



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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Unmasking of crucial structural fragments for coronavirus protease inhibitors and its implications in COVID-19 drug discovery



Kalyan Ghosh<sup>a,#</sup>, Sk. Abdul Amin<sup>b,#</sup>, Shovanlal Gayen<sup>a,\$,\*</sup>, Tarun Jha<sup>b,\*</sup>

<sup>a</sup>Laboratory of Drug Design and Discovery, Department of Pharmaceutical Sciences, Dr. Harisingh Gour University, Sagar, MP, India.

<sup>b</sup>Natural Science Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India.

## ARTICLE INFO

### Article history:

Received 20 August 2020

Revised 19 March 2021

Accepted 20 March 2021

Available online 26 March 2021

### Keywords:

SARS-CoV

PLpro

3CLpro

SARpy

Bayesian model

Fragment based drug discovery

## ABSTRACT

Fragment based drug discovery (FBDD) by the aid of different modelling techniques have been emerged as a key drug discovery tool in the area of pharmaceutical science and technology. The merits of employing these methods, in place of other conventional molecular modelling techniques, endorsed clear detection of the possible structural fragments present in diverse set of investigated compounds and can create alternate possibilities of lead optimization in drug discovery. In this work, two fragment identification tools namely SARpy and Laplacian-corrected Bayesian analysis were used for previous SARS-CoV PLpro and 3CLpro inhibitors. A robust and predictive SARpy based fragments identification was performed which have been validated further by Laplacian-corrected Bayesian model. These comprehensive approaches have advantages since fragments are straight forward to interpret. Moreover, distinguishing the key molecular features (with respect to *ECFP\_6* fingerprint) revealed good or bad influences for the SARS-CoV protease inhibitory activities. Furthermore, the identified fragments could be implemented in the medicinal chemistry endeavors of COVID-19 drug discovery.

© 2021 Elsevier B.V. All rights reserved.

## 1. Introduction

World Health Organization (WHO) has declared COVID-19 disease as a global pandemic. To date over 120 million confirmed cases along with near 3 million COVID-19-related mortalities were reported worldwide [1,2]. Studies on COVID-19 revealed that betacoronaviruses also known as SARS-CoV-2 usually produce several proteases during their life cycle [3]. Among them two key proteases essential for viral replications are papain-like protease (PLpro) and 3C like protease (3CLpro). The papain-like protease discharge several diUbLys48 products by cleaving ISG15, a two-domain Ub-like protein, and Lys48-linked polyUb chains [4-6]. Thus, they function by hijacking the host ubiquitin enzyme that is responsible for the host defence mechanism [7,8]. Whereas, in 3C like protease (3CLpro) also known as SARS-CoV-2 main protease (Mpro) encompasses of 306 amino acid long polypeptide chains and plays a significant role in enzymatic activity leading to its

post-translational processing of replicase polyproteins. The 3CLpro monomer mainly consists of three domains [9]. Both PLpro and 3CLpro are equally important for viral life cycle and play a vital role in the transcription/replication during the infection. Hence, pinpointing these two proteases (*i.e.*, PLpro and 3CLpro) may serve as important targets for designing of several antiviral drugs against the highly contagious COVID-19 [10,11].

Discovering an effective inhibitor against the PLpro and 3CLpro are ongoing [12-17]. The cost of bringing a single effective drug candidate to market not only requires billions of dollars but also requires huge time to accomplish [18]. In order to make this process faster, researchers around the globe has shifted their focus to the various emerging *in silico* lead discovery methods (Figure 1) [19,20]. Presently, among all the known computational techniques, the fragment based drug discovery (FBDD) proved to be very efficient method of drug design and discovery [21,22]. The identified important structural fragments will be helpful for efficient designing of lead molecule against SARS-CoV-2 main proteases - PLpro and 3CLpro. This way of lead designing process based on smaller fragments may remove the usual ADME associated problems with protease inhibitors based drug discovery [23].

In the past 20 years, the concept of FBDD has recognized itself as a strategic approach for finding high-quality lead candidates [19]. It has a potential to address inflexible biological tar-

E-mail addresses: [shovanlal.gayen@gmail.com](mailto:shovanlal.gayen@gmail.com) (S. Gayen), [tjupharm@yahoo.com](mailto:tjupharm@yahoo.com) (T. Jha).

# Authors have equal contribution.

\$ Present address: Laboratory of Drug Design and Discovery, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India.

\* Corresponding authors.

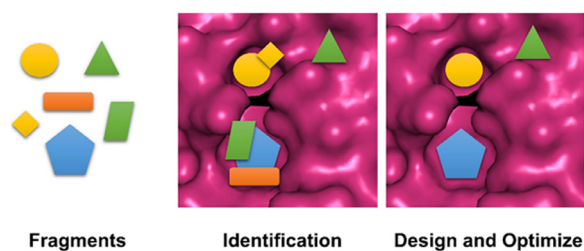


Fig. 1. Fragment based drug discovery process.

gets involving intracellular protein–protein interactions also. Today, more than ten FBDD based lead molecules targeting diverse protein families in different diseases have been progressed into clinical trials [20]. In FBDD the identification of very small molecules or so-called “fragments” which are responsible for binding to their specific target proteins are carried out [21]. Typically, screening of a small library of low molecular weight compounds for binding to a specific part of the target/receptor is the starting point of FBDD [22]. The structural evidence and thermodynamics of fragment binding proposed that they tends to bind to the crucial regions of the protein sites, mainly contribute to the enthalpy driven free energy change of ligand binding [24]. Thus, the fragments can contribute to the crucial part of a drug molecule.

In this study different fragment based computational tools like SARpy analysis and Bayesian classification was applied on to the dataset of previous SARS-CoV PLpro and 3CLpro inhibitors. The methods were successfully used previously to identify different fragments of various pharmacological and toxicological properties [25–29]. In these methods, the whole molecular structure was fragmented in a certain way into small parts and finally the contribution to increase or decrease the pharmacological activity/toxicological properties was identified. Thus, the identified fragments can give potential clues in the lead optimization process. Since the binding site of PLpro and 3CLpro enzymes were conserved between SARS-CoV and SARS-CoV-2 the fragments will be very useful for COVID-19 drug discovery [3,14,30,31]. The current study will offer an idea to in-depth qualification of fragment hits. This will stimulate further research by providing valuable guidance to the medicinal chemists for designing of novel PLpro and 3CLpro inhibitors against previous SARS as well as recent COVID-19 diseases.

## 2. Materials and methods

### 2.1. Dataset

The dataset comprising of diverse SARS-CoV PLpro and 3CLpro inhibitors with their biological activities were retrieved from different literatures [32–40]. The duplicate compounds and the compounds with no inhibitory activities were removed from the dataset. Finally, 91 PLpro and 88 3CLpro inhibitors were selected for the determination of structural alerts using SARpy analysis [26–28]. The activity threshold was considered by taking the average of SARS-CoV PLpro  $pIC_{50}$  values (*i.e.*,  $IC_{50}$  value of 6,200 nM), whereas in 3CLpro the activity threshold was set to the  $IC_{50}$  value of 10,000 nM. Thus, the molecules having the  $pIC_{50}$  values higher than the active threshold value were selected as *actives* compounds, whereas the molecules with  $pIC_{50}$  values lower than the active thresholds were considered as *inactives*. Finally, out of 91 PLpro inhibitors, 40 molecules were found to be *active* and 51 *inactives* **Tables S1**. Among 88 3CLpro inhibitors, 27 molecules were considered as *actives* and 61 *inactives* as depicted in the **Table S2**.

