Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of Molecular Structure 1251 (2022) 132041

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstr

Ligand-based quantitative structural assessments of SARS-CoV-2 3CL^{pro} inhibitors: An analysis in light of structure-based multi-molecular modeling evidences

Nilanjan Adhikari[a,1], Suvankar Banerjee[a,1], Sandip Kumar Baidya[a], Balaram Ghosh[b,*], Tarun Jha[a,*]

[a] Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India
[b] Epigenetic Research Laboratory, Birla Institute of Technology and Science-Pilani Hyderabad Campus, Shamirpet, Hyderabad, India, 500078

ARTICLE INFO

Article history:
Received 2 October 2021
Revised 10 November 2021
Accepted 26 November 2021
Available online 29 November 2021

Keywords:
COVID-19
SARS-CoV-2 3CL^{pro}
QSAR
ANN
SVM
HQSAR

ABSTRACT

Due to COVID-19, the whole world is undergoing a devastating situation, but treatment with no such drug candidates still has been established exclusively. In that context, 69 diverse chemicals with potential SARS-CoV-2 3CL^{pro} inhibitory property were taken into consideration for building different internally and externally validated linear (SW-MLR and GA-MLR), non-linear (ANN and SVM) QSAR, and HQSAR models to identify important structural and physicochemical characters required for SARS-CoV-2 3CL^{pro} inhibition. Importantly, 2-oxopyrrolidinyl methyl and benzylester functions, and methylene (hydroxy) sulphonic acid warhead group, were crucial for retaining higher SARS-CoV-2 3CL^{pro} inhibition. These GA-MLR and HQSAR models were also applied to predict some already repurposed drugs. As per the GA-MLR model, curcumin, ribavirin, saquinavir, sepimostat, and remdesivir were found to be the potent ones, whereas according to the HQSAR model, lurasidone, saquinavir, lopinavir, elbasvir, and paritaprevir were the highly effective SARS-CoV-2 3CL^{pro} inhibitors. The binding modes of those repurposed drugs were also justified by the molecular docking, molecular dynamics (MD) simulation, and binding energy calculations conducted by several groups of researchers. This current work, therefore, may be able to find out important structural parameters to accelerate the COVID-19 drug discovery processes in the future.

1. Introduction

In December 2019, many people of Wuhan, China, have been affected with a serious, unknown pneumonia-like disorder that did not respond to any antibiotics [1]. Certainly, the disease spread at lightning speed to the rest of the world, and therefore, the World Health Organization (WHO) announced this infection as coronavirus disease 2019 (COVID-19) [2]. This disease has been produced by the infection of the novel coronavirus or severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) [3–4]. Consequently, this novel coronavirus is found to be closely related to a similar virus, i.e., SARS-CoV spread during 2003 [1]. Currently, COVID-19 has greatly affected the whole world and around 222 countries and territories throughout the world have been found to have the COVID-19 infection. More than 232 million COVID-19 cases and over 4.7 million deaths were reported due to this CoVID-19 disease as of May 2021 [5]. The severe COVID-19 is diagnosed as acute respiratory distress syndrome (ARDS) and multiple organ failure possibly induced by the uncontrolled immune response in the host cells [6–7]. Coronavirus (CoV) infection in humans, as well as in other animals, has resulted in a variety of highly frequent and serious diseases, along with the SARS and the Middle East respiratory syndrome (MERS) [8]. The SARS was found first in China in November 2002. In 2003, the WHO found out the causative agent of SARS, and they identified the SARS-CoV [9]. Again, SARS-CoV can also be found among some animals like civet cats and horseshoe bats [10]. However, the source of the MERS-CoV was primarily searched in bats, but the dromedary camels of Oman and Canary Islands showed to have a high frequency of MERS-CoV-neutralizing antibodies [9]. Interestingly, in line with the debate on the actual source, the SARS-CoV-2 was found in bats (RaTG13) and pangolins

* Corresponding authors.
E-mail addresses: balaram@hyderabad.bits-pilani.ac.in (B. Ghosh), tjupharm@yahoo.com (T. Jha).
[1] Authors have equal contributions.

https://doi.org/10.1016/j.molstruc.2021.132041
0022-2860/

[11]. The pangolins and the civet cats served as an intermediary host, but researchers detect the similarity among the strains of CoVs which are closely related to SARS-CoV-2 [1].

The structure of *coronaviruses* consists of single-stranded positive-sense RNA that acquires a greater number of viral RNA genomes [12–13]. Some of the newer studies demonstrated that SARS-CoV-2 has a similar genomic resemblance with other *Betacoronaviruses*. The main structure of SARS-CoV-2 contains spike protein (S), membrane protein (M), open reading frame 1ab (ORF1ab) which encodes non-structural proteins (nsps), envelope protein (E), 50-untranslated region (UTR), 30-untranslated region (UTR), nucleocapsid protein (N) and other unidentified non-structural ORFs that are useful toward the construction of viral particles [14]. A common coronavirus genome gives rise to more than 20 proteins and at least six open reading frames (ORFs) [13,15]. Moreover, two-third of a genome contains first ORF (ORF1a/b) and 16 non-structural proteins (nsp1-nsp16) but the *Gammacoronavirus* lacks nsp1 [13]. Several coronaviral proteins have already been established as promising targets for COVID-19 drug discovery. These include helicase, RNA-dependent RNA polymerase (RdRp), hemagglutinin esterase along with the structural and other non-structural proteins which can be targeted for anti-SARS-CoV-2 drug discovery [3,16].

Among these proteins, there are two proteases, namely papain-like protease (PL^{pro}) and main protease or 3-chymotrypsin-like protease (M^{pro}/$3CL^{pro}$), that are essential components for replication of SARS-CoV-2. These PL^{pro} and $3CL^{pro}$ are responsible for cleaving the two polyproteins, i.e., PP1A and PP1AB into several functional components. PL^{pro} split the N-terminal domain to viral precursor protein at the three sites whereas $3CL^{pro}$ splits the C-terminal domain to precursor protein at the 11 sites [17]. However, viral proteases such as $3CL^{pro}$/M^{pro} along with PL^{pro} are responsible for the nsp preparation by processing these polyproteins [13]. Thus, M^{pro} and PL^{pro} are auspicious targets for antiviral drug development. Enormous efforts have been made to analyze their significance and to develop effective therapeutics against the SARS-CoV. Similar approaches had been adopted for other pathogenic coronaviruses (i.e., MERS-CoV) because of the similarity of active sites and enzymatic mechanisms with SARS-CoV [18–19]. Interestingly, the sequence homology of SARS-CoV-2 $3CL^{pro}$ is 96% structurally closer to the SARS-CoV $3CL^{pro}$ [20]. These structural similarities point out that targeting SARS-CoV-2 $3CL^{pro}$ as a potential therapeutic target is quite a feasible approach for anti-CoVID-19 drug development.

Besides, the development of newer anti-SARS-CoV-2 agents, the study of previously reported SARS-CoV inhibitors can be helpful. In this scenario, drug repurposing has emerged out as a major approach in the development of anti-SARS-CoV-2 treatment [21]. Also, several repurposed drugs such as lopinavir, ritonavir, azvudine, favipiravir, remdesivir, chloroquine, hydroxychloroquine, methylprednisolone, tocilizumab, ribavirin, and oseltamivir are being tested in the clinical studies against SARS-CoV-2 treatment [22–23].

Our group has already performed several extensive studies on the previous SARS-CoV inhibitors [3,18–19,24–30]. In this study, we have reported the structural analysis of 69 diverse SARS-CoV-2 $3CL^{pro}$ inhibitors by the regression-based quantitative structure-activity relationship (QSAR) methodologies to identify the fundamental structural features having crucial effects on SARS-CoV-2 $3CL^{pro}$ inhibition. The outcomes of this current research were also validated using the available SARS-CoV-2 $3CL^{pro}$ crystal structure-bound ligand interactions. This may be beneficial in the newer anti-coronaviral drug development, optimization of previous SARS-CoV inhibitors, and screening of newer drugs for CoVID-19 treatment.

2. Experimental

2.1. Dataset preparation

A total of 69 structurally diverse chemical entities (Supplementary Table S1) with a wide range of SARS-CoV-2 $3CL^{pro}$ inhibitory activity *in vitro* (IC_{50} value ranging from 0.01 μM - 124.93 μM) were mustered together from the literature [20,31–38]. To maintain the uniformity of the dataset, the mean SARS-CoV-2 $3CL^{pro}$ inhibitory activity (IC_{50} in μM) values were transformed into their negative logarithmic scale. A set of 1444 2D molecular descriptors was calculated for each compound using PaDEL descriptor software [39] followed by the dataset pre-treatment technique to eliminate the highly correlated descriptors. The dataset division was carried out using the Kennard-Stone (*KS*) method using DTC Lab software [40] where a 3:1 ratio was preserved for the training and the test set ($N_{Train} = 51$, $N_{Test} = 18$).

2.2. Feature selection and MLR model development

For the multiple linear regression (MLR) model development, two different techniques such as stepwise (SW) and genetic algorithm (GA) methods were utilized to reduce the dimension of predictor parameters [41–42]. The best subset selection process was applied to this reduced data to identify the best regression model consisting of five molecular descriptors using DTC Lab software [40,42]. The best models were selected based on their squared correlation coefficient (R^2), Leave-One-Out (*LOO*)-cross-validated R^2 (Q^2), and externally validated R^2 (R^2_{Pred}) values [42–43]. Both the internal and external cross-validations were conducted to justify the reliability and predictive ability of selected MLR models. Statistical validation parameters, namely adjusted R^2 (R^2_A), standard error of estimate (*SEE*), predicted residual sum of squares (*PRESS*), and variance ratio (F) at a specified degree of freedom value were estimated. To rationalize the robustness of the selected MLR models, the internal cross-validation parameter such as Q^2 was calculated that indicates the internal predictability of the MLR models for the training population. The external predictability of these MLR models was evaluated using the R^2_{Pred} values [43]. Additionally, employment of the Y-randomization test (cRp^2), calculation of the r_m^2 metrics, and the Golbraikh-Tropsha model acceptability criteria were also determined for these MLR models [42–44].

2.3. Non-linear QSAR model development

Here, the development of non-linear QSAR models such as the artificial neural network (ANN) and the support vector machine (SVM) was carried out by the descriptors that were used to build the MLR models. This was performed to validate these MLR models further as well as to investigate the machine learning capability of those selected molecular descriptors.

2.3.1. Artificial Neural Network (ANN) model

The artificial neural network (ANN) method is a popular machine learning technique for QSAR model development. It imitates the biological neuronal functions and deals with feed-forward and back-propagation of error algorithms [45–46]. The typical construction of an ANN model generally comprises three different layers namely, the input, hidden, and output layers. Compared to the feed-forward network of the brain neural nodes, in ANN method, the input is provided in the input layer which is forwarded to the successive hidden layer. In this hidden layer, the information is conveyed to the nodes present in it, and the output is finally sent from the hidden layer to the output layer [45–47].

In this study, to investigate the machine learning capability of these MLR models, the Autoweka software [48] was used to opti-

mize the learning process (optimization of parameters like learning time, number of epochs, momentum, and learning rate). The final ANN model was constructed in Weka 3.8 software [49] using the backpropagation of error algorithm and was both internally and externally cross-validated.

2.3.2. Support Vector Machine (SVM) model

The support vector machine (SVM) model was first introduced by Vapnik [50] which follows the structure risk minimization principles of statistical learning theory [46]. SVM is used to solve problems through either regression or classification-based analyses. In SVM, to create a maximum margin, the input samples are grooved in separate classes by hyperplane construction into the linear data. Here, the non-linear conversion of the linear data is carried out using the Kernel function [$K_{(x,\ y)}$] and project the variable matrix into a higher dimensional feature space employing the Kernel function [46–47,51]. Here, the radial basis function (*RBF*) kernel was utilized to transform the data. Like the ANN model development, the optimization of the SVM model parameters such as the complexity of the kernel (C) and kernel width (γ) was done using Autoweka software [48]. The final SVM models were constructed using Weka 3.8 software [49].

2.4. Hologram QSAR (HQSAR) model

The utilization of fragment-based approaches in the arena of drug design and discovery has provided several successes over time and has become more popular over the last decade [52–53]. The hologram QSAR (HQSAR) method utilizes the molecular hologram or specialized molecular fingerprints of different lengths as the independent variable to correlate them with the biological response of compounds with the help of the partial least square (PLS) technique [54]. In this study, the HQSAR models were constructed on the molecules of the training set by SYBYL-X 2.0 software [55] using various combinations of fragments/parameters. These fragment distinction parameters include the atom number (A), atomic connections (C), chirality (Ch), bonds (B), number of hydrogens (H), and donor-acceptor (DA) features. Depending on the combination of fragment distinction parameters, the HQSAR models were constructed on the training set molecules. The best model among these constructed HQSAR models was chosen depending on the highest Leave-One-Out (*LOO*) cross-validated R^2 (Q^2) and the lower standard error (*SE*) value. Moreover, this best HQSAR model was subjected to external validation on test set molecules [43,47].

3. Results and discussion

3.1. Multiple linear regression (MLR) models

Both of the final MLR models (SW-MLR and GA-MLR) models were constructed using 5 molecular descriptors and were selected as the final models depending on their Q^2 and R^2_{pred} values. The developed SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models are shown below:

$$pIC_{50} = 5.270(\pm0.125) + 0.164(\pm0.035)maxHBint10 - 0.133(\pm0.019)naasC - 0.368(\pm0.077)nssssC + 0.551(\pm0.155)MDEN{-}22 - 3.812(\pm1.193)MATS1s \quad \text{(Eq. 1)}$$

$N_{train} = 51$, $N_{test} = 18$, $R = 0.838$, $R^2 = 0.703$, $R^2_A = 0.670$, $Q^2 = 0.636$, $SEE = 0.544$, $PRESS = 13.308$, $F(5, 45) = 21.258$, $Avg.\ r^2_{mLOO} = 0.515$, $cR^2_p = 0.656$, $R^2_{pred} = 0.565$; $Avg.\ r^2_{mtest} = 0.543$, $p < 0.05$.

$$pIC_{50} = 4.358(\pm0.202) + 4.942(\pm0.966)MATS3e + 0.124(\pm0.039)maxHBint10 + 0.198(\pm0.044)maxHBint8 + 1.367(\pm0.490)FMF - 0.5653(\pm0.093)MDEO{-}22 \quad \text{(Eq. 2)}$$

$N_{train} = 51$, $N_{test} = 18$, $R = 0.860$, $R^2 = 0.740$, $R^2_A = 0.711$, $Q^2 = 0.674$, $SEE = 0.508$, $PRESS = 11.618$, $F(5, 45) = 25.660$, $Avg.\ r^2_{mLOO} = 0.566$, $cR^2_p = 0.692$, $R^2_{pred} = 0.654$; $Avg.\ r^2_{mtest} = 0.502$, $p < 0.05$.

