



Tutorial article

Structure features of peptide-type SARS-CoV main protease inhibitors: Quantitative structure activity relationship study

Vijay H. Masand^{a,*}, Siddhartha Akasapu^b, Ajaykumar Gandhi^c, Vesna Rastija^d, Meghshyam K. Patil^e

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

^b Corden Pharma, Colorado, USA

^c Department of Chemistry, Government College of Arts and Science, Aurangabad, Maharashtra, 431 004, India

^d Department of Agroecology and Environmental Protection, Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia

^e Department of Chemistry, Osmanabad Sub-Centre Dr. Babasaheb Ambedkar Marathwada University, Osmanabad, Maharashtra, India

ARTICLE INFO

Keywords:

QSAR
COVID-19
SARS-CoV
SARS-CoV-2
Peptide-type compounds

ABSTRACT

In the present work, an extensive QSAR (Quantitative Structure Activity Relationships) analysis of a series of peptide-type SARS-CoV main protease (MPro) inhibitors following the OECD guidelines has been accomplished. The analysis was aimed to identify salient and concealed structural features that govern the MPro inhibitory activity of peptide-type compounds. The QSAR analysis is based on a dataset of sixty-two peptide-type compounds which resulted in the generation of statistically robust and highly predictive multiple models. All the developed models were validated extensively and satisfy the threshold values for many statistical parameters (for e.g. $R^2 = 0.80-0.82$, $Q^2_{100} = 0.74-0.77$, $Q^2_{LMO} = 0.66-0.67$). The developed QSAR models identified number of sp^2 hybridized Oxygen atoms within seven bonds from aromatic Carbon atoms, the presence of Carbon and Nitrogen atoms at a topological distance of 3 and other interrelations of atom pairs as important pharmacophoric features. Hence, the present QSAR models have a good balance of Qualitative (Descriptive QSARs) and Quantitative (Predictive QSARs) approaches, therefore useful for future modifications of peptide-type compounds for anti-SARS-CoV activity.

1. Introduction

Coronaviruses (subfamily *Coronavirinae*, family *Coronaviridae*, order *Nidovirales*) have been classified into four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus* [1]. Of these, especially *Betacoronavirus*, have been found to cause respiratory, enteric, hepatic, and neurological diseases in many animals, and also in humans [2]. The two *Betacoronaviruses* (β CoVs) viz. severe acute respiratory syndrome CoV (SARS-CoV) and Middle East respiratory syndrome CoV (MERS-CoV) spread in 2003 and 2012, respectively [2,3]. SARS-CoV and MERS-CoV have fatality rate of 10% and 35%, respectively [2,3]. Unfortunately, till this date, there is no appropriate treatment for SARS-CoV and MERS-CoV [2,3]. The situation has worsened with the recent outbreak of more contagious novel coronavirus SARS-CoV-2, which is the

causative agent for COVID-19 pandemic. This disease has a long-term socio-economic impact on many countries due to high infection and mortality rate. At present, this disease is responsible for more than 732, 499 deaths and 20 million confirmed infected cases [4]. Therefore, there is an urgent need to curb this deadly disease.

The recent outbreak of COVID-19 pandemic, with no approved treatment for infections, has spread quickly in many countries. Even though, COVID-19 is a new disease caused by the novel coronavirus SARS-CoV-2, but a good number of studies suggest that it has significant similarity with SARS-CoV [3,5–7]. This similarity is reflected from the facts that SARS-CoV and SARS-CoV-2 have a 79% similarity at the genome level, in addition, a lot of proteins like glycosylated spike (S) protein (76% of sequence similarity and a highly conserved receptor-binding domain), papain-like protease (PLpro) (83% similarity with similar

Abbreviations: SMILES, Simplified molecular-Input Line-Entry System; GA, Genetic algorithm; MLR, Multiple linear Regression; QSAR, Quantitative structure-activity analysis; WHO, World health organization; ADMET, Absorption, Distribution, Metabolism, Excretion and Toxicity; OLS, Ordinary Least Square; QSARINS, QSAR Insurbria; OECD, Organisation for Economic Co-operation and Development; OFS, Objective Feature Selection; SFS, Subjective Feature Selection.

* Corresponding author.

E-mail addresses: vijaymasand@gmail.com (V.H. Masand), asidhu09@gmail.com (S. Akasapu), gascajay18@gmail.com (A. Gandhi), vrastija@fazos.hr (V. Rastija).

<https://doi.org/10.1016/j.chemolab.2020.104172>

Received 26 April 2020; Received in revised form 9 September 2020; Accepted 2 October 2020

Available online 3 October 2020

0169-7439/© 2020 Elsevier B.V. All rights reserved.

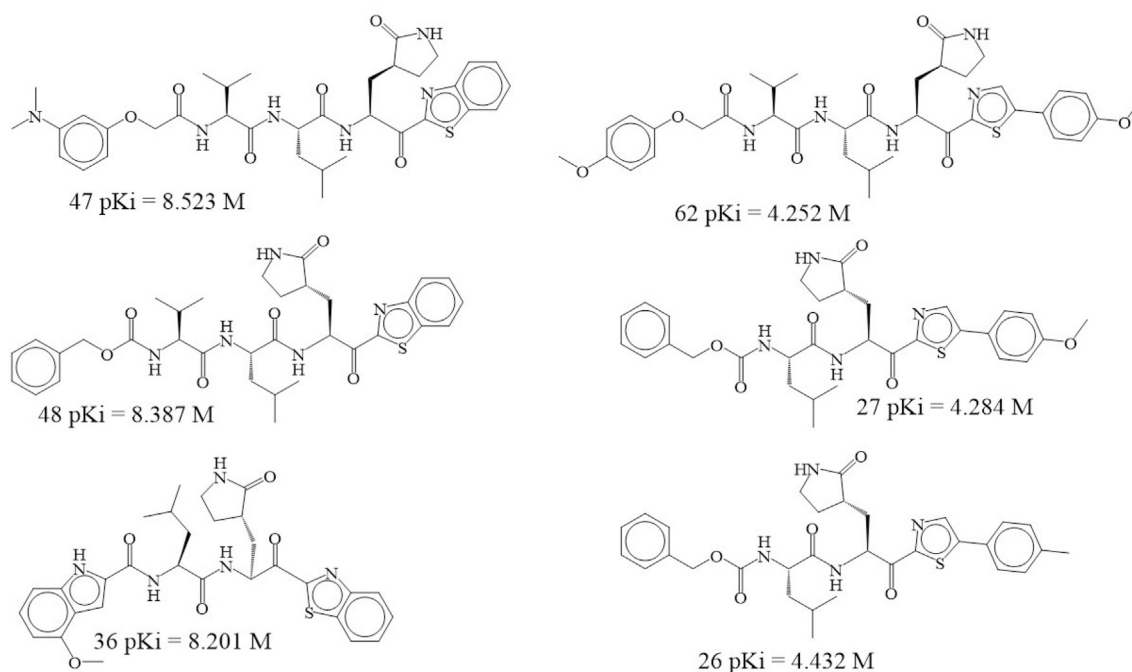


Fig. 1. Variations in activity and chemical structure in the present dataset.

