[Short title + Author Name - P&H title] 32 (2024) 101914

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

Saudi Pharmaceutical Journal

journal homepage: www.sciencedirect.com

Inhibition of SARS–CoV–2 NSP–15 by Uridine–5′–Monophosphate Analogues Using QSAR Modelling, Molecular Dynamics Simulations, and Free Energy Landscape

Mohammed Merae Alshahrani

Department of Clinical Laboratory Sciences, Faculty of Applied Medical Sciences, Najran University, 1988, Najran 61441, Saudi Arabia

ARTICLE INFO

Keywords:
SARS–CoV–2
NSP15
Uridine 5′–monophosphate
QSAR
Molecular dynamics

ABSTRACT

SARS–CoV–2 is accountable for severe social and economic disruption around the world causing COVID–19. Non–structural protein–15 (NSP15) possesses a domain that is vital to the viral life cycle and is known as uridylate–specific endoribonuclease (EndoU). This domain binds to the uridine 5′–monophosphate (U5P) so that the protein may carry out its native activity. It is considered a vital drug target to inhibit the growth of the virus. Thus, in this current study, ML–based QSAR and virtual screening of U5P analogues targeting Nsp15 were performed to identify potential molecules against SARS–CoV–2. Screening of 816 unique U5P analogues using ML–based QSAR identified 397 compounds ranked on their predicted bioactivity (pIC50). Further, molecular docking and hydrogen bond interaction analysis resulted in the selection of the top three compounds (**53309102**, **57398422**, and **76314921**). Molecular dynamics simulation of the most promising compounds showed that two molecules **53309102** and **57398422** acted as potential binders of Nsp15. The compound was able to inhibit nsp15 activity as it was successfully bound to the active site of the nsp15 protein. This was achieved by the formation of relevant contacts with enzymatically critical amino acid residues (His235, His250, and Lys290). Principal component analysis and free energy landscape studies showed stable complex formation while MM/GBSA calculation showed lower binding energies for **53309102** (ΔG_{TOTAL} = –29.4 kcal/mol) and **57398422** (ΔG_{TOTAL} = –39.4 kcal/mol) compared to the control U5P (ΔG_{TOTAL} = –18.8 kcal/mol). This study aimed to identify analogues of U5P inhibiting the NSP15 function that potentially could be used for treating COVID–19.

1. Introduction

As per international assessments, the 2019 coronavirus pandemic, characterized by its severity and global dissemination, was attributed to the SARS–CoV–2 pathogen (severe acute respiratory syndrome coronavirus 2), resulting in widespread fatalities on a global scale (Malik et al., 2020; Wang et al., 2020). In the most recent 28 days (17 July to 13 August 2023), the WHO's six regions recorded over 1.4 million new cases of COVID–19 and over 2300 fatalities. This represents a rise of 63 % and a decline of 56 %, respectively, when compared to the most recent 28 days prior. As of August 13, 2023, the global tally of verified COVID–19 cases has exceeded 769 million, with the number of recorded deaths surpassing 6.9 million ("Weekly epidemiological update on COVID–19 – 17 August 2023," n.d.). According to the findings of an investigation of the genomic data, SARS–CoV–2 is most comparable to earlier strains of SARS and MERS and expresses 75 and 85 % of sequence similarity (Hu et al., 2021; Mercatelli and Giorgi, 2020; Sanders et al., 2020; Zhu et al., 2020). Humans, birds, and other animals are all susceptible to the gastroenteritis, hepatitis, and even death that can be caused by coronaviruses (Chafekar and Fielding, 2018). Multiple comorbidities were found to significantly elevate susceptibility to this syndrome. SARS–CoV infections are more common in the elderly because of the prevalence of comorbidities such as Parkinson's disease, cancer, stroke, high blood pressure, chronic bronchitis, diabetes, and chronic obstructive pulmonary disease (Deng and Peng, 2020; Guan et al., 2020; Huang et al., 2020). Constant surveillance of SARS–CoV–2 transmission across the world is an important aspect of the worldwide plan to limit the development of mutations. Nevertheless, it is also essential to monitor their propagation in animal populations and individuals who are chronically infected with the virus ("Tracking SARS–CoV–2 variants," n.d.).

The SARS–CoV–2 virus is a member of the Nidovirales order, the family Coronaviridae, and the smallest known RNA virus genome (Gorbalenya et al., 2006; Kim et al., 2020a; V'kovski et al., 2021).

E-mail address: mmalshahrani@nu.edu.sa.

https://doi.org/10.1016/j.jsps.2023.101914
Received 9 September 2023; Accepted 9 December 2023
Available online 10 December 2023

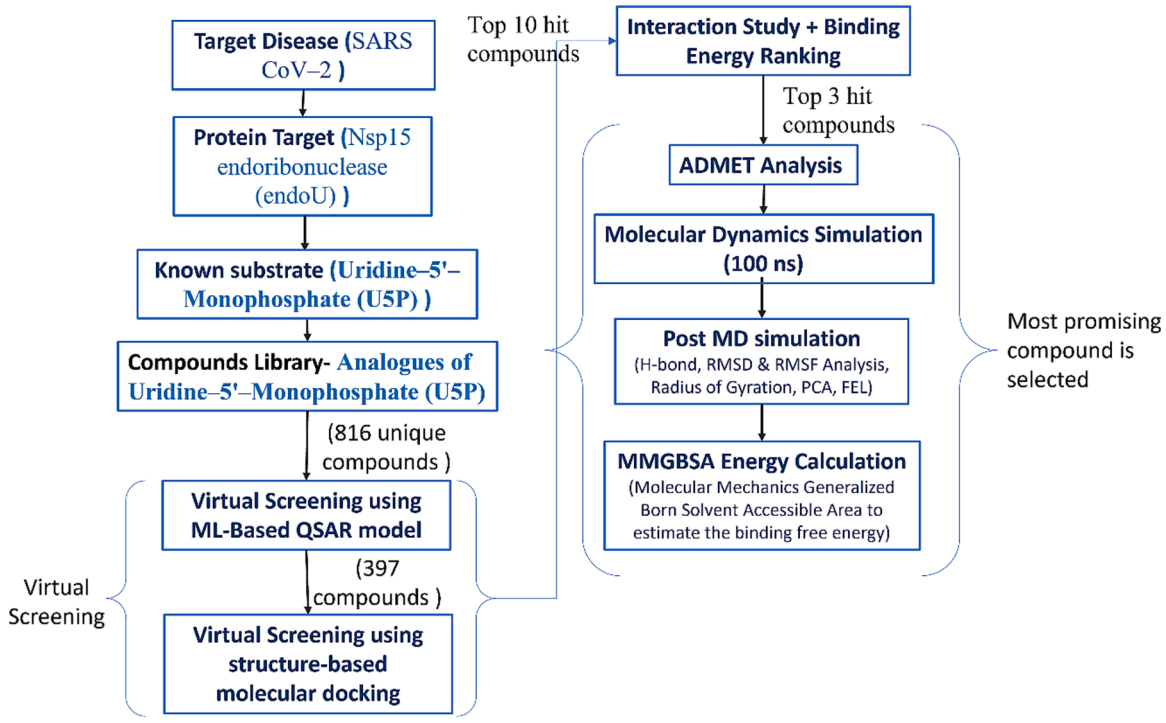

Fig. 1. Workflow of the methodology used in the present study.

Positive, non–segmented, and single–stranded genomic RNA is found in the coronaviruses' genomic makeup (Singh et al., 2021). The translation of the structural proteins and accessory proteins, which are derived from nested sub–genomic mRNA, is necessary for the expression and replication of the genome (Sawicki et al., 2007). The replication–translation complex (RTC) is formed when the non–structural proteins (NSPs) interact with one another to form a complex. The RTC is responsible for proofreading, capping, and RNA synthesis (Snijder et al., 2016). NSPs stand as the central component within the replication–transcription complex. Their primary role involves encoding the functional region situated within nsp12, a molecule more frequently denoted as RNA–dependent RNA polymerase (RdRp) (Cheng et al., 2005). The role of nsp3 and nsp5 revolves around facilitating the assembly of functional protein complexes and regulating the operational behavior of the remaining enzymes (Subissi et al., 2014). The virus's growth and replication mechanisms rely on a multitude of critical constituents, which encompass RdRp (nsp12/nsp7/nsp8), RNA helicase (nsp13), RNA exoribonuclease (nsp14/nsp10), RNA endoribonuclease (Nsp15), and RNA Cap methyltransferases (nsp14 and nsp16/nsp10). These elements collaboratively contribute to the virus's reproduction process. Therefore, in contrast to the majority of other viruses, coronaviruses contain a significant number of enzymes that have the potential to be employed as targets in the development of new antiviral drugs.

It is well known that Nsp15 (nonstructural protein–15) is crucial to the development of the pathology caused by coronaviruses. The C–terminal 'EndoU' domain of Nsp15 serves as a uridylate–specific endoribonuclease in the context of single– and double–stranded RNA molecules by cleaving the 3′ end of pyrimidines, specifically uridylates (Deng and Baker, 2018; Godoy et al., 2023; Ivanov et al., 2004; Kang et al., 2007). Evidence suggests that this RNase degrades dsRNA intermediates produced by viruses, thereby blocking their recognition by the host (Deng et al., 2017; Kindler et al., 2017). Additionally, Nsp15 plays a crucial part in the processing of the viral genome, which is required for the replication of coronaviruses (Pillon et al., 2021). The elimination of Nsp15 leads to a dramatic reduction in viral replication (Gao et al., 2021; Kang et al., 2007; Kindler et al., 2017). Nsp15 interferes with the identification of the viral RNA, which directly delays the type I interferon response. The anti–interferon role of Nsp15 has been confirmed by multiple studies using primary cells infected with live viruses (Deng et al., 2017; Kindler et al., 2017). In the context of single– and double–stranded RNA molecules, the C–terminal 'EndoU' domain of Nsp15 functions as a uridylate–specific endoribonuclease that cleaves the 3′ end of pyrimidines, specifically uridylates (Hackbart et al., 2020). Nsp15 is an endoribonuclease unique to nidoviruses and has no close human homologs, thus can be considered as an excellent target for the development of new drugs to treat SARS–CoV–2 infections. Several studies showed identification of successful inhibitors against SARS–CoV–2 infections that targeted Nsp15 protein (Chandra et al., 2021; Hong et al., 2021; Kim et al., 2021; Saramago et al., 2022; Sinha et al., 2020). The catalytic triad of Nsp15′s active site consists of two histidines and a lysine, and it bears a striking resemblance to the active site of the well–studied endoribonuclease RNase A (Godoy et al., 2023; Kim et al., 2020b; Pillon et al., 2021). Since Nsp15 appears to be the endoribonuclease responsible for suppressing the innate immune response, it presents itself as a promising therapeutic target.

In a study, uridine 5′–monophosphate (U5P) bound to Nsp15 active site showed interactions with important residues while it is also crystalized with the nsp15 structure (Pillon et al., 2021). In a previous study, it was found that Tipiracil, a uracil analogue, binds to the Nsp15 uracil site (uridine 5′–monophosphate binding site) that inhibits Nsp15 RNA nuclease activity tested in vitro (Kim et al., 2021). In this current study, analogues of U5P which acts as a substrate of Nsp15 were screened for their ability to inhibit SARS–CoV–2 Nsp15. Both ML–based quantitative structural analysis and molecular docking–based screening were used in the discovery of putative inhibitors of Nsp15. The compounds were stable when docked with the nsp15 protein, as demonstrated by molecular dynamics modelling, principal component analysis, and free energy landscape. In addition, the binding free energy of the complexes was computed so that the binding affinity of the molecule could be verified. As a result, the purpose of this investigation was to deploy computational approaches in order to screen and analyse the efficiency of U5P analogues against Nsp15, the SARS–CoV–2 binding protein. Fig. 1 showed the workflow of the methodology used in this study.

2. Materials and Methods

2.1. Protein and compound library

The Nsp15 of SAR–CoV–2 complexed with uridine–5′–monophosphate (U5P) was selected as a target protein in this study. The crystal structure was retrieved using the PDB ID: 6WLC (Kim et al., 2021) from the Protein Data Bank (PDB) (Berman et al., 2000). In this study, the protein's chain A was obtained due to the protein's homodimeric nature. The binding site residues, which were located around the U5P molecule at a distance of 6 Å, were utilised to generate the grid box for conducting the subsequent molecular docking analysis. The substrate, U5P was used as a control for this study, and was used to generate analogues using the Polypharmacology Browser 2 server (PPB2) (Awale and Reymond, 2018). PPB2 has determined that the compounds connected with the target in ChEMBL are the ones that are most similar to U5P. In this instance, the server generates analogs through the utilization of a Naive Bayes Machine Learning model, specifically the ECfp4 model, incorporating 2000 of its closest neighbors. For this study, we adopted three distinct methodologies: (a) employing the Extended Connectivity fingerprint (ECfp4) in combination with NN (ECfp4) + NB(ECfp4), (b) using Molecular Quantum Numbers (MQN) along with NN(MQN) + NB(ECfp4), and (c) incorporating Shape and Pharmacophore fingerprint (Xfp) together with NN(Xfp) + NB(ECfp4). The SMILES of the substrate was provided as input which resulted in 20 "targets" with nearest neighbor which were used for virtual screening. Further the SMILES of these compounds were converted to 3D SDF structures using the "SMILES Translator and Structure File Generator" from Online SMILES Translator tool (https://cactus.nci.nih.gov/translate/index.html#Form) (Weininger, 1988). After minimising each of the 3D–SDF files of the compounds, Open–babel was used to convert each of the files to PDBQT (O'Boyle et al., 2011).

