DRUG DISCOVERY

*15(36), 2021*

**To Cite:**
Ambrose GO, Enya J, AbelJack-Soala T, Fabunmi BT, Temidayo AK, Olusola BO. Lipophilic Efficiency as an Important Metric in the Design of SARS coronavirus 3C-like proteinase (3CL-pro) Inhibitors: Guidepost towards Lead Selection and Optimization in the Treatment of COVID-19. *Drug Discovery*, 2021, 15(36), 131-148

**Author Affiliation:**
[1]Department of Biochemistry, Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria
[2]Department of Microbiology, Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria
[3]Department of Anatomy, Basic Medical Sciences, University of Ilorin, Ilorin, Nigeria
[4]Biochemistry Unit, Department of Science Laboratory Technology, Federal Polytechnic Ilaro, Ogun State, Nigeria
[5]Biological Sciences Department, Achievers University Owo, P.M.B. 1030 Ondo State, Nigeria

✉**Corresponding author:**
Department of Biochemistry, Faculty of Life Sciences, University of Ilorin, Ilorin, Nigeria
E-mail: ocheab1@gmail.com

**Peer-Review History**
Received: 27 April 2021
Reviewed & Revised: 31/April/2021 to 1/July/2021
Accepted: 03 July 2021
Published: July 2021

**Peer-review**
External peer-review was done through double-blind method.



DISCOVERY
SCIENTIFIC SOCIETY

# Lipophilic Efficiency as an Important Metric in the Design of SARS coronavirus 3C-like proteinase (3CL-pro) Inhibitors: Guidepost towards Lead Selection and Optimization in the Treatment of COVID-19

**George Oche Ambrose[1]✉, Joseph Enya[3], Tamonokorite AbelJack-Soala[1], Bukola Temitope Fabunmi[5], Ajayi Kolawole Temidayo[2], Bankole Olukayode Olusola[1,4]**

## ABSTRACT

Pfizer researchers reported in 2018 that lipophilic efficiency (LipE) is an important metric that is increasingly being applied to medicinal chemistry drug discovery programs. In this perspective, drug discovery examples have been strictly applied when adopting LipE to guide medicinal chemistry lead optimization toward candidate drugs with exceptional efficacy and safety *in vivo* potential, especially when guided by optimization based on physicochemical properties. In general, most medicinal chemists only consider potency and try to increase it during hits and lead optimization or when studying the structure-activity relationship. It should be noted that lipophilicity should be considered in conjunction with potency variations to ensure both the safety (drug-likeness) and the efficacy of the candidate drug. Therefore, the aim of this study is to identify successful potential leads against 3CL-pro and optimize them for maximum potency and safety in COVID-19 treatment with a design strategy approach. 3CL-pro inhibitors with lipophilic efficacy and related bioactivity data were obtained from the Chembl database and analyzed based on relationship between LipE and logP (lipophilic). The 2D physicochemical descriptors of the compounds were calculated. Quantitative Structural-Activity Relationships (QSAR) model was built and bioactivities of novel compounds were predicted while molecular mechanism was inferred by docking assay. Based on analysis, 80 novel compounds were found, 6 of the novel compounds (36, 37, 46, 47, 77 and 79) revealed an increase in both LipE and potency with logP decrease, which makes them better alternatives to existing 3CL-pro inhibitors in the treatment of COVID-19.



ISSN 2278–540X EISSN 2278–5396 | OPEN ACCESS

**Keywords:** COVID-19, Lipophilic, QSAR model, 3CL-pro inhibitors, drug

## 1. INTRODUCTION

From a global public health and socioeconomic perspective, there is an urgent expectation to come-up with a rapid intervention that includes effective vaccines and antiviral drugs to stop the spread of the current pandemic virus [1] designated as coronavirus disease 2019 (COVID-19), broadcasted by the World Health Organization a pandemic [2, 3]. The infection inhibits the liver, respiratory and central nervous systems, the digestive system of humans and animals. Thus the resent name, SARS-CoV-2 agreed upon by scientist due to 82% similarity with the known Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV) [4-6]. The SARS-CoV-2 principle protease (Mpro), also identified as chymotrypsin-like (3CL) protease (3CLpro), is a non-structural protein playing crucial role through cleavage of the virus-encoded polyproteins for replication and maturation (Coro 21). By so doing, it is rendered as an alluring quarry for the upshot of efficacious antiviral drugs against SARS-CoV-2 virus [7, 8].

The 3CLpro homodimer enzyme (EC: 3.4.22.69; optimal activity at pH 7.5 and 42 ° C) is largely maintained amid members of Coronaviridae with approximately 40-44% sequence homology. It has three structural segments consisting of domain I (residue 8 - 101) and domain II (residue 102 - 184), with the pair having beta-barrel motifs representing the catalytic region of chymotrypsin, as well as domain III (residue 185 - 200), having helical structure participating in the protein dimerization and enzyme activation [9-13]. A valid approach necessary to design potent drugs for SARS-CoV-2 infection requires the interaction mechanism of protein-ligand. These protein-ligands exhibit an important role in structure-based drug design, while enhancing steric compatibility of drug agents is an effective strategy for generating energy-efficient binding of drug agents to target receptors [14].

Leeson and Springthrope [16] initially proposed ligand lipophilicity efficiency (LLE or LipE) [15]. And various studies conducted on this ligand have proofed it to be an effective and direct means of evaluating the quality of research compounds. Safety of the compounds is accessed by relating lipophilicity and potency. LipE tries to maximize acceptable minimum lipophilicity per unit *in vitro* or in basic terms, to enhance the activity while sustaining low lipophilicity [16]. LipE is estimated as logP (or log *D*) minus logarithm of ligand potency (pKd, pKi or pEC50) [17]. Practically, instead of measuring log P, the computed value of clog P is commonly adopted together with the most pertinent *in vitro* strength to predict *in vivo* efficacy [18]. LipE can be adopted to recognize small sized molecules which would otherwise be overlooked due to its low potency and lipophilic value. This is advantageous in view of the fact that size and lipophilicity are largely increased following lead optimization. Thus using small quantity of LipE, makes less lipophilitic compounds a beneficial starting point [17, 19, 20].

In recent time, LipE has gained popularity as a direct and important method to modulate lipophilicity, with examples to show for its use in drug optimization. Some notable examples worth mentioning include; CB2 agonists [24], CB2 agonists/CB1 agonists [25], soluble epoxide hydrolase inhibitors [35], twofold PI3K/mTOR inhibitors [27], HIV non-nucleoside reverse transcriptase inhibitors [28], ATP-competitive Akt inhibitors [26], design of a potent cyclin-dependent kinase 2 (CDK2) inhibitor [21, 22] and protein kinase B inhibitor [23] (see [29, 30] for review). Hence this study aims to select successful leads as starting points towards optimization to design more potent and drug-like clinical candidates as inhibitors of 3CL-pro.

## 2. MATERIALS AND METHODS

### 2.1. Collection of data and dataset groundwork

The trivial name, source, lipophilic efficiency metrics (LipE), experimental lipophilicity (AlogP) and biological activities ($IC_{50}$) of 3CLpro inhibitors were obtained from the ChEMBL database. A total of 15 coronavirus 3C-like proteinase Inhibitors were retrieved on the basis of avail chemical structures with corresponding bioactivities ($IC_{50}$) (Figure 1). Bioactivities ($IC_{50}$) were subsequently adjusted to $pIC_{50}$ adopting the expression $pIC_{50}$ = (-log ($IC_{50}$ X)). 3CLpro inhibitors chemical structures were downloaded from the ChEMBL database as smiles which were then changed to (2D) SDF format on DataWarrior v5.0.0. [31].

### 2.2. Analysis of Lipophilic efficiency

The lipophilic efficiency of retrieved dataset in relation to their corresponding potency ($pIC_{50}$) and lipophilicity (logP) was carried out using DataWarrior v5.0.0



ISSN 2278–540X EISSN 2278–5396 | OPEN ACCESS

**2.3. 2D QSAR study**

The potency of a compound must be quantified by molecular descriptors in order to build a QSAR model [32] and so based on this, CDK descriptor version 1.0 was adopted for the computation of varied descriptors in the following groups: Hybrid features, Constitutional properties, Topological properties, Electronic properties and Geometric descriptors. The computed properties were organized in a matrix format. These computed properties were preprocessed to determine the correlation coefficient cut-off of 0.99 based on a variance cut-off of 0.0001 and take-away invariables (constant column) by using JFrameVWSP version 1.0. The dataset of 28 molecules retrieved from literature [33] was separated into the test and training dataset by adopting Kennard-Stone method [34] QSAR model validation is essential to assess how reliable a developed model is [35]. This is usually achieved by evaluating the internal stability and predictive ability of the QSAR models. In this study, the QSAR model developed was authenticated using leave-one-out (LOO) method to achieve internal validation. This was performed by removing a molecule, creating and authenticating the model against the individual molecule for all the $Q^2$ ($rCV^2$) values and documented. Equation (1) was used to calculate $rCV^2$ (cross-validation regression coefficient) to describe the internal stability of the model.

$$rCV^2 = 1 - \frac{\sum\left(Y_{obs} - Y_{pred}\right)^2}{\sum\left(Y_{obs} - \bar{Y}\right)^2} \quad (1)$$

Where Y in the stated equation represents the training dataset average activity value. $Y_{pred}$ and $Y_{obs}$ stands for the predicted and observed activity values correspondingly. It is worthy of note that, rCV greater than 0.5 proposes a realistically robust model [36]. Sequel to the process of internal validation, QSAR model's high predictive power was projected from an external test set of compounds not applied in building the QSAR model. The predictive capacity or external validation obtained was determined by predictive $R^2$ ($R_{pred}^2$) based on equation (2).

$$r^2_{prod} = 1 - \frac{\sum\left(Y_{pred(test)} - Y_{(test)}\right)^2}{\sum\left(Y_{(test)} - \overline{Y_{(training)}}\right)^2} \quad (2)$$

$Y_{(test)}$ and $Y_{pred(test)}$ is the observed activity and predicted values for test set compounds respectively.
$Y_{(training)}$ is the average bioactivity of compound in the training set.
QSAR model ($R_{pred}^2$) greater than 0.6 is the acceptable predictive power for the test set molecules [37-39].