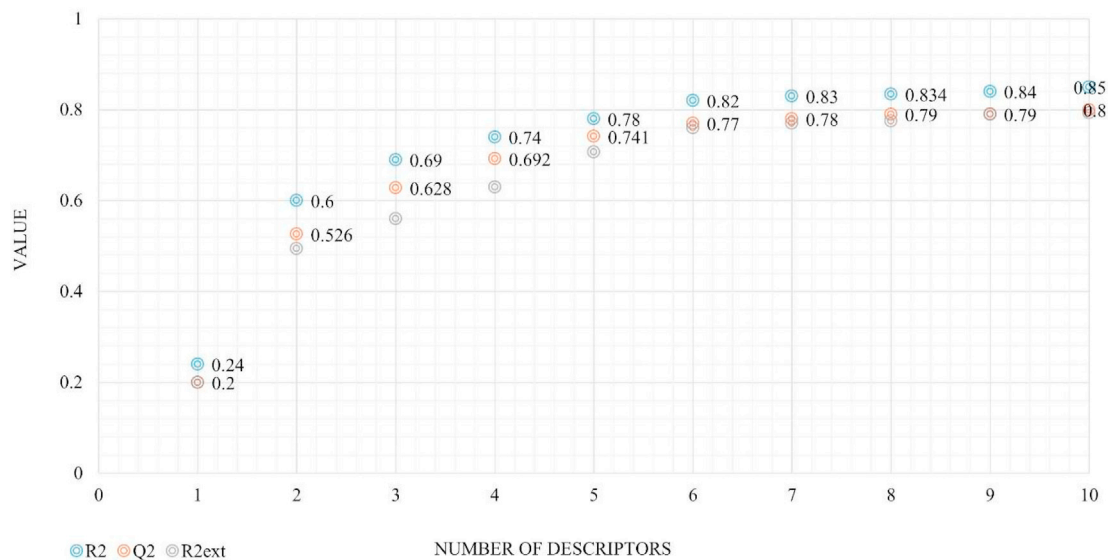


Fig. 2. Correlation of statistical parameters with number of molecular descriptors.

active sites), RNA-dependent RNA polymerase (RdRp), and coronavirus main protease (3CLpro) are essential for both. To add further, RdRp and 3CLpro protease of SARS-CoV-2 share over 95% of sequence similarity with those of SARS-CoV [3,5–7]. In short, viral proteins essential for SARS-CoV-2 entry into host cells and subsequent replication are highly similar to those associated with SARS-CoV. Consequently, research and development on SARS-CoV could be useful for the development of a therapeutic or preventive agent for COVID-19. To add further, designing and synthesis of a new drug and its congeners followed by their bio-screening need a lot of time. Therefore, high similarity of SARS-CoV-2 with SARS-CoV and considering the potential threat of this pandemic, it is reasonable that a therapeutic drug, which has been previously tested against SARS-CoV could be easily optimized to be effective

against SARS-CoV-2.

To optimize a compound to become a lead or a drug, a researcher needs to follow an easy, efficient, economical, and eco-friendly approach (e-Chemistry approach). A feasible solution to achieve these goals is to use Computer-Aided Drug Designing (CADD). CADD is a contemporary approach with a good number of benefits like cheaper, result oriented, minimizes animal testing as well as trials and errors, less time-consuming, a few advantages to mention. In recent time, the thriving branches of CADD such as QSAR (Quantitative Structure Activity Relationships), Molecular docking, Pharmacophore modelling, etc. have contributed significantly in optimization of lead and drug candidates [8, 9].

QSAR analysis involves finding mathematical correlation between the

Table 2
Statistical parameters for developed QSAR models 1.1 and 1.2

Statistical Parameter	Model-1.1	Model-1.2
Fitting		
R^2_{tr}	0.801	0.824
R^2_{adj}	0.78	0.8
$R^2_{tr} - R^2_{adj}$	0.022	0.025
LOF	0.326	0.324
Kxx	0.241	0.271
ΔK	0.06	0.05
RMSE _{tr}	0.461	0.433
MAE _{tr}	0.375	0.328
RSS _{tr}	13.15	9.366
CCC _{tr}	0.89	0.904
s	0.489	0.467
F	36.953	33.606
Internal validation		
R^2_{cv} (Q^2_{loo})	0.741	0.769
$R^2 - R^2_{cv}$	0.06	0.055
RMSE _{cv}	0.526	0.496
MAE _{cv}	0.427	0.378
PRESS _{cv}	17.14	12.312
CCC _{cv}	0.857	0.874
Q^2_{LMO}	0.673	0.665
R^2_{Yscr}	0.097	0.124
Q^2_{Yscr}	-0.182	-0.22
External validation		
RMSE _{ex}	-	0.527
MAE _{ex}	-	0.461
PRESS _{ext}	-	3.332
R^2_{ex}	-	0.758
$Q^2 - F^1$	-	0.741
$Q^2 - F^2$	-	0.74
$Q^2 - F^3$	-	0.739
CCC _{ex}	-	0.841
$R^2 - ExPy$	0.743	0.758
R^2_o	0.684	0.623
k'	0.996	0.975
$1 - (R^2/R^2_o)$	0.08	0.178
$r^{-2}m$	0.562	0.48
R^2_o	0.741	0.753
k	0.997	1.019
$1 - (R^2 - ExPy/R^2_o)$	0.002	0.006
r^2m	0.713	0.709

structural features of congeneric molecules with bio-activity. The main steps followed during a typical QSAR analysis are (1) Selection of an appropriate dataset (2) 3D-structure generation and their optimization using suitable technique (3) Molecular descriptor calculation followed by their pruning (4) Generation of QSAR model using a proper algorithm for feature (that is, molecular descriptor) selection and (5) Adequate validation of developed QSAR model [10].

QSAR is an established CADD branch, which has been used successfully to identify noticeable and hidden structural patterns/features having correlation with a desired activity/property (Qualitative QSAR). In addition, it is helpful to predict the activity/property before the actual synthesis and testing of a molecule (Quantitative QSAR). Therefore, QSAR analysis is performed routinely by the researchers to get important qualitative and quantitative idea about the congeneric molecules for optimization. An adequately validated and statistically robust QSAR analysis offers in-depth knowledge about the structural patterns that have good correlation with the desired activity/toxicity/property of a drug candidate and enhanced intuition for the mechanism of drug action [11–18].

A good number of researchers have synthesized peptide-type compounds and tested against SAR-CoV [19–22]. Recently, Zhang et al. [23] synthesized and tested α -Ketoamides (peptide-type compounds) as broad-spectrum inhibitors of coronavirus (SARS-CoV). Despite these efforts, optimization of peptide-type compounds is in pipeline to get a drug candidate. Henceforth, in the present work, we have developed

Qualitative cum Quantitative SAR models for a series of sixty-two peptide-type compounds for anti- SARS-CoV activity. The results could be beneficial for future optimizations of peptide-type compounds with better activity profile.