2.2. QSAR modelling

Machine learning approaches were applied to build QSAR (Quantitative structure–activity relationship) model for virtual screening of the natural organic molecules. Six regression models were utilised for QSAR which include Linear Regression model, Random Forest regressor, Bayesian Ridge Regression model, Decision tree regressor, Support Vector Regression model, and Gradient Boosting Regression model. The compound library for constructing the models was obtained by employing the ChEMBL database, which can be accessed at (https://www.ebi.ac.uk/chembl/) (Davies et al., 2015). ChEMBL database was searched for "SARS CoV–2″ in all "Targets", from where the top ranked "Target" was retrieved with the highest number of compounds with their IC50. After removing blanks from IC50, compounds with IC50 in units of ug.ml^{-1} (microgram per millilitre) were converted into nM (nano–Molar) using the equation (1).

$$\mathrm{nM} = \left(10^6 * \mathrm{ug.ml}^{-1}\right)/\mathrm{MW} \quad (1)$$

Here, MW is molecular weight. After the conversion of the unit, the $\log_{10}$ of IC50 value was used for QSAR. QSAR models were developed after compound properties were calculated using the RDKit programme (Landrum, 2014). Using the python packages, Morgan Fingerprints were used to construct ML models to predict pIC50 based on physicochemical descriptors. Finally, 70 % of these compounds were used to train the models, while 30 % were used as test compounds. For the purpose of validating the trained models, the coefficient of determination (R^2) of each model was computed, and the model with the highest value was chosen for application in the screening procedure. Since a high value for R^2 indicates a strong correlation between the model's predictions and the data, it is generally agreed that the model is accurate. After selecting the best QSAR model, the analogues of U5P were screened by using the most accurate QSAR models that had been developed to estimate the bioactivity (IC50). In this part of the process, the compounds that demonstrated greater activity in comparison to the control were chosen for additional screening using molecular docking.

2.3. Molecular docking

The AutoDock 4.2 tool (Morris et al., 2009) was used for constructing the grid box including the binding site residues surrounding the uridine–5′–monophosphate (U5P) of the Nsp15 of SARS–CoV–2 with a distance of 6 Å. Here, the grid box dimensions were 20 Å × 20 Å × 20 Å and centered at 91.23 Å, –21.51 Å, –28.3 Å for x, y and z axis, respectively. Hydrogen and Gasteiger charges were introduced into the target protein, and water molecules and heteroatoms were removed, as part of the docking protocol. The molecular docking screening of these compounds was carried out with the help of the programme, AutoDock Vina (Eberhardt et al., 2021). The virtual screening process fixed the values of certain docking parameters, primarily, binding modes set at 20, exhaustiveness set at 10, and a maximum energy difference of 4 kcal/mol. In the AutoDock protocol, the spacing used was 1 Å. Docking was proceeded with the ML–based QSAR model screened compounds along with the control. Based on the binding scores, the compounds were ranked, and the top 10 compounds were analysed for hydrogen bond interactions. Furthermore, the most promising compounds were put through dynamics simulation analysis.

2.4. Interaction analysis

The representation of interactions in 2D and 3D was done using Biovia Discovery Studio and the Protein-Ligand Interaction Profiler (PLIP) service (Adasme et al., 2021; "BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1.0.20298, San Diego: Dassault Systèmes, 2020.," n.d.). The selected compound was also used for ADME-Toxicity analysis to validate the druglike properties of the top hit compound. The ADME-Toxicity properties were calculated using SwissADME server (Daina et al., 2017) and the ProTox-II - Prediction Of Toxicity Of Chemicals server (Banerjee et al., 2018) to study the physico-chemical behaviours.

2.5. Molecular dynamics simulation

We utilised molecular dynamics simulations (MD) to investigate the dynamic stability and adaptability of protein–ligand complexes. The selection of the top hit compounds and a control ligand was based on these simulations. GROMACS 2021.2 and the CHARMM27 force field were used to run MD simulations on all complexes (Berendsen et al., 1995; Hess et al., 2008). The topologies, hit compound parameter files, and standard inhibitor parameter files were all generated using the SwissParam server (Zoete et al., 2011). In addition, electrostatic forces were calculated using the Ewald Particle Mesh method (Darden et al., 1993). Following the neutralisation of the system using Na+ and Cl– ions, the solvation box had simulation using the TIP3P transferable intermolecular potential, which incorporates a three–point (TIP3P) water model (Izadi et al., 2014). The complexes were simulated within a dodecahedron box, with a buffer spacing of 1 Å. The protein–ligand solvated complex underwent energy minimization using the steepest descent (SD) technique for a total of 5000 iterations. After heating the system to 310 K and using the LINCS algorithm to remove all hydrogen bonds, the system is stable (Hess et al., 1997). In order to achieve equilibrium, the system was exposed to a uniform ensemble of constant temperature (NVT) and pressure (NPT) conditions for 1 ns at 310 K and 1 atm, respectively. In this case, the 100 ns production run utilised the equilibrated system. The Parrinello–Rahman pressure method (Parrinello and Rahman, 1981) was used to keep the pressure constant, while the velocity–rescaling method (Bussi et al., 2007) was used to couple the temperature. The analysis of post–MD data was conducted utilising the GROMACS tools to investigate the parameters such as root mean square

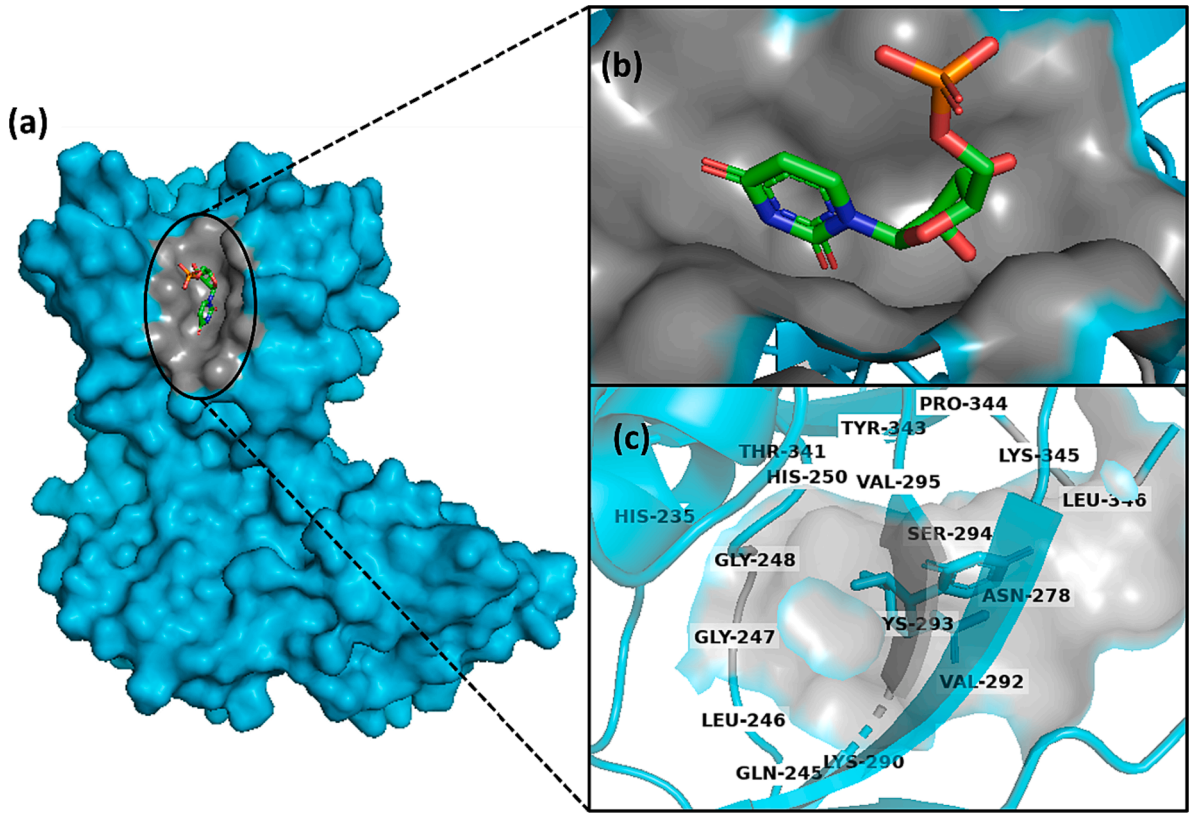

Fig. 2. Binding site residues surrounding Uridine–5′–Monophosphate (U5P) of NSP15 Endoribonuclease (PDB ID: 6WLC) (a) Native protein structure with U5P (b) U5P in the binding pocket of NSP15 (c) Binding site residues surrounding U5P.

deviation (RMSD), root mean square fluctuation (RMSF), and hydrogen bonding.

2.6. PCA: Principal component analysis

PCA was performed using the default settings in GROMACS (Berendsen et al., 1995; Hess et al., 2008)on protein–ligand complexes. The trajectory was preprocessed to remove periodic conditions in preparation for principal component analysis. The covariance matrix for the transformed trajectory was determined with the help of the 'gmx covar' tool in GROMACS. The covariance matrix characterises the relationships between the protein–ligand complex's atomic fluctuations. The 'gmx anaeig' function was used to compute the covariance matrix's eigenvalues and eigenvectors. To visualise the trajectory on the PCs, we used the 'gmx anaproj' GROMACS tool to calculate the PC coordinates for each frame.

2.7. Free energy landscape (FEL)

Protein conformational changes associated with different energy states can be effectively revealed using free energy landscape (FEL) (Maisuradze et al., 2010). By analysing the steady state, represented by the FEL minima, and the transient state, represented by the FEL barriers, the dynamics of biological systems can be studied to gain insight into phenomena such as biomolecule recognition, aggregation, and folding. The FEL was calculated by the following equation:

$$\Delta G(X) = -k\mathrm{B}T\ln P(X)$$

In this scenario, ΔG stands for the change in Gibbs free energy, kB stands for the Boltzmann constant, T stands for the absolute temperature, X corresponds to the reaction coordinate, and P (X) indicates for the system's probability distribution along the reaction coordinate.

2.8. MM/GBSA calculations

Employing the GROMACS plugin known as gmx MMPBSA, we determined the binding free energy of the complex over the last 20 ns of the simulation, utilizing the MM/GBSA (Molecular Mechanics Generalised Born Surface Area) approach. The computation of the binding free energy was carried out via the MM/GBSA method, which utilizes the following equations:

$$\Delta G = G_{complex} - \left[G_{receptor} + G_{ligand}\right] \quad (2)$$

$$\Delta G_{binding} = \Delta H - T\Delta S \quad (3)$$

$$\Delta H = \Delta G_{GAS} + \Delta G_{SOLV} \quad (4)$$

$$\Delta G_{GAS} = \Delta E_{EL} + \Delta E_{VDWAALS} \quad (5)$$

$$\Delta G_{SOLV} = \Delta E_{GB} + \Delta E_{SURF} \quad (6)$$

$$\Delta E_{SURF} = \gamma.\mathrm{SASA} \quad (7)$$

where, $G_{complex}$, $G_{receptor}$ and G_{ligand} are, in the context of Eq (2), the total free energies of the protein–ligand complex, the free form of protein, and the ligand in the solvent, respectively. The other equations in Eqs (3)–(7) showed that the change in solvation free energy, ΔG_{SOLV}, the change in conformational entropy, –TΔS, the change in enthalpy, ΔH, and the change in gas–phase energy, ΔG_{GAS}. SASA represented the Solvent–accessible surface area, while γ represented the surface tension of the solvent. $\Delta E_{VDWAALS}$ and ΔE_{EL} represented the van der Waals and electrostatic energy changes, while ΔE_{GB} and ΔE_{SURF} reflected the polar and nonpolar solvation energy changes, respectively.

3. Results

3.1. Binding site

Nsp15 endoribonuclease (endoU) from SARS CoV–2 was retrieved with the PDB ID: 6WLC which is complexed with uridine–5′–monophosphate (U5P). Here, U5P is the substrate and considered as a control in the study, that included the molecular docking and molecular dynamics simulation protocols. Thus, the grid box for molecular docking involved the surrounding residues of the U5P bound to Nsp15 in the experimental structure. Binding site residues surrounding the control U5P were His235, Gln245, His246, Gly247, Gly248, His250, Asn278, Lys290, Val292, Lys293, Ser294, Val295, Thr341, Tyr343, Pro344, Lys345, and Leu346 as shown in the Fig. 2. Previous study showed that uridine 5′–monophosphate (U5P) bound to the Nsp15 endoU active site had the residues His235, His250, Lys290, Ser294, and Tyr343, which were also found to be part of the binding pocket depicted by the current study (Pillon et al., 2021). Additionally, another study showed the catalytic triad His235, His250, and Lys290 of the Nsp15 were also found to be part of the binding site residues (Ricagno et al., 2006). All the binding site residues including these active site residues and catalytic triad were used for building the grid box for molecular docking.

3.2. Analogues of Uridine–5′–Monophosphate (U5P)

Here, the substrate (Uridine–5′–Monophosphate) was used to generate analogues of the compound and screen them against the protein target. Polypharmacology Browser 2 (PPB2) was used for generating analogues of Uridine–5′–Monophosphate using the three methods. The ECfp4 Naive Bayes Machine Learning model is utilised by the Polypharmacology Browser 2 (PPB2), which provides 2000 of its nearest neighbours.

