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 1224 (2021) 129026

Contents lists available at ScienceDirect

Journal of Molecular Structure

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

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

Kalyan Ghosh[a,1], Sk. Abdul Amin[b,1], Shovanlal Gayen[a,*], Tarun Jha[b,*]

[a] Laboratory of Drug Design and Discovery, Department of Pharmaceutical Sciences, Dr. Harisingh Gour University, Sagar, Madhya Pradesh, 470003, India
[b] Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, P. O. Box 17020, Kolkata, 700032, India

ARTICLE INFO

Article history:
Received 9 June 2020
Revised 9 July 2020
Accepted 4 August 2020
Available online 5 August 2020

Keywords:
COVID-19
SARS-CoV-2
SARS-CoV Mpro
SPCI analysis
Monte Carlo based optimization
Natural product

ABSTRACT

As the world struggles against current global pandemic of novel coronavirus disease (COVID-19), it is challenging to trigger drug discovery efforts to search broad-spectrum antiviral agents. Thus, there is a need of strong and sustainable global collaborative works especially in terms of new and existing data analysis and sharing which will join the dots of knowledge gap. Our present chemical-informatics based data analysis approach is an attempt of application of previous activity data of SARS-CoV main protease (Mpro) inhibitors to accelerate the search of present SARS-CoV-2 Mpro inhibitors. The study design was composed of three major aspects: (1) classification QSAR based data mining of diverse SARS-CoV Mpro inhibitors, (2) identification of favourable and/or unfavourable molecular features/fingerprints/substructures regulating the Mpro inhibitory properties, (3) data mining based prediction to validate recently reported virtual hits from natural origin against SARS-CoV-2 Mpro enzyme. Our Structural and physico-chemical interpretation (SPCI) analysis suggested that heterocyclic nucleus like diazole, furan and pyridine have clear positive contribution while, thiophen, thiazole and pyrimidine may exhibit negative contribution to the SARS-CoV Mpro inhibition. Several Monte Carlo optimization based QSAR models were developed and the best model was used for screening of some natural product hits from recent publications. The resulted *active* molecules were analysed further from the aspects of fragment analysis. This approach set a stage for fragment exploration and QSAR based screening of *active* molecules against putative SARS-CoV-2 Mpro enzyme. We believe the future *in vitro* and *in vivo* studies would provide more perspectives for anti-SARS-CoV-2 agents.

1. Introduction

Severe acute respiratory syndrome (SARS) coronavirus-2 (SARS-CoV-2) has been spreading alarmingly by causing tremendous social and economic disruption [1–3]. This zoonatic infection has spread over 216 countries and territories [4,5].

SARS-CoV-2 is the 7th human coronavirus (HCoV) and the 3rd HCoV which posted hefty means in the 21st century [2] after SARS-CoV in 2002 and Middle East respiratory syndrome (MERS) coronavirus (MERS-CoV) in 2012. The MERS-CoV infected 1,700 people with a fatality rate of ~36% and the SARS-CoV infected 8422 people with a fatality rate of ~10% [2,6]. SARS-CoV-2 created an unprecedented health emergency around the world and till date 11 591 595 confirmed cases and 537 859 deaths have been documented [5].

Notably, the genome of SARS-CoV-2 comprises ~30,000 nucleotides with 10 Open Reading Frames (ORFs). The 3′ terminal regions encode viral structural proteins including spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins. On the other hand, the 5′ terminal ORF1ab encodes two viral replicase polyproteins pp1a and pp1b [1]. A number of 16 non-structural (ns) proteins (nsp1 to nsp16) are raised upon proteolytic cleavage of pp1a and pp1b. The nsp5 (Chymotrypsin-like protease 3CLpro also called Main protease Mpro) is a prerequisite enzyme of the viral replication and maturation. Mpro turns into a charismatic target for anti-SARS-CoV-2 drug discovery and development [7–12].

The research work in terms of molecular docking and target based virtual screening studies on Mpro have moved at a much

* Corresponding authors.
E-mail addresses: pharmacist.amin@gmail.com, skabdulamin.rs@jadavpuruniversity.in (Sk.A. Amin), shovanlal.gayen@gmail.com (S. Gayen), tjupharm@yahoo.com (T. Jha).
[1] Authors have equal contribution.

https://doi.org/10.1016/j.molstruc.2020.129026
0022-2860/

faster pace [7–28] after releasing of the several covalent and non-covalent inhibitor bound crystal structures. Despite the availability of inhibitor-bound SARS-CoV-2 Mpro crystal structures and lots of proteomic knowledge, the significant fragment/feature which modulates the structure-activity relationships (SARs) pattern is still not known. Therefore, ligand-based molecular modelling approaches are necessary to gather knowledge about the favourable and/or unfavourable molecular features/fingerprints/substructures regulating the Mpro inhibitory properties.

Quantitative structure-activity relationship (QSAR) study is a very significant ligand-based molecular modelling technique that easily recognised the effect of structural and physico-chemical features of ligands on the biological activity [29,30]. Not only that, it offers prediction of particular compounds to their biological activities of interest.

As the SARS-CoV-2 genome has over 80% identity to SARS-CoV (about 96% sequence similarity for their Mpro), previously reported SARS-CoV Mpro inhibitors may have huge prospect to show their efficacy against SARS-CoV-2 also. In this connection, we design our current research, a part of our rational molecular modelling studies [3,31–34], by covering three major characteristics- (1) classification QSAR based data mining of diverse SARS-CoV Mpro inhibitors, (2) identification of favourable and/or unfavourable molecular features/fingerprints/substructures regulating the Mpro inhibitory properties, (3) QSAR based prediction to validate recently reported virtual hits from natural origins.

This study encompasses the effect of structural and physico-chemical features required for Mpro inhibition. The study provides useful information to the medicinal chemists for design effective Mpro/3CLpro inhibitors in future. It set the stage for molecule identification and QSAR based screening of active molecules against putative SARS-CoV-2 Mpro enzyme which surely claim a momentous attention to the scientific audiences.

2. Methods and materials

2.1. Dataset collection

The fighting against COVID-19 disease requires strong and sustainable global collaborative works especially in terms of data sharing which will join the dots of knowledge gap [35]. A set of 113 compounds were retrieved from the data as shared by Bobrowski and co-workers [35,36]. We considered only compounds having half-maximal inhibitory concentration (IC_{50}) and eliminated the compounds with binding affinity (K_i) value. Thus, 88 compounds were kept for this current modelling study (**Table S1**).

2.2. Classification based QSAR

The classification modelling assists to discriminate the *Active* and *Inactive* molecules in terms of their investigated biological significance. The 'activity threshold' for the current work was set to the IC_{50} of 10,000 nM. Here, we performed Structural and physico-chemical interpretation (SPCI) analysis [37,38] and Monte Carlo based Coral QSAR studies [39–42]. Performing Monte Carlo based Coral QSAR study not only offer a graphical visualization of critical fingerprint or fragments attributed to enhance/decrease the SARS-CoV Mpro inhibitory activity but also it allows the chance of screening external set compounds.

2.3. Structural and physico-chemical interpretation (SPCI) analysis

The SPCI analysis was explored to identify and approximate the contributions of different fragments that are important for Mpro inhibition. Initially, the descriptor calculation was performed with the help of SiRMS tool. Further, these descriptors were used for model development and validation in our study [37,38].

In SPCI analysis, four diverse classification-based QSAR models were generated by using machine learning approaches like: Gradient boosting classification (GBC), Random Forest (RF), Support Vector Machine (SVM) and *k*-nearest neighbour (*k*NN). These models were further evaluated by different statistical parameters like: balanced accuracy, sensitivity, and specificity [38]. Additionally, all the fragments comprising of at most three attachment points were preferred and subsequently, favoured fragments were counted by RDKit in amalgamation with SMARTS pattern [38]. Finally, the overall contribution of the different fragments obtained from four machine learning models are shown in median fragment contribution graphs generated by using rspciR software package [43].