1.1. Experimental methodology

Selection of dataset: The dataset selected for the present work comprises sixty-two peptide-type compounds having moderate to high activity against SARS-CoV [19,21,22]. The selected peptide-type compounds possess $K_i = 3$ to 56,000 nM. The reported activity K_i values were converted to pK_i ($pK_i = -\log K_i$) before actual QSAR analysis. For the sake of ease and understanding the chemical space covered by the present dataset, most and least active three molecules have been depicted in Fig. 1. The SMILES notation for all selected molecules along with their reported activity values K_i , and pK_i are present in Table S1 in supplementary material.

1.2. Structure optimization and molecular descriptor calculation

The structures were drawn using ChemSketch 12 Freeware (www.acdlabs.com) followed by their conversion to 3D-structures using OpenBabel 2.4. Then, the force field MMFF94 was used to optimize the structures. The 3D-structures were optimized using default settings available in TINKER. The molecules in each set were then aligned using Open3DAlign. In the next step, the optimized and aligned molecules were then used for calculation of molecular descriptors using PyDescriptor [24] and PaDEL [25].

1.3. Molecular descriptor pruning

PyDescriptor and PaDEL provided more than 30,000 molecular descriptors for each molecule in all sets. Therefore, molecular descriptor pruning was essential to remove redundant molecular descriptors. For this, molecular descriptors with high co-linearity ($|R| > 0.90$) and nearly constant ($> 95\%$) were removed to avoid the inclusion of multi-collinear and spurious variables in GA-MLR (Genetic Algorithm–Multi-linear Regression) model, using objective feature selection in QSARINS ver. 2.2.2 [26–29]. The resulting reduced molecular descriptor pool comprised of only 668 molecular descriptors only but large enough to cover 1D-to 3D-descriptor space.

QSAR model building and their validation:

The reduced molecular descriptor set comprises zero-, one-, two- and three-dimensional descriptors, charge descriptors and molecular properties, thereby covering broad descriptor space. Subjective feature selection (SFS) was executed to build the statistically acceptable GA-MLR based QSAR models using QSARINS ver. 2.2.2. The developed models were subjected to thorough statistical validation (internal and external validation) according to OECD principles; models with high internal and external predictive ability have been reported.

The general procedure for building the QSAR models is as follows:

- (1) The set was bifurcated randomly, using random splitting option in QSARINS, into a training and a prediction set of 52 (i.e. 80 % training set) and 12 (i.e. 20 % prediction set), respectively. Then, the training set was used for model development, and the prediction set for the external validation, that is, to judge the predictive ability on new chemicals.
- (2) QSARINS was used to build GA-MLR based QSAR models using default settings. The selected fitness function to maximize in GA was Q^2 , this ensured the double cross-validation as well [26,27,30]. During model development it was observed that there was growth in the value of Q^2 up to six variables, but then, it had visible and significant reduction (see Fig. 2). Therefore, molecular descriptor selection was restricted to a set of six descriptors to

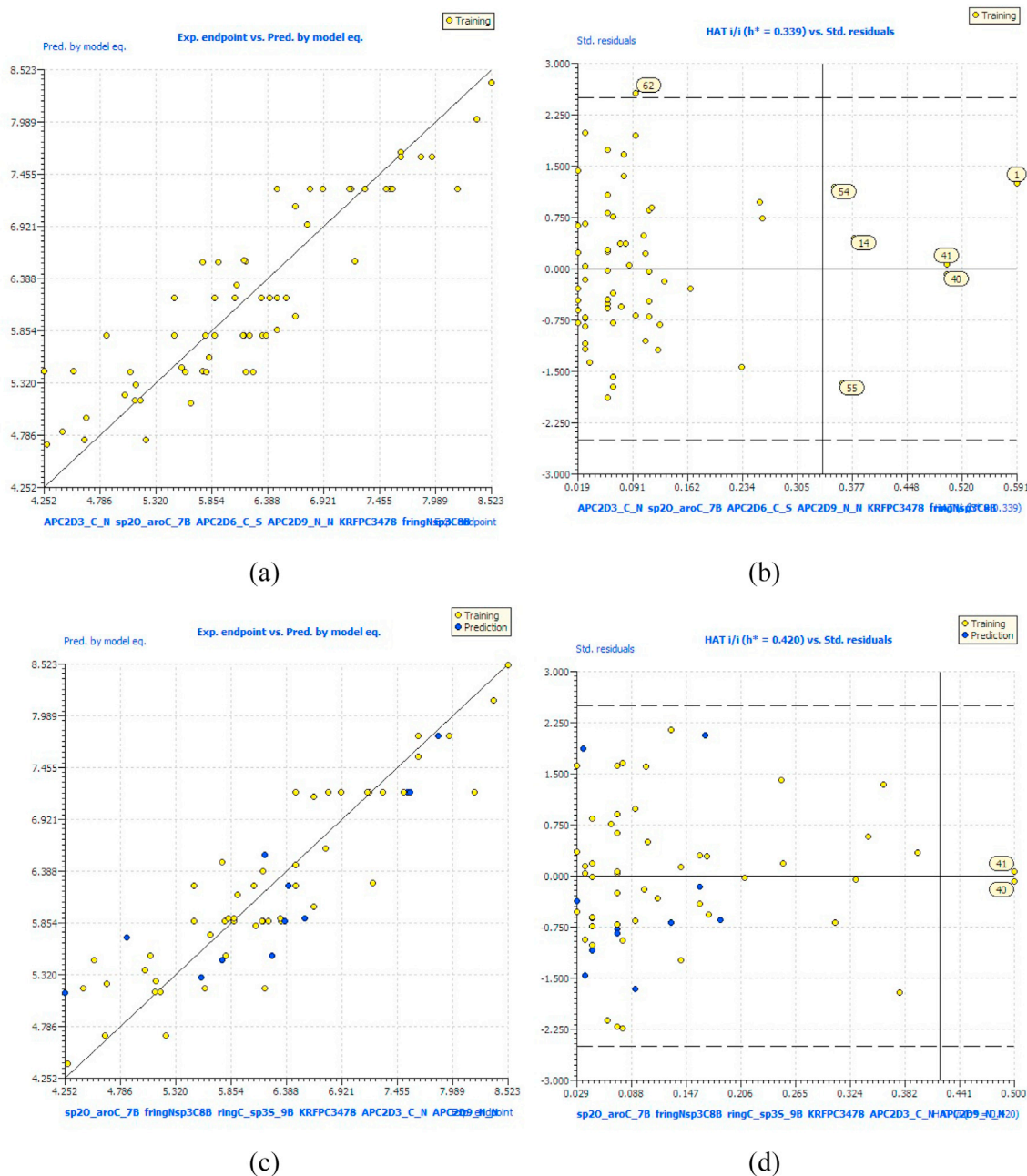


Fig. 3. (a) Graph of experimental vs Predicted pKi values for model 1.1 (b) Williams plot for model 1.1 (c) Graph of experimental vs Predicted pKi values for model 1.2 (d) Williams plot for model 1.2.

avoid over-fitting and develop easy and informative QSAR models.

