Extending the identification of structural features responsible for anti-SARS-CoV activity of peptide-type compounds using QSAR modelling

V.H. Masand , V. Rastija , M.K. Patil , A. Gandhi & A. Chapolikar

View supplementary material

Published online: 27 Aug 2020.

Submit your article to this journal

Article views: 996

View related articles

View Crossmark data

Citing articles: 1 View citing articles

Taylor & Francis
Taylor & Francis Group

Extending the identification of structural features responsible for anti-SARS-CoV activity of peptide-type compounds using QSAR modelling

V.H. Masand[a], V. Rastija[b], M.K. Patil[c], A. Gandhi[d] and A. Chapolikar[d]

[a]Department of Chemistry, Vidya Bharati Mahavidyalaya, Amravati, India; [b]Department of Chemistry, Faculty of Agrobiotechnical Sciences, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia; [c]Department of Chemistry, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India; [d]Department of Chemistry, Government College of Arts and Science, Aurangabad, India

ABSTRACT

A quantitative structure–activity relationship (QSAR) model was built from a dataset of 54 peptide-type compounds as SARS-CoV inhibitors. The analysis was executed to identify prominent and hidden structural features that govern anti-SARS-CoV activity. The QSAR model was derived from the genetic algorithm–multi-linear regression (GA-MLR) methodology. This resulted in the generation of a statistically robust and highly predictive model. In addition, it satisfied the OECD principles for QSAR validation. The model was validated thoroughly and fulfilled the threshold values of a battery of statistical parameters (e.g. $r^2 = 0.87$, $Q^2_{loo} = 0.82$). The derived model is successful in identifying many atom-pairs as important structural features that govern the anti-SARS-CoV activity of peptide-type compounds. The newly developed model has a good balance of descriptive and statistical approaches. Consequently, the present work is useful for future modifications of peptide-type compounds for SARS-CoV and SARS-CoV-2 activity.

Introduction

Coronaviruses, which generally infect the respiratory system of humans, other mammals and birds, belong to a large group of enveloped non-segmented positive-sense, single-stranded RNA viruses with unusually large genomes. These are identified on the basis of club-like spikes, which project from their surface. They also possess a unique replication strategy [1–5]. When these viruses were first observed under the electron microscope, the characteristic appearance of a fringe of large and bulbous surface projections, reminiscent of the 'crown', was found. Hence, the name 'coronaviruses' was given to them [6]. Later, in 1975, the International Committee on the Taxonomy of Viruses established the *Coronaviridae* family [7]. Coronaviruses are responsible for various clinical manifestations such as acute and chronic respiratory, enteric, gastrointestinal, liver, hepatic, and central nervous system diseases [8,9]. The emergence of severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and Middle East respiratory syndrome coronavirus (MERS-CoV) caused global outbreaks [10,11]. It

Supplemental data for this article can be accessed at https://doi.org/10.1080/1062936X.2020.1784271

was found that these two viruses were transmitted from other animals to humans [12]. Another four human coronaviruses, HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HUK1, are responsible for the common cold, but may cause severe lower respiratory tract infections in patients with underlying medical conditions, and in young children and aged people. In this family, a seventh human-infecting coronavirus, SARS-CoV-2, also known as novel coronavirus, has been reported from China [13]. As of 8 April 2020, officially 1,519,012 cases have been diagnosed and 88,526 deaths due to SARS-CoV-2 have been reported in 225 countries. The most affected countries are the United States, Spain, Italy, Germany, France, China, Iran, and the United Kingdom. The disease caused by SARS-CoV-2 virus is now known as COVID-19 [14].

SARS-CoV-2 is transmitted by contact or respiratory droplets. Thereby, personal hygiene and physical distancing are recommended public health measures to stop further transmission or spread of this virus [15,16]. Unfortunately, the high rate of transmission and relatively long incubation period create a higher risk of transmission of COVID-19 [16,17]. Therefore, there is an urgent need to develop a cheap and efficient drug for this deadly disease.

A good strategy to combat COVID-19 is to leverage previously reported drugs or compounds that have activity against similar coronaviruses. It has been established that the viral proteins necessary for SARS-CoV-2 to enter into host cells and consequently replicate are very much similar to those linked with SARS-CoV. In addition, synthesis and bio-screening of a new drug and its analogues need a lot of time. Therefore, utilising the high resemblance of SARS-CoV-2 with SARS-CoV is a good approach to develop a drug candidate. Thus, a therapeutic or preventive agent tested against SARS-CoV could be useful for the development of a drug for COVID-19. It is rational to believe that a therapeutic agent that has been earlier tested for anti-SARS-CoV activity could be a good candidate to be effective against SARS-CoV-2. Some previously reported SARS-CoV inhibitors are shown in Figure 1. In short, a faster, inexpensive and effective strategy is to leverage drugs and biomolecules previously reported to have activity against related coronaviruses such as SARS-CoV.