2.4. Monte Carlo optimization based QSAR

Monte Carlo optimization method was used to identify the important structural fingerprints that are solely responsible for endorsing or deterring of activity [3]. Different descriptors that are generally employed in the study are: SMILES-based descriptors, Graph-based descriptors and Hybrid descriptors.

The SMILES-based descriptors are calculated by the following equation:

$$^{SMILES}\,DCW\,(T,N) = a\,CW\,(ATOMPAIR) + b\,CW\,(NOSP) + c\,CW\,(BOND) + d\,CW\,(HALO) + \alpha\Sigma CW\,(S_k) + \beta\Sigma CW\,(SS_k) + \gamma\,\Sigma CW\,(SSS_k)$$

Where, T and N represent threshold value and number of epoch, respectively. The correlation weights are represented by CW. The different coefficients like a, b, c, d, α, β and γ are used for descriptor modification. NOSP, HALO, BOND and ATOMPAIR represent global SMILES attributes and the local smile attributes are denoted by S_k, SS_k and SSS_k [40–42].

Further, different Graph-based descriptors like: GAO (graph of atomic orbital), HSG (hydrogen-suppressed graph) and HFG (hydrogen-filled graph) are calculated by following equation:

$$^{Graph}DCW\,(T,\ N) = \alpha\Sigma CW\,(A_k) + \beta\Sigma CW\,(^0EC_k) + \Gamma\Sigma CW(^1EC_k) + \delta\Sigma CW(^2EC_k) + \epsilon\Sigma CW(^3EC_k)$$

Where, ^0EC_k, ^1EC_k and ^3EC_k represent different Morgan's connectivity indices. A_k denotes different chemical atoms like: C, N, O etc. α, β and γ are the coefficients with 0 and 1 value. The coefficients having value 0 and 1 are denoted by α, β and γ [39,42].

The mixture of SMILES and graph-based descriptors forms hybrid descriptors which are represented as:

$$^{Hybrid}DCW(T,N) = {}^{SMILES}\,DCW(T,N) + {}^{Graph}\,DCW(T,N)$$

Finally, the model development and validation step was performed by using balance of correlation method, where twenty-one classification models were developed from three different splits. The dataset containing 88 Mpro inhibitors were distributed into training (52 compounds), calibration (18 compounds) and test (18 compounds) sets which were considered for the study. Additionally, optimization of T (threshold) and N (epoch) were also accomplished separately for individual model [3,39]. The sensitivity, specificity, accuracy along with the MCC values was calculated as a measure of internal and external validation [3]. Lastly, the important structural attributes that are exclusively liable for promoting or hindering of Mpro activity were identified.

3. Result and discussions

Compounds having the SARS-CoV Mpro IC_{50} value less than the 'activity threshold' were yielded to lower Mpro inhibitors or *inac-*

tives (0) and those with Mpro IC_{50} value higher than the 'activity threshold' (IC_{50} = 10,000 nM) were classified as promising Mpro inhibitors or *actives* (1). Thus, 27 molecules were identified as *actives* (1) while, 61 compounds were distinguished as lower SARS-CoV Mpro inhibitors (0) in the classification analysis (**Table S1**).

At first structural and physico-chemical interpretation (SPCI) analysis was performed [38]. These machine-learning based models were utilised for a fragment/feature analysis to estimate the contributions of different fragments towards Mpro inhibition. It enables an extensive interpretation of the structural and physico-chemical properties responsible for SARS-CoV Mpro inhibitory activities. The Monte Carlo optimization-based QSAR modelling was also employed by the aid of SMILES and graph-based descriptors to justify fragment contributions [39]. The best Monte Carlo optimization-based QSAR model was used for screening the recently reported docking based natural product hits. Together with the fragment/fingerprint analysis results justified the selection of the potential hits retrieved through such QSAR derived prediction.

3.1. Structural and physico-chemical interpretation (SPCI) analysis

With the aim to construct an interpretable QSAR model, gradient boosting machine (GBM), random forest (RF), support vector machine (SVM) and *k*-nearest neighbor (*k*NN) were established using structural and physico-chemical interpretation (SPCI) analysis (Table 1). The parameter settings used for the individual models (GBM, RF, SVM and *k*NN) development are given in **Table S2**.

Table 1 demonstrates that *k*NN model produced significant five-fold cross-validation performance as far as the values of balanced accuracy and specificity values were concerned. However, the sensitivity value is comparatively poor. Moreover, RF model shared similar statistical parameters.

A consensus model was also developed to eliminate biasness of individual models. The fragments obtained from different models are depicted in Fig. 1. These fragments were found to have different positive and negative contributions towards 3CLpro inhibitory activity.

Table 1
Five-fold cross-validation performance for classification model built in this study.

Model	Balanced accuracy	Sensitivity	Specificity
GBM	0.62	0.41	0.84
RF	0.64	0.41	0.87
SVM	0.59	0.19	1.00
*k*NN	0.67	0.41	0.93

As per the consensus model, heterocyclic like diazole, furan and pyridine have clear positive contribution to the SARS-CoV Mpro inhibition. This can be explained by comparing the order of activities in molecules **001**-**003** where compounds possessed both furan and pyridine rings in their structure (Fig. 2) and showed higher Mpro inhibitory activity (IC_{50} in between 50 and 63 nM). The importance of five member heterocyclic furan ring was also found in compounds **016, 025, 027**and these compounds displayed promising inhibitory activities. In spite of having furan ring in their structures, compounds **036** and **039** (Fig. 2) were found in the category of less active molecules or *inactives* due to the presence of higher negatively contributing attributes in their structures.

Meanwhile, the positive contribution of diazoles was justified by observing the SARS-CoV Mpro *active* compounds **004, 007** and **027**. Notably, compound **027** (Fig. 2) possessed two positively contributed features including furan, diazole rings and one negatively contributed thiophene ring, therefore, it just crossed the 'activity threshold' to be *active* (IC_{50} = 10,000 nM). Hence, the contributions of all the groups or moieties present in a structure collectively decide the inhibition potential.

As can be seen from Fig. 1, the aliphatic OH substituent would lead to increase of Mpro IC_{50} values in the compounds **004, 007**, etc. On the other hand, thiophen, thiazole and pyrimidine may exhibit negative contribution. Furthermore, the fragments such as