The SW-MLR model (***Eq. (1)***) was found to predict 63.6% and explain 67% SARS-CoV-2 3CL^{pro} inhibition of the dataset molecules, whereas the GA-MLR model (***Eq. (2)***) was capable of explaining 71.1% and predict 67.4% activity variation of the dataset molecules. The meaning of the descriptors used to build the SW-MLR and GA-MLR models are provided in Table 1. Additionally, the MLR models were also able to pass the Golbraikh and Tropsha model acceptability criteria [44] (Table 2).

The observed versus predicted activity plots for the SW-MLR and GA-MLR models are given in Fig. 1A and 1B, respectively.

Furthermore, the applicability domain (AD) of the constructed MLR models has been checked using the Euclidean distance-based method [40] where the normalized mean distance of molecules for each of the developed MLR models had been investigated (Fig. 1C and 1D). Additional information regarding ***Eq. (1)*** and ***Eq. (2)*** are provided in Supplementary Tables S2-S7.

3.1.1. Robust validation of the MLR models

The robustness of both these final SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models was further validated. 50 different training and test set combinations were made and keeping the same descriptors, these models were tried for their statistical evaluation. Interestingly, all these 50 models (Supplementary Tables S8 and S9) generated were statistically validated as evidenced by their internal and external cross-validation parameters. This further suggests that the selection of these descriptors used to develop the SW-MLR and GA-MLR models is validated.

3.1.2. Interpretation of MLR model descriptors

In the SW-MLR model (***Eq. (1)***), five descriptors (*naasC, nssssC, MDEN-22, MATS1s, maxHBint10*) were found to be correlated with the SARS-CoV-2 3CL^{pro} inhibition. The descriptor *naasC* denotes the count of atom-type E-state:C:. It was noticed that compounds having higher values of the descriptor possess low activity (Compounds **39, 50, 52, 64, 54–57**, and **60–63**). On the other hand, compounds with zero value of the descriptor possess higher SARS-CoV-2 3CL^{pro} inhibitors (Compounds **11, 19, 21, 25, 27**, and **30–35**). The other descriptor *nssssC* denotes the count of the atom-type E-state:>C<. It represents the E-state of the SP^3 hybridized carbon atom. The negative coefficient of *nssssC* indicates that molecules possessing higher values of this descriptor disfavor SARS-CoV-2 3CL^{pro} inhibition. Compounds having such type of carbon functions are less potent inhibitors (Compounds **2–3, 38, 49, 54, 58, 60**, and **65**). Another descriptor *MDEN-22* represents the molecular distance edge between all secondary nitrogen atoms. The positive contribution of *MDEN-22* implies that compounds with higher values of this descriptor possess higher SARS-CoV-2 3CL^{pro} inhibitory activity. Compounds containing a higher number of secondary nitrogen atoms are effective SARS-CoV-2 3CL^{pro} inhibitors (Compounds **9, 19, 21–22**, and **25–26**). Again, *MATS1s* denotes the Moran autocorrelation-lag 1/weighted by I state. The negative coefficient of this descriptor reveals that compounds with lower values of *MATS1s* are effective inhibitors of SARS-CoV-2 3CL^{pro} (Compounds **22, 25–27, 30**, and **32**). In the SW-MLR model (***Eq. (1)***), the positive impact of the descriptor *maxHBint10* has been implicated in SARS-CoV-2 3CL^{pro} inhibition. It denotes the maximum E-state descriptor of strength for potential hydrogen

Table 1
Meaning of the descriptors used to construct SW-MLR and GA-MLR models.

Sl No.	Descriptor	Description
1	*naasC*	Count of atom-type E-state:C:
2	*nssssC*	Count of atom-type E-state:>C<
3	*MDEN-22*	Molecular distance edge between all secondary nitrogen
4	*MATS1s*	Moran autocorrelations-lag 1/weighted by I state
5	*maxHBint10*	Maximum E-state descriptor of strength for potential hydrogen bonds of path length 10
6	*maxHBint8*	Maximum E-state descriptor of strength for potential hydrogen bonds of path length 8
7	*MDEO-22*	Distance edge between all secondary oxygens
8	*FMF*	Complexity of a molecule
9	*MATS3e*	Moran autocorrelation-lag 3/weighted by Sanderson electronegativities

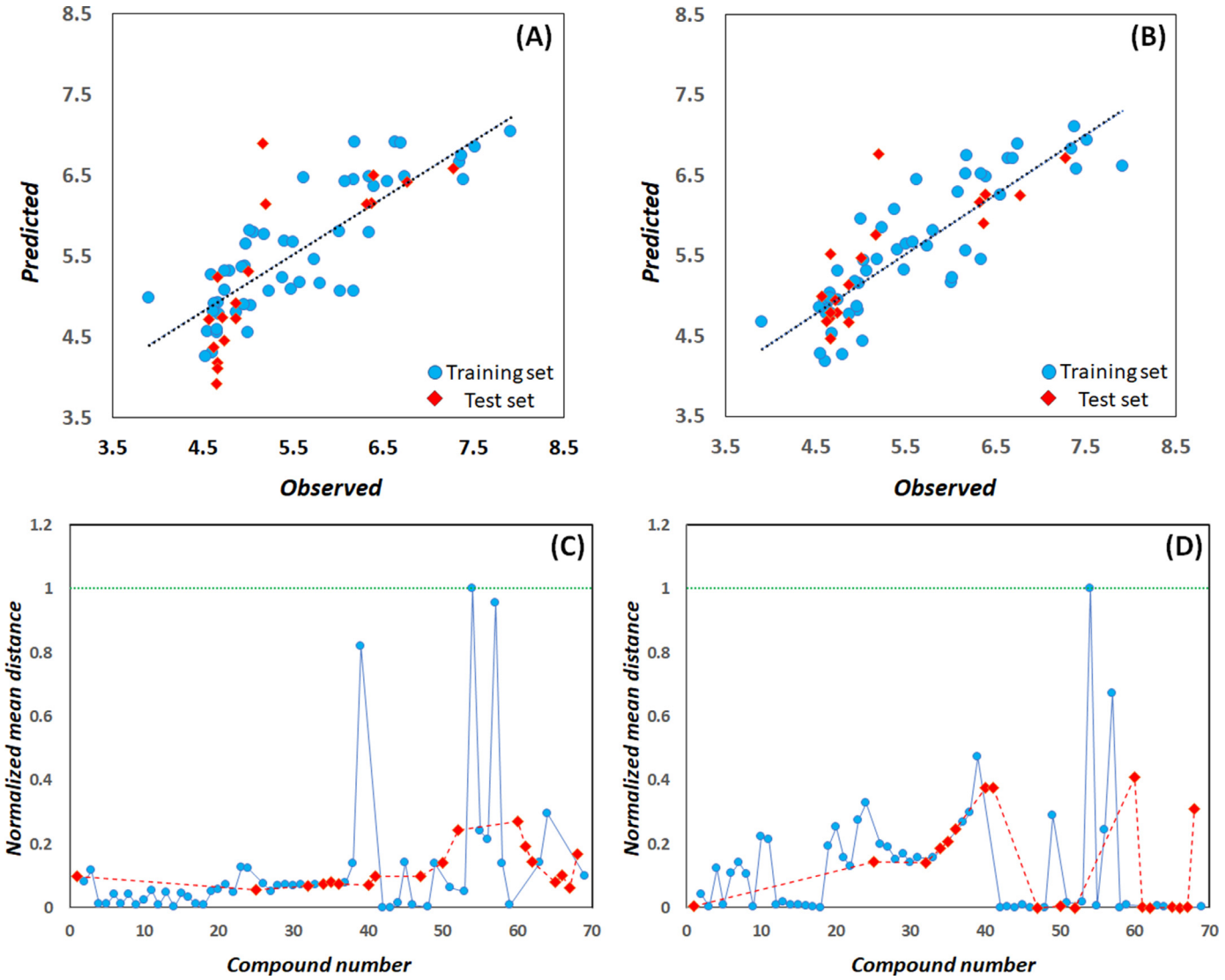

Fig. 1. (A) The observed *vs* predicted activity of the training and test sets as per the SW-MLR model; (B) The observed *vs* predicted activity of the training and test sets as per the GA-MLR (C) Normalized mean distance values *vs* compound numbers as per the SW-MLR model. (D) Normalized mean distance values *vs* compound numbers as per the GA-MLR model.

bonds of path length 10. Several compounds of this dataset having higher activity display higher values of this descriptor (Compounds **11, 19–22, 25–26, 30–31**, and **33**). These compounds contain a 2–2-oxopyrrolidinyl methyl ring associated with other functional groups (such as amide, carbonyl) capable of forming potential hydrogen bonding interactions. A large number of compounds devoid of such structural feature possess a zero value of *maxHBint10* and hence, are lower active SARS-CoV-2 $3CL^{pro}$ inhibitors (Compounds **1, 4–10, 12–18, 41–47, 50–52, 58**, and **61–66**). This result supports our earlier observation [26] where it was noticed from the robust classification-dependent analysis (namely, linear discriminant analysis, SARpy analysis, Bayesian classification technique, and recursive partitioning) that compounds possessing 2-oxopyrrolidine moiety with adjacent electronegative groups are crucial for imparting higher SARS-CoV-2 $3CL^{pro}$ inhibition. The crystallographic data also showed that the 2-oxopyrrolidine ring enters the S1 subsite along with forming favorable hydrogen bonding interactions with Phe140, His164, and Glu166.

Table 2
Golbraikh and Tropsha criteria's for the constructed MLR models.

Parameter	*Threshold*	*Eq.1*	*Eq.2*	*Eq.3*
Q^2	$Q^2>0.5$	0.636	0.674	0.655
r^2	$r^2>0.6$	0.627	0.649	0.918
r_0^2-$r_0'^2$	$\|r0^2-r'0^2\| < 0.3$	0.089	0.154	0.070
k	$0.85 < k < 1.15$	0.991	0.974	0.975
k'	$0.85 < k' < 1.15$	0.997	1.018	1.023
$(r^2-r_0^2)/r^2$	$[(r^2-r0^2)/r^2] < 0.1$	0.160	0.000	0.033
$(r^2-r_0'^2)/r^2$	$[(r^2-r'0^2)/r^2] < 0.1$	0.018	0.238	0.1098

r^2, Squared correlation coefficient between observed vs predicted response of the test set compounds; r_0^2, The values for regression through the origin (observed *vs* predicted); $r_0'^2$, The values for regression through the origin (predicted *vs* observed); k, Slope of the regression lines through the origin for observed *vs* predicted; k', Slope of the regression lines through the origin for predicted *vs* observed.

Similar to the SW-MLR model (***Eq. (1)***), the GA-MLR model (***Eq. (2)***) also displays the positive impact of *maxHBint10* on SARS-CoV-2 3CL^{pro} inhibition. Apart from that, four other descriptors, i.e., *maxHBint8, FMF, MDEO-22*, and *MATS3e* (Table 1) have a direct correlation with the biological activity. The descriptor *maxHBint8* denotes the maximum E-state descriptor of strength for potential hydrogen bonds of path length 8. The positive contribution of *maxHBint8* implies that compounds having higher values of this descriptor favor SARS-CoV-2 3CL^{pro} inhibition (Compounds **11, 19–22, 25–26**, and **30–33**). All these compounds exert the positive impact of the 2-oxopyrrolidinyl methyl group along with associated electronegative functions (such as carbonyl, amide, and hydroxy) for forming effective hydrogen bonding interactions at the enzyme active site. On the other hand, similar to the former descriptor, i.e., *maxHBint10*, a number of compounds having zero value of this descriptor are devoid of such feature and therefore, are less active SARS-CoV-2 3CL^{pro} inhibitors. In the GA-MLR model (***Eq. (2)***), the descriptor *FMF* has a positive contribution towards biological activity. The descriptor *FMF* denotes the complexity of a molecule. The positive contribution of *FMF* points out that molecules with higher complexity values are higher active than molecules with lower complexity. As per the values of this descriptor, several compounds (Compounds **11, 19–22, 25–26**) have higher values of this parameter. These compounds contain aryl (benzyloxy, benzyl) and heteroaryl (2-oxopyrrolidinyl, indolyl) rings in their molecular structure. However, several compounds possess a value of 0 (Compounds **6, 8, 13**) of this descriptor or lower values of this descriptor (Compounds **4, 7, 10**, and **40**). Compounds **6, 8**, and **13** bear a linear structure with no aryl/heteroaryl functions whereas compounds **4, 7, 10**, and **40** contain only one phenyl ring in their structure. Therefore, it may be inferred that compounds with a higher number of aryl/heteroaryl ring structures are better effective compared to compounds with no aryl/heteroaryl groups as well as compounds with a lower number of aryl/heteroaryl ring structures. On the other hand, *MDEO-22* has a negative influence on the SARS-CoV-2 3CL^{pro} inhibitory activity. It denotes the distance edge between all secondary oxygen atoms. Compounds with a higher number of esters and ether functions display higher values of this descriptor (Compounds **1, 38, 40, 49, 54, 57**, and **68**) and therefore, are lower active compounds. On the other hand, the higher active molecules do not possess such types of features. *MATS3e*, an autocorrelation descriptor, denotes Moran autocorrelation-lag 3/weighted by Sanderson electronegativities. It is correlated positively with the SARS-CoV-2 3CL^{pro} inhibitory activity. Several molecules possess higher values of this descriptor due to the presence of higher electronegative functionalities at their structures (Compounds **11, 20, 22, 25, 31**, and **33**). These electronegative functionalities are responsible for forming favorable hydrogen-bonding interactions with the active site amino acid residues. On the other hand, a large number of molecules contain negative values of *MATS3e* descriptor and therefore, are lower active SARS-CoV-2 3CL^{pro} inhibitors.