### 2.2. SARpy analysis

SARpy is a python based standalone software programme for automated QSAR model development [41–44]. The software uses user-defined SMILES notations for generating substructures in the set and tries to correlate between the particular molecular structures and their biological activity using three different steps. These are:

- Fragmentation:** In this step, recursive simple fragmentation algorithm is used to detect the chemical substructures present in the training set molecules. It iterates over every bond in the input structures and tries to generate possible pair of fragments.
- Evaluation:** Once all the substructures have been generated, their individual evaluation is done in order to detect the possible structural alerts (SAs) present in them.
- Extraction:** Finally, from the large collection of structural alerts that were generated, only the reduced sets of predicted rules were applied. In our current study, the rule sets were generated using two different settings they are: standard settings and modified settings. Further, the parameters selected in standard settings are minimum 2, maximum 18 atoms and occurring in a minimum 3 training set substances. Whereas, in modified settings minimum 3 and maximum 18 atoms and occurring in a minimum 5 training set substances was selected [45].

Additionally, each of the above SA settings was verified with two different single alert precision parameters they are: Auto MIN and Auto MAX. In the auto MIN settings the false negatives are minimized thus responsible for increasing the sensitivity values. Whereas, the auto MAX settings is considered for increasing the specificity values by minimizing the false positives in the training set. Thus, the aforementioned settings were first applied to split 1 to identify the settings generating the best overall result. Finally, the best optimal settings were then applied to both split 2 and split 3 for model development and evaluation [44–47].

Evaluation of developed models is one of the key features for the prediction of model performances. From the three different splits, the number of true positives, true negatives, false positives and false negatives were determined. This information was further used for prediction of performances such as sensitivity, specificity, accuracy, Matthew's correlation coefficients (MCC), error rate (**Table 1**). The unpredicted rate was also calculated as prediction performance measure for different models by using the formula shown in **Table 1**.

### 2.3. Bayesian classification study

Bayesian classification study by the aid of Biovia Discovery Studio (DS) software [48] was conducted. Before conduct this Bayesian classification study, several fundamental molecular features namely,  $ALogP$ , Molecular weight ( $MW$ ), Number of hydrogen bond donors ( $nHBD$ ), Number of hydrogen bond acceptors ( $nHBA$ ), Number of rotatable bonds ( $nRB$ ), Number of rings ( $nRings$ ), Number of aromatic rings ( $nAR$ ), Molecular fractional polar surface area ( $MFPSA$ ) of the dataset molecules have been calculated [48]. Alongside those molecular properties extended connectivity fingerprint of diameter 6 ( $ECFP_6$ ), a topological fingerprint descriptor was also considered for this study [49]. The quality of this classification model was evaluated using the Receiver operating characteristics (ROC) [50]. Furthermore, the sensitivity, specificity and concordance were calculated for both the training and the test sets [25].

**Table 1**  
The equations of statistical validation parameters.

Entry	Parameter	Equation
1	Sensitivity	$TP/(TP+FN)$
2	Specificity	$TN/(TN+FP)$
3	Accuracy	$(TP+TN)/(TP+FP+TN+FN)$
4	MCC	$(TP*TN)-(FP*FN)/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$
5	Error rate	$(FP+FN)/Total$
6	Unpredicted rate	Number of unpredicted compounds / number of compounds in the dataset

TP = True positive, TN = True negative, FP = False positive, FN = False negative.

**Table 2**  
Performance of model building settings as obtained in case of PLpro and 3CLpro inhibitors.

PLpro inhibitors Settings	Rules (#) (Positive, Negative)	Training set			Test set		
		Error rate	Unpredicted rate	Structures matched	Error rate	Unpredicted rate	Structures matched
Standard-Auto MIN	14 (7, 7)	0.19	0.03	71	0.06	0.39	11
Standard-Auto MAX	9 (3, 6)	0.00	0.53	34	0.00	0.61	7
<b>Modified-Auto MIN</b>	<b>10 (5, 5)</b>	<b>0.23</b>	<b>0.04</b>	<b>70</b>	<b>0.22</b>	<b>0.22</b>	<b>14</b>
Modified-Auto MAX	5 (2, 3)	0.00	0.62	28	0.22	0.22	0
<b>3CLpro inhibitors</b>	<b>14 (4, 10)</b>	<b>0.19</b>	<b>0.03</b>	<b>68</b>	<b>0.17</b>	<b>0.06</b>	<b>17</b>
Standard-Auto MIN	7 (1, 6)	0.00	0.41	41	0.39	0.00	7
Standard-Auto MAX	8 (2, 6)	0.11	0.11	62	0.17	0.00	17
Modified-Auto MIN	3 (1, 2)	0.00	0.60	28	0.39	0.00	7
Modified-Auto MAX							

### 3. Results and discussions

#### 3.1. Training Set and Test Set division

In this study 91 PLpro and 88 3CLpro inhibitors were selected as dataset in order to identify the fragments important for controlling the protease inhibition. The inhibitors were distributed into training set (80%) and test set (20%) randomly by using CORAL software. Again, the dataset was divided into training set A (80%) and test set A (20%) using the rational division algorithm (Kennard-Stone algorithm) [51-52] by using the DatasetDivisionGUI 1.2 tool [53]. This algorithm selects the objects so they are divided evenly throughout the descriptor space of the original data set. The descriptors such as *AlogP*, Molecular weight (*MW*), Number of hydrogen bond donors (*nHBD*), Number of hydrogen bond acceptors (*nHBA*), Number of rotatable bonds (*nRB*), Number of rings (*nRings*), Number of aromatic rings (*nAR*), Molecular fractional polar surface area (*MFPSA*) and biological activity (*active/inactive*).

After the division of training and test set the key fragments from developed models were obtained by SARpy as well as Bayesian classification analysis and were interpreted for their importance in protease inhibition.

#### 3.2. Model Development

##### 3.2.1. SARpy analysis

In SARpy analysis, model building setting was considered as the initial step in order to build robust classification models. Different model buildings settings like: Standard-Auto MIN; Standard-Auto MAX; Modified-Auto MIN; Modified-Auto MAX was applied to the dataset. After model development, the three different statistical parameters like: error rate, unpredicted rate and number of structured matched along with the number of generated rules were considered for model evaluation for both training and test

set. The performance of different model building settings for both PLpro and 3CLpro inhibitors are shown in Table 2. Among these, the Modified-Auto MIN setting was found to have acceptable parameters such as: lowest unpredicted rate 0.04 and 0.22; highest structures matched 70 and 14 for both training and test set respectively in case of PLpro inhibitors, whereas, Standard-Auto MIN settings was found to be suitable for 3CLpro inhibitors. So, these settings were considered for further model development.

Further, to evaluate the capability of the selected methods the dataset was split again into training and test set two times (Split 2 and 3) separately and was used for the model development. The uniform distribution of the training and test set SARS-CoV PLpro and 3CLpro inhibitors in the PCA three dimensional plots referred a proper division of the training and the test sets from the split 2 and 3. After the division, by using Modified-Auto MIN settings for PLpro inhibitors and Standard-Auto MIN settings for 3CLpro inhibitors, total three different models from each setting were generated as shown in Table 3.

As expected, all the developed models were found to have significant statistical parameters. Among them, the model developed by using split 3 were found to have better statistically parameters in both the cases.