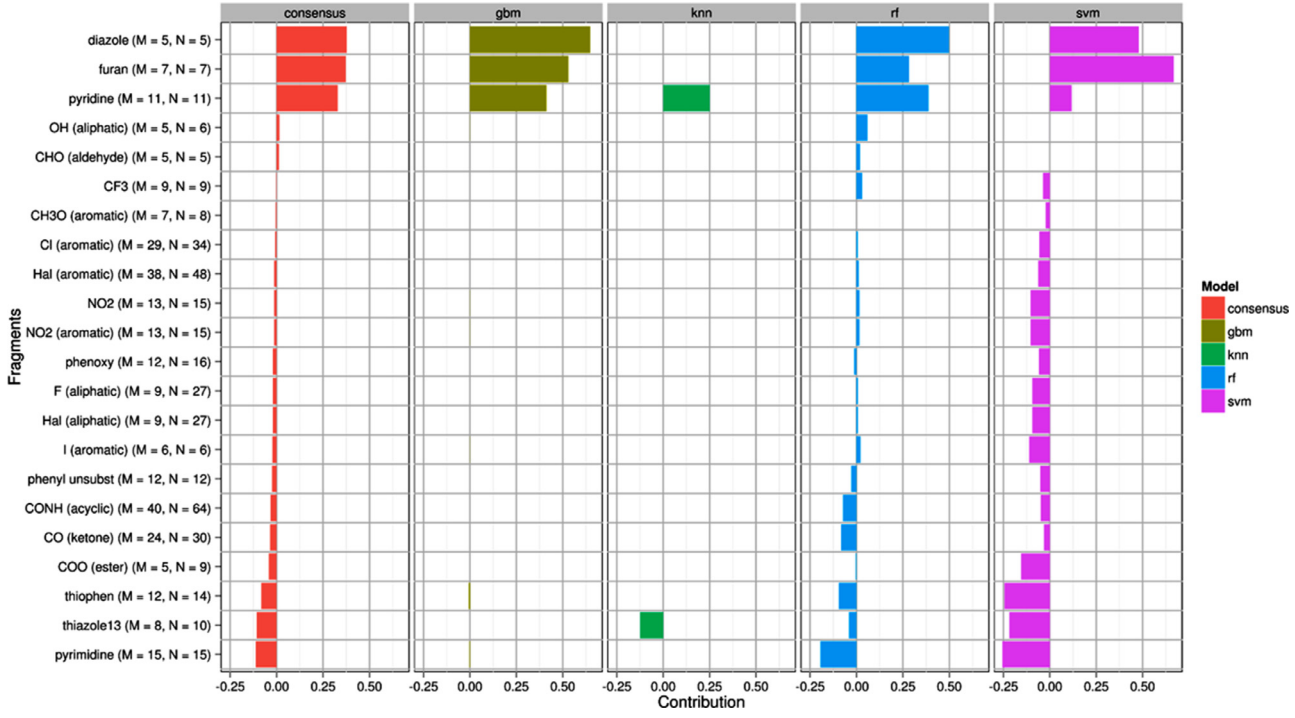

Fig. 1. Contribution plot of different fragments (present in at least 5 compounds) identified by using consensus (red), GBM (dark green), kNN (light green), RF (cyan), and SVM (purple) models. The numbers M and N signify the number of compounds containing a fragment and the number of fragments present in the dataset, respectively. (For interepretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

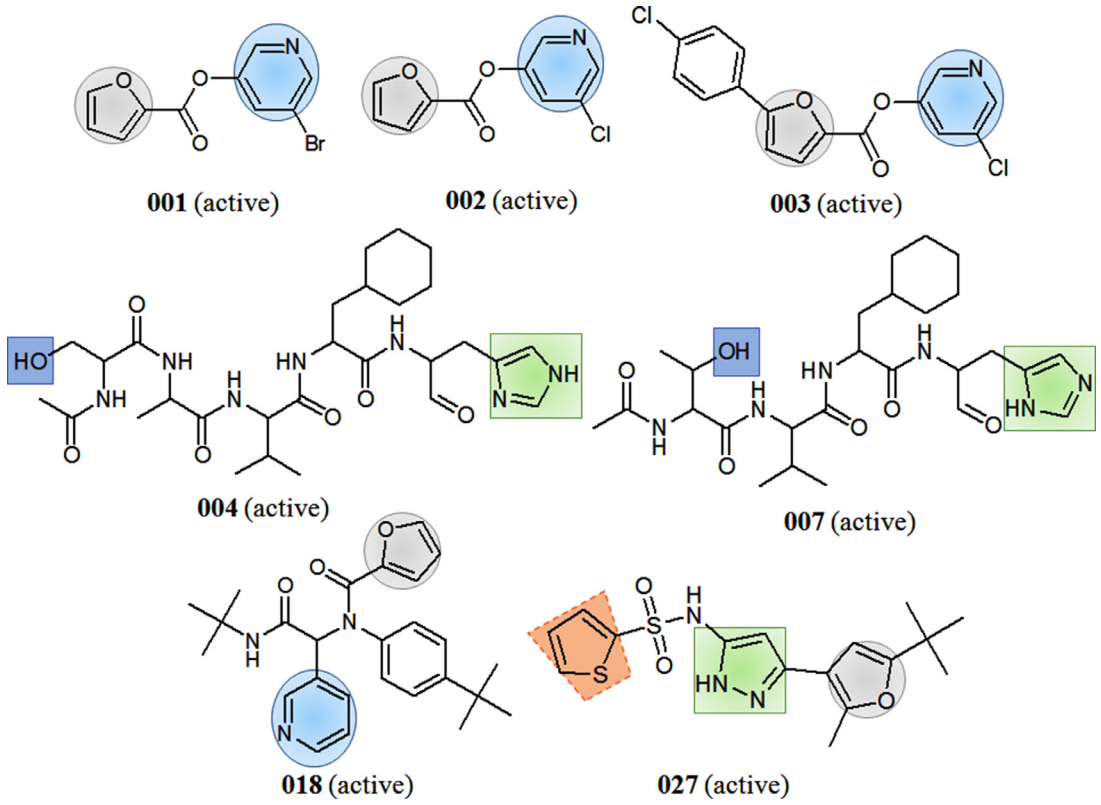

Fig. 2. Examples of prototype higher active Mpro inhibitors with good and bad molecular fingerprints obtained from SPCI models.

phenoxy, I-aromatic, F-aliphatic, halo-aliphatic, acyclic CONH may negatively contributed to the SARS-CoV Mpro inhibition. For better interpretation, examples of prototype lower active Mpro inhibitors with good and bad molecular fingerprints obtained from SPCI models is depicted in Fig. 3.

This effect can also be noticed for molecules **061** to **088** where their lower activities can be related to the presence of pyrimidine ring (Fig. 3). Thus, it may be supposed that six member pyridine ring would be effective for the interaction and could be used as replacement of pyrimidine to improve the biological potency. Similarly, five member heterocyclic rings such as thiophene and thiazole should be replaced with furan ring to trigger the Mpro inhibitory activity. Surprisingly, compound **044** (Fig. 3) which possessed two diazole and one thiophene moieties was found as *inactive*. The probable reason behind this phenomenon will be discussed with the structural attributes predicted by Monte Carlo optimization based classification QSAR.

3.2. Monte Carlo optimization based classification QSAR

The 88 compounds with diverse structural features were also used in Monte Carlo optimization based classification QSAR analysis [3,42]. Twenty-one different models from three different splits were generated using SMILES and graph-based descriptors with a combination of different connectivity indices were generated for construction of different Monte Carlo optimization based QSAR study [39]. Overall statistical characteristics of twenty-one different models are given in Table 2.

From the Table 2, it may be observed that the model **M21** showed satisfactory statistical property. The sensitivity, specificity and accuracy as well as MCC values obtained for different sets were statistically significant. The values attained for the test set such as sensitivity, specificity, accuracy and MCC of 100%, 92.31%, 94.44% and 87.71%, respectively) justified the acceptable predictability of classification based QSAR model. Therefore, the model **M21** (SMILES and HSG with $^{1}EC_{k}$) from split-3 was found to be the best among 21 developed models. The model was also successfully passed Y-randomization test (*MCCr*^{2} *i.e.*, average randomized $^{C}R^{2}_{p}$ > 0.5). The end point value calculates for **M21** is follows:

$$\text{SARS-CoV Mpro} = 0.017(\pm 0.006) + 0.009(\pm 0.00009) * DCW(6,3)$$

Moreover, different structural attributes of the model **M21** (SMILES and HSG with $^{1}EC_{k}$) from split-3 is given in **Table S3**. Positive structural attribute like **n...c...c...** (representing the presence of nitrogen atom attached to two sp^2carbon atoms) was found to have influence on the activity of the Mpro inhibitors. This feature was found in compound **005**, **009**, **018** (Fig. 4) possessing good SARS-CoV Mpro inhibitory activities. As can be observed from Fig. 4, the presence of sulphur atom attached to doubly bonded oxygen atom represented by the structural attribute **++++O—B2==**,would lead to moderate increase in Mpro pIC_{50} values in the compounds **010**, **012** and **022** etc.

On the other hand, the presence of **N...C...(...** (nitrogen attached to an sp^3 carbon atom having branching), also found to play a crucial role in promoting the inhibitory activity of compound **018**. Similarly, **NOSP11000** (existence of nitrogen and oxygen atom) in compound **020** might induce inhibitory activity. Not surprisingly, the compounds containing two or more positive structural attributes were found to have better SARS-CoV Mpro pIC_{50} values and considered to be higher actives as can be observed from Fig. 4. The analysis of structural attributes from Monte Carlo Optimization results also highlighted similar fragments obtained from our current SPCI analysis. Similar fragments between the two analyses are highlighted in Table 3.