3.1.3. Analysis to extract the key structural features

As both these SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models reveal the importance of different structural features responsible for regulating SARS-CoV-2 3CL^{pro} inhibition, another SW-MLR model (***Eq. (3)***) was tried with well-known physicochemical parameters (namely *ClogP, CMR*, dipole moments along X, Y and Z axis, total dipole moment, polar surface area, polar volume, surface area, and volume) calculated with the help of the SYBYL-X 2.0 software [55] along with several indicator variables to extract out as well as to validate the crucial features required for modulating the SARS-CoV-2 3CL^{pro} inhibitory potency of these molecules.

$$pIC_{50} = 4.999(\pm0.094) + 1.555(\pm0.167)I_{2-oxo-pyr} + 0.542(\pm0.185)I_{Benzylester} \quad \text{(Eq. 3)}$$

$N_{train}=51, N_{test}=18, R=0.833, R^2=0.695, R_A^2=0.682, Q^2=0.655, SEE=0.533, PRESS=13.637, F(2,48)=54.738, Avg.r_{mLOO}^2=0.536, cR_p^2=0.679, R_{pred}^2=0.874; Avg.r_{mtest}^2=0.648, p<0.05.$

Constructed with two indicator variables ($I_{2\text{-}oxo\text{-}pyr}$ and $I_{Benzylester}$),***Eq. (3)*** was able to explain 68.2% and predict 65.5% activity variation of the dataset molecules. $I_{2\text{-}oxo\text{-}pyr}$ and $I_{Benzylester}$ denote the presence or absence of the 2-oxopyrrolidinyl methyl group and benzyl acetate amide function, respectively. It was observed from the model (***Eq. (3)***) that the presence of the 2-oxopyrrolidinyl methyl group and the benzyl acetate amide function have a positive impact on SARS-CoV-2 3CL^{pro} inhibition. The observed versus predicted activity plot for ***Eq. (3)*** is given in Fig. 2A.

The AD of the ***Eq. (3)*** was also tested similarly to the SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models (Fig. 2B). The Golbraikh and Tropsha model acceptability parameters [42] for ***Eq. (3)*** are provided in Table 1. Additional information regarding ***Eq. (3)*** is provided in Supplementary Tables S10-S12.

Several compounds (Compounds **4–5, 7, 9–11, 20–21, 27, 40**, and **42**) are found to possess the benzyl acetate amide function in their molecular structure and therefore, are effective SARS-CoV-2 3CL^{pro} inhibitors. On the other hand, compounds bearing the 2-oxopyrrolidinyl methyl group (Compounds **11, 19–37**) are also potent SARS-CoV-2 3CL^{pro} inhibitors. However, it was interesting to note that compounds containing both these functional moieties are the most potent compounds in this series (Compounds **19–21**). Therefore, both these functional moieties must be taken into consideration during designing highly potent SARS-CoV-2 3CL^{pro} inhibitors.

3.2. Non-linear QSAR model development

3.2.1. Artificial neural network (ANN) model

Here, the descriptors used to construct the SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models were employed in the generation of the ANN models (SW-ANN and GA-ANN). To achieve the optimal parameters for ANN model development, an extensive search regarding the different ANN model parameters, i.e., the number of nodes in the hidden layer, learning rate, momentum, and the number of learning epochs was performed using the Autoweka software [48].

For both the SW-ANN (***Eq. (1)***) and GA-ANN (***Eq. (2)***) models, the parameter search provided an optimal learning epoch of 0.1. Besides, for the SW-ANN model, the parameter search provided an optimal number of hidden layer nodes of 03 with an optimal training time of 300 (Fig. 3A and 3C). Also, an optimal number of nodes

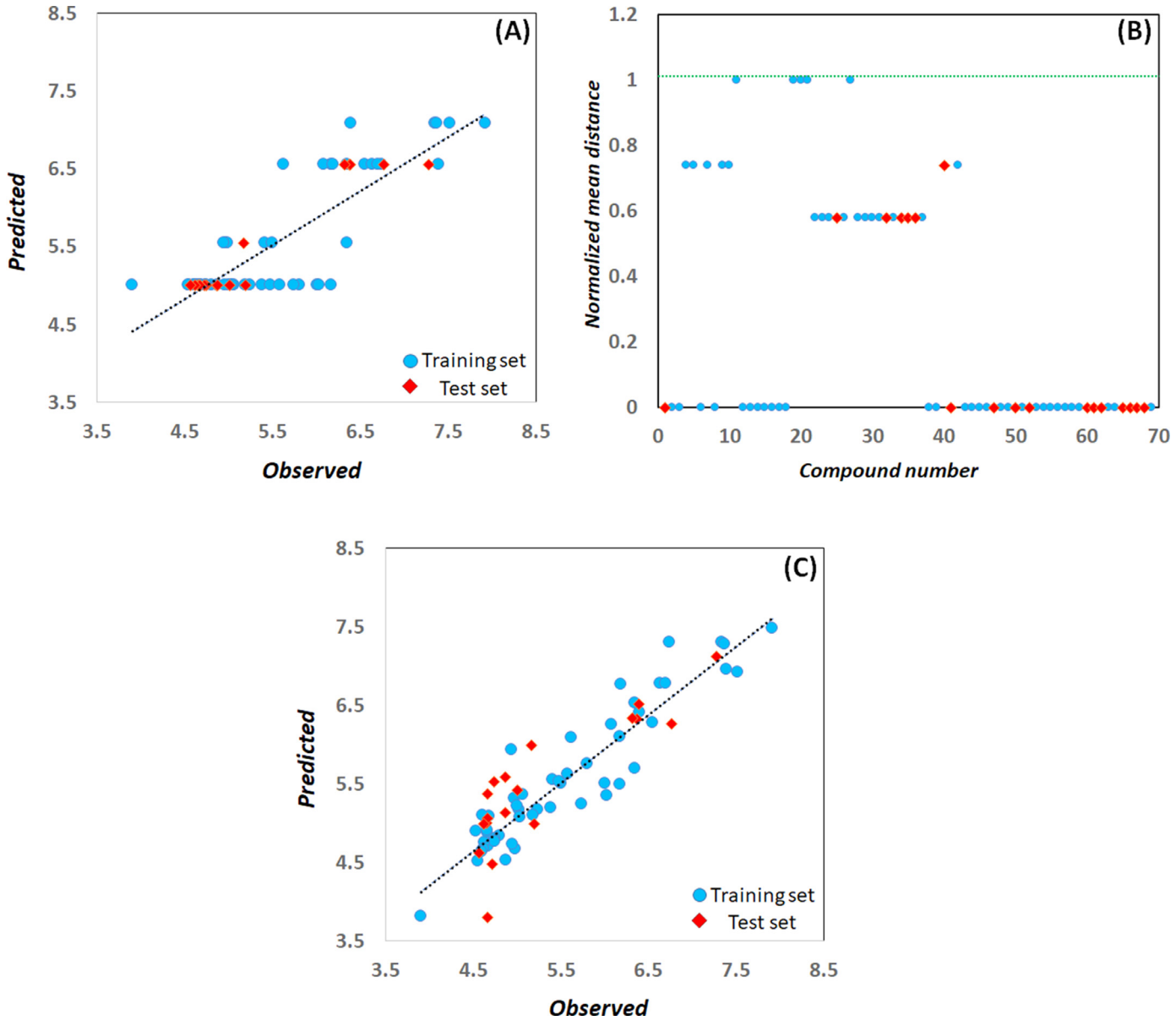

Fig. 2. (A) The observed *vs* predicted activity of the training and test sets as per the SW-MLR model; (B) Normalized mean distance values *vs* compound numbers as per the SW-MLR model; (C) The observed *vs* predicted activity of the training and test sets as per the HQSAR model (model **47B**).

in the hidden layer was found to be 01 for the GA-ANN model along with an optimal learning time of 100 (Fig. 3B and 3D). Additionally, the predictability of the ANN models has been depicted in Table 3.

From the perspective of the model performance, the SW-ANN model delivered a correlation coefficient (R) and LOO-cross-validated correlation coefficient (R_{CV}) value of 0.872 and 0.779, respectively. For the test set, the SW-ANN model exhibited an R_{Pred} value of 0.745. On the other hand, the GA-ANN model provided an R and R_{CV} value of 0.865 and 0.833, respectively. The GA-ANN model also exhibited an R_{Pred} value of 0.814 for the test set population. The observed versus predicted activity of the ANN models has been depicted in (Fig. 4A and 4B).

3.2.2. Support vector machine (SVM) model

Similar to the ANN models, the SW-SVM and GA-SVM models were also constructed by using the descriptors of the SW-MLR and GA-MLR models. The optimal SVM parameters, i.e., the kernel width (γ) and complexity (C) was also done using Autoweka software [48] by grid searching (Fig. 4C and 4D) and exponential adjustments of C and γ values (Values ranging from −19 to 19) with an incremental step of 2.0. The SVM parameter search showed an optimal C of 128.0 and 0.125 along with γ of 0.0078125 and 8.0 for the SW-SVM and GA-SVM models, respectively.

The SW-SVM model showed an R and R_{CV} of 0.839 and 0.828 along with an R_{Pred} value of 0.812 for the test set. Also, the GA-SVM model delivered an R value of 0.913 with an R_{CV} of 0.828 while exhibiting an R_{Pred} value of 0.814 for the test set (Table 3). The observed versus predicted activity of the ANN models has been depicted in (Fig. 4C and 4D). Also, the predicted activities from the SW-ANN and SW-SVM models are given in Supplementary Table S2 whereas the predicted activities from the GA-ANN and GA-SVM models are given in Supplementary Table S4.

3.3. Hologram QSAR (HQSAR) model

Regarding the HQSAR analysis, all the probable model combinations were investigated by using different fragment distinction

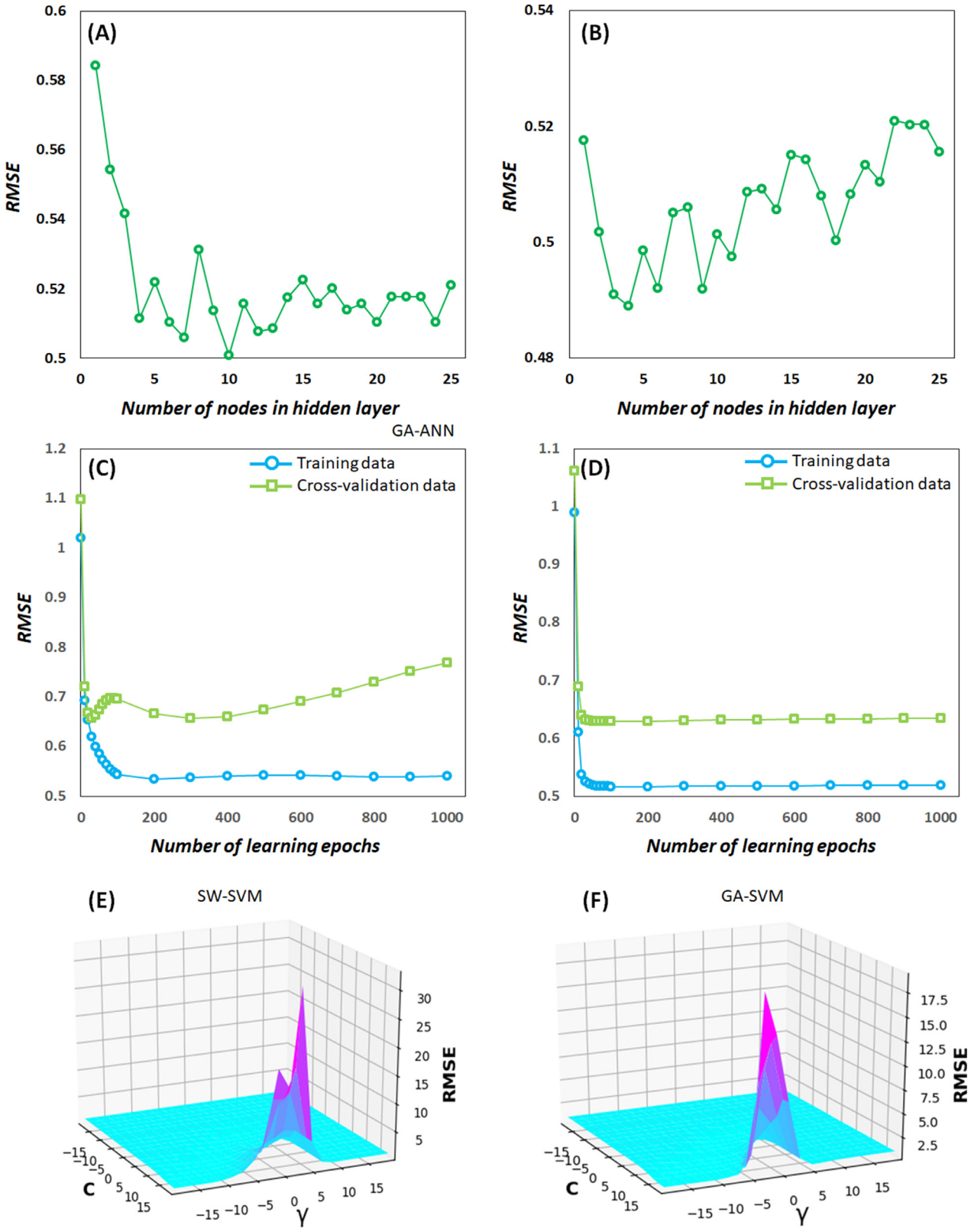

Fig. 3. (A) ANN parameter optimization involving the number of hidden nodes for the SW-ANN model; (B) ANN parameter optimization involving the number of hidden nodes for the GA-ANN model; (C) ANN parameter optimization involving learning epochs for the SW-ANN model; (D) ANN parameter optimization involving learning epochs for the GA-ANN model; (E) SVM parameter optimization involving a grid search for the SW-SVM model; (F) SVM parameter optimization involving a grid search for the GA-SVM model.

Table 3
Statistical parameters of the constructed ANN and SVM models.

