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 Chhetri[1], **Vishal Singh Bhandari**[2], **Rajesh Maharjan**[1], **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.

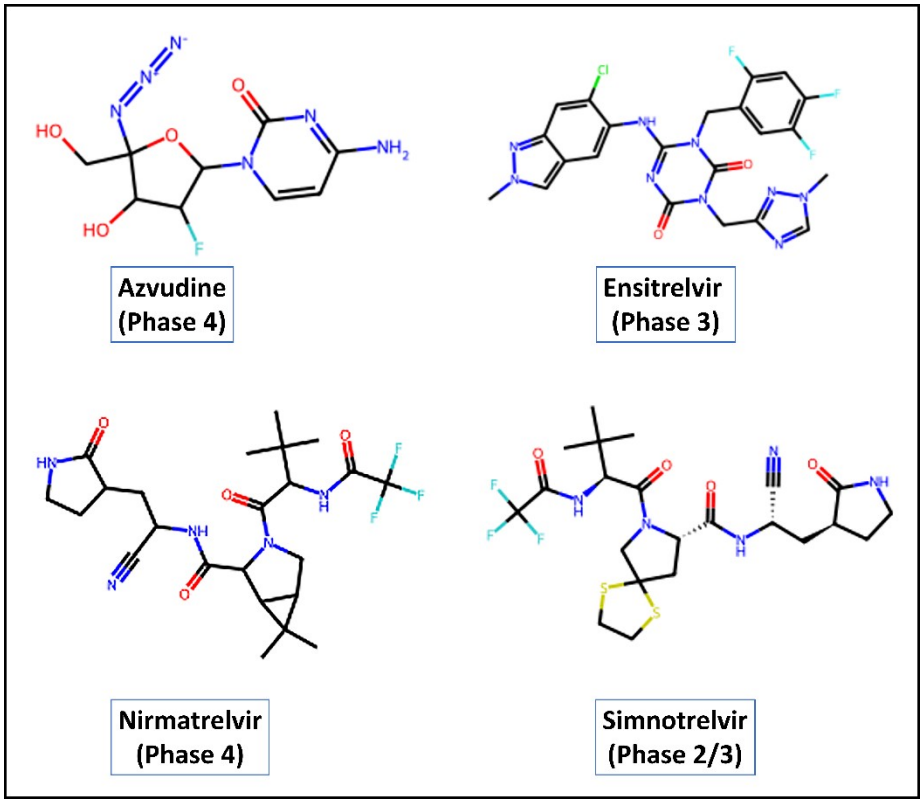

Figure-S1: Molecular Structure of Azvudine, Ensitrelvir, Simnotrelvir, and Nirmatrelvir.

Table-S1: General performance of 20 different classifiers.

Classifiers	Accuracy	ROC AUC
NuSVC	0.89	0.90
ExtraTreesClassifier	0.89	0.89
RandomForestClassifier	0.88	0.88
LGBMClassifier	0.88	0.88
SVC	0.87	0.88
XGBClassifier	0.87	0.87
BaggingClassifier	0.86	0.86