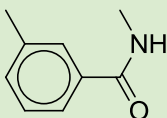
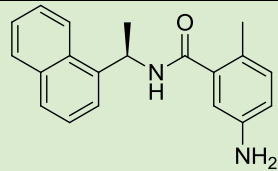
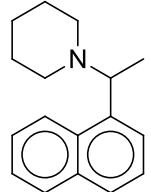
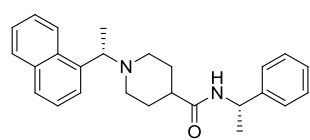
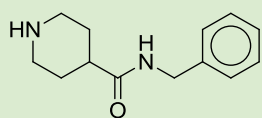
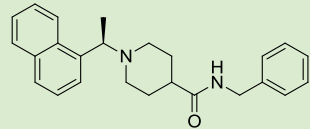
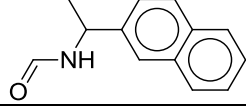
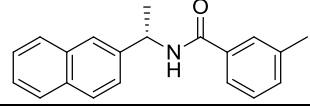
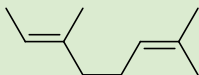
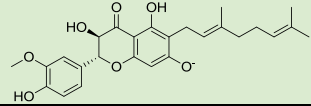
In case of PLpro inhibitors, the best model with sensitivity = 0.67, 0.83; specificity = 0.93, 0.75; and accuracy = 0.79, 0.79 were found for both training and test set respectively. Whereas, for 3CLpro inhibitors the training and test sets were found to have sensitivity = 0.73, 1.00; specificity = 1.00, 0.92; and accuracy = 0.86, 0.94, respectively. Additionally, the MCC value 0.61, 0.75 for training and 0.58, 0.86 for test in case of PLpro and 3CLpro respectively indicates that the classifier have good performance for the identification of structural fragments from the dataset. These fragments are high chemical reactivity molecular fragments responsible for modulating the protease inhibition. Therefore, the fragments obtained from compounds can be used to flag the potential chemi-

**Table 3**  
Performance of prediction models of PLpro and 3CLpro inhibitors developed from 3 different splits.

PLpro inhibitors	Training set			Test set		
	split 1	split 2	split 3	split 1	split 2	split 3 <sup>a</sup>
Sensitivity	0.64	0.74	<b>0.67</b>	0.71	0.56	<b>0.83</b>
Specificity	1.00	0.89	<b>0.93</b>	0.71	1.00	<b>0.75</b>
Accuracy	0.76	0.80	<b>0.79</b>	0.71	0.71	<b>0.79</b>
MCC	0.61	0.62	<b>0.61</b>	0.43	0.56	<b>0.58</b>
Error rate	0.23	0.18	<b>0.21</b>	0.22	0.22	<b>0.17</b>
Unpredicted rate	0.04	0.11	<b>0.04</b>	0.22	0.22	<b>0.22</b>
Structures matched (#)	70	65	<b>70</b>	14	14	<b>14</b>
<b>3CLpro inhibitors</b>						
Sensitivity	0.83	0.81	<b>0.70</b>	0.50	0.66	<b>1.00</b>
Specificity	0.80	0.92	<b>1.00</b>	1.00	1.00	<b>0.92</b>
Accuracy	0.80	0.89	<b>0.86</b>	0.82	0.77	<b>0.94</b>
MCC	0.53	0.73	<b>0.75</b>	0.60	0.62	<b>0.86</b>
Error rate	0.19	0.10	<b>0.13</b>	0.17	0.17	<b>0.06</b>
Unpredicted rate	0.03	0.06	<b>0.06</b>	0.06	0.28	<b>0.06</b>
Structures matched (#)	68	66	<b>66</b>	17	13	<b>17</b>

<sup>a</sup> The best model obtained from split 3 is marked in bold.

**Table 4**  
Important identified structural alerts of PLpro inhibitors as obtained from SARpy analysis.

SMARTS	Structural alert (SA)	Structure
<chem>Cc1cc(ccc1)C(=O)NC</chem>		
<chem>N1(CCC(CC1))C(C)c1cccc2ccccc12</chem>		
<chem>N1(CCC(CC1)C(=O)NCc1cc(ccc1))</chem>		
<chem>C(=O)NC(c1cc2c(cc1)cccc2)C</chem>		
<chem>CC=C(CCC=C(C)C)C</chem>		

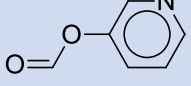
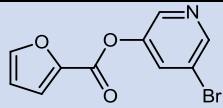
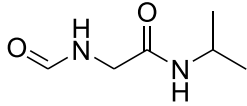
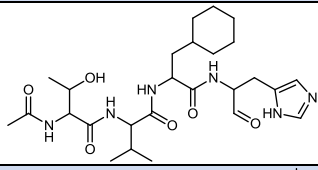
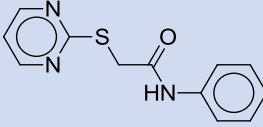
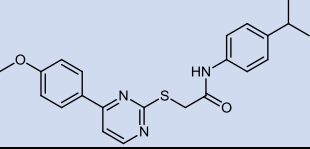
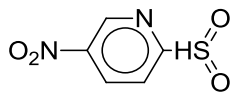
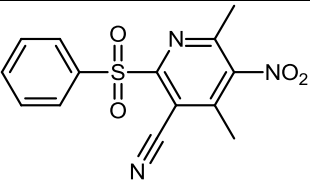
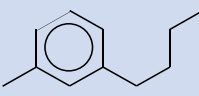
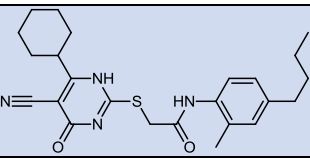
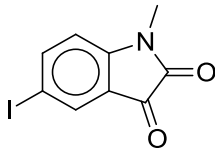
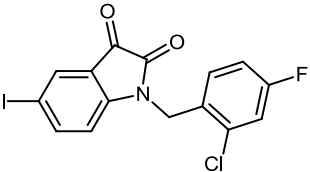
cal compounds responsible for protease inhibition. The fragments along with the SMARTS patterns as obtained from the best model split 3 of both PLpro and 3CLpro inhibitors are shown below in [Table 4](#) and [5](#), respectively.

Furthermore, the outcomes from the Kennard-Stone rational division method yielded comparable results to the random division ([Supplementary Table S3-S4](#)). Similar fingerprints such as N1(CCC(CC1)C(=O)NCc1cc(ccc1)) and CC=C(CCC=C(C)C)C for PLpro inhibitors and N(C(=O)CSc1cccc(n1))c1cccc1, S(=O)(=O)c1nc(c(c1))N(=O)=O for 3CLpro inhibitors were also obtained from both Random division and Kennard-Stone rational division methods as can be seen from [Supplementary Table S5-S6](#).

### 3.2.2. Bayesian classification analysis

The processes of Laplacian-corrected Bayesian classification model development was done on training set and externally validated on test set compounds. The selection of the training and test sets was considered from the best model generated from SARpy model. The results of Laplacian-corrected Bayesian classification model analysis derived from training set compounds are summarized in [Table 6](#). It showed that the Bayesian classification model with sensitivity, specificity and concordance values of 0.767, 0.930 and 0.863, respectively, was obtained for the training set of SARS-CoV PLpro inhibitors. Likewise, the Bayesian classification model for the training set of SARS-CoV 3CLpro inhibitors displays sensitivity, specificity and concordance values of 1.000, 0.729 and 0.814, respectively.

**Table 5**  
Important identified structural alerts of 3CLpro inhibitors as obtained from SARpy analysis.

SMARTS	Structural alert (SA)	Structure
<chem>C(=O)Oc1nc1</chem>		
<chem>CC(C)NC(=O)CNC(=O)</chem>		
<chem>N(C(=O)CSc1nc1)C1CCCC1</chem>		
<chem>S(=O)(=O)c1nc(c(c1))N(=O)=O</chem>		
<chem>c1ccc(cc1C)CCCC</chem>		
<chem>CN1C(=O)C(=O)c2cc(I)ccc12</chem>		

**Table 6**  
Statistics of training and test set of SARS-CoV PLpro and 3CLpro inhibitors.

Target	Set	ROC	TP	FN	FP	TN	Sensitivity	Specificity	Concordance
PLpro	Training#	0.751	23	7	3	40	0.767	0.930	0.863
	Test	0.662	8	2	3	5	0.800	0.625	0.722
3CLpro	Training#	0.840	22	0	13	35	1.000	0.729	0.814
	Test	0.908	5	0	7	6	1.000	0.462	0.611

Receiver operating characteristics (ROC); # a 5-fold cross validation is performed for the training set to calculate the statistics.

The ROC of 0.751, 0.840 and 0.662, 0.908 were revealed for the training, test set of SARS-CoV PLpro and SARS-CoV 3CLpro inhibitors, respectively. These results demonstrate that the both models were robust and also provide a good predictive value.

### 3.3. Mining of the fragments those modulate biological activity

In order to identify the potential SARS-CoV protease inhibitor activity regulators, particular focus is given on the molecular modelling techniques like SARpy and Bayesian classification analysis to identify the fragments controlling the protease inhibition. The fragments can increase or reduce the protease inhibitions are classified as good or bad fragments, respectively. For this rational division method and their corresponding QSAR methodologies, random selection provides essentially equivalent results.