Computer-aided drug designing is an economical and viable solution to achieve these goals, with a good number of benefits such as minimal animal testing, reduced trial and error,
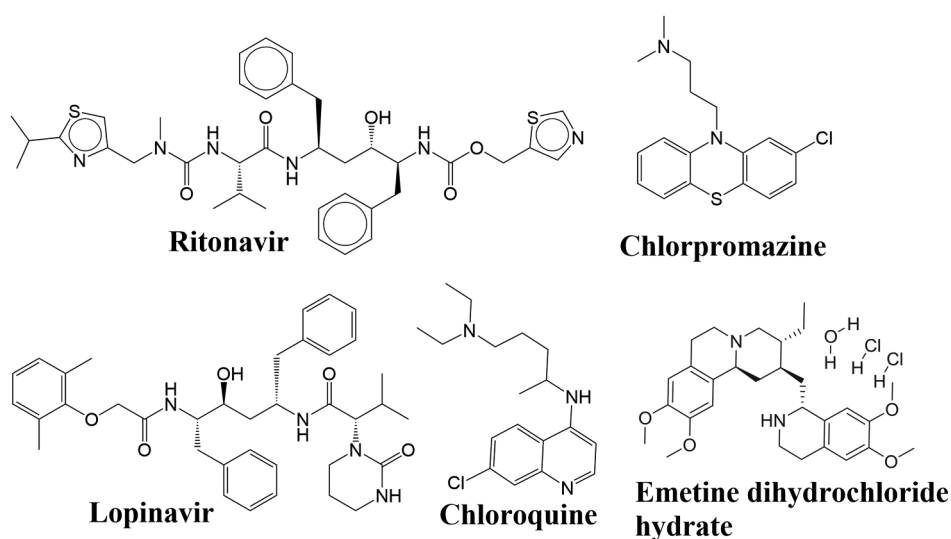

Figure 1. Different inhibitors reported for SARS-CoV.

and time saving [18]. It has gained popularity among researchers as a result of the significant contribution made by its thriving branches including QSAR (quantitative structure–activity relationship), molecular docking, pharmacophore modelling, etc., in development of lead and drug candidates.

QSAR analysis comprises identification of relationships between the structural features of analogous molecules with bio-activity using statistical methods. A typical QSAR analysis involves the use of different branches of science such as chemistry, computer science, mathematics, statistics and biology. A general protocol for QSAR analysis involves the following steps [19–22]: (1) Collection of molecules (referred as a dataset) that have the desired activity/property; (2) drawing of structures and optimization to 3D using a suitable method; (3) calculation of a large number of molecular descriptors, followed by their pruning using a suitable statistical method; (4) generation of a QSAR model using an appropriate feature (that is, molecular descriptor) selection algorithm; and (5) appropriate validation of the developed QSAR model.

A main objective for executing a QSAR analysis is to identify salient and concealed structural patterns/features that have a relationship with the desired activity/property (i.e. descriptive QSAR). Another objective of QSAR analysis is to help to predict the activity/ property before the wet lab synthesis and bio-testing of a molecule (i.e. statistical QSAR) [23]. Nowadays, essential descriptive and statistical aspects of the congeneric molecules for future optimization are regularly obtained. A QSAR model that has a good balance of descriptive and statistical aspects not only provides additional knowledge about the structural patterns that have good relationship with the desired activity/property of a drug candidate, but also improves insight into the mechanism of drug action.

Peptide-type compounds have received the attention of researchers because of their activity against SAR-CoV [24–27]. Consequently, a large number of peptide-type compounds have been synthesized and tested. Recently, α-Ketoamides (peptide-type analogues, which are used in the present work) were synthesized and screened for inhibitory activity against SARS-CoV and SARS-CoV-2 [28,29]. One of the compounds, 13b, revealed equivalent activity against SARS-CoV (IC_{50} = 0.90 μM) and SARS-CoV-2 (IC_{50} = 0.67 μM) [28]. Despite these efforts, further optimizations of peptide-type compounds are necessary to obtain a drug candidate. Moreover, as the analogues of peptide-type compounds used in the present work have a variety of substituents as well as positional isomers, it is necessary that important hidden relationships of structural features, which cannot be identified by visual inspection during structure–activity relationship analysis, should be identified using QSAR or a similar technique.