Smiles of the compound Uridine–5′–Monophosphate (U5P) (C1 = CN (C(=O)NC1 = O)C2C(C(C(O2)COP(=O)(O)O)O)O) was used as input and the output produced 20 targets with the nearest neighbors. PPB2 determined that the molecules associated with the target in ChEMBL were the most similar to U5P, and it selected those compounds for further process. Three methods, (a) Extended Connectivity fingerprint ECfp4 NN(ECfp4) + NB(ECfp4), (b) Molecular Quantum Numbers MQN NN(MQN) + NB(ECfp4), (c) Shape and Pharmacophore fingerprint Xfp NN(Xfp) + NB(ECfp4) were used in this study. Extended Connectivity fingerprint ECfp4 NN(ECfp4) + NB(ECfp4) generated 726 compounds, Shape and Pharmacophore fingerprint Xfp NN(Xfp) + NB(ECfp4) generated 509 compounds and Molecular Quantum Numbers MQN NN (MQN) + NB(ECfp4) generated 459 compounds. A total of 1694 analogues were generated. After removing duplicates and blanks in SMILES of analogues, 816 unique compounds remained. These analogues were used for first phase of screening using the ML–based QSAR model. A similar study showed the use of Polypharmacology Browser 2 for generating target–associated ChEMBL compounds as the closest analogues of triazines for identifying the target Lysophosphatidic Acid Acyltransferase β (LPAAT–β) (Poirier et al., 2019). Few other studies also showed the successful use of the Polypharmacology Browser 2 server (Montaruli et al., 2019; Reymond, 2022).

3.3. ML–based QSAR model

The ML–based QSAR model was performed using six regression models, Linear Regression model, Random Forest regressor, Bayesian Ridge Regression model, Decision tree regressor, Support Vector Regression model, and Gradient Boosting Regression model. Here, for construction the QSAR model, the database of known inhibitors was searched in the ChEMBL database. The "SARS CoV–2″ in all "Targets" were searched which returned 21 target ids. By sorting according to activity, target "CHEMBL4303835" had the highest compound activity with 37,665 compounds, thus was selected. In target "CHEMBL4303835", 10,086 compounds were found with "IC50" which was used further in the building the QSAR model. SMILES of the compounds were stored, while their IC50 in ug.ml^{-1} were converted into nM. After converting the unit, the $\log_{10}$ of IC50 value was used for QSAR. Here, 30 % of these compounds were used as a test set and the remaining 70 % were used to train the models. The coefficient of determination (R^2) was calculated for the validation of the trained models as shown in the Table 1. It was observed that the Random Forest regressor model was predicted to have the highest value possible, which was 0.72. This demonstrates that there is a significant correlation between the values that were predicted and those that were actually observed; consequently, the model is considered to be a good fit to the data. In addition, it was previously established that those models with an $R^2 > 0.6$ were deemed to be acceptable (Guerra et al., 2016). The best predicted QSAR model (Random Forest regressor model) was used used to perform QSAR for 816 and control to determine activity. As indicated in the supplementary Table S1, the QSAR analysis demonstrated that 397 of the compounds exhibited higher levels of bioactivity when compared to the control molecule. These 397 compounds, in addition to the control compound, were chosen for further molecular docking investigation. Previous studies virtually screened a vast array of compounds utilising QSAR models that were comparable to the models used in this work (Bommu et al., 2019; Fan et al., 2019; Rudrapal and Chetia, 2020). In a separate investigation, the application of QSAR–based virtual screening led to the discovery of a hit molecule demonstrating efficacy against SARS 3CLpro. (Jawarkar et al., 2022). In a separate study,

Table 1
Correlation Co-efficient and Coefficient of determination (R^2) of the six models from ML–based QSAR.

Regression models	R^2	Correlation Co-efficient
Random Forest	0.72	0.84
Bayesian Ridge	0.56	0.74
Linear Regression	0.21	0.45
Decision tree	0.07	0.26
Support Vector	0.16	0.4
Gradient Boosting	0.16	0.4

Table 2
The docking scores and the hydrogen bonds of the top 10 hits after virtual screening with the target protein Nsp15 along with the control.

Ligand	Docking scores (exhaustiveness = 10)	Docking scores (exhaustiveness = 100)	H–Bond
control_U5P_6030	–5.9	–6.0	GLY248, HIS250, LYS290, THR341, TYR343
lig165	–8.1	–8.1	NA
lig302	–7.9	–7.9	TRP333, TYR343
lig149	–7.8	–7.9	NA
lig200	–7.7	–7.9	GLY248, THR341
lig27	–7.6	–7.8	THR341, GLN245, SER294
lig32	–7.6	–7.6	GLY248, SER294
lig46	–7.5	–7.7	ASP240, GLN245
lig106	–7.5	–7.6	NA
lig103	–7.5	–7.7	GLY248, LYS290, TYR343
lig118	–7.4	–7.5	NA

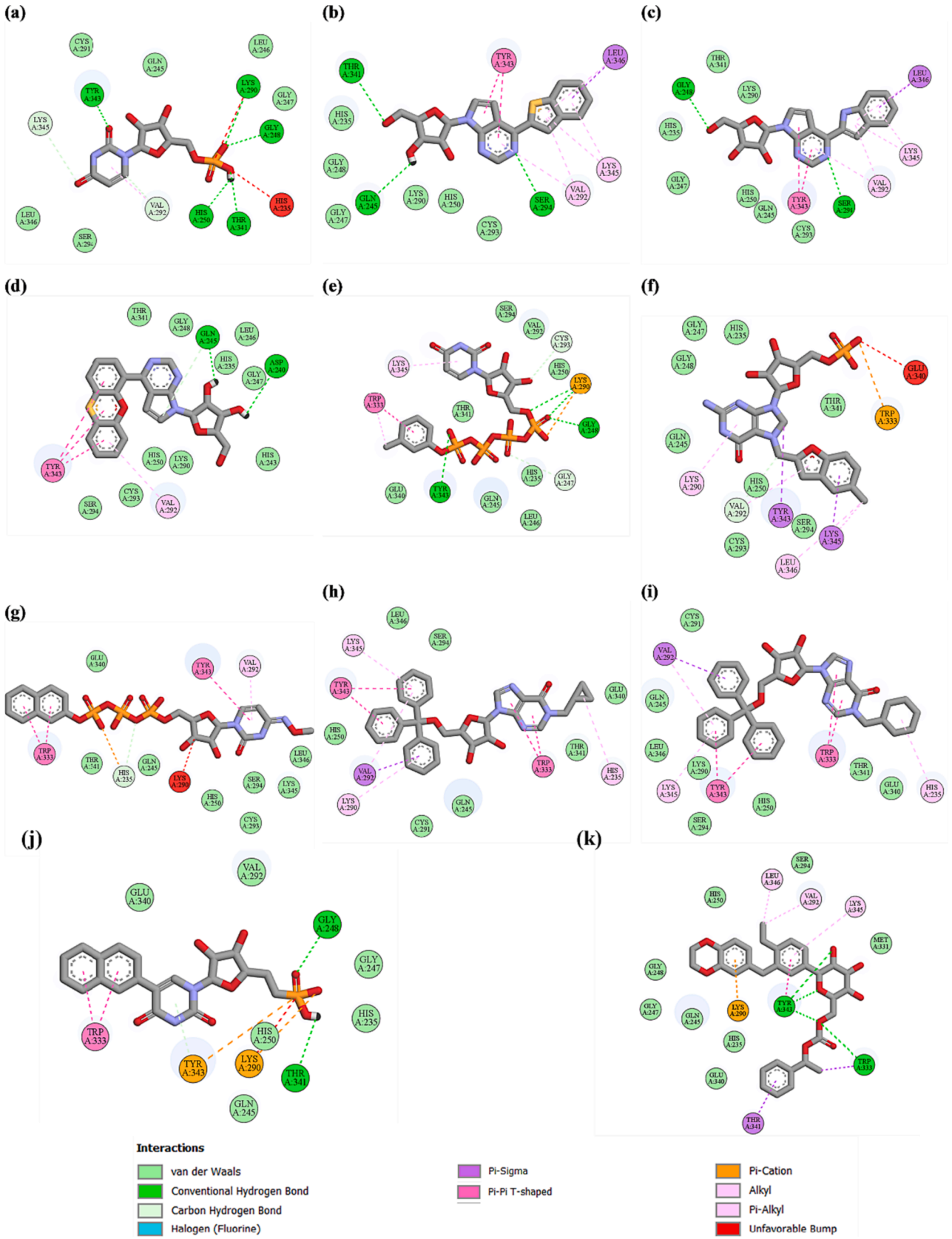

Fig. 3. 2D interactions of the protein–ligand complexes for (a) control, (b) lig27, (c) lig32, (d) lig46, (e) lig103, (f) lig106, (g) lig118, (h) lig149, (i) lig165, (j) lig200, and (k) lig302.

the inhibitory activity of 52 different fenarimol analogues against *Trypanosoma cruzi* was analysed using QSAR model (Cotuá et al., 2021). These studies advocated the application of QSAR in drug designing tasks for developing antivirals.

3.4. Molecular docking

The screened QSAR based compounds were further screened using molecular docking against the target protein Nsp15 along with the control U5P. Further, the docked structure of Nsp15 with U5P has been aligned with the native structure of protein with U5P (PDB ID: 6WLC). It was observed that the alignment showed no change of conformation of the docked structure to that of the native as shown in supplementary Figure S1. There was minimum deviation from the native structure with RMSD of 0.28 Å. Here, a total of 397 analogues of U5P were docked and it was found that 236 compounds interacted better than control. Moreover, the top 10 hits were used for further investigation of the hydrogen bonds interactions and compared to the control. Table 2 showed the top 10 hit compounds with their best docking scores and the hydrogen bond interactions. It was observed that the top 10 hits had the docking scores in the range −7.4 kcal/mol to −8.1 kcal/mol which were better than the control (−5.9 kcal/mol). Moreover, the docking process was carried out with a higher exhaustiveness setting (100), and the obtained results were found to be in line with the earlier findings. This indicates that variations in exhaustiveness did not yield any notable differentiation among the ligands. Previous studies also showed the docking scores of −6.9 kcal/mol for EGCG which was docked against the Nsp15 of Sars–CoV–2(Hong et al., 2021). Another study showed the docking scores range from –5.1 kcal/mol to –9.8 kcal/mol from docking in Vina for the FDA approved compounds against Nsp15(Chandra et al., 2021). Saikosaponin V and Saikosaponin U which showed high binding affinity for Nsp15 of SARS–CoV–2 had docking scores of −8.35 Kcal/mol and −7.27 kcal/mol, respectively (Sinha et al., 2020). The top 10 hits in this current study also showed comparable binding energy as the hits from the earlier studies against nsp15 of SARS–CoV–2.

The 2D interactions of the protein–ligand complexes were visualized using the Biovia Discovery Studio (Biovia, 2019) as shown in the Fig. 3. The structural stability of the protein–ligand complex is greatly affected by hydrogen bonding, making it an important factor in drug development (Cichero et al., 2021). Compared to the control which showed hydrogen bonds with the residues Gly248, His250, Lys290, Thr341, and Tyr343, the compounds lig103 (PubChem CID – 53309102) and lig200 (PubChem CID – 57398422) showed similar hydrogen bonds. The compound **57398422** showed hydrogen bonds with residues Gly248, Thr341 while **53309102** showed hydrogen bonds with Gly248, Lys290, Tyr343 which were found in the control. These promising compounds were selected for further analysis using MD. Moreover, the compound lig27 (PubChem CID – 76314921) also showed three hydrogen bonds (Thr341, Gln245, Ser294), where one (Thr341) was found in the control. Thus, **76314921** was also selected as it had maximum number of hydrogen bonds and showed interaction with key residue. All the other compounds were not as promising as the three selected compounds as they showed less hydrogen bonds.

Numerous prior investigations have delved into the intricate interactions between the identified compounds and specific amino acid residues, namely Gln245, Gly248, Lys290, Ser294, Thr341, and Tyr343. These insights were gleaned from the collective findings of three distinct studies (Godoy et al., 2023; Hong et al., 2021; Sinha et al., 2020). In a distinct analysis, researchers explored the binding of uridine 5′–monophosphate (5′–UMP) to the active site of Nsp15, elucidating the formation of hydrogen bonds with residues His235, His250, Lys290, Ser294, and Tyr343 (Pillon et al., 2021). Additionally, a separate investigation unveiled the pivotal role of the residues His235, His250, and Lys290 within Nsp15, constituting a crucial His–His–Lys triad responsible for its endoribonuclease activity, as outlined by (Ricagno et al., 2006). Furthermore, various other residues within the active sites were found to contribute to oligomer formation and enzymatic activity, as corroborated by the works of many scientists (Bhardwaj et al., 2008; Kim et al., 2020b; Zhang et al., 2018). Consequently, the robust binding affinity exhibited by the three compounds substantiates their selection for further in–depth scrutiny.

Table 3
The bioactivity using the ADMET properties study of the top three hit compounds.

Molecule	53309102	57398422	76314921
MW	674.66	448.36	383.42
Rotatable bonds	12	5	3
H-bond acceptors	18	8	6
H-bond donors	7	5	3
MR	127.77	111.84	101.29
TPSA	339.14	171.89	128.87
iLOGP	−0.21	0.8	2.23
Predicted Toxicity Class	6	6	2
Solubility	Very soluble	Soluble	Soluble
PAINS alerts	0	0	0

3.5. ADMET properties

In order to evaluate the drug's ADME (absorption, distribution, metabolism, and excretion) properties, as well as its possible toxicological impact on biological systems, an analysis was conducted to identify the ADME-toxicity features of the selected compounds (53309102, 57398422 and 76314921). The anticipated toxicity class for the selected compounds 53309102, and 57398422 listed in Table 2, was considered acceptable by the Hazard Communication Standard (HCS) with toxicity class 6. The GHS (Globally Harmonised System of Classification and Labelling of Chemicals) provides the foundation for this toxicity class. However, 76314921 had high toxicity level which was not acceptable. The compounds 57398422 and 76314921 was found to have significant molecular weights (MW) that was less than 500 Da and there were less than 10 hydrogen bond acceptors and less than 5 hydrogen bond donors. This showed acceptable druglike property of the compounds. The number of rotatable bonds was 3–12 for the compounds, signifying good flexibility of the compound. It was observed that the compound was soluble in water, as shown in Table 3, while GI absorption was lower. The molar refractivity (MR) value in the range 101.29 to 127.77 indicates that the compound has a moderate to high molecular volume and polarizability. The topological polar surface area (TPSA) value of 339.14 suggests that the compound 53309102 has a relatively higher polar surface area, which can also influence its ability to cross biological membranes and interact with biological targets. For drug-like molecules, an iLOGP between 1 and 5 was considered suitable, thus these compounds 90470472 and 74977521 were favorable. It was observed that there was 0 alert in PAINS (pan assay interference compounds) for all the compounds, suggesting promising outcomes in in vitro and in vivo studies. Overall, the three compounds showed promising ADMET properties and was selected for further analysis using MD simulation.