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

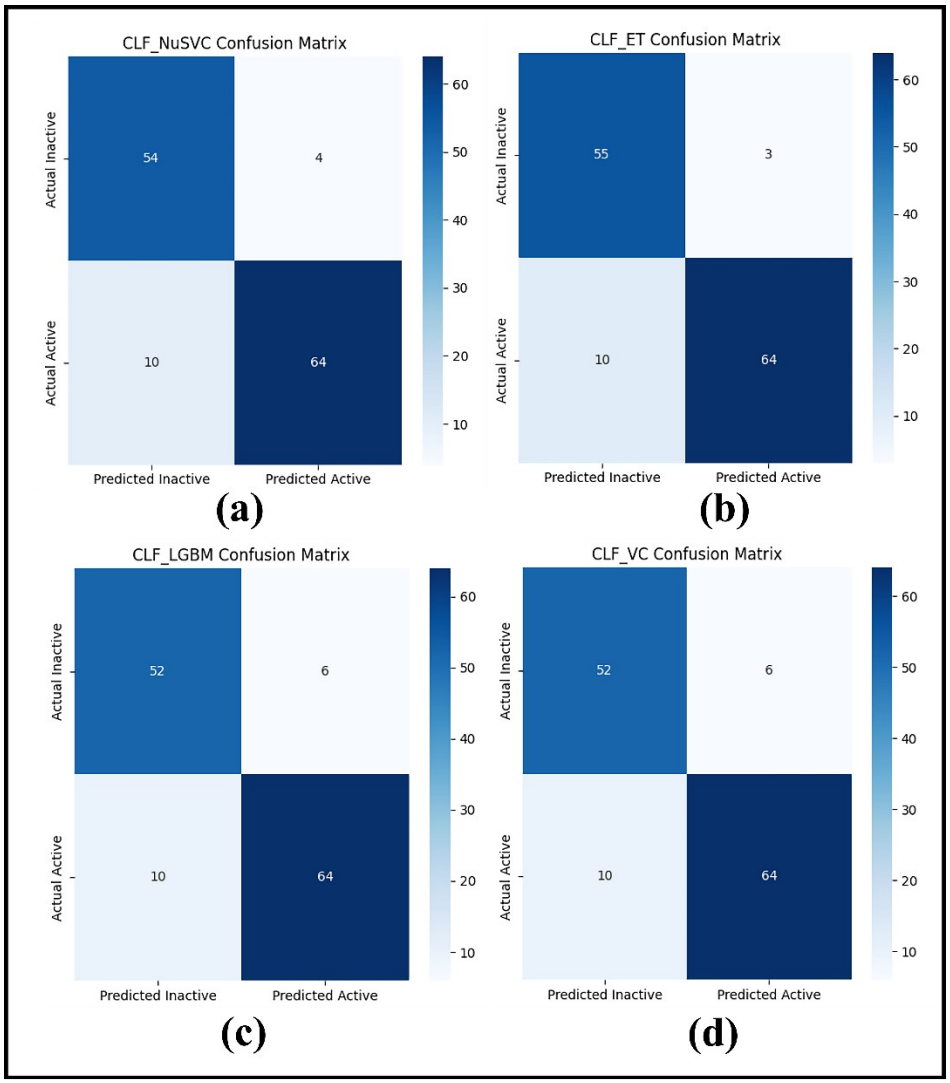

Figure-S2: Confusion matrix for (**a**) NuSVC (**b**) ExtraTreesClassifier (**c**) LGBM Classifier (**d**) Voting Classifier.

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

Compound	Physiochemical Properties		Pharmacokinetics		Druglikeness	
Azvudine	Molecular weight (g/mol)	286.22	GI absorption	Low	Lipinski	Yes;0 violations
	Num. heavy atoms	20	BBB period	No	Ghose	No;1
	Num. arom. heavy atoms	6	P-gp substrate	No	Veber	No;1
	Fraction Csp3	0.56	CYP1A2 inhibitor	No	Egan	No;1
	Num. rotatable bonds	3	CYP2C9 inhibitor	No	Muegge	No;1
	Num. H-bond acceptors	9	CYP2C9 inhibitor	No	Bioavailability score	0.55
	Num. H-bond donors	3	CYP2D6 inhibitors	No		
	Molar Refractivity	59.60	CYP3A4 inhibitor	No		
	TPSA(Å^2)	160.35	Log K_p (cm/s)	-8.62		
Nirmatrelvir	Molecular weight (g/mol)	499.53	GI absorption	High	Lipinski	Yes
	Num. heavy atoms	35	BBB period	No	Ghose	No;1
	Num. arom. heavy atoms	0	P-gp substrate	Yes	Veber	No;1
	Fraction Csp3	0.78	CYP1A2 inhibitor	No	Egan	Yes
	Num. rotatable bonds	11	CYP2C9 inhibitor	No	Muegge	Yes
	Num. H-bond acceptors	8	CYP2C9 inhibitor	No	Bioavailability score	0.55
	Num. H-bond donors	3	CYP2D6 inhibitors	No		
	Molar Refractivity	125.68	CYP3A4 inhibitor	Yes		
	TPSA(Å^2)	131.40	Log K_p (cm/s)	-7.81		
M1	Molecular weight (g/mol)	359.38	GI absorption	High	Lipinski	Yes
	Num. heavy atoms	26	BBB period	No	Ghose	No;1
	Num. arom. heavy atoms	12	P-gp substrate	Yes	Veber	Yes
	Fraction Csp3	0.44	CYP1A2 inhibitor	No	Egan	Yes
	Num. rotatable bonds	4	CYP2C9 inhibitor	No	Muegge	Yes