Therefore, in the present work, we have developed a QSAR model for a series of 54 peptide-type compounds for their anti-SARS-CoV activity. The outcomes could be useful to optimize peptide-type compounds for a better activity profile for SARS-COV-2 as well.

Materials and methods

Dataset selection

A dataset of 54 peptide-type compounds with moderate to high activity against main protease (MPro) of SARS-CoV has been selected for the present work [25–27,29]. The selected peptide-type compounds have IC_{50} values ranging from 0.23 to 279 μM. The
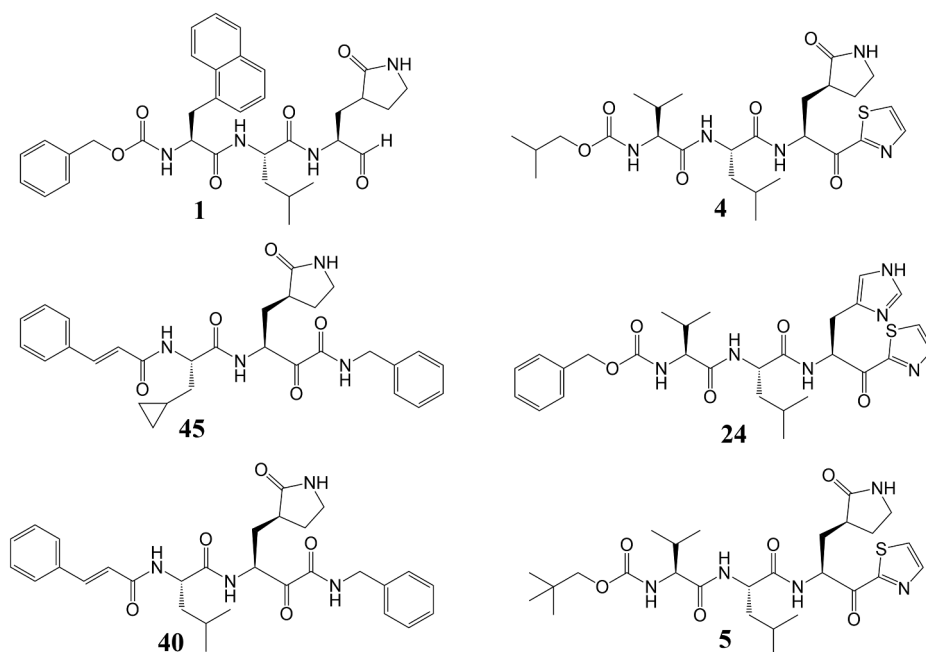
Figure 2. Variations in activity and chemical structure in the studied dataset of peptide-type compounds.

reported activity IC_{50} values were converted to pIC_{50} ($pIC_{50} = -\log IC_{50}$) before QSAR analysis. For the sake of simplicity, understanding and to demonstrate the chemical space covered by the present dataset, the three most and least active molecules are shown in Figure 2. The SMILES notation for all the molecules along with their reported activity values IC_{50}, and pIC_{50} are presented Table S1 (supplementary material).

Structure drawing, optimization and molecular descriptor calculation

ChemSketch 12 Freeware (www.acdlabs.com) and OpenBabel 2.4 [30] were used for structure drawing and conversion to 3D structures, respectively. Then, the force field MMFF94 available in TINKER [31] was used to optimize the structures' default settings. Open3DAlign [32] was executed for aligning all the molecules in the dataset. In the next step, PyDescriptor [33] and PaDEL [34] were used for calculation of molecular descriptors.

Molecular descriptor pruning

Molecular descriptor pruning was essential as PyDescriptor and PaDEL provide more than 30,000 molecular descriptors including 1D to 3D molecular descriptors such as charge, atom-pair, fingerprint descriptors, etc., for each molecule. For this, objective feature selection in QSARINS ver. 2.2.2 [35–38] was performed. As a rule, molecular descriptors with high co-linearity ($|r|>0.90$) and nearly constant ($> 95\%$) were excluded. This step is necessary to avoid the inclusion of multi-collinear and spurious variables in the genetic

algorithm–multi-linear regression (GA-MLR) model. The resultant molecular descriptor set comprised 603 molecular descriptors only but adequate to cover 1D- to 3D-descriptor space with the presence of fingerprint, atom-pair and other molecular descriptors.

QSAR model building and validation

The statistically robust GA-MLR-based QSAR model was derived using QSARINS ver. 2.2.2. The derived model was subjected to thorough statistical validation (internal and external), Y-randomization and applicability domain analysis as per the OECD principles. The general procedure for building the QSAR model is as follows [19–21,35,39]:

(1) For the QSAR model derived using divided dataset, the dataset was split randomly, using the random splitting option in QSARINS, into a training and a prediction set of 44 (i.e. 80% training set) and 10 (i.e. 20% prediction set), respectively. After that, the training set was used for model development, and the prediction set for the external validation.