MLR method was used to develop QSAR model from the dataset to exam potential leads against 3CLpro inside a training dataset (28 compounds). CDK algorithm was used to calculate the total molecular descriptor (108) for individual compound.

**2.4. Molecular docking**

The three-dimension crystal structure of (PDB: 1uj1) from the protein databank was retrieved to prepare the target SARS coronavirus 3C-like proteinase (3CLpro) [40]. Discovery studio 2017R2 was employed to remove all heteroatoms while Pymol tool for non-essential water molecules. Subsequent to receptor and ligands preparation, this study utilized PyRx, AutoDock Vina option based on scoring functions to perform molecular docking analysis. PyRx, AutoDock Vina exhaustive search docking function was used for the analysis. Upon successful minimization process, the resolution of the grid box was centered at 76.0065 ×-11.5107 × 18.0445 on the x, y and z axes correspondingly at grid dimension 25x 25 x 25 Å to specify the binding site (Figure 14b).



ISSN 2278–540X EISSN 2278–5396 | OPEN ACCESS

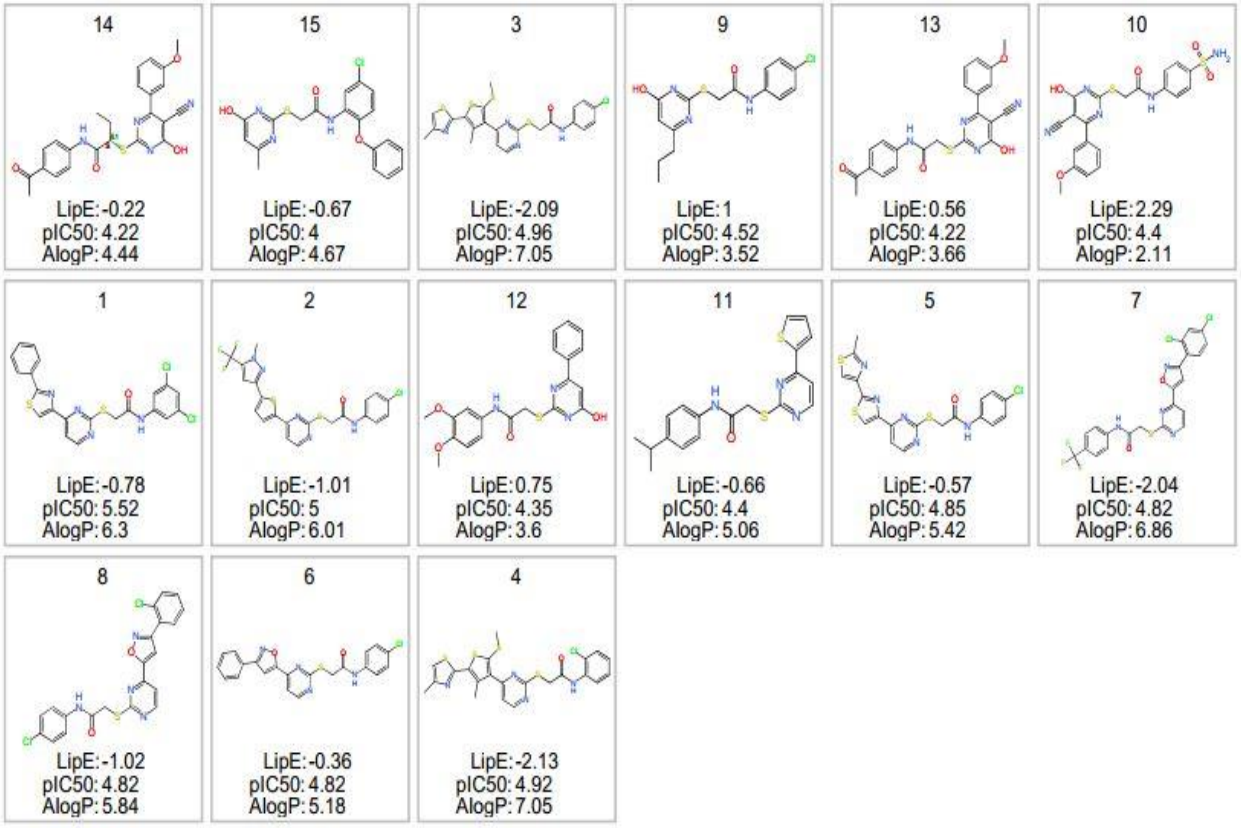


**Figure 1**: Raw data with corresponding bioactivities ($IC_{50}$)

## 3. RESULTS AND DISCUSSION

Figure 2 shows the relationship between potency ($pIC_{50}$), lipophilicity (logP) and lipophilic efficiency (LipE) of 3CL-pro inhibitors, which is useful in lead selection and optimization as discussed in the section below:

### 3.1. Iso-LipE 3CLpro inhibitors

Multiple combinations of $pIC_{50}$ and logP can lead to the same LipE. On this account, having the knowledge of LipE alone for molecule is not informative [50]. It is necessary to interpret 3CLpro inhibitors in the LipE plot format, to give insight into required procedure for optimization-oriented designs. Based on our analysis, compound 2 had a $pIC_{50}$ value of 5 with a measured logP of 6.01, resulting in a LipE of -1.01 (Figure 3). Although compound 8 is slightly less potent with a $pIC_{50}$ value of 4.82, it had a logP of 5.84 which resulted in a comparable lipophilic efficiency of -1.02.

If we consider only the potency, compound 2 ought to appear superior to compound 8. These two compounds are comparably attractive and isoefficient for follow-up. In addition, taking into consideration that both compounds 2 and 8 have logP values of 6.01 and 5.84 respectively, which exceed the required optimum for ADME (Absorption, Distribution, Metabolism, and Excretion) properties (logP: ≤ 5), none is considered a better starting place than the other. Also, compound 11 had a $pIC_{50}$value of 4.4 with a measured logP of 5.06, resulting in a lipophilic efficiency of -0.66 while compound 15, with similar potency of $pIC_{50}$ valueof 4, had a logP of 4.67 resulting in a similar lipophilic efficiency of -0.67 (Figure 3). Given that compound 11 had a logP of 5.06 which exceed the required optimum for ADME properties (logP: ≤ 5), compound 15 is a better starting place. In fact, both compounds are isopotent i.e. they have similar $pIC_{50}$ values. This highlights the essence of analyzing LipE changes in the context of logP for selecting a lead compounds [41].