3.6. Molecular dynamics simulation

Molecular Dynamics (MD) Simulations of the protein–ligand complexes were performed to verify the stability and adaptability of the analogues of U5P with the protein Nsp15. The extracted complexes' dynamic trajectories were thoroughly explored in a computer simulation. The root–mean–square deviation was one of the metrics used in this study (RMSD). In order to measure the complex's stability and convergence during an MD simulation, RMSD calculation was employed. After an MD simulation, the RMSD, RMSF, and Hydrogen bonds of the protein–ligand complex reveal interesting information about the system's dynamics over time. The protonation state of the titrable amino acids

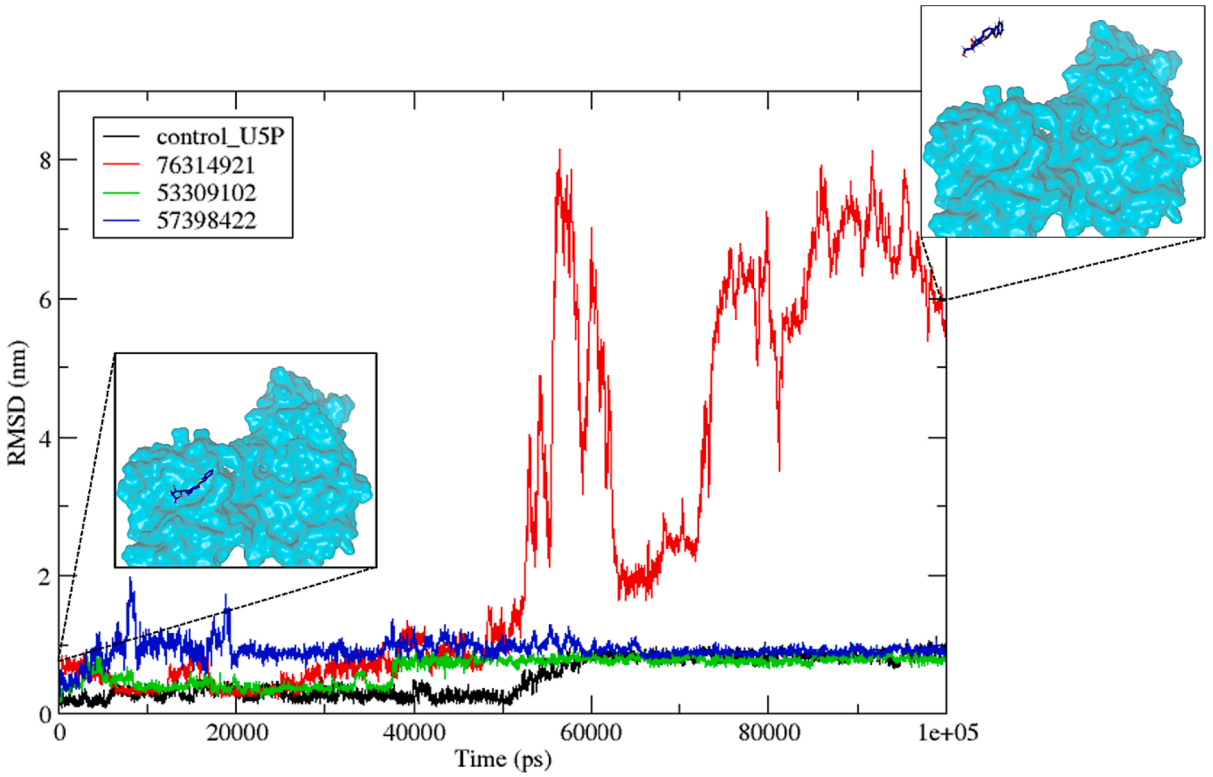

Fig. 4. RMSD of the ligands 53309102, 57398422, and 76314921 along with control (U5P) when bound to the protein Nsp15 of SARS–CoV–2.

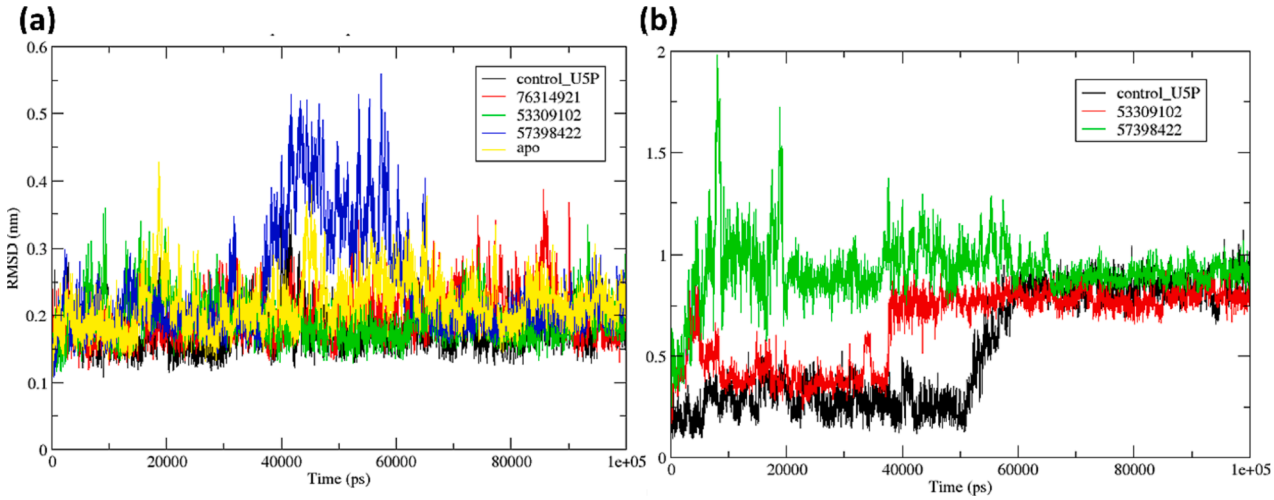

Fig. 5. RMSD of the protein and ligand complexes for the compounds 53309102 and 57398422, in addition to the control (U5P) (a) RMSD of protein C alpha atoms aligned over the original structure in bound and unbound form; (b) RMSD of the ligands.

were calculated for the protein. The pKa values for these amino acids are listed in the Supplementary Table S2. The isoelectric point (pI) of the system was 4.72 and at pH 7.4, the total charge on the system was −12. As observe in Supplementary Table S1, positively charged residues with higher pKa than the physiological pH were protonated during the simulation, while negatively charged residues that showed a lower pKa than the set pH, were deprotonated. This accurately model the ionization states of the amino acid residues in the system at the specified pH conditions. Further the protonation state of the compounds were also calculate as shown in the Figure S2.

3.7. Trajectory analysis

The RMSD of the ligands 53309102, 57398422, and 76,314,921 along with control (U5P) when bound to the protein Nsp15 of SARS–CoV–2 as shown in Fig. 4. Here, the RMSD of the ligand was calculated by first aligning the protein structure, and then calculating the deviation that resulted from that alignment. It was observed that the compound 76314921 showed high RMSD post 50 ns of the simulation shown in Fig. 4. The conformation of the protein–ligand complex for 76314921 at 0 ns and 100 ns was observed. It was found that the 76314921 was stable at the 0 ns, however it moved out of the protein binding site at the 100 ns, showing no binding affinity with the protein. Thus, 76314921 was not selected for further investigation.

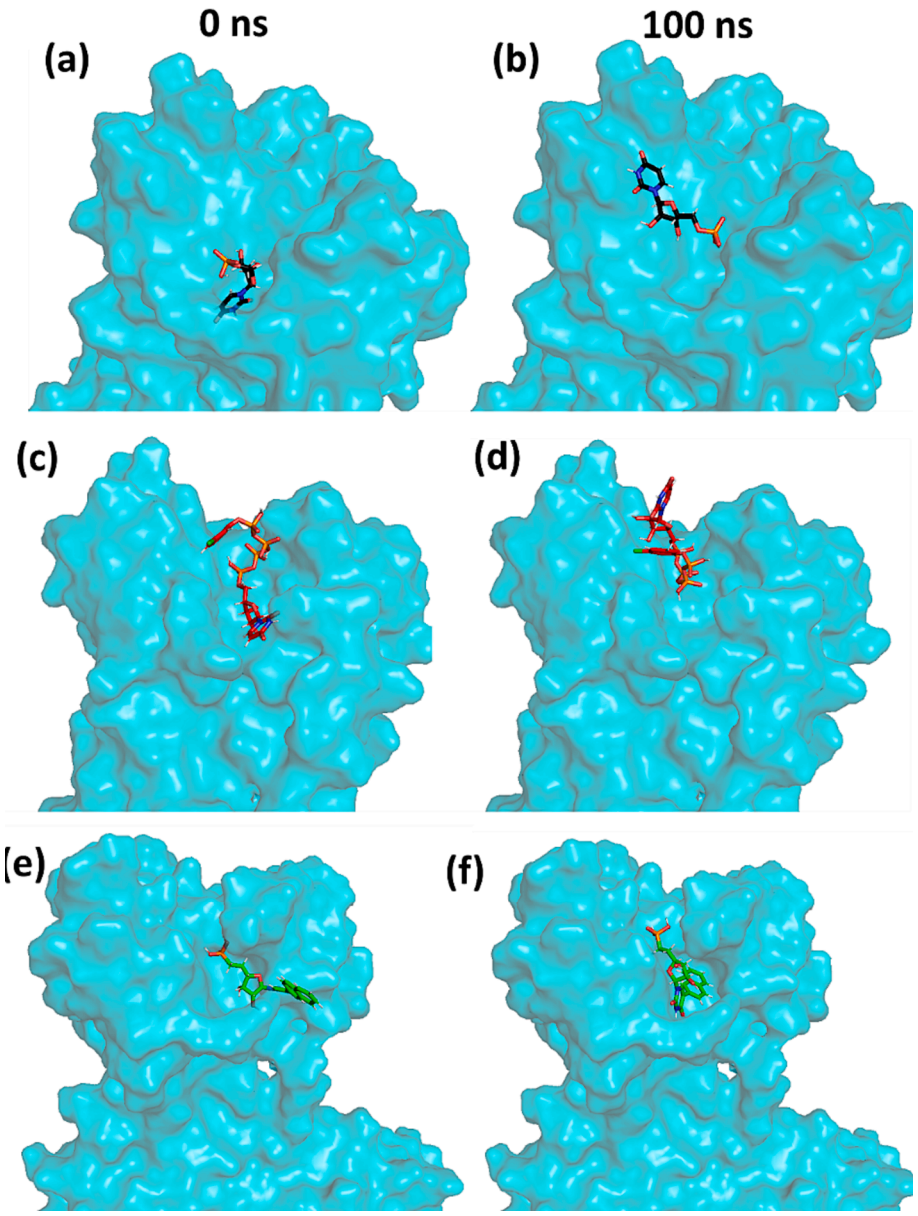

Fig. 6. Conformational structure of the protein–ligand complexes at 0 ns and 100 ns of the MD simulation for (a) Control (U5P), (b) 53309102, and (c) 57398422.

3.8. RMSD

The root–mean–square deviation (RMSD) was calculated in this study by first aligning each frame of the trajectory with the protein's initial equilibrated structure. RMSD of the protein was calculated by aligning the Cα–atoms aligned over the initial structure. The RMSD of the protein shown in the Fig. 5(a) depicted that the conformation of the protein bound to U5P and 53309102 was stable for the entire 100 ns simulation. Protein bound to 57398422 showed deviations between the trajectories of 40 ns to 60 ns simulation, however the RMSD became stable post 60 ns. RMSD of 0.2 nm to 0.3 nm was observed for protein bound to the two compounds, along with the control ligand. The unbound (apo) protein showed identical RMSD pattern with no major fluctuations. It showed RMSD in the 0.2 nm to 0.3 nm which was stable and consistent. Similar RMSD trend was observed by a previous study for NSP15-ZINC000104379474 complex, which showed promising results against SARS-CoV-2(Ibrahim et al., 2022). RMSD values within this range represent a relatively stable structure, with minor fluctuations and deviations from the starting conformation. The protein is likely undergoing some conformational changes when bound to 57398422, such as local fluctuations or small rearrangements, but it remains overall stable.

RMSD of the ligands was shown in Fig. 5(b) for the compounds 53309102, and 57398422 along with the control (U5P) when bound to the Nsp15 protein. U5P showed the lowest deviation with RMSD of <0.5 nm for the first 50 ns. However, it increased to 1 nm and became stable for the rest of the 50 ns simulation. Similar trend was observed for 53309102, with RMSD of <0.5 nm for the first 35 ns and it increased 0.7 nm where it became stable for the rest of the simulation. Contrastingly, 57398422 showed high fluctuations in RMSD for the first 50 ns simulation, however post 50 ns, it became stable with RMSD of 1 nm similar to the control. It suggested that the ligands shifted from the initial binding site of the protein, however, the ligand was still bound to protein and was stable. Furthermore, the conformation at 0 ns and 100 ns of the complexes was investigated.