The values for molecular descriptors, which are present in QSAR models, are available in the supplementary information for each molecule.

- (3) One of the OECD guideline suggests to thoroughly validating a QSAR model, therefore all the models were subjected to internal and external validation, Y-scrambling along with model applicability domain (AD) analysis using QSARINS. The statistical quality and strength of a GA-MLR based QSAR model was judged on the basis of: (a) internal validation based on leave-one-out (LOO) and leave-many-out (LMO) procedure (i.e. cross-validation (CV)); (b)

using External validation; (c) Y-randomization (or Y-scrambling) and (d) fulfilling of respective threshold value for the statistical parameters [31,32]: $R^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2 > Q^2$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $Q^2 - F^a \geq 0.60$, $r^2_m \geq 0.6$, $(1 - r^2/r_0^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1 - r^2/r_0^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $|r_0^2 - r_0'^2| < 0.3$ with $RMSE$ and MAE close to zero. A QSAR model that did not satisfy above mentioned criteria was consequently excluded.

Thus, the complete procedure involving molecular descriptor calculation and their pruning, followed by subjective feature selection along with model building and validation was performed on all the four sets. It was observed that the MMFF94 optimized set resulted in development of statistically better QSAR models, which have been reported in the present

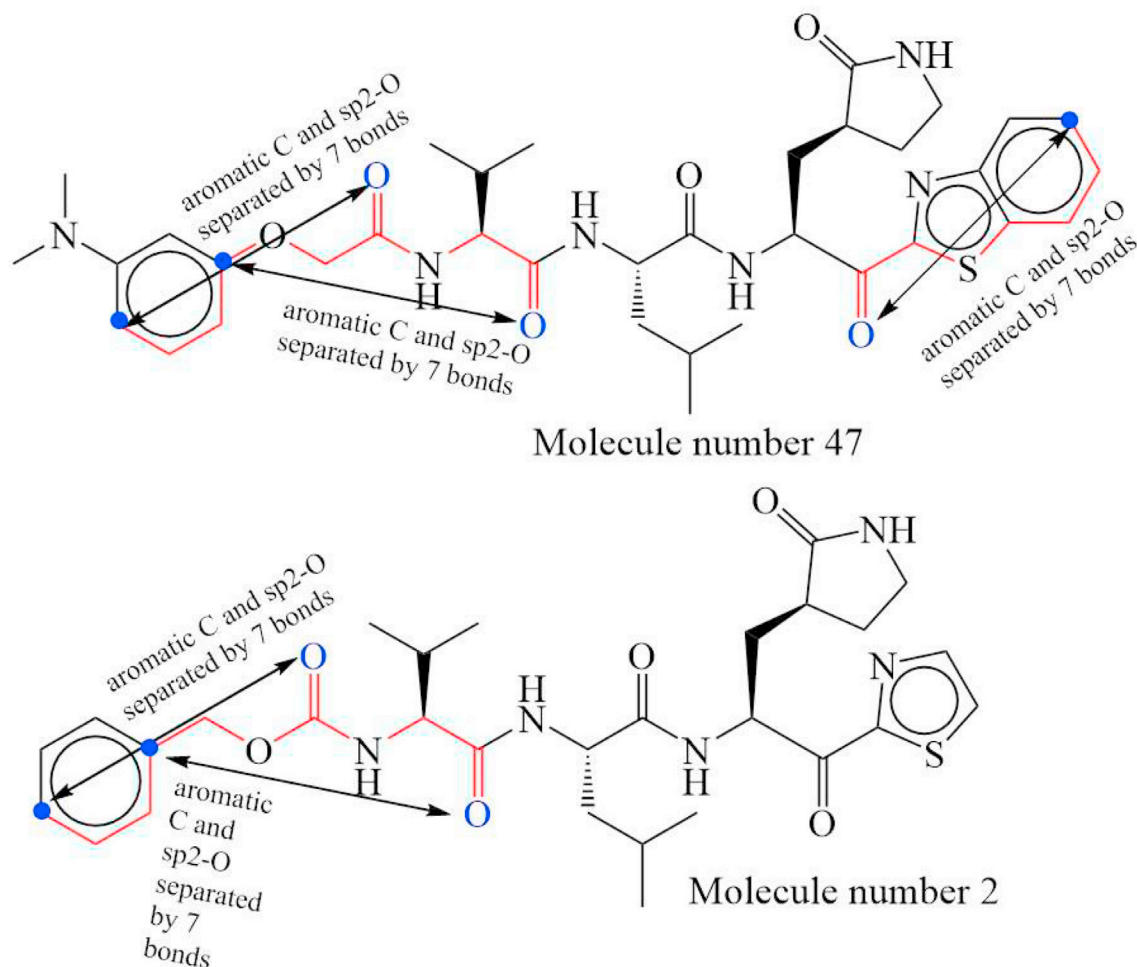


Fig. 4. Representation of sp2O_aroC7_B using molecule number 47 and 2 as representatives only.

work.

2. Result and discussions

Though, the dataset used in the present study contains only sixty-two compounds but the presence of positional isomers, heterocyclic rings, etc. significantly augment the chemical space covered by the peptide-type compounds. In our previous work on QSAR analysis related to small and moderate sized datasets [29], we have demonstrated that a QSAR model built using undivided whole dataset provides advantages like identification of maximum useful information, capturing of maximum relevant molecular descriptors, and benchmark for comparison and assessment of QSAR models constructed using divided datasets. In addition, this approach also helps to capture unrevealed structural features, which govern the bio-activity profile of congeneric molecules. Therefore, in the present work, models have been derived using divided and undivided datasets.

The derived QSAR models are as follow:

Model-1.1 (Undivided Set model):

$$\text{pKi} = 1.781 (\pm 1.357) + 1.073 (\pm 0.317) * \text{sp2O_aroC7B} + 0.377 (\pm 0.111) * \text{APC2D3_C_N} - 1.264 (\pm 0.647) * \text{APC2D9_N_N} + -2.538 (\pm 0.751) * \text{KRFPC3478} + -0.743 (\pm 0.233) * \text{fringNsp3C8B} - 0.61 (\pm 0.246) * \text{APC2D6_C_S}$$

Model-1.2 (Divided Set model: Training: 80%, Prediction: 20%):

$$\text{pKi} = 1.7 (\pm 1.658) + 1.182 (\pm 0.353) * \text{sp2O_aroC7B} + 0.363 (\pm 0.115) * \text{APC2D3_C_N} - 1.351 (\pm 0.642) * \text{APC2D9_N_N} - 2.422 (\pm 0.731) * \text{KRFPC3478} - 0.772 (\pm 0.269) * \text{fringNsp3C8B} - 0.169$$

$$(\pm 0.085) * \text{ringC_sp3S_9B}$$

The statistical parameters for developed models 1.1 and 1.2 have been presented in Table 2. The symbols have their usual meaning, which are available in the supplementary material also.