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

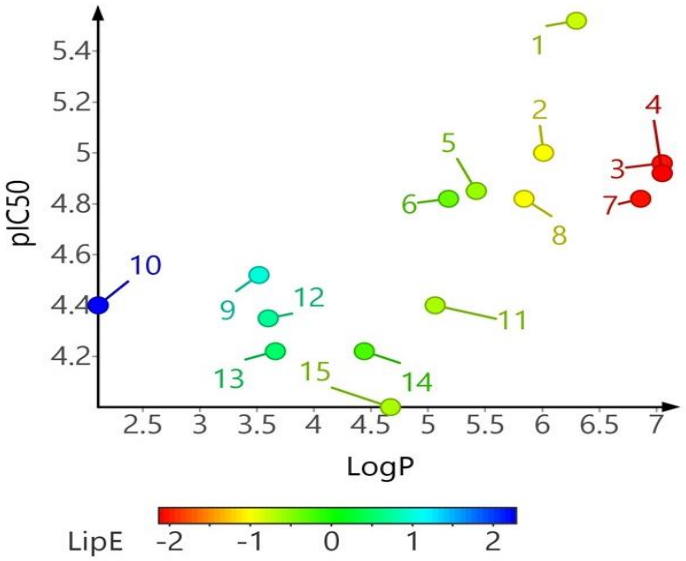


**Figure 2**: Relationship between $pIC_{50}$ and logP of 3CL-pro inhibitors in relation to their respective LipE

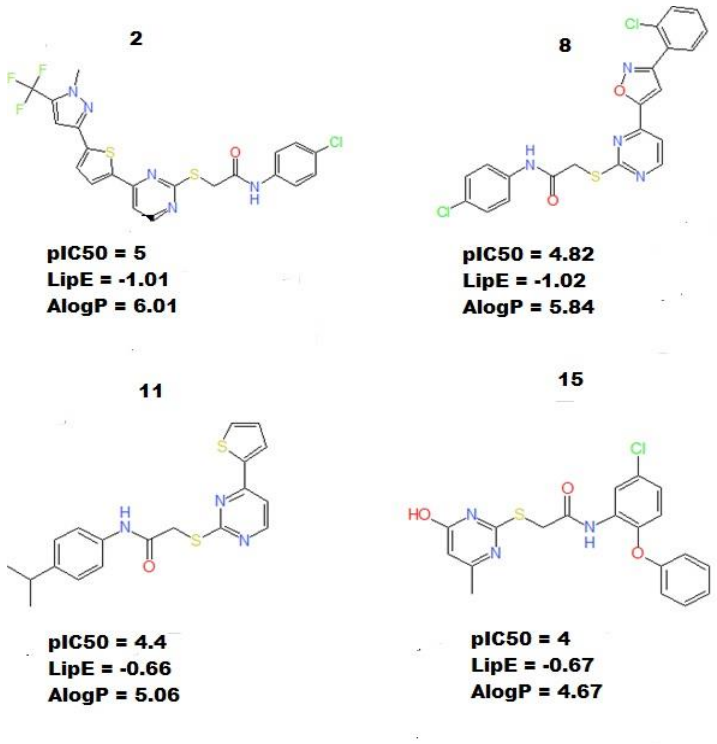


**Figure 3:** Isoefficient 3CL-pro inhibitors

**3.2. Iso-potent Efficiency Changes of 3CLpro inhibitors**

Unlike compound 2 and 8; 11 and 15, the compounds with comparable $pIC_{50}$ can have different LipE. Not realizing logP would make then look alike but then again, the compounds with lower logP will possess a higher lipophilic efficiency [42]. This is an isopotent efficiency change. It is illustrated by the potency of compounds 6 and 8; 13 and 14, in which reduction in LipE are seen while $pIC_{50}$ is sustained, which lead to an increase in LipE (Figure 4).Compounds 6 and 8; 13 and 14 are same in terms of $pIC_{50}$, irrespective of logP ranging from 3 to 6. Of all these compounds, 13 have appreciably high LipE when compared with the other compounds, thus presenting it a valuable lead which would have been lost in the subset of equipotent compounds without a LipE-based analysis.



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

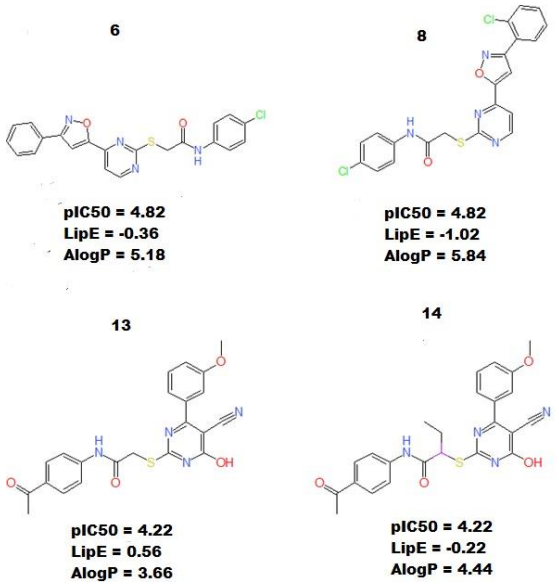


**Figure 4:** Isopotent 3CL-pro inhibitors

### 3.3. Isolipophilic LipE Changes of 3CL-pro inhibitors

Invariant logP in relation with $pIC_{50}$ changes throughout a set of derivatives resulting in isolipophilic modifications of LipE. Generally, isolipophilic variation in efficacy occurs with twosome of enantiomers, even though they are not limited to such examples [42]. This is observed in compound 12 and 13 in which an increase in LipE are observed while lipophilicity is maintained. This resulted in an insignificant increase in potency (Figure 5). Compound 12 and 13 are same in terms of their logP values, despite $pIC_{50}$ values of 4.35 and 4.22 respectively. 12 possessed slightly higher LipE, making it an indispensable lead. Because logP remain the same, assessment of isolipophilic LipE variations cannot be distinguished from a potency-centric analysis.

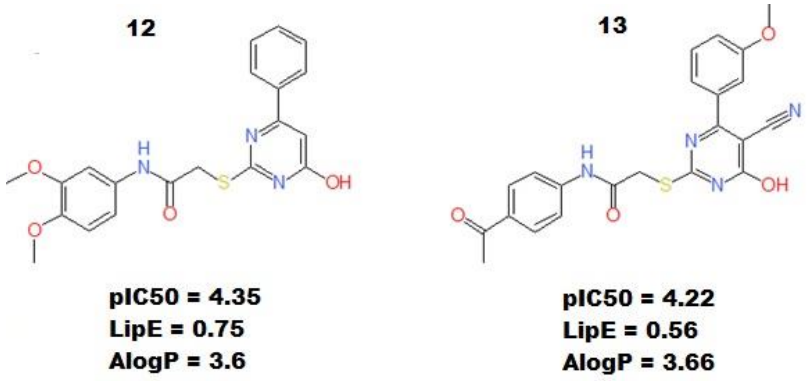


**Figure 5**: Isolipophilic 3CL-pro Inhibitors

At the same time, decreasing logP and improving potency results in a very large increase in lipophilic efficiency [42-44]. This was achieved by generating novel compounds based on structural modification of the starting chemical entity. Therefore, in a search for a starting chemical specie from our dataset, Compound 1 (with the lowest logP value) and 10 (with the highest potency) reveals an efficiency increase (3-fold) of -0.78 and 2.29 respectively with potency and logP decrease (figure 6). Given that compound 1 had a logP of 6.03 which is over what is optimal for ADME properties (logp: ≤ 5), compound 10 with logP value of 2.11 is a better starting chemical entity.



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

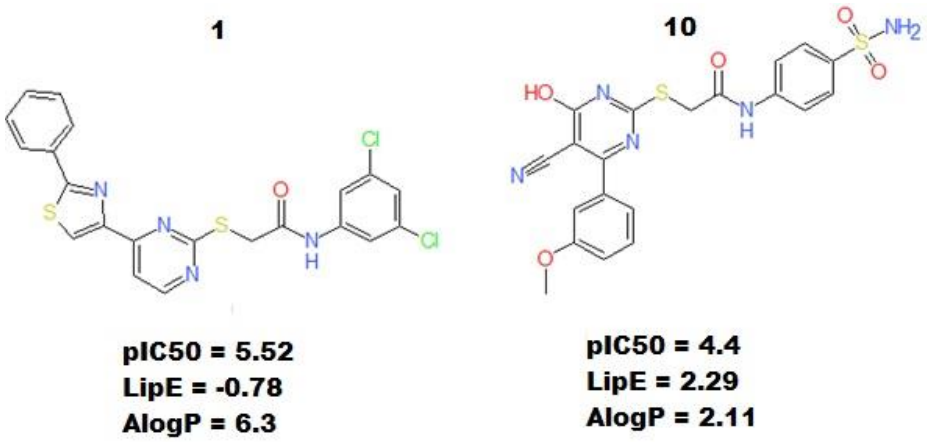


**Figure 6**: 10-1 showing an efficiency Increase decrease in potency and logP

Quantitative structural-activity relationship analysis (QSAR) is a successful method that has been adopted in reasonable drug design and in understanding drug action mechanism. Bioactivities of compounds are rated as a function of various physicochemical properties in QSAR studies. This makes clear how the variation of biological activity is based on alteration in the chemical structures [45]. These physicochemical properties are quantified in the form of descriptors. Given this, we designed a set of new compounds from compound 10 with a *de novo* design approach using DataWarrior v5.0.0. DataWarrior adopted an evolutionary method that mimics nature by randomly transforming known molecular configurations having small changes to establish novel generations with possibly better structures. Each generation of molecules are tested for robustness with a set of modifiable principles and the most auspicious structures serve as a starting point for subsequent generation. The mutation algorithm executes vicissitudes such as bond order changes, ring aromatisations, replacing an atom, atom insertions, substituent migrations just to mention a few.

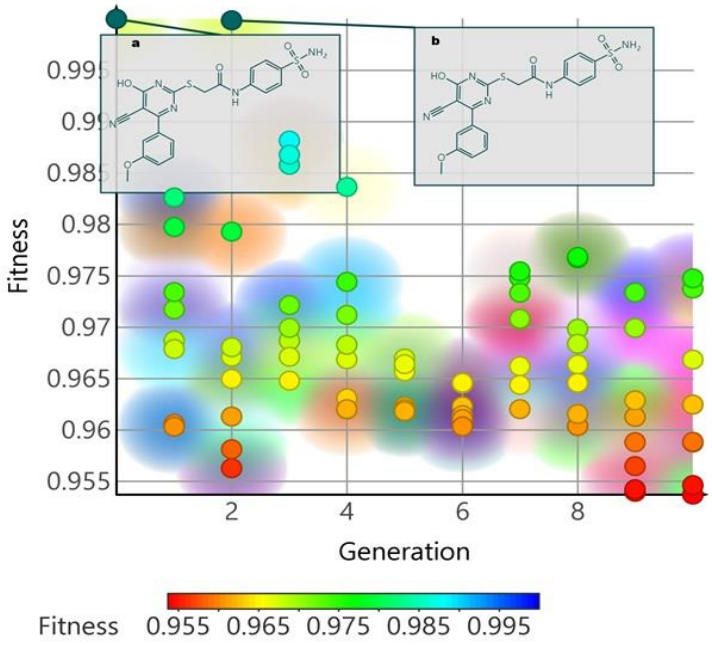


**Figure 7**: Novel compounds generated based on the structure of the parent molecule (compound 10) (a = Compound 10; b = Compound 17)

Every structure to be mutated is firstly evaluated for all possible mutations concerning how extensively the alteration will increase or decrease the drug-likeness. Mutations with alteration in the required direction are assigned a higher probability than mutations that reduce drug-likeness. Mutations that would create high ring tension are eliminated from the list. Based on this



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

approach, 81 structures were created that retain the scaffold of compound 10 with the corresponding fitness scores (Figure 7). Compound 17 of the second generation holds the highest fitness scores (0.98817), which makes it closest to the original structure (compound 10, fitness score = 1,000).