In addition, multiple molecular dynamics (MD) simulation replicates were conducted following a 10 ns equilibration period. Supplementary Figure S3 illustrates the RMSD patterns observed in these replicates,

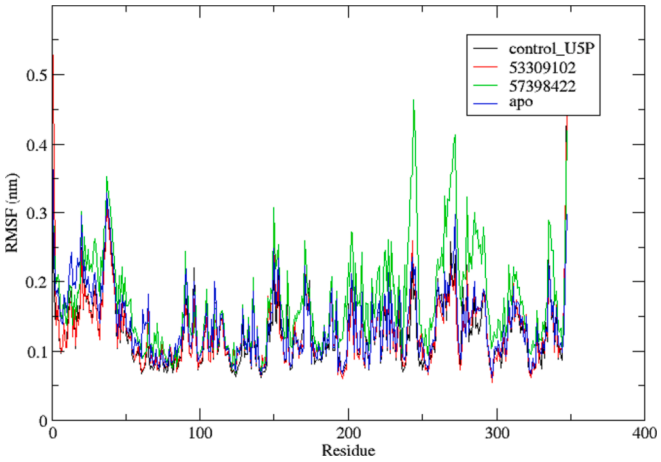

Fig. 7. RMSF of the protein during the 100 ns simulation residues for the unbound (apo protein) and protein–ligand complexes of Control (U5P), 53309102, and 57398422.

demonstrating a consistent trend. Notably, the RMSD values for the protein remained stable throughout both the initial and subsequent MD runs, consistently fluctuating within the narrow range of 0.2 to 0.3 nm. Moreover, the RMSD values for the ligand displayed a similar pattern in both runs, indicating reproducibility and reliability of the MD simulation protocol. It's noteworthy that the ligand with the identifier 76314921 exhibited a notably high RMSD, suggesting its displacement from the protein's binding site, which was also observed in the 1st run. This observation further validates the robustness and accuracy of the MD simulation procedure employed in this study.

3.9. Conformational change

The conformational change of the protein–ligand complex during the 100 ns simulation was illustrated in the Fig. 6. The change of ligand position when bound to the protein at 0 ns and 100 ns simulation was compared. It was observed that the control showed conformation change from its original binding pocket and moved to surface of the protein as shown in the Fig. 6(a, b). However, it was still bound to the protein near the binding pocket as also depicted by the RMSD trend. 53309102 showed similar RMSD trend as the control, it slightly shifted from its binding pocket to the nearby pocket of the protein as shown in the Fig. 6(c, d). 57398422 changed its conformation slightly and moved in the inner side of the pocket shown in Fig. 6(e, f). This indicated that both 53309102 and 57398422 was strongly bonded to the protein during the 100 ns simulation which was also observed in the RMSD plot shown in Fig. 5(b). The conformational changes were due to the formation of more stable bonds as shown further in the hydrogen bond prediction.

3.10. RMSF

The RMSF of the protein was calculated over the course of 100 ns simulation for the control (U5P), 53309102, and 57398422 along with the control. The Fig. 7 showed that majority of the residues of the protein bound to the ligands had a stable RMSF of <0.3 nm. However, few residues showed high fluctuations with few peaks. The residue Asp37 showed high fluctuations over 0.3 nm for both of the compounds (53309102, and 57398422) along with the control (U5P). Interestingly, this residue also showed RMSF >0.3 nm for the apo protein. These findings underscore the dynamic nature of Asp37 within the protein structure and its potential significance in ligand binding and structural stability. However, 57,398,422 showed high fluctuations (RMSF > 0.3 nm) for the 18 residues. The study showed similar RMSF trend to that of a prior study targeting NSP15 (Ibrahim et al., 2022). These residues exhibited significant fluctuations during the simulation could be the cause of ligand binding which induced conformational changes and increased flexibility in the binding site or nearby regions of the protein.

3.11. Radius of gyration and COM (Center of Mass)

The radius of gyration (Rg) is a measure of the protein's compactness as shown in the Fig. 8(a). A higher Rg indicates a more extended conformation, while a lower Rg suggests a more compact structure. The control protein–ligand complex (black line) maintains a relatively consistent Rg throughout the simulation, suggesting a stable conformation. The protein–ligand complex with compound 53309102 (green line) exhibits a higher variability in its Rg values, particularly at the beginning and end of the simulation. This indicates potential conformational changes or fluctuations during these periods. The protein-57398422 complex (red line) has a similar trend to the 53309102, though it seems to be more stable in the middle of the simulation. The apo state of the protein (blue line) without the ligand remains relatively consistent but slightly more compact compared to the other complexes. Similar results were observed by the prior investigated protein–ligand complex (NSP15-ZINC000104379474 complex) (Ibrahim et al., 2022). The protein–ligand complexes exhibit varying degrees of stability, with the control appearing the most stable in terms of Rg. The apo state indicates that the absence of a ligand leads to a marginally more compact protein structure.

The Fig. 8(b) plot signifies the average distance between the center of

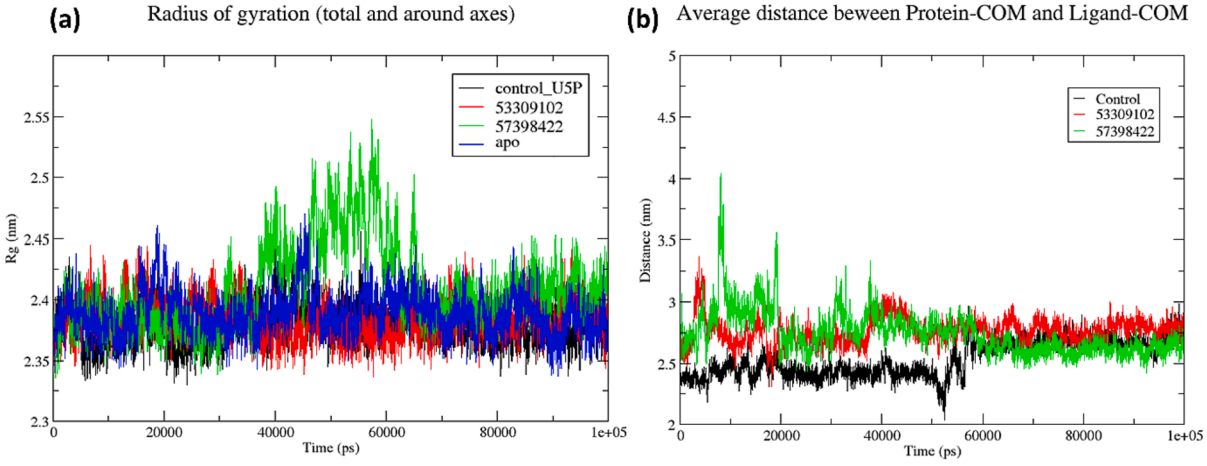

Fig. 8. (a) Radius of gyration (b) Distance between the centre of mass (COM) of the protein and the ligand.

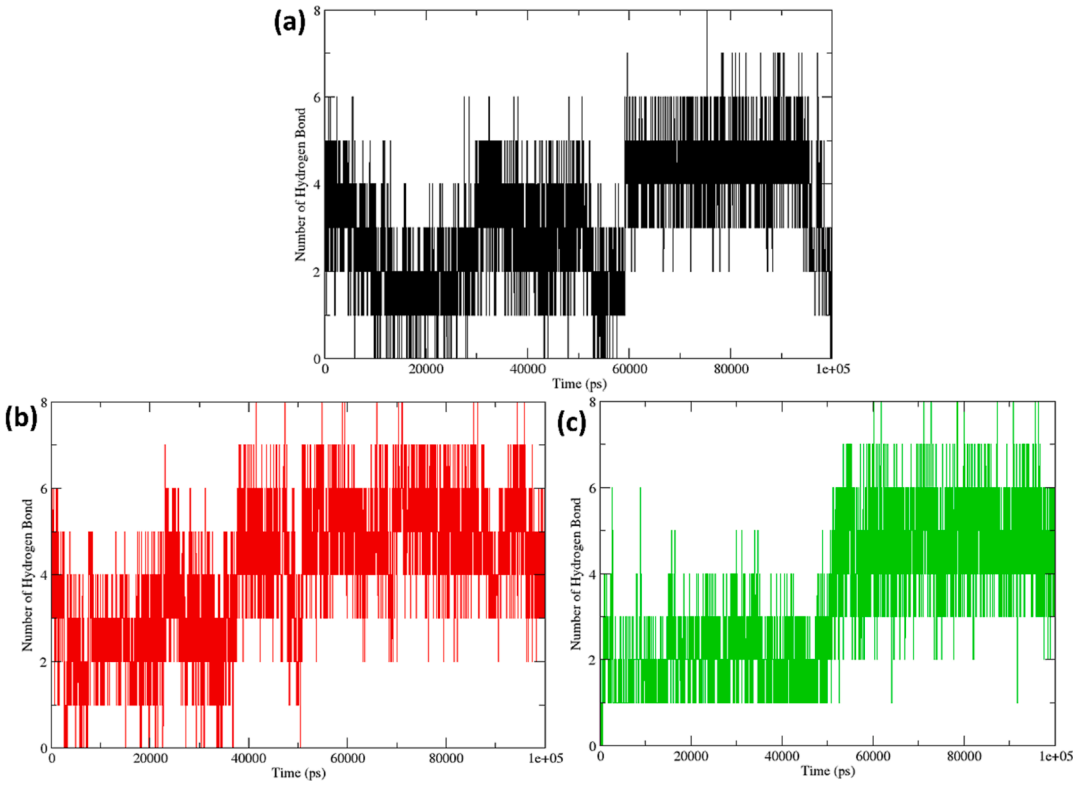

Fig. 9. Hydrogen bond plots for the 100 ns simulation for the protein–ligand complexes of (a) Control, (b) 53309102, and (c) 57398422.

mass (COM) of the protein and the ligand. A shorter distance indicates a tighter binding, while a longer distance can imply weaker interactions or potential dissociation events. The control-protein complex (black line) maintains a relatively stable distance between the protein and ligand COM, suggesting consistent interactions throughout the simulation. The protein-53309102 complex (red line) displays a more fluctuating pattern, with several peaks indicating potential periods of reduced binding affinity or minor dissociation events. The protein-57398422 complex (green line) exhibits a similar trend to 53309102, though with slightly lesser fluctuations. The control-protein complex seems to have the most consistent protein–ligand interactions throughout the simulation. Both 53309102 and 57398422 complexes show variable binding patterns, indicating potential conformational changes affecting ligand binding.

3.12. Hydrogen bond

The formation of folded proteins is dependent on the stability that is provided by hydrogen bonds formed between the protein and the ligand. The hydrogen bonds formed by the protein–ligand complexes during the 100 ns simulation were plotted in the Fig. 9. It was observed that the control and the hit compounds showed similar hydrogen bond trend. The number of hydrogen bonds were 2–4 for the first half of the 100 ns simulation for all the protein–ligand complexes. Later the number of hydrogen bonds increased to 5–6 for all the complexes as shown in the Fig. 9. Few frames showed 7–8 hydrogen bonds during the 100 ns simulation thus, showing strong binding with the protein. Further, the hydrogen bond interactions with the significant key residues were depicted for the three complexes with the compounds control (U5P), 53309102, and 57398422.

3.13. Interaction analysis

In most cases, the interaction between a ligand and a protein is mediated by hydrogen bonding between the two molecules. It's possible that the building blocks of proteins, the essential amino acids, have an important part to play in these correlations. When a ligand binds to a particular amino acid in the active site or function of a protein, it is possible for the ligand to have opposite effects on the protein's activity as a result of the interaction. Thus, the binding of the ligands with the key residues are important for the inhibiting the proteins' function. Here, post MD simulation, the complexes were analysed for the intermolecular interactions in 2D and 3D as shown in the Fig. 10. The control showed a single conventional hydrogen bond with Tyr343 and a single pi–cation bond with Lys290. It also showed multiple van der waals contact with Glu340 and Thr341 and Met331 while it also showed an unfavourable bump with the residue Trp333. 57398422 showed multiple conventional hydrogen bond with Gln347, Tyr343, Leu346 and carbon hydrogen bond with residues Gln245 and Lys345. Additionally, it also showed van der waals contacts with residues Gly247, Gly248, Leu246, His250, Lys290, Cys293, Ser294, Val292, His243, Thr341 and also pi–cation bonds with Glu340 and Trp333. Similarly, 53309102 showed multiple conventional hydrogen bonds with residues Gly248, Thr341, His235, Tyr343, His250 and multiple carbon hydrogen bonds with Gly247. It also showed van der waals contacts with residues Leu249, His243, Glu340, Lys290, Met331, Lys345 and Trp333. Based on the fact that His235, His250, and Lys290 are arranged in a manner that is strikingly similar to that of the active site of ribonuclease A, it has been suggested that these three residues make up the catalytic triad (Ricagno et al., 2006). The active site has carries six key residues conserved among SARS–CoV–2 proteins including His235, His250, Lys290, Thr341, Tyr343, and Ser294 (Kim et al., 2020b). Oligomer formation and enzymatic activity are both influenced by these residues that are located at the active site (Bhardwaj et al., 2008; Kim et al., 2020b; Zhang et al., 2018). Interestingly, it was observed that 53309102 and 57398422 showed interactions with these catalytic site residues His235, His250 and Lys290 post MD simulation, indication significant binding to the functional site of the protein. Additionally, 53309102 and 57398422 showed interactions with the residues Thr341, Tyr343, and Ser294 which are key residues in the active site. Previous also showed the

