**Structure features of peptide-type SARS-CoV main protease inhibitors: Quantitative Structure Activity Relationship study**

Vijay H. Masand*, Ajaykumar Gandhi², Vesna Rastija³, Meghshyam K. Patil⁴

¹ Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati, Maharashtra, India-444 602  
² Department of Chemistry, Government College of Arts and Science, Aurangabad, Maharashtra, India-431 004 (email: gascajay18@gmail.com)  
³ Department of Agroecology and Environmental Protection, Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia (email: vrastija@fazos.hr)  
⁴ Department of Chemistry, Osmanabad Sub-Centre Dr. Babasaheb Ambedkar Marathwada University, Osmanabad, Maharashtra, India

**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. $R^2 = 0.80–0.82$, $Q^2_{loo} = 0.74–0.77$). The developed models identified interrelations of atom pairs as important molecular descriptors. Therefore, the present QSAR models have a good balance of Qualitative and Quantitative approaches, thereby, useful for future modifications of peptide-type compounds for anti-SARS-CoV activity.

**Keywords:** QSAR; COVID-19; SARS-CoV; SARS-CoV-2; peptide-type compounds

**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 Insubria, OECD- Organisation for Economic Co-operation and Development, OFS- Objective Feature Selection, SFS- Subjective Feature Selection

**Introduction:** Coronaviruses (subfamily Coronavirinae, family Coronaviridae, order Nidovirales) have been classified into four genera: Alphacoronavirus, Betacoronavirus,
**Gammacoronavirus** and **Deltacoronavirus** (Zumla et al., 2016). Of these, especially **Betacoronavirus**, have been found to cause respiratory, enteric, hepatic, and neurological diseases in many animals, and also in humans (Pillaiyar et al., 2016). The two **Betacoronaviruses** (βCoVs) viz. severe acute respiratory syndrome CoV (SARS-CoV) and Middle East respiratory syndrome CoV (MERS-CoV) spread in 2003 and 2012, respectively (Pillaiyar et al., 2016; Wu et al., 2020). SARS-CoV and MERS-CoV have fatality rate of 10% and 35%, respectively (Pillaiyar et al., 2016; Wu et al., 2020). Unfortunately, till this date, there is no appropriate treatment for SARS-CoV and MERS-CoV (Pillaiyar et al., 2016; Wu et al., 2020). 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 1,17,000 deaths and 1.8 million confirmed infected cases (WHO, 2020). 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 (Dhama et al., 2020; Liu et al., 2020; Wu et al., 2020; Zhou et al., 2020). This similarity is reflected from the facts that SARS-CoV and SAR-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 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(Dhama et al., 2020; Liu et al., 2020; Wu et al., 2020; Zhou et al., 2020). 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 (Baig et al., 2016; Macalino et al., 2015).

QSAR analysis involves finding mathematical correlation between the 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 (Masand et al., 2019a; Masand et al., 2016; Pourbasheer et al., 2015).

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 (Cherkasov et al., 2014; Chirico and Gramatica, 2012; Fujita and Winkler, 2016; Gramatica, 2013; Gramatica et al., 2012; Huang and Fan, 2011; Martin et al., 2012; Masand et al., 2017b; Masand et al., 2014; Masand et al., 2015b).

A good number of researchers have synthesized peptide-type compounds and tested against SAR-CoV (Konno et al., 2013; Regnier et al., 2009; Thanigaimalai et al., 2013a; Thanigaimalai et al., 2013b). Recently, Zhang et al. (Zhang et al., 2020) 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.

**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 (Konno et al., 2013; Thanigaimalai et al., 2013a; Thanigaimalai et al., 2013b). The selected peptide-type compounds possess Ki = 3 to 56,000 nM. The reported activity Ki values were converted to pKi (pKi = –logKi) 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 figure 1. The SMILES notation for all selected molecules along with their reported activity values Ki, and pKi are present in the Table S1 in supplementary material.

![Figure 1. Variations in activity and chemical structure in the present dataset](image)

**Structure optimization and molecular descriptor calculation:**

The structures were drawn using ChekSketch 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* (Masand and Rastija, 2017) and PaDEL (Yap, 2011).