A good number of statistical parameters for model 1.1 and 1.2, which are related to fitting, internal and external validation and Y-scrambling, have been tabulated in Table 2. From Table 2, it is clear that R^2_{tr} , CCC_{tr} , CCC_{cv} , R^2_{adj} , and F satisfy the recommended threshold value, which shows that the QSAR models are statistically robust with adequate number of molecular descriptors in the models. The values for different cross-validation parameters such as R^2_{cv} , $RMSE_{cv}$, MAE_{cv} , CCC_{cv} , and Q^2_{LMO} support the statistical robustness of the QSAR models. The external predictive ability of the models is established by the high values of R^2_{ex} , Q^2F^1 , Q^2F^2 , Q^2F^3 , and CCC_{ex} [12,14,15,33–35].

In short, the developed QSAR models fulfill the recommended threshold values for many internal and external validation parameters. In addition, for a better validation of derived models, the model applicability domain (AD) was assured by plotting Williams plots for models 1.1 and 1.2 (see Fig. 3). Therefore, these models are statistically robust and possess good external predictive ability. To add further, satisfaction of recommended threshold values for these parameters along with low correlation among the molecular descriptors point out that these models are not developed by chance (see supplementary information).

2.1. Interpretation of QSAR models

The models 1.1 and 1.2 have been built using the undivided and

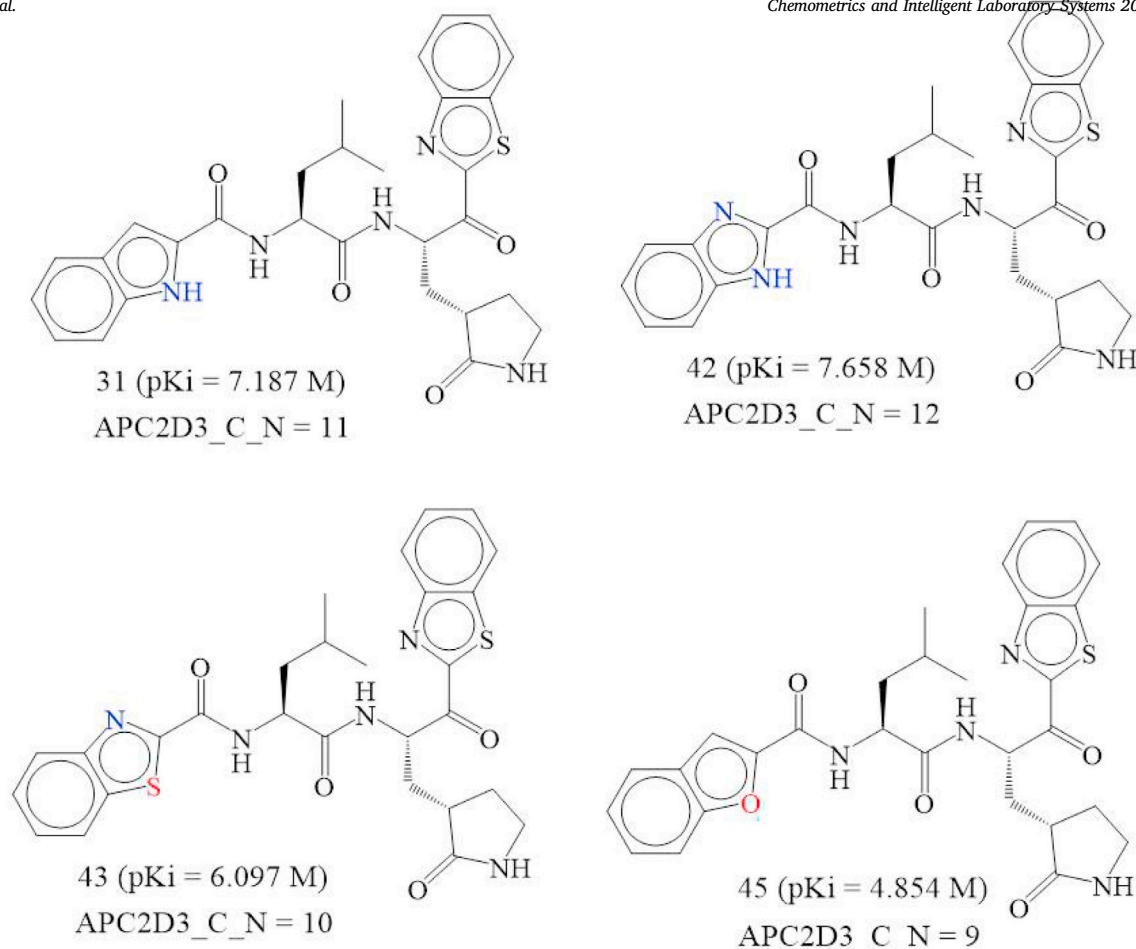


Fig. 5. Exemplification of molecular descriptor APC2D3_C_N.

divided dataset, respectively. They comprise four common molecular descriptors. Therefore, the approach to develop QSAR models using undivided and divided dataset has been successful in identification of greater number of important molecular descriptors, which is useful to capture maximal information. Though we have compared the activities of the molecules of the dataset in terms of a single descriptor, we make it clear that the combined or converse effect of confounding factors/descriptors do have additional influence on the activity profile of the compounds.

A molecular descriptor with a positive coefficient in both the model is $sp2O_aroC_7$ B (number of sp^2 hybridized Oxygen atoms within seven bonds from aromatic Carbon atoms). Since, in the present series of compounds, sp^2 hybridized Oxygen atoms are always present as a part of carbonyl group ($>C=O$), therefore it is rational to consider that this molecular descriptor also points out toward the presence of number of carbonyl groups in conjugation with aromatic Carbon atoms, that is, aromatic rings. Therefore, increasing the number of carbonyl groups within seven bonds from aromatic Carbon atoms could increase the anti-SARS activity of peptide-type of compounds. This observation is supported by the fact that molecule number 2 (pKi = 5.658 M) possesses only two such carbonyl groups, whereas the molecule number 47 (pKi = 8.523 M) and 48 (pKi = 8.387 M) have three such carbonyl groups. Another such comparison is possible between molecule number 30 (pKi = 5.77 M) and 31 (pKi = 7.174 M). This descriptor has been depicted in Fig. 4. The sp^2 -hybridized Oxygen and aromatic Carbon atoms have been shown using blue colour and the seven bonds are red coloured.

The atom-pair molecular descriptor APC2D3_C_N stands for the presence of Carbon and Nitrogen atoms at a topological distance of 3.

This molecular descriptor has a positive coefficient in both the models. Therefore, higher value of this descriptor could lead to better activity profile for a molecule. A comparison of 31 (pKi = 7.187 M) with 45 (pKi = 4.854 M) as well as among 42 (pKi = 7.658 M), 43 (pKi = 6.097 M) and 45 (pKi = 4.854 M) also support this observation. These molecules along with their APC2D3_C_N values have been depicted in Fig. 5.

