



OPEN

Targeting SARS-CoV-2 endoribonuclease: a structure-based virtual screening supported by in vitro analysis

Ibrahim M. Ibrahim¹, Abdo A. Elfiky¹, Mohamed M. Fathy¹, Sara H. Mahmoud² & Mahmoud ElHefnawi³✉

Researchers are focused on discovering compounds that can interfere with the COVID-19 life cycle. One of the important non-structural proteins is endoribonuclease since it is responsible for processing viral RNA to evade detection of the host defense system. This work investigates a hierarchical structure-based virtual screening approach targeting NSP15. Different filtering approaches to predict the interactions of the compounds have been included in this study. Using a deep learning technique, we screened 823,821 compounds from five different databases (ZINC15, NCI, Drug Bank, Maybridge, and NCI Diversity set III). Subsequently, two docking protocols (extra precision and induced fit) were used to assess the binding affinity of the compounds, followed by molecular dynamic simulation supported by the MM-GBSA free binding energy. Interestingly, one compound (ZINC000104379474) from the ZINC15 database has been found to have a good binding affinity of -7.68 kcal/Mol. The VERO-E6 cell line was used to investigate its therapeutic effect in vitro. Half-maximal cytotoxic concentration and Inhibitory concentration 50 were determined to be 0.9 mg/ml and 0.01 mg/ml, respectively; therefore, the selectivity index is 90. In conclusion, ZINC000104379474 was shown to be a good hit for targeting the virus that needs further investigations in vivo to be a drug candidate.

The coronavirus belongs to the Coronaviridae family and the Riboviria realm, according to the International Committee on Virus Taxonomy (ICTV)¹. It was discovered in humans for the first time in the 1960s, and there are currently seven strains of the virus that may cause illness in people. Out of them, four (OC43, HKU1, 229E, and NL63) produce moderate illness, while the other three (severe acute respiratory syndrome coronavirus (SARS-CoV), Middle-East respiratory syndrome coronavirus (MERS-CoV), and more recently, SARS-CoV-2) are more dangerous and can cause deadly diseases in humans². SARS-CoV-2 has a 79% and 50% sequence similarity to SARS-CoV and MERS-CoV, respectively^{3,4}. On the other hand, two bat CoV strains share a higher similarity with SARS-CoV-2 than that of the SARS-CoV. The bat coronavirus 'RaTG13' discovered in *Rhinolophus affinis* is the closest relative to SARS-CoV-2 known to date, with 96.2% similarity. Following it is the bat coronavirus 'RmYN02', which was recently discovered in a *Rhinolophus malayanus*, with 93.3% similarity to the SARS-CoV-2 genome^{5,6}. These two strains were discovered in Yunnan, a Chinese province. The new beta-coronavirus SARS-CoV-2 is a (+) sense RNA virus with a genome of nearly 30 kb. This genome encodes four structural genes: spike (S), envelope (E), membrane (M), and nucleocapsid (N) and nine accessory proteins (ORF3(a, b), ORF6, ORF7(a, b), ORF8, ORF9(b, c), and ORF10) which are responsible for the formation of the 16 Non-structural proteins (NSPs) numbered from 1 to 16^{7,8}. Endoribonuclease (EndoU or NSP15) is a 346 residue protein with three domains namely, the N-terminal domain, middle domain, and the C-terminal domain. The N-terminal domain was found to be responsible for the formation of a hexamer, while the C-terminal domain contains the active site of the protein that facilitates replication and processing of sub-genomic RNAs⁹⁻¹². It was found that NSP15 hydrolyzes the phosphodiester bond that is located at the uridine (U) sites of single and double-stranded RNA molecules. This cleavage produces 2', 3' cyclic phosphodiester and 5' hydroxyl terminal¹³. Without the role of NSP15 in the replication of the SARS-CoV-2 virus, the host innate immune system would utilize the pattern recognition receptor MDA5 to recognize the PolyU sequence of the negative-sense viral RNA that is replicated

¹Biophysics Department, Faculty of Science, Cairo University, Giza 12613, Egypt. ²Centre of Scientific Excellence for Influenza Viruses (CSEIV), National Research Centre, Cairo 12622, Egypt. ³Biomedical Informatics and Chemoinformatics Group (BICG), Informatics and Systems Department, National Research Centre, Cairo 12622, Egypt. ✉email: mahef@aucegypt.edu; mahef111@gmail.com

from the poly Adenine sequence of the viral RNA¹⁴. This identification would lead to the RNA-activated antiviral response by inhibiting the formation of cytoplasmic stress granules with antiviral functions. Moreover, it has been reported that the deletion of NSP15 decreased viral replication significantly^{15,16}. In addition, NSP15 was essential in interfering with interferon-beta (IFN- β) production and is associated with retinoblastoma tumor suppressor protein¹⁷. Based on its vital role in the replication of the virus, this protein is considered one of the potential targets of the SARS-CoV-2 virus.

Artificial neurons are used to process data in deep learning, a branch of machine learning. Text mining and picture pattern recognition are two sectors where deep learning has been used. The approach is also utilized to speed up the drug development processes in drug discovery, such as virtual screening, molecular docking, and QSAR models^{18–21}. Deep learning (DL) algorithms have been used to predict liver injuries from drugs, detect coronavirus infection from radiography images, find new antibacterial compounds, and predict drug-target interactions^{22–25}. DL in drug discovery requires the drugs to be vectorized before finding a mapping between these vectors and their properties. One of the advantages of deep learning in drug discovery is that it can be used to predict large-scale data quickly compared to the traditional ways²⁶. Wang et al. have developed a deep learning model to screen drugs against NSP5 of SARS-CoV-2 using a directed message passing neural network. They first trained the model on experimental data of several beta CoVs, then fine-tuned it with newly discovered active and inactive drugs against SARS-CoV-2. The final model was used to screen approximately 5 million druglike compounds from the ZINC15 dataset. Moreover, they tested the top seven compounds, and one of them had a unique chemical backbone with an IC₅₀ of 37.0 μ M²⁶.

In this study, we have targeted NSP15, the endoribonuclease, using five Different drug databases. A state-of-the-art Deep Learning library (Deep Purpose)²⁷ was used to screen and filter the compounds before molecular docking, which was done in two steps using the Extra Precision (XP) mode and Induced Fit (IF) mode of the Glide tool under Schrödinger. This was followed by a triplicate Molecular Dynamic (MD) Simulation of 100 ns each. Then, the calculation of binding energy using the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) approach was done to find the binding strength of each compound to the receptor after MD Simulation and to find the contribution of each amino acid to the binding. In vitro assessments followed this to find its 50% inhibitory concentration (IC₅₀), 50% cytotoxic concentration (CC₅₀), and selectivity index (SI).

Methods

In silico methods. A workflow of analysis comprising six steps for structure-based drug design was adopted (Fig. 1). We utilized the Deep Purpose library to filter the compounds with predicted IC₅₀ \leq 100 nM. In the second step, we utilized the Glide tool in Schrödinger to filter the compounds using XP and IF protocols which scored below (better) than the threshold of the redocked cocrystallized compound (citrate). In the third and fourth steps, we utilized the MD simulation to assert the binding of the remaining compounds. Then the MM-GBSA was calculated to find the binding energy and its components. The fifth and sixth steps included the clustering of the trajectory and the in vitro analyses to obtain IC₅₀, CC₅₀, and selectivity index.