**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 (Gramatica et al., 2014; Gramatica et al., 2013; Masand et al., 2015a; Masand et al., 2015c). 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 for all the four sets 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 (Gramatica et al., 2014; Gramatica et al., 2013; Masand et al., 2017a). 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. Therefore,
molecular descriptor selection was restricted to a set of six descriptors to avoid overfitting 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(Masand et al., 2019a; Masand et al., 2018; Masand et al., 2019b): $R^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $R^2_{tr} > Q^2_{tr}$, $R^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $Q^2-F_n \geq 0.60$, $r^2_m \geq 0.6$, $(1-r^2/r^2_o) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1-r^2/r^2_o) < 0.1$, $0.9 \leq k' \leq 1.1$, $|r^2_o - r^2_o'| < 0.3$ 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 work.

Result and Discussions:

Though, the dataset used in the present study is moderate sized 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 (Masand et al., 2015c), 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 bioactivity 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):
\[
p_{Ki} = 1.781 \pm 1.357 + 1.073 \pm 0.317 \times \text{sp2O}_\text{aroC}_7B + 0.377 \pm 0.111 \times \text{APC2D3}_C_N -1.264 \pm 0.647 \times \text{APC2D9}_N_N -2.538 \pm 0.751 \times KRFPC3478 -0.743 \pm 0.233 \times \text{fringNsp3C8B} -0.61 \pm 0.246 \times \text{APC2D6}_C_S
\]

Model-1.2 (Divided Set model: - Training: 80%, Prediction: 20%):
\[
p_{Ki} = 1.7 \pm 1.658 + 1.182 \pm 0.353 \times \text{sp2O}_\text{aroC}_7B + 0.363 \pm 0.115 \times \text{APC2D3}_C_N -1.351 \pm 0.642 \times \text{APC2D9}_N_N -2.422 \pm 0.731 \times KRFPC3478 -0.772 \pm 0.269 \times \text{fringNsp3C8B} -0.169 \pm 0.085 \times \text{ringC}_\text{sp3S}_9B
\]
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.

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     |
| \(K_{xx}\)            | 0.241     | 0.271     |
| \(\Delta K\)          | 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**|           |           |
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, \( CCC_{cv} \), \( R^2_{adj} \) 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, MAE, \( CCC_{cv} \), 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 \( R^2_{ex} \), \( Q^2_{F1} \), \( Q^2_{F2} \), \( Q^2_{F3} \), and \( CCC_{ex} \).
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 figure 2). 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).

**Interpretation of QSAR models:**

The models 1.1 and 1.2 have been built using the undivided and 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_arO_C_7B (number of sp2 hybridized Oxygen atoms within seven bonds from aromatic Carbon atoms). Since, in the present series of compounds, sp2 hybridized Oxygen atoms are always present as a part of carbonyl group (\(>\text{C}=\text{O}\)), 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 figure 3. The sp2-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 figure 4.
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. Therefore, such a combination of Nitrogen atoms should be avoided for better activity.
KRFPC3478 is a fingerprint 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.
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 sp3-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 figure 7 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.

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

**Acknowledgments:** 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).

**References:**
Baig, M.H., Ahmad, K., Roy, S., Ashraf, J.M., Adil, M., Siddiqui, M.H., Khan, S., Kamal, M.A., Provaznik, I., Choi, I., 2016. Computer Aided Drug Design: Success and Limitations. Curr Pharm Des 22, 572-581.

Cherkasov, A., Muratov, E.N., Fourches, D., Varnek, A., Baskin, II, Cronin, M., Dearden, J., Gramatica, P., Martin, Y.C., Todeschini, R., Consonni, V., Kuz'min, V.E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., 2014. QSAR modeling: where have you been? Where are you going to? J Med Chem 57, 4977-5010.

Chirico, N., Gramatica, P., 2011. 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, 2320-2335.

Chirico, N., Gramatica, P., 2012. 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, 2044-2058.

Consonni, V., Ballabio, D., Todeschini, R., 2009. Comments on the definition of the Q2 parameter for QSAR validation. J Chem Inf Model 49, 1669-1678.

Consonni, V., Todeschini, R., Ballabio, D., Grisoni, F., 2019. On the Misleading Use of QF32 for QSAR Model Comparison. Molecular Informatics 38.

