periodNoatoms0P-gp substrateNoheavy atomsCYP1A2 inhibitorNobeavy atomsCYP2C9 inhibitorNobonds0.56CYP1A2 inhibitorNoNum. rotatable3CYP2C9 inhibitorNobonds0CYP2C9 inhibitorNoacceptors0CYP2D6 inhibitorsNoMolar59.60CYP3A4 inhibitorNoRefractivity160.35Log Kp (cm/s)-8.62Molecular499.53GI absorptionHighweight (g/mol)0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNoatoms0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNonum. neavy35BBB periodNoatoms0P-gp substrateYesheavy atoms1CYP2C9 inhibitorNoNum. rotatable11CYP2C9 inhibitorNoNum. H-bond3CYP2D6 inhibitorsNoNum. H-bond3CYP2C9 inhibitorNoMolar125.68CYP2C9 inhibitorNoMolar125.68CYP2A1A1 inhibitorYesRefractivity1Log Kp (cm/s)-7.81Molacular359.38GI abs	Physiochemical PropertiesPharmacokineticsDruglikenessMolecular weight (g/mol)286.22GI absorptionLowLipinskiNum. heavy atoms20BBB periodNoGhoseNum. arom. heavy atoms6P-gp substrateNoVeberheavy atoms7CYP1A2 inhibitorNoEganFraction Csp30.56CYP1A2 inhibitorNoBioavailabilityNum. rotatable3CYP2C9 inhibitorNoBioavailabilityscoreptors78CYP2D6 inhibitorsNoBioavailabilityMolar59.60CYP3A4 inhibitorNoKoseScoreMolecular499.53GI absorptionHighLipinskiweight (g/mol)160.35Log K_p (cm/s)-8.62-Num. heavy35BBB periodNoGhoseatoms0P-gp substrateYesVeberheavy atoms710CYP2C9 inhibitorNoWum. neavy35BBB periodNoMueggebonds0P-gp substrateYesVeberheavy atoms0P-gp substrateNoMueggeNum. H-bond8CYP2C9 inhibitorNoMueggebonds11CYP2C9 inhibitorNoMueggebonds11CYP2C9 inhibitorNoMueggeNum. H-bond3CYP2D6 inhibitorsNoMueggeNum. H-bond3CYP2C9 inhibitorNoMuegge

Table-S2: Physicochemical properties, pharmacokinetics, druglikeness, and medicinal chemistry of the molecules.

	Num. H-bond	5	CYP2C9 inhibitor	No	Bioavailability	0.55
	acceptors				score	
	Num. H-bond	2	CYP2D6 inhibitors	No		
	donors					
	Molar	100.84	CYP3A4 inhibitor	No		
	Refractivity					
	$TPSA(\dot{A}^2)$	118.29	$Log K_p (cm/s)$	-8.88		
	Molecular	345.35	GI absorption	High	Lipinski	Yes
	weight (g/mol)					
	Num. heavy	25	BBB period	No	Ghose	Yes
	atoms					
	Num. arom.	12	P-gp substrate	No	Veber	Yes
	heavy atoms					
	Fraction Csp3	0.35	CYP1A2 inhibitor	No	Egan	Yes
M2	Num. rotatable	5	CYP2C9 inhibitor	No	Muegge	Yes
	bonds					
	Num. H-bond	5	CYP2C9 inhibitor	No	Bioavailability	0.55
	acceptors				score	
	Num. H-bond	2	CYP2D6 inhibitors	No		
	donors					
	Molar	94.97	CYP3A4 inhibitor	No		
	Refractivity					
	TPSA(Å ²)	104.49	$Log K_p (cm/s)$	-8.04		
	Molecular	337.37	GI absorption	High	Lipinski	Yes
	weight (g/mol)					
	Num. heavy	24	BBB period	No	Ghose	No;1
	atoms					
	Num. arom.	8	P-gp substrate	No	Veber	Yes
	heavy atoms					
	Fraction Csp3	0.69	CYP1A2 inhibitor	No	Egan	Yes
M3	Num. rotatable	3	CYP2C9 inhibitor	No	Muegge	Yes
	bonds					
	Num. H-bond	5	CYP2C9 inhibitor	No	Bioavailability	0.55
	acceptors				score	
	Num. H-bond	3	CYP2D6 inhibitors	No		
	donors					
	Molar	90.82	CYP3A4 inhibitor	No		
	Refractivity					
	$TPSA(\dot{A}^2)$	115.49	Log K _p (cm/s)	-8.77		



Figure-S3: Bioavailability radar diagram for Azvudine and Nirmatrelvir as references and candidates M1, M2, and M3.

