



Article

Identification of Anti-SARS-CoV-2 Compounds from Food Using QSAR-Based Virtual Screening, Molecular Docking, and Molecular Dynamics Simulation Analysis

Magdi E. A. Zaki ^{1,*}, Sami A. Al-Hussain ¹, Vijay H. Masand ^{2,*}, Siddhartha Akasapu ³, Sumit O. Bajaj ³, Nahed N. E. El-Sayed ⁴, Arabinda Ghosh ⁵ and Israa Lewaa ⁶

¹ Department of Chemistry, Faculty of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 13318, Saudi Arabia; sahusain@imamu.edu.sa

² Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati, Maharashtra 444 602, India

³ Corden Pharma, Boulder, CO 80301, USA; asidhu09@gmail.com (S.A.); sumitbajaj16@gmail.com (S.O.B.)

⁴ Egyptian Drug Authority, Agouza, Giza 35521, Egypt; nnelsayed@gmail.com

⁵ Microbiology Division, Department of Botany, Gauhati University, Guwahati, Assam 781014, India; dra.ghosh@gauhati.ac.in

⁶ Department of Business Administration, Faculty of Business Administration, Economics and Political Science, British University in Egypt, Cairo 11837, Egypt; Israa.lewaa@bue.edu.eg

* Correspondence: Mezaki@imamu.edu.sa (M.E.A.Z.); vijaymasand@gmail.com (V.H.M.)



Citation: Zaki, M.E.A.;

Al-Hussain, S.A.; Masand, V.H.;

Akasapu, S.; Bajaj, S.O.;

El-Sayed, N.N.E.; Ghosh, A.; Lewaa, I.

Identification of Anti-SARS-CoV-2

Compounds from Food Using

QSAR-Based Virtual Screening,

Molecular Docking, and Molecular

Dynamics Simulation Analysis.

Pharmaceuticals **2021**, *14*, 357.

[https://doi.org/10.3390/](https://doi.org/10.3390/ph14040357)

[ph14040357](https://doi.org/10.3390/ph14040357)

Academic Editor:

Jean Jacques Vanden Eynde and

Annie Mayence

Received: 20 March 2021

Accepted: 5 April 2021

Published: 13 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Due to the genetic similarity between SARS-CoV-2 and SARS-CoV, the present work endeavored to derive a balanced Quantitative Structure–Activity Relationship (QSAR) model, molecular docking, and molecular dynamics (MD) simulation studies to identify novel molecules having inhibitory potential against the main protease (Mpro) of SARS-CoV-2. The QSAR analysis developed on multivariate GA–MLR (Genetic Algorithm–Multilinear Regression) model with acceptable statistical performance ($R^2 = 0.898$, $Q^2_{loo} = 0.859$, etc.). QSAR analysis attributed the good correlation with different types of atoms like non-ring Carbons and Nitrogens, amide Nitrogen, sp^2 -hybridized Carbons, etc. Thus, the QSAR model has a good balance of qualitative and quantitative requirements (balanced QSAR model) and satisfies the Organisation for Economic Co-operation and Development (OECD) guidelines. After that, a QSAR-based virtual screening of 26,467 food compounds and 360 heterocyclic variants of molecule **1** (benzotriazole–indole hybrid molecule) helped to identify promising hits. Furthermore, the molecular docking and molecular dynamics (MD) simulations of Mpro with molecule **1** recognized the structural motifs with significant stability. Molecular docking and QSAR provided consensus and complementary results. The validated analyses are capable of optimizing a drug/lead candidate for better inhibitory activity against the main protease of SARS-CoV-2.

Keywords: SARS-CoV; SARS-CoV-2; COVID-19; QSAR; molecular docking; QSAR-based virtual screening; machine learning

1. Introduction

The current epidemic involving the coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has completely disturbed the global health system and the world economy. The first patient of COVID-19 was reported in Wuhan, China, in December, 2019 [1], and now the virus has reached 215 countries in the world [2]. In some countries, it is in stage one or stage two, but in some countries, it has reached stage three, i.e., community spread [2]. Due to its high spreading rate and mortality, WHO (World Health Organization) has declared it a Public Health Emergency of International Concern [2].

SARS-CoV-2 belongs to the family of Coronaviridae and the genus Betacoronavirus [3]. These are enveloped (enclosed) viruses having a single positive-stranded RNA genome

(~26–32 kb in length) [4]. The name “Corona” has been given to them due to the crown-like presence of a fringe of large and bulbous surface projections when observed under the microscope [5]. SARS-CoV-2 is the seventh member [6] from this family to infect the humans, other mammals, and birds [7]. Coronavirus infections mainly cause respiratory tract infections, and some of them are responsible for severe infectivity and damage to public health, e.g., in 2002, severe acute respiratory syndrome (SARS); in 2012, Middle East respiratory syndrome (MERS); and now, COVID-19 [7]. It has been established that SARS-CoV-2 has ~80% similarity in genetic sequence with other coronaviruses, especially with SARS-CoV [8]. It can easily spread through contact or respiratory droplets. Most of the infected patients are asymptomatic, which creates a real challenge to search such infected persons and control further spreading of COVID-19. Therefore, travel restrictions, social distancing, personal hygiene, and quarantine measures have been implemented by many countries to curb further spread of this virus [9].

As this global pandemic has recently appeared and developing a new medicine for a disease usually takes a lot of time, to date, there is no side-effect-free vaccine or drug available for the prevention or the treatment of COVID-19. However, certain active pharmaceuticals, developed for other diseases, such as anti-malarial (Chloroquine and Hydroxychloroquine), anti-Ebola (Remdesivir, etc.), and vaccines from Moderna, Pfizer, etc., have gained emergency use authorization from drug regulatory agencies in many countries for management of critical COVID-19 infections [10–12]. The appearance of new variants of SARS-CoV-2 with higher rate of spread is a concern [13]. In another approach, convalescent plasma containing antibodies from COVID-19 recovered some of the patients; oxygenation [14] and ventilation are also among the possible treatment strategies. However, these therapies are not ultimately effective. Therefore, the scientific community is still working on the development of specific COVID-19 drugs, and studies are going on in several directions for different type of patients. In this regard, a good number of researchers employed molecular docking and Quantitative Structure–Activity Relationship (QSAR) for virtual screening to identify novel hits for SARS-CoV-2 [11,15,16].

For the fast and cheaper development of medicines treating SARS-CoV-2, drug repositioning strategy is proposed. This approach is based on exploring the anti-SARS-CoV-2 potential of leverage molecules earlier reported to have inhibitory activity against genetically related viruses such as SARS-CoV.

A high semblance of SARS-CoV-2 with SARS-CoV is a primary hope to develop a safer and cheaper drug candidate. That is, the existing knowledge about SARS-CoV, a high similarity of SARS-CoV-2 with SARS-CoV, and a compound previously tested for SARS-CoV is a viable candidate to develop a therapeutic drug. It is rational to believe that a SARS-CoV inhibitor could also be equipotent against SARS-CoV-2. In other words, a cheaper, time-saving, feasible strategy is to leverage molecules earlier reported to have inhibitory activity against similar viruses such as SARS-CoV.

The e-Chemistry approach [17] (easy, efficient, economical, and eco-friendly) is to be followed to make the process of drug designing cheaper, more result oriented, and less time-consuming, with less trials and errors. The field of drug design has witnessed the rise of Computer Aided Drug Designing (CADD) as a fast, efficient, and cheaper tool for lead/drug optimization [18] with the advent of computers and allied resources. The CADD analysis is useful to identify the pharmacophoric features, predict the bio-chemical profile of a compound before its actual synthesis and bio-screening, and to understand the mechanism of action of molecules. The success of QSAR, molecular docking, and pharmacophore modeling has led to their regular utilization, to develop effective strategies to optimize a lead/drug candidate.

The field of QSAR has emerged as an attractive methodology to identify the important structural features and predict the desired activity/property, using a series of molecules. For a QSAR analysis to be successful, it should have a balance of mechanistic (Descriptive/Qualitative) and predictive (Quantitative) approaches [19–22]. In Qualitative QSAR, the main impetus is to understand the influence of molecular structure on the mode of

action or the biological mechanism of studied chemicals, whereas, in Quantitative QSAR, the main focus is on statistical validation of developed QSAR models, to verify their performance, i.e., statistical robustness and predictive ability for new chemicals [23].

Generally, researchers use molecular docking based HTS (high-throughput screening) for identifying hits, but its hit rate ranges between 0.1 and 0.01% [24]. Therefore, the overall cost of drug discovery substantially increases as most of the tested compounds are regularly found to be false positives toward the desired bioactivity [25]. On the other hand, QSAR-based virtual screening (VS) has been used to identify hits for drug designing, and its typical hit rates ranges from 1 to 40% [24]. This higher success rate not only highlights the advantages of using QSAR-based VS, it also brings QSAR modeling in tune with the final needs of experimental medicinal chemists to discover bioactive compounds. In other words, this approach extends the focus of QSAR modeling from accomplishing statistically significant training set models toward utilizing the validated models to rank molecules for subsequent biological evaluation. In the present work, a database of food compounds and a good number of heterocyclic variants of most active compound 1 were used to perform the QSAR-based VS. An obvious reason for choosing the food compounds is the innate ability of many foods compounds to have immune stimulatory effects without toxicity [26,27].

Molecular docking is a contemporary structure-based drug design approach, which involves simulations of interactions between a lead/drug molecule with a target enzyme. It provides an in-depth knowledge and understanding of binding patterns to identify the important structural features of a small molecule that are responsible for its binding with the enzyme [28].

Thus, a concomitant use of QSARs and molecular docking not only provides consensus and complementary information about prominent structural features but it is useful also to predict the bio-activity of a compound before its wet lab synthesis and testing. Therefore, in the present work, we employed QSARs, QSAR-based virtual screening, and molecular docking analyses to achieve the desired goals.