Dhama, K., Sharun, K., Tiwari, R., Dadar, M., Malik, Y.S., Singh, K.P., Chaicumpa, W., 2020. COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics. Human Vaccines & Immunotherapeutics, 1-7.

Fujita, T., Winkler, D.A., 2016. Understanding the Roles of the “Two QSARs”. Journal of Chemical Information and Modeling 56, 269-274.

Gramatica, P., 2013. On the development and validation of QSAR models. Methods Mol Biol 930, 499-526.

Gramatica, P., Cassani, S., Chirico, N., 2014. QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. Journal of Computational Chemistry 35, 1036-1044.

Gramatica, P., Cassani, S., Roy, P.P., Kovarich, S., Yap, C.W., Papa, E., 2012. QSAR Modeling is not Push a Button and Find a Correlation: A Case Study of Toxicity of (Benzo-)triazoles on Algae, Molecular Informatics, pp. 817-835.
Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. Journal of Computational Chemistry 34, 2121-2132.
Huang, J., Fan, X., 2011. Why QSAR fails: an empirical evaluation using conventional computational approach. Mol Pharm 8, 600-608.
Konno, S., Thanigaimalai, P., Yamamoto, T., Nakada, K., Kakiuchi, R., Takayama, K., Yamazaki, Y., Yakushiji, F., Akaji, K., Kiso, Y., Kawasaki, Y., Chen, S.-E., Freire, E., Hayashi, Y., 2013. Design and synthesis of new tripeptide-type SARS-CoV 3CL protease inhibitors containing an electrophilic arylketone moiety. Bioorganic & Medicinal Chemistry 21, 412-424.
Liu, C., Zhou, Q., Li, Y., Garner, L.V., Watkins, S.P., Carter, L.J., Smoot, J., Gregg, A.C., Daniels, A.D., Jervey, S., Albaiu, D., 2020. Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases. ACS Central Science 6, 315-331.
Macalino, S.J.Y., Gosu, V., Hong, S., Choi, S., 2015. Role of computer-aided drug design in modern drug discovery. Archives of Pharmacal Research 38, 1686-1701.
Martin, T.M., Harten, P., Young, D.M., Muratov, E.N., Golbraikh, A., Zhu, H., Tropsha, A., 2012. Does rational selection of training and test sets improve the outcome of QSAR modeling? J Chem Inf Model 52, 2570-2578.
Masand, V.H., El-Sayed, N.N.E., Bambole, M.U., Patil, V.R., Thakur, S.D., 2019a. Multiple quantitative structure-activity relationships (QSARs) analysis for orally active trypanocidal N-myristoyltransferase inhibitors. Journal of Molecular Structure 1175, 481-487.
Masand, V.H., El-Sayed, N.N.E., Bambole, M.U., Quazi, S.A., 2018. Multiple QSAR models, pharmacophore pattern and molecular docking analysis for anticancer activity of α, β-unsaturated carbonyl-based compounds, oxime and oxime ether analogues. Journal of Molecular Structure 1157, 89-96.
Masand, V.H., El-Sayed, N.N.E., Mahajan, D.T., Mercader, A.G., Alafeefy, A.M., Shibi, I.G., 2017a. QSAR modeling for anti-human African trypanosomiasis activity of substituted 2-Phenylimidazopyridines. Journal of Molecular Structure 1130, 711-718.
Masand, V.H., El-Sayed, N.N.E., Mahajan, D.T., Rastija, V., 2017b. QSAR analysis for 6-arylpyrazine-2-carboxamides as Trypanosoma brucei inhibitors. SAR and QSAR in Environmental Research 28, 165-177.
Masand, V.H., Elsayed, N.N., Thakur, S.D., Gawhale, N., Rathore, M.M., 2019b. Quinoxalinones Based Aldose Reductase Inhibitors: 2D and 3D-QSAR Analysis. Molecular Informatics.

Masand, V.H., Mahajan, D.T., Alafeefy, A.M., Bukhari, S.N., Elsayed, N.N., 2015a. Optimization of antiproliferative activity of substituted phenyl 4-(2-oxoimidazolidin-1-yl) benzenesulfonates: QSAR and CoMFA analyses. European journal of pharmaceutical sciences : official journal of the European Federation for Pharmaceutical Sciences 77, 230-237.