An atom pair molecular descriptor with a negative coefficient in both the models is APC2D9_N_N, which stands for the number of Nitrogen atoms separated from each other by a topological distance of nine. It appears that increasing its value could cause diminish the anti-SARS activity of peptide-type compounds. A comparison of molecule number 51 (pKi = 7.658 M) with 54 (pKi = 6.658 M) and 55 (pKi = 6.658 M) supports this observation (see Fig. 6). Therefore, such a combination of Nitrogen atoms should be avoided for better activity.

KRFPC3478 is a finger print molecular descriptor, which represents the presence of Carbon atom at the position number 3 of an Indole ring. This descriptor with a negative coefficient in both models has negative contribution towards the anti-SARS activity of peptide-type compounds. An analysis of molecule 31–42 indicates that the presence of Indole moiety does not have negative contribution each time. Hence, it is rational to consider that the presence of Carbon atom at the position number 3 of an Indole ring has negative contribution. Therefore, it must be avoided to have better activity. The molecule 40 (pKi = 5.174 M) and 41 (pKi = 5.125 M) have relatively lower activity than other analogues bearing Indole ring (see Fig. 7).

Another molecular descriptor with a negative coefficient in model 1.1 and 1.2 and hence with a negative correlation with activity is fringNsp3C8B. This descriptor represents frequency of occurrence of sp^3 -

Absence of Nitrogen atoms separated from each other by a topological distance of nine

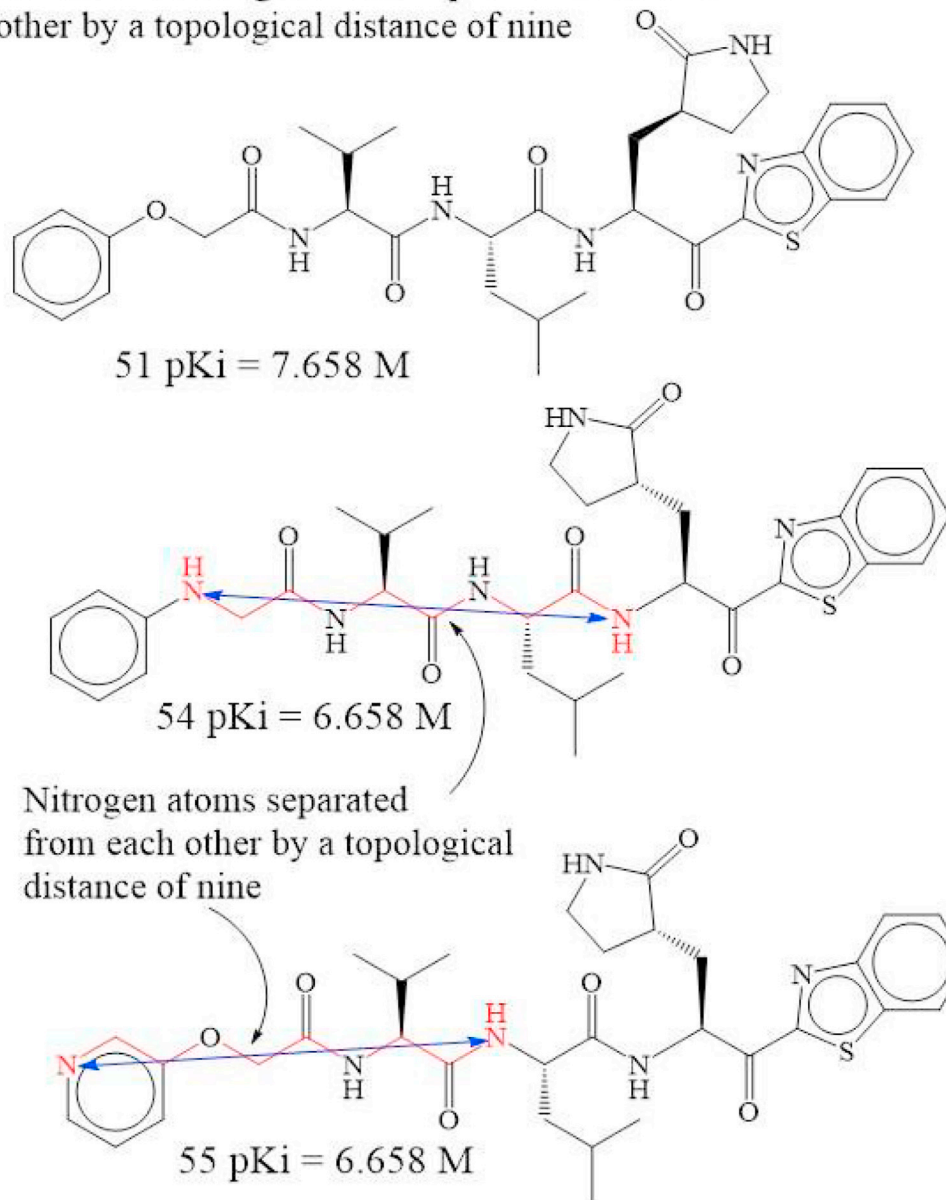


Fig. 6. Pictorial representation of molecular descriptor APC2D9_N_N.

hybridized Carbon atoms exactly at eight bonds from the ring Nitrogen atoms. As the number of such Nitrogen atoms increases, the activity decreases. This observation is supported on comparing molecule 31 (pKi = 7.187 M) and 46 (pKi = 6.167 M). These two molecules are positional isomers of each other, but they have good difference in their activity. This could be attributed to higher frequency of occurrence of sp³-hybridized Carbon atoms exactly at eight bonds from the ring Nitrogen atoms in case of molecule 46 than the molecule 31. This descriptor has been shown in Fig. 8 using molecule number 31 and 46 as representatives. The sp³-hybridized Carbon atoms and ring nitrogen atoms have been highlighted using blue colour, while the eight bonds by red colour.