2. Results

In the present work, the QSAR analysis and molecular docking were performed to identify important structural features. The QSAR model was build using easily interpretable molecular descriptors to correlate them with structural features. The five-parametric GA-MLR model has good external predictive ability with the presence of easily understandable molecular descriptors along with interpretation in terms of structural features. Even though, in the present analysis, a straight evaluation of Ki values of the molecules of the dataset was performed to explain the effect of a specific descriptor, it is important to note that the combined or converse effect of unknown factors or other molecular descriptors could have a substantial effect on the Ki value of a molecule.

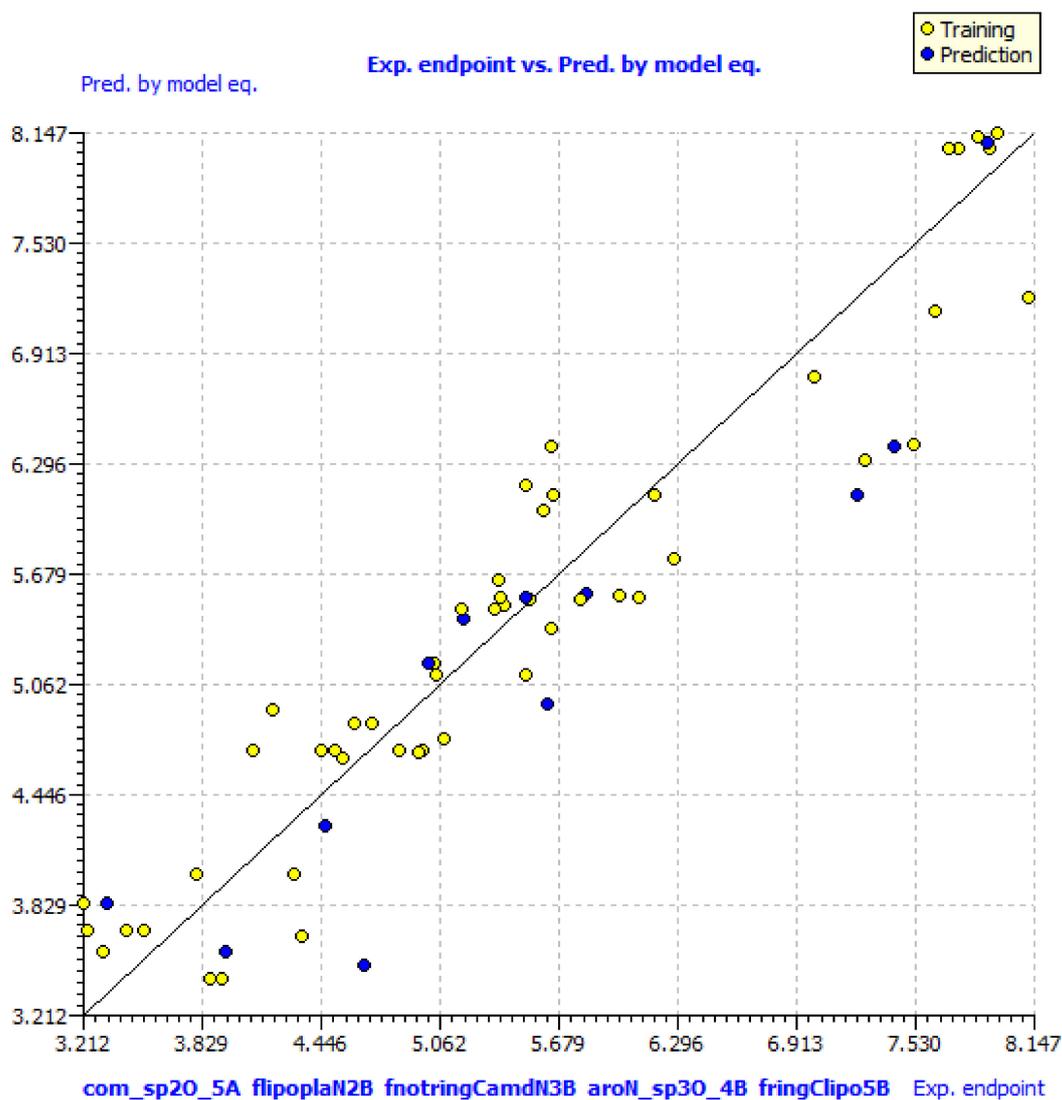
GA-MLR QSAR Model

$$\text{pKi} = 4.618 (\pm 0.415) + 2.774 (\pm 0.475) * \text{fnotringCamdN3B} + 0.762 (\pm 0.135) * \text{aroN_sp3O_4B} + 0.035 (\pm 0.021) * \text{fringClipo5B} + 0.962 (\pm 0.262) * \text{flipoplaN2B} - 0.279 (\pm 0.089) * \text{com_sp2O_5A}$$

$$R^2_{tr} = 0.898, R^2_{adj.} = 0.886, LOF = 0.291, RMSE_{tr} = 0.432, MAE_{tr} = 0.349, CCC_{tr} = 0.947, s = 0.460, F = 77.120, R^2_{cv} (Q^2_{loo}) = 0.859, RMSE_{cv} = 0.507, MAE_{cv} = 0.402, CCC_{cv} = 0.926, Q^2_{LMO} = 0.848, R^2_{Yscr} = 0.102, Q^2_{Yscr} = -0.164, RMSE_{ex} = 0.648, MAE_{ex} = 0.537, R^2_{ex} = 0.799, Q^2 - F^1 = 0.795, Q^2 - F^2 = 0.795, Q^2 - F^3 = 0.769, CCC_{ex} = 0.893$$

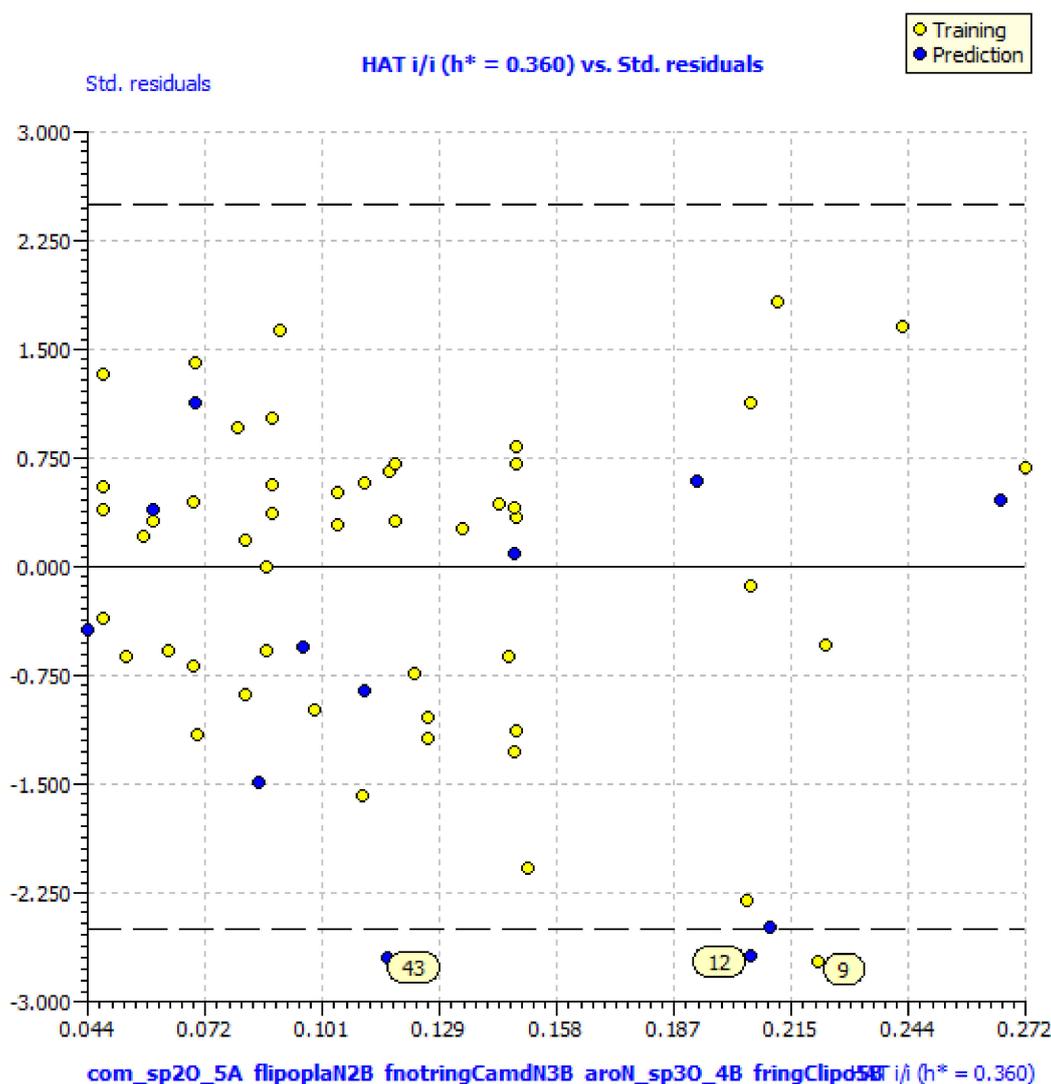
The different graphs associated with the developed QSAR model, i.e., experimental vs. predicted pKi and the Williams plot to assess applicability domain of model, are depicted in Figure 1. The developed QSAR model satisfies the recommended threshold values for a good number of statistical parameters (see Supplementary Materials for formulae and meaning of statistical parameters) [17,20–22,29–32]. The close values of R^2_{tr} (coefficient of determination), $R^2_{adj.}$ (adjusted coefficient of determination), LOF (lack of fit), and R^2_{cv} (Q^2_{loo}) (cross-validated coefficient of determination for leave-one-out) indicate that the

model comprises optimum number of variables and free from over-fitting. The high value of Q^2_{LMO} (cross-validated coefficient of determination for leave-many-out) confirms the acceptable internal validation of a model. The low value of R^2_{Yscr} and Q^2_{Yscr} (Y-scrambling related parameters) suggests that the developed model is free from chance correlations. The high value of R^2_{ex} (external coefficient of determination), Q^2-F^n and CCC_{ex} (Concordance Correlation Coefficient) vindicate the acceptable external predictive ability. The F (Fischer F-ratio) value vindicates the high statistical significance of the model. In addition, an analysis of the correlation matrix (see Supplementary Materials Table S3) indicates that the molecular descriptors have low correlation with each other.