The lower activities of molecules **061-080** (Fig. 4) may be related to the presence of pyrimidine ring in their structure represented by the combined structural attributes **n...1...c...** and **n...c...(.....**Thus, it may be anticipated that the replacement of

Table 2
The statistical characteristics of twenty-one different classification models obtained from Monte Carlo optimization method.

Parameter	Set	TP	TN	FP	FN	N_{Total}	Sensitivity	Specificity	Accuracy	MCC
Split-1										
M1 SMILES	Sub-Training	15	33	1	3	52	0.8333	0.9706	0.9231	0.8287
	Calibration	5	13	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	12	2	1	18	0.7500	0.8571	0.8333	0.5635
M2 SMILES, GAO (^0EC_k)	Sub-Training	13	31	3	5	52	0.7222	0.9118	0.8462	0.6535
	Calibration	5	12	1	0	18	1.0000	0.9231	0.9444	0.8771
	Test	3	12	2	1	18	0.7500	0.8571	0.8333	0.5635
M3 SMILES, GAO (^1EC_k)	Sub-Training	15	33	1	3	52	0.8333	0.9706	0.9231	0.8287
	Calibration	5	13	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	10	4	1	18	0.7500	0.7143	0.7222	0.3959
M4 SMILES, HFG (^0EC_k)	Sub-Training	16	34	0	2	52	0.8889	1.0000	0.9615	0.9162
	Calibration	5	13	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	8	6	0	18	1.0000	0.5714	0.6667	0.4781
M5 SMILES, HFG (^1EC_k)	Sub-Training	16	34	0	2	52	0.8889	1.0000	0.9615	0.9162
	Calibration	5	13	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	9	5	1	18	0.7500	0.6429	0.6667	0.3287
M6 SMILES, HSG (^0EC_k)	Sub-Training	17	34	0	1	52	0.9444	1.0000	0.9808	0.9578
	Calibration	5	12	1	0	18	1.0000	0.9231	0.9444	0.8771
	Test	4	11	3	0	18	1.0000	0.7857	0.8333	0.6701
M7 SMILES, HSG (^1EC_k)	Sub-Training	16	33	1	2	52	0.8889	0.9706	0.9423	0.8717
	Calibration	5	13	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	11	3	1	18	0.7500	0.7857	0.7778	0.4725
Split-2										
M8 SMILES	Sub-Training	15	34	2	1	52	0.9375	0.9444	0.9423	0.8677
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	9	2	3	18	0.5714	0.8182	0.7222	0.4029
M9 SMILES, GAO (^0EC_k)	Sub-Training	12	34	2	4	52	0.7500	0.9444	0.8846	0.7226
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	10	1	3	18	0.5714	0.9091	0.7778	0.5230
M10 SMILES, GAO (^1EC_k)	Sub-Training	16	36	0	0	52	1.0000	1.0000	1.0000	1.0000
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	6	5	3	18	0.5714	0.5455	0.5556	0.1140
M11 SMILES, HFG (^0EC_k)	Sub-Training	14	35	1	2	52	0.8750	0.9722	0.9423	0.8631
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	11	0	3	18	0.5714	1.0000	0.8333	0.6701
M12 SMILES, HFG (^1EC_k)	Sub-Training	11	33	3	5	52	0.6875	0.9167	0.8462	0.6287
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	5	11	0	2	18	0.7143	1.0000	0.8889	0.7774
M13 SMILES, HSG (^0EC_k)	Sub-Training	12	36	0	4	52	0.7500	1.0000	0.9231	0.8216
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	10	1	4	18	0.4286	0.9091	0.7222	0.3959
M14 SMILES, HSG (^1EC_k)	Sub-Training	14	33	3	2	52	0.8750	0.9167	0.9038	0.7789
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	5	10	1	2	18	0.7143	0.9091	0.8333	0.6447
Split-3										
M15 SMILES	Sub-Training	15	33	1	3	52	0.8333	0.9706	0.6231	0.8287
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	13	0	1	18	0.8000	1.0000	0.9444	0.8619
M16 SMILES, GAO (^0EC_k)	Sub-Training	18	34	0	0	52	1.0000	1.0000	1.0000	1.0000
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	10	3	1	18	0.8000	0.7692	0.7778	0.5230
M17 SMILES, GAO (^1EC_k)	Sub-Training	17	34	0	1	52	0.9444	1.0000	0.9808	0.9578
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	12	1	1	18	0.8000	0.9231	0.8889	0.7231
M18 SMILES, HFG (^0EC_k)	Sub-Training	16	33	1	2	52	0.8889	0.9706	0.9423	0.8717
	Calibration	2	14	0	2	18	0.5000	1.0000	0.8889	0.6614
	Test	5	12	1	0	18	1.0000	0.9231	0.9444	0.8771
M19 SMILES, HFG (^1EC_k)	Sub-Training	16	33	1	2	52	0.8889	0.9706	0.9423	0.8717
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	4	12	1	1	18	0.8000	0.9231	0.8889	0.7231
M20 SMILES, HSG (^0EC_k)	Sub-Training	16	33	1	2	52	0.8889	0.9706	0.9423	0.8717
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	3	11	2	2	18	0.6000	0.8462	0.7778	0.4462
M21 SMILES, HSG (^1EC_k)	Sub-Training	16	32	2	2	52	0.8889	0.9412	0.9231	0.8301
	Calibration	4	14	0	0	18	1.0000	1.0000	1.0000	1.0000
	Test	5	12	1	0	18	1.0000	0.9231	0.9444	0.8771

The selected model is shown in bold face; True positive (TP); False negative (FN); False positive (FP); True negative (TN); Total number of compounds (N_{Total}).

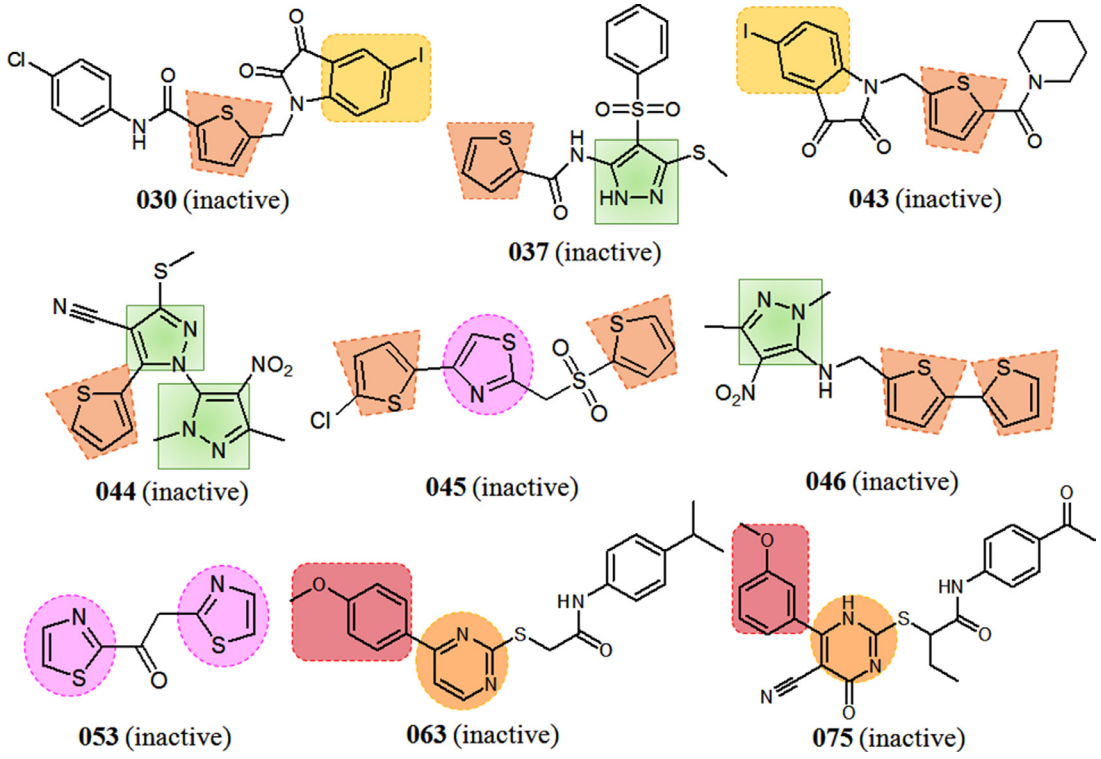

Fig. 3. Examples of prototype lower active Mpro inhibitors with good and bad molecular fingerprints obtained from SPCI models.