(2) QSARINS 2.2.2 was used to build the GA-MLR-based QSAR model using default settings using Q^2_{LOO} as a selected fitness function for feature selection and to ensure double cross-validation. During model development it was observed that there was an increase in the value of Q^2_{LOO} up to six variables, but after that it had only minor augmentation (Figure 3). Consequently, molecular descriptor selection was limited to a set of six descriptors. This helped to circumvent over-fitting and derive an easy and informative QSAR model.

The values for these selected molecular descriptors, which are present in QSAR model, are available in the supplementary information.
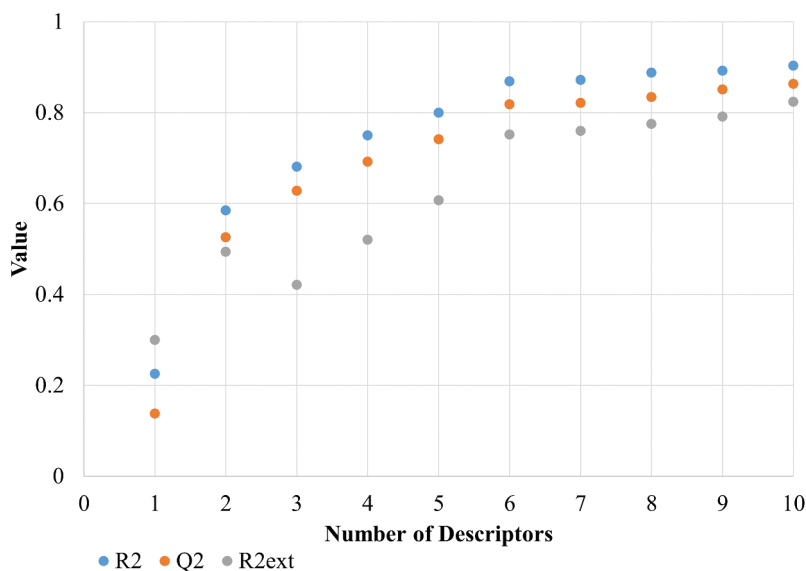

Figure 3. Correlation of statistical parameters with number of molecular descriptors.

(1) The model was subjected to internal and external validation, Y-randomization and model applicability domain (AD) analysis using QSARINS 2.2.2 as per OECD guidelines. This ensured the proper validation of the model. 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 values for the statistical parameters: $r^2_{tr} \geq 0.6$, $Q^2_{loo} \geq 0.5$, $Q^2_{LMO} \geq 0.6$, $r^2 > Q^2$, $r^2_{ex} \geq 0.6$, $RMSE_{tr} < RMSE_{cv}$, $\Delta K \geq 0.05$, $CCC \geq 0.80$, $Q^2\text{-}F^n \geq 0.60$, $r^2_m \geq 0.6$, $(1\text{-}r^2/r_o^2) < 0.1$, $0.9 \leq k \leq 1.1$ or $(1\text{-}r^2/r_o'^2) < 0.1$, $0.9 \leq k' \leq 1.1$, $| r_o^2 - r_o'^2 | < 0.3$ with $RMSE$ and MAE close to zero. Any QSAR model that did not satisfy any of these criteria was therefore omitted.

Results and discussion

The dataset used in the present study is small in size; however, the presence of a variety of substituents, different rings, etc., significantly increased the chemical space covered by the peptide-type compounds. In our previous work on QSAR analysis of small and moderate-sized datasets [39], we have established that the composition of the training and prediction set is very important. The derived QSAR model is as follows:

pIC_{50} = −0.951 (±0.928) + 0.192 (±0.16) *com_sp3C_2A* + 0.058 (±0.019) *notringCplus_SASA* + 0.15 (±0.032) *SubFPC274* + 0.554 (±0.082) *KRFPC3733* − 1.141 (±0.247) *fringNsp2O9B* − 0.398 (±0.156) *fsp2OringC3B*