(a)

Figure 1. Cont.



(b)

Figure 1. Different graphs associated with the developed Quantitative Structure–Activity Relationship (QSAR) model: (a) experimental vs. predicted pKi and (b) Williams plot to assess applicability domain of model. (molecules out of applicability domain have been shown with their serial numbers).

3. Discussion

fnotringCamdN3B: The molecular descriptor fnotringCamdN3B signifies the frequency of occurrence of amide Nitrogen atoms exactly at three bonds from the non-ring Carbon atoms. If the same amide Nitrogen atom was simultaneously present at one or two bonds from any other non-ring Carbon atom, then it was excluded during the calculation of fnotringCamdN3B. This molecular descriptor has positive coefficient in the developed model, therefore increasing its value could lead to a better Ki value. As the amide Nitrogen is always a constituent of amide group, therefore it is reasonable to believe that this molecular descriptor highlights the importance of amide groups. A simple analysis reveals that the peptidomimetic molecules (10–13, 15, 20, and 21) of the present dataset contain such a combination of amide Nitrogen and non-ring Carbon atoms as pyrrolidin-2-one ring. Konno et al., and Regnier et al. [33,34] have also reported that the pyrrolidin-2-one ring occupies the S1 pocket. This indicates the crucial role of pyrrolidin-2-one ring for potent inhibitory activity against SARS.

aroN_sp3O_4B: The molecular descriptor aroN_sp3O_4B points out the presence of aromatic Nitrogen atoms within four bonds from sp³-hybridized Oxygen atoms. It is another molecular descriptor with positive coefficient in the developed model, therefore increasing the number of such combinations could augment the Ki value. Only eight molecules, which are also the top eight active candidates, possess a higher value of this molecular descriptor. These eight molecules are benzotriazole derivatives. This could be a reason for better Ki value for benzotriazole derivatives. The docking pose for molecule **1** supports the importance of aromatic Nitrogen atoms (see the docking Section 3.2). The three Nitrogen atoms of the Benzotriazole ring are present within four bonds from the sp³-hybridized Oxygen atom.

fringClipo5B: A molecular descriptor which increases the Ki value, due to positive coefficient in the developed QSAR model, is fringClipo5B. It represents the frequency of occurrence of lipophilic atoms exactly at five bonds from ring Carbon atoms. If the same lipophilic atom is simultaneously present at four or less bonds from any other ring Carbon atom, then it was excluded during the calculation of fringClipo5B. Thus, this descriptor highlights the importance of ring Carbon atoms and lipophilic atoms. Another molecular descriptor which indicates the importance of lipophilic atoms is **flipoplaN2B**. It stands for the frequency of occurrence of planer Nitrogen atoms exactly at two bonds from the lipophilic atoms. If the same planer Nitrogen atom is simultaneously present at one bond from any other lipophilic atom, then it was excluded during the calculation of flipoplaN2B. These two molecular descriptors indicate that specific combinations of lipophilic atoms with ring Carbon atoms and planer Nitrogen atoms could be useful to increase the Ki value for a molecule.

com_sp2O_5A: The molecular descriptor com_sp2O_5A represents the total number of sp²-hybridized Oxygen atoms present within 5 Å from the center of mass of the molecule. It has a negative coefficient in the developed QSAR model; therefore, increasing the number of such Oxygen atoms could result in decreased Ki value for a molecule for main protease. A poor Ki value for molecule **43** (4.68 M), **49** (4.34 M), **53** (3.95 M), **54** (3.95 M), **55** (3.87 M), and **60** (3.32 M) could be attributed to high frequency of occurrence of such Oxygen atoms within 5 Å from center of mass of molecules (com_sp2O_5A = 5). Interestingly, all these molecules are peptidomimetic derivatives. There are five peptidomimetic molecules present in the top 15 molecules from the present dataset, which contain two to four such Oxygen atoms within 5 Å from center of mass of molecules. Therefore, it is reasonable to say that the value of **com_sp2O_5A** should be less than five to have good Ki value.

In the present QSAR analysis, the constituent molecular descriptors of the GA-MLR QSAR model have provided salient and hidden information about the structural features related to diverse set of molecules tested for their activity for the main protease (M_{pro}). It is essential to understand that no single molecular descriptor can completely explain the observed distribution of Ki for such a diverse set of molecules. That is, the performance of the developed QSAR model relies on the concomitant use of constituent molecular descriptors.

3.1. QSAR-Based Virtual Screening

The SMILES notations, calculated values of molecular descriptors, pKi and Ki for different variants of compound **1** and food compounds used for virtual screening are available in Supplementary Materials. For the sake of convenience, herein, we show the ten most active molecules from different variants of compound **1** and the ten most active molecules from the food database, as predicted by the developed QSAR model. From Table 1, it is clear that an increase in the number of aromatic Nitrogen atoms could lead to a better Ki value for a molecule.

Table 1. Predicted (Pred) pKi and Ki for different variants of molecule 1 (#1 to 10) and food compounds (#11 to 20) used for QSAR-based virtual screening.

SN	Pred-pKi (M)	Pred-Ki (nM)
S341	9.226	0.594
S337	9.121	0.757
S342	9.121	0.757
S338	9.086	0.82
S339	8.842	1.439
S340	8.807	1.56
S293	8.499	3.17
S161	8.464	3.436
S251	8.464	3.436
S317	8.464	3.436
FoodS8291	8.46	3.467
FoodS6189	8.355	4.416
FoodS677	8.32	4.786
FoodS3568	8.32	4.786
FoodS4426	8.251	5.61
FoodS6919	8.251	5.61
FoodS4135	8.181	6.592
FoodS7495	8.181	6.592
FoodS1368	7.971	10.691
FoodS4841	7.936	11.588

3.2. Docking Analysis

Main protease (Mpro), also known as Nsp5 and 3CLpro, is a cysteine protease enzyme active in homodimer form only [17,35,36]. It is an essential enzyme for SARS-CoV-2, which participates in cleavage process of H-CoV polyproteins [6,17,35–38]. It has been established that it consists of three domains, domain I (residues 8–101), domain II (residues 102–184), and domain III (residues 201–303). A long loop (residues 185–200) connects domains II and III. The active site, which is highly conserved among all CoV's Mpros and usually composed of four sites (S1', S1, S2, and S4), is situated in the gap between domains I and II [6,17,35–38]. The catalytic dyad of Cys145 and His41 is an important feature of active site of this protein [6,17,35–38]. This protein is necessary for the processing of polyproteins and operates at 11 cleavage sites on the large polyprotein 1ab [8,35,39,40]. It exclusively breaks polypeptide sequences after a glutamine residue, whereas no human host-cell proteases are known to have such substrate specificity; therefore, an inhibitor of Mpro could be safe for humans [8,35,39,40]. In the current study, the dataset molecules and a known inhibitor 13b were docked inside the active site of Mpro. The X-ray resolution of 13b is available, which was used to validate the docking protocol. Figure 2a shows the active site of the Mpro protein. The validation is performed by removing the crystal ligand 13b from the active site and reloading it again. The alignment of Mpro with the crystal ligand 13b and the redocked ligand is depicted in Figure 2b–d, which indicates that the docking protocol is acceptable (see Figure 2).

The docking scores for all the molecules are present in the Supplementary Materials. Table 2 contains the docking scores for the most active ten molecules. Even though molecule 1 is most active, molecule 10 has a better docking score. In fact, molecule 10 has the sixth best docking score (see Table 3). This means that the molecule 10 fits better inside the big active site of Mpro and fills it completely. This could be attributed to the larger size of molecule 10, as compared to molecule 1.

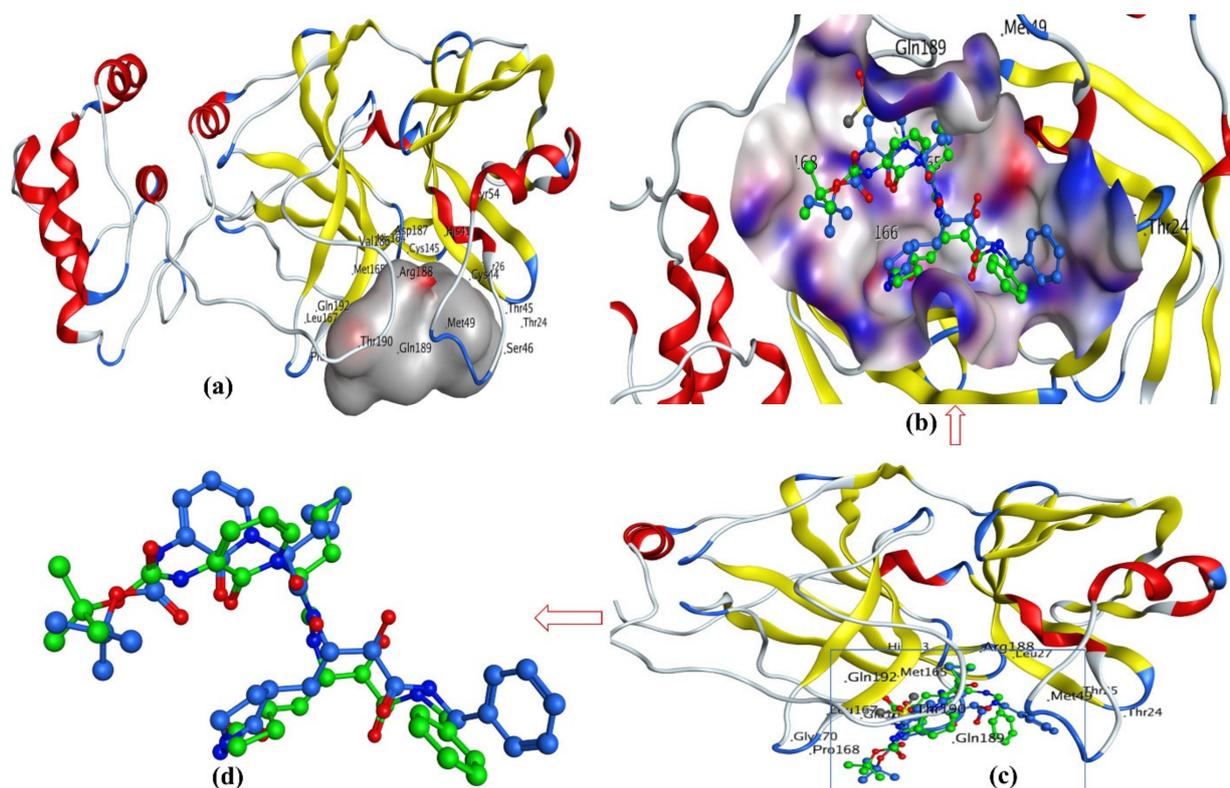


Figure 2. Docking and X-ray determined pose for **13b** to validate the docking protocol: (a) main protease (Mpro) with active site (gray-colored contour), (b) ligand **13b** in active site with surface, (c) without surface, and (d) comparison of docking pose for **13b** with X-ray determined pose.