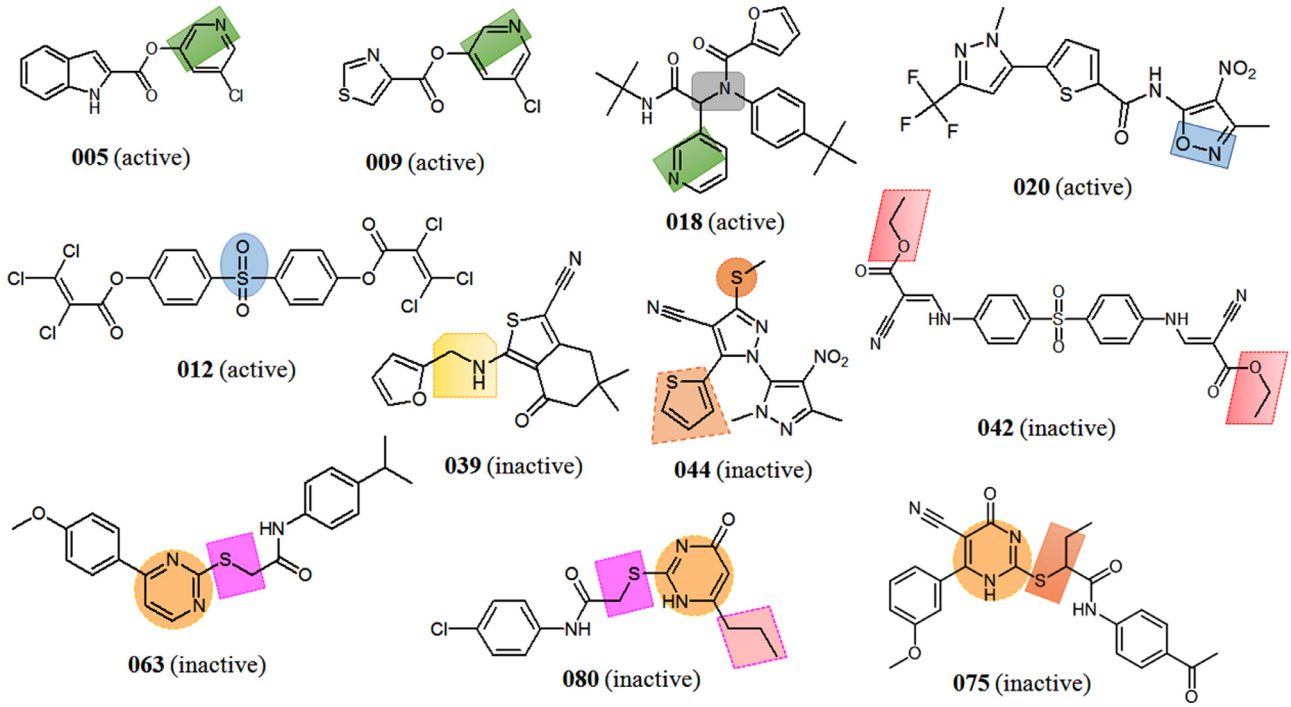

Fig. 4. Examples of some prototype SARS-CoV Mpro inhibitors with structural attributes obtained from best model M21.

pyrimidine led to improve in the SARS-CoV Mpro inhibitory activities. Further, the structural attribute **S...C....** signified the presence of sulphur atom bonded to a sp^3 carbon atom in compound **063** and **080** was also answer able for hindrance of activity. Additionally, the structural attribute **S...C...(...** found in compound **075** was also responsible for lowering the activity as shown in Fig. 4. In compounds, the structural attribute **S.....** suggested the presence of sulphur atom outside a ring is also responsible for lowering the activity of Mpro inhibitors. This may be one of the reason for lower inhibition potential of compound **044** (Fig. 4). For instance, compound **044** (Fig. 3) was found as lower active Mpro inhibitors, although having two diazole and one thiophene moieties. In this case a thiophene moiety and the presence of NO_2 (aromatic) negatively contributed towards biological activity.

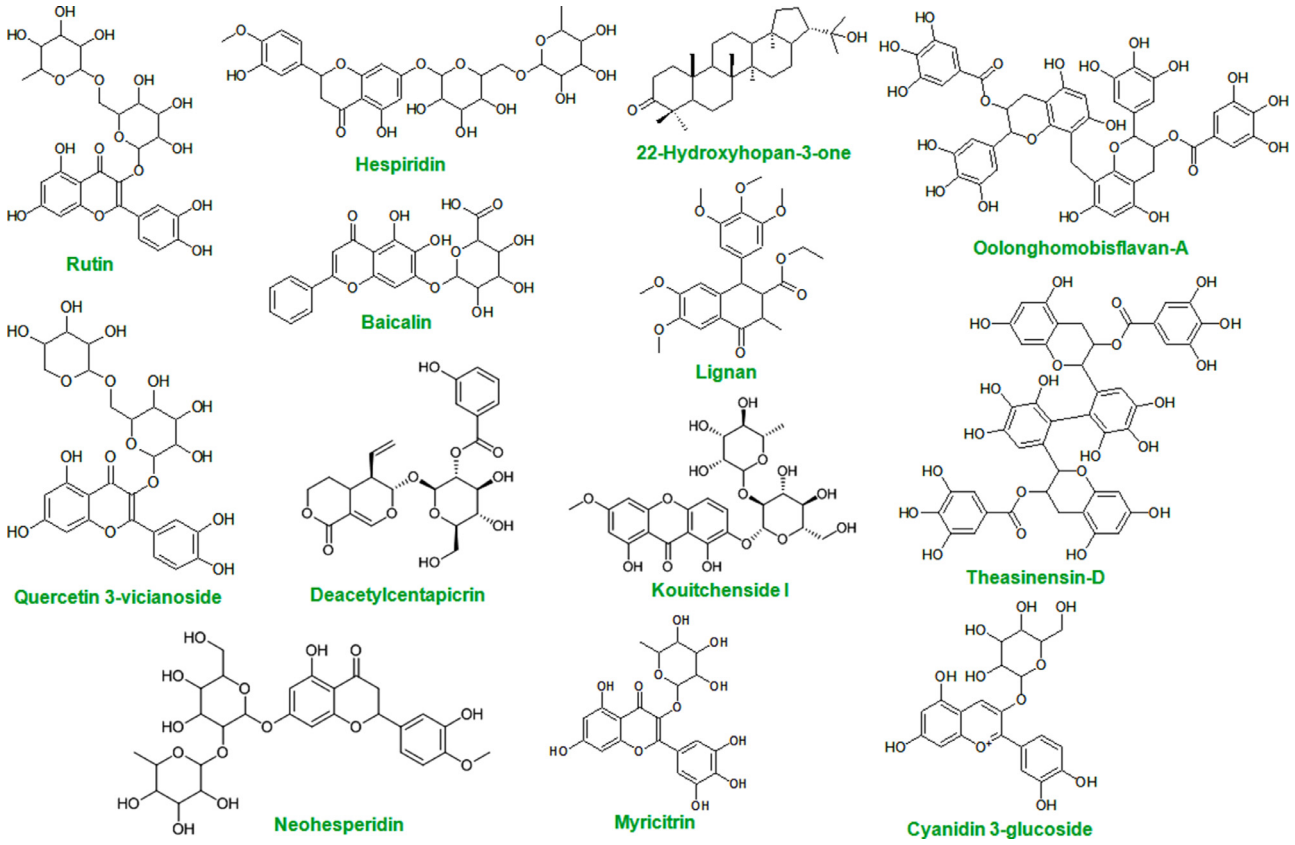

Fig. 5. The chemical structure of the most potent virtual hits for corona virus Mpro inhibition.