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

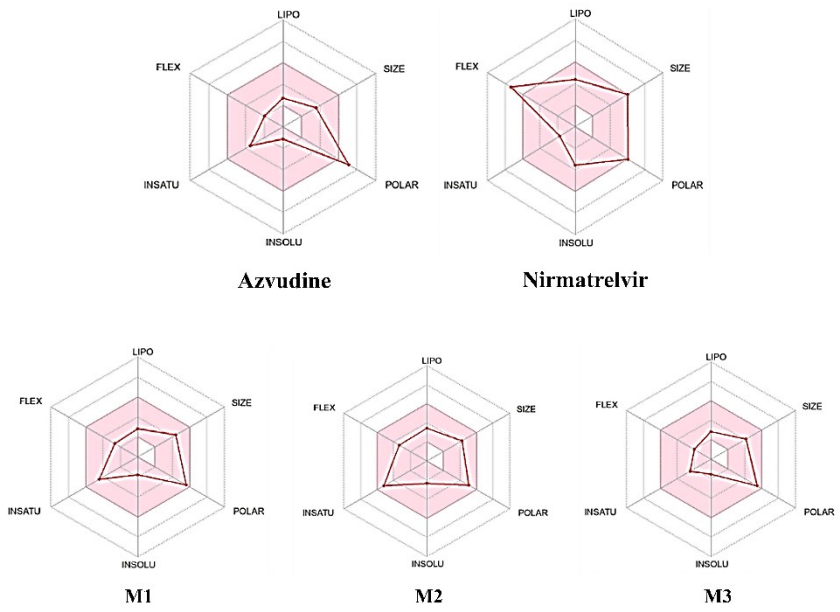

Figure-S3: Bioavailability radar diagram for Azvudine and Nirmatrelvir as references and candidates M1, M2, and M3.

Table-S3: General performance of 20 different regressors.