Table 2. Docking score for top ten active molecules from the selected dataset of 62 molecules.

SN	Structure	Docking Score (kcal/mol)
1		−5.997
2		−7.008
3		−6.42

Table 2. Cont.

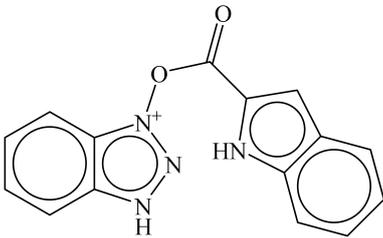
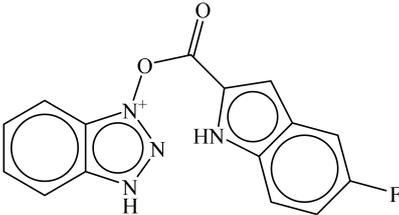
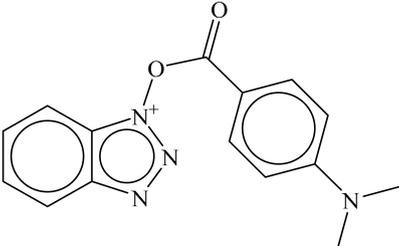
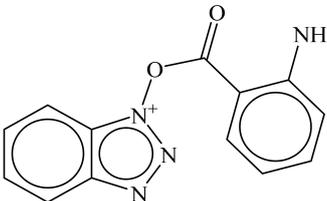
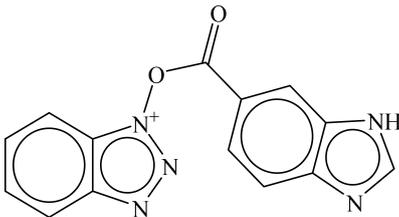
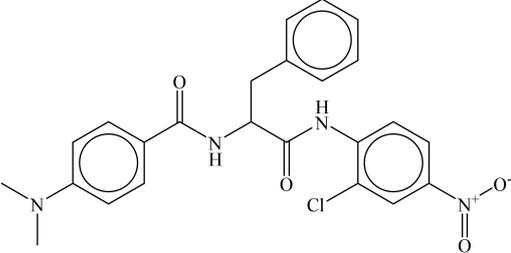
SN	Structure	Docking Score (kcal/mol)
4		-5.886
5		-6.041
6		-6.48
7		-5.866
8		-6.331
9		-7.774

Table 2. Cont.

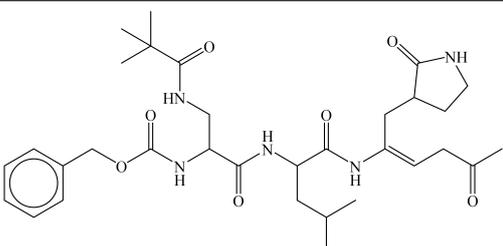
SN	Structure	Docking Score (kcal/mol)
10		-9.931

Table 3. Structures and docking score for molecules having higher docking score in the present dataset.

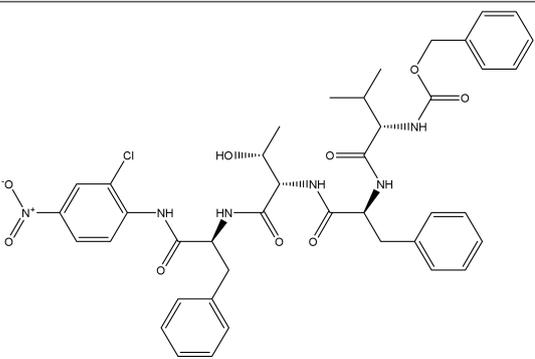
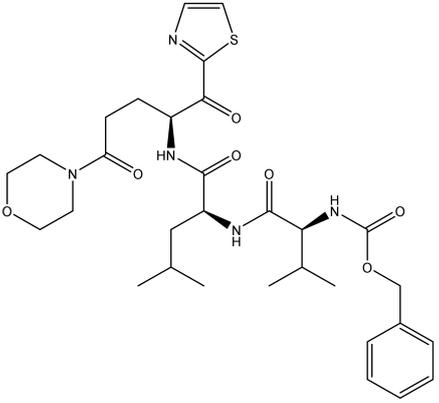
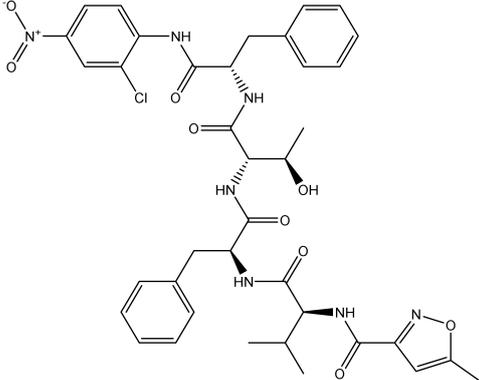
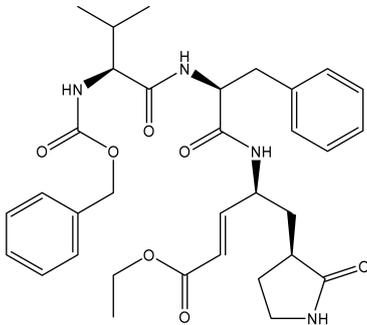
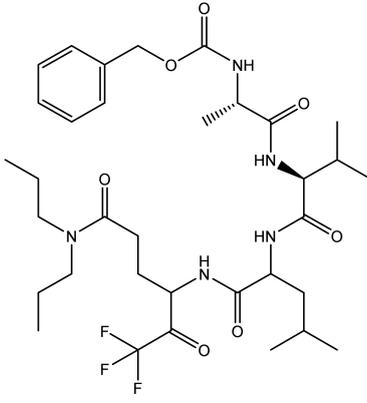
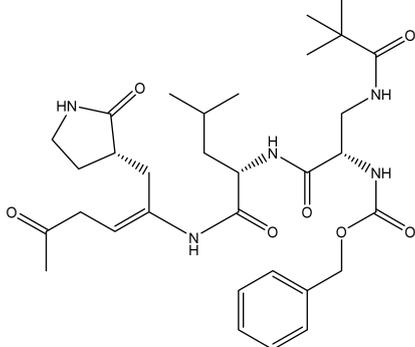
SN	Structure	Docking Score (kcal/mol)
18		-10.285
60		-10.259
19		-10.159

Table 3. Cont.

SN	Structure	Docking Score (kcal/mol)
21	 Chemical structure of molecule 21, featuring a benzotriazole core, a benzyl group, a propyl chain, and a pyrrolidine ring.	−10.026
57	 Chemical structure of molecule 57, featuring a benzotriazole core, a benzyl group, a propyl chain, and a trifluoromethyl group.	−9.975
10	 Chemical structure of molecule 10, featuring a benzotriazole core, a benzyl group, a propyl chain, and a pyrrolidine ring.	−9.931

For the sake of convenience, we present the docking pose and pharmacophore model for most active molecule **1** as a representative example (see Figure 3). In addition, the docking scores and the interacting amino acids for the least and most active molecules **1** and **62** are tabulated in Table 4.

3.3. Docking Pose for Most Active Molecule

From the docking pose, it appears that most active molecule **1** has occupied the S1 and S2 pockets of the active site. The benzene ring of benzotriazole moiety is present inside S1 pocket in proximity of Leu141 and responsible for lipophilic interactions. The polar triazole ring of benzotriazole moiety is oriented toward Met165 due to polar interactions. The importance of triazole has been also highlighted by the **aroN_sp3O_4B** molecular descriptor in QSAR analysis. The linker -COO- moiety is responsible for the establishment of H-bonding (2.22 Å) with Glu166. The benzene part of indole moiety is near to Gln189 due to lipophilic interactions. The same observation is vindicated by the presence of two molecular descriptors **flipoplaN2B** and **fringClipo5B**, which highlight the importance of

lipophilic atoms in the newly developed QSAR model. The -NH- part of Indole ring is accountable for the H-bond formation with Asp187 (2.31 Å). Thus, QSAR and molecular docking analyses provide similar, as well as complementary, results.