Parameters	*SW-ANN*			*GA-ANN*		
	Training set	*LOO–CV*	*Test set*	*Training set*	*LOO–CV*	*Test set*
R	0.872	0.779	0.745	0.839	0.828	0.812
RMSE	0.462	0.605	0.647	0.527	0.540	0.509
MAE	0.377	0.475	0.419	0.390	0.409	0.370
Parameters	*SW-SVM*			*GA-SVM*		
	Training set	*LOO–CV*	*Test set*	*Training set*	*LOO–CV*	*Test set*
R	0.865	0.833	0.789	0.913	0.828	0.814
RMSE	0.476	0.521	0.536	0.397	0.532	0.499
MAE	0.349	0.400	0.368	0.255	0.422	0.387

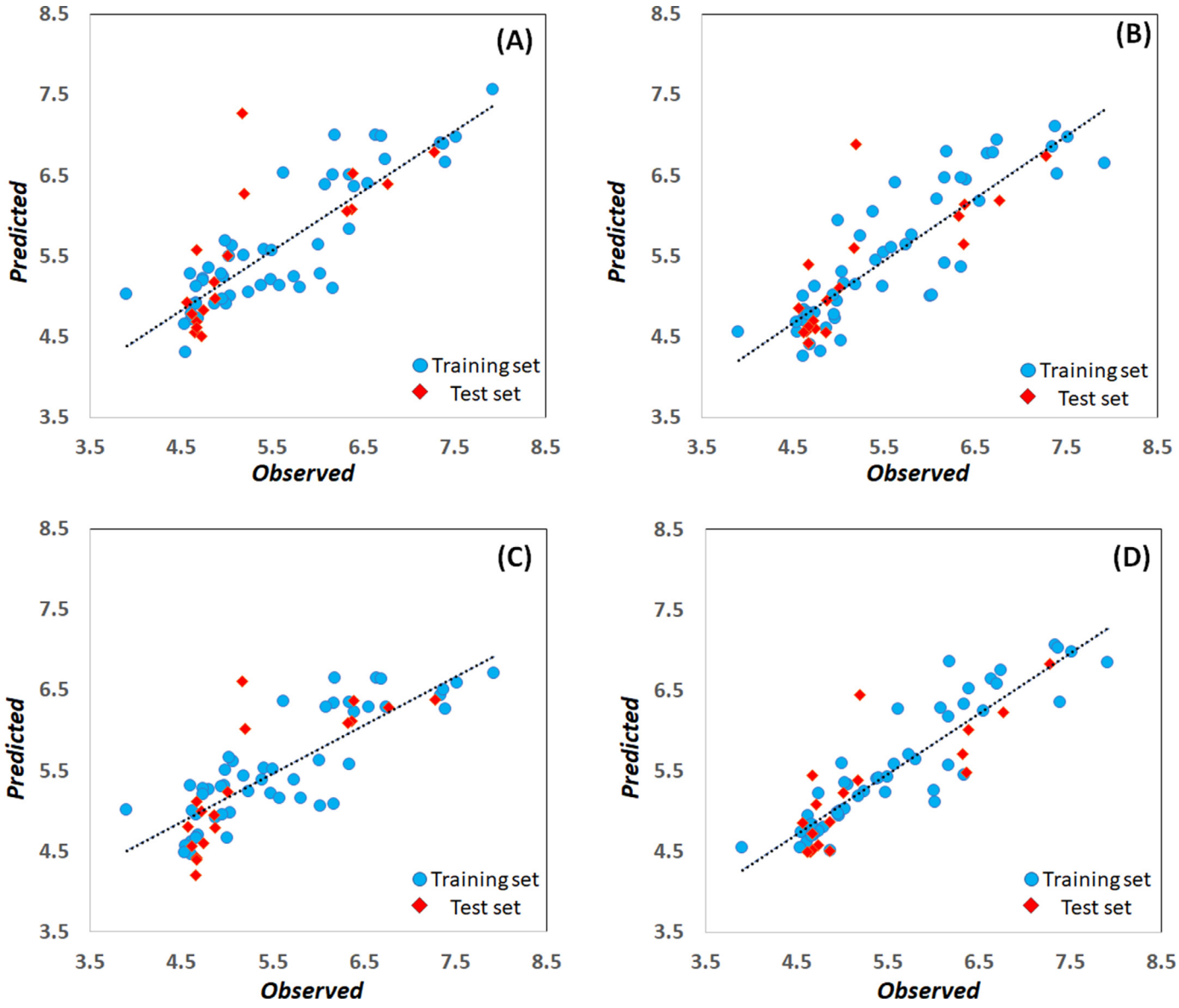

Fig. 4. (A) The observed *vs* predicted activity of the training and test sets as per the SW-ANN model; (B) The observed *vs* predicted activity of the training and test sets as per the GA-ANN; (C) The observed *vs* predicted activity of the training and test sets as per the SW-SVM model; (D) The observed *vs* predicted activity of the training and test sets as per the GA-SVM model.

Table 4
HQSAR models constructed using different fragment distinction parameters.

Model	Fragment distinction	Q^2	R^2	*SE*	Length	Component
1	A	0.301	0.633	0.604	199	5
2	B	0.358	0.516	0.686	61	4
3	C	0.372	0.726	0.528	59	6
4	H	0.184	0.352	0.794	401	4
5	Ch	0.184	0.352	0.794	401	4
6	DA	0.303	0.574	0.643	401	4
7	A/B	0.411	0.695	0.550	401	5
8	A/C	0.339	0.810	0.440	401	6
9	A/H	0.451	0.772	0.482	307	6
10	A/Ch	0.368	0.649	0.585	83	4
11	A/DA	0.384	0.581	0.624	353	3
12	B/C	0.442	0.689	0.55	151	4
13	B/H	0.258	0.516	0.686	61	4
14	B/Ch	0.258	0.516	0.686	61	4
15	B/DA	0.335	0.737	0.512	307	5
16	C/H	0.372	0.726	0.528	59	6
17	C/Ch	0.372	0.726	0.528	59	6
18	C/DA	0.300	0.748	0.501	257	5
19	H/Ch	0.184	0.352	0.794	401	4
20	H/DA	0.303	0.574	0.643	401	4
21	Ch/DA	0.279	0.551	0.661	401	4
22	A/B/C	0.450	0.742	−0.507	257	5
23	A/B/H	0.515	0.798	0.453	307	6
24	A/B/Ch	0.403	0.696	0.550	401	5
25	A/B/DA	0.426	0.745	0.504	401	5
26	B/C/H	0.442	0.689	0.55	151	4
27	B/C/Ch	0.442	0.689	0.55	151	4
28	B/C/DA	0.34	0.706	0.535	353	4
29	C/H/Ch	0.372	0.726	0.528	59	6
30	C/Ch/DA	0.312	0.727	0.521	257	5
31	A/B/C/H	0.580	0.805	0.445	199	6
32	A/B/C/Ch	0.402	0.671	0.566	257	4
33	A/B/C/DA	0.355	0.789	0.459	83	5
34	A/B/H/Ch	0.515	0.836	0.408	307	6
35	A/B/H/DA	0.586	0.831	0.414	71	6
36	A/B/Ch/DA	0.391	0.753	0.495	401	5
37	A/C/H/Ch	0.551	0.817	0.432	83	6
38	A/C/H/DA	0.564	0.829	0.413	257	5
39	A/C/Ch/DA	0.228	0.575	0.636	53	3
40	A/H/Ch/DA	0.439	0.739	0.509	61	5
41	B/C/H/Ch	0.442	0.689	0.550	151	4
42	B/C/H/DA	0.340	0.706	0.535	353	4
43	B/H/Ch/DA	0.372	0.732	0.516	307	5
44	C/H/Ch/DA	0.312	0.727	0.521	257	5
45	A/B/C/H/Ch	0.595	0.824	0.423	53	6
46	A/B/C/Ch/DA	0.265	0.700	0.540	257	4
47	**A/B/C/H/DA**	**0.628**	**0.843**	**0.400**	**257**	**6**
48	A/B/H/Ch/DA	0.614	0.801	0.450	71	6
49	B/C/H/Ch/DA	0.319	0.693	0.547	353	4
50	A/B/C/H/Ch/DA	0.530	0.832	0.413	353	6

The best HQSAR model (Model 47) is shown in bold face.

parameters (A, B, C, H, Ch, and DA) to generate 50 HQSAR models (Table 4).

Among these 50 models, 10 models were found to fulfill the internal cross-validation criteria as their $Q^2 > 0.50$ (model 23, 31, 34, 35, 37, 38, 45, 47, 48 and 50). Among these 10 HQSAR models, the best model (model 47) was taken into account depending on the highest Q^2 value (0.628) and the lowest *SE* value (0.40).

The best model (model **47**) was optimized further with the help of different atom counts to investigate whether a higher Q^2 value would be achieved or not. It was observed that six out of seven HQSAR models (model **47A**-**47F**) fulfilled the Q^2 criteria ($Q^2 > 0.50$). Among these models, model **47B** (atom count: 2–5, Length: 353, Component: 6) exhibited the highest Q^2 value (0.670) and the lowest *SE* value (0.367) (Table 5). The observed and predicted activity values of the molecules as per the best HQSAR model (model **47B**) are provided in Supplementary Table S13. The observed versus predicted activity from the HQSAR model (model **47B**) has been provided in Fig. 2C.

Table 5
Sub-models of HQSAR model **47** constructed using different atom counts.

Model	Atom Count	Q^2	R^2	*SE*	Length	Component
47A	1 to 4	0.644	0.87	0.363	353	6
47B*	**2 to 5**	**0.67**	**0.867**	**0.367**	**353**	**6**
47C	3 to 6	0.622	0.844	0.399	353	6
47D	4 to 7	0.628	0.843	0.400	257	6
47E	5 to 8	0.609	0.827	0.419	257	6
47F	6 to 9	0.53	0.805	0.440	71	5
47G	7 to 10	0.369	0.67	0.560	71	3

The best HQSAR model (Model **47B**) is shown in bold face ($^*R^2_{Pred} = 0.715$).

3.3.1. Interpretation of HQSAR model

According to the best HQSAR model (model **47B**), the importance of the good and bad fragments of the molecules can be obtained. The good fragments of these compounds are shown in green and blue-green colors, whereas the bad fragments are shown in red and orange-red colors. However, the white-colored frag-

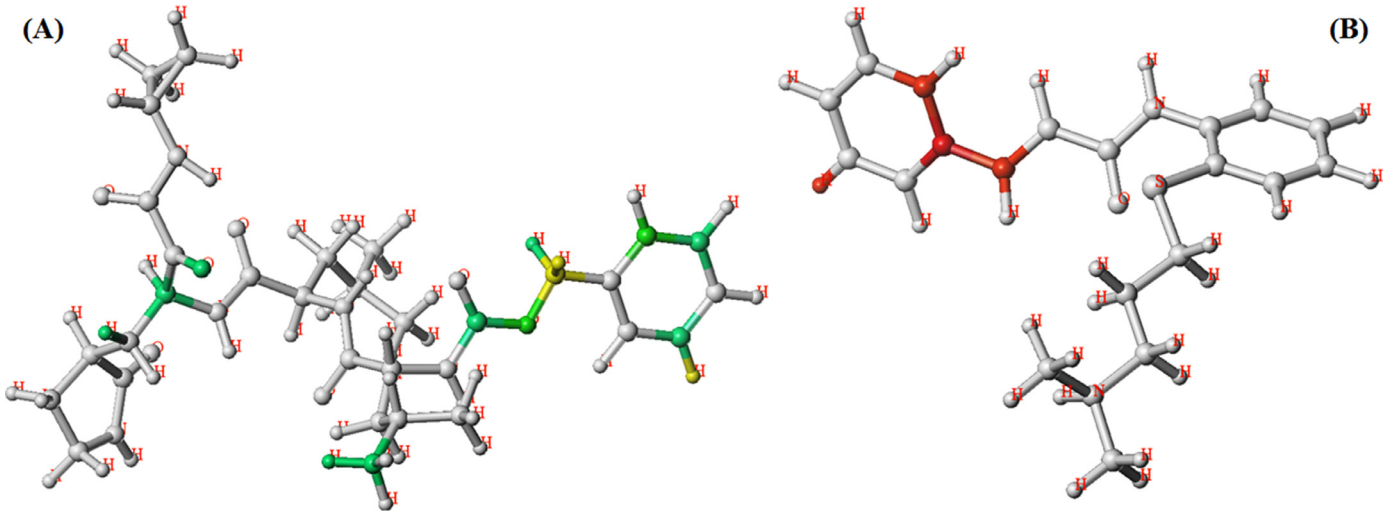

Fig. 5. Contour map of the (A) compound **21** and (B) compound **18** was obtained from the HQSAR study (bad=> red: less than −0.1939675; red-orange: −0.1939675 to −0.1163805; orange: −0.1163805 to −0.077587; white: −0.077587 to 0.073580667; yellow: 0.073580667 to 0.110371; green-blue: 0.110371 to 0.18395167; good-> green: 0.18395167 and above).

ments display moderate contributions towards biological activity. The best active compound (Compound **21**) shows the significance of the good fragments as depicted in Fig. 5A.

Three carbon atoms at the terminal phenyl group along with the adjacent carboxamido ester moiety of compound **21** display their positive contributions towards SARS-CoV-2 3CL^{pro} inhibition as evidenced by the green (0.18395167 and above) and green-blue (0.110371 to 0.18395167) fragments. Moreover, the methylene group attached with the phenyl ring shows a good contribution as depicted by the yellow (0.073580667 to 0.110371) fragments (Fig. 5A). Again, the chiral carbon atom where the 2-oxopyrrolidinyl methyl group is attached shows good contributions. Moreover, the adjacent carbonyl oxygen atom shows a good contribution towards SARS-CoV-2 3CL^{pro} inhibitory activity. One of the terminal methyl carbon atoms of the *i*–butyl moiety, adjacent to the benzyloxy carboxamido function, projects its good contribution towards SARS-CoV-2 3CL^{pro} inhibition. The positive influence of the benzyl and benzyloxy carboxamido functions on the biological activity is also noticed for several other molecules (Compounds **5, 7, 19–20, 40**, and **43**). This result is in agreement with our earlier observation [30] where it was observed from the crystallographic data analysis that the *i*–butyl group enters the S2 subsite of the SARS-CoV-2 3CL^{pro} enzyme surrounded by the amino acid residues His41, Met49, and Met169. Apart from that, the caboxybenzyl group occupies the S4 pocket, whereas the carboxy group can form potential hydrogen bonding interaction.