Tsai *et al*., (2006) [33] discovered 28 novel family of SARS-CoV Protease Inhibitors with their corresponding $IC_{50}$ values and shared the same scaffold as compound 10 (Figure 8). In order to predict the bioactivities of the newly generated datasets, a QSAR model was built from the 28 datasets having the same scaffold as the novel compounds.

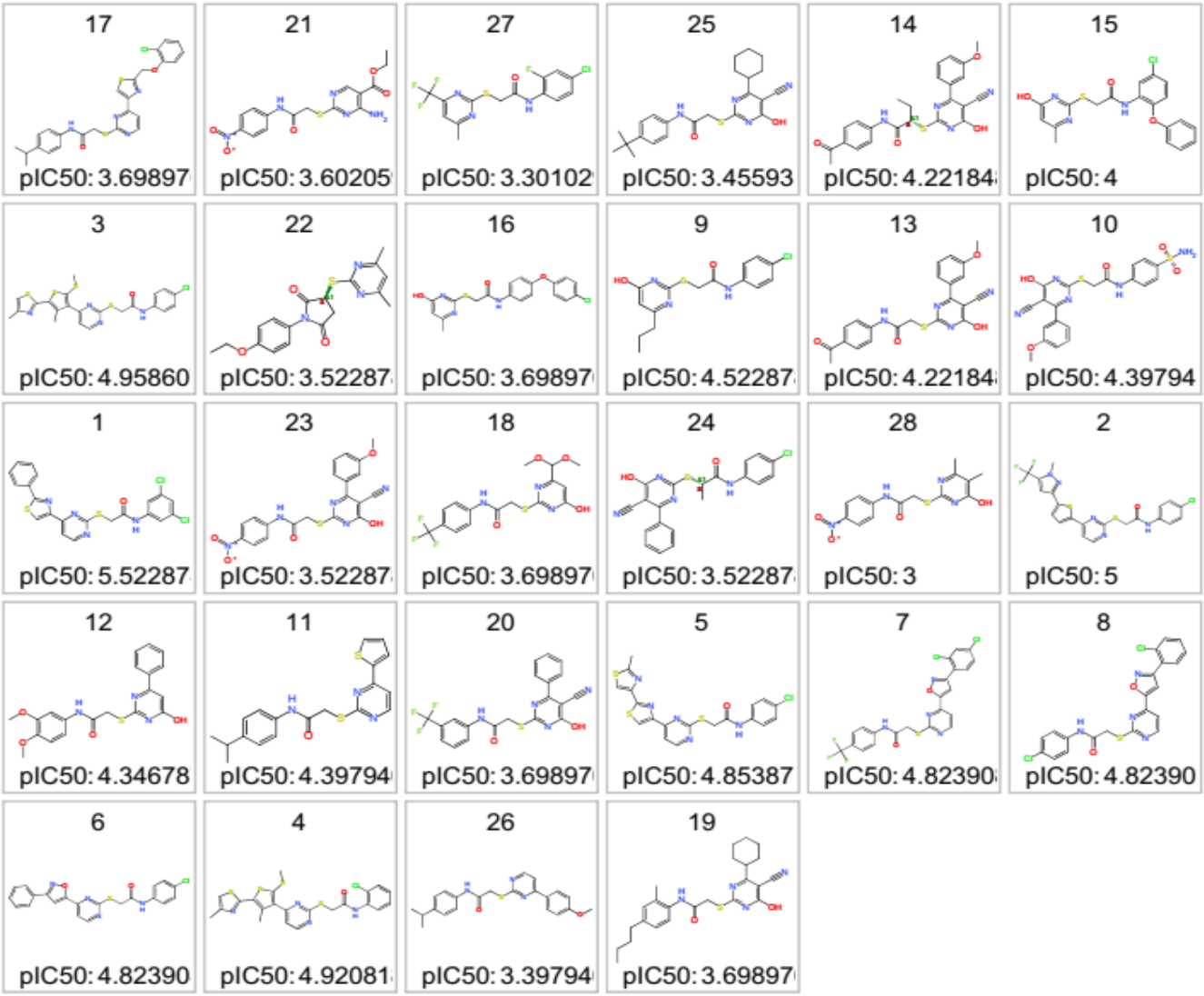


**Figure 8**: Retrieved chemical structures with same scaffold as compound 10

To analyze the multiplicity of the testing and training set, the Principal Component Analysis approach (PCA) was adopted and the PCA was executed with structural descriptors evaluated for the entire data set. This approach helped to identify homogeneities in the total data, as well as to describe the spatial location of the samples to help in dividing the data into train and test sets.

The PCA result revealed three main components (PC1, PC2 and PC3), elucidated as 99.998% of the entire variables which are as follows: PC1 = 41.975%, PC2 = 33.541% and PC3 = 24.482%. Given that the first three principal components can account for most of the variables, the different score plot is a dependable exemplification of the spatial allotment of points for the whole data. The compounds distribution over the initial three principal components space is shown by the plot of PC1, PC2 and PC3 (Figure 9) with PC1 and PC2 covering the largest variability in the total data (Figure 10). This number revealed that the samples of the test and training sets appear to be uniformly strewed in the three-dimensional space and consequently, it was possible to divide the dataset. In addition, the compounds in the training sets are represented in the whole dataset.



ISSN 2278–540X EISSN 2278–5396 | OPEN ACCESS

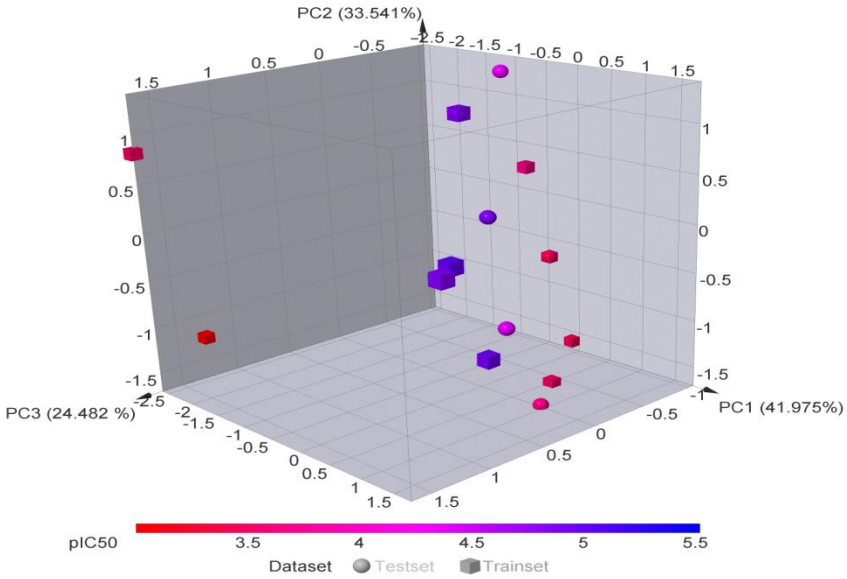


**Figure 9**: Analysis of the primary component of the test and train sets

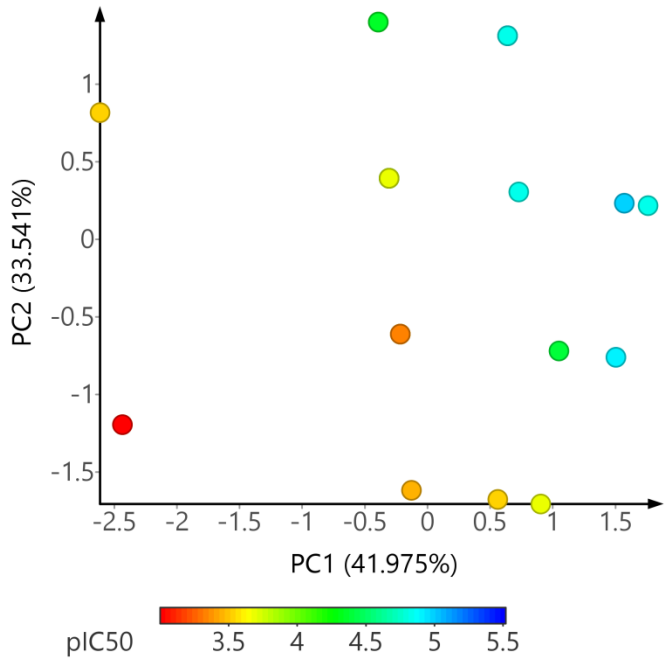


**Figure 10**: Analysis of principle component with PC1 and PC2

The next step after analyzing the separation of the dataset into the test and training set was to identify and choose the major factors that are most important for the SARS-CoV protease inhibition activity of the 28 new inhibitors. Genetic algorithm (GA) was applied as the method for selecting variables to select only the most important (relevant) combinations to obtain the model with the utmost predictive power by employing training dataset. The three (3) most important descriptors according to the GA approach, are SCH-5, C1SP3 and khs.ddsN based on the variance cut-off of 0.01 and inter correlation cut-off of 0.9. In this study, these descriptors have been shown to be related to the studied biological response and are described in Table 1.