r^2_{tr} = 0.869, $r^2_{adj.}$ = 0.848, $r^2_{tr} - r^2_{adj.}$ = 0.021, LOF = 0.134, Kxx = 0.283, ΔK = 0.035, $RMSE_{tr}$ = 0.266, MAE_{tr} = 0.226, RSS_{tr} = 3.111, CCC_{tr} = 0.93, s = 0.29, F = 40.835, r^2_{cv} (Q^2_{loo}) = 0.818, r^2-r^2_{cv} = 0.051, $RMSE_{cv}$ = 0.313, MAE_{cv} = 0.268, $PRESS_{cv}$ = 4.315, CCC_{cv} = 0.903, Q^2_{LMO} = 0.803, r^2_{Yscr} = 0.138, Q^2_{Yscr} = −0.23, $RMSE_{ex}$ = 0.424, MAE_{ex} = 0.361, $PRESS_{ext}$ = 1.796, r^2_{ex} = 0.752, Q^2-F^1 = 0.745, Q^2-F^2 = 0.744, Q^2-F^3 = 0.667, CCC_{ex} = 0.867, r^2-ExPy = 0.752, $r_o'^2$ = 0.722, k' = 0.996, 1-$(r^2/r_o'^2)$ = 0.04, r'^2_m = 0.622, r_o^2 = 0.744, k = 0.998, 1-$(r^2$-ExPy$/r_o^2)$ = 0.011, r^2_m = 0.685

The symbols have their usual meaning; the details for same are available in the supplementary material. Many statistical parameters have been computed for the derived model, which are associated to fitting, internal and external validation and Y-scrambling. It is clear that the model satisfies the recommended threshold value for r^2_{tr}, CCC_{tr}, $r^2_{adj.}$ and F, which confirms that the QSAR model is statistically acceptable with a sufficient number of molecular descriptors in it [40–42]. The values for various internal validation parameters such as r^2_{cv}, CCC_{cv}, $RMSE_{cv}$, MAE_{cv}, CCC_{cv}, and Q^2_{LMO} vindicate the statistical robustness of the QSAR model. The high values of r^2_{ex}, Q^2F^1, Q^2F^2, Q^2F^3, and CCC_{ex} confirm the external predictive ability of the model [42–48].

The derived QSAR model satisfies the suggested threshold values for a good number of internal and external validation parameters. In addition, the model AD was ensured by plotting the Williams plot (Figure 4). Hence, the model is statistically robust and has good external predictive ability. Moreover, fulfilment of recommended threshold values for many parameters as well as low r^2 value after Y-scrambling indicate that the model is not developed by chance (see Figure S1 in supplementary information).
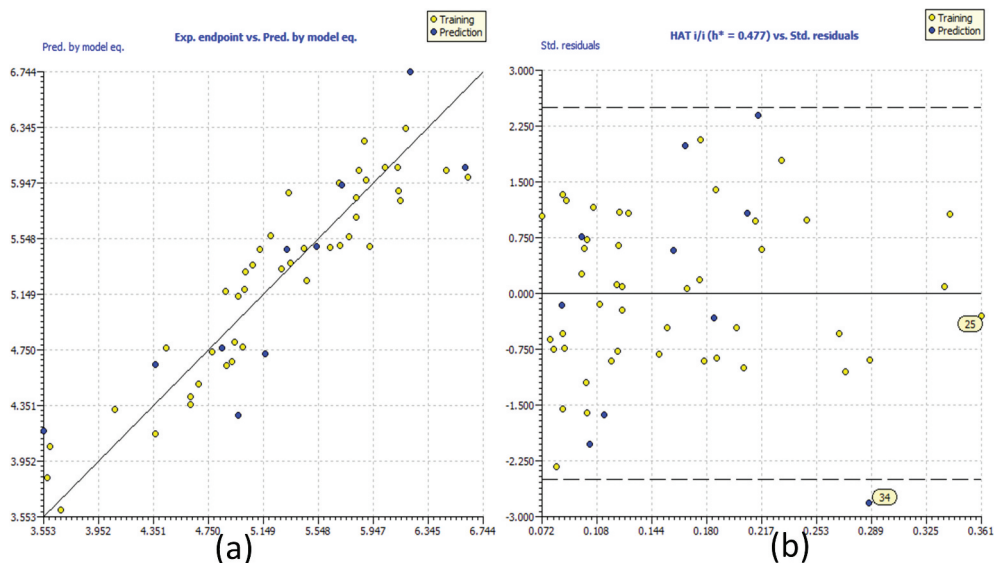
Figure 4. (a) Graph of experimental vs. predicted pIC_{50} values; (b) Williams plot.

QSAR model interpretation

Although 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 unknown factors or other molecular descriptors could have a significant influence on the activity profile of the compounds.