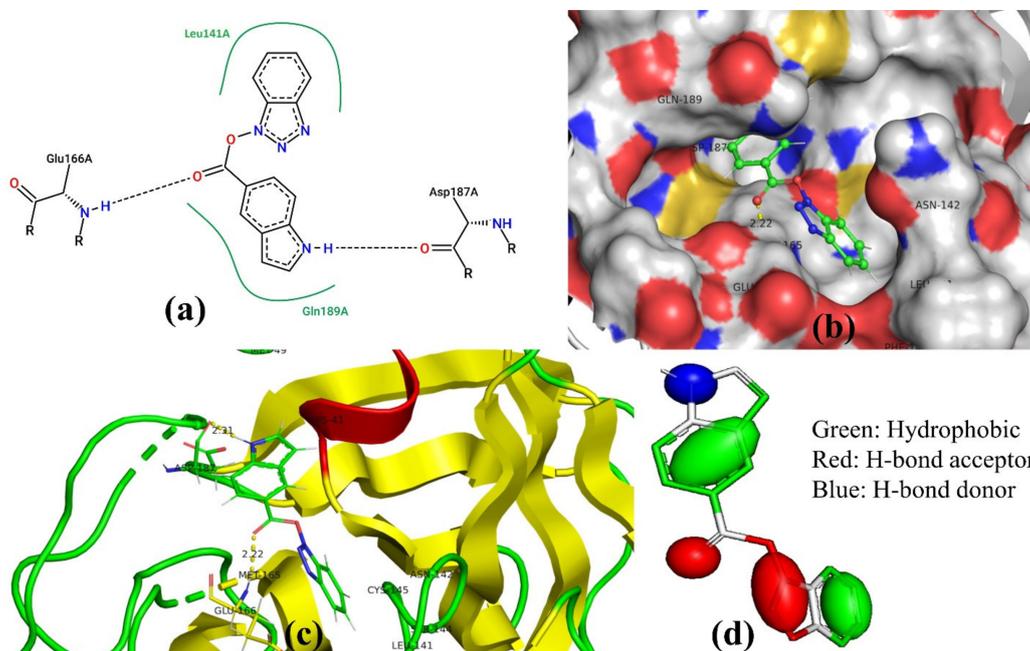


Figure 3. The docking pose for most active molecule **1** inside the active site of Mpro: (a) 2D representation of interactions, (b) with molecular surface, (c) without molecular surface, and (d) pharmacophore model.

Table 4. Docking scores along with the interacting amino acids for the least and most active molecules **1** and **62**.

SN	List of Interacting Amino Acids	Docking Score
1	His41, Met49, Tyr54, Phe140, Leu141, Asn142, Ser144, Cys145, His163, His164, Met165, Glu166 (strong H-bond), Asp187 (strong H-bond), Arg188, Gln189	−10.285
62	His41, Met49, Tyr54, Phe140, Leu141, Asn142, Ser144, Cys145, His163, His164, Met165 (weak H-bond), Glu166, Pro168, Val186, Asp187, Arg188, Gln189 (weak H-bond)	−9.358

The pharmacophore model generated by using the docking pose of most active compound **1** (see Figure 3c,d) indicates that a lipophilic region concomitant with a large H-bond acceptor region could avail in establishing useful interactions with the residues of S1 pocket, whereas a combination of large lipophilic region with H-bond donor is helpful to have key interactions with S2 pocket of Mpro.

3.4. MD Simulations and MMGBSA Binding Free-Energy Calculations

The final convergence and the stability of apo-Mpro and **1** bound Mpro were assessed in MD simulation studies. After 50 ns of convergence, apo-Mpro displayed RMSD of C α backbone 2.2 Å. Initially, until 25 ns the backbone seemed to be stable but later from 25 to 50 ns of simulation the RMSD enumerated by average deviation 1.0 Å (Figure 4a, red). Whereas, RMSD plot of molecule **1** bound Mpro displayed the average deviation of 0.8 Å owing to its stable conformation over apo-Mpro protein. The interactions of protein and the molecule **1** displayed little RMSD differences (~0.8 Å) is acceptable conferred its stable conformation (Figure 4a, black). Radius of gyration is the indicator of size and compactness of the protein in the ligand bound state displayed in Figure 4b. The Rg plot of C α -backbone of apo-Mpro (Figure 4b, red) has least compactness due to lowering of fluctuations 22.8 to 22.2 Å with an average of 22.35 Å from the beginning to end of the 50 ns simulation. On the other hand, Mpro bound complex with **1** displayed a constant gyration of the C α backbone 22.4 Å with least deviations (Figure 4b, black). This signifies that the molecule **1**

bound to Mpro complex is highly stable in comparison to the apo-Mpro. The RMSF plot displayed the stable conformation of each amino acid residues during of the simulation displaying fewer fluctuations in each amino acid positions in the complex of 1 bound Mpro (Figure 4c, black).

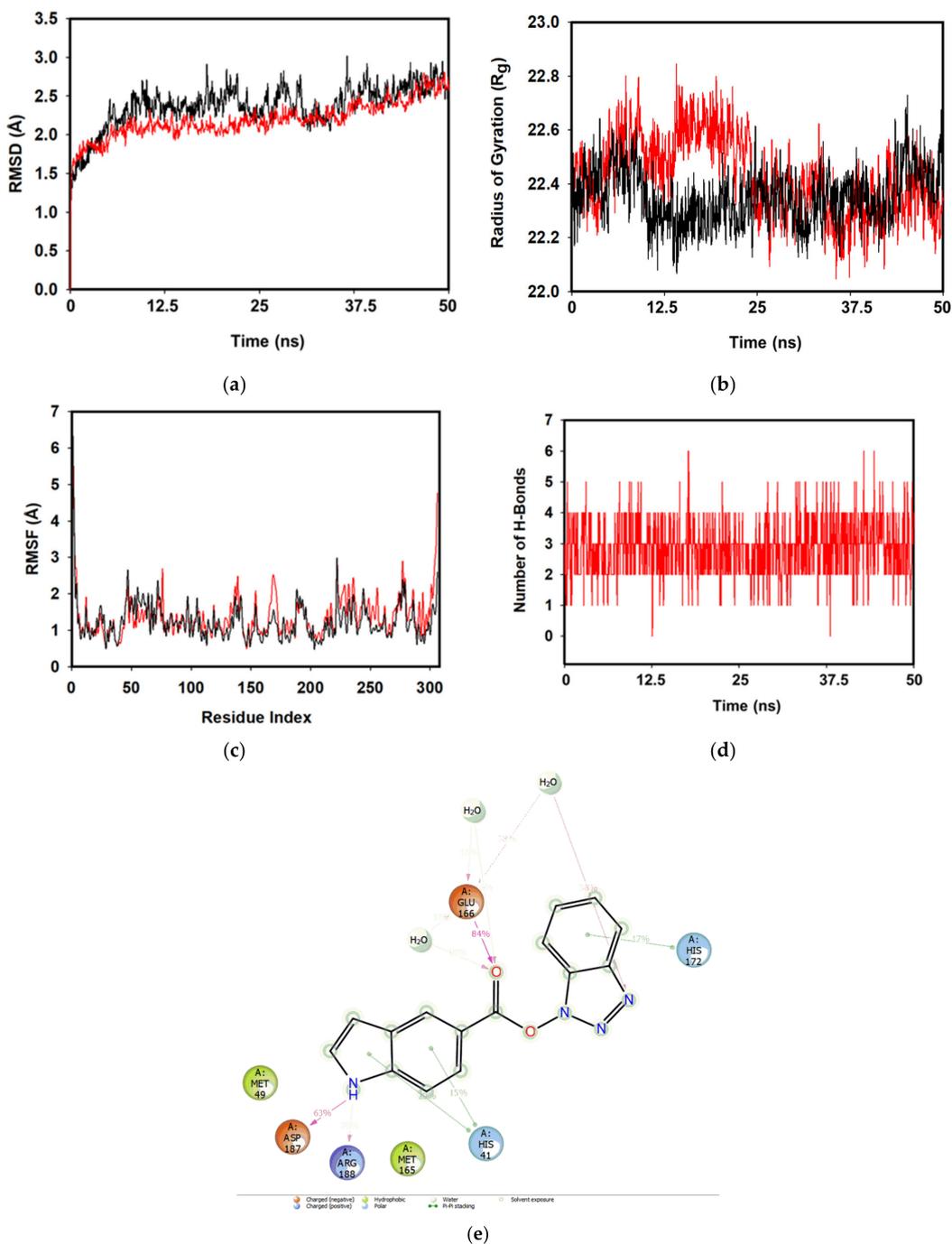


Figure 4. RMSD plots of (a) apo-Mpro (red) and 1 bound Mpro complex (black) implying overlapping vibrations and smooth interactions between protein and ligand. (b) Radius of gyration of apo-Mpro (red) and 1 bound Mpro complex (black), owing to their compactness after 50 ns simulation. (c) RMSF plots apo-Mpro (red) and 1-Mpro (black) displaying least fluctuations throughout the 50 ns convergence during simulation. (d) H-bonds plot displaying the number of H-bonds formed during to total time scale of 50 ns of simulation. (e) Two-dimensional interaction plot of Mpro with 1 displaying the involvement of amino acids making varying interaction after 50 ns of simulation. The average H-bonds count displayed three numbers of bonds formation from the beginning to end of the simulation with the ligand 1 molecule (Figure 4d).

Mpro binding site for **1** displayed the major interactions, such as forming water bridges with negatively charged Glu166 and H-bond (Figure 4e). Asp187 and Arg 188 residues were also involved in H-bonding formation with the **1** molecule. Other residues include His41 and His172 involved in polar contacts, as well as pi-pi stacking, respectively. Therefore, the varying interactions played a critical role in stabilizing the whole complex (Figure 4e).

Structural superimposition of initial and final frames of the **1** bound Mpro displayed mere changes at the binding site. The secondary fold of the protein at the converged state (at 50 ns) showed alteration at the α -Helical turn (Figure 5a, arrow) that facilitated better orientation of the molecule **1** geometry (red). Moreover, there was a tilt of the ligand (arrow, yellow = 0 ns; and arrow, red = 50 ns) during the course of simulation (Figure 5a). This signifies that the ligand was docked well at the binding site of the Mpro. The free energy change of ligand binding in the main protease Mpro of SARS-CoV-2 using MMGBSA calculations displayed the average (dG) binding = -52.54 ± 4 kcal/mol. The significantly high binding energy implied the higher affinity of molecule **1** with Mpro and perhaps opened the new arena for a novel inhibitor drug against SARS-CoV-2.