#### 3.3.1. Interpretation of fragments for SARS-CoV PLpro inhibitors

The important fragments for the PLpro inhibitors were identified as shown in Table 4. The possible role of these fragments controlling the PLpro inhibition is shown in Figure 2.

The molecular fragment Cc1cc(ccc1)C(=O)NC signifies the presence of any benzamide moiety found in compounds **A009** (SARS-CoV PLpro IC<sub>50</sub> = 0.6 μM) is responsible for inducing activity. Hence, this feature is responsible for modulating activities of SARS-CoV PLpro inhibitors. This fingerprint is marked in the Figure 2. Not surprisingly, this observation is in agreement with fingerprints G9, G10, G12, G14 (Figure S1) predicted by the Bayesian classification model on SARS-CoV PLpro inhibitors. The inhibitor **A009** (GRL0617) and SARS-CoV PLpro interaction in Discovery Studio (DS) [48] suggests a pair of hydrogen bonds and some hydrophobic interactions which stabilized the complex (PDB: 3E9S). Notably, the amide group of benzamide moiety of **GRL0617** helps in the interaction with the side chain of D165 and the backbone nitrogen of Q270 to form hydrogen bonds (Figure 3A).

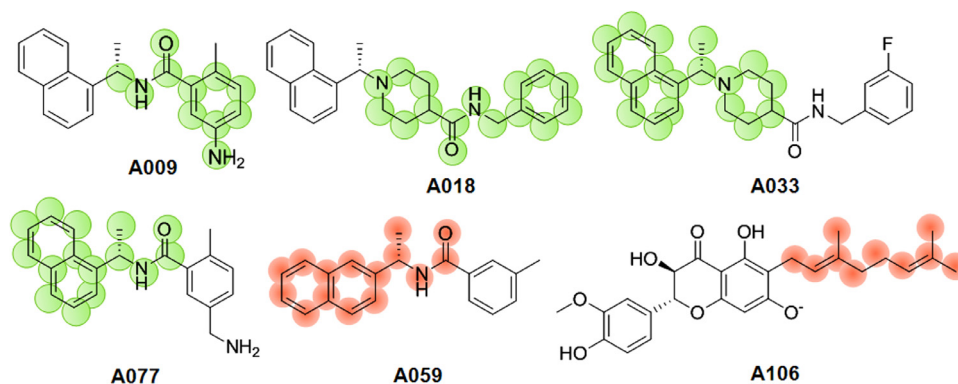


Fig. 2. Important structural alerts for PLpro inhibitors.

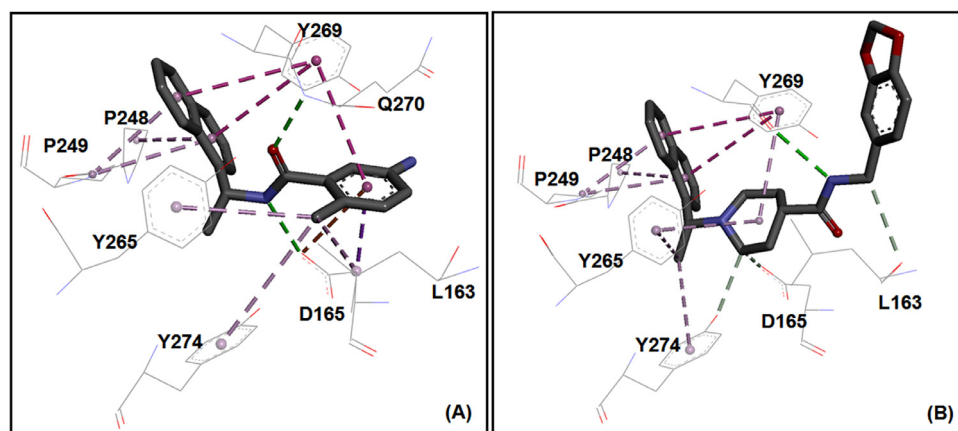


Fig. 3. 3D interaction plots of naphthyl based PLpro inhibitors with SARS-CoV PLpro active site amino acid residues (A) PDB: 3E9S and (B) PDB: 3MJ5.

Next, fragment N1(CCC(CC1))C(C)c1cccc2ccccc12 found in compound **A018**, **A033** and others (**Figure 2**), **A022-A025**, **A027**, **A030**, **A032**, **A033**, **A040-A042**, **A047**, **A048**, **A053**, **A054** signify the presence of 1-(1-(naphthalen-1-yl)ethyl)piperidine group is also answerable for SARS-CoV PLpro inhibitory activities. Meanwhile, we have analyzed the ligand-receptor interaction in Discovery Studio (DS) [48] where the piperidine ring engages in  $\pi$ -sigma interaction with the Y265 while the naphthalene ring forms several interactions with amino acids P248, P249, Y269 at the solvent exposed site of the enzyme (PDB: 3MJ5), as illustrated in **Figure 3B**. Therefore, this indicates that naphthalene and piperidine rings are pivotal for SARS-CoV PLpro inhibition which also supports Bayesian modelling result in which G1-G7 fingerprints predicted these features as positive regulator of the PLpro inhibitory activities. Recent study by Freitas and co-workers suggested that 1-naphthalene based derivatives are also capable of inhibiting SARS-CoV-2 PLpro up to  $IC_{50}$  value of 2.4  $\mu$ M [54].

Meanwhile, it should be noted that the 1-naphthyl containing PLpro inhibitors are more effective than the corresponding 2-naphthyl analogue (**A003**: SARS-CoV PLpro  $IC_{50}$  = 14.8  $\mu$ M). Bad fragment B2 (**Figure S2**) predicted by the Bayesian classification model on SARS-CoV PLpro inhibitors illustrated the negative contribution of 2-naphthyl analogues. In addition, fragment SA N1(CCC(CC1)C(=O)Nc1cc(ccc1)) representing the N-benzylpiperidine-4-carboxamide group found in several active compounds (**A022-A025**, **A027**, **A030**, **A032**, **A033**, etc.) and C(=O)NC(c1cc2c(cc1)cccc2)C representing the N-(1-(naphthalen-2-yl)ethyl)formamide group found in compounds (example **A077** in **Figure 2**) are the potential fragment that are liable for increasing the PLpro inhibitory activities of these compounds. Interestingly, this information is aligned with the experimental results by

means of X-ray structural analyses of inhibitor bound SARS-CoV PLpro enzyme (PDB: 4OW0) [32]. The carboxamide nitrogen of the inhibitor (compound **A033**) engages in the formation of a 3 Å hydrogen bond with the backbone carbonyl of Y269 at the active site (PDB: 4OW0) as depicted in **Figure 4** drawn by PyMOL tool [55].

Further, fragment like 2,6-dimethylocta-2,6-dieneas found in compounds **A106-A112** is represented by the molecular fragment CC=C(CCC=C(C)C)C (**Figure 2**). Likewise, bad fingerprints B8, B14 and B15 (as depicted in **Figure S2**) also supported the negative contribution of 2,6-dimethylocta-2,6-dienefunction (compounds **A106-A112**) to SARS-CoV PLpro inhibitory activities. Additionally, the study was also performed with Kennard-Stone rational dataset division method, similar types of good fingerprints like G5, G7, G8, G9, G10, G20 etc. (**Figure S5**) and bad fingerprints like B8, B13, B14 etc. (**Figure S6**) were obtained.

### 3.3.2. Interpretation of fragments for 3CLpro inhibitors

The fragments crucial for inducing activity to 3CLpro inhibitors were predicted and were further used for interpretation as shown in **Table 5**. The structural alert C(=O)Oc1cnccc1 defines the presence of pyridin-3-yl formate group in compounds **B001-B003** and **B005-B011** is responsible for inducing biological activities against SARS-CoV 3CLpro (**Figure 5**).