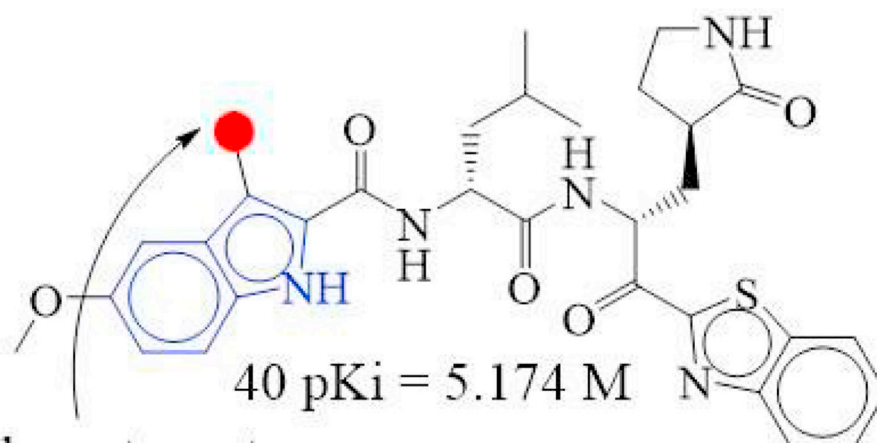
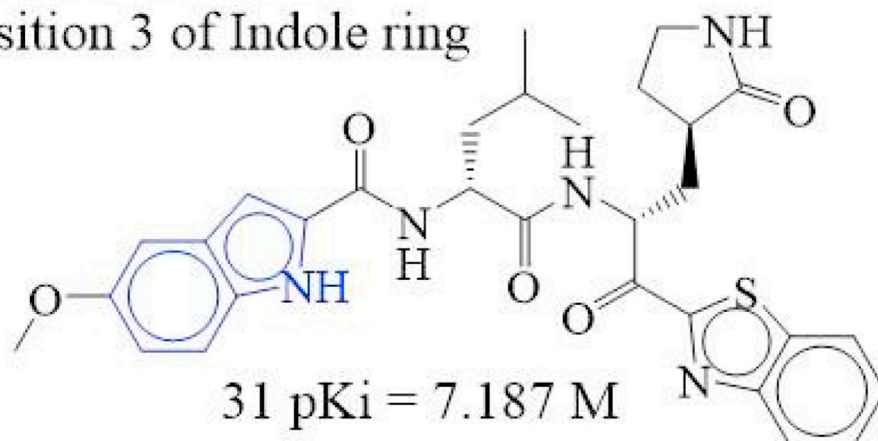
Though, the molecular descriptors APC2D6_C_S (presence of Carbon and Sulfur at a topological distance of six) and ringC_sp3S_9B (number of ring Carbon atoms within nine bonds from sp³-hybridized Sulfur atoms) convey interrelation of Carbon and Sulfur and their subsequent effect on activity, but both molecular descriptors provide different level and type

of information in varying details. Both molecular descriptors have a negative coefficient. Therefore, increasing their value could reduce the anti-SARS activity profile. A comparison of molecule number 42 (pKi = 7.658 M) with 43 (pKi = 6.097 M) vindicates this observation. The molecule 42 has a lower number of ringC_sp3S_9B and a higher value of APC2D3_C_N than 43.

3. Conclusions

In conclusion, statistically robust QSAR models with good external predictive ability have been developed, which have successfully highlighted a good number of molecular features. The developed models 1.1 to 2.2 satisfy the threshold values for many statistical parameters that are necessary to establish the quality and usefulness of a QSAR model. Thus, the developed QSAR models have a good balance of Quantitative and Qualitative aspects. Therefore, the developed models could be useful for

No Carbon atom at
position 3 of Indole ring



Carbon atom at
position 3 of Indole ring

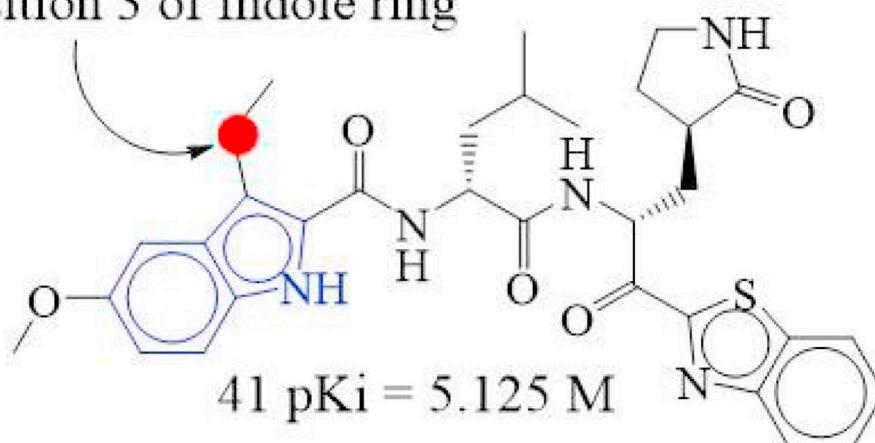


Fig. 7. Representation of molecular descriptor KRFP3478.

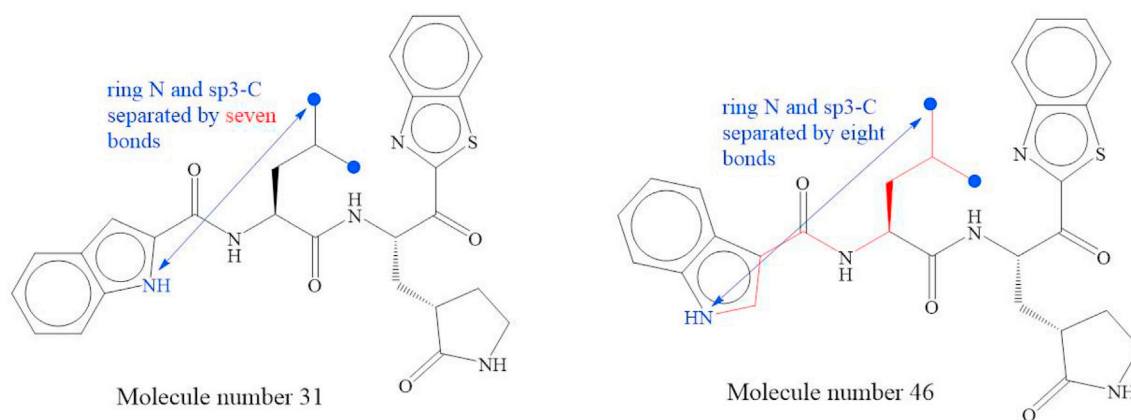


Fig. 8. Depiction of molecular descriptor fringNsp3C8B using 31 and 46 as representatives.

future optimization of the activity profile of the molecules used in the present dataset.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are thankful to TINKER, ChemSketch 12 Freeware (ACD labs), and PaDEL developers for providing the free versions of their software. Authors are thankful to Dr. Paola Gramatica, Italy and her team for providing QSARINS-2.2.2 (www.qsar.it).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemolab.2020.104172>.