In general, the quality of a model is described by its predictability in a QSAR study. The techniques adopted in this study were performed to correlate physicochemical descriptors of 28 3CL-pro inhibitors from their inhibitory training set. Physicochemical descriptors were taken as individual variables and the inhibitory activity of the 3CL-pro target was taken as dependent variables.

The data set of 28 compounds was divided into training set of 19 compounds in order to create and test the model. The datasets were subsequently utilized to construct the model and a test set of 9 compounds, which were used to test the structured model. The resulting linear equation based on the MLR is as follows (equation 3):

$$pIC_{50} = 4.52085(\pm 0.2233) - 0.2688(\pm 0.08067)\ C1SP3 + 4.92257(\pm 0.79286)\ SCH-5 - 0.51868(\pm 0.18816)\ khs.ddsN \quad (3)$$

$R^2$ (regression coefficient) = 0.82776,
SEE (standard error of estimate) = 0.28645,
$Q^2$ (cross-validation regression coefficient) =0.6468,
SDEP (standard deviation of error of prediction) = 0.3829
$r^2$ (LOO) ((Leave out one cross validation) = 0.7077
PRESS (predicted residual sum of squares) = 1.23081
F (variance ratio) = 24.02849 (DF: 3, 15), $r^2_{pred}$ (external predictive power) = 0.80913

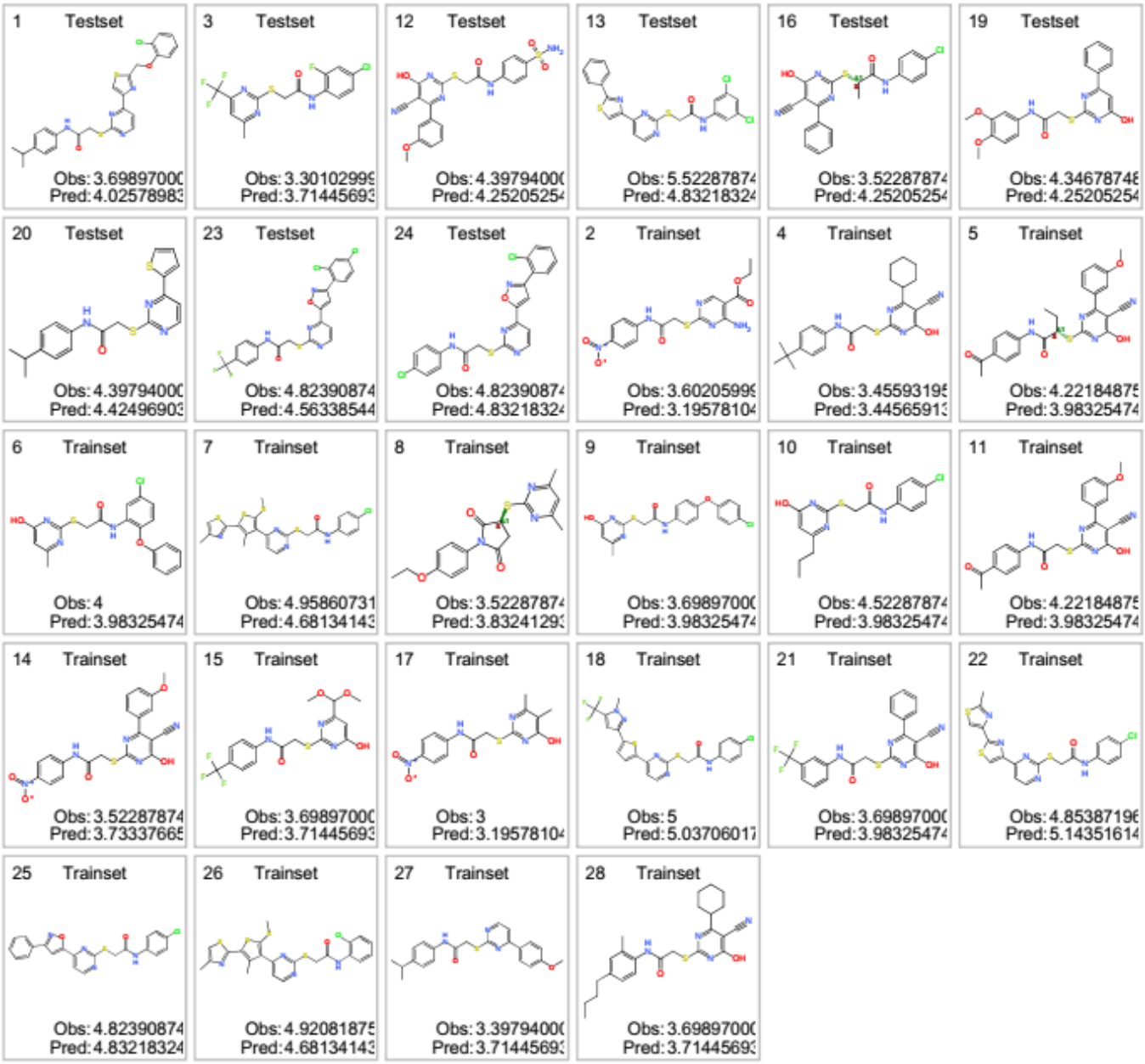


**Figure 11**: Experimental and Predicted bioactivities of the Train and Test set

Equation (5) suggests that the model established with GA-MLR presented an excellent tetragonal correlation coefficient ($R^2$) value with both internal and external predictive power ($r^2pred$) having excellent values. The developed QSAR model derived by GA-MLR showed a noteworthy connection between dependent variable ($pIC_{50}$ values) and the carefully chosen descriptors (independent variables).

An 82.8% correlation exists among the activity and selected descriptors in the training dataset, determined based on value of the regression ($R^2$=0.82776). Meanwhile, the value of the cross-validation regression coefficient ($Q^2$ = 0.6468) put forward ~64.7% prediction exactitude of this QSAR model.

Figure 11 shows the predicted and observed biological activities of the training and test datasets. Figure 12 is the predicted $pIC_{50}$ plot versus the experimental $pIC_{50}$ which revealed that there is a good agreement between the predicted and experimental activity values.

**Table 1**: GA selected descriptors with their corresponding description

| S/N | Descriptors | Description | Contribution |
|---|---|---|---|
| 1 | C1SP3 | Singly bound carbon bound to one other carbon | Negative Contribution |
| 2 | SCH-5 | Simple chain, order 5 | Positive Contribution |
| 3 | khs.ddsN | Description not available in the database | Negative Contribution |

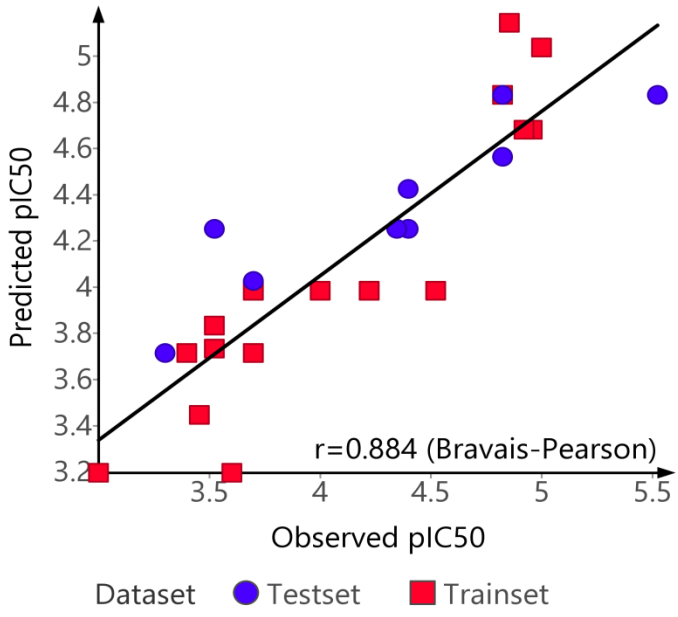


**Figure 12:** The predicted $pIC_{50}$ values using MLR modeling against the experimental (observed) $pIC_{50}$ values

The three descriptors for the best model (GA-MLR) used for PC1–PC2 loadings plot is shown in Figure 13. For the loadings, it was affirmed that the compounds with greater biological activity values, situated on the right side presented a larger influence on the SCH-5 descriptor located on the same side as in Figure 8. Conversely, compounds with lesser biological activity values on the left side have extra distinct contributions from the other descriptors (mostly from khs.ddsN and C1SP3).



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

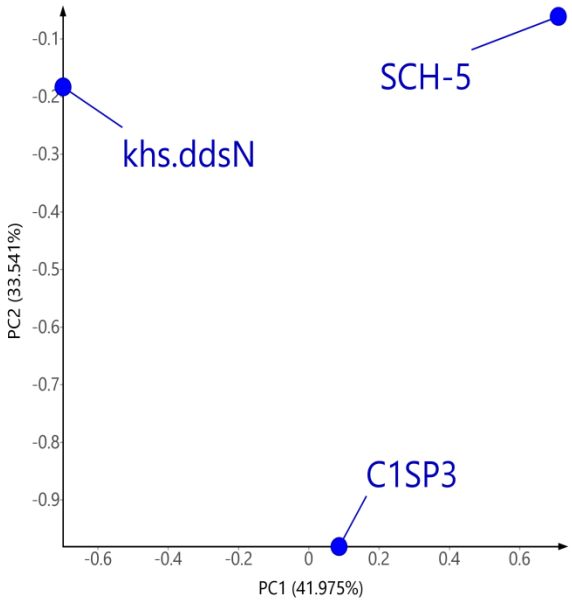


**Figure 13**: PC1–PC2 loadings plot using the three descriptors for the best model (GA-MLR)

The Y randomization test performed to guarantee that there are no random correlations was used in this study to validate the recognized QSAR model and also to ensure that the selected descriptors are not random. Therefore, results of the model should be of low statistical quality. Random MLR models created was done randomly by rearranging the dependent variable while maintaining the individual variables. The recently built QSAR models will predictably have significant $R^2$ and $Q^2$ low values for more than a few trials, thus confirming that developed QSAR models are robust. Approximately hundred trials of Y-randomization were conducted in this study which gave lesser values for $R^2$ and $Q^2$, thereby authenticating the initial model (the established GA-MLR model) (Figure 14).