Structural preparation. Protein data bank (PDB) (<https://www.rcsb.org/>) was used to obtain the target protein (PDB ID: 6W01, Chain A) (endoribonuclease) as a FASTA sequence and 3D structure. First, unwanted atoms (water, co-crystallized ligand (citrate), polyethylene glycol, and ethylene glycol molecules) were removed from the structure before preparation using the protein preparation wizard in Maestro V 11.8 (Schrödinger). Next, missing side chains and loops were filled using PRIME, while missing Hydrogen atoms were added using the protein preparation wizard. In addition, PROPKA was used to generate protonation states at pH 7.0. Finally, a minimization step was initialized with 0.3 Å as a convergent metric for heavy atoms²⁸. Figure 2 shows the 3D structure of the NSP15 after preparation and its three domains (N-terminal domain, Middle domain, and C-terminal domain). The total number of atoms was 5482 atoms.

Compounds in NCI Diversity set III (<https://wiki.nci.nih.gov/display/ncidtpdata/compound+sets>), NCI (<https://cactus.nci.nih.gov/download/nci/>), Drug Bank (<https://go.drugbank.com/>), Maybridge (<http://www.maybridge.com/>), and ZINC15 (<https://zinc.docking.org/>) (only drugs that are in stock and have 3D conformations) databases were downloaded for virtual screening (VS) using a python DL library called Deep Purpose²⁷ and for docking using Maestro GLIDE²⁸. For VS, Deep Purpose required the drugs to be in SMILES format; therefore, the drugs were converted into SMILES using open babel²⁹ in the cases when the database did not provide a downloadable SMILES format. Deep Purpose has an encoder-decoder architecture with a total of 15 encoders for both the drugs and the proteins. In our case, we used Convolutional Neural Network (CNN), Morgan fingerprints, DayLight fingerprints, and Message passing neural network (MPNN) encoders to encode the drugs, while for the protein, we used a CNN and Amino Acid Composition (AAC) encoding. These encoders were combined (CNN_CNN, Daylight_AAC, Morgan_AAC, Morgan_CNN, MPNN_CNN; the first encoder is for the compound SMILES and the second is for the protein sequence) to generate five predictions. These predictions were further aggregated using mean_max aggregation ($\frac{\text{value}_{\text{mean}} + \text{value}_{\text{max}}}{2}$) to produce one value for each compound. The distribution of the SMILES lengths used in this study is shown in Supplementary Fig. S1 of Supplementary Material 1, which resembles the distribution of the SMILES used to train the models in the DeepPurpose library²⁷. For the docking study, the 3D conformations of the drugs were downloaded. In the case of missing 3D structures, the 2D structure of the drugs was downloaded and then prepared using LigPrep in Maestro. Different ionization states were generated using Epik at a pH range of 7.0+/-2.0, and tautomers for each ionized or neutral compound were generated. For each drug, a maximum of 32 stereoisomers were generated.

Docking. The co-crystallized compound was redocked using GLIDE docking with the XP mode to get a docking score that will be used as a threshold for filtering drugs. Then, GLIDE docking with XP precision was used to

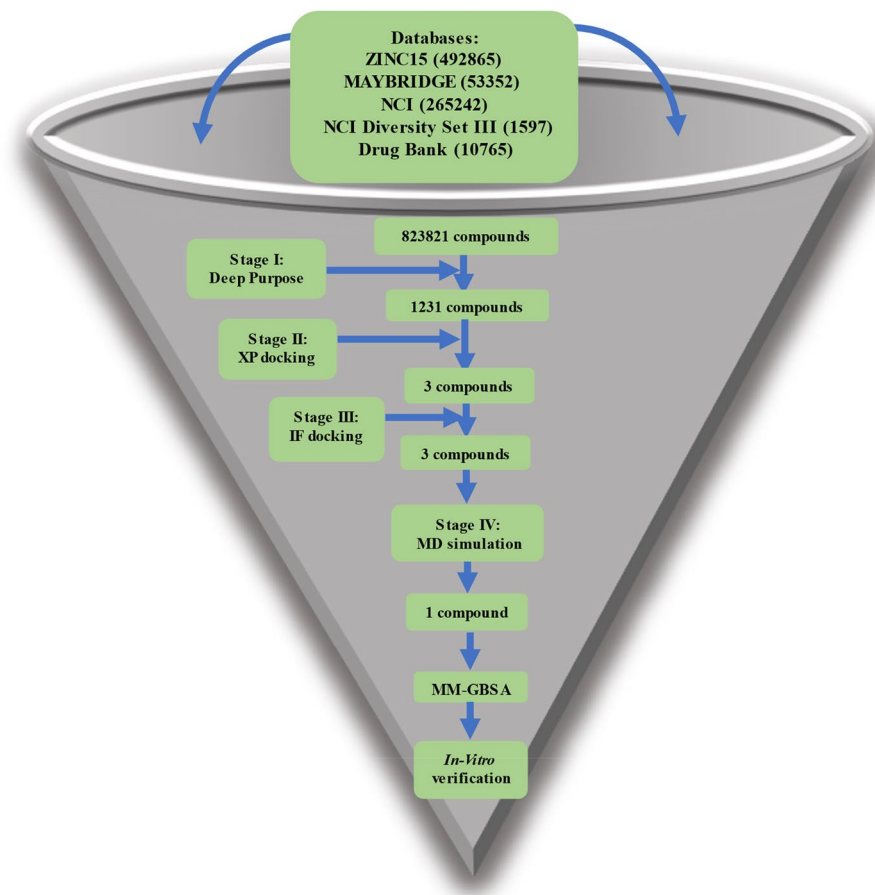


Figure 1. A workflow showing the steps performed in this study.

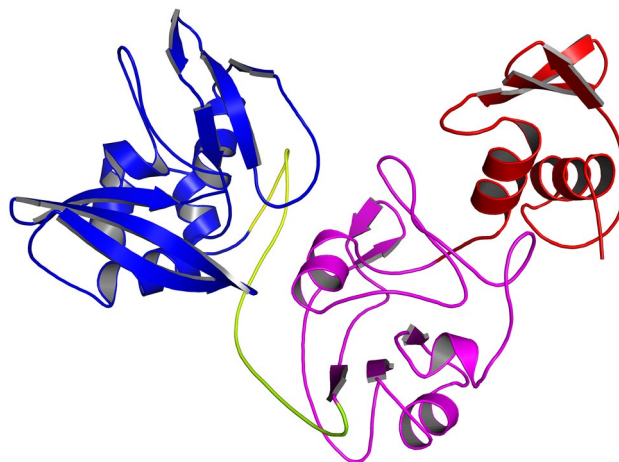


Figure 2. Shows the NSP15 domains. Red: The N-terminal domain. Magenta: The middle domain. Blue: The C-terminal domain which contains the active site of the protein.

dock the conformations of the prepared drugs to each protein, and the best results were selected using the score from the docking of the co-crystallized compound. The active site for endoribonuclease was H235, Q245, H250, K290, V292, T341, and Y343³⁰. The default setting was used in XP docking. After that, an IF docking for the best compounds from XP docking was performed using the default settings and the same active site for each protein.