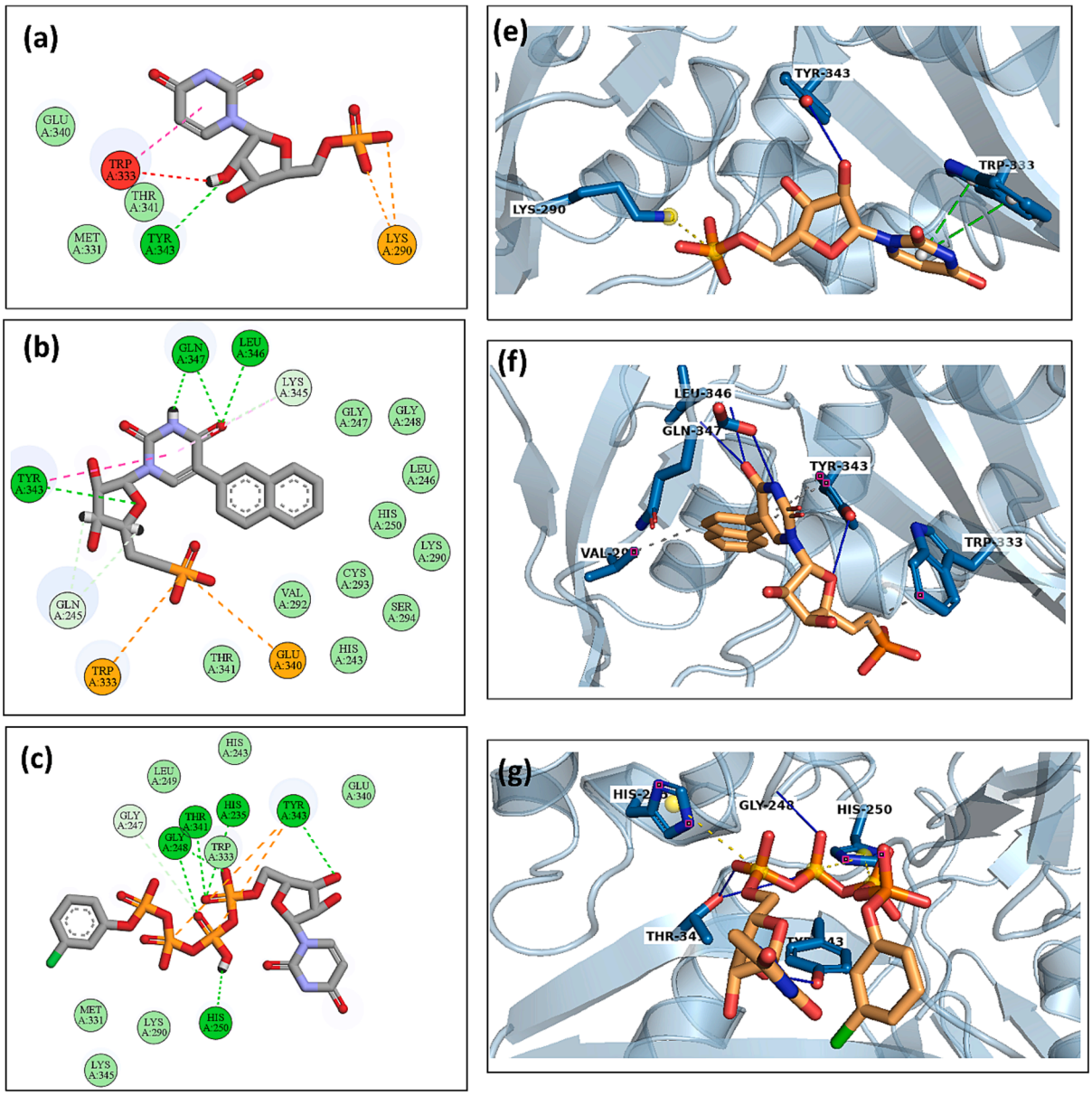

Fig. 10. 2D and 3D Intermolecular interactions between protein–ligand complexes with the compounds (a, e) Control, (b, f) 57398422, and (c, g) 53309102.

inhibitor Saikosaponin V formed bonds with Thr341, Ser294, Trp333, Tyr343, Val292, Glu340, and Lys345 amino acids which were also observed in this current study for the compounds 53309102 and 57398422 (Sinha et al., 2020). Overall, both 53309102 and 57398422 showed hydrogen bonds with important residues indicating strong binding affinity with the protein.

3.14. Principal component analysis (PCA)

Principal component analysis is a popular dimensionality reduction method for extracting important observations from large datasets. The movement of the control, 53309102, and 57398422 complexes along two main components (PC1 and PC2) were studied using PCA. Fig. 11 showed the PCA plots of the complexes. Here, all the complexes had a similar range of motion modes along primary components. Moreover, complexes showed low dispersion indicating less conformational variation. Stable complex formation was indicated by the fact that binding of a compound restricted the free motion of the corresponding complex. Compared to the control, 53309102 showed identical dispersion while 57398422 slightly more dispersion. The free energy landscape analyses for 100 ns MD simulation for the control, 53309102, and 57398422 system were performed to study the energy conformations.

3.15. Freeenergy landscape (FEL)

The free–energy landscape (FEL) is an essential tool in studying the dynamic behaviour of intricate systems in molecular dynamics simulations. Within this particular context, the Free Energy Landscape (FEL) is denoted by two variables, namely PC1 and PC2, which appropriately encapsulate the system's conformational motion. PC1 and PC2 represent distinct system features and are crucial in determining the stability of the complex. In doing an analysis of the FEL, the regions that are represented in the colour blue provide significant and meaningful perspectives. Significantly, the dimensions and configuration of these blue patches can yield insights into the stability of various components within the complex. The smaller and more concentrated blue zones are indicative of very stable parts within the complex. The notion is further elucidated by the three–dimensional projections of the FEL, which

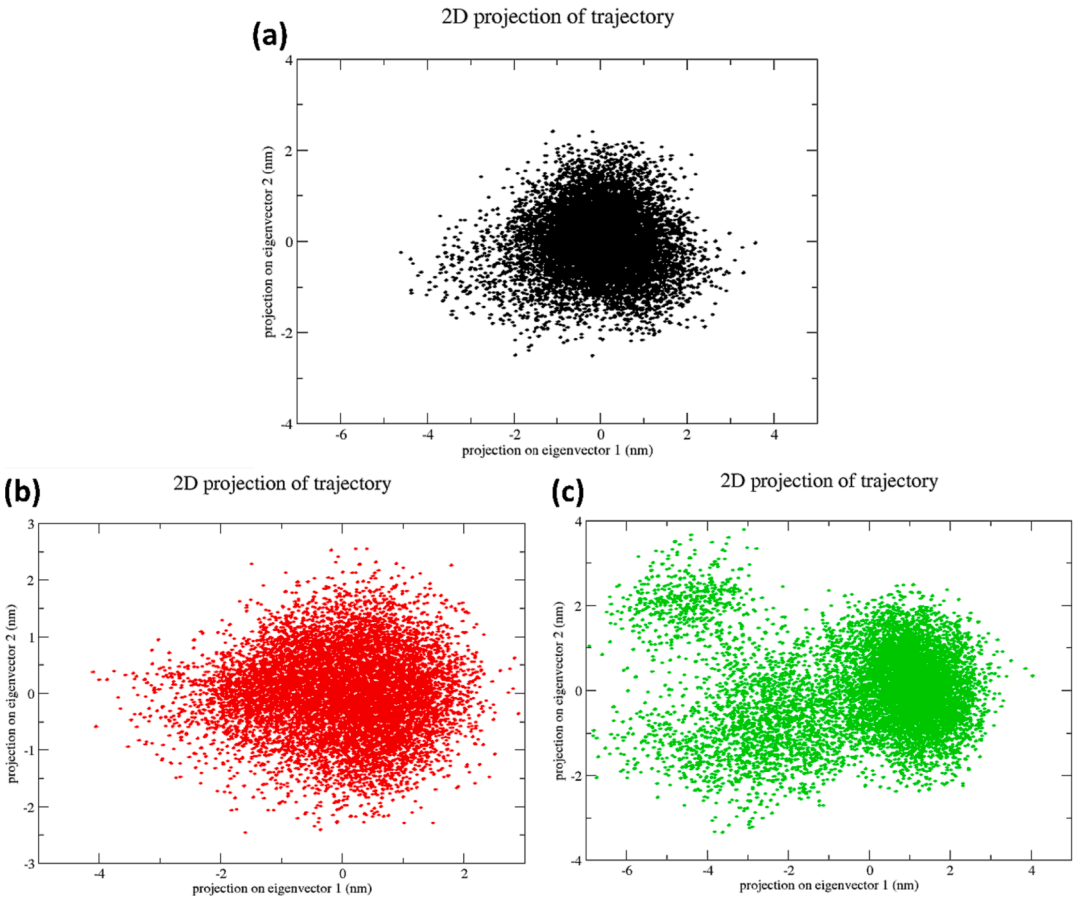

Fig. 11. Principal component analysis of the two primary components (PC1 and PC1) for the complexes of (a) Control, (b) 53309102, and (c) 57398422.

visually demonstrate the emergence of a narrow funnel shape. The shown funnel illustrates the dynamic progression of conformational alterations that unfold during the simulation, culminating in the development of a structure characterised by low energy (Baidya et al., 2022). write the following paragraph in man writing phrase to remove plagiarism (Kannan and Kolandaivel, 2018).

As shown in the Fig. 12, the complexes formed a single funnel, which indicated that they have one local minima. Each of the complexes displayed local energy minima across a more extensive free energy landscape, denoted by a vivid violet hue. These pronounced basins within the graph signify that the complexes have reached a state of minimal energy within their respective structures. Thus, all had stable complex formation. Compared to the control, 53309102 showed identical free energy landscape plot, while 57398422 another small basin with centralized blue areas but was comparable to the control.

3.16. Binding free energy

The evaluation of the binding free energy of the protein–ligand complex was conducted using the MM/GBSA approach. The present study was performed using trajectory data acquired from the 100 ns simulation. Total solvation energy of the system of the compounds 53309102 and 57398422 was positive but the higher negative value of gas phase energy was responsible for average total binding free energy as shown in Fig. 13. However, the total solvation energy of the system of the control showed positive gas phase energy while the negative total solvation energy was responsible for the average total binding free energy. Higher negative value of total binding free energy implies more stability in the interaction of the complex. The cumulative binding free energy (ΔG_{TOTAL}) of the control group was observed as −17.8 kcal/mol. Compounds 53309102 and 57398422 exhibited cumulative binding free energies (ΔG_{TOTAL}) of −29.2 kcal/mol and –23.7 kcal/mol, respectively. The data clearly demonstrated that the hit compounds displayed lower binding free energies (ΔG_{TOTAL}) in comparison to the control, suggesting a stronger and more persistent binding association with the protein. This observation indicates a positive and long-lasting interaction between these chemicals and the protein.

4. Discussion

The findings reported in the present study, provide valuable insights into the interaction between SARS–CoV–2 Nsp15 and its ligands, with a focus on uridine–5′–monophosphate (U5P) and its analogues. We employed a multi–step approach, including analogues generation, machine learning–based QSAR modeling, molecular docking, molecular dynamics simulation, FEP, and free–binding energy analyses to assess the stability and binding of these compounds to the target protein.

First, the Nsp15 binding site was characterised, emphasising critical residues crucial for ligand interactions, some of which overlap with the previously discovered catalytic triad. Following the creation of U5P analogues, the study evaluated these compounds using machine learning–based QSAR models, selecting 397 candidates for additional research. Further molecular docking was used to screening these compounds. Three compounds (53309102, 57398422, and 76314921) stood out because of their potent and reliable interactions with Nsp15. In order to qualify as potential drug candidates, drug molecules must adhere to defined parameters, even though occasional deviations may occur (Van De Waterbeemd and Gifford, 2003). Consequently, the compounds underwent rigorous evaluation for Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) properties, which revealed their suitability for pharmacological applications.

The root–mean–square deviation (RMSD), root–mean–square

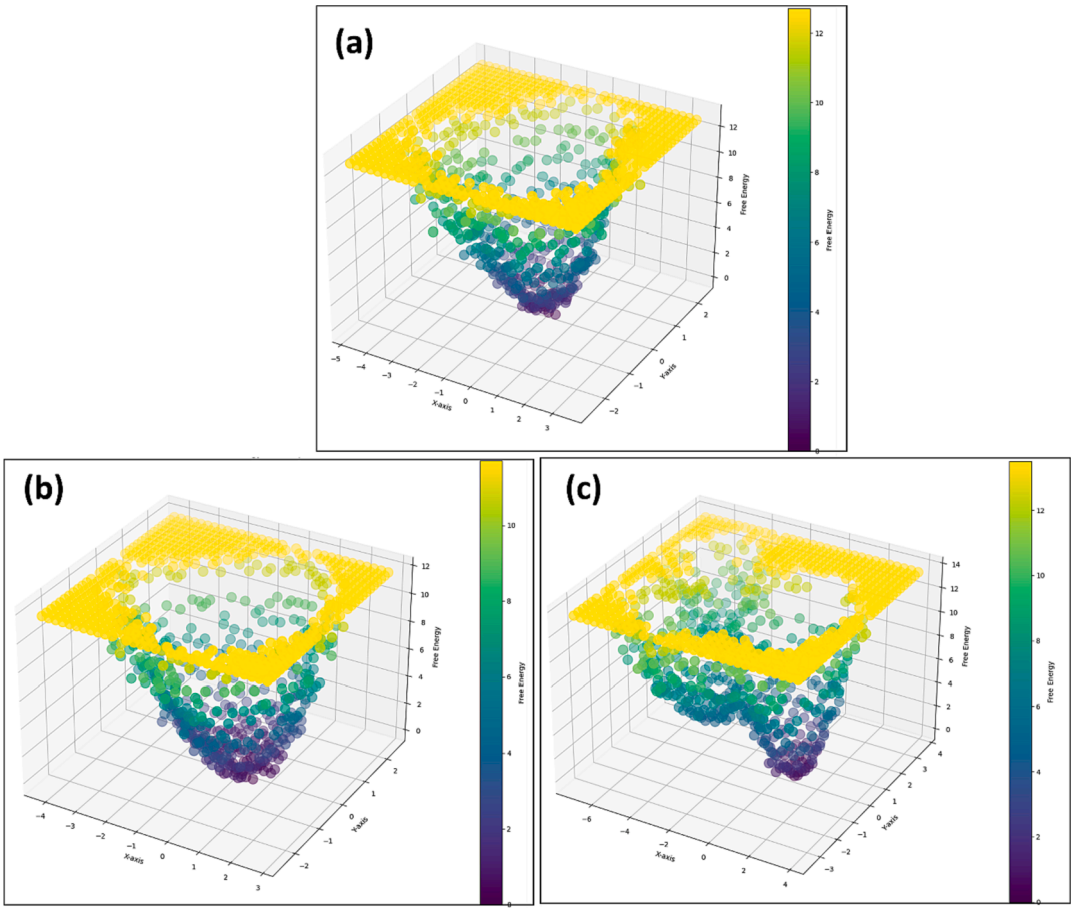

Fig. 12. 3D free–energy landscape diagrams indicated by color (lighter to darker represents higher free energy to lower free energy) for the complexes with ligands (a) Control, (b) 53309102, and (c) 57398422.