The other potent compounds of this series (Compounds **9, 11, 19–20, 26–27**, and **37**) only show the importance of several good fragments but do not show any bad fragments. Compound **11** displays the importance of the 2-oxopyrrolidinyl methyl function as well as the adjacent methylene (hydroxy) sulphonic acid warhead. Similar observations are also noticed for other potent SARS-CoV-2 3CL^{pro} inhibitors (Compounds **29, 31, 33**, and **35**). Many of the dataset compounds having moderate to the less SARS-CoV-2 3CL^{pro} inhibitory activity show neither good nor bad fragments, but they exhibit the influence of moderate fragments towards the SARS-CoV-2 3CL^{pro} inhibition (Compounds **14–17, 39, 41, 45, 47–48, 50–54, 56–59, 61, 63–64** and **66–69**).

Regarding the least active molecule (Compound **18**), red (−0.1939675 to −0.1163805) fragments at the phenyl ring and the adjacent unsaturated carbon atom have been observed (Fig. 5B). Apart from that, many of these compounds (Compounds **10, 23, 44, 46, 60**, and **65**) exhibit red/red-orange (less than −0.1939675) fragments which show their negative influence on SARS-CoV-2 3CL^{pro} inhibition.

4. External validation of QSAR models by repurposed drugs

Several FDA-approved drugs and other compounds which have already been repurposed by several groups of researchers [3,56–76] were considered here to judge whether our QSAR models can explain/predict these molecules or not.

A number of compounds were found to be highly predicted (IC_{50} < 1 μM) by our SW-MLR (***Eq. (1)***) and GA-MLR (***Eq. (2)***) models. Among these compounds, five compounds were predicted as higher active SARS-CoV-2 3CL^{pro} inhibitors by both of these models (Figs. 6 and 7).

As per the best MLR model (i.e., GA-MLR model), curcumin was predicted as the most active SARS-CoV-2 3CL^{pro} inhibitor (Predicted IC_{50} = 93.52 nM) whereas by the SW-MLR model (***Eq. (1)***) curcumin resulted in a predicted IC_{50} of 848.47 nM. Ribavirin was the second most active molecule predicted by the GA-MLR model (***Eq. (2)***) (Predicted IC_{50} = 129.89 nM) whereas the predicted IC_{50} by the SW-MLR model (***Eq. (1)***) was 411.34 nM. Saquinavir was the third most effective molecule predicted by the GA-MLR model (***Eq.2***) (predicted IC_{50} = 135.21 nM) and SW-MLR model (***Eq. (1)***) (predicted IC_{50} = 564.06 nM). Apart from these molecules, sepimostat (predicted IC_{50} = 643.57 nM) and remdesivir (predicted IC_{50} = 677.75 nM) were predicted well by GA-MLR models.

However, in case of HQSAR model (model **47b**), the best predicted compounds (predicted IC_{50}<1 μM) are completely different (Fig. 7). The best five compounds predicted well by HQSAR model are lurasidone (predicted IC_{50} = 63.53 nM), saquinavir (predicted IC_{50} = 123.03 nM), lopinavir (predicted IC_{50} = 246.60 nM), elbasvir (predicted IC_{50} = 456.04 nM) and paritaprevir (predicted IC_{50} = 538.27 nM).

Interestingly, saquinavir is the only molecule predicted well by all three models. Sepimostat is another molecule predicted slightly lower (IC_{50} = 1702 nM) by the HQSAR model compared to the SW-MLR and GA-MLR model (***Eq.1*** and ***Eq.2***, respectively).

Various molecular docking studies have already been performed with curcumin and SARS-CoV-2 3CL^{pro} enzyme [74,77–81]. Khaerunnisa et al. [82] showed that curcumin displayed strong binding energy (ΔG = −7.05 kcal/mol) with the SARS-CoV-2 3CL^{pro} enzyme (PDB: 6LU7). It forms potential hydrogen bond interac-

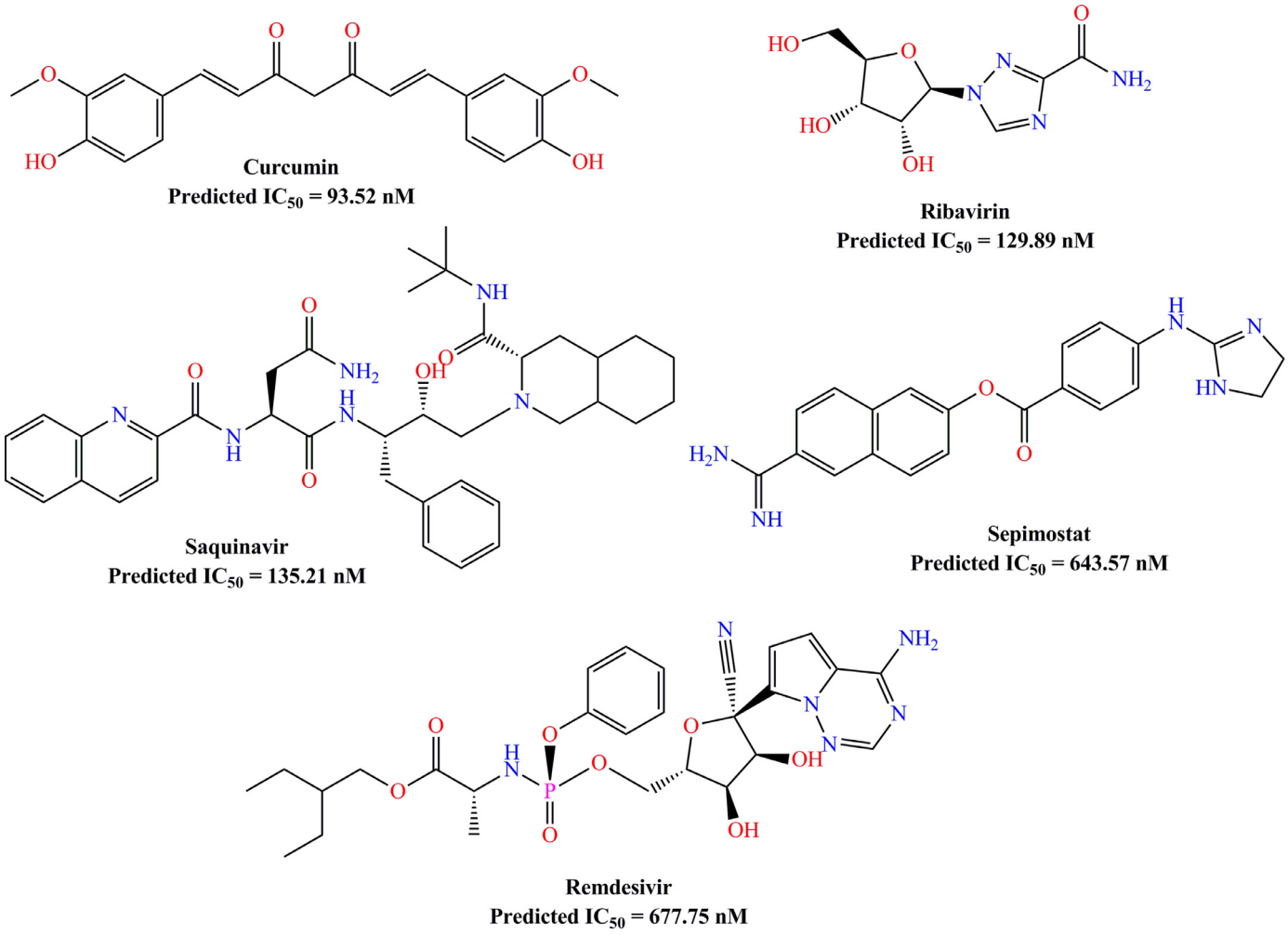

Fig. 6. The best predicted FDA approved drugs by GA-MLR model.

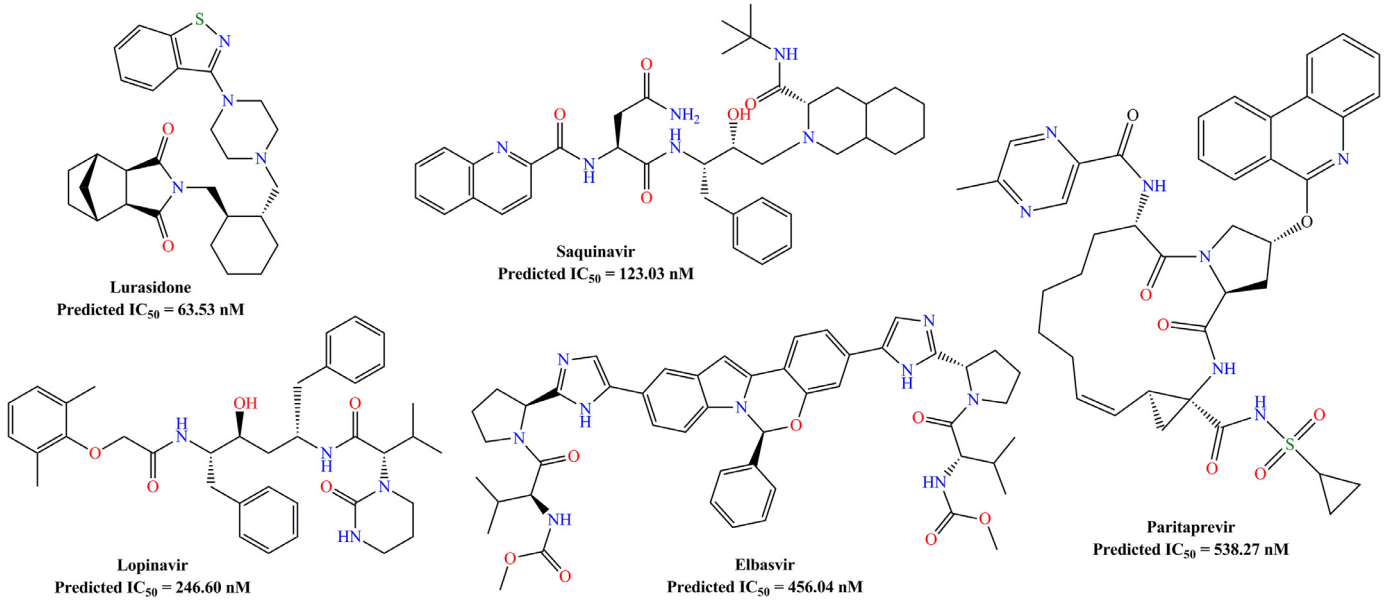

Fig. 7. The best predicted FDA approved drugs by HQSAR model.

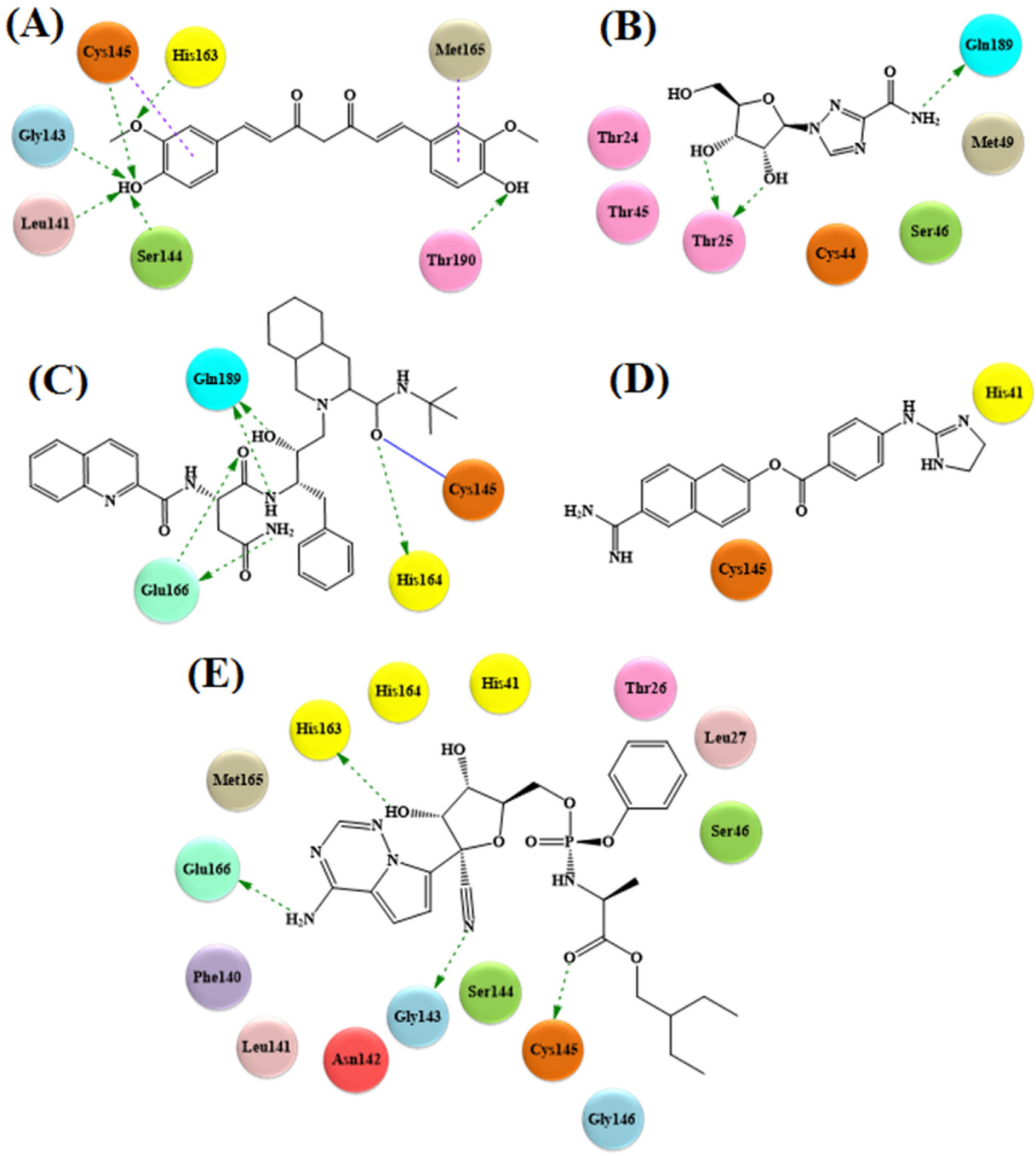

Fig. 8. Molecular docking interaction of (A) Curcumin, (B) Ribavirin, (C) Saquinavir, (D) Sepimostat, and (E) Remdesivir with SARS-CoV-2 $3CL^{pro}$ enzyme (Hydrogen bonding interaction: green dotted arrow; π-sulfur interaction: violet dotted line; covalent interaction: blue line).

tion with Cys145, Leu141, Gly143, Ser144, and Thr190 and π-sulfur interaction with Met165. The results are also in agreement with the observation of Ibrahim and co-workers [78] showing a docking score of −9.2 kcal/mol (Fig. 8A).