This indicates that pyridin-3-yl formate group is favoured for SARS-CoV 3CLpro inhibition which is also supported by G3, G4, G5 and G6 good fingerprints (**Figure S3**) as derived from the *ECFP\_6* fingerprint. In contrast, the presence of pyridine nucleus having branching with  $NO_2$  and  $SO_2$  group represented by the fragment S(=O)(=O)c1nc(c(c1))N(=O)=O is also considered as negative fragment in compounds **B032** and **B033** (**Figure 5**). This in-

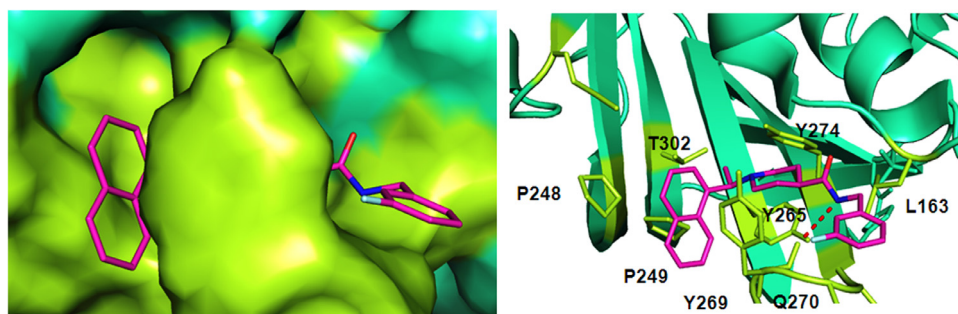


Fig. 4. Binding mode of compound **A033** with SARS-CoV PLpro active site amino acid residues (PDB: 4W0W).

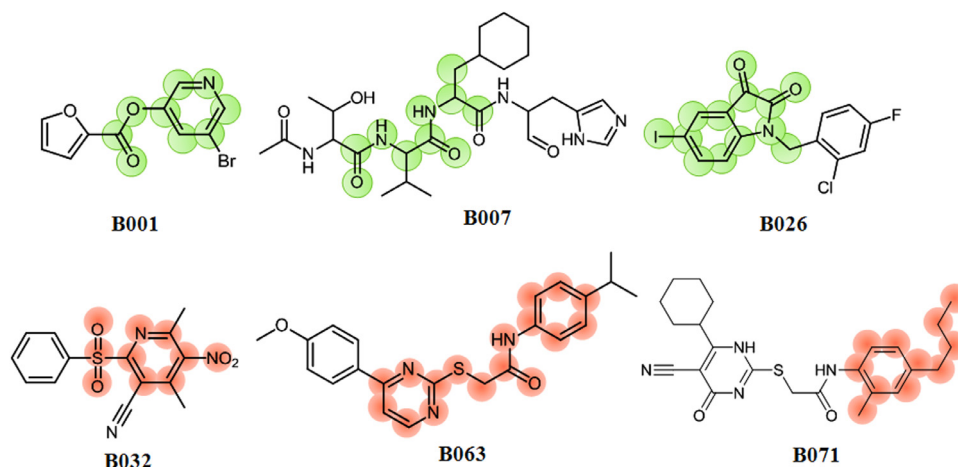


Fig. 5. Important structural alerts for SARS-CoV 3CLpro inhibitors.

indicates that pyridine nucleus having branching with  $\text{NO}_2$  and  $\text{SO}_2$  group is unfavoured which also support previously published results [56]. The fragments like CC(C)NC(=O)C(=O)c2cc(I)ccc12 defines the presence of 5-iodo indoline-2,3-dione group is accountable for SARS-CoV 3CLpro inhibitory activity. Active compounds **B013**, **B026** bearing the fragment are illustrated in Figure 5. Further, the fragments like CC(C)NC(=O)CNC(=O) and CN1C(=O)C(=O)c2cccc12 define the presence of 2-formamido-N-isopropylacetamide and 1-methylindoline-2,3-dione groups, respectively are also accountable for SARS-CoV 3CLpro inhibitory activity. These observations can be justified by observing the compounds **B007**, **B004**, **B017** for fragment CC(C)NC(=O)CNC(=O) and **B013**, **B014**, **B015**, **B019**, **B024** for SA CN1C(=O)C(=O)c2cccc12. Notably, the CC(C)NC(=O)CNC(=O) feature of compound **B007** formed several interactions with SARS-CoV 3CLpro active site amino acid residues (PDB: 3ATW) as illustrated in Figure 6.

By contrast, compounds **B052**, **B057** bearing fragment CC(C)NC(=O)CNC(=O) possess lower SARS-CoV 3CLpro inhibitory activities. Likewise, the fragment, N(C(=O)CSc1nccc(n1))c1cccc1 found in compounds **B063**-**B068** and **B070**-**B077** represent the N-phenyl-2-(pyrimidin-2-ylthio)acetamide group is a significant negative fragment for protease inhibition. This inspection is in agreement with bad fingerprints (Figure S4) as constructed from the ECFP<sub>6</sub> fingerprint. This observation is also verified by the experimental results such as the activity of compound compounds **B063**-**B068** and **B070**-**B077** bearing such fragment exhibited poor SARS-CoV 3CLpro inhibitory activities. Hence, this sub-structural feature is responsible for lowering the proteolytic activity against SARS-CoV Mpro (Figure 7). Further, with Kennard-Stone rational dataset division method, comparable good fingerprints including G1, G8, G9 etc. (Figure S7) and bad fingerprints like: B3, B7, B8 etc. (Figure S8) were also obtained. Thus, these fingerprints will

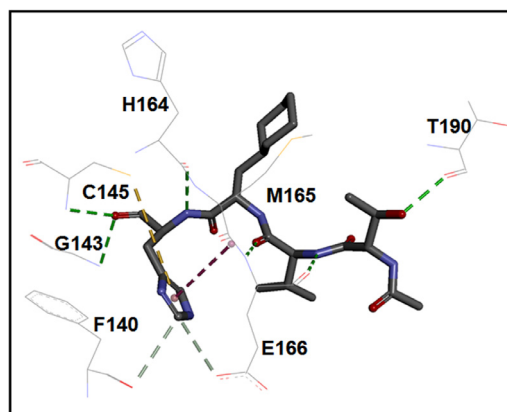


Fig. 6. 3D interaction plot of compound **B007** with SARS-CoV 3CLpro active site amino acid residues (PDB: 3ATW).

be important in optimizing the activity of the 3CLpro inhibitors in future.

Further, the fragment c1ccc(cc1)CCCC represents the presence of 1-butyl benzene. This function should also consider as potential modulator of protease inhibition and is responsible for hindrance of activity in compounds. Not surprisingly, the Bayesian modelling based Good fragments G7, G18, G19 and G20 clearly indicated the importance of furan ring against SARS-CoV 3CLpro inhibition. This observation is in agreement with our previous observation [56] in which compounds bearing both furan and pyridine exhibited effective proteolytic activities ( $\text{IC}_{50}$  in between 50 to 63 nM). In addition, compound **B018** having a furan ring in its structure exhibits promising SARS-CoV 3CLpro inhibitory activities. In a structure-