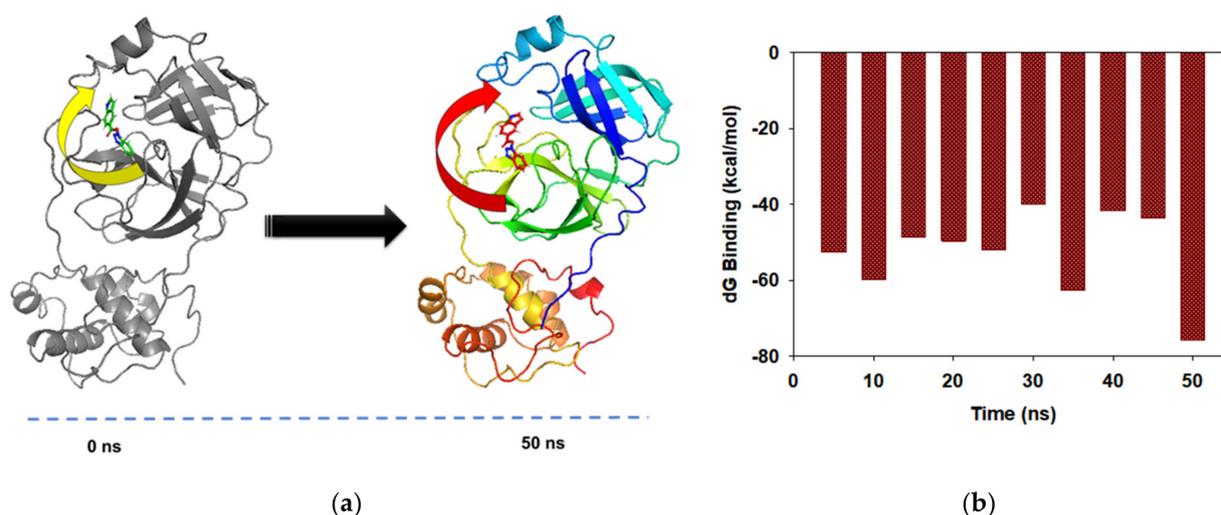


Figure 5. (a) Structural superimposition of first frame (0 ns) and last frame (50 ns) of **1** bound Mpro complex after simulation. The conformational change of secondary structure (arrow) observed at the **1** bound site and geometry of ligand displayed at 0 ns (yellow) and 50 ns (red). (b) Free-energy decomposition of binding energies at every 5.0 ns frame in MMGBSA calculations for 50 ns simulation.

4. Materials and Methods

4.1. QSAR Analysis and Model Building

In the present work, a dataset of 351 molecules was downloaded from Binding database (<https://www.bindingdb.org/bind/index.jsp> (accessed on 28 September 2020)). The dataset contains diverse compounds, thus covering enough chemical space. Then, dataset curation involving the removal of salts, duplicates, and entries with ambiguous K_i values led to a reduced dataset of sixty-two compounds. However, the remaining sixty-two compounds have a wide range of activities against SARS-CoV ($K_i = 7.5$ to 614,000 nM), thus justifying the composition of the dataset. Afterward, prior to subsequent QSAR analysis, the reported K_i values were converted to pK_i ($pK_i = -\log_{10}K_i$). The Simplified Molecular-Input Line-Entry System (SMILES) notation for all selected molecules, along with their reported activity values K_i and pK_i are present in the Supplementary Materials. Figure 6 contains the representative examples of different classes of molecules used in the present work. Moreover, Table 5 contains the SMILES notations, K_i (nM), and pK_i (M) for the top five most active and least active molecules.

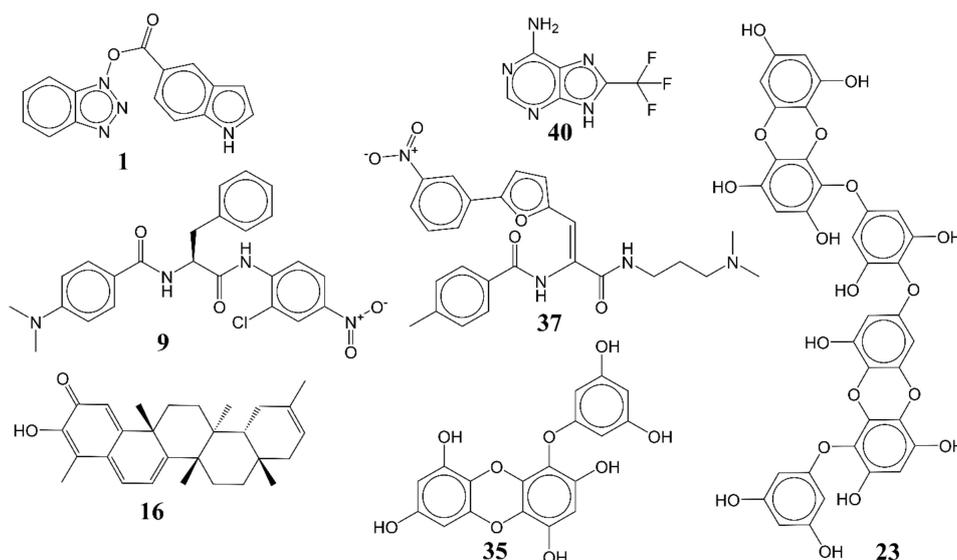


Figure 6. Representative examples of different classes of molecules used in the present work. (Bold numbers indicate serial number of molecules in the data set).

Table 5. Simplified Molecular–Input Line–Entry System (SMILES) notations, Ki(nM), and pKi(M) for top five most active and least active molecules.

SN	SMILES	Ki (nM)	pKi (M)
1	<chem>c1cccc(c12)n(nn2)OC(=O)c(c3)ccc(c34)[nH]cc4</chem>	7.5	8.125
2	<chem>c1cccc(c12)n(nn2)OC(=O)c3ccc(cc3)N(CC)CC</chem>	11.1	7.955
3	<chem>CNc(cc1)ccc1C(=O)On(nn2)c(c23)cccc3</chem>	12.1	7.917
4	<chem>c1cccc(c12)n(nn2)OC(=O)c(c3)[nH]c(c34)cccc4</chem>	12.3	7.91
5	<chem>c1cccc(c12)n(nn2)OC(=O)c(c3)[nH]c(c34)ccc(F)c4</chem>	13.8	7.86
58	<chem>CCCN(CCC)C(=O)CC[C@@H](C(=O)C(F)(F)F)NC(=O)[C@H](CC(C)C)NC(=O)[C@H](C(C)C)NC(=O)OCc1cccc1</chem>	363,000	3.44
59	<chem>CCN(CC)C(=O)CC[C@@H](C(=O)c1nccs1)NC(=O)[C@H](CC(C)C)NC(=O)OCc2cccc2</chem>	462,000	3.335
60	<chem>C1COCCN1C(=O)CC[C@@H](C(=O)c2nccs2)NC(=O)[C@H](CC(C)C)NC(=O)[C@H](C(C)C)NC(=O)OCc3cccc3</chem>	478,000	3.321
61	<chem>CCCN(CCC)C(=O)CC[C@@H](C(=O)C(F)(F)F)NC(=O)[C@H](CC(C)C)NC(=O)OCc1cccc1</chem>	584,000	3.234
62	<chem>CCN(CC)C(=O)CC[C@@H](C(=O)c1nccs1)NC(=O)[C@H](C(C)C)NC(=O)OCc2cccc2</chem>	614,000	3.212

SN = serial number.

OpenBabel [41] ver. 2.4 was used to convert SMILES notation to 3D structures, using MMFF94 force field. The 3D structures were then used to calculate a myriad of molecular descriptors using “PyDescriptor” [42]. This led to a cluster of more than 15,000 molecular descriptors for each molecule. In the next step, QSARINS-2.2.4 [43,44] was used to remove constant, nearly constant and highly correlated ($|R| > 0.90$) molecular descriptors to avoid the inclusion of multi-collinear and redundant molecular descriptors in the process of development of a robust QSAR model. This significantly reduced the size of molecular descriptor pool (11,299 molecular descriptors), still containing a variety of 1D to 3D molecular descriptors, thus covering a broad descriptor space.

After that, GA-MLR (Genetic Algorithm–Multilinear Regression) [45] was used for subjective feature selection (SFS), using QSARINS-2.2.4 with default settings, except that the number of generations was set to 10,000. During SFS, the dataset was split randomly

into the training set (80%) and the prediction (test or external) set (20%). The prediction set was not used during model building. It was only used for validation of the developed model. The heuristic search was limited to five variables to avoid over-fitting and enhance simplicity of the model. The GA-MLR module of QSARINS-2.2.4 uses Q^2 as a fitness function. The developed QSAR model was subjected to thorough statistical validation (internal and external validation) according to Organisation for Economic Co-operation and Development (OECD) principles. Model with a high internal and external predictive ability has been reported. Further details about the QSAR model development are available in the literature [20,29–31].

4.2. QSAR-Based Virtual Screening

For QSAR-based virtual screening, the most active compound **1** served as a template to generate a good number of heterocyclic variants using RDKit (RDKit: open-source cheminformatics; <http://www.rdkit.org> (accessed on 28 September 2020)). This resulted in a pool of 360 different heterocyclic variants. Moreover, a database of 26,467 food compounds was downloaded from FooDB (<http://foodb.ca/> (accessed on 28 September 2020)), followed by the application of rule of five; the removal of duplicates, salts, and metal derivatives led to a reduced dataset of 8453 molecules. Thus, 8813 (360 + 8453) molecules were used for QSAR-based VS. Prior to molecular descriptor calculations, the 3D structures of molecules were prepared in the same way as modeling set. Then, the molecular descriptors were calculated, and the properly validated five-parametric QSAR model was used to predict the biological property of novel compounds.

4.3. Molecular Docking Analysis

The pdb file for main protease was fetched from SWISS-MODEL (<https://swissmodel.expasy.org/repository/species/> (accessed on 28 September 2020)). The pdb 6lu7 [37] was selected on the basis of X-ray resolution and completion of the sequence. Before actual docking simulations, the health of the protein was checked by plotting Ramachandran's plot [46] (see Figure 7). The optimized protein is acceptable for docking analysis (see Figure 7). Prior to actual docking analysis, the native ligand N3 (a peptidomimetic inhibitor) was removed. All the compounds were docked in the active site, but for the sake of convenience, herein, the docking pose for most active molecule as a representative is depicted.