The molecular dynamics (MD) simulation and MM-GBSA binding energy calculation revealed that curcumin strongly binds to the active site ($\Delta G = -34.2$ kcal/mol) with stable binding interaction during 40 ns MD simulation run.

On the other hand, the second-highest active repurposed drug ribavirin displayed good docking interaction (relative docking score = 2.01, relative ligand efficiency = 3.21, relative glide lipo = 0.37, and relative glide Hbond = 4.36) at the SARS-CoV-2 $3CL^{pro}$ active site [74]. It formed potential hydrogen bonding interaction with the backbone amino acid residues Thr25 and Gln189 (Fig. 8B). Deshpande et al. [83] also showed that ribavirin was effectively bound to the active site of SARS-CoV-2 $3CL^{pro}$ (PDB: 6Y84). Again, Gupta et al. [84] demonstrated that ribavirin strongly binds to SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7) and produced favorable hydrogen bonding with Leu141, Gly143, Arg188, and Gln189. The result was slightly varied in the case of the results obtained by Kumar et al. [85]. It displayed a docking score of −6.813 and binding energy of −35.63 kcal/mol during docking with SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7). Moreover, it showed potential hydrogen bonding with His164, Glu166, Gln189, and Thr190 at the active site.

The 3rd highest active compound saquinavir was studied through molecular docking analysis by different groups of re-

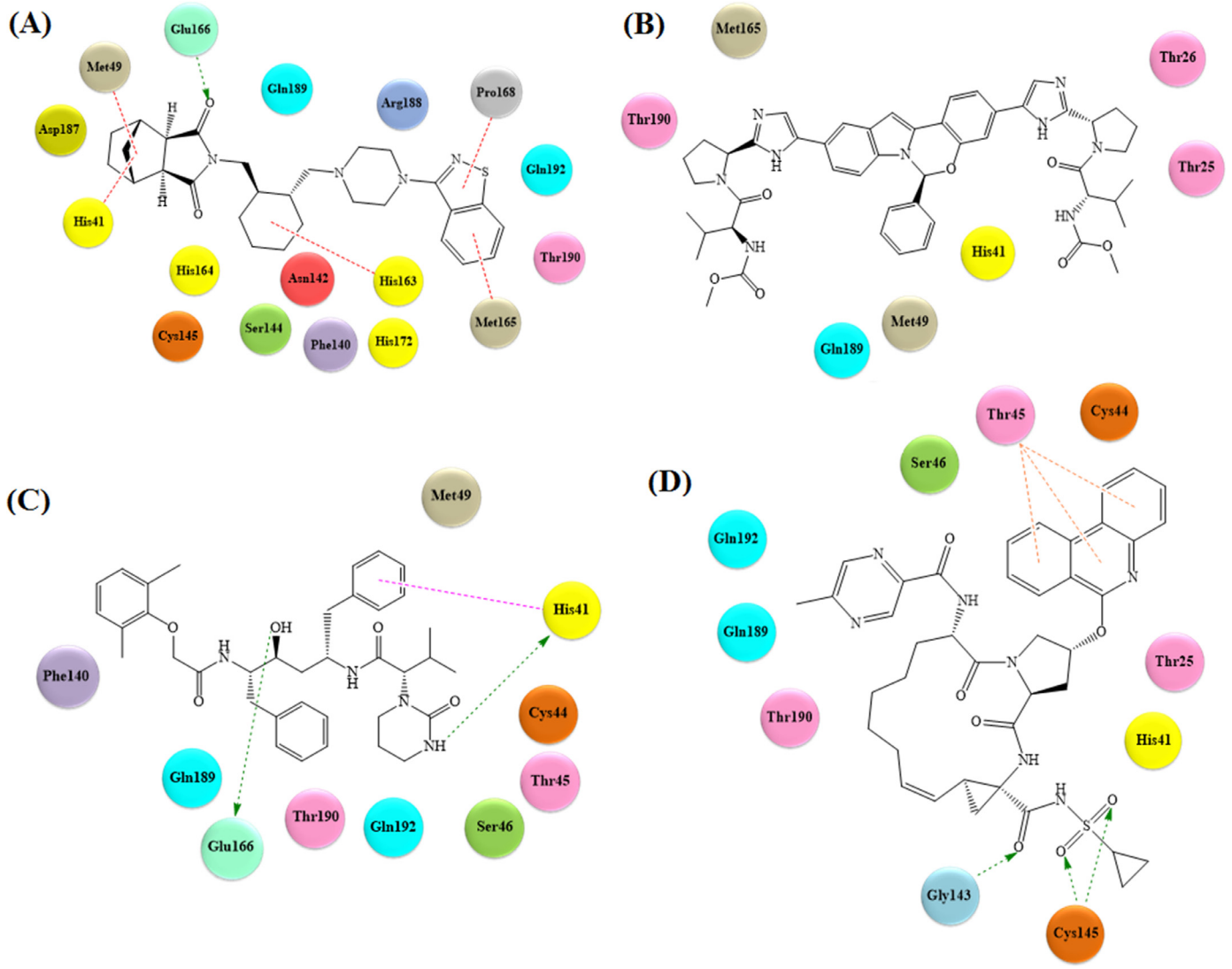

Fig. 9. Molecular docking interaction of (A) Lurasidone, (B) Elbasvir, (C) Lopinavir, and (D) Paritaprevir with SARS-CoV-2 $3CL^{pro}$ enzyme (Hydrogen bonding interaction: green dotted arrow; π-π stacking interaction: magenta dotted line; π-alkyl interaction: red dotted line; amide-π stacking: orange dotted line).

searchers [58,63,86–94]. Hall et al. [86] reported a docking score of −7.285 kcal/mol while binding to the SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7). Again, Talluri et al. [87] showed that saquinavir had a binding affinity of −9.2 kcal/mol during the active site binding (PDB: 6LU7). It formed hydrogen bonding interactions with His163, His164, Gly143, Ser144, Cys145, and Glu166. Saquinavir also showed similar binding energy (−9.6 kcal/mol) while binding to SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7) that was noticed by the molecular docking study conducted by Ortega et al. [88]. Raphael et al. [58] examined that saquinavir had binding energy of −9.0 kcal/mol with SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7) making several hydrogen bonding interactions with Glu166, His41, His164, Gln189 along with several π-alkyl interactions.

Al-Khafaji et al. [63] showed that saquinavir, while docked into SARS-CoV-2 $3CL^{pro}$ enzyme (PDB: 6LU7) displayed the highest docking score (−9.856 kcal/mol) and the lowest MM-GBSA binding energy (−72.17 kcal/mol). Saquinavir formed five hydrogen bonds with Glu166, Gln189, His164 as well as covalent bonding with Cys145 (Fig. 8C). Nevertheless, the molecular dynamics (MD) simulation study revealed that saquinavir has an average RMSD of 0.0186 as well as lower fluctuation and the binding became stable at 50 ns.

Sepimostat was the 4^{th} highest predicted molecule as per the GA-MLR model. Tsuji et al. [64] showed that sepimostat effectively fit into the active site of SARS-CoV-2 $3CL^{pro}$ enzyme (PDB: 6Y2G) having an RDOCK score of −58.121 kcal/mol and a Vina score of −7.9 kcal/mol. The carbonyl moiety of sepimostat was closely located at the Cys145 whereas the dihydroimidazole ring is located closely towards His41 at the enzyme active site (Fig. 8D).

Various molecular modeling studies [83,95–101] were conducted on remdesivir, which was found to be the 5^{th} highest molecule in our analysis. Murugan et al. [96] displayed that remdesivir showed a strong binding affinity with viral SARS-CoV-2 $3CL^{pro}$ enzyme (−44.4 kcal/mol). Alajmi [97] showed that remdesivir had a docking score of −8.246 kcal/mol and prime MM-GBSA binding energy of −52.23 kcal/mol with the SARS-CoV-2 $3CL^{pro}$ enzyme. As per the observation of Rehman et al. [98], remdesivir had a docking energy of −7.5 kcal/mol with SARS-CoV-2 $3CL^{pro}$ (PDB: 6LU7). It formed hydrogen bonding with Gly143, Thr24, His164, Glu166, and Gln189 along with hydrophobic interactions with His41 and Met49. As per the observation of Abdelrheem et al. [99], remdesivir displayed binding energy of −9.24 kcal/mol, ligand efficiency of −0.24, and intermolecular energy of −16.32. Hiremanth et al. [96] showed that remdesivir forms hydrogen bonding interaction

with His41, Arg188, and Thr190 with SARS-CoV-2 3CL^{pro} (PDB: 6LU7). The post MD simulation study by Naik et al. [101] depicted that remdesivir displayed potential hydrogen bonding with His163, Glu166, Cys145, Gly143 (Fig. 8E) along with several water molecules to make stable interactions. The RMSF values showed lower atomic fluctuations in binding site residues, which suggested smaller conformational changes and established stable binding.

As far as the HQSAR model was concerned, it also predicted several repurposed drug molecules as potential SARS-CoV-2 3CL^{pro} inhibitors. Among these compounds, the five highly active molecules were lurasidone (predicted IC_{50} = 63.53 nM), saquinavir (predicted IC_{50} = 123.03 nM), lopinavir (predicted IC_{50} = 246.60 nM), elbasvir (predicted IC_{50} = 456.04 nM) and paritaprevir (predicted IC_{50} = 538.27 nM).

Thurakkal et al. [102] showed that lurasidone had strong binding energy (−8.4 kcal/mol) during binding to the SARS-CoV-2 3CL^{pro} enzyme (PDB: 6Y84). The MD simulation study revealed that the complex had a lower average backbone RMSD (1.54 Å). Elmezayen et al. [103] showed that lurasidone had a binding energy score (−11.17 kcal/mol) and inhibition constant (K_i = 6.52 nM) while coordinating to the active site of SARS-CoV-2 3CL^{pro} enzyme (PDB: 6LU7). Lurasidone formed hydrogen bonding interaction with His41, Glu166 as well as alkyl interactions with Met165 and Met49 along with π-alkyl interactions with Pro168, Met165, and His41 (Fig. 9A).

Wang et al. [62] showed that elbasvir had a docking score of −9.9 and the MM-PBSA-WSAS binding free energy (ΔG = −6.5 kcal/mol). Elbasvir is strongly bound to the active site of SARS-CoV-2 3CL^{pro} showing interaction with Thr25, Thr26, His41, Met49, Met165, Glu166, Gln189, and Thr190 (Fig. 9B).

As far as the observation of Tripathi et al. [104] was concerned, lopinavir exhibited effective SARS-CoV-2 3CL^{pro} inhibition at 16 µM dose. Deshpande et al. [83] showed that lopinavir had good binding energy (ΔG = −9.9 kcal/mol) with SARS-CoV-2 3CL^{pro} (PDB: 6Y84). Similarly, Gyebi et al. [105] showed that lopinavir had a docking score of −8.3 with SARS-CoV-2 3CL^{pro} (PDB: 6LU7). It showed several hydrogen bonding and hydrophobic interactions at the active site. Shah et al. [106] carried out a comparative molecular docking analysis and showed that lopinavir had good dock scores of five different SARS-CoV-2 3CL^{pro} enzymes. It had docking scores of −6.834, −6.968, −7.331, −8.44, and 7.58 with five different SARS-CoV-2 3CL^{pro} enzymes, namely 5R7Y, 5R7Z, 5R80, 5R81, and 5R82 respectively. The molecular docking study of lopinavir with SARS-CoV-2 3CL^{pro} (PDB: 5R81) showed that lopinavir formed hydrogen bonds with Glu166 and His41. Moreover, His41 formed a π-π stacking interaction with the phenyl ring of lopinavir (Fig. 9C).

Paritaprevir is another molecule that was predicted well by the HQSAR modeling conducted in this study. Bahadur Gurung et al. [60] displayed that paritaprevir formed hydrogen bonds with Glu166 and Asn142 along with several hydrophobic interactions with Thr25, Thr26, His41, Met49, Gly143, Cys145, Met165, Gln189, and Gln192. Alamri et al. [107] showed that paritaprevir formed hydrogen bonding interactions with Gly143 and Cys145. It also formed amide-π stacking interaction with Thr45 and π-alkyl interaction with Met49, Met165, and Pro168. The MD simulation study resulted in an RMSD value of 1.2 Å. Again, the total binding free energy resulted in a favorable MM-GBSA total energy of −47.15 kcal/mol (Fig. 9D).

5. Conclusion

The world is currently running through a devastating situation due to COVID-19, and SARS-CoV-2 has badly affected the whole world. To date, no such small molecule drug candidate has been approved exclusively for the treatment of COVID-19, and thus, there is an extreme requirement for such drugs. In this current work, a robust, regression-dependent molecular modeling study was conducted on 69 diverse compounds possessing SARS-CoV-2 3CL^{pro} inhibitory activity. Different linear (SW-MLR and GA-MLR) and non-linear (ANN and SVM) QSAR, as well as HQSAR models, were constructed in this present work to extract out important structural features responsible for imparting SARS-CoV-2 3CL^{pro} inhibition. All these QSAR models were validated by internal and external cross-validation methods in a robust fashion. For the SW-MLR and GA-MLR models, 50 different test and training set combinations were used to validate the robustness of these models. These QSAR models suggest that the E-State indices related to SP^3 hybridized carbon atoms, molecular distance edge between all secondary nitrogen atoms, molecular distance edge between all secondary oxygen atoms, molecular complexity, maximum E-state descriptor of strength for potential hydrogen bonds of path length 10 and 8 are crucial for pertaining the SARS-CoV-2 3CL^{pro} inhibitory activity. Another SW-MLR model with well-known physicochemical parameters and indicator variables showed that the 2-oxopyrrolidine and the benzylester functions are crucial for higher SARS-CoV-2 3CL^{pro} inhibition. The non-linear ANN models (SW-ANN, GA-ANN) and SVM (SW-SVM, GA-SVM) models were successfully able to optimize all the linear QSAR models (SW-MLR and GA-MLR) as suggested by the statistical validation parameters. The HQSAR model was also able to identify important good and bad structural fragments essential for modulating the SARS-CoV-2 3CL^{pro} inhibition. It also showed that 2-oxopyrrolidinyl methyl function, benzylester function, and methylene (hydroxy) sulphonic acid warhead group are important for retaining higher SARS-CoV-2 3CL^{pro} inhibitory activity. Nevertheless, the GA-MLR and HQSAR models were also tried to predict externally several already repurposed drugs by a different group of researchers. The GA-MLR and HQSAR models predicted 5 already repurposed compounds each as potent inhibitors of SARS-CoV-2 (GA-MLR: curcumin, ribavirin, saquinavir, sepimostat, and remdesivir; HQSAR: lurasidone, saquinavir, lopinavir, elbasvir, and paritaprevir). The binding modes of interactions of all these selected compounds were justified by molecular docking studies, binding energy calculation, and molecular dynamics (MD) simulation studies conducted by various groups of researchers. Therefore, this current linear and non-linear QSAR along with HQSAR models can accelerate the discovery of newer, effective, and promising SARS-CoV-2 3CL^{pro} inhibitors to combat the dreaded COVID-19 in the future.