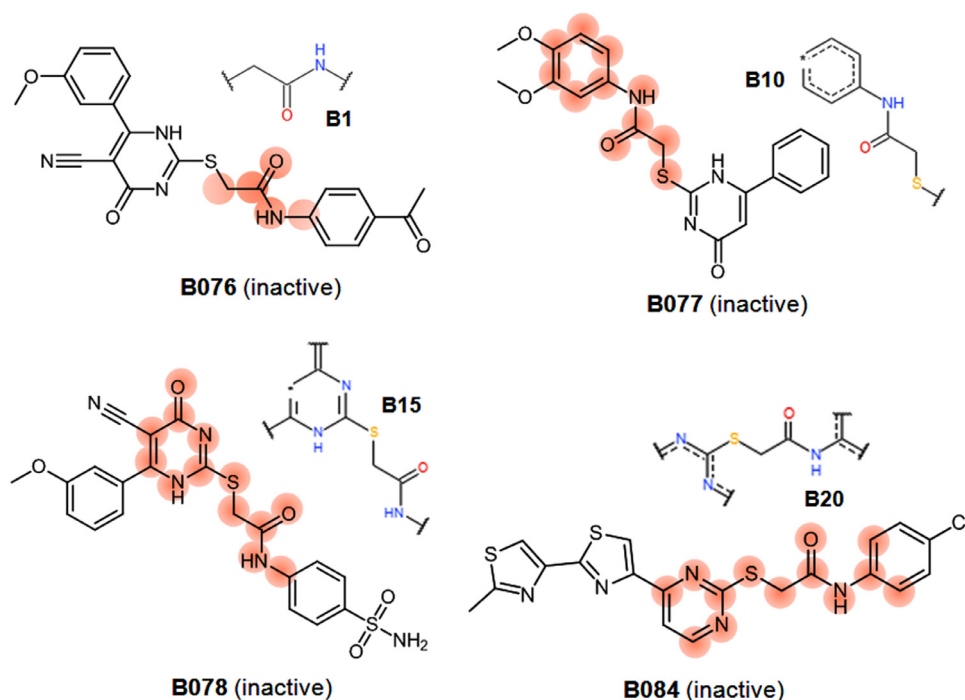


Fig. 7. Structures of some inactive inhibitors containing bad Bayesian fragments.

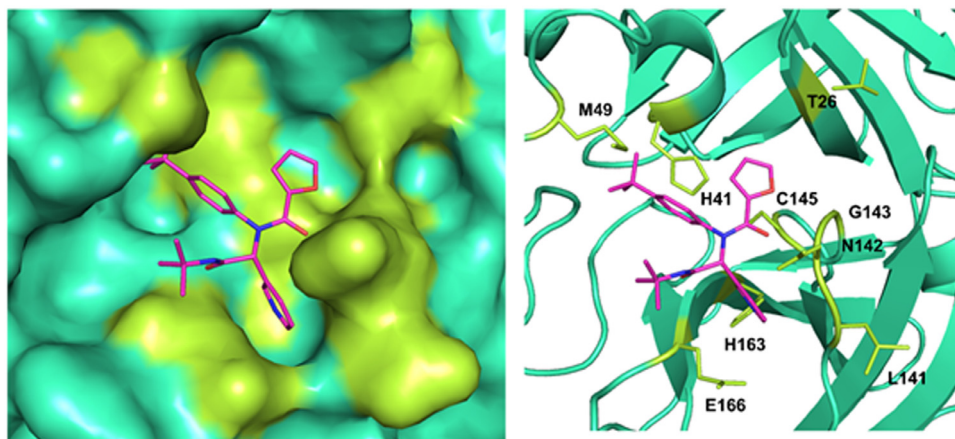


Fig. 8. Binding mode of compound **B018** with SARS-CoV 3CLpro active site amino acid residues as derived from the crystal structure (PDB: 3V3M) [57].

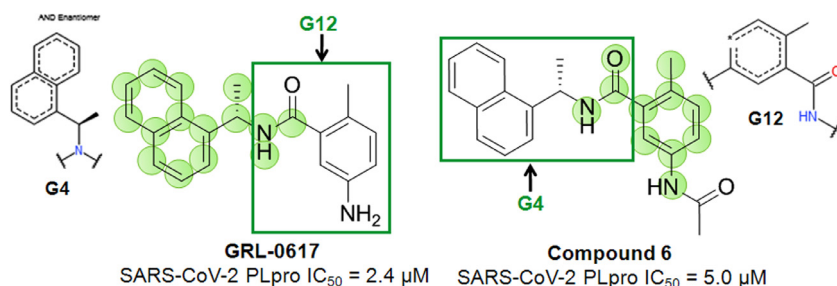
based analysis, Jacobs *et al* [57] found that the furan oxygen atom of compound **B018** forms hydrogen bonding interaction the backbone NH of G143. At the S1' site, the catalytic C145 resides beneath the furan oxygen atom at a short distance (Figure 8). Further, we have analyzed the ligand (**B018**)-receptor (SARS-CoV 3CLpro) interaction in DS [48] where a  $\pi$ -donor hydrogen bond is noticed between the furan ring and the catalytic C145 [57] (Figure S9).

In summary, SARpy analysis based structural fragments for SARS-CoV 3CLpro inhibitory activity suggested that heterocyclic rings such as pyridine, pyrimidine and indoline modulate the biological properties. Moreover, the results of Bayesian modelling study are closely associated with the SARpy analysis. Additionally, the Bayesian modelling study predicted the positive influence of furan ring in the SARS-CoV 3CLpro inhibitory property. Therefore, to achieve an attractive potency level, several heterocyclic rings controlled the 3CLpro inhibitory activity with respect to their polarity, size and hydrogen bonding capabilities.

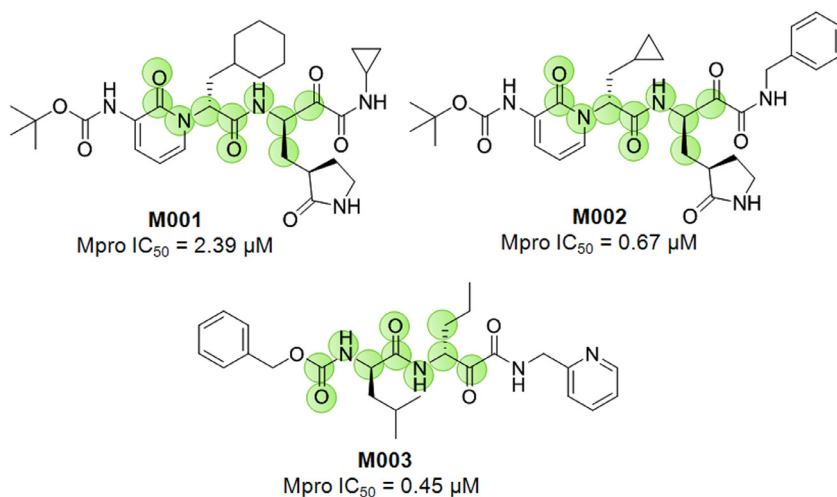
#### 4. Implications for COVID-19 Drug Discovery

The medicinal chemistry outlook of novel coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, is just few months old. Molecular modelling driven anti-viral study can generate several possibilities in drug discovery against COVID-19. The fragments studied and extensively interpreted in this current study surely leaving much room for follow-up investigations against SARS-CoV-2. Notably, the fragments identified by the SARpy analysis and Bayesian classification study may be an effective approach to accelerate drug design against SARS-CoV-2.

Since the binding site of PLpro and 3CLpro enzymes were conserved between SARS-CoV and SARS-CoV-2 the fragments will be very useful for COVID-19 drug discovery. Here, the identified fragments can also found in recently reported compounds (Figures 9 and 10) [7,9,17,54]. The current study will offer an idea to explore the important existing structural data and an in-depth qualification



**Fig. 9.** Structures of two active SARS-CoV-2 PLpro inhibitors highlighting important fragments. The activity values were taken from literature [54].



**Fig. 10.** Structures of some active SARS-CoV-2 3CLpro inhibitors highlighting important fragments. The activity values were taken from literatures [9,17].

of fragment hits. This will stimulate further research by providing valuable guidance to the medicinal chemists for designing of novel PLpro and 3CLpro inhibitors against previous SARS as well as recent COVID-19 diseases.

Meanwhile, looking the exact crystal binding position of an inhibitor and changing/replacing the different fragment(s) may offers the possibility of modifying the lead compounds, hence, their inhibitory potential against the proteases. Besides, potential protease inhibition activities are also accompanied by further increases in scaffold diversity. However, care should be taken to design effective protease inhibitors by keeping ADME properties in mind. Nevertheless, this study may offer a bigger direction in broad spectrum anti-viral drug design and discovery.