The descriptor KRFPC3733, which represents the fragment -C-C-C-C-N- (moiety A), has a positive coefficient in the model. This indicates that the presence of moiety A in a molecule increases the activity. This observation is supported when molecule 24 (pIC_{50} = 3.585) is compared with 29 (pIC_{50} = 4.365), 28 (pIC_{50} = 6.215) and 40
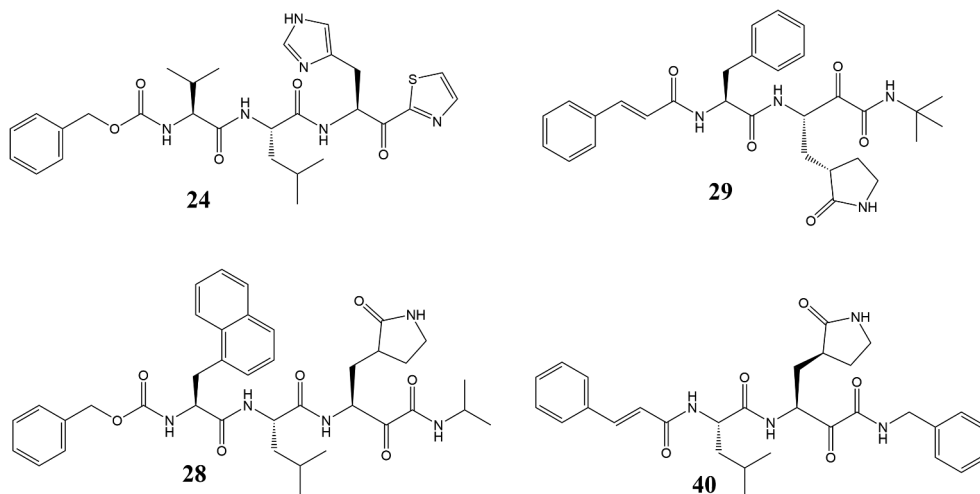

Figure 5. Comparison of molecules with respect to KRFPC3733.

(pIC$_{50}$ = 6.481) (Figure 5). Another such comparison is possible between pairs of compounds: molecule number 37 (pIC$_{50}$ = 5.014) with 21 (pIC$_{50}$ = 5.022) and molecule number 30 (pIC$_{50}$ = 4.944) with 29 (pIC$_{50}$ = 4.365).

The descriptor notringCplus_SASA, which represents solvent accessible surface area due to non-ring carbon atoms, is present with a positive coefficient. Therefore, a higher value could lead to better activity. This observation is supported by the following pairs of molecules: 33 (pIC$_{50}$ = 5.445) with 27 (pIC$_{50}$ = 7.71) and 6 (pIC$_{50}$ = 4.886) with 8 (pIC$_{50}$ = 5.167).

fringNsp2O9B represents the number of sp^2-hybridized oxygen atoms exactly at nine bonds from ring nitrogen atoms. Since, in the present series of peptide-type molecules, sp^2-hybridized oxygen atoms are present due to the presence of carbonyl group (>C = O), it is therefore rational to consider that this molecular descriptor also points out towards the presence of number of carbonyl groups in conjugation with ring nitrogen atoms. It has negative coefficient. Therefore, as the value of such oxygen atoms increases in a molecule, the activity decreases.

The descriptor fsp2OringC3B, which characterizes the number of ring carbon atoms at exactly three bonds from sp^2-hybridized oxygen atom, provides a different level and type of useful information. As stated earlier, the presence of sp^2-hybridized oxygen atoms signifies the presence of a carbonyl group (>C = O), therefore it is rational to consider that this molecular descriptor also points towards the presence of number of carbonyl groups in connection with aromatic carbon atoms. It has a negative coefficient in the model, which means increasing its value could cause a lower activity profile. This is supported when we compare molecule 8 (pIC$_{50}$ = 5.167) with 11 (pIC$_{50}$ = 5.638), and 2 (pIC$_{50}$ = 4.444) with 18 (pIC$_{50}$ = 5.538) (Figure 6).

A similar molecular descriptor with positive correlation with activity is SubFPC274, which symbolizes the presence of aromatic rings. It has a positive coefficient, meaning that the higher the value, the better the activity. In short, the presence of aromatic rings is beneficial for activity.

The descriptor com_sp3C_2A denotes the number of sp^3-hybrizdized carbon atoms within 2 Angstrom from the centre of mass of a molecule. This descriptor has a positive