Declaration of Competing Interest

The authors declare no conflict of interests.

CRediT authorship contribution statement

Nilanjan Adhikari: Conceptualization, Data curation, Methodology, Visualization, Investigation, Writing – original draft. **Suvankar Banerjee:** Conceptualization, Data curation, Methodology, Visualization, Investigation, Writing – original draft. **Sandip Kumar Baidya:** Data curation, Visualization, Writing – original draft. **Balaram Ghosh:** Conceptualization, Writing – review & editing, Supervision. **Tarun Jha:** Conceptualization, Writing – review & editing, Supervision.

Acknowledgements

Suvankar Banerjee is thankful to the Swami Vivekananda Merit-cum-Means (SVMCM) scholarship, Govt. of West Bengal, India for awarding the fellowship. Sandip Kumar Baidya is thankful to the Indian Council of Medical Research (ICMR), New Delhi, India for awarding the senior research fellowship (SRF) [FILE NO.:

45/29/2019-PHA, dated: 21–06–2019]. The research has been supported by the research fund provided by the Council of Scientific and Industrial Research (CSIR- 37(1722)/19/EMR-II), New Delhi, India to Dr. Balaram Ghosh. The authors are sincerely thankful to the Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India, and the Department of Pharmacy, BITS-Pilani, Hyderabad, India for providing research facilities.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.molstruc.2021.132041.

References

[1] G.K. Goh, A.K. Dunker, J.A. Foster, V.N. Uversky, Shell disorder analysis suggests that pangolins offered a window for a silent spread of an attenuated SARS-CoV-2 precursor among humans, J. Proteome. Res. 19 (2020) 4543–4552.
[2] Q. Li, C.B. Kang, Progress in developing inhibitors of SARS-CoV-2 3C-like protease, Microorganisms 8 (2020) 1250.
[3] N. Adhikari, S.A. Amin, T. Jha, Dissecting the drug development strategies against SARS-CoV-2 through diverse computational modeling techniques, in: K. Roy (Ed.), Methods in Pharmacology and Toxicology, Springer, New York, 2020, pp. 1–103.
[4] Y.C. Liu, R.L. Kuo, S.R. Shih, COVID-19: the first documented coronavirus pandemic in history, Biomed. J. 43 (2020) 328–333.
[5] Accessed on May 2021 (2021) https://www.worldometers.info/coronavirus/.
[6] F. Ciceri, L. Beretta, A.M. Scandroglio, S. Colombo, G. Landoni, A. Ruggeri, J. Peccatori, A. D'Angelo, F. De Cobelli, P. Rovere-Querini, M. Tresoldi, Microvascular COVID-19 lung vessels obstructive thromboinflammatory syndrome (MicroCLOTS): an atypical acute respiratory distress syndrome working hypothesis, Crit. Care Resusc. 22 (2000) 95–97.
[7] F. Yazdanpanah, M.R. Hamblin, N. Rezaei, The immune system and COVID-19: friend or foe? Life Sci. 256 (2020) 117900.
[8] H.C. Hung, Y.Y. Ke, S.Y. Huang, P.N. Huang, Y.A. Kung, T.Y. Chang, K.J. Yen, T.T. Peng, S.E. Chang, C.T. Huang, Y.R. Tsai, Discovery of M protease inhibitors encoded by SARS-CoV-2, Antimicrob. Agents Chemother. 64 (2020) e00872 -20.
[9] E. De Wit, N. Van Doremalen, D. Falzarano, V.J. Munster, SARS and MERS: recent insights into emerging coronaviruses, Nat. Rev. Microbiol. 14 (2016) 523–534.
[10] S.K. Lau, P.C. Woo, K.S. Li, Y. Huang, H.W. Tsoi, B.H. Wong, S.S. Wong, S.Y. Leung, K.H. Chan, K.Y. Yuen, Severe acute respiratory syndrome coronavirus–like virus in Chinese horseshoe bats, Proc. Natl. Acad. Sci. USA 102 (2005) 14040–14045.
[11] T. Zhang, Q. Wu, Z. Zhang, Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak, Curr. Biol. 30 (2020) 1346–1351.
[12] M.P. Egloff, F. Ferron, V. Campanacci, S. Longhi, C. Rancurel, H. Dutartre, E.J. Snijder, A.E. Gorbalenya, C. Cambillau, B. Canard, The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world, Proc. Natl. Acad. Sci. USA 101 (2004) 3792–3796.
[13] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, J. Med. Virol. 92 (2020) 418–423.
[14] M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.L. Chen, Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants, J. Pharm. Anal. 10 (2020) 313–319.
[15] Y.W. Chen, C.P. Yiu, K.Y. Wong, Prediction of the SARS-CoV-2 (2019-nCoV) 3C–like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates, F1000Res 9 (2020) 129.
[16] M. Prajapat, P. Sarma, N. Shekhar, P. Avti, S. Sinha, H. Kaur, S. Kumar, A. Bhattacharyya, H. Kumar, S. Bansal, B. Medhi, Drug targets for corona virus: a systematic review, Ind. J. Pharmacol. 52 (2020) 56–65.
[17] K. Akaji, H. Konno, Design and evaluation of anti-SARS-coronavirus agents based on molecular interactions with the viral protease, Molecules 25 (2020) 3920.
[18] S.A. Amin, T. Jha, Fight against novel coronavirus: a perspective of medicinal chemists, Eur. J. Med. Chem. 201 (2020) 112559.
[19] S.A. Amin, S. Banerjee, S. Gayen, T. Jha, Protease targeted COVID-19 drug discovery: what we have learned from the past SARS-CoV inhibitors? Eur. J. Med. Chem. 215 (2021) 113294.
[20] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors, Science 368 (2020) 409–412.
[21] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, G. Xiao, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro, Cell Res. 30 (2020) 269–271.
[22] P. De, I. Chakraborty, B. Karna, N. Mazumder, Brief review on repurposed drugs and vaccines for possible treatment of COVID-19, Eur. J. Pharmacol. 898 (2021) 173977.
[23] C. Harrison, Coronavirus puts drug repurposing on the fast track, Nat. Biotech. 38 (2020) 379–381.
[24] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors, J. Mol. Struct. 1224 (2021) 129026.
[25] S.A. Amin, K. Ghosh, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors, J. Biomol. Struct. Dyn. 39 (2021) 4764–4773.
[26] S.A. Amin, S. Banerjee, K. Ghosh, S. Gayen, T. Jha, Protease targeted COVID-19 drug discovery and its challenges: insight into viral main protease (Mpro) and papain-like protease (PLpro) inhibitors, Bioorg. Med. Chem. 29 (2020) 115860.
[27] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Unmasking of crucial structural fragments for coronavirus protease inhibitors and its implications in COVID-19 drug discovery, J. Mol. Struct. 1237 (2021) 130366.
[28] S.A. Amin, K. Ghosh, S. Singh, I.A. Qureshi, T. Jha, S. Gayen, Exploring naphthyl derivatives as SARS-CoV papain-like protease (PLpro) inhibitors and its implications in COVID-19 drug discovery, Mol. Divers. (2021) 1–14 *In Press*, doi:10.1007/s11030-021-10198-10203.
[29] S.A. Amin, S. Banerjee, S. Singh, I.A. Qureshi, S. Gayen, T. Jha, First structure-activity relationship analysis of SARS-CoV-2 virus main protease (Mpro) inhibitors: an endeavor on COVID-19 drug discovery, Mol. Divers. 25 (2021) 1827–1838.
[30] N. Adhikari, S. Banerjee, S.K. Baidya, B. Ghosh, T. Jha, Robust classification-based molecular modelling of diverse chemical entities as potential SARS-CoV-2 3CLpro inhibitors: theoretical justification in light of experimental evidences, SAR QSAR Env. Res. 32 (2021) 473–493.
[31] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Wang, Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, Cell Res. 30 (2020) 678–692.
[32] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, Structure of M pro from SARS-CoV-2 and discovery of its inhibitors, Nature 582 (2020) 289–293.
[33] M.D. Sacco, C. Ma, P. Lagarias, A. Gao, J.A. Townsend, X. Meng, P. Dube, X. Zhang, Y. Hu, N. Kitamura, B. Hurst, Structure and inhibition of the SARS–CoV-2 main protease reveal strategy for developing dual inhibitors against Mpro and cathepsin L, Sci. Adv. 6 (2020) eabe0751.
[34] W. Dai, B. Zhang, X.M. Jiang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, J. Peng, F. Liu, C. Li, Structure-based design of antiviral drug candidates targeting the SARS–CoV-2 main protease, Science 368 (2020) 1331–1335.
[35] W. Vuong, M.B. Khan, C. Fischer, E. Arutyunova, T. Lamer, J. Shields, H.A. Saffran, R.T. McKay, M.J. van Belkum, M. Joyce, H.S. Young, Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication, Nat. Commun. 11 (2020) 4282.
[36] A.D. Rathnayake, J. Zheng, Y. Kim, K.D. Perera, S. Mackin, D.K. Meyerholz, M.M. Kashipathy, K.P. Battaile, S. Lovell, S. Perlman, W.C. Groutas, 3C-like protease inhibitors block coronavirus replication in vitro and improve survival in MERS-CoV–infected mice, Sci. Transl. Med. 12 (2020) eabc5332.
[37] W. Zhu, M. Xu, C.Z. Chen, H. Guo, M. Shen, X. Hu, P. Shinn, C. Klumpp-Thomas, S.G. Michael, W. Zheng, Identification of SARS-CoV-2 3CL protease inhibitors by a quantitative high-throughput screening, ACS Pharmacol. Transl. Sci. 3 (2020) 1008–1016.
[38] H. Su, S. Yao, W. Zhao, M. Li, J. Liu, W. Shang, H. Xie, C. Ke, M. Gao, K. Yu, H. Liu, Discovery of baicalin and baicalein as novel, natural product inhibitors of SARS-CoV-2 3CL protease in vitro, bioRxiv (2020), doi:10.1101/2020.04.13.038687.
[39] C.W. Yap, PaDEL descriptor: an open source software to calculate molecular descriptors and fingerprints, J. Comput. Chem. 32 (2011) 1466–1474.
[40] as accessed in (May 2021) http://dtclab.webs.com/software-tools.
[41] P. De, K. Roy, QSAR and QSAAR modeling of nitroimidazole sulfonamide radiosensitizers: application of small dataset modeling, Struct. Chem. 32 (2021) 631–642.
[42] V. Kumar, K. Roy, Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3C-like protease (3CLpro) enzyme inhibitors against SARS-CoV diseases, SAR QSAR Environ. Res. 31 (2020) 511–526.
[43] K. Roy, S. Kar, R.N. Das, A Primer On QSAR/QSPR modeling: Fundamental Concepts, Springer, New York, 2015.
[44] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Model. 20 (2002) 269–276.
[45] C. Nantasenamat, C.I. Ayudhya, N. Tansila, T. Naenna, V. Prachayasittikul, Prediction of GFP spectral properties using artificial neural network, J. Comput. Chem. 28 (2007) 1275–1289.
[46] C. Nantasenamat, T. Naenna, C.I. Ayudhya, V. Prachayasittikul, Quantitative prediction of imprinting factor of molecularly imprinted polymers by artificial neural network, J. Comput. Aid. Mol. Des. 19 (2005) 509–524.
[47] S.A. Amin, N. Adhikari, S. Gayen, T. Jha, First report on the structural exploration and prediction of new BPTES analogs as glutaminase inhibitors, J. Mol. Struct. 1143 (2017) 49–64.
[48] C. Nantasenamat, A. Worachartcheewan, S. Jamsak, L. Preeyanon, W. Shoombuatong, S. Simeon, P. Mandi, C.I. Ayudhya, V. Prachayasittikul, AutoWeka: toward an automated data mining software for QSAR and QSPR studies, in: Artificial Neural Networks, Springer, New York, 2015, pp. 119–147.