MD simulation. MD simulation on the best three complexes (ZINC000104379474, ZINC000004715217, ZINC000408720658) from the IF docking was done as a triplicate using NAMD for 100 ns each to make sure that the results we obtain are not a coincidence or caused by the artificial forcefield used³¹. The systems were prepared using the input generator module in the CHARMM-GUI web server^{32–35}. For each complex, the system was solvated in the TIP3P water model. The temperature was set to 310 K, and the system was neutralized by adding Na⁺ and Cl⁻ ions with a concentration of 0.154 M. Furthermore, the periodic boundary conditions (PBC) were applied during the simulation. Each system was minimized for 10,000 steps using the Conjugate Gradient algorithm, and the time step was set to 2 femtoseconds. Following minimization, an equilibration step for a nanosecond in a constant number of atoms, constant pressure, and constant temperature (NPT) ensemble was performed. Temperature and pressure were maintained at 310 K and 1 atm using Langevin dynamics and Nose–Hoover Langevin piston. Finally, a production run of 100 ns was performed in a constant number of atoms, constant volume, and constant temperature (NVT) ensemble.

Molecular mechanics-generalized born surface area (MM-GBSA). Gmx_MMPBSA³⁶ was used to calculate the binding affinity using MM-GBSA with interaction entropy. The salt concentration was adjusted to 0.154 M, and the method of generalized born (igb) was set to 5. Additionally, decomposition of the free energy was obtained to determine the contribution of amino acids within 1 nm of the ligand.

Clustering and finding interactions. TtClust python library was used to cluster the trajectories and obtain a representative frame for each cluster. Alignment was performed first on the backbone before clustering, and the number of clusters was determined automatically using the elbow method³⁷. Protein–Ligand Interaction Profiler (PLIP) was used with each representative frame to find the number and types of interactions³⁸.

In vitro methods. Cytotoxicity (CC₅₀) determination. The compound that showed a good result in MD simulation was continued for in vitro studies to assess its therapeutic effect. The compound was obtained from NCI and dissolved in dimethyl sulfoxide (DMSO, Sigma-Aldrich) at a 1 mg/ml concentration and stored at 4 °C. To assess the CC₅₀, the stock solution of the compound was diluted further to the working solutions with DMEM (1–0.001 mg/ml). Cytotoxic activity was tested in VERO-E6 cells using a crystal violet assay as previously described by Feoktistova et al.³⁹ with minor modifications. Briefly, the cells were seeded in 96-well plates (100 µl/well at a density of 3 × 10⁵ cells/ml) and incubated for 24 h at 37 °C in 5% CO₂. After 24 h, the cells were treated with various concentrations of the compound in quadruplicate. At 72 h post-treatment, the supernatant was discarded, and cell monolayers were fixed with 10% formaldehyde for 1 h at room temperature (RT). The fixed monolayers were then dried thoroughly and stained with 50 µl of 0.1% crystal violet for 20 min on a bench rocker at room temperature. The monolayers were then washed and dried overnight. The crystal violet dye in each well was dissolved in 200 µl methanol for 20 min on a bench rocker at room temperature. The absorbance of the crystal violet solutions was measured at λ_{max} 570 nm as a reference wavelength using a multi-well plate reader. The CC₅₀ value was calculated using nonlinear regression analysis using GraphPad Prism software (version 5.01) by plotting log concentrations of the compound versus normalized response (variable slope).

$$\text{cytotoxicity \%} = \frac{\text{the absorbance of cells without treatment} - \text{the absorbance of cells with treatment}}{\text{the absorbance of cells without treatment}} \times 100$$

The concentration that displayed CC₅₀ was calculated using a plot of percent cytotoxicity vs sample concentration as previously described by Mosmann⁴⁰.

Inhibitory concentration 50 (IC₅₀) determination. The IC₅₀ value for the compound was determined as previously described⁴¹, with slight modifications. 2.4 × 10⁴ Vero-E6 cells were placed in each well of 96-well tissue culture plates and cultured overnight at 37 °C in a humidified 5% CO₂ incubator. After that, the cell monolayers were rinsed once in 1 × PBS. An aliquot of the SARS-CoV-2 "NRC-03-nhCoV" virus⁴² containing 100 TCID₅₀ was incubated with serially diluted concentrations of the tested compound and kept at 37 °C for 1 h. Another set of Vero-E6 cells was treated with virus/compound mix and co-incubated at 37 °C in a total volume of 200 µl per well. Untreated cells infected with a virus represent virus control; however, cells that have not been treated and have not been infected represent cell control. The cells were fixed with 100 µl of 10% paraformaldehyde for 20 min and stained with 0.5% crystal violet in distilled water for 15 min at room temperature after being incubated for 72 h at 37 °C in a 5% CO₂ incubator. After that, 100 µl absolute methanol per well was used to dissolve the crystal violet dye, and the optical density of the color was measured at 570 nm using an Anthos Zenyth 200rt plate reader (Anthos Labtec Instruments, Heerhugowaard, Netherlands). The IC₅₀ of the compound is that required to reduce the virus-induced cytopathic effect (CPE) by 50%, relative to the virus control. The IC₅₀ value was calculated using nonlinear regression analysis of GraphPad Prism software (version 5.01) by plotting log concentrations of ZINC000104379474 versus normalized response (variable slope).

Results

Screening results. All compounds that were downloaded from different databases, a total of 823,821 compounds, were converted into SMILES format for screening using DeepPurpose with the NSP15. The results from the library are continuous values indicating the binding score. The best compounds were selected using a threshold of 100 nM, which resulted in 1231 compounds for NSP15 (Supplementary Material 2). These compounds were prepared for docking, as mentioned in the methods section.