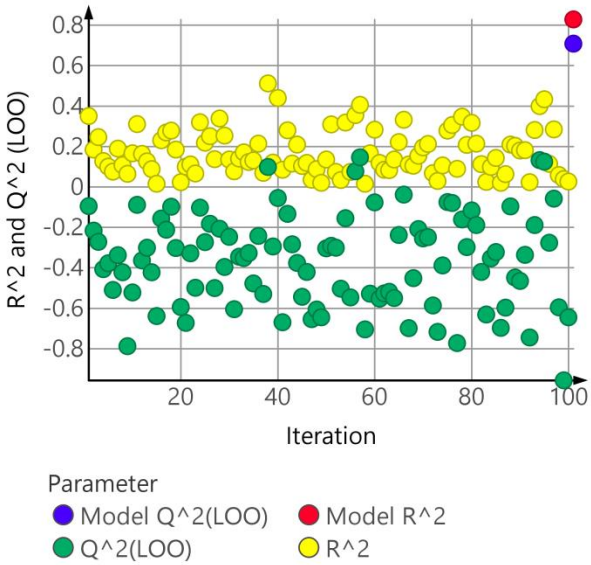


**Figure 14**: $R^2$ train and $Q^2$LOO values following numerous Y-randomization tests for GA-MLR

The residue for the predicted values of $pIC_{50}$ for the training and test sets against the experimental $pIC_{50}$values is plotted as presented in Figure 15. It was observed that the model did not indicate relative and systematic error at all, since the propagation of the residues on the horizontal lines is random.



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

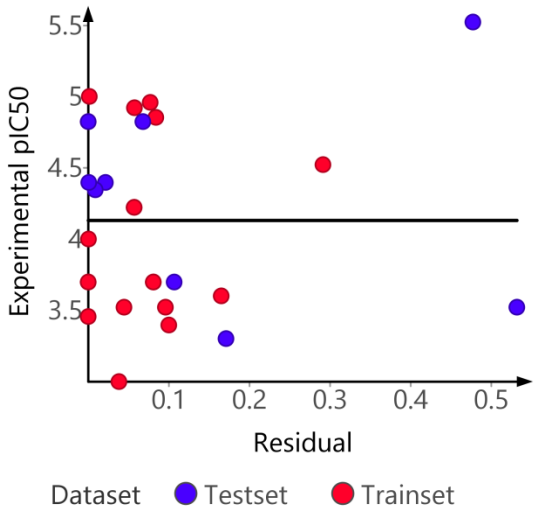


**Figure 15**: The residual against the experimental $pIC_{50}$ by adopting GA-MLR

Furthermore, QSAR model was utilized to envisage the bioactivities of the novel compounds from their respective 2D physicochemical properties (SCH-5, khs.ddsN and C1SP3). The compound in generation 0 represents the parent molecule (compound 10), and has a predicted $pIC_{50}$ value of 4.49. Twenty compounds across generation 3-10 have a predicted $pIC_{50}$ values higher than the parent molecule, which makes them more potent than the parent molecule (Figure 16). From the predicted $pIC_{50}$ we computed LipE and clogP for each of the novel compounds using Datawarrior v5.0.

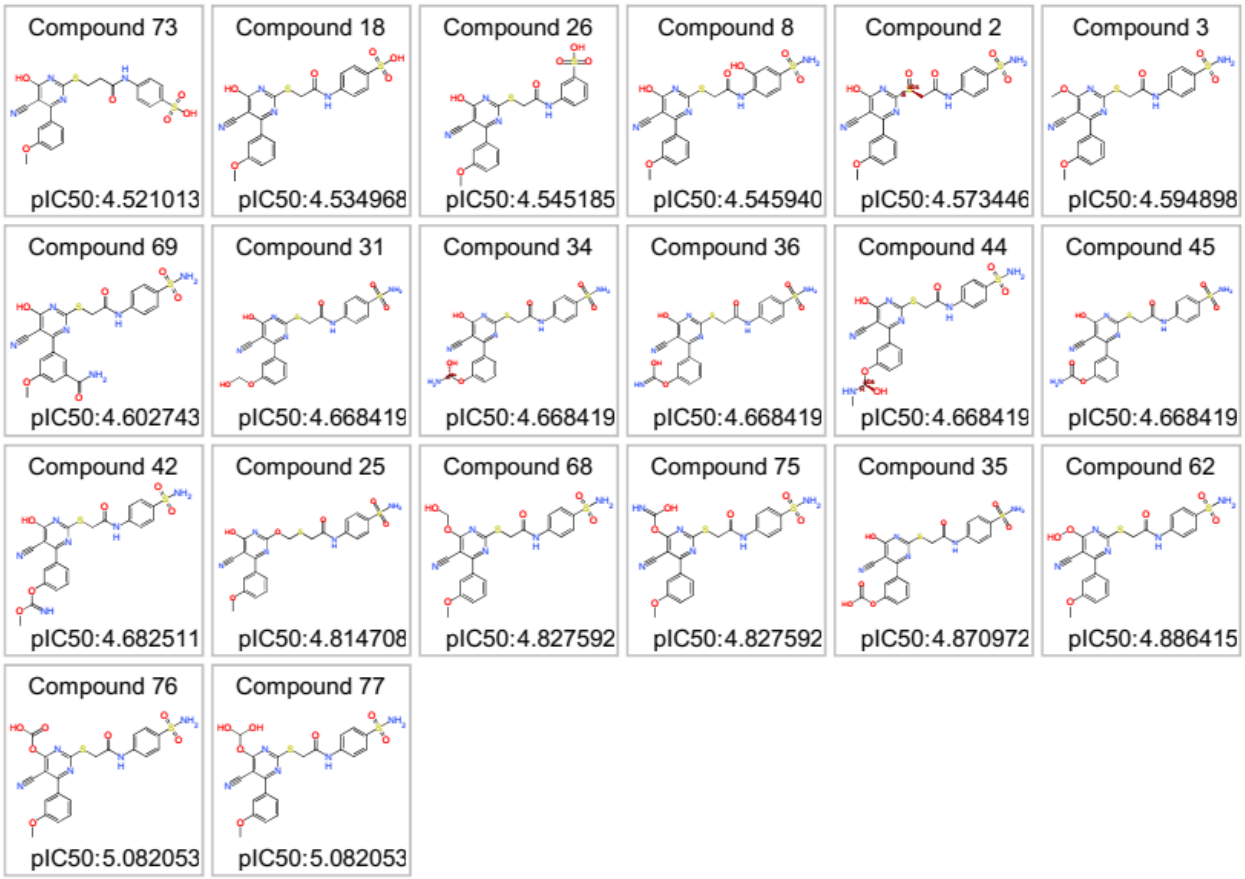


**Figure 16**: *De novo* synthesized compounds with higher predicted bioactivities



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

Increase in LipE can be attained by modifying the ligand to impact lipophilicity, potency or both. The biggest assured impact of LipE frequently occurs when adjustments improve potency and concurrently lower lipophilicity [33]. Based on our design strategy, 6 of the novel compounds (36, 37, 46, 47, 77 and 79) revealed an increase in both LipE and potency with logP decrease (Figure 17).

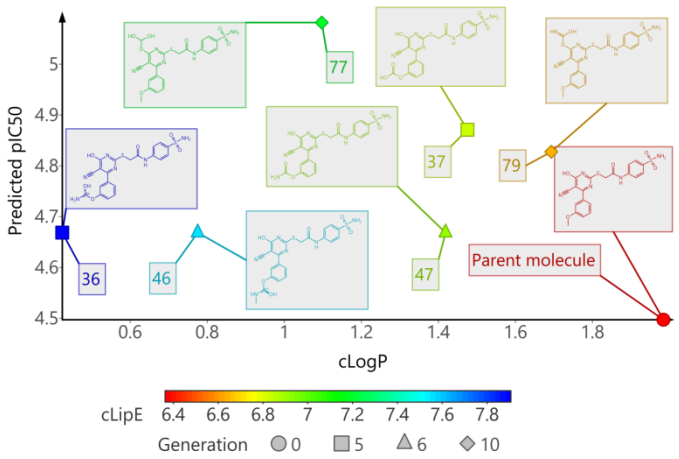


**Figure 17**: Optimized compounds based on bioactivities, logP and LipE

In order to have a better understanding of the molecular mechanism underlying the action of the unique compounds selected on the basis of their bioactivities and LipE, molecular docking was employed in this study. In the present study, the six selected compounds and their parent molecule were docked into the binding pocket of SARS coronavirus 3C-like proteinase for their inhibitory (antagonistic) (Figure 18a and b) properties. Compound 77 showed a better binding affinity, -7.4kcal/mol, when compared with the parent molecule (compound 10; -7.3kcal/mol) and thus is considered as the lead compound (Figure 19).