References

- Zumla, J.F.W. Chan, E.I. Azhar, D.S.C. Hui, K.-Y. Yuen, Coronaviruses — drug discovery and therapeutic options, *Nat. Rev. Drug Discov.* 15 (2016) 327–347.
- Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S.-H. Jung, An overview of severe acute respiratory syndrome–coronavirus (SARS-CoV) 3CL protease inhibitors: peptidomimetics and small molecule chemotherapy, *J. Med. Chem.* 59 (2016) 6595–6628.
- Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B* 10 (2020) 766–788, <https://doi.org/10.1016/j.apsb.2020.02.008>.
- WHO. <https://covid19.who.int/>, 2020.
- Dhama, K. Sharun, R. Tiwari, M. Dadar, Y.S. Malik, K.P. Singh, W. Chaicumpa, COVID-19, an Emerging Coronavirus Infection: Advances and Prospects in Designing and Developing Vaccines, Immunotherapeutics, and Therapeutics, *Human Vaccines & Immunotherapeutics*, 2020, pp. 1–7.
- Liu, Q. Zhou, Y. Li, L.V. Garner, S.P. Watkins, L.J. Carter, J. Smoot, A.C. Gregg, A.D. Daniels, S. Jervey, D. Albaiu, Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases, *ACS Cent. Sci.* 6 (2020) 315–331.
- Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2, *Cell Discov.* 6 (2020) 14.
- M.H. Baig, K. Ahmad, S. Roy, J.M. Ashraf, M. Adil, M.H. Siddiqui, S. Khan, M.A. Kamal, I. Provaznik, I. Choi, Computer aided drug design: success and limitations, *Curr. Pharmaceut. Des.* 22 (2016) 572–581.
- S.J.Y. Macalino, V. Gosu, S. Hong, S. Choi, Role of computer-aided drug design in modern drug discovery, *Arch Pharm. Res. (Seoul)* 38 (2015) 1686–1701.
- V.H. Masand, N.N.E. El-Sayed, M.U. Bambole, V.R. Patil, S.D. Thakur, Multiple quantitative structure-activity relationships (QSARs) analysis for orally active trypanocidal N-myristoyltransferase inhibitors, *J. Mol. Struct.* 1175 (2019a) 481–487.
- A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, Baskin II, M. Cronin, J. Dearden, P. Gramatica, Y.C. Martin, R. Todeschini, V. Consonni, V.E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha, QSAR modeling: where have you been? Where are you going to? *J. Med. Chem.* 57 (2014) 4977–5010.
- N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection, *J. Chem. Inf. Model.* 52 (2012) 2044–2058.
- T. Fujita, D.A. Winkler, Understanding the roles of the “two QSARs”, *J. Chem. Inf. Model.* 56 (2016) 269–274.
- P. Gramatica, On the development and validation of QSAR models, *Methods Mol. Biol.* 930 (2013) 499–526.
- P. Gramatica, S. Cassani, P.P. Roy, S. Kovarich, C.W. Yap, E. Papa, QSAR modeling is not push a button and find a correlation: a case study of toxicity of (Benzo-) triazoles on algae, *Molecular Informatics* (2012) 817–835.
- J. Huang, X. Fan, Why QSAR fails: an empirical evaluation using conventional computational approach, *Mol. Pharm.* 8 (2011) 600–608.
- 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.
- V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T. Ben Hadda, V. Rastija, A.M. Alfeefy, Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model, *Med. Chem. Res.* 24 (2015b) 1241–1264.
- S. Konno, P. Thanigaimalai, T. Yamamoto, K. Nakada, R. Kakiuchi, K. Takayama, Y. Yamazaki, F. Yakushiji, K. Akaji, Y. Kiso, Y. Kawasaki, S.-E. Chen, E. Freire, Y. Hayashi, Design and synthesis of new tripeptide-type SARS-CoV 3CL protease inhibitors containing an electrophilic arylketone moiety, *Bioorg. Med. Chem.* 21 (2013) 412–424.
- T. Regnier, D. Sarma, K. Hidaka, U. Bacha, E. Freire, Y. Hayashi, Y. Kiso, New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CLpro inhibitors, *Bioorg. Med. Chem. Lett.* 19 (2009) 2722–2727.
- P. Thanigaimalai, S. Konno, T. Yamamoto, Y. Koivai, A. Taguchi, K. Takayama, F. Yakushiji, K. Akaji, S.-E. Chen, A. Naser-Tavakolian, A. Schön, E. Freire, Y. Hayashi, Development of potent dipeptide-type SARS-CoV 3CL protease inhibitors with novel P3 scaffolds: design, synthesis, biological evaluation, and docking studies, *Eur. J. Med. Chem.* 68 (2013a) 372–384.
- P. Thanigaimalai, S. Konno, T. Yamamoto, Y. Koivai, A. Taguchi, K. Takayama, F. Yakushiji, K. Akaji, Y. Kiso, Y. Kawasaki, S.-E. Chen, A. Naser-Tavakolian, A. Schön, E. Freire, Y. Hayashi, Design, synthesis, and biological evaluation of novel dipeptide-type SARS-CoV 3CL protease inhibitors: structure–activity relationship study, *Eur. J. Med. Chem.* 65 (2013b) 436–447.
- Zhang, D. Lin, Y. Kusov, Y. Nian, Q. Ma, J. Wang, A. von Brunn, P. Leyssen, K. Lanko, J. Neyts, A. de Wilde, E.J. Snijder, H. Liu, R. Hilgenfeld, Alpha-ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment, *J. Med. Chem.* 63 (2020) 4562–4578.
- V.H. Masand, V. Rastija, PyDescriptor : a new PyMOL plugin for calculating thousands of easily understandable molecular descriptors, *Chemometr. Intell. Lab. Syst.* 169 (2017) 12–18.
- C.W. Yap, PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints, *J. Comput. Chem.* 32 (2011) 1466–1474.
- P. Gramatica, S. Cassani, N. Chirico, QSARINS-chem: insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS, *J. Comput. Chem.* 35 (2014) 1036–1044.
- P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: a new software for the development, analysis, and validation of QSAR MLR models, *J. Comput. Chem.* 34 (2013) 2121–2132.
- V.H. Masand, D.T. Mahajan, A.M. Alafeefy, S.N. Bukhari, N.N. Elsayed, Optimization of antiproliferative activity of substituted phenyl 4-(2-oxoimidazolidin-1-yl) benzenesulfonates: QSAR and CoMFA analyses, *Eur. J. Pharmaceut. Sci. Off. J. Eur. Federat. Pharma. Sci.* 77 (2015a) 230–237.
- V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T.B. Hadda, V. Rastija, A.M. Alfeefy, Effect of information leakage and method of splitting (rational and random) on

- external predictive ability and behavior of different statistical parameters of QSAR model, *Med. Chem. Res.* 24 (2015c) 1241–1264.
- [30] V.H. Masand, N.N.E. El-Sayed, D.T. Mahajan, A.G. Mercader, A.M. Alafeefy, I.G. Shibi, QSAR modeling for anti-human African trypanosomiasis activity of substituted 2-Phenylimidazopyridines, *J. Mol. Struct.* 1130 (2017) 711–718.
- [31] V.H. Masand, N.N.E. El-Sayed, M.U. Bambole, S.A. Quazi, Multiple QSAR models, pharmacophore pattern and molecular docking analysis for anticancer activity of α , β -unsaturated carbonyl-based compounds, oxime and oxime ether analogues, *J. Mol. Struct.* 1157 (2018) 89–96.
- [32] V.H. Masand, N.N. Elsayed, S.D. Thakur, N. Gawhale, M.M. Rathore, Quinoxalinones based aldose reductase inhibitors: 2D and 3D-QSAR analysis, *Mol. Inf.* (2019b).
- [33] N. Chirico, P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? comparison of different validation criteria and proposal of using the concordance correlation coefficient, *J. Chem. Inf. Model.* 51 (2011) 2320–2335.
- [34] V. Consonni, D. Ballabio, R. Todeschini, Comments on the definition of the Q2 parameter for QSAR validation, *J. Chem. Inf. Model.* 49 (2009) 1669–1678.
- [35] V. Consonni, R. Todeschini, D. Ballabio, F. Grisoni, On the misleading use of Q2F3 for QSAR model comparison, *Mol. Inf.* 38 (2019), e1800029.