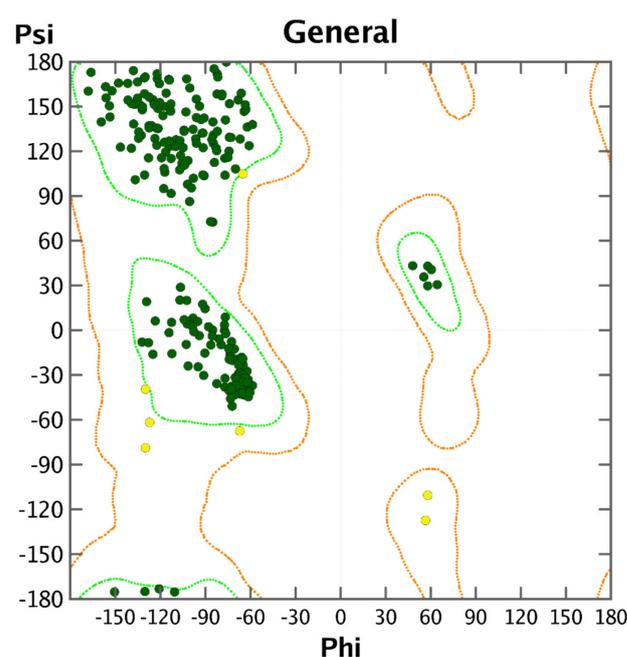


Figure 7. Ramachandran plot for main protease of SARS-CoV-2.

For the molecular docking analysis, the software NRGSuite [47] was used. This free software is available as a plugin for PyMOL (www.pymol.org (accessed on 28 September 2020)). It has the ability to detect the surface cavities in a protein and use them as target binding-sites for docking simulations with the help of FlexAID [48]. It uses genetic algorithm for conformational search, simulates ligand and side-chain flexibility and allows for the simulation of covalent docking. In the present work, flexible–rigid docking protocol was employed with following default settings to get optimum performance from NRGsuite: binding sites input method—spherical shape (diameter: 18Å); spacing of three dimensional grid—0.375Å; side chain flexibility—no; ligand flexibility—yes; ligand pose as reference—no; constraints—no; Hetero groups—included water molecules; van der Waals permeability—0.1; solvent types—no type; number of chromosomes—1000; number of generations—1000; fitness model—share; reproduction model—population boom; and number of top complexes—5 [22].

For validation of molecular docking, the molecule **13b**, a known peptidomimetic inhibitor of Mpro [35], was used to validate the docking protocol.

4.4. Molecular Dynamics and Binding Energy Calculations

Molecular Dynamics Simulation (MDS) studies were carried out in order to determine stability and convergence [49] of main protease main protease (Mpro) with and without molecule **1**. To set up the simulations, initially, the systems were built for complex **1**-Mpro and apo-Mpro, respectively, in the system builder. For this purpose, Desmond 2018-4 was used to set up the initial parameters within explicit SPC water model orthorhombic box $4.0 \times 4.0 \times 4.0$ Å. The protease ligand complex and apo-Mpro were neutralized with NaCl salt by adding 0.15 M Na⁺ ions. The ASL module was used to select the specific residues of ligand and protein molecules for the better prepared systems, which were relaxed by using the Desmond default protocol of relaxation [50]. MDS run of 20 ns was set up at a constant temperature and constant pressure (NPT) for the final production run. The NPT ensemble was set up by using the Nosé–Hoover chain coupling scheme [51], at a temperature of 300 K. for final production and throughout the dynamics with relaxation time 1 ps. RESPA integrator was used to calculate the bonding interactions for a time step 2 fs [52]. All other parameters were associated in the settings followed as described elsewhere [53]. After the final production run, the simulation trajectories of main protease alone (apo-Mpro) and complexed with molecule **1** were analyzed for the final outcome of root mean square deviation (RMSD), root mean square fluctuation (RMSF), and number of hydrogen bonds formation derived from simulation. Binding energies of the complexes were calculated by using MMGBSA [42] for every 1 ns trajectories, till 20 ns and the average binding energies with standard deviations were measured for accurate binding approximation and stability described elsewhere [49].

5. Conclusions

In the present study, the repositioning approach of SARS-CoV inhibitors, alongside QSARs, QSAR-based virtual screening, and molecular docking–molecular dynamics analyses were performed to identify new potent inhibitor candidates against main protease of SARS-CoV-2. A five parametric GA–MLR QSAR model was developed to identify the main pharmacophoric features that govern the Mpro inhibitory activity. Internal and external validation and other stringent tests according to OECD principles were performed for the developed model. From the present analysis, pharmacophoric features like non-ring Carbons and Nitrogens, amide Nitrogen, sp²-hybridized Carbons, lipophilic atoms, etc., appeared as prominent features that govern the Mpro inhibitory activity. The developed QSAR model possesses high external predictive ability and robustness for fitting and internal validation. In addition, virtual screening successfully offered new derivatives of molecule **1** and from food database with improved Ki values in the range from 0.59 to 11.59 nM. Additionally, molecular docking of active candidate **1** within the active pocket of Mpro shed the light about the important pharmacophoric moieties involved in the binding

interactions which are responsible for the inhibitory potential. It appeared that molecule **1** occupied the S1 and S2 pockets of the active site. Molecular docking and MD analysis identified the crucial role of triazole and benzene rings in establishing lipophilic and H-bonding with the important residues like Leu141, Met165, Glu166, Asp187, Gln189, etc., of the active site Mpro. QSAR and molecular docking provided consensus, as well as complementary pharmacophoric features, which therefore have to be retained in developing potent and selective SARS-CoV-2 inhibitors. Lastly, the significantly high binding energy of compound **1** with Mpro vindicates the higher affinity and opens the new arena for a novel inhibitor drug against SARS-CoV-2. The results could be highly useful to develop a therapeutic agent for SARS-CoV-2.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ph14040357/s1> Table S1: SMILES notations, Ki, pKi, and values of all molecular descriptors used in the model. Table S2: Details regarding performance of model. Table S3: Correlation matrix for molecular descriptors present in the developed QSAR model, statistical parameters for used for validation of QSAR models and the formulae for the statistical parameters. Table S4: SMILES notations, pKi, Ki, and values of molecular descriptors for molecules used for QSAR-based virtual screening. Figure S1: Different graphs associated with model (a) graph of experimental vs. residual values (b) Y-scrambling plot.

Author Contributions: Conceptualization, M.E.A.Z. and S.A.A.-H.; formal analysis, V.H.M., S.A. and I.L.; data curation, S.A. and S.O.B.; writing, M.E.A.Z., V.H.M., A.G., and N.N.E.E.-S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors extend their sincere appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, KSA, for its support of this research through the Research Grant No. 20-13-12-011.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available in the supplementary section.

Acknowledgments: The authors acknowledge the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University, Riyadh, KSA, for its support of this research through the Research Grant No. 20-13-12-011. V.H.M. is thankful to Paola Gramatica and her team for providing the free copy of QSARINS 2.2.4.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

SMILES	Simplified Molecular-Input Line-Entry System
GA	Genetic Algorithm
MLR	Multiple Linear Regression
QSAR	Quantitative Structure–Activity Relationship
WHO	World Health Organization
ADMET	Absorption, Distribution, Metabolism, Excretion, and Toxicity
OLS	Ordinary least square
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
QSARINS	QSAR Insubria
HTS	High-throughput screening
MDS	Molecular Dynamics Simulation
Mpro	Main protease
OECD	Organisation for Economic Co-operation and Development

References

1. Gao, K.; Nguyen, D.D.; Chen, J.; Wang, R.; Wei, G.-W. Repositioning of 8565 Existing Drugs for COVID-19. *J. Phys. Chem. Lett.* **2020**, *11*, 5373–5382. [[CrossRef](#)]