Table 3
The fragments/fingerprints modulating the SARS-CoV Mpro inhibitory activities.

Entry	SPCI analysis	Monte Carlo Optimization based QSAR	Contribution
1	Pyridine	n.....	Positive
2	Furan	o.....	Positive
3	CHO (aldehyde)	c...(...O...	Positive
4	CH_3O (aromatic)	O...(...C...	Negative
5	NO_2 (aromatic)	++++N−B3==	Negative
6	F (aliphatic)	F.....	Negative
7	Phenyl unsubst	1.....	Negative
8	CO (ketone)	O...(...C...	Negative
9	Thiophene	s...1.....	Negative
10	Cl (aromatic)	c...(...Cl..	Negative

Moreover, the compound **044** also possessed $-SCH_3$ which is also negative contributor as suggested by negative structural attribute **S...C.....** towards the inhibition (Fig. 4).

Meanwhile, structural attributes **O...(...C...**(oxygen atom having branching attached to sp^3 carbon) and **C...C.....** (two consecutive sp^3 carbon actom) exhibited negative contribution in compounds **042** and **080**, respectively (Fig. 4). Furthermore, the attribute **N...(...C...** (nitrogen atom having branching attached to sp^3 carbon) shown in many compounds including **035**, **039**, **058** etc. also negatively contributed to the SARS-CoV Mpro inhibition (Fig. 4). Notably, some good structural attributes such as **s...(...c...)**, **++++O−B2==** etc. are also found in some lower active Mpro inhibitors (**039**, **042**, **087** and **067**, Fig. 4), but their strong negatively contributing groups further reduces their SARS-CoV Mpro inhibitory activities.

3.3. QSAR derived prediction

Since the SARS-CoV-2 Mpro shares about 96% sequence similarity with SARS-CoV Mpro (while genome has over 80% identity), previously reported SARS-CoV Mpro inhibitors may have huge prospect to show their efficacy against SARS-CoV-2 Mpro also. Thus, considering high statistical significance of the best Monte Carlo optimization based QSAR model, we applied the model **M21** (SMILES and HSG with ^1EC_k) from split-3 to perform QSAR derived prediction of a library of nature product hits from recent publications [7–9,13,16–20,23,24,26–28]. The lists of nature product hits are depicted in **Table S4**.

After screening with the model **M21**, a number of 13 molecules from natural origin were predicted as *actives* (Table 4). These 13 compounds including Rutin, Hespiridine, 22-Hydroxyhopan-3-one, Oolonghomobisflavan-A, Theasinensin-D, Quercetin 3-vicianoside, Deacetylcentapicrin, Kouitchenside I, Neohesperidin, Lignan, Myricitrin, Baicalin, Cyanidin 3-glucoside were considered as the most potent virtual hits for coronavirus Mpro inhibition Fig. 5.

From the Fig. 5, it may be conferred that all these *actives* were structurally similar to each other. Most of these compounds were polyphenols. Maximum of these molecules contain ring fragments while the oxygen atom were common in all these structures. The different structural fragments such as o..... (presence of oxygen in a ring), (..... (presence of branching), ++++O−B2== (presence of double bonded oxygen), O...c...1...(presence of OH/OCH_3 attached with ring) etc. are responsible for their predicted Mpro inhibitory activity. The molecular docking interactions analysis also reported that these hits found to potentially bind with active site amino acid residues of SARS-CoV-2 Mpro [8,13,16,18,24,27,28]. The molecular docking study performed by Das and co-workers sug-

Table 4
The name and common sources of the most potent virtual hits for corona virus Mpro inhibition.

Cpd[a]	Name	Common sources	Pharmacological actions	Reference
N1	Rutin	Passion flower, buckwheat, tea, and apple	Antioxidant, cytoprotective,anticarcinogenic, antiviral, vasoprotective, neuroprotective and cardioprotective activities	[44]
N2	Hespiridine	Citrus fruits including such as oranges, lemons, grapefruits, pomelos and limes	Antimicrobial, cardiovascular function, type II diabetes, anti-inflammation, wound healing, UV protection, antiskin cancer and skin lightening.	[45]
N9	22-Hydroxyhopan-3-one	*Cassia siamea* (Fabaceae)	–	[18]
N11	Oolonghomobisflavan-A	It is a constituent of oolong tea.	Antioxidant, antiobesity activity	[46]
N13	Theasinensin-D	Black tea and oolong tea	Antioxidant and antimicrobial effect	[47]
N14	Quercetin 3-vicianoside	–	–	[9]
N28	Deacetylcentapicrin	*Swertia macrosperma*	Anti-virus, antiinflammatory action	[27]
N30	Kouitchenside I	*Swertia kouitchensis*	α-Glucosidase inhibitory activity; anti-virus, antiinflammatory action	[27,48]
N31	Neohesperidin	Citrus fruits including such as oranges, lemons, grapefruits, pomelos and limes	Food supplement	[49]
N48	Lignan	Kiwi fruit, asparagus, pineapple, grapes, lemons, oranges and even in tea and coffee	Antioxidant, antiinflammatory effect	[50]
N55	Myricitrin	Bayberry	Nitric oxide (NO) and protein kinase C (PKC) inhibitor. Also shows antipsychotic-like and anxiolytic-like actions	[51]
N57	Baicalin	*Scutellaria baicalensis*	Anxiolytic effects without sedative or myorelaxant effects, induces cancer cell apoptosis	[52]
N58	Cyanidin 3-glucoside	Blackberry	Chemopreventive and chemotherapeutic activity	[53]

[a] Compound number.

gested that rutin (also known as vitamin P) forms non-covalent interactions with the SARS-CoV-2 Mpro active site residues [13]. It interacts with H41, L141, N142, E166, T190 and Q192 by forming hydrogen bonding. Moreover, a π-sulphur and π-alkyl interactions were noticed with C145 and P168, respectively. Hesperidin forms amide-π stacked interaction with T45, π-alkyl interactions with M49 and C145 as well as hydrogen bonding interactions with T24, T25, T45, S46 and C145 [13]. Both these two dietary polyphenols (rutin and hesperidin) having low systemic toxicity indicate promising potential for the treatment of COVID-19.

Apart from hesperidin, another active constituent, neohesperidin, from *Citrus aurantium* was also found to be *active* as per our QSAR based prediction. Moreover, kouitchenside I and deacetylcentapicrin from the plants of Swertia genus were also predicted as *actives*.

A pentacyclic triterpene, 22-hydroxyhopan-3-one from *Cassia siamea* (Fabaceae) showed AutoDock Vina 4.2 promising binding affinity (-8.6 kcal/mol) against Mpro of SARS-CoV-2 (PDB: 6LU7) [18]. 22-Hydroxyhopan-3-one forms conventional hydrogen bond with K137 along with alkyl and π-alkyl interactions with L275, L287, L286 and Y239, respectively [18].

Oolonghomobisflavan-A is an important polymerized polyphenol present in Tea. The semi-flexible docking tool CDOCKER utility of Discovery Studio suggested that Oolonghomobisflavan-A possesses two π-alkyl (M165, H41), and one π-π T-shaped interaction (H41) as well as forms several hydrogen bonds with T25, N142, H163, E166, R188, and H164 [16]. It showed the binding free energy of -256.875 kJ/mol better than the drug Lopinavir (binding free energy of -250.585kJ/mol) as per MM-PBSA calculations [16].

Quercetin 3-vicianoside interact with catalytic amino acid residues (PDB: 6LU7) by forming hydrogen bonds with L141, G143, S144, H163, E166 and hydrophobic bonds with T25, His41, F140, N142, C145, H164, M165, D187, R188, Q189 [9].