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

KNeighboursRegressor	0.54	0.79
NuSVR	0.54	0.79
SVR	0.54	0.79

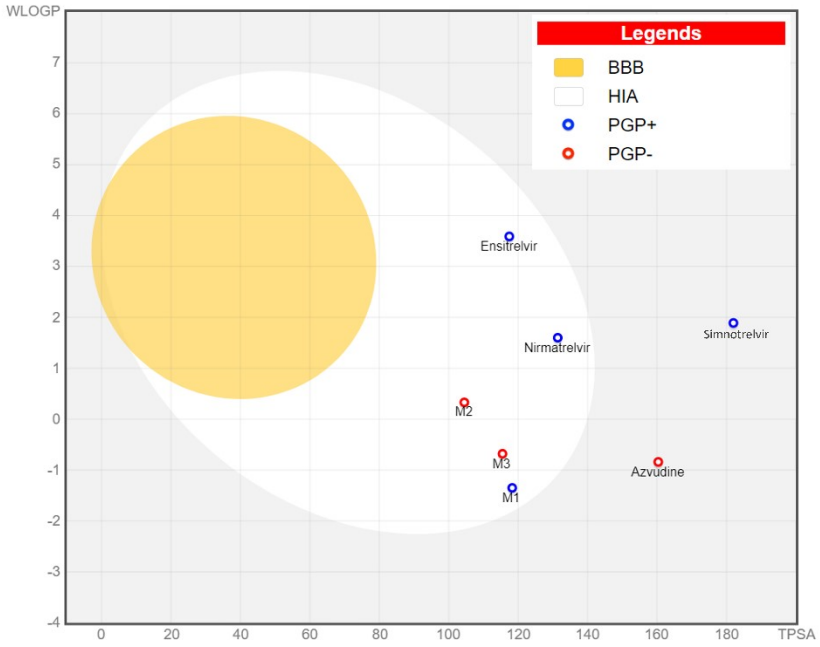

Figure-S4: Boiled Egg graph of known inhibitors and potential inhibitors.

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

Compounds	Final intermolecular energy [vdW+H-bond+desolv+electrostatic energy] (kcal/mol)	Final total internal energy (kcal/mol)	Torsional free energy (kcal/mol)	Unbound system's energy (kcal/mol)	Estimated free energy of binding (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

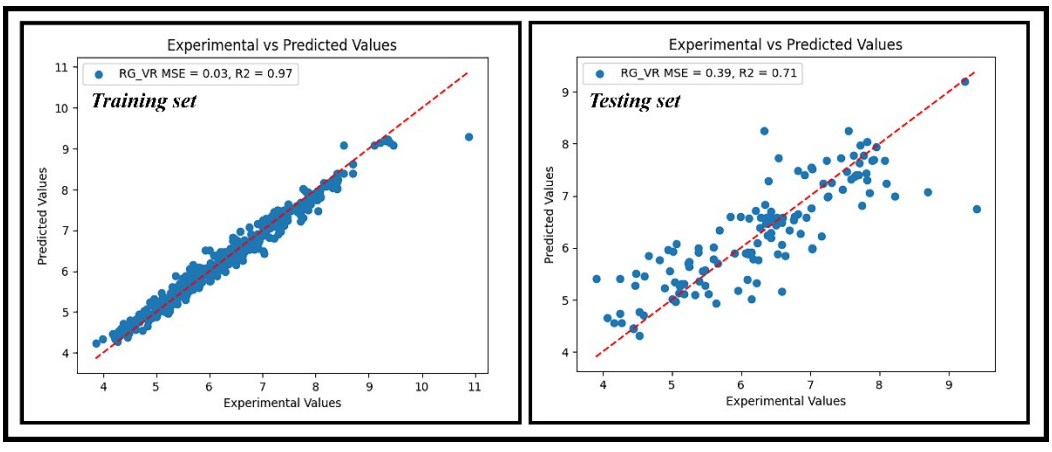

Figure-S5: Experimental versus predicted pIC50 for training and testing set.

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

Compounds	Hydrophobic Interactions	Distance (Å)	Hydrogen Bonds	Distance Donor-Acceptor (Å)	Angle (°)
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