[49] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software: an update, ACM SIGKDD Explor. Newslet. 11 (2009) 10–18.
[50] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (1999) 988–999.
[51] C. Nantasenamat, C.I. Ayudhya, T. Naenna, V. Prachayasittikul, Prediction of bond dissociation enthalpy of antioxidant phenols by support vector machine, J. Mol. Graph. Model. 27 (2008) 188–196.
[52] C.W. Murray, M.L. Verdonk, D.C. Rees, Experiences in fragment-based drug discovery, Trends Pharmacol. Sci. 33 (2012) 224–232.
[53] A.K. Romasanta, P. van der Sijde, I. Hellsten, R.E. Hubbard, G.M. Keseru, J. van Muijlwijk-Koezen, I.J. de Esch, When fragments link: a bibliometric perspective on the development of fragment-based drug discovery, Drug Discov. Today 23 (2018) 1596–1609.
[54] T.W. Heritage, D.R. Lowis, Molecular Hologram QSAR, in: A.L. Parrill, M.R. Reddy (Eds.), Rational Drug Design: Novel Methodology and Practical Applications, ACS Symposium Series, American Chemical Society, Washington DC, 1999, pp. 212–225.
[55] SYBYL-X 2.0 (2012) Tripos Inc., St Louis, MO.
[56] S.K. Maurya, A.K. Maurya, N. Mishra, H.R. Siddique, Virtual screening, ADME/T, and binding free energy analysis of anti-viral, antiprotease, and antiinfectious compounds against NSP10/NSP16 methyltransferase and main protease of SARS CoV-2, J. Receptor Signal. Transduct. 40 (2020) 605–612.
[57] K.A. Peele, C.P. Durthi, T. Srihansa, S. Krupanidhi, A.V. Sai, D.J. Babu, M. Indira, A.R. Reddy, T.C. Venkateswarulu, Molecular docking and dynamic simulations for antiviral compounds against SARS-CoV-2: a computational study, Inform. Med. Unlocked 19 (2020) 100345.
[58] V.P. Raphael, S.K. Shanmughan, Computational Evaluation of the Inhibition Efficacies of HIV Antivirals on SARS-CoV-2 (COVID-19) Protease and Identification of 3D Pharmacophore and hit compounds, Adv. Pharmacol. Pharm. Sci. 2020 (2020) 8818008.
[59] B. Shah, P. Modi, S.R. Sagar, In silico studies on therapeutic agents for COVID-19: drug repurposing approach, Life Sci. 252 (2020) 117652.
[60] A.B. Gurung, M.A. Ali, J. Lee, M.A. Farah, K.M. Al-Anazi, Structure-based virtual screening of phytochemicals and repurposing of FDA approved antiviral drugs unravels lead molecules as potential inhibitors of coronavirus 3C-like protease enzyme, J. King Saud Univ. Sci. 32 (2020) 2845–2853.
[61] N. Lobo-Galo, M. Terrazas-López, A. Martıńez-Martıńez, G A, Dıáz-Sańchez, FDA-approved thiol-reacting drugs that potentially bind into the SARS-CoV-2 main protease, essential for viral replication, J. Biomol. Struct. Dyn. 39 (2020) 3419–3427.
[62] J. Wang, Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study, J. Chem. Inf. Model. 60 (2020) 3277–3286.
[63] K. Al-Khafaji, D. Al-Duhaidahawi, T.T. Tok, Using integrated computational approaches to identify safe and rapid treatment for SARS-CoV-2, J. Biomol. Struct. Dyn. 39 (2020) 3387–3395.
[64] M. Tsuji, Potential anti-SARS-CoV-2 drug candidates identified through virtual screening of the ChEMBL database for compounds that target the main coronavirus protease, FEBS Open Biol. 10 (2020) 995–1004.
[65] S.A. Khan, K. Zia, S. Ashraf, R. Uddin, Z. Ul-Haq, Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach, J. Biomol. Struct. Dyn. 39 (2021) 2607–2616.
[66] S. Pant, M. Singh, V. Ravichandiran, U.S.N. Murty, H.K. Srivastava, Peptide–like and small-molecule inhibitors against Covid-19, J. Biomol. Struct. Dyn. 39 (2020) 2904–2913.
[67] S. Das, S. Sarmah, S. Lyndem, R.A. Singha, An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study, J. Biomol. Struct. Dyn. 39 (2020) 3347–3357.
[68] Z. Reiner, M. Hatamipour, M. Banach, M. Pirro, K. Al-Rasadi, T. Jamialahmadi, D. Radenkovic, F. Montecucco, A. Sahebkar, Statins and the COVID-19 main protease: in silico evidence on direct interaction, Arch. Med. Sci. 16 (2020) 490–496.
[69] R. Islam, M.R. Parves, A.S. Paul, N. Uddin, M.S. Rahman, A.A. Mamun, M.N. Hossain, M.A. Ali, M.A. Halim, A molecular modeling approach to identify effective antiviral phytochemicals against the main protease of SARS CoV-2, J. Biomol. Struct. Dyn. 39 (2020) 3213–3224.
[70] A. Shamsi, T. Mohammad, S. Anwar, M.F. AlAjmi, A. Hussain, M.T. Rehman, A. Islam, M.I. Hassan, Glecaprevir and Maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy, Biosci. Rep. 40 (2020) 1–8.
[71] V.K. Bhardwaj, R. Singh, J. Sharma, V. Rajendran, R. Purohit, S. Kumar, Identification of bioactive molecules from tea plant as SARS-CoV-2 main protease inhibitors, J. Biomol. Struct. Dyn. 39 (2020) 3449–3458.
[72] A.D. Elmezayen, A. Al-Obaidi, A.T. Sahin, K. Yelekc, Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes, J. Biomol. Struct. Dyn. 39 (2020) 2980–2992.
[73] A. Fischer, M. Sellner, S. Neranjan, M. Smies˙ko, M.A. Lill, Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds, Int. J. Mol. Sci. 21 (2020) 3626.
[74] M. Kandeel, M. Al-Nazawi, Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease, Life Sci 251 (2020) 117627.
[75] P. Calligari, S. Bobone, G. Ricci, A. Bocedi, Molecular investigation of SARS—CoV-2 proteins and their interactions with antiviral drugs, Viruses 12 (2020) 445.
[76] X. Liu, X.J. Wang, Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines, J. Genet. Genomics 47 (2020) 119–121.
[77] K. Rajagopal, P. Varakumar, A. Baliwada, G. Byran, Activity of phytochemical constituents of Curcuma longa (turmeric) and Andrographis paniculata against coronavirus (COVID-19): an in silico approach, Fut. J. Pharm. Sci. 6 (2020) 104.
[78] M.A.A. Ibrahim, A.H.M. Abdelrahman, T.A. Hussien, E.A.A. Badr, T.A. Mohamed, H.R. El-Seedi, P.W. Pare, T. Efferth, M.F. Hegazy, In silico drug discovery of major metabolites from spices as SARS-CoV-2 main protease inhibitors, Comput. Biol. Med. 126 (2020) 104046.
[79] S. Basu, B. Veeraraghavan, S. Ramaiah, A. Anbarasu, Novel cyclohexanone compound as a potential ligand against SARS-CoV-2 main-protease, Microb. Pathog. 149 (2020) 104546.
[80] D.M. Teli, M.B. Shah, M.T. Chhabria, In silico Screening of Natural Compounds as Potential Inhibitors of SARS-CoV-2 Main Protease and Spike RBD: targets for COVID-19, Front. Mol. Biosci. 7 (2021) 599079.
[81] S. Mahmud, M.A.R. Uddin, G.K. Paul, M.S.S. Shimu, S. Islam, E. Rahman, A. Islam, M.S. Islam, M.M. Promi, T.B. Emran, M.A. Saleh, Virtual screening and molecular dynamics simulation study of plant-derived compounds to identify potential inhibitors of main protease from SARS-CoV-2, Brief Bioinform 22 (2021) 1402–1414.
[82] S. Khaerunnisa, H. Kurniawan, R. Awaluddin, S. Suhartati, S. Soetjipto, Potential inhibitor of COVID-19 main protease (Mpro) from several medicinal plant compounds by molecular docking study, Preprints (2020) 2020030226, doi:10.20944/preprints202003.0226.v1.
[83] R.R. Deshpande, A.P. Tiwari, N. Nyayanit, M. Modak, In silico molecular docking analysis for repurposing therapeutics against multiple proteins from SARS-CoV-2, Eur. J. Pharmacol. 886 (2020) 173430.
[84] S. Gupta, V. Singh, P.K. Varadwaj, N. Chakravartty, A.V.S.K.M. Katta, S.P. Lekkala, G. Thomas, S. Narasimhan, A.R. Reddy, V.B.R. Lachagari, Secondary metabolites from spice and herbs as potential multitarget inhibitors of SARS-CoV-2 proteins, J. Biomol. Struct. Dyn. (2020) 1–20 *In press*, doi:10.1080/07391102.2020.1837679.
[85] P. Kumar, T. Bhardwaj, A. Kumar, B.R. Gehi, S.K. Kapuganti, N. Garg, G. Nath, R. Giri, Reprofiling of approved drugs against SARS-CoV-2 main protease: an in-silico study, J. Biomol. Struct. Dyn. (2020) 1–15 *In press*, doi:10.1080/07391102.2020.1845976.
[86] D.C. Hall Jr., H.F. Ji, A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease, Travel Med. Infect. Dis. 35 (2020) 101646.
[87] S. Talluri, Molecular Docking and Virtual Screening based prediction of drugs for COVID-19, Comb. Chem. High Throughput. Screen. 24 (2021) 716–728.
[88] J.T. Ortega, M.L. Serrano, F.H. Pujol, H.R. Rangel, Unrevealing sequence and structural features of novel coronavirus using in silico approaches: the main protease as molecular target, EXCLI J. 19 (2020) 400–409.
[89] M. Bello, A. Martínez-Muñoz, I. Balbuena-Rebolledo, Identification of saquinavir as a potent inhibitor of dimeric SARS-CoV2 main protease through MM/GBSA, J. Mol. Model. 26 (2020) 340.
[90] S. Keretsu, S.P. Bhujbal, S.J. Cho, Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation, Sci. Rep. 10 (2020) 17716.
[91] R.O. Barros, F.L.C.C. Junior, W.S. Pereira, N.M.N. Oliveira, R.M. Ramos, Interaction of drug candidates with various SARS-CoV-2 receptors: an in silico study to combat COVID-19, J. Proteom. Res. 19 (2020) 4567–4575.
[92] M. Mondal, C. Sarkar, S. Jamaddar, A.B.R. Khalipha, M.T. Islam, A. Mahafzah, M.S. Mubarak, Evaluation of the binding affinity of anti-viral drugs against main protease of SARS-CoV-2 through a molecular docking study, Infect. Disord. Drug Targets. (2020) *In Press*, doi:10.2174/1871526520666201207124408.
[93] A.A.A. Abu-Saleh, I.E. Awad, A. Yadav, R.A. Poirier, Discovery of potent inhibitors for SARS-CoV-2′s main protease by ligand-based/structure-based virtual screening, MD simulations, and binding energy calculations, Phys. Chem. Chem. Phys. 22 (2020) 23099–23106.
[94] X. Zhao, R. Liu, Z. Miao, N. Ye, W. Lu, A study of potential SARS-CoV-2 antiviral drugs and preliminary research of their molecular mechanism, based on Anti-SARS-CoV drug screening and molecular dynamics simulation, J. Comput. Biol. (2020), doi:10.1089/cmb.2020.0112.
[95] R. Yu, L. Chen, R. Lan, R. Shen, P. Li, Computational screening of antagonists against the SARS-CoV-2 (COVID-19) coronavirus by molecular docking, Int. J. Antimicrob. Agents 56 (2020) 106012.
[96] N.A. Murugan, S. Kumar, J. Jeyakanthan, V. Srivastava, Searching for target-specific and multi-targeting organics for Covid-19 in the Drugbank database with a double scoring approach, Sci. Rep. 10 (2020) 19125.
[97] M.F. AlAjmi, A. Azhar, M. Owais, S. Rashid, S. Hasan, A. Hussain, M.T. Rehman, Antiviral potential of some novel structural analogs of standard drugs repurposed for the treatment of COVID-19, J. Biomol. Struct. Dyn. 39 (2021) 6676–6688.
[98] M.T. Rehman, M.F. AlAjmi, Hussain A, Natural Compounds as Inhibitors of SARS-CoV-2 Main Protease (3CLpro): a molecular docking and simulation approach to combat COVID-19, Curr. Pharm. Des. (2020) *In Press*, doi:10.2174/1381612826999201116195851.
[99] D.A. Abdelrheem, S.A. Ahmed, H.R.A. El-Mageed, H.S. Mohamed, A.A. Rahman, K.N.M. Elsayed, S.A. Ahmed, The inhibitory effect of some natural bioactive compounds against SARS-CoV-2 main protease: insights from molecular docking analysis and molecular dynamic simulation, J. Environ. Sci. Health A Tox. Hazard. Subst. Environ. Eng. 55 (2020) 1373–1386.

[100] S. Hiremath, H.D.V. Kumar, M. Nandan, M. Mantesh, K.S. Shankarappa, V. Venkataravanappa, C.R.J. Basha, C.N.L. Reddy, In silico docking analysis revealed the potential of phytochemicals present in Phyllanthus amarus and Andrographis paniculata, used in Ayurveda medicine in inhibiting SARS–CoV-2, 3 Biotech 11 (2021) 44.
[101] S.R. Naik, P. Bharadwaj, N. Dingelstad, S. Kalyaanamoorthy, S.C. Mandal, A. Ganesan, D. Chattopadhyay, P. Palit, Structure-based virtual screening, molecular dynamics and binding affinity calculations of some potential phytocompounds against SARS-CoV-2, J. Biomol. Struct. Dyn. (2021) 1–18 *In Press*, doi:10.1080/07391102.2021.1891969.
[102] L. Thurakkal, S. Singh, R. Roy, P. Kar, S. Sadhukhan, M. Porel, An in-silico study on selected organosulfur compounds as potential drugs for SARS-CoV-2 infection via binding multiple drug targets, Chem. Phys. Lett. 763 (2021) 138193.
[103] A.D. Elmezayen, A. Al-Obaidi, A.T. Şahin, K. Yelekçi, Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes, J. Biomol. Struct. Dyn. 39 (2020) 2980–2992.
[104] P.K. Tripathi, S. Upadhyay, M. Singh, S. Raghavendhar, M. Bhardwaj, P. Sharma, A.K. Patel, Screening and evaluation of approved drugs as inhibitors of main protease of SARS-CoV-2, Int. J. Biol. Macromol. 164 (2020) 2622–2631.
[105] G.A. Gyebi, O.B. Ogunro, A.P. Adegunloye, O.M. Ogunyemi, S.O. Afolabi, Potential inhibitors of coronavirus 3-chymotrypsin-like protease (3CLpro): an in silico screening of alkaloids and terpenoids from African medicinal plants, J. Biomol. Struct. Dyn. 39 (2020) 3396–3408.
[106] B. Shah, P. Modi, S.R. Sagar, In silico studies on therapeutic agents for COVID-19: drug repurposing approach, Life Sci. 252 (2020) 117652.
[107] M.A. Alamri, M.T. Ul Qamar, M.U. Mirza, R. Bhadane, S.M. Alqahtani, I. Muneer, M. Froeyen, O.M. Salo-Ahen, Pharmacoinformatics and molecular dynamics simulation studies reveal potential covalent and FDA-approved inhibitors of SARS-CoV-2 main protease 3CLpro, J. Biomol. Struct. Dyn. 39 (2021) 4936–4948.