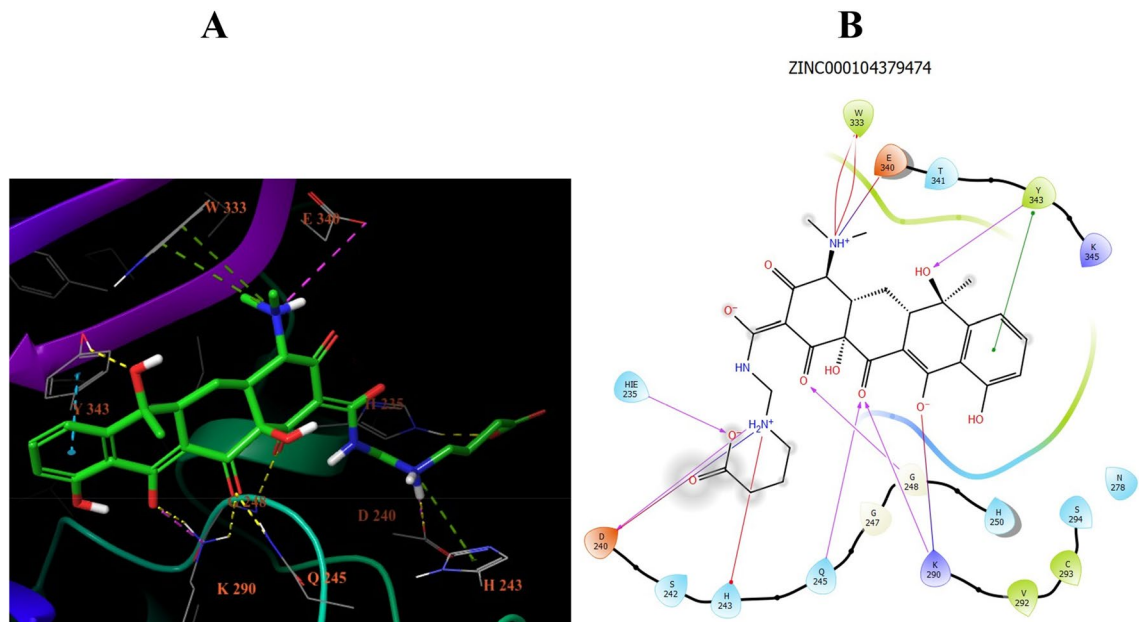


Figure 3. The 2D and 3D interactions between the IF docking of ZINC000104379474 and NSP15. There are 12 interactions in total. Half of them (six interactions) are H-bonds between H235, D240, Q245, G248, K290, and Y343, followed by three salt bridges (D240, K290, and E340), then one Pi-cation (H243) interaction and two Pi–Pi stacking (Y343).

Docking results. Before the XP docking of prepared compounds, a redocking step was performed with the co-crystallized compound to get a threshold docking score to filter the docking results of the prepared compounds. The threshold was -7.225 kcal/mol for NSP15. After filtering the XP docking results, three compounds were found to have a docking score better (more negative) than the threshold values. These compounds were further filtered using IF docking mode based on the same docking threshold from the redocking step. This produced the same three compounds for NSP15, and they were selected for MD simulation. Their IDs are ZINC000104379474 (73 atoms), ZINC000004715217 (71 atoms), and ZINC000408720658 (53 atoms). The 2D interactions from the IF docking of ZINC000104379474 with the NSP15 is shown in Fig. 3. The interaction consists mostly of H-bonds (six interactions) with H235, D240, Q245, G248, K290, and Y343. In addition, there are three salt bridges (D240, K290, and E340). Moreover, two Pi–Pi stacking interactions emerged with Y343 and a Pi–Cation interaction with H243.

MD simulation results. NAMD was used to perform the MD simulation of the best three complexes. For NSP15-ZINC000004715217 and NSP15-ZINC000408720658, the ligands were found to dissociate from the binding site during the production run for two replicates out of three (at 10 and 88 ns, respectively, for the first replicate and 43 and 58 ns, respectively, for the second replicate); therefore, these complexes were excluded from further analysis. On the other hand, ZINC000104379474 remained bound to NSP15 for the whole duration of the simulation in two out of the three replicates. Only the first replicate's results are reported in the main text, while the second replicate's results can be found in the Supplementary Material 1 file (Supplementary Figs. S2–S4, and Supplementary Tables S1 and S2). The NSP15-ZINC000104379474 complex shows average values of 23.7 Å, 2.34 Å, 18,669 Å², and 83 for the radius of gyration (RoG), root-mean-square displacement (RMSD), surface accessible surface area (SASA), and the total number of H-bonds, respectively, during the 100 ns simulation period as depicted in Fig. 4A. The stability of SASA, RoG, and H-bonds indicates the stability of the NSP15-ZINC000104379474 system. The RMSD values indicate system equilibration over the trajectory. The first 58 ns show a stable fluctuation except for the duration from 12.5 to 15 ns which shows a spike reaching above 4 Å before returning to stability. The RMSF plot (Fig. 4B) shows small fluctuations except for the N-terminal region (RMSF 6.5 \AA).

Figure 5 and Table 1 show the interactions in each cluster representative with ZINC000104379474 that were obtained from the PLIP webserver. The interaction between ZINC000104379474 and NSP15 is primarily by H-bonding. G248 and K290 are the most common residues that contribute to drug–protein interaction in all clusters, followed by V292, S294, and Y343. The most common amino acid responsible for hydrophobic contacts is Y343. The binding free energy calculation using MM–GBSA shows that the NSP15-ZINC000104379474 complex has an average binding energy of -7.68 kcal/mol. The decomposition of binding energy for the NSP15-ZINC000104379474 complex (Fig. 6) shows that amino acids H235, G247, G248, H250, K290, V292, W333, Y343 have binding affinity of -1.68 , -1.41 , -1.58 , -1.25 , -3.98 , -1.09 , -1.25 , -3.54 kcal/mol, respectively. Table 2 shows the different energy terms calculated by MM–GBSA and the binding energy.