## 5. Conclusion

Here, classification-based QSAR models for diverse set of SARS-CoV PLpro and 3CLpro were developed and validated. All models were successfully used to hunt promising fragments. The fragments can be further utilised for the virtual screening of chemical libraries or FDA approved drugs to identify effective protease inhibitors.

The closely related fragment concepts are experiencing increasing interest in medicinal chemistry. In our analysis, we have refined the assessment of fragments by exploring structure activity relationships (SARs). The fragments encode an attractive knowledge base for compound design as well as utilized for lead optimization.

Since the size of the database of SARS-CoV protease inhibitor continues to grow, collectively our modeling studies emphasize that our approach could be used to aid in the process of lead optimization against the proteases. In diseases like COVID-19 with the lack of potential drug and large-scale inhibitors, it is quite difficult

to connect the dots to *quarere* significant therapeutics. Reciprocally, one can apply these identified fragments to increase the protease inhibition of weak hits resulting from QSAR/VS studies etc. against targeted proteases.

To summarize, we have developed robust and validated models which may be applicable to drug design and lead optimization, opening up opportunities to design small molecules targeting the coronavirus proteases.

## CRediT author statement

**Kalyan Ghosh:** Data curation, Methodology, Software, Investigation, Writing- Original draft preparation.

**Sk. Abdul Amin:** Conceptualization, Methodology, Software, Investigation, Visualization, Writing- Original draft preparation.

**Shovanlal Gayen:** Conceptualization, Writing- Reviewing and Editing.

**Tarun Jha:** Writing- Reviewing and Editing, Supervision.

## Declaration of Competing Interest

The authors have no conflict of interests.

## Acknowledgment

Sk. Abdul Amin is thankful to the Council of Scientific and Industrial Research (CSIR), New Delhi, India for providing financial assistance in form of a Fellowship [FILE NO.: 09/096(0967)/2019-EMR-I, Dated: 01-04-2019]. Tarun Jha is thankful for the financial support from RUSA 2.0 of UGC, New Delhi, India to Jadavpur University, Kolkata, India. 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.2021.130366.

## References

- [1] World Health Organization WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020, 2000 <https://www.who.int/dg/speeches/detail/who-director-general-s-openingremarks-at-the-media-briefing-on-covid-19> (Date of Access: Aug 18, 2020).
- [2] WHO data <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (Date of Access: March 30, 2021).
- [3] S.A. Amin, T. Jha, Fight against novel coronavirus: A perspective of medicinal chemists, *Eur. J. Med. Chem.* 201 (2020) 112559, doi:10.1016/j.ejmech.2020.112559.
- [4] N. Adhikari, S.K. Baidya, A. Saha, T. Jha, Structural insight into the viral 3Clike protease inhibitors: Comparative SAR/QSAR approaches, in: S.P. Gupta (Ed.), *Viral Proteases and their inhibitors*, Academic Press, USA, 2017, pp. 317–402. Chapter 11.
- [5] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: Genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (2020) 418–423.
- [6] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, W.L. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, *Science* 368 (2020) 779–782.
- [7] W. Dai, B. Zhang, H. Su, J. Li, Y. Zhao, X. Xie, Z. Jin, F. Liu, C. Li, Y. Li, F. Bai, H. Wang, X. Cheng, X. Cen, S. Hu, X. Yang, J. Wang, X. Liu, G. Xiao, H. Jiang, H. Liu, Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease, *Science* 368 (2020) 1331–1335.
- [8] B. Goyal, D. Goyal, Targeting the Dimerization of the Main Protease of Coronaviruses: A Potential Broad-Spectrum Therapeutic Strategy, *ACS Comb. Sci.* (2020), doi:10.1021/acscombsci.0c00058.
- [9] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, R. Hilgenfeld, Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors, *Science* 368 (2020) 409–412.
- [10] C.T. Keng, S. Åkerström, C.S. Leung, L.L. Poon, J.M. Peiris, A. Mirazimi, Y.J. Tan, SARS coronavirus 8b reduces viral replication by down-regulating E via an ubiquitin dependent proteasome pathway, *Microbes Infect* 13 (2011) 179–188.
- [11] V. Thiel, K.A. Ivanov, A. Putics, T. Hertzog, B. Schelle, S. Bayer, B. Weißbrich, E.J. Snijder, H. Rabenau, H.W. Doerr, A.E. Gorbalenya, Mechanisms and enzymes involved in SARS coronavirus genome expression, *J. Gen. Virol.* 84 (2003) 2305–2315.
- [12] A. Zumla, J.F. Chan, E.I. Azhar, D.S. Hui, K.Y. Yuen, Coronaviruses—drug discovery and therapeutic options, *Nat. Rev. Drug Discov.* 15 (2016) 327–347.
- [13] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 382 (2020) 727–733, doi:10.1056/NEJMoa2001017.
- [14] 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), doi:10.1080/07391102.2020.1780946.
- [15] S.T. Ngo, N. Quynh Anh Pham, L. Thi Le, D.H. Pham, V.V. Vu, Computational determination of potential inhibitors of SARS-CoV-2 main protease, *J. Chem. Inf. Model.* (2020), doi:10.1021/acs.jcim.0c00491.
- [16] J. Wang, Fast Identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study, *J. Chem. Inf. Model.* (2020), doi:10.1021/acs.jcim.0c00179.
- [17] C. Ma, M.D. Sacco, B. Hurst, J.A. Townsend, Y. Hu, T. Szeto, X. Zhang, B. Tarbet, M.T. Marty, Y. Chen, J. Wang, Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease, *BioRxiv* (2020), doi:10.1101/2020.04.20.051581.
- [18] I. Khanna, Drug discovery in pharmaceutical industry: productivity challenges and trends, *Drug Discov. Today* 17 (2012) 1088–1102.
- [19] C.W. Murray, D.C. Rees, The rise of fragment-based drug discovery, *Nat. Chem.* 1 (2009) 187–192.
- [20] D.A. Erlanson, S.W. Fesik, R.E. Hubbard, W. Jahnke, H. Jhoti, Twenty years on: the impact of fragments on drug discovery, *Nat. Rev. Drug Discov.* 15 (2016) 605–619.
- [21] H. Chen, X. Zhou, A. Wang, Y. Zheng, Y. Gao, J. Zhou, Evolutions in fragment-based drug design: The deconstruction–reconstruction approach, *Drug Discov. Today* 20 (2014) 105–113.
- [22] D.A. Erlanson, W. Jahnke, *Fragment-based Drug Discovery. Lessons and Outlook*, Wiley-VCH, Weinheim, Germany, 2016 ISBN 9783527337750.
- [23] M. Drag, G.S. Salvesen, Emerging principles in protease-based drug discovery, *Nat. Rev. Drug Discov.* 9 (2010) 690–701.
- [24] G.G. Ferenczy, G.M. Keserü, Thermodynamics of fragment binding, *J. Chem. Inf. Model.* 52 (2012) 1039–1045.
- [25] S.A. Amin, N. Adhikari, T. Jha, Diverse classes of HDAC8 inhibitors: in search of molecular fingerprints that regulate activity, *Future Med. Chem.* 10 (2018) 1589–1602, doi:10.4155/fmc-2018-0005.
- [26] T. Ferrari, G. Gini, N.G. Bakhtyari, E. Benfenati, in: Mining toxicity structural alerts from SMILES: A new way to derive Structure Activity Relationships, IEEE, 2000, pp. 120–127. In 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM) 2011 Apr 11.
- [27] A. Lombardo, F. Pizzo, E. Benfenati, A. Manganaro, T. Ferrari, G. Gini, A new in silico classification model for ready biodegradability, based on molecular fragments, *Chemosphere* 108 (2014) 10–16.
- [28] A. Golbamaki, E. Benfenati, N. Golbamaki, A. Manganaro, E. Merdivan, A. Roncaglioni, G. Gini, New clues on carcinogenicity-related substructures derived from mining two large datasets of chemical compounds, *J. Environ. Sci. Health C* 34 (2016) 97–113.
- [29] R. Gaikwad, S. Ghorai, S.A. Amin, N. Adhikari, T. Patel, K. Das, T. Jha, S. Gayen, Monte Carlo based modelling approach for designing and predicting cytotoxicity of 2-phenylindole derivatives against breast cancer cell line MCF7, *Toxicol. In Vitro* 52 (2018) 23–32.
- [30] 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).
- [31] T. Bobrowski, V.M. Alves, C.C. Melo-Filho, D. Korn, S. Auerbach, C. Schmitt, EN. Muratov, A. Tropsha, Computational models identify several FDA approved or 1 experimental drugs as putative agents against SARS-CoV-2, *ChemRxiv* (2020), doi:10.26434/chemrxiv.12153594.v1.
- [32] Y.M. Báez-Santos, S.J. Barraza, M.W. Wilson, M.P. Agius, A.M. Mielech, N.M. Davis, S.C. Baker, S.D. Larsen, A.D. Mesecar, X-ray structural and biological evaluation of a series of potent and highly selective inhibitors of human coronavirus papain-like proteases, *J. Med. Chem.* 57 (2014) 2393–2412.
- [33] K.W. Cheng, S.C. Cheng, W.Y. Chen, M.H. Lin, S.J. Chuang, I.H. Cheng, C.Y. Sun, C.Y. Chou, Thiopurine analogs and mycophenolic acid synergistically inhibit the papain-like protease of Middle East respiratory syndrome coronavirus, *Antiviral Res* 115 (2015) 9–16.
- [34] C.Y. Chou, C.H. Chein, Y.S. Han, M.T. Prebanda, H.P. Hsieh, B. Turk, G.G. Chang, X. Chen, Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus, *Biochem. Pharmacol.* 75 (2008) 1601–1609.
- [35] M. Frieman, D. Basu, K. Matthews, J. Taylor, G. Jones, R. Pickles, R. Baric, D.A. Engel, Yeast based small molecule screen for inhibitors of SARS-CoV, *PLoS One* 6 (2011) e28479.
- [36] A.K. Ghosh, J. Takayama, K.V. Rao, K. Ratia, R. Chaudhuri, D.C. Mulhearn, H. Lee, D.B. Nichols, S. Baliji, S.C. Baker, M.E. Johnson, Severe acute respiratory syndrome coronavirus papain-like novel protease inhibitors: design, synthesis, protein–ligand X-ray structure and biological evaluation, *J. Med. Chem.* 53 (2010) 4968–4979.
- [37] A.K. Ghosh, J. Takayama, Y. Aubin, K. Ratia, R. Chaudhuri, Y. Baez, K. Sleeman, M. Coughlin, D.B. Nichols, D.C. Mulhearn, B.S. Prabhakar, Structure-Based Design, Synthesis, and Biological Evaluation of a Series of Novel and Reversible Inhibitors for the Severe Acute Respiratory Syndrome— Coronavirus Papain-Like Protease, *J. Med. Chem.* 52 (2009) 5228–5240.
- [38] J.Y. Park, J.H. Kim, Y.M. Kim, H.J. Jeong, D.W. Kim, K.H. Park, H.J. Kwon, S.J. Park, W.S. Lee, Y.B. Ryu, Tanshinones as selective and slow-binding inhibitors for SARS-CoV cysteine proteases, *Bioorg. Med. Chem.* 20 (2012) 5928–5935.
- [39] K. Ratia, K.S. Saikatendu, B.D. Santarsiero, N. Barretto, S.C. Baker, R.C. Stevens, A.D. Mesecar, Severe acute respiratory syndrome coronavirus papain-like protease: structure of a viral deubiquitinating enzyme, *Proc. Nat. Acad. Sci* 103 (2006) 5717–5722.
- [40] [https://chemrxiv.org/articles/Computational\\_Models\\_Identify\\_Several\\_FDA\\_Approved\\_or\\_Experimental\\_Drugs\\_as\\_Putative\\_Agents\\_Against\\_SARS-CoV-2/12153594/1](https://chemrxiv.org/articles/Computational_Models_Identify_Several_FDA_Approved_or_Experimental_Drugs_as_Putative_Agents_Against_SARS-CoV-2/12153594/1)
- [41] A.P. Toropova, A.A. Toropov, E. Benfenati, A quasi-QSPR modelling for the photo catalytic decolorization rate constants and cellular viability (CV%) of nanoparticles by CORAL, *SAR QSAR Environ. Res.* 26 (2015) 29–40.
- [42] 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.
- [43] M. Van Bossuyt, G. Raitano, M. Honma, E. Van Hoeck, T. Vanhaecke, V. Rogiers, B. Mertens, E. Benfenati, New QSAR models to predict chromosome damaging potential based on the in vivo micronucleus test, *Toxicol.Lett* (2020).
- [44] T. Ferrari, D. Cattaneo, G. Gini, N. Golbamaki Bakhtyari, A. Manganaro, E. Benfenati, Automatic knowledge extraction from chemical structures: the case of mutagenicity prediction, *SAR QSAR Environ. Res.* 24 (2013) 365–383.
- [45] H. Yang, J. Li, Z. Wu, W. Li, G. Liu, Y. Tang, Evaluation of different methods for identification of structural alerts using chemical ames mutagenicity data set as a benchmark, *Chem. Res. Toxicol.* 30 (2017) 1355–1364.
- [46] D. Baderna, D. Gadaleta, E. Lostaglio, G. Selvestrel, G. Raitano, A. Golbamaki, A. Lombardo, E. Benfenati, New in silico models to predict in vitro micronucleus induction as marker of genotoxicity, *J. Hazard. Mater.* 385 (2020) 121638.
- [47] J. Hemmerich, F. Troger, B.; F. Füzi, G. Ecker, Using machine learning methods and structural alerts for prediction of mitochondrial toxicity, *Mol. Inform.* 39 (2020) 2000005.
- [48] D.S. Biovia, *Discovery Studio* 2016, 2000 San Diego, CA, USA.
- [49] R. David, H. Mathew, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* 50 (2010) 742–754.
- [50] T. Fawcett, An introduction to ROC analysis, *Pattern Recog. Lett.* 27 (2006) 861–874.