Masand, V.H., Mahajan, D.T., Gramatica, P., Barlow, J., 2014. Tautomerism and multiple modelling enhance the efficacy of QSAR: antimalarial activity of phosphoramidate and phosphorothioamidate analogues of amiprophos methyl. Medicinal Chemistry Research 23, 4825-4835.

Masand, V.H., Mahajan, D.T., Maldhure, A.K., Rastija, V., 2016. Quantitative structure–activity relationships (QSARs) and pharmacophore modeling for human African trypanosomiasis (HAT) activity of pyridyl benzamides and 3-(oxazolo[4,5-b]pyridin-2-yl)anilides. Medicinal Chemistry Research 25, 2324-2334.

Masand, V.H., Mahajan, D.T., Nazeruddin, G.M., Ben Hadda, T., Rastija, V., Alfeefy, A.M., 2015b. Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. Medicinal Chemistry Research 24, 1241-1264.

Masand, V.H., Mahajan, D.T., Nazeruddin, G.M., Hadda, T.B., Rastija, V., Alfeefy, A.M., 2015c. Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. Medicinal Chemistry Research 24, 1241-1264.

Masand, V.H., Rastija, V., 2017. PyDescriptor : A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors. Chemometrics and Intelligent Laboratory Systems 169, 12-18.

Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., Jung, S.-H., 2016. An Overview of Severe Acute Respiratory Syndrome–Coronavirus (SARS-CoV) 3CL Protease Inhibitors: Peptidomimetics and Small Molecule Chemotherapy. Journal of Medicinal Chemistry 59, 6595-6628.

Pourbasheer, E., Shokouhi Tabar, S., Masand, V.H., Aalizadeh, R., Ganjali, M.R., 2015. 3D-QSAR and docking studies on adenosine A2A receptor antagonists by the CoMFA method. SAR QSAR Environ Res 26, 461-477.
Regnier, T., Sarma, D., Hidaka, K., Bacha, U., Freire, E., Hayashi, Y., Kiso, Y., 2009. New developments for the design, synthesis and biological evaluation of potent SARS-CoV 3CLpro inhibitors. Bioorganic & Medicinal Chemistry Letters 19, 2722-2727.

Thanigaimalai, P., Konno, S., Yamamoto, T., Koiwai, Y., Taguchi, A., Takayama, K., Yakushiji, F., Akaji, K., Chen, S.-E., Naser-Tavakolian, A., Schön, A., Freire, E., Hayashi, Y., 2013a. Development of potent dipeptide-type SARS-CoV 3CL protease inhibitors with novel P3 scaffolds: Design, synthesis, biological evaluation, and docking studies. European Journal of Medicinal Chemistry 68, 372-384.

Thanigaimalai, P., Konno, S., Yamamoto, T., Koiwai, Y., Taguchi, A., Takayama, K., Yakushiji, F., Akaji, K., Kiso, Y., Kawasaki, Y., Chen, S.-E., Naser-Tavakolian, A., Schön, A., Freire, E., Hayashi, Y., 2013b. Design, synthesis, and biological evaluation of novel dipeptide-type SARS-CoV 3CL protease inhibitors: Structure–activity relationship study. European Journal of Medicinal Chemistry 65, 436-447.

WHO, 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200414-sitrep-85-covid-19.pdf?sfvrsn=7b8629bb_4.

Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., Zheng, M., Chen, L., Li, H., 2020. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharmaceutica Sinica B.

Yap, C.W., 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of Computational Chemistry 32, 1466-1474.

Zhang, L., Lin, D., Kusov, Y., Nian, Y., Ma, Q., Wang, J., von Brunn, A., Leyssen, P., Lanko, K., Neyts, J., de Wilde, A., Snijder, E.J., Liu, H., Hilgenfeld, R., 2020. α-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. Journal of Medicinal Chemistry.

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F., 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discovery 6.

Zumla, A., Chan, J.F.W., Azhar, E.I., Hui, D.S.C., Yuen, K.-Y., 2016. Coronaviruses — drug discovery and therapeutic options. Nature Reviews Drug Discovery 15, 327-347.