Myricitrin (Plant source: *Myrica cerifera*) showed a docking score of -15.64 and binding affinity of -22.13 kcal/mol [24]. It forms several hydrogen bonding and other interactions with amino acid residues T24, T25, T26, L27, H41, C44, S46, M49, L141, N142, G143, S144, C145, H163, E166 and Q189 [24].

Baicalin (Plant source: *S. baicalensis*) is an experimentally manifested antiviral representative against SARS-CoV [54], SARS-CoV-2 [22]. Notably, baicalin exhibited an IC_{50} of 6.41 μM against SARS-CoV-2 Mpro along with K_d of 11.50 μM [21]. In addition, the docking study of baicalin performed by Islam et al. [8] showed interaction through one hydrophobic, one π-sulfur and six hydrogen bonding interactions with the catalytic residues of SARS-CoV-2 Mpro (AutoDock Vina score of -8.1 kcal/mol and GOLD score of 59.19) [8]. Cyanidin 3-glucoside exhibits numerous hydrogen bonding interactions and hydrophobic interactions (AutoDock Vina score of -8.4 kcal/mol) in which one hydrophobic interaction is noticed with the catalytic C145 [8].

These hits could be tested for their *in-vitro* and *in-vivo* inhibition potential against SARS-CoV-2 Mpro. Further, the backbone structure of these molecules could be exploited to develop more potent Mpro inhibitors in future.

4. Conclusion

Quantitative structure-activity relationship (QSAR) study is an efficient technique that extracts crucial information from complex datasets. Recently, QSAR modelling truly recognised the effect of structural and physicochemical features of compounds on the investigated biological activity and also offers simultaneous prediction of virtual libraries.

In this current study, we developed multiple classification QSAR models with a diverse dataset of compounds possessing SARS-CoV Mpro inhibitory properties. On one hand, the Structural and physico-chemical interpretation (SPCI) analysis was accomplished to perform fragment/fingerprint analysis, where the contributions of different molecular features modulating Mpro inhibition were estimated. On the other hand, Monte Carlo optimization based QSAR was constructed and the best model was further used for screening of a library of nature product hits from recent publications. Lastly, the resulted *active* molecules analysed from the aspects of fragment analysis and highlighted their natural sources. This study surely motivate medicinal chemists to rejuvenate potential chemicals by joining fragments/features together or attaching with other scaffolds in hopes to trigger antiviral potency against Mpro of SARS-CoV-2 and other coronavirus as well as efficacy without accruing much toxicities.

Declaration of Competing Interest

The authors have no conflict of interests.

CRediT authorship contribution statement

Kalyan Ghosh: Data curation, Methodology, Software, Investigation. **Sk. Abdul Amin:** Conceptualization, Visualization, Investigation, Writing - original draft. **Shovanlal Gayen:** Conceptualization, Writing - review & editing. **Tarun Jha:** Writing - review & editing, Supervision.

Acknowledgment

Financial assistance from the Council of Scientific and Industrial Research (CSIR), New Delhi, India in the form of a Senior Research Fellowship (SRF) [FILE NO.: 09/096(0967)/2019-EMR-I, Dated: 01-04-2019] to Sk. Abdul Amin is thankfully acknowledged. Tarun Jha is also thankful for the financial support from RUSA 2.0 of UGC, New Delhi, India to Jadavpur University, Kolkata, India. Authors are thankful to Prof. Alla P. Toropova and Prof. Andrey Toropov of Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Italy for their useful suggestions during preparation of the revised manuscript. We are very much thankful to the Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India and Department of Pharmaceutical Sciences, Dr. Harisingh Gour University, India for providing the research facilities.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.molstruc.2020.129026.

References

[1] A.K. Ghosh, M. Brindisi, D. Shahabi, M.E. Chapman, A.D. Mesecar, Drug development and medicinal chemistry efforts toward SARS-coronavirus and Covid-19 therapeutics, ChemMedChem (2020) https://doi.org/10.1002/cmdc.202000223.
[2] T. Pillaiyar, S. Meenakshisundaram, M. Manickam, Recent discovery and development of inhibitors targeting coronaviruses, Drug Discov. Today (2020) https://doi.org/10.1016/j.drudis.2020.01.015.
[3] 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. (2020) https://doi.org/10.1080/07391102.2020.1780946.
[4] https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020 (as accessed on 7th June 2020).
[5] https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (as accessed on 8th July 2020).
[6] http://www.who.int/csr/sars/archive/2003_05_07a/en (as accessed on 10th May 2020).
[7] S.K. Enmozhi, K. Raja, I. Sebastine, J. Joseph, Andrographolide as a potential inhibitor of SARS-CoV-2 main protease: an in silico approach, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1760136.
[8] R. Islam, 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. (2020) https://doi.org/10.1080/07391102.2020.1761883.
[9] T. Joshi, T. Joshi, P. Sharma, H.Pundir S.Mathpal, V. Bhatt, S. Chandra, In silico screening of natural compounds against COVID-19 by targeting Mpro and ACE2 using molecular docking, Eur. Rev. Med. Pharmaco. 24 (2020) 45294536.
[10] R.S. Joshi, S.S. Jagdale, S.B. Bansode, S.S. Shankar, M.B. Tellis, V.K. Pandya, A. Chugh, A.P. Giri, M.J. Kulkarni, Discovery of potential multi-target-directed ligands by targeting host-specific SARS-CoV-2 structurally conserved main protease, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1760137.
[11] M.T. Khan, A. Ali, Q. Wang, M. Irfan, A. Khan, M.T. Zeb, Y.J. Zhang, S. Chinnasamy, D.Q. Wei, Marine natural compounds as potents inhibitors against the main protease of SARS-CoV-2. A molecular dynamic study, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1769733.
[12] R.J. Khan, R.K. Jha, G.M. Amera, M. Jain, E. Singh, A. Pathak, R.P. Singh, J. Muthukumaran, A.K. Singh, 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) https://doi.org/10.1080/07391102.2020.1753577.
[13] S. Das, S. Sarmah, S. Lyndem, A.S. Roy, An investigation into the identification of potential inhibitors of SARS-CoV-2 main protease using molecular docking study, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1763201.
[14] 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. (2020) https://doi.org/10.1080/07391102.2020.1751298.
[15] K. Al-Khafaji, D. AL-DuhaidahawiL, T.T. Tok, Using integrated computational approaches to identify safe and rapid treatment for SARS -CoV-2, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1764392.
[16] 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. (2020) https://doi.org/10.1080/07391102.2020.1766572.
[17] A.B. Gurung, M.A. Ali, J. Lee, M.A. Farah, K.M. Al-Anazi, Unravelling lead antiviral phytochemicals for the inhibition of SARS-CoV-2 Mpro enzyme through in silico approach, Life Sci. (2020) 117831.
[18] 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. (2020) https://doi.org/10.1080/07391102.2020.1764868.
[19] D. Kumar, K. Kumari, A. Jayaraj, V. Kumar, R.V. Kumar, S.K. Dass, R. Chandra, P. Singh, Understanding the binding affinity of noscapines with protease of SARS-CoV-2 for COVID-19 using MD simulations at different temperatures, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1752310.
[20] X. Ren, X. Shao, X.X. Li, X.H. Jia, T. Song, W.Y. Zhou, P. Wang, Y. Li, X.L. Wang, Q.H. Cui, P.J. Qiu, Identifying potential treatments of COVID-19 from traditional chinese medicine (TCM) by using a data-driven approach, J. Ethnopharmacol. (2020) 112932 https://doi.org/10.1016/j.jep.2020.1932.
[21] H. Su, S. Yao, W. Zhao, M. Li, J. Liu, W. Shang, C.Ke H.Xie, 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) https://doi.org/10.1101/2020.04.13.038687.
[22] H. Liu, F. Ye, Q. Sun, H. Liang, C. Li, R. Lu, B. Huang, W. Tan, L. Lai, Scutellariabaicalensis extract and baicalein inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro, BioRxiv (2020) https://doi.org/10.1101/2020.04.10.035824.
[23] L. Mittal, A. Kumari, M. Srivastava, M. Singh, S. Asthana, Identification of potential molecules against COVID-19 main protease through structure-guided virtual screening approach, J. Biomol. Struct. Dyn. (2020) https://doi.org/10.1080/07391102.2020.1768151.
[24] M.T. ulQamar, 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. (2020) https://doi.org/10.1016/j.jpha.2020.03.009.
[25] 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. (2020) https://doi.org/10.1080/07391102.2020.1757510Z.
[26] D. Umesh, C. Kundu, Selvaraj, S.K. Singh, V.K. Dubey, Identification of new anti-nCoV drug chemical compounds from Indian spices exploiting SARS-CoV-2 main protease as target, J. Biomol. Struct. Dyn. (2020) 1–9.
[27] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, Acta Pharm. Sin. B 10 (2020) 766–788.
[28] Z.J. Zhang, W.Y. Wu, J.J. Hou, L.L. Zhang, F.F. Li, L. Gao, X.D. Wu, J.Y. Shi, R. Zhang, H.L. Long, M Lei, Active constituents and mechanisms of respiratory detox shot, a traditional Chinese medicine prescription, for COVID-19 control and prevention: network-molecular docking-LC-MSE analysis, Integr. Med. (2020) https://doi.org/10.1016/j.joim.2020.03.004.
[29] B.J. Neves, R.C. Braga, C.C. Melo-Filho, J.T. Moreira-Filho, E.N. Muratov, C.H. Andrade, QSAR-based virtual screening: advances and applications in drug discovery, Front. Pharmacol. 9 (2018) 1275.
[30] N. Brown, P. Ertl, R. Lewis, T. Luksch, D. Reker, N. Schneider, Artificial intelligence in chemistry and drug design, J. Comput. Aided Mol. Des. 34 (2020) 709–715.
[31] N. Adhikari, S.K. Baidya, A. Saha, T. Jha, Structural insight into the viral 3C-like protease inhibitors: comparative SAR/QSAR approaches, in: In Viral Proteases and their Inhibitors, Academic Press, U.S.A, 2017, pp. 317–409.
[32] S.A. Amin, N. Adhikari, T. Jha, Design of aminopeptidase N inhibitors as anti–cancer agents, J. Med. Chem. 61 (2018) 6468–6490.
[33] S.A. Amin, N. Adhikari, T. Jha, Exploration of histone deacetylase 8 inhibitors through classification QSAR study: Part II*, J. Mol. Struct. 1204 (2020) 127529.
[34] S.A. Amin, T. Jha, Fight against novel coronavirus: A perspective of medicinal chemists, Eur. J. Med. Chem. 201 (2020) 112559.
[35] T. Bobrowski, V.M. Alves, C.C. Melo-Filho, D. Korn, S. Auerbach, C. Schmitt, E.N. Muratov, A. Tropsha, Computational models identify several FDA approved or 1 experimental drugs as putative agents against SARS-CoV-2, ChemRxiv (2020) https://doi.org/10.26434/chemrxiv.12153594.v1.
[36] https://chemrxiv.org/articles/Computational_Models_Identify_Several_FDA_Approved_or_Experimental_Drugs_as_Putative_Agents_Against_SARS-CoV-2/12153594/1 (as accessed on 21st May 2020).