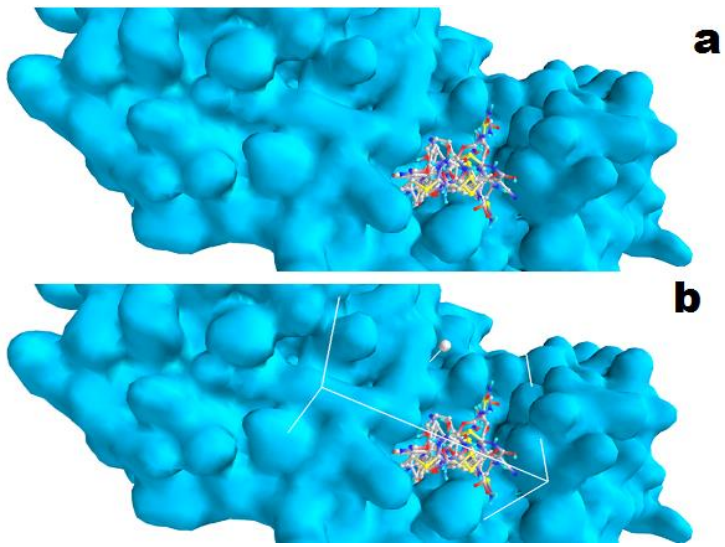


**Figure 18**: (a) Grid box within which the ligand binds 76.0065 * -11.5107 * 18.0445 along the x, y and z axes correspondingly (b) Compounds within binding pockets

The highest binding energy of -7.4kcal/mol attributed to compound 77 is considered to be as a result of chemical interactions at the receptor's active site (Figure 20a) which includes: Five (5) hydrogen bonds involving GLN110, ASN151 and HIS246 residues; Three (3) hydrophobic interactions involving VAL104, LEU202 and ILE249 residues. However, that of compound 10 serving as the reference molecule presents the following chemical interactions at the binding pocket (Figure 20b): Four (4) hydrogen bonds involving ASN151, ASP295 and THR111 residues with two (2) hydrophobic interactions involving ILE249 and PRO293. Therefore,



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

the consequence of higher binding affinity of compound 77 within SARS coronavirus 3C-like proteinase drug-able pocket is due to the presence of more hydrophobic interactions and more hydrogen bonds when compared to compound 10.

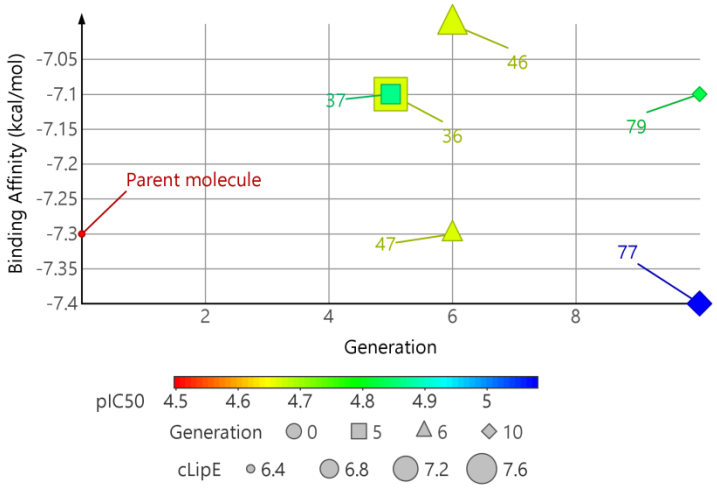


**Figure 19**: Binding affinities of selected 3CL-pro inhibitors

Hydrogen (H)-bonds potentiates varied cellular functions by accelerating molecular interactions. In order words, hydrogen bonds are considered to be facilitators of protein-ligand binding [48, 49]. Previous studies have shown that interdependent receptor-ligand H-bond pairings potentiate high-affinity binding, which corresponds to an increase in binding affinity [50]. Additionally, ASN151 was predicted to be universally implicated in hydrogen bonding with the ligands within SARS coronavirus 3C-like proteinase drug-able pocket.

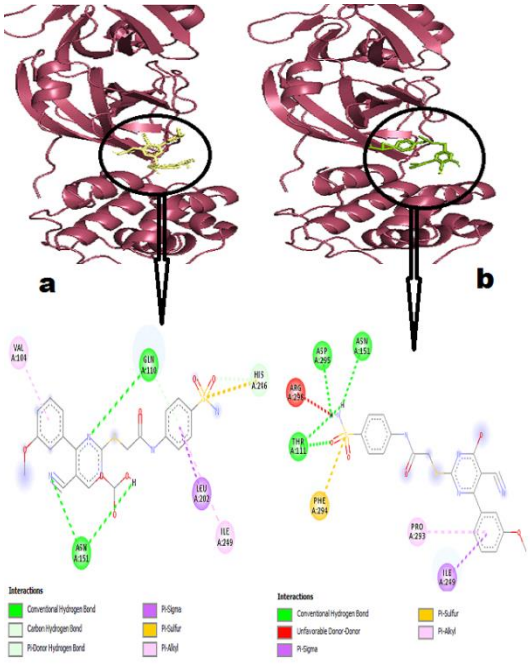


**Figure 20**: 3D and 2D interactions (a) compound 77 (b) compound 10

## 4. CONCLUSION

This study has maximized the combination of LipE and logP to select potent leads against SARS coronavirus 3C-like proteinase Inhibitors and to optimize these leads based on changes in their physicochemical properties in order to discover highly efficient and



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS

reliable clinical candidates for the cure of COVID-19. Six novel compounds (36, 37, 46, 47, 77 and 79) were suggested for further *in vitro* and preclinical testing based on their predicted efficiencies values and potency.

**Funding:**
This study has not received any external funding.

**Ethical approval**
Not applicable.

**Conflict of Interest**
The authors declare that there are no conflicts of interests.

**Data and materials availability**
All data associated with this study are present in the paper.

## REFERENCES AND NOTES

1. Martinez, M.A., 2020. Compounds with therapeutic potential against novel respiratory 2019 coronavirus. *Antimicrobial agents and chemotherapy*, *64*(5).
2. Bhowmik, D., Nandi, R. and Kumar, D., 2020. Evaluation of flavonoids as 2019-nCoV cell entry inhibitor through molecular docking and pharmacological analysis.
3. Gao, J., Tian, Z. and Yang, X., 2020. Breakthrough: Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Bioscience trends*.
4. Gorbalenya, A.E., Baker, S.C., Baric, R., Groot, R.J.D., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W. and Penzar, D., 2020. Severe acute respiratory syndrome-related coronavirus: The species and its viruses–a statement of the Coronavirus Study Group..
5. Chen, Y., Liu, Q. and Guo, D., 2020. Emerging coronaviruses: genome structure, replication, and pathogenesis. *Journal of medical virology*, *92*(4), pp.418-423.
6. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R. and Niu, P., 2020. A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*.
7. Ubani, A., Agwom, F., Morenikeji, O.R., Shehu, N.Y., Luka, P., Umera, E.A., Umar, U., Omale, S., Nnadi, E. and Aguiyi, J.C., 2020. Molecular Docking Analysis of Some Phytochemicals On Two SARS-CoV-2 Targets. *bioRxiv*.
8. Khan, S.A., Zia, K., Ashraf, S., Uddin, R. and Ul-Haq, Z., 2020. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, pp.1-10.
9. Dayer, M.R., Taleb-Gassabi, S. and Dayer, M.S., 2017. Lopinavir; a potent drug against coronavirus infection: insight from molecular docking study. *Arch Clin Infect Dis*, *12*(4), p.e13823.
10. Chen, Y.W., Yiu, C.P.B. and Wong, K.Y., 2020. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research*, *9*.
11. Khan, S.A., Zia, K., Ashraf, S., Uddin, R. and Ul-Haq, Z., 2020. Identification of chymotrypsin-like protease inhibitors of SARS-CoV-2 via integrated computational approach. *Journal of Biomolecular Structure and Dynamics*, pp.1-10.
12. Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., Zhang, B., Li, X., Zhang, L., Peng, C. and Duan, Y., 2020. Structure of M pro from SARS-CoV-2 and discovery of its inhibitors. *Nature*, pp.1-5.
13. Ceccarelli, M., Berretta, M., Rullo, E.V., Nunnari, G. and Cacopardo, B., 2020. Editorial–Differences and similarities between Severe Acute Respiratory Syndrome (SARS)-CoronaVirus (CoV) and SARS-CoV-2. Would a rose by another name smell as sweet?. *European review for medical and pharmacological sciences*, *24*, pp.2781-2783.
14. Virupakshaiah, D.B.M., Kelmani, C., Patil, R. and Hegade, P., 2007, June. Computer aided docking studies on antiviral drugs for SARS. In *Proceedings of world academy of science, engineering and technology* (Vol. 24, pp. 307-6884).
15. Ryckmans, T., Edwards, M.P., Horne, V.A., Correia, A.M., Owen, D.R., Thompson, L.R., Tran, I., Tutt, M.F. and Young, T., 2009. Rapid assessment of a novel series of selective CB2 agonists using parallel synthesis protocols: A Lipophilic Efficiency (LipE) analysis. *Bioorganic & medicinal chemistry letters*, *19*(15), pp.4406-4409.
16. Leeson, P.D. and Springthorpe, B., 2007. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nature Reviews Drug Discovery*, *6*(11), pp.881-890.