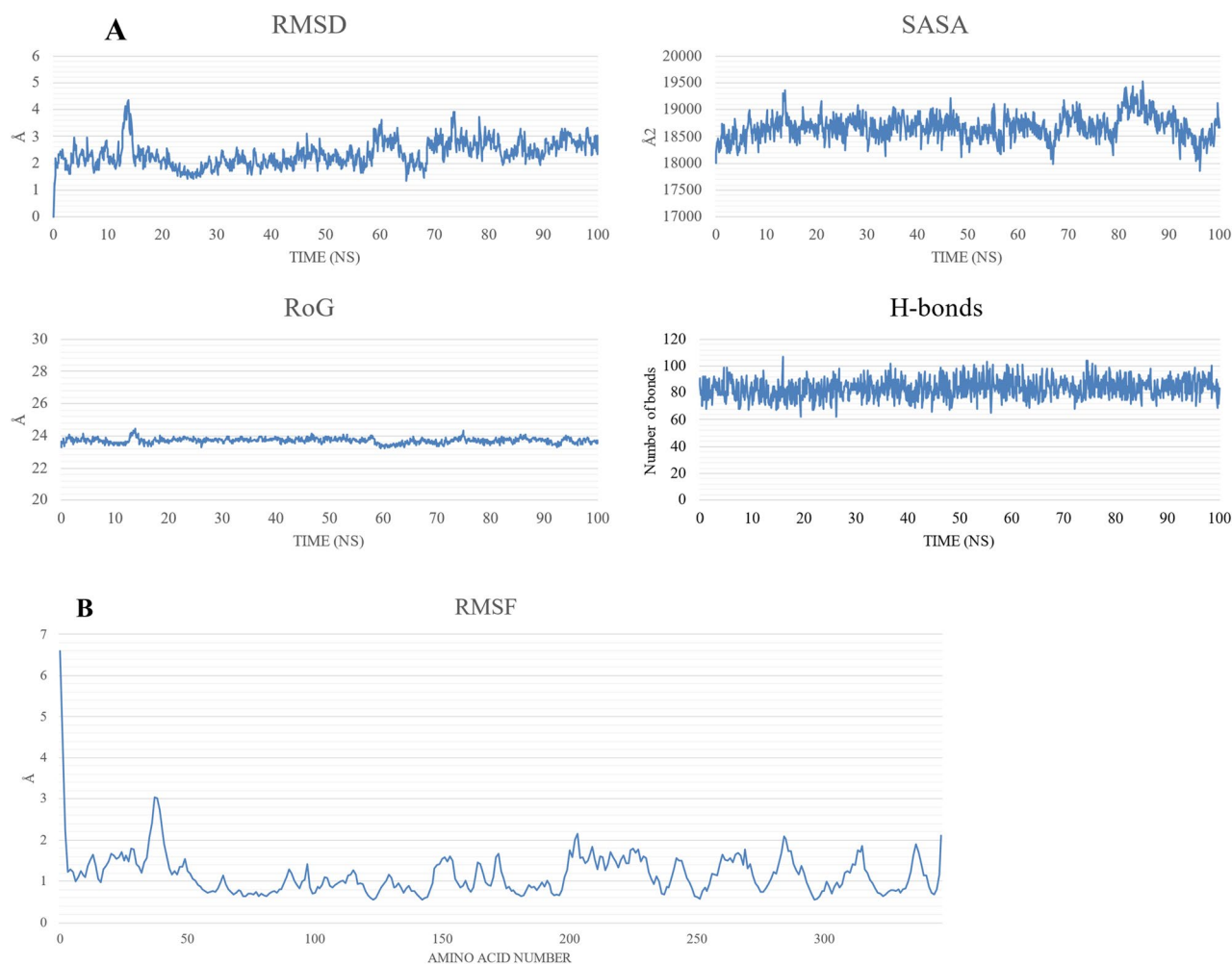


Figure 4. The molecular dynamics simulation analysis of the NSP15-ZINC000104379474 complex. (A) The radius of gyration (RoG) in Å, root-mean-square displacement (RMSD) in Å, surface accessible surface area (SASA) in Å², and the total number of H-bonds versus the simulation time in ns. (B) The per-residue root-mean-square fluctuation (RMSF) in Å.

ADMET prediction. The Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties for ZINC000104379474 were predicted using the pkCSM web server⁴³. Table 3 shows the predicted values for each property. ZINC000104379474 shows a water solubility value of $-2.899 \log(\text{mol/l})$ which is in the same region as other druggable compounds⁴⁴. The fraction unbound indicates the predicted fraction of the compound that will be unbound to serum proteins. High unbound fraction and low Blood-Brain Barrier (BBB) permeability ($\log \text{BBB}$) indicate a good distribution. Inhibitors of Cytochrome P450 can activate the drug metabolism and, therefore, can remove the compound from the market. ZINC000104379474 compound was used to predict whether it is an inhibitor for different isoforms (CYP1A2, CYP2C19, CYP2C9, CYP2D6, and CYP3A4). The server predicted that the compound is not an inhibitor for any of the isoforms. For Excretion, the model predicts that the compound is not a substrate for renal organic cation transporter 2. The interaction with this transporter helps in the clearance of the compound and may produce adverse interactions; therefore, a negative prediction is considered good. Finally, four indicators are used to predict the toxicity of ZINC000104379474. Ames toxicity is a test that indicates whether the compound is a carcinogen. Inhibition of hERG I/II is the principal cause of fatal ventricular arrhythmia and has resulted in the withdrawal of many substances. As its name implies, Hepatotoxicity indicates whether the compound may disrupt the liver's normal function. Fortunately, the server predicts that the compound does not cause any toxicity.

In vitro results. The tested sample of the ZINC000104379474 compound against the SARS-CoV-2 virus in VERO E6 cells showed IC₅₀ and CC₅₀ values of 0.01 and 0.9 mg/ml (Fig. 7), respectively. Therefore, its selectivity index is 90. Since the CC₅₀ value is 90-fold greater than the IC₅₀, this compound shows a potential to be used against the SARS-CoV-2 virus. (CC₅₀ > IC₅₀; SI > 1 (90)).

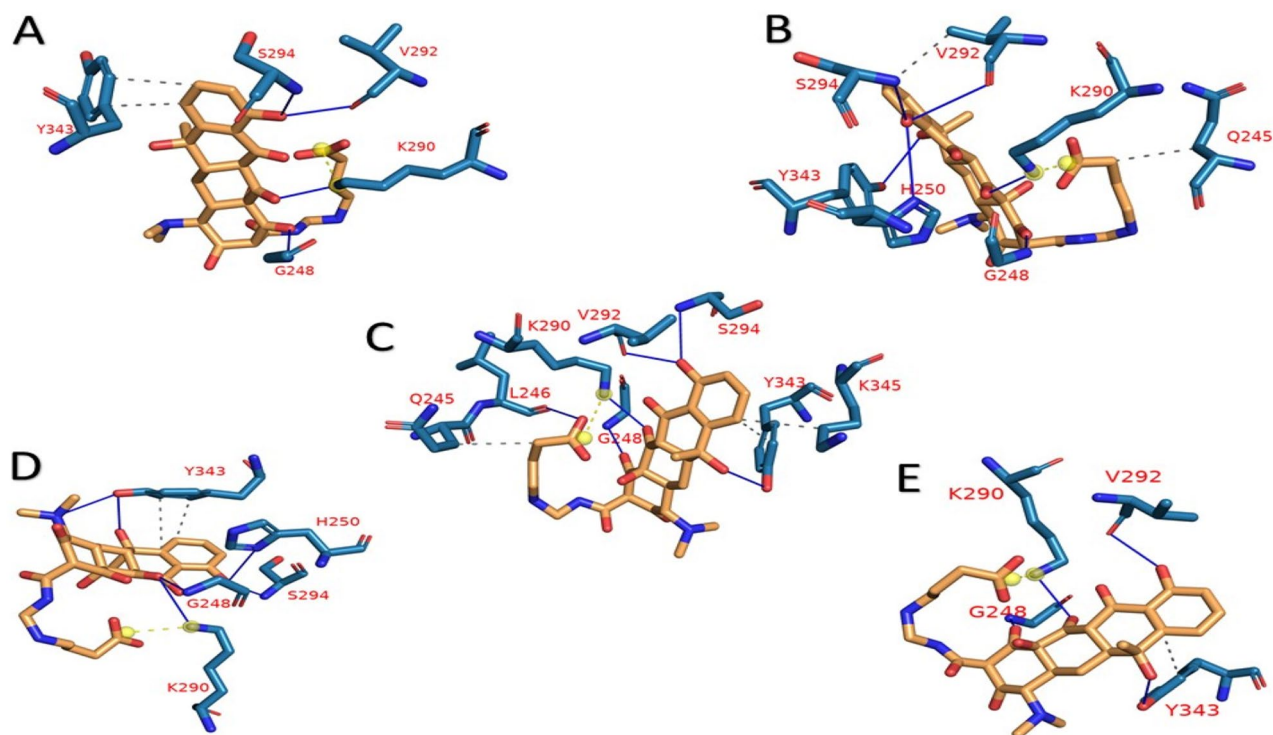


Figure 5. the representative frames (A–E) from clustering using TtClust, and the types and number of interactions calculated and depicted using PLIP webserver for NSP15-ZINC000104379474 complex. Dashed-Grey lines: Hydrophobic interactions, blue lines: H-bond, dashed-yellow lines: salt-bridges.