sp3-Carbon at three bonds from non-ring Oxygen atom

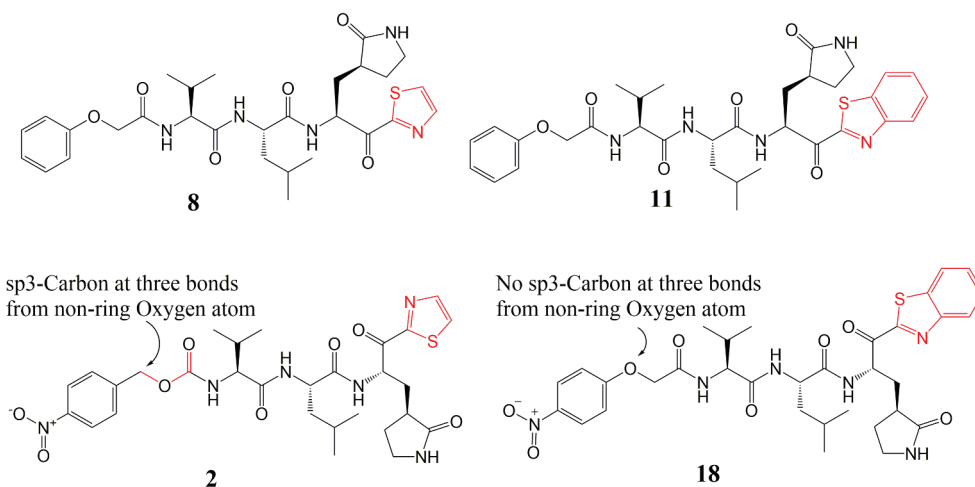No sp3-Carbon at three bonds from non-ring Oxygen atom

Figure 6. Pictorial depiction of molecular descriptor fsp2OringC3B.

coefficient, which indicates that as the number of such carbon atoms increases, the activity also increases. This descriptor has been shown in Figure S2 in the supplementary material.

The different molecular descriptors representing different structural features have provided meaningful insights into the reasons for differences in the anti-SARS-CoV activity of peptide-type compounds. Meanwhile, it is essential to accept that none of the descriptors alone could explain the observed distribution of anti-SARS-CoV activities. Thus, each individual model depends on the combined use of all selected molecular descriptors.

Conclusion

In the present work, the developed model has a good balance of external predictive ability (statistical QSAR), which is indicated by high value of r^2_{tr} = 0.869, r^2_{cv} (Q^2loo) = 0.818, r^2_{ex} = 0.752, and Q^2-F^3 = 0.667. The model also has descriptive ability in terms of molecular features (descriptive QSAR). The developed model is successful in identifying not only salient but hidden correlations of structural features with each other as well as with the activity. It could be useful to optimize peptide-type molecules for better activity against SARS-CoV and SARS-CoV-2.

Acknowledgements

Disclosure statement

References

[1] A.R. Fehr and S. Perlman, *Coronaviruses: An overview of their replication and pathogenesis*, in *Coronaviruses: Methods and Protocols*, H.J. Maier, E. Bickerton, and P. Britton, eds., Springer New York, New York, 2015, pp. 1–23.

[2] L. Zhao, B.K. Jha, A. Wu, R. Elliott, J. Ziebuhr, A.E. Gorbalenya, R.H. Silverman, and S.R. Weiss, *Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology*, Cell Host Microbe 11 (2012), pp. 607–616. doi: 10.1016/j.chom.2012.04.011.

[3] X.-Y. Ge, J.-L. Li, X.-L. Yang, A.A. Chmura, G. Zhu, J.H. Epstein, J.K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.-Y. Zhang, L.-F. Wang, P. Daszak, and Z.-L. Shi, *Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor*, Nature 503 (2013), pp. 535–538. doi: 10.1038/nature12711.

[4] Y. Chen and D. Guo, *Molecular mechanisms of coronavirus RNA capping and methylation*, Virol. Sin. 31 (2016), pp. 3–11. doi: 10.1007/s12250-016-3726-4.

[5] J. Xu, S. Zhao, T. Teng, A.E. Abdalla, W. Zhu, L. Xie, Y. Wang, and X. Guo, *Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV*, Viruses 12 (2020), pp. 244. doi: 10.3390/v12020244.

[6] Anonymous, *Virology: Coronaviruses*, Nature 220 (1968), pp. 650. doi: 10.1038/220650b0.

[7] S.R. Weiss and S. Navas-Martin, *Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus*, Microbiol. Mol. Biol. Rev. 69 (2005), pp. 635–664. doi: 10.1128/mmbr.69.4.635-664.2005.