- [51] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, A. Tropsha, Does rational selection of training and test sets improve the outcome of QSAR modeling? *J. Chem. Inf. Model.* 52 (2012) 2570–2578.
- [52] A. Golbraikh, M. Shen, Z. Xiao, Y.D. Xiao, K.H. Lee, A. Tropsha, Rational selection of training and test sets for the development of validated QSAR models, *J. Comput. Aided Mol. Des.* 17 (2003) 241–253.
- [53] [http://teqip.jdvu.ac.in/QSAR\\_Tools/](http://teqip.jdvu.ac.in/QSAR_Tools/) (as accessed on 17th Feb, 2021)
- [54] B.T. Freitas, I.A. Durie, J. Murray, J.E. Longo, H.C. Miller, D. Crich, R.J. Hogan, R.A. Tripp, S.D. Pegan, Characterization and Noncovalent Inhibition of the Deubiquitinase and deISGylase Activity of SARS-CoV-2 Papain-Like Protease, *ACS Infect. Dis.* (2020), doi:10.1021/acsinfecdis.0c00168.
- [55] , PyMOL is a molecular visualization system on an open-source foundation, 2000.
- [56] K. Ghosh, S.A. Amin, S. Gayen, T. Jha, Chemical-informatics approach to COVID-19 drug discovery: Exploration of important fragments and data mining based prediction of some hits from natural origins as main protease (Mpro) inhibitors, *J. Mol. Struct.* 1224 (2020) 129026.
- [57] J. Jacobs, V. Grum-Tokars, Y. Zhou, M. Turlington, S.A. Saldanha, P. Chase, A. Eggler, E.S. Dawson, Y.M. Baez-Santos, S. Tomar, A.M. Mielech, Discovery, synthesis, and structure-based optimization of a series of N-(tert-butyl)-2-(N-arylamido)-2-(pyridin-3-yl) acetamides (ML188) as potent noncovalent small molecule inhibitors of the severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL protease, *J. Med. Chem.* 56 (2013) 534–546.