Cluster	Number of hydrogen bonds	Amino acids in NSP15	Number of hydrophobic interactions	Amino acids in NSP15	Number of salt bridge interactions	Amino acids in NSP15
A	4	G248 – K290 – V292 – S294	2	Y343 (2)	1	K290
B	6	G248 – H250 – K290 – V292 – S294 – Y343	2	Q245– V292	1	K290
C	6	L246 – G248 – K290 – V292 – S294 – Y343	3	Q245– Y343 – K345	1	K290
D	6	G248 – H250 – K290 – S294 – Y343 (2)	2	Y343 (2)	1	K290
E	4	G248 – K290 – V292 – Y343	1	Y343	1	K290

Table 1. PLIP web server analysis for NSP15-ZINC000104379474 complex. Bold amino acids indicate that they are most common in all cluster representatives.

Discussion

Since the emergence of SARS-CoV-2, many research groups worldwide have started searching for drugs that can affect the virus life cycle utilizing different in silico techniques ranging from classical virtual screening and molecular docking to the usage of state-of-the-art deep learning models. Some of the recent work done by other research groups to find new inhibitors for SARS-CoV-2 proteins is mentioned here. El-Demerdash, A. et al. have screened 15 guanidine alkaloids against two NSPs (NSP10 and NSP5) and three structural proteins (S–N–M), calculated their binding affinities, determined their structure–activity relationship (SAR), and predicted their Absorption, Distribution, Metabolism, Excretion, and Toxicity properties. Two of the screened compounds (crambescidin 786 and crambescidin 826) showed promising results against their target proteins⁴⁵. El Hassab et al. have screened 48 million druglike compounds obtained from the ZINC database against NSP16 using a pharmacophore-based screening which resulted in only 24 compounds. These compounds were subsequently filtered into four based on the target protein's docking scores before running an MD simulation. Their results showed that only one compound had a good binding affinity compared to the pan-methyl-transferase inhibitor Sinefungin and might act as an inhibitor for NSP16⁴⁶. Alazmi et al. have screened approximately 100 thousand natural compounds from the ZINC database against NSP4, NSP15, RdRp, and ACE2, followed by docking and MD simulation. Their work showed that Baicalin could bind to the three viral proteins, while Limonin can bind ACE2, rendering it inefficient in binding SARS-CoV-2 RBD⁴⁷. Elgohary et al. targeted the main protease of SARS-CoV 2 virus with 15 batzelladine marine alkaloids. They started their approach by performing docking

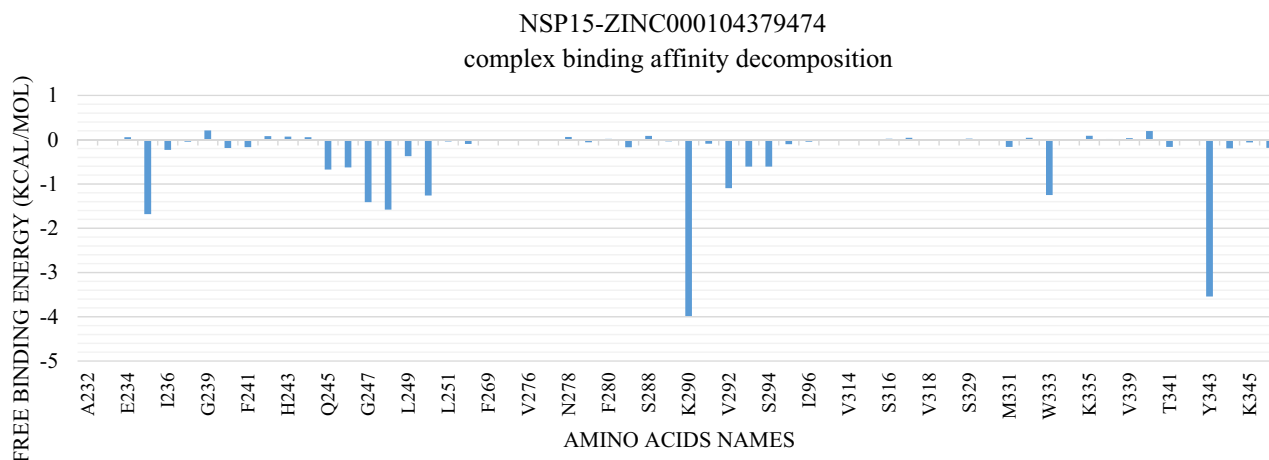


Figure 6. Molecular Mechanics Generalized Born Surface Area (MM-GBSA) decomposition of NSP15-ZINC000104379474 complex. It shows the contribution of amino acids within 1 nm of the ZINC000104379474 compound.

System	ΔE_{VDW}	ΔE_{ELE}	ΔG_{GB}	ΔG_{SA}	ΔG_{GAS}	ΔG_{SOLV}	ΔG_{TOTAL}	-TAS	$\Delta G_{BINDING}$
NSP15-ZINC000104379474 complex	-31.08 ± 3.68	-251.31 ± 15.34	260.34 ± 14.42	-4.90 ± 0.29	-282.39 ± 14.63	255.44 ± 14.35	-26.95 ± 5.18	19.28 ± 2.1	-7.68 ± 5.59

Table 2. Binding free energies and individual energy terms and their standard deviation of NSP15-ZINC000104379474 complex calculated by MM-GBSA. Units are in kcal/mole.

Property	Model name	Predicted value
Absorption	Water solubility	-2.899
Distribution	Fraction unbound (human)	0.353
	Blood-brain barrier permeability	-1.676
Metabolism	CYP1A2 inhibitor	No
	CYP2C19 inhibitor	No
	CYP2C9 inhibitor	No
	CYP2D6 inhibitor	No
	CYP3A4 inhibitor	No
Excretion	Renal OCT2 substrate	No
Toxicity	AMES toxicity	No
	hERG I inhibitor	No
	hERG II inhibitor	No
	Hepatotoxicity	No

Table 3. pkCSM webservice analysis of ADMET properties for ZINC000104379474 compound.

followed by MD simulation and SAR⁴⁸. Kumar et al. have performed high throughput screening of an in-house database against NSP15 followed by MD simulation for the high-ranked compounds. A derivative of piperazine was found to bind to the active site of NSP15 and demonstrated inhibition of viral replication by in vitro assays². Mahmud et al. have performed a repurposing study of 23 compounds against NSP15 and found three compounds with a predicted binding affinity better than -9.0 kcal/mol¹⁷. Savale et al. have screened 8722 small molecules from the Asinex antiviral database against NSP15 protein. They prepared the compounds and the protein using the LigPrep module in the Schrödinger suite and protein preparation wizard before docking using a multi-step docking workflow. The top-ranked molecules were selected based on the Prime-MMGBSA approach and Glide-XP score. The top five molecules were then checked for druglike properties and toxicity using SwissADME and PkCSM webservers. Finally, they performed MD simulation for the complexes and calculated binding free energy using the MM-PBSA approach and found that five compounds showed strong binding affinity⁴⁹.