Regressors	R-Squared	RMSE
HistGradientBoostingRegressor	0.70	0.64
LGBMRegressor	0.70	0.64
Tweedieregressor	0.65	0.68
GammaRegressor	0.65	0.69
RandomForestRegressor	0.65	0.69
BayesianRidge	0.64	0.70
GradientBoostingRegressor	0.63	0.71
BaggingRegressor	0.62	0.72
PoissonRegressor	0.62	0.72
ElasticNetCV	0.61	0.72
LassoCV	0.61	0.73
XGBRegressor	0.60	0.74
LassoLarsCV	0.59	0.74
RidgeCV	0.58	0.75
passiveAggressiveRegressor	0.58	0.75
HuberRegressor	0.58	0.75
Ridge	0.56	0.77

Table-S3: General performance of 20 different regressors.

KNeighboursRegressor	0.54	0.79
NuSVR	0.54	0.79
SVR	0.54	0.79



Figure-S4: Boiled Egg graph of known inhibitors and potential inhibitors.

Table-	S4:	Mo	lecular	· docki	ng r	esults	of t	hree	candida	ate	comp	ounds	with	3CL	pro.
					ω										1

Compounds	Final intermolecular energy	Final total	Torsional	Unbound	Estimated
	[vdW+H-	internal	free	system's	free energy
	bond+desolv+electrostatic	energy	energy	energy	of binding
	energy] (kcal/mol)	(kcal/mol)	(kcal/mol)	(kcal/mol)	(kcal/mol)
M1	-9.53	-0.37	+0.89	-0.37	-8.64
M2	-9.41	-0.74	+1.19	-0.74	-8.22
M3	-8.90	-0.62	+0.89	-0.62	-8.00



Figure-S5: Experimental versus predicted pIC50 for training and testing set.

Compounds	Hydrophobic	Distance	Hydrogen	Distance	Angle (°)
_	Interactions	(Å)	Bonds	Donor-	
				Acceptor (Å)	
M1	MET165	3.76	PHE140	2.63	122.12
	GLU166	3.68	ASN142	2.71	145.26
			HIS163	2.08	133.42
			GLU166	3.61	142.77
			GLU166	2.59	110.28
			ARG188	3.75	152.30
			GLN189	4.07	123.79
M2	-	-	GLY143	2.80	165.60
			HIS163	3.02	173.20
			GLU166	3.91	138.06
			GLU166	2.85	129.94
			ARG188	3.00	130.99
M3	MET165	3.30	ASN142	2.70	147.65
	GLU166	3.72	GLY143	4.02	151.29
			HIS163	2.93	135.18
			GLU166	3.13	123.06
			GLN189	2.78	102.37

Table-S5: Active residues of 3CLpro and their interaction with best docking pose of candidate compounds.



Figure-S6: Redocking of Z219104216 in 3CLpro. The original and redocked pose of native ligand is denoted by blue and red color respectively.

Table-S6: Average binding free energies of 3CLpro-M1 complex using the MM-PBSA method (all values are in kcal/mol).

Van der Waals	Electrostatic	Polar solvation	Non-polar solvation	Total binding
energy	energy	energy	energy	energy
-37.85 ± 3.24	-25.86 ± 8.88	48.86 ± 8.09	-4.00 ± 0.17	-18.86 ± 4.38



Figure-S7: RMSF plot for apo and M1 binding forms of 3CLpro during 200ns MD simulation.

Table-S7: The binding free energies contributed by active amino acid residues of 3CLpro and M1 (LIG).

Residues	Average energy \pm SD (kcal/mol)	
HIS:41	-0.69 ± 0.92	
MET:49	$\textbf{-0.59}\pm0.35$	
LEU:50	$\textbf{-0.59}\pm0.30$	
PHE:140	-1.19 ± 0.52	
LEU:141	0.56 ± 0.51	
ASN:142	-0.56 ± 0.33	
GLY:143	$\textbf{-0.56} \pm 0.49$	
SER:144	-1.15 ± 0.87	
CYS:145	-1.7 ± 0.54	
GLU:166	0.69 ± 0.79	
LIG:307	-8.13 ± 2.28	



Figure-S8: Secondary structure evaluation of (**a**) apo and (**b**) M1 binding forms of 3CLpro during 200 ns MD simulation.



Figure-S9: Heatmap of the binding free energy contribution by active residues and ligand during last 20 ns MD simulations.