ISSN 2278–540X EISSN 2278–5396 | OPEN ACCESS

17. Keserü, G.M. and Makara, G.M., 2009. The influence of lead discovery strategies on the properties of drug candidates. *nature reviews Drug Discovery*, *8*(3), pp.203-212.
18. Arnott, J.A. and Planey, S.L., 2012. The influence of lipophilicity in drug discovery and design. *Expert opinion on drug discovery*, *7*(10), pp.863-875.
19. Hann, M.M. and Keserü, G.M., 2012. Finding the sweet spot: the role of nature and nurture in medicinal chemistry. *Nature reviews Drug discovery*, *11*(5), pp.355-365.
20. Oka, R., Engkvist, O. and Chen, H., 2013. Investigation of the influence of molecular topology on ligand binding. *Journal of Molecular Graphics and Modelling*, *40*, pp.22-29.
21. Wyatt, P.G., Woodhead, A.J., Berdini, V., Boulstridge, J.A., Carr, M.G., Cross, D.M., Davis, D.J., Devine, L.A., Early, T.R., Feltell, R.E. and Lewis, E.J., 2008. Identification of N-(4-piperidinyl)-4-(2, 6-dichlorobenzoylamino)-1 H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. *Journal of medicinal chemistry*, *51*(16), pp.4986-4999.
22. Squires, M.S., Feltell, R.E., Wallis, N.G., Lewis, E.J., Smith, D.M., Cross, D.M., Lyons, J.F. and Thompson, N.T., 2009. Biological characterization of AT7519, a small-molecule inhibitor of cyclin-dependent kinases, in human tumor cell lines. *Molecular cancer therapeutics*, *8*(2), pp.324-332.
23. Imig JD, Hammock BD. Soluble epoxide hydrolase as a therapeutic target for cardiovascular diseases. Nature reviews Drug discovery. 2009 Oct;8(10) pp 794-805.
24. Ryckmans, T., Edwards, M.P., Horne, V.A., Correia, A.M., Owen, D.R., Thompson, L.R., Tran, I., Tutt, M.F. and Young, T., 2009. Rapid assessment of a novel series of selective CB2 agonists using parallel synthesis protocols: A Lipophilic Efficiency (LipE) analysis. *Bioorganic & medicinal chemistry letters*, *19*(15), pp.4406-4409.
25. Lange, J.H., van der Neut, M.A., Wals, H.C., Kuil, G.D., Borst, A.J., Mulder, A., den Hartog, A.P., Zilaout, H., Goutier, W., van Stuivenberg, H.H. and van Vliet, B.J., 2010. Synthesis and SAR of novel imidazoles as potent and selective cannabinoid CB2 receptor antagonists with high binding efficiencies. *Bioorganic & medicinal chemistry letters*, *20*(3), pp.1084-1089.
26. Freeman-Cook, K.D., Autry, C., Borzillo, G., Gordon, D., Barbacci-Tobin, E., Bernardo, V., Briere, D., Clark, T., Corbett, M., Jakubczak, J. and Kakar, S., 2010. Design of selective, ATP-competitive inhibitors of Akt. *Journal of medicinal chemistry*, *53*(12), pp.4615-4622.
27. Liu, K.K.C., Bagrodia, S., Bailey, S., Cheng, H., Chen, H., Gao, L., Greasley, S., Hoffman, J.E., Hu, Q., Johnson, T.O. and Knighton, D., 2010. 4-methylpteridinones as orally active and selective PI3K/mTOR dual inhibitors. *Bioorganic & medicinal chemistry letters*, *20*(20), pp.6096-6099.
28. Mowbray, C.E., Burt, C., Corbau, R., Gayton, S., Hawes, M., Perros, M., Tran, I., Price, D.A., Quinton, F.J., Selby, M.D. and Stupple, P.A., 2009. Pyrazole NNRTIs 4: selection of UK-453,061 (lersivirine) as a development candidate. *Bioorganic & medicinal chemistry letters*, *19*(20), pp.5857-5860.
29. Edwards, M.P. and Price, D.A., 2010. Role of physicochemical properties and ligand lipophilicity efficiency in addressing drug safety risks. In *Annual Reports in Medicinal Chemistry* (Vol. 45, pp. 380-391). Academic Press.
30. Meanwell, N.A., 2011. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. *Chemical research in toxicology*, *24*(9), pp.1420-1456.
31. O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T. and Hutchison, G.R., 2011. Open Babel: An open chemical toolbox. *Journal of cheminformatics*, *3*(1), p.33.
32. Mahabeleshwar, G.H. and Kundu, G.C., 2003. Tyrosine kinase p56lck regulates cell motility and nuclear factor κB-mediated secretion of urokinase type plasminogen activator through tyrosine phosphorylation of IκBα following hypoxia/reoxygenation. *Journal of Biological Chemistry*, *278*(52), pp.52598-52612.
33. Tsai, K.C., Chen, S.Y., Liang, P.H., Lu, I.L., Mahindroo, N., Hsieh, H.P., Chao, Y.S., Liu, L., Liu, D., Lien, W. and Lin, T.H., 2006. Discovery of a novel family of SARS-CoV protease inhibitors by virtual screening and 3D-QSAR studies. *Journal of medicinal chemistry*, *49*(12), pp.3485-3495.
34. Paul, M.K. and Mukhopadhyay, A.K., 2004. Tyrosine kinase–role and significance in cancer. *International journal of medical sciences*, *1*(2), p.101.
35. Roy, K., Das, R.N., Ambure, P. and Aher, R.B., 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, *152*, pp.18-33.
36. Veber, D.F., Johnson, S.R., Cheng, H.Y., Smith, B.R., Ward, K.W. and Kopple, K.D., 2002. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*, *45*(12), pp.2615-2623.
37. Yadav, D.K. and Khan, F., 2013. QSAR, docking and ADMET studies of camptothecin derivatives as inhibitors of DNA topoisomerase-I. *Journal of Chemometrics*, *27*(1-2), pp.21-33.
38. Yadav, D.K., Kumar, S., Saloni, H.S., Kim, M.H., Sharma, P., Misra, S. and Khan, F., 2017. Molecular docking, QSAR and ADMET studies of withanolide analogs against breast cancer. *Drug Design, Development and Therapy*, *11*, p.1859.
39. Kumar Yadav, D., Kalani, K., Khan, F. and Kumar Srivastava, S., 2013. QSAR and docking based semi-synthesis and in vitro evaluation of 18 β-glycyrrhetinic acid

derivatives against human lung cancer cell line A-549. *Medicinal Chemistry*, *9*(8), pp.1073-1084.

40. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E., 2000. The protein data bank. *Nucleic acids research*, *28*(1), pp.235-242.
41. Brockunier, L.L., He, J., Colwell Jr, L.F., Habulihaz, B., He, H., Leiting, B., Lyons, K.A., Marsilio, F., Patel, R.A., Teffera, Y. and Wu, J.K., 2004. Substituted piperazines as novel dipeptidyl peptidase IV inhibitors. *Bioorganic & medicinal chemistry letters*, *14*(18), pp.4763-4766.
42. Johnson, T.W., Gallego, R.A. and Edwards, M.P., 2018. Lipophilic efficiency as an important metric in drug design. *Journal of medicinal chemistry*, *61*(15), pp.6401-6420.
43. Johnson, T.W., Tanis, S.P., Butler, S.L., Dalvie, D., DeLisle, D.M., Dress, K.R., Flahive, E.J., Hu, Q., Kuehler, J.E., Kuki, A. and Liu, W., 2011. Design and synthesis of novel N-hydroxy-dihydronaphthyridinones as potent and orally bioavailable HIV-1 integrase inhibitors. *Journal of medicinal chemistry*, *54*(9), pp.3393-3417.
44. Freeman-Cook, K.D., Hoffman, R.L. and Johnson, T.W., 2013. Lipophilic efficiency: the most important efficiency metric in medicinal chemistry. *Future Medicinal Chemistry*, *5*(2), pp.113-115.
45. Bayat, Z. And Yazdan Abad, M.F., 2011. Quantitative Structure-Property Relationship (Qspr) Study Of Kovats Retention Indices Of Some Of Adamantane Derivatives By The Genetic Algorithm And Multiple Linear Regression (Ga-Mlr) Method. *Petroleum & Coal*, *53*(2).
46. Sander, T., Freyss, J., von Korff, M. and Rufener, C., 2015. DataWarrior: an open-source program for chemistry aware data visualization and analysis. *Journal of chemical information and modeling*, *55*(2), pp.460-473.
47. Cramer, R.D., 1993. Partial least squares (PLS): its strengths and limitations. *Perspectives in Drug Discovery and Design*, *1*(2), pp.269-278.
48. Salentin, S., Haupt, V.J., Daminelli, S. and Schroeder, M., 2014. Polypharmacology rescored: Protein–ligand interaction profiles for remote binding site similarity assessment. *Progress in biophysics and molecular biology*, *116*(2-3), pp.174-186.
49. Sawada, T., 2012. Chemical insight into the influenza a virus hemagglutinin binding to the sialoside revealed by the fragment molecular orbital method. *Open Glycoscience*, *5*(1).
50. Chen, D., Oezguen, N., Urvil, P., Ferguson, C., Dann, S.M. and Savidge, T.C., 2016. Regulation of protein-ligand binding affinity by hydrogen bond pairing. *Science advances*, *2*(3), p.e1501240.



ISSN 2278–540X EISSN 2278–5396 I OPEN ACCESS