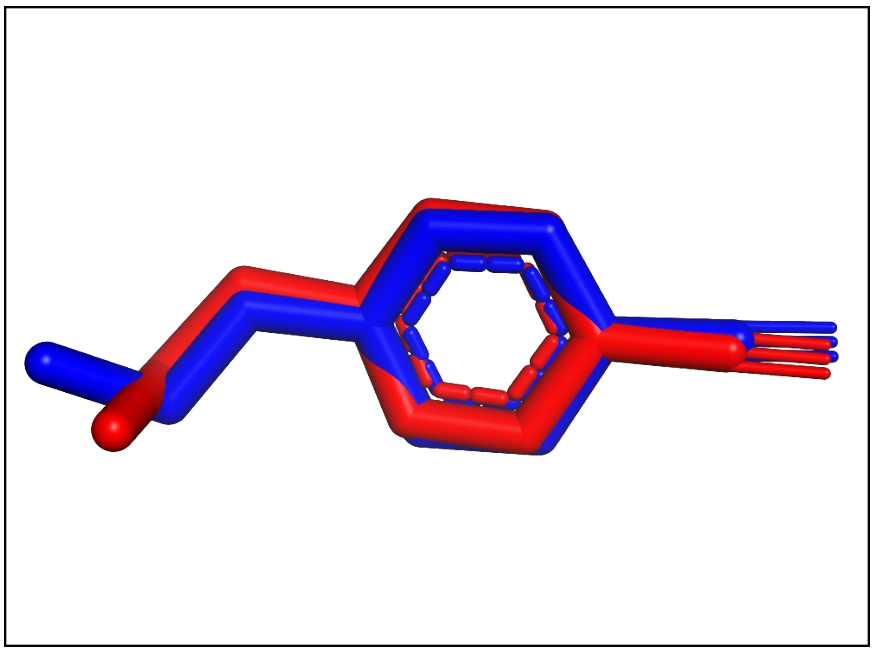

Figure-S6: Redocking of Z219104216 in 3CLpro. The original and redocked pose of native ligand is denoted by blue and red color respectively.

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

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

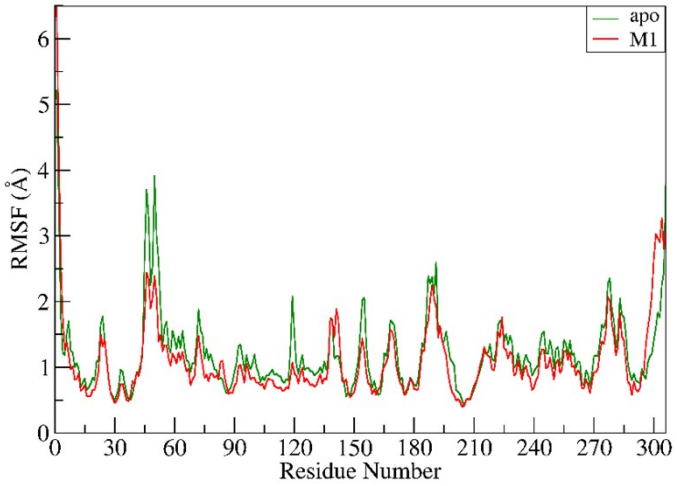

Figure-S7: RMSF plot for apo and M1 binding forms of 3CLpro during 200ns MD simulation.

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

Residues	Average energy ± SD (kcal/mol)
HIS:41	-0.69 ± 0.92
MET:49	-0.59 ± 0.35
LEU:50	-0.59 ± 0.30
PHE:140	-1.19 ± 0.52
LEU:141	0.56 ± 0.51
ASN:142	-0.56 ± 0.33
GLY:143	-0.56 ± 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