fluctuation (RMSF), and hydrogen bond formation between the ligands and the protein were the main topics of a thorough examination of molecular dynamics simulations. The analyses of RMSD, RMSF, and Radius of Gyration (Rg) plots in this study exhibited patterns consistent with a previous investigation targeting Nsp15, as conducted by Ibrahim et al. in 2022 (Ibrahim et al., 2022). During the 100 ns molecular dynamics (MD) simulation, both compounds 53309102 and 57398422 demonstrated robust and sustained binding interactions with the protein, a phenomenon congruent with the RMSD plot. The observed conformational changes were attributed to the formation of more stable bonds, as corroborated by hydrogen bond predictions. Notably, post-MD simulation, both 53309102 and 57398422 displayed interactions with critical catalytic site residues, including His235, His250, and Lys290, underscoring their substantial binding to the protein's functional site. Furthermore, these compounds exhibited interactions with key residues Thr341, Tyr343, and Ser294 within the active site, establishing crucial hydrogen bonds with these pivotal residues, as well as with the catalytic triad residues His235, His250, and Lys290. These findings provide compelling evidence of the compounds' strong and specific interactions with the Nsp15 protein, emphasizing their potential as potent inhibitors. These insights aligned with prior studies (Godoy et al., 2023; Hong et al., 2021; Sinha et al., 2020). It also coincides with the binding site residues (residues His235, His250, Lys290, Ser294, and Tyr343) of uridine 5′–monophosphate (5′–UMP) in the active site of Nsp15 (Pillon et al., 2021).

The stability of these complexes was shown by principal component analysis (PCA) and free energy landscape (FEL) plots, with two ligands (53309102, 57398422) behaving similarly to the control ligand U5P. The higher binding affinity of the selected compounds was further substantiated by the binding free energy estimates, with ΔG_{TOTAL} values indicating more beneficial and stable interactions in comparison to the control.

In conclusion, this comprehensive investigation provides compelling evidence that compounds 53309102 and 57398422 hold promise as inhibitors of Nsp15 endoribonuclease from SARS-CoV-2. The study offers a thorough understanding of the binding mechanisms, stability, and interactions with crucial residues, along with intriguing possibilities for drug development. These findings lay a strong foundation for further experimental validation and drug design efforts, contributing to the development of antiviral medications specifically targeting the SARS-CoV-2 virus.

5. Conclusion

In this study, analogues of the substrate (Uridine–5′–Monophosphate) of Nsp15, were screened for their ability to inhibit the growth of SARS–CoV–2 by binding to and inactivating the virus's target protein, Nsp15 that mediated the replication of the virus. This investigation employed a strategy involved in both QSAR–based ML models and physical mechanisms based molecular docking. The top three compounds were selected after a two–stage virtual screening process that relied on bioactivity pIC50, binding scores, and significant hydrogen bonds. The compounds 53309102, and 57398422 were found to have the best binding with the protein indicated in MD simulation, MM/GBSA, PCA, and FEL analysis. Intriguingly, this study revealed that post-MD simulation, compounds 53309102 and 57398422 exhibited interactions with the catalytic site residues, His235, His250, and Lys290, which signified their substantial binding affinity for the functional site

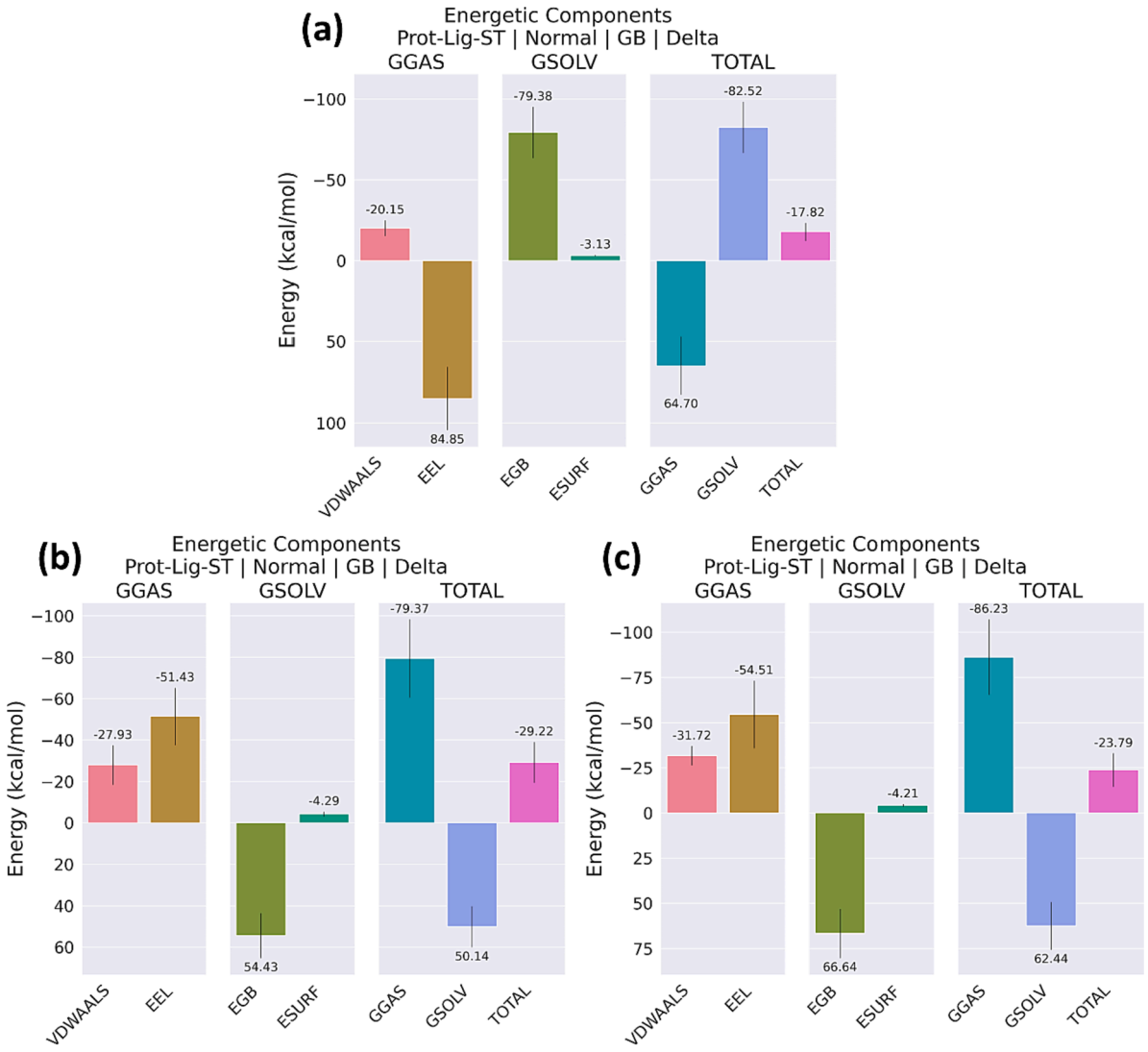

Fig. 13. Total Binding Free Energy (ΔG_{TOTAL}) of the complexes bound to the compounds (a) Control (U5P), (b) 53309102, and (c) 57398422.

of the Nsp15 protein. Moreover, during the 100 ns simulation, both compounds exhibited robust and sustained binding interactions with the protein, as corroborated by the RMSD plot. The cumulative binding free energies (ΔG_{TOTAL}) for compounds 53309102 and 57398422 were calculated to be −29.2 kcal/mol and –23.7 kcal/mol, respectively, further confirming their superior binding affinity when compared to the control. In conclusion, this study strongly advocates the utilization of compounds 53309102 and 57398422 as specific agents targeting Nsp15 of SARS-CoV-2. These compounds present a promising avenue for experimental validation aimed at combating SARS-CoV-2. The comprehensive insights provided by this research, supported by the significant binding interactions, stability, and robust binding free energy estimates, underscore the potential of these compounds as potent inhibitors of the viral replication machinery. These findings serve as a foundation for further investigations and drug development efforts in the quest to develop effective treatments for SARS-CoV-2.

Declaration of Competing Interest

The author declares that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Author is thankful to the Najran University for providing research facilities.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jsps.2023.101914.

References

Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J., Schroeder, M., 2021. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucl. Acids Res. 49, W530–W534. https://doi.org/10.1093/nar/gkab294.

Awale, M., Reymond, J.-L., 2018. The polypharmacology browser PPB2: target prediction combining nearest neighbors with machine learning (preprint). Chemistry. https://doi.org/10.26434/chemrxiv.6895646.v1.

Baidya, A.T., Kumar, A., Kumar, R., Darreh-Shori, T., 2022. Allosteric binding sites of Aβ peptides on the acetylcholine synthesizing enzyme ChAT as deduced by in silico molecular modeling. Int. J. Mol. Sci. 23, 6073. https://doi.org/10.3390/ijms23116073.

Banerjee, P., Eckert, A.O., Schrey, A.K., Preissner, R., 2018. ProTox-II: a webserver for the prediction of toxicity of chemicals. Nucleic Acids Res. 46, W257–W263. https://doi.org/10.1093/nar/gky318.