Nidoviral RNA uridylylate-specific endoribonuclease (NendoU or NSP15) is one of the crucial 16 NSPs of SARS-CoV-2. Its C-terminal has a catalytic site specific to uridine. It was recently discovered that NSP15 has a role in the innate immune response, and therefore it is considered an important target for combatting the infection^{50,51}. NSP15 consists of the N-terminal, middle, and catalytic C-terminal domains. It specifically cuts the

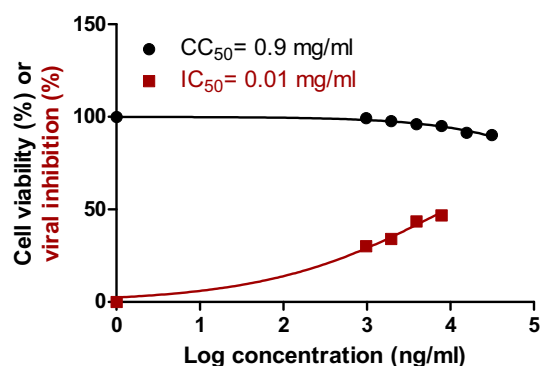


Figure 7. Dose–response and inhibition curves for ZINC000104379474 compound on VERO-E6 cells. Half maximal cytotoxic concentration (CC₅₀) in Vero E6 cells (black) and inhibitory concentration 50% (IC₅₀) against NRC-03-nhCoV were calculated using nonlinear regression analysis of GraphPad Prism.

double-stranded RNA through an Mn²⁺-dependent activity. The C-terminal domain consists of two antiparallel β -sheets that hold six key amino acids (H235, H250, K290, S294, T341, and Y343)⁵¹.

In this work, we performed a multi-database virtual screening with 823,821 compounds using a Deep Learning approach followed by docking using the XP and IF docking implemented in Schrödinger Suite²⁸. Only 1231 compounds were selected from the results of the deep learning library based on a threshold of 100 nM. After their preparation using LigPrep in Maestro, 1314 different conformations were obtained. These conformations were docked using the XP method, and the top compounds (3 compounds) were selected for the IF docking protocol. The compounds with a better docking score (3 compounds) than that of the co-crystallized drug were selected for the subsequent analyses. The three molecules were selected for three replicates of 100 ns MD simulation each. After removing the PBC, the trajectory was analyzed for its stability using the TK console in VMD⁵², then the binding affinity with entropy was calculated using the MM-GBSA approach in gmx_MMPBSA³⁶. Figure 4 shows the plots of the analyses done on the trajectories.

Furthermore, the trajectories were clustered to obtain representative frames using TtClust³⁷. These frames were used with PLIP to detect the types of interactions between NSP15 and ZINC000104379474³⁸. This compound was found to interact with most of the key amino acids responsible for the catalytic activity of NSP15 with an average binding affinity of -7.68 kcal/Mol. H250 forms one H-bond in just two clusters with -1.25 kcal/mol contribution to the total binding energy. Besides, K290 is a common amino acid among the five clusters used to analyze the interaction. It interacts with ZINC000104379474 via one H-bond and one salt bridge in each cluster. It shows the most substantial binding among the other amino acids with a -3.98 kcal/mol contribution. On the other hand, S294 interacts with ZINC000104379474 through both H-bond and hydrophobic interaction with only -0.6 kcal/mol in the first four clusters. Additionally, Y343 forms two types of interactions with ZINC000104379474. In cluster 1, it forms one hydrophobic interaction, while in the second cluster, it forms one H-bond. In addition, it forms both types of interaction in the last three clusters, with only four interactions (two H-bonds and two hydrophobic interactions) in cluster 4. It has a contribution of -3.54 kcal/mol to the total binding affinity. Although H235 is not found to have any interaction with the compound in these five clusters, it has a contribution of -1.68 kcal/mol. On the other hand, T341 has a minimal contribution of only -0.16 kcal/mol. Moreover, G248 forms one hydrogen bond in all cluster representatives with a contribution of -1.58 kcal/Mol. In addition to this, V292 forms one hydrogen bond in four out of five cluster representatives with a contribution of -1.09 kcal/Mol. Although G247 and W333 were not found to form any interaction in the cluster representatives, they have a contribution of -1.41 kcal/Mol, and -1.25 kcal/Mol, respectively.

Based on the *in silico*, the CC₅₀, and the IC₅₀ results, ZINC000104379474 is expected to have potent inhibitory activity against SARS-CoV-2 NSP15. Therefore, it is recommended to be used as an antiviral agent after *in vivo* validation.

SARS-CoV-2 is still affecting our daily lives despite many vaccine shots. Therefore, an urgent need for possible therapeutics is mandatory to inhibit the viral spread. In this study, we screened five different databases for possible anti-SARS-CoV-2 endoribonuclease. The compounds are then filtered using different computational methods giving us three compounds that bind the NSP15. The MDS further select the compound ZINC000104379474 as the best potential NSP15 inhibitor. The *in vitro* study further validates this result as the ZINC000104379474 has IC₅₀ and CC₅₀ values of 0.01 and 0.9 mg/ml. Further *in vivo* validation is suggested as future work.

Data availability

Data is available upon request from the corresponding author.