[37] P. Polishchuk, Interpretation of quantitative structure–activity relationship models: past, present, and future, J. Chem. Inf. Model. 57 (2017) 2618–2639.
[38] P. Polishchuk, O. Tinkov, T. Khristova, L. Ognichenko, A. Kosinskaya, A. Varnek, V. Kuz'min, Structural and physico-chemical interpretation (SPCI) of QSAR models and its comparison with matched molecular pair analysis, J. Chem. Inf. Model. 56 (2016) 1455–1469.
[39] A.P. Toropova, A.A. Toropov, E. Benfenati, A quasi-QSPR modelling for the photocatalyticdecolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL, SAR QSAR Environ. Res. 26 (2015) 29–40.
[40] A.A. Toropov, A.P. Toropova, T. Puzyn, E. Benfenati, G. Gini, D. Leszczynska, J. Leszczynski, QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells, Chemosphere 92 (2013) 31–37.
[41] A.A. Toropov, A.P. Toropova, G. Raitano, E. Benfenati, CORAL: building up QSAR models for the chromosome aberration test, Saudi J. Biol. Sci. (2018) https://doi.org/10.1016/j.sjbs.2018.05.013.
[42] A. Worachartcheewan, P. Mandi, V. Prachayasittikul, A.P. Toropova, A.A. Toropov, C. Nantasenamat, Large-scale QSAR study of aromatase inhibitors using SMILES-based descriptors, Chemometr.Intell. Lab. 138 (2014) 120–126.
[43] https:https://github.com/DrrDom/rspci (as accessed on 21st May 2020).
[44] A. Ganeshpurkar, A.K. Saluja, The pharmacological potential of rutin, Saudi Pharm. J. 25 (2017) 149–164.
[45] M.Q. Man, B. Yang, P.M. Elias, Benefits of hesperidin for cutaneous functions, Evidence-Based Complementary and Alternative Medicine, 2019 https://doi.org/10.1155/2019/2676307.
[46] E. Sukhbold, S. Sekimoto, E. Watanabe, A. Yamazaki, L. Yang, M. Takasugi, K. Yamada, R. Hosomi, K. Fukunaga, H. Arai, Effects of oolonghomobisflavanA on oxidation of low-density lipoprotein, Biosci. Biotechnol. Biochem. 81 (2017) 1569–1575 https://doi.org/10.1080/09168451.2017.1314758.
[47] M. Weerawatanakorn, W.L. Hung, M.H. Pan, S. Li, D. Li, X. Wan, C.T. Ho, Chemistry and health beneficial effects of oolong tea and theasinensins, Food Sci. Hum. Wellness 4 (2015) 133–146.
[48] J. Ruan, C. Zheng, Y. Liu, L. Qu, H. Yu, L. Han, Y. Zhang, T. Wang, Chemical and biological research on herbal medicines rich in xanthones, Molecules 22 (2017) 1698.
[49] M. Winnig, B. Bufe, N.A. Kratochwil, J.P. Slack, W. Meyerhof, The binding site for neohesperidindihydrochalcone at the human sweet taste receptor, BMC Struct. Biol. 7 (2007) 66 https://doi.org/10.1186/1472-6807-7-66.
[50] L. Korkina, V. Kostyuk, C. De Luca, S. Pastore, Plant phenylpropanoids as emerging anti-inflammatory agents, Mini Rev. Med. Chem. 11 (2011) 823–835 https://doi.org/10.2174/138955711796575489.
[51] M. Pereira, I.P. Siba, L.R. Chioca, D. Correia, M.A. Vital, M.G. Pizzolatti, A.R. Santos, R. Andreatini, Myricitrin, a nitric oxide and protein kinase C inhibitor, exerts antipsychotic-like effects in animal models, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 35 (2011) 1636–1644.
[52] Y. Tao, S. Zhan, Y. Wang, G. Zhou, H. Liang, X. Chen, H. Shen, Baicalin, the major component of traditional Chinese medicine Scutellariabaicalensis induces colon cancer cell apoptosis through inhibition of onco miRNAs, Sci. Rep. 8 (2018) 1 https://doi.org/10.1038/s41598-018-32734-2.
[53] M. Ding, R. Feng, S.Y. Wang, L. Bowman, Y. Lu, V.Castranova Y.Qian, B.H. Jiang, X. Shi, Cyanidin-3-glucoside, a natural product derived from blackberry, exhibits chemopreventiveand chemotherapeutic activity, J. Biol. Chem. 281 (2006) 17359–17368 https://www.jbc.org/content/281/25/17359.full.
[54] F. Chen, K.H. Chan, Y. Jiang, R.Y. Kao, H.T. Lu, K.W. Fan, V.C. Cheng, W.H. Tsui, I.F. Hung, T.S. Lee, Y. Guan, In vitro susceptibility of 10 clinical isolates of SARS coronavirus to selected antiviral compounds, J. Clin. Virol. 31 (2004) 69–75.