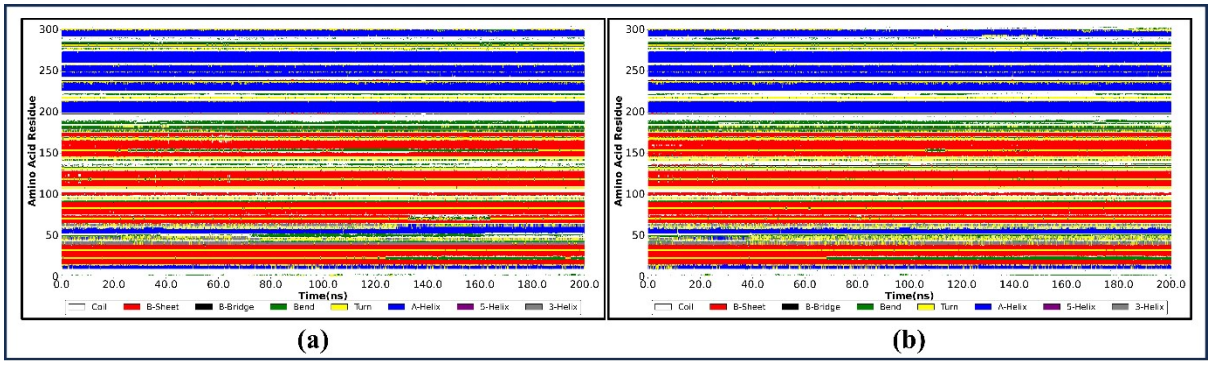

Figure-S8: Secondary structure evaluation of (**a**) apo and (**b**) M1 binding forms of 3CLpro during 200 ns MD simulation.

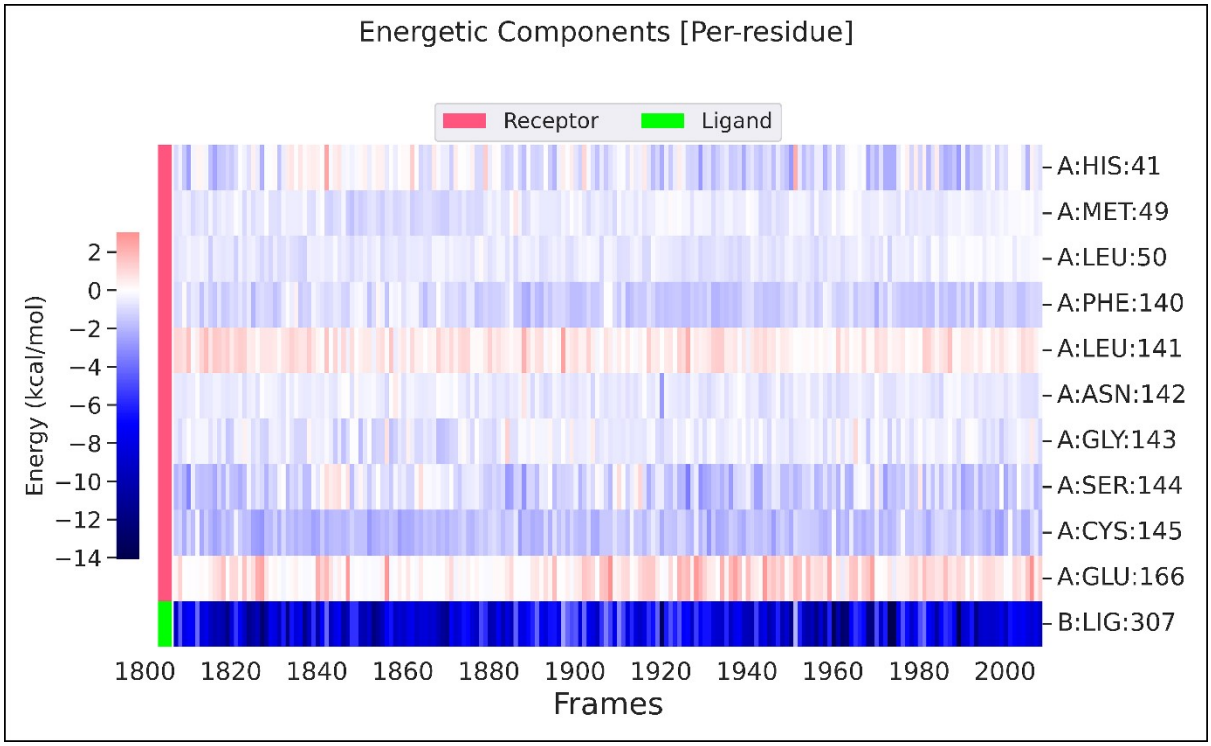

Figure-S9: Heatmap of the binding free energy contribution by active residues and ligand during last 20 ns MD simulations.