Received: 30 May 2022; Accepted: 27 July 2022

Published online: 03 August 2022

References

1. ICTV Code. <https://talk.ictvonline.org/information/w/ictv-information/383/ictv-code>. Accessed on 1 May 2022.

2. Kumar, S. *et al.* A novel compound active against SARS-CoV-2 targeting Uridylate-specific endoribonuclease (NendoU/NSP15): In silico and in vitro investigations. *RSC Med. Chem.* **12**, 1757–1764 (2021).
3. Zhu, Z. *et al.* From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **21**, 1–14 (2020).
4. Al-Qahtani, A. A. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): Emergence, history, basic and clinical aspects. *Saudi J. Biol. Sci.* **27**, 2531–2538 (2020).
5. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
6. Zhou, H. *et al.* A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**, 2196–2203 (2020).
7. Gordon, D. E. *et al.* A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
8. Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M. & Bergeron, J. R. C. Structural characterization of SARS-CoV-2: Where we are, and where we need to be. *Front. Mol. Biosci.* **7**, 344 (2020).
9. Pillon, M. C. *et al.* Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat. Commun.* **12**, 1–12 (2021).
10. Hong, S. *et al.* Epigallocatechin gallate inhibits the uridylate-specific endoribonuclease Nsp15 and efficiently neutralizes the SARS-CoV-2 strain. *J. Agric. Food Chem.* **69**, 5948–5954 (2021).
11. Yoshimoto, F. K. A biochemical perspective of the nonstructural proteins (NSPs) and the spike protein of SARS CoV-2. *Protein J.* **40**, 260–295 (2021).
12. Mishra, G. P. *et al.* The interaction of the bioflavonoids with five SARS-CoV-2 proteins targets: An in silico study. *Comput. Biol. Med.* **134**, 104464 (2021).
13. Ulferts, R. & Ziebuhr, J. Nidovirus ribonucleases: Structures and functions in viral replication. *RNA Biol.* **8**, 295–304 (2011).
14. Hackbart, M., Deng, X. & Baker, S. C. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc. Natl. Acad. Sci.* **117**, 8094–8103 (2020).
15. Kindler, E. *et al.* Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLoS Pathog.* **13**, e1006195 (2017).
16. Gao, B. *et al.* Inhibition of anti-viral stress granule formation by coronavirus endoribonuclease nsp15 ensures efficient virus replication. *PLoS Pathog.* **17**, e1008690 (2021).
17. Mahmud, S. *et al.* Targeting SARS-CoV-2 nonstructural protein 15 endoribonuclease: An in silico perspective. *Future Virol.* **16**, 467–474 (2021).
18. Kumari, M. & Subbarao, N. Deep learning model for virtual screening of novel 3C-like protease enzyme inhibitors against SARS coronavirus diseases. *Comput. Biol. Med.* **132**, 104317 (2021).
19. Carpenter, K. A., Cohen, D. S., Jarrell, J. T. & Huang, X. Deep learning and virtual drug screening. *Future Med. Chem.* **10**, 2557–2567 (2018).
20. McNutt, A. T. *et al.* GNINA 1.0: Molecular docking with deep learning. *J. Cheminform.* **13**, 1–20 (2021).
21. Ghasemi, F., Mehridehnavi, A., Perez-Garrido, A. & Perez-Sanchez, H. Neural network and deep-learning algorithms used in QSAR studies: Merits and drawbacks. *Drug Discov. Today* **23**, 1784–1790 (2018).
22. Xu, Y. *et al.* Deep learning for drug-induced liver injury. *J. Chem. Inf. Model* **55**, 2085–2093 (2015).
23. Gianchandani, N., Jaiswal, A., Singh, D., Kumar, V. & Kaur, M. Rapid COVID-19 diagnosis using ensemble deep transfer learning models from chest radiographic images. *J. Ambient Intell. Humaniz. Comput.* **16**, 1–13 (2020).
24. Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **181**, 475–483 (2020).
25. Hu, S. *et al.* Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks. *BMC Bioinform.* **20**, 1–12 (2019).
26. Wang, S., Sun, Q., Xu, Y., Pei, J. & Lai, L. A transferable deep learning approach to fast screen potential antiviral drugs against SARS-CoV-2. *Brief Bioinform.* **22**, 1–11 (2021).
27. Huang, K. *et al.* DeepPurpose: A deep learning library for drug–target interaction prediction. *Bioinformatics* **36**, 5545–5547 (2020).
28. Maestro, Schrödinger, LLC. (2021).
29. O’Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminform.* **3**, 33 (2011).
30. Kim, Y. *et al.* Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* **29**, 1596–1605 (2020).
31. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).
32. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
33. Lee, J. *et al.* CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
34. Jo, S. *et al.* in *Biomolecular Modelling and Simulations* (ed. Karabencheva-Christova, T.) **96**, 235–265 (Academic Press, 2014).
35. Brooks, B. R. *et al.* CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **30**, 1545–1614 (2009).
36. Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A. & Moreno, E. gmx_MMPBSA: A new tool to perform end-state free energy calculations with GROMACS. *J. Chem. Theory Comput.* **17**, 6281–6291 (2021).
37. Tubiana, T., Carvaille, J.-C., Boulard, Y. & Bressanelli, S. TtClust: A versatile molecular simulation trajectory clustering program with graphical summaries. *J. Chem. Inf. Model* **58**, 2178–2182 (2018).
38. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: Fully automated protein–ligand interaction profiler. *Nucleic Acids Res.* **43**, W443–W447 (2015).
39. Feoktistova, M., Geserick, P. & Leverkus, M. Crystal violet assay for determining viability of cultured cells. *Cold Spring Harb. Protoc.* **2016**, pdb-prot087379 (2016).
40. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J. Immunol. Methods* **65**, 55–63 (1983).
41. Mostafa, A. *et al.* FDA-approved drugs with potent in vitro antiviral activity against severe acute respiratory syndrome coronavirus 2. *Pharmaceuticals* **13**, 443 (2020).
42. Kandeil, A. *et al.* Coding-complete genome sequences of two sars-cov-2 isolates from egypt. *Microbiol. Resour. Announc.* **9**, e00489–e520 (2020).
43. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. pkCSM: Predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.* **58**, 4066–4072 (2015).
44. Wang, J. *et al.* Development of reliable aqueous solubility models and their application in druglike analysis. *J. Chem. Inf. Model* **47**, 1395–1404 (2007).
45. El-Demerdash, A. *et al.* Comprehensive virtual screening of the antiviral potentialities of marine polycyclic guanidine alkaloids against SARS-CoV-2 (COVID-19). *Biomolecules* **11**, 460 (2021).
46. El Hassab, M. A. *et al.* In silico identification of novel SARS-COV-2 2'-O-methyltransferase (nsp16) inhibitors: Structure-based virtual screening, molecular dynamics simulation and MM-PBSA approaches. *J. Enzyme Inhib. Med. Chem.* **36**, 727–736 (2021).
47. Alzami, M. & Motwalli, O. In silico virtual screening, characterization, docking and molecular dynamics studies of crucial SARS-CoV-2 proteins. *J. Biomol. Struct. Dyn.* **39**, 6761–6771 (2021).
48. Elgohary, A. M. *et al.* Investigating the structure-activity relationship of marine polycyclic batzelladine alkaloids as promising inhibitors for SARS-CoV-2 main protease (Mpro). *Comput. Biol. Med.* **147**, 105738 (2022).

49. Savale, R. U. *et al.* Pharmacoinformatics approach based identification of potential Nsp15 endoribonuclease modulators for SARS-CoV-2 inhibition. *Arch Biochem. Biophys.* **700**, 108771 (2021).
50. Deng, X. *et al.* Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci.* **114**, E4251–E4260 (2017).
51. Sinha, S. K. *et al.* An in-silico evaluation of different Saikosaponins for their potency against SARS-CoV-2 using NSP15 and fusion spike glycoprotein as targets. *J. Biomol. Struct. Dyn.* **39**, 3244–3255 (2021).
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph* **14**, 33–38 (1996).

Acknowledgements

Bibliotheca Alexandrina HPC in Alexandria, Egypt, is utilized to perform the MDS calculations. We acknowledge Grant Number 5202 PRISM from the Academy of scientific research and technology, Ministry of Scientific Research.

Author contributions

I.I. performed the calculations and drafted the manuscript. S.M. performed the in vitro analysis. M.E. designed the study, and methodology, and funded the study. A.E. revised the manuscript and draw figures. M.F. revised the manuscript. A.E., M.F. & M.E. supervised I.I. All the authors approve the final version of the document.

Funding

Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). This work was supported by the Academy of scientific research and technology, Ministry of Scientific Research [Grant number 5202 PRISM].

Competing interests

All the authors have currently submitted a patent to the Egyptian Patent Office with patent number EG/P/2022/413.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-17573-6>.

Correspondence and requests for materials should be addressed to M.E.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022