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1995. GROMACS: a message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The protein data bank. Nucleic Acids Res. 28, 235–242. https://doi.org/10.1093/nar/28.1.235.
Bhardwaj, K., Palaninathan, S., Alcantara, J.M.O., Li Yi, L., Guarino, L., Sacchettini, J.C., Kao, C.C., 2008. Structural and functional analyses of the severe acute respiratory syndrome coronavirus endoribonuclease Nsp15. J. Biol. Chem. 283, 3655–3664. https://doi.org/10.1074/jbc.M708375200.
Biovia, D.S., 2019. Discovery Studio Visualizer. San Diego.
BIOVIA, Dassault Systèmes, Discovery Studio Visualizer, v21.1.0.20298, San Diego: Dassault Systèmes, 2020., n.d.
Bommu, U.D., Konidala, K.K., Pabbaraju, N., Yeguvapalli, S., 2019. QSAR modeling, pharmacophore-based virtual screening, and ensemble docking insights into predicting potential epigallocatechin gallate (EGCG) analogs against epidermal growth factor receptor. J. Recept. Signal Transduct. 39, 18–27. https://doi.org/10.1080/10799893.2018.1564151.
Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126, 014101 https://doi.org/10.1063/1.2408420.
Chafekar, A., Fielding, B.C., 2018. MERS-CoV: understanding the latest human Coronavirus threat. Viruses 10, 93. https://doi.org/10.3390/v10020093.
Chandra, A., Gurjar, V., Qamar, I., Singh, N., 2021. Identification of potential inhibitors of SARS-COV-2 endoribonuclease (EndoU) from FDA approved drugs: a drug repurposing approach to find therapeutics for COVID-19. J. Biomol. Struct. Dyn. 39, 4201–4211. https://doi.org/10.1080/07391102.2020.1775127.
Cheng, A., Zhang, W., Xie, Y., Jiang, W., Arnold, E., Sarafianos, S.G., Ding, J., 2005. Expression, purification, and characterization of SARS coronavirus RNA polymerase. Virology 335, 165–176. https://doi.org/10.1016/j.virol.2005.02.017.
Cichero, E., Calautti, A., Francesconi, V., Tonelli, M., Schenone, S., Fossa, P., 2021. Probing in silico the benzimidazole privileged scaffold for the development of drug-like anti-RSV agents. Pharmaceuticals 14, 1307. https://doi.org/10.3390/ph14121307.
Cotuá, J., Llinás, H., Cotes, S., 2021. Virtual screening based on QSAR and molecular docking of possible inhibitors targeting chagas CYP51. J. Chem. 2021, e6640624.
Daina, A., Michielin, O., Zoete, V., 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci. Rep. 7, 42717. https://doi.org/10.1038/srep42717.
Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. https://doi.org/10.1063/1.464397.
Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., Overington, J.P., 2015. ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res. 43, W612–W620. https://doi.org/10.1093/nar/gkv352.
Deng, X., Baker, S.C., 2018. An "Old" protein with a new story: coronavirus endoribonuclease is important for evading host antiviral defenses. Virology 517, 157–163. https://doi.org/10.1016/j.virol.2017.12.024.
Deng, X., Hackbart, M., Mettelman, R.C., O'Brien, A., Mielech, A.M., Yi, G., Kao, C.C., Baker, S.C., 2017. Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. Proc. Natl. Acad. Sci. 114 https://doi.org/10.1073/pnas.1618310114.
Deng, S.-Q., Peng, H.-J., 2020. Characteristics of and public health responses to the Coronavirus disease 2019 outbreak in China. J. Clin. Med. 9, 575. https://doi.org/10.3390/jcm9020575.
Eberhardt, J., Santos-Martins, D., Tillack, A.F., Forli, S., 2021. AutoDock Vina 1.2.0: new docking methods, expanded force field, and python bindings. J. Chem. Inf. Model. 61, 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.
Fan, F., Warshaviak, D.T., Hamadeh, H.K., Ii, R.T.D., 2019. The integration of pharmacophore-based 3D QSAR modeling and virtual screening in safety profiling: a case study to identify antagonistic activities against adenosine receptor, A2A, using 1,897 known drugs. PLoS One 14, e0204378.
Gao, B., Gong, X., Fang, S., Weng, W., Wang, H., Chu, H., Sun, Y., Meng, C., Tan, L., Song, C., Qiu, X., Liu, W., Forlenza, M., Ding, C., Liao, Y., 2021. Inhibition of anti-viral stress granule formation by coronavirus endoribonuclease nsp15 ensures efficient virus replication. PLOS Pathog. 17, e1008690.
Godoy, A.S., Nakamura, A.M., Douangamath, A., Song, Y., Noske, G.D., Gawriljuk, V.O., Fernandes, R.S., Pereira, H.D.M., Oliveira, K.I.Z., Fearon, D., Dias, A., Krojer, T., Fairhead, M., Powell, A., Dunnet, L., Brandao-Neto, J., Skyner, R., Chalk, R., Bajusz, D., Bege, M., Borbás, A., Keserű, G.M., von Delft, F., Oliva, G., 2023. Allosteric regulation and crystallographic fragment screening of SARS-CoV-2 NSP15 endoribonuclease. Nucleic Acids Res. 51, 5255–5270. https://doi.org/10.1093/nar/gkad314.
Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J., Snijder, E.J., 2006. Nidovirales: evolving the largest RNA virus genome. Virus Res. 117, 17–37. https://doi.org/10.1016/j.virusres.2006.01.017.
Guan, W., Ni, Z., Hu, Y.u., Liang, W., Ou, C., He, J., Liu, L., Shan, H., Lei, C., Hui, D.S.C., Du, B., Li, L., Zeng, G., Yuen, K.-Y., Chen, R., Tang, C., Wang, T., Chen, P., Xiang, J., Li, S., Wang, J.-L., Liang, Z., Peng, Y., Wei, L., Liu, Y., Hu, Y.-H., Peng, P., Wang, J.-M., Liu, J., Chen, Z., Li, G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., Zhong, N., 2020. Clinical characteristics of Coronavirus disease 2019 in China. N. Engl. J. Med. 382, 1708–1720. https://doi.org/10.1056/NEJMoa2002032.
Guerra, C.J., López, J.M., Figueredo, S.F., Muñoz, A.E., Robles, J.R., 2016. 2D-QSAR analysis of derivatives of quinoxalinE 1,4-DI- *N* -oxides with activity against chagas' disease. Quím. Nova. https://doi.org/10.5935/0100-4042.20160078.
Hackbart, M., Deng, X., Baker, S.C., 2020. Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. Proc. Natl. Acad. Sci. 117, 8094–8103. https://doi.org/10.1073/pnas.1921485117.
Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theory Comput. 4, 435–447. https://doi.org/10.1021/ct700301q.
Hong, S., Seo, S.H., Woo, S.-J., Kwon, Y., Song, M., Ha, N.-C., 2021. Epigallocatechin gallate inhibits the uridylate-specific endoribonuclease Nsp15 and efficiently neutralizes the SARS-CoV-2 strain. J. Agric. Food Chem. 69, 5948–5954. https://doi.org/10.1021/acs.jafc.1c02050.
Hu, B., Guo, H., Zhou, P., Shi, Z.-L., 2021. Characteristics of SARS-CoV-2 and COVID-19. Nat. Rev. Microbiol. 19, 141–154. https://doi.org/10.1038/s41579-020-00459-7.
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., Cheng, Z., Yu, T., Xia, J., Wei, Y., Wu, W., Xie, X., Yin, W., Li, H., Liu, M., Xiao, Y., Gao, H., Guo, L., Xie, J., Wang, G., Jiang, R., Gao, Z., Jin, Q., Wang, J., Cao, B., 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.
Ibrahim, I.M., Elfiky, A.A., Fathy, M.M., Mahmoud, S.H., ElHefnawi, M., 2022. Targeting SARS-CoV-2 endoribonuclease: a structure-based virtual screening supported by in vitro analysis. Sci. Rep. 12, 13337. https://doi.org/10.1038/s41598-022-17573-6.
Ivanov, K.A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A.E., Ziebuhr, J., 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. Proc. Natl. Acad. Sci. 101, 12694–12699. https://doi.org/10.1073/pnas.0403127101.
Izadi, S., Anandakrishnan, R., Onufriev, A.V., 2014. Building water models: a different approach. J. Phys. Chem. Lett. 5, 3863–3871. https://doi.org/10.1021/jz501780a.
Jawarkar, R.D., Bakal, R.L., Zaki, M.E.A., Al-Hussain, S., Ghosh, A., Gandhi, A., Mukerjee, N., Samad, A., Masand, V.H., Lewaa, I., 2022. QSAR based virtual screening derived identification of a novel hit as a SARS CoV-229E 3CLpro Inhibitor: GA-MLR QSAR modeling supported by molecular Docking, molecular dynamics simulation and MMGBSA calculation approaches. Arab. J. Chem. 15, 103499 https://doi.org/10.1016/j.arabjc.2021.103499.
Kang, H., Bhardwaj, K., Li, Y., Palaninathan, S., Sacchettini, J., Guarino, L., Leibowitz, J. L., Kao, C.C., 2007. Biochemical and genetic analyses of murine Hepatitis Virus Nsp15 endoribonuclease. J. Virol. 81, 13587–13597. https://doi.org/10.1128/JVI.00547-07.
Kannan, S., Kolandaivel, P., 2018. The inhibitory performance of flavonoid cyanidin-3-sambubiocide against H274Y mutation in H1N1 influenza virus. J. Biomol. Struct. Dyn. 36, 4255–4269. https://doi.org/10.1080/07391102.2017.1413422.
Kim, Y., Jedrzejczak, R., Maltseva, N.I., Wilamowski, M., Endres, M., Godzik, A., Michalska, K., Joachimiak, A., 2020b. Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV -2. Protein Sci. 29, 1596–1605. https://doi.org/10.1002/pro.3873.
Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J.W., Kim, V.N., Chang, H., 2020a. The Architecture of SARS-CoV-2 Transcriptome. Cell 181, 914–921.e10. https://doi.org/10.1016/j.cell.2020.04.011.
Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., Kang, S., Nicolaescu, V., Randall, G., Michalska, K., Joachimiak, A., 2021. Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. Commun. Biol. 4, 193. https://doi.org/10.1038/s42003-021-01735-9.
Kindler, E., Gil-Cruz, C., Spanier, J., Li, Y., Wilhelm, J., Rabouw, H.H., Züst, R., Hwang, M., V'kovski, P., Stalder, H., Marti, S., Habjan, M., Cervantes-Barragan, L., Elliot, R., Karl, N., Gaughan, C., Van Kuppeveld, F.J.M., Silverman, R.H., Keller, M., Ludewig, B., Bergmann, C.C., Ziebuhr, J., Weiss, S.R., Kalinke, U., Thiel, V., 2017. Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. PLOS Pathog. 13, e1006195.
Landrum, G., 2014. Rdkit: Open-Source Cheminformatics. Release 2014.03.1. Zenodo. https://doi.org/10.5281/ZENODO.10398.
Maisuradze, G.G., Liwo, A., Scheraga, H.A., 2010. Relation between free energy landscapes of proteins and dynamics. J. Chem. Theory Comput. 6, 583–595. https://doi.org/10.1021/ct9005745.
Malik, Y.S., Kumar, N., Sircar, S., Kaushik, R., Bhat, S., Dhama, K., Gupta, P., Goyal, K., Singh, M.P., Ghoshal, U., El Zowalaty, M.E., Yatoo, M.I., Tiwari, R., Pathak, M., Patel, S.K., Sah, R., Rodriguez-Morales, A.J., Ganesh, B., Kumar, P., Singh, R.K., 2020. Coronavirus Disease Pandemic (COVID-19): challenges and a global perspective. Pathogens 9, 519. https://doi.org/10.3390/pathogens9070519.
Mercatelli, D., Giorgi, F.M., 2020. Geographic and genomic distribution of SARS-CoV-2 mutations. Front. Microbiol. 11.
Montaruli, M., Alberga, D., Ciriaco, F., Trisciuzzi, D., Tondo, A.R., Mangiatordi, G.F., Nicolotti, O., 2019. Accelerating drug discovery by early protein drug target prediction based on a multi-fingerprint similarity search †. Molecules 24, 2233. https://doi.org/10.3390/molecules24122233.
Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791. https://doi.org/10.1002/jcc.21256.
O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open babel: an open chemical toolbox. J. Cheminformatics 3, 33. https://doi.org/10.1186/1758-2946-3-33.
Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. J. Appl. Phys. 52, 7182–7190. https://doi.org/10.1063/1.328693.

Pillon, M.C., Frazier, M.N., Dillard, L.B., Williams, J.G., Kocaman, S., Krahn, J.M., Perera, L., Hayne, C.K., Gordon, J., Stewart, Z.D., Sobhany, M., Deterding, L.J., Hsu, A.L., Dandey, V.P., Borgnia, M.J., Stanley, R.E., 2021. Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. Nat. Commun. 12, 636. https://doi.org/10.1038/s41467-020-20608-z.

Poirier, M., Awale, M., Roelli, M.A., Giuffredi, G.T., Ruddigkeit, L., Evensen, L., Stooss, A., Calarco, S., Lorens, J.B., Charles, R.-P., Reymond, J.-L., 2019. Identifying Lysophosphatidic Acid Acyltransferase β (LPAAT-β) as the target of a nanomolar angiogenesis inhibitor from a phenotypic screen using the polypharmacology browser PPB2. ChemMedChem 14, 224–236. https://doi.org/10.1002/cmdc.201800554.

Reymond, J.-L., 2022. Molecular similarity for drug discovery, target prediction and chemical space visualization. Chimia 76, 1045. https://doi.org/10.2533/chimia.2022.1045.

Ricagno, S., Egloff, M.-P., Ulferts, R., Coutard, B., Nurizzo, D., Campanacci, V., Cambillau, C., Ziebuhr, J., Canard, B., 2006. Crystal structure and mechanistic determinants of SARS coronavirus nonstructural protein 15 define an endoribonuclease family. Proc. Natl. Acad. Sci. 103, 11892–11897. https://doi.org/10.1073/pnas.0601708103.

Rudrapal, M., Chetia, D., 2020. Virtual screening, molecular docking and QSAR studies in drug discovery and development programme. J. Drug Deliv. Ther. 10, 225–233. https://doi.org/10.22270/jddt.v10i4.4218.

Sanders, J.M., Monogue, M.L., Jodlowski, T.Z., Cutrell, J.B., 2020. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): a review. JAMA 323, 1824–1836. https://doi.org/10.1001/jama.2020.6019.

Saramago, M., Costa, V., Souza, C., Bárria, C., Domingues, S., Viegas, S., Lousa, D., Soares, C., Arraiano, C., Matos, R., 2022. The nsp15 nuclease as a good target to combat SARS-CoV-2: mechanism of action and its inactivation with FDA-approved drugs. Microorganisms 10, 342. https://doi.org/10.3390/microorganisms10020342.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. J. Virol. 81, 20–29. https://doi.org/10.1128/JVI.01358-06.

Singh, N.A., Kumar, P., Jyoti, N.K., 2021. Spices and herbs: potential antiviral preventives and immunity boosters during COVID-19. Phytother. Res. 35, 2745–2757. https://doi.org/10.1002/ptr.7019.

Sinha, S.K., Shakya, A., Prasad, S.K., Singh, S., Gurav, N.S., Prasad, R.S., Gurav, S.S., 2020. 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. 1–12 https://doi.org/10.1080/07391102.2020.1762741.

Snijder, E.J., Decroly, E., Ziebuhr, J., 2016. Chapter Three - The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. In: Ziebuhr, J. (Ed.), Advances in Virus Research, Coronaviruses. Academic Press, pp. 59–126. https://doi.org/10.1016/bs.aivir.2016.08.008.

Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B., Imbert, I., 2014. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc. Natl. Acad. Sci. 111, E3900–E3909. https://doi.org/10.1073/pnas.1323705111.

Tracking SARS-CoV-2 variants [WWW Document], n.d. URL https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed 8.23.23).

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., Thiel, V., 2021. Coronavirus biology and replication: implications for SARS-CoV-2. Nat. Rev. Microbiol. 19, 155–170. https://doi.org/10.1038/s41579-020-00468-6.

Van De Waterbeemd, H., Gifford, E., 2003. ADMET in silico modelling: towards prediction paradise? Nat. Rev. Drug Discov. 2, 192–204. https://doi.org/10.1038/nrd1032.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet 395, 470–473. https://doi.org/10.1016/S0140-6736(20)30185-9.

Weekly epidemiological update on COVID-19 - 17 August 2023 [WWW Document], n.d. URL https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19—17-august-2023 (accessed 8.23.23).

Weininger, D., 1988. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. J. Chem. Inf. Comput. Sci. 28, 31–36. https://doi.org/10.1021/ci00057a005.

Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z., Rao, Z., 2018. Structural and biochemical characterization of endoribonuclease Nsp15 encoded by middle east respiratory syndrome Coronavirus. J. Virol. 92, e00893–e918. https://doi.org/10.1128/JVI.00893-18.

Zhu, Z., Lian, X., Su, X., Wu, W., Marraro, G.A., Zeng, Y., 2020. 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, 224. https://doi.org/10.1186/s12931-020-01479-w.

Zoete, V., Cuendet, M.A., Grosdidier, A., Michielin, O., 2011. SwissParam: a fast force field generation tool for small organic molecules. J. Comput. Chem. 32, 2359–2368. https://doi.org/10.1002/jcc.21816.