2. WHO. Available online: <https://covid19.who.int/> or https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200414-sitrep-85-covid-19.pdf?sfvrsn=7b8629bb_4 (accessed on 4 January 2021).
3. Ko, M.; Jeon, S.; Ryu, W.S.; Kim, S. Comparative analysis of antiviral efficacy of FDA-approved drugs against SARS-CoV-2 in human lung cells. *J. Med. Virol.* **2020**. [[CrossRef](#)] [[PubMed](#)]
4. Zumla, A.; Chan, J.F.W.; Azhar, E.I.; Hui, D.S.C.; Yuen, K.-Y. Coronaviruses—Drug discovery and therapeutic options. *Nat. Rev. Drug Discov.* **2016**, *15*, 327–347. [[CrossRef](#)] [[PubMed](#)]
5. Peiris, J.S.M.; Guan, Y.; Yuen, K.Y. Severe acute respiratory syndrome. *Nat. Med.* **2004**, *10* (Suppl. S12), S88–S97. [[CrossRef](#)]
6. Fu, L.; Ye, F.; Feng, Y.; Yu, F.; Wang, Q.; Wu, Y.; Zhao, C.; Sun, H.; Huang, B.; Niu, P.; et al. Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease. *Nat. Commun.* **2020**, *11*, 4417. [[CrossRef](#)]
7. Pillaiyar, T.; Meenakshisundaram, S.; Manickam, M. Recent discovery and development of inhibitors targeting coronaviruses. *Drug Discov. Today* **2020**, *25*, 668–688. [[CrossRef](#)]
8. Jeon, S.; Ko, M.; Lee, J.; Choi, I.; Byun, S.Y.; Park, S.; Shum, D.; Kim, S. Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob. Agents Chemother.* **2020**, *64*. [[CrossRef](#)]
9. Leung, N.H.L.; Chu, D.K.W.; Shiu, E.Y.C.; Chan, K.-H.; McDevitt, J.J.; Hau, B.J.P.; Yen, H.-L.; Li, Y.; Ip, D.K.M.; Peiris, J.S.M.; et al. Respiratory virus shedding in exhaled breath and efficacy of face masks. *Nat. Med.* **2020**, *26*, 676–680. [[CrossRef](#)]
10. Chilamakuri, R.; Agarwal, S. COVID-19: Characteristics and Therapeutics. *Cells* **2021**, *10*, 206. [[CrossRef](#)]
11. Mulholland, A.J.; Amaro, R.E. COVID19—Computational Chemists Meet the Moment. *J. Chem. Inf. Model.* **2020**, *60*, 5724–5726. [[CrossRef](#)]
12. Tripathi, N.; Tripathi, N.; Goshisht, M.K. COVID-19: Inflammatory responses, structure-based drug design and potential therapeutics. *Mol. Divers.* **2021**. [[CrossRef](#)]
13. Kirby, T. New variant of SARS-CoV-2 in UK causes surge of COVID-19. *Lancet Respir. Med.* **2021**. [[CrossRef](#)]
14. Gil, C.; Ginex, T.; Maestro, I.; Nozal, V.; Barrado-Gil, L.; Cuesta-Geijo, M.Á.; Urquiza, J.; Ramírez, D.; Alonso, C.; Campillo, N.E.; et al. COVID-19: Drug Targets and Potential Treatments. *J. Med. Chem.* **2020**. [[CrossRef](#)]
15. Mohamed, N.M.; Ali, E.M.H.; AboulMagd, A.M. Ligand-based design, molecular dynamics and ADMET studies of suggested SARS-CoV-2 Mpro inhibitors. *RSC Adv.* **2021**, *11*, 4523–4538. [[CrossRef](#)]
16. Yang, Y.; Zhu, Z.; Wang, X.; Zhang, X.; Mu, K.; Shi, Y.; Peng, C.; Xu, Z.; Zhu, W. Ligand-based approach for predicting drug targets and for virtual screening against COVID-19. *Brief. Bioinform.* **2021**. [[CrossRef](#)]
17. Masand, V.H.; Akasapu, S.; Gandhi, A.; Rastija, V.; Patil, M.K. Structure features of peptide-type SARS-CoV main protease inhibitors: Quantitative structure activity relationship study. *Chemom. Intell. Lab. Syst.* **2020**, *206*. [[CrossRef](#)]
18. Macalino, S.J.Y.; Gosu, V.; Hong, S.; Choi, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharmacol. Res.* **2015**, *38*, 1686–1701. [[CrossRef](#)]
19. Fujita, T.; Winkler, D.A. Understanding the Roles of the “Two QSARs”. *J. Chem. Inf. Model.* **2016**, *56*, 269–274. [[CrossRef](#)]
20. Masand, V.H.; El-Sayed, N.N.E.; Bambole, M.U.; Patil, V.R.; Thakur, S.D. Multiple quantitative structure-activity relationships (QSARs) analysis for orally active trypanocidal N-myristoyltransferase inhibitors. *J. Mol. Struct.* **2019**, *1175*, 481–487. [[CrossRef](#)]
21. Masand, V.H.; Elsayed, N.N.; Thakur, S.D.; Gawhale, N.; Rathore, M.M. Quinoxalinones Based Aldose Reductase Inhibitors: 2D and 3D-QSAR Analysis. *Mol. Inform.* **2019**, *38*, e1800149. [[CrossRef](#)]
22. Masand, V.H.; El-Sayed, N.N.E.; Bambole, M.U.; Quazi, S.A. Multiple QSAR models, pharmacophore pattern and molecular docking analysis for anticancer activity of α , β -unsaturated carbonyl-based compounds, oxime and oxime ether analogues. *J. Mol. Struct.* **2018**, *1157*, 89–96. [[CrossRef](#)]
23. Gramatica, P. Principles of QSAR Modeling. *Int. J. Quant. Struct. Prop. Relatsh.* **2020**, *5*, 61–97. [[CrossRef](#)]
24. Neves, B.J.; Braga, R.C.; Melo-Filho, C.C.; Moreira-Filho, J.T.; Muratov, E.N.; Andrade, C.H. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front. Pharmacol.* **2018**, *9*, 1275. [[CrossRef](#)]
25. Thorne, N.; Auld, D.S.; Inglese, J. Apparent activity in high-throughput screening: Origins of compound-dependent assay interference. *Curr. Opin. Chem. Biol.* **2010**, *14*, 315–324. [[CrossRef](#)] [[PubMed](#)]
26. Teodoro, A.J. Bioactive Compounds of Food: Their Role in the Prevention and Treatment of Diseases. *Oxidative Med. Cell. Longev.* **2019**, *2019*, 1–4. [[CrossRef](#)]
27. Childs, C.E.; Calder, P.C.; Miles, E.A. Diet and Immune Function. *Nutrients* **2019**, *11*, 1933. [[CrossRef](#)]
28. Baig, M.H.; Ahmad, K.; Roy, S.; Ashraf, J.M.; Adil, M.; Siddiqui, M.H.; Khan, S.; Kamal, M.A.; Provaznik, I.; Choi, I. Computer Aided Drug Design: Success and Limitations. *Curr. Pharm. Des.* **2016**, *22*, 572–581. [[CrossRef](#)] [[PubMed](#)]
29. Masand, V.H.; El-Sayed, N.N.E.; Mahajan, D.T.; Rastija, V. QSAR analysis for 6-arylpyrazine-2-carboxamides as Trypanosoma brucei inhibitors. *SAR QSAR Environ. Res.* **2017**, *28*, 165–177. [[CrossRef](#)]
30. Masand, V.H.; El-Sayed, N.N.E.; Mahajan, D.T.; Mercader, A.G.; Alafeefy, A.M.; Shibi, I.G. QSAR modeling for anti-human African trypanosomiasis activity of substituted 2-Phenylimidazopyridines. *J. Mol. Struct.* **2017**, *1130*, 711–718. [[CrossRef](#)]
31. Masand, V.H.; Mahajan, D.T.; Maldhure, A.K.; Rastija, V. Quantitative structure-activity relationships (QSARs) and pharmacophore modeling for human African trypanosomiasis (HAT) activity of pyridyl benzamides and 3-(oxazol[4,5-b]pyridin-2-yl)anilides. *Med. Chem. Res.* **2016**, *25*, 2324–2334. [[CrossRef](#)]
32. Masand, V.H.; Rastija, V.; Patil, M.K.; Gandhi, A.; Chapolikar, A. Extending the identification of structural features responsible for anti-SARS-CoV activity of peptide-type compounds using QSAR modelling. *SAR QSAR Environ. Res.* **2020**, *31*, 643–654. [[CrossRef](#)]

33. Konno, S.; Thanigaimalai, P.; Yamamoto, T.; Nakada, K.; Kakiuchi, R.; Takayama, K.; Yamazaki, Y.; Yakushiji, F.; Akaji, K.; Kiso, Y.; et al. Design and synthesis of new tripeptide-type SARS-CoV 3CL protease inhibitors containing an electrophilic arylketone moiety. *Bioorg. Med. Chem.* **2013**, *21*, 412–424. [[CrossRef](#)]
34. Regnier, T.; Sarma, D.; Hidaka, K.; Bacha, U.; Freire, E.; Hayashi, Y.; Kiso, Y. New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CLpro inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 2722–2727. [[CrossRef](#)] [[PubMed](#)]
35. Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* **2020**. [[CrossRef](#)] [[PubMed](#)]
36. Jo, S.; Kim, S.; Shin, D.H.; Kim, M.-S. Inhibition of SARS-CoV 3CL protease by flavonoids. *J. Enzym. Inhib. Med. Chem.* **2019**, *35*, 145–151. [[CrossRef](#)] [[PubMed](#)]
37. Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; et al. α -Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.* **2020**, *63*, 4562–4578. [[CrossRef](#)] [[PubMed](#)]
38. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628. [[CrossRef](#)] [[PubMed](#)]
39. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [[CrossRef](#)] [[PubMed](#)]
40. Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* **2020**. [[CrossRef](#)] [[PubMed](#)]
41. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [[CrossRef](#)]
42. Masand, V.H.; Rastija, V. PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. *Chemom. Intell. Lab. Syst.* **2017**, *169*, 12–18. [[CrossRef](#)]
43. Gramatica, P.; Cassani, S.; Chirico, N. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. *J. Comput. Chem.* **2014**, *35*, 1036–1044. [[CrossRef](#)]
44. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121–2132. [[CrossRef](#)]
45. Saxena, A.K.; Prathipati, P. Comparison of MLR, PLS and GA-MLR in QSAR analysis*. *SAR QSAR Environ. Res.* **2003**, *14*, 433–445. [[CrossRef](#)]
46. Ramachandran, G.N.; Ramakrishnan, C.; Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **1963**, *7*, 95–99. [[CrossRef](#)]
47. Gaudreault, F.; Morency, L.P.; Najmanovich, R.J. NRGsuite: A PyMOL plugin to perform docking simulations in real time using FlexAID. *Bioinformatics* **2015**, *31*, 3856–3858. [[CrossRef](#)] [[PubMed](#)]
48. Gaudreault, F.; Najmanovich, R.J. FlexAID: Revisiting Docking on Non-Native-Complex Structures. *J. Chem. Inf. Model.* **2015**, *55*, 1323–1336. [[CrossRef](#)] [[PubMed](#)]
49. Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461. [[CrossRef](#)] [[PubMed](#)]
50. Khan, R.J.; Jha, R.K.; Amera, G.M.; Jain, M.; Singh, E.; Pathak, A.; Singh, R.P.; Muthukumaran, J.; Singh, A.K. Targeting SARS-CoV-2: A systematic drug repurposing approach to identify promising inhibitors against 3C-like proteinase and 2'-O-ribose methyltransferase. *J. Biomol. Struct. Dyn.* **2020**, 1–14. [[CrossRef](#)]
51. Dayer, M.R.; Taleb-Gassabi, S.; Dayer, M.S. Lopinavir; A Potent Drug against Coronavirus Infection: Insight from Molecular Docking Study. *Arch. Clin. Infect. Dis.* **2017**, *12*, e13823. [[CrossRef](#)]
52. Kaiser, J. SCIENCE RESOURCES: Chemists Want NIH to Curtail Database. *Science* **2005**, *308*, 774. [[CrossRef](#)] [[PubMed](#)]
53. Chowdhury, P. In silico investigation of phytoconstituents from Indian medicinal herb ‘*Tinospora cordifolia* (giloy)’ against SARS-CoV-2 (COVID-19) by molecular dynamics approach. *J. Biomol. Struct. Dyn.* **2020**, 1–18. [[CrossRef](#)] [[PubMed](#)]