[8] K. McIntosh, *Coronaviruses: A comparative review*, in *Current Topics in Microbiology and Immunology/Ergebnisse Der Mikrobiologie Und Immunitätsforschung*, W. Arber, R. Haas, W. Henle, P.H. Hofschneider, N.K. Jerne, P. Koldovský, H. Koprowski, O. Maaløe, R. Rott, H. G. Schweiger, M. Sela, L. Syruček, P.K. Vogt, and E. Wecker, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 1974, pp. 85–129

[9] J.M. González, P. Gomez-Puertas, D. Cavanagh, A.E. Gorbalenya, and L. Enjuanes, *A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae*, Arch. Virol. 148 (2003), pp. 2207–2235. doi: 10.1007/s00705-003-0162-1

[10] J. Cui, F. Li, and Z.-L. Shi, *Origin and evolution of pathogenic coronaviruses*, Nat. Rev. Microbiol. 17 (2019), pp. 181–192. doi: 10.1038/s41579-018-0118-9

[11] S. Cauchemez, M.D. Van Kerkhove, S. Riley, C.A. Donnelly, C. Fraser, and N.M. Ferguson, *Transmission scenarios for Middle East respiratory syndrome coronavirus (MERS-CoV) and how to tell them apart*, Eurosurveillance 18 (2013), pp. 20503. doi: 10.2807/ese.18.24.20503-en

[12] L. Liu, T. Wang, and J. Lu, *The prevalence, origin, and prevention of six human coronaviruses*, Virol. Sin. 31 (2016), pp. 94–99. doi: 10.1007/s12250-015-3687-z

[13] A.E. Gorbalenya, S.C. Baker, R.S. Baric, R.J. de Groot, C. Drosten, A.A. Gulyaeva, B.L. Haagmans, C. Lauber, A.M. Leontovich, B.W. Neuman, D. Penzar, S. Perlman, L.L.M. Poon, D. Samborskiy, I. A. Sidorov, I. Sola, and J. Ziebuhr, *Severe acute respiratory syndrome-related coronavirus: The species and its viruses – A statement of the Coronavirus Study Group*, bioRxiv (2020). doi: 10.1101/2020.02.07.937862

[14] K. Dhama, K. Sharun, R. Tiwari, M. Dadar, Y.S. Malik, K.P. Singh, and W. Chaicumpa, *COVID-19, an emerging coronavirus infection: Advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics*, Hum. Vacc. Immunother. (2020), pp. 1–7. doi: 10.1080/21645515.2020.1735227

[15] N.H.L. Leung, D.K.W. Chu, E.Y.C. Shiu, K.-H. Chan, J.J. McDevitt, B.J.P. Hau, H.-L. Yen, Y. Li, D.K. M. Ip, J.S.M. Peiris, W.-H. Seto, G.M. Leung, D.K. Milton, and B.J. Cowling, *Respiratory virus shedding in exhaled breath and efficacy of face masks*, Nat. Med. 26 (2020), pp. 676–680. doi: 10.1038/s41591-020-0843-2

[16] W. Luerang, T. Khammee, W. Kumpum, S. Suksamrarn, V. Chatsudthipong, and C. Muanprasat, *Hydroxyxanthone as an inhibitor of cAMP-activated apical chloride channel in human intestinal epithelial cell*, Life Sci. 90 (2012), pp. 988–994. doi: 10.1016/j.lfs.2012.05.001.

[17] Similarities and differences – COVID-19 and influenza. Available at https://www.paho.org/hq/index.php?option=com_content&view=article&id=15760:similarities-and-differences-covid-19-and-influenza&catid=740&lang=pt&Itemid=1926, 2020.

[18] S.J.Y. Macalino, V. Gosu, S. Hong, and S. Choi, *Role of computer-aided drug design in modern drug discovery*, Arch. Pharm. Res. 38 (2015), pp. 1686–1701. doi: 10.1007/s12272-015-0640-5.

[19] V.H. Masand, N.N.E. El-Sayed, D.T. Mahajan, and V. Rastija, *QSAR analysis for 6-arylpyrazine-2-carboxamides as* Trypanosoma brucei *inhibitors*, SAR QSAR Environ. Res. 28 (2017), pp. 165–177. doi: 10.1080/1062936x.2017.1292407

[20] V.H. Masand, N.N.E. El-Sayed, D.T. Mahajan, A.G. Mercader, A.M. Alafeefy, and I.G. Shibi, *QSAR modeling for anti-human African trypanosomiasis activity of substituted 2-Phenylimidazopyridines*, J. Mol. Struc. 1130 (2017), pp. 711–718. doi: 10.1016/j.molstruc.2016.11.012.

[21] V.H. Masand, D.T. Mahajan, A.K. Maldhure, and V. Rastija, *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*, Med. Chem. Res. 25 (2016), pp. 2324–2334. doi: 10.1007/s00044-016-1664-1

[22] E. Pourbasheer, S. Shokouhi Tabar, V.H. Masand, R. Aalizadeh, and M.R. Ganjali, *3D-QSAR and docking studies on adenosine A2A receptor antagonists by the CoMFA method*, SAR QSAR Environ. Res. 26 (2015), pp. 461–477. doi: 10.1080/1062936X.2015.1049666.

[23] T. Fujita and D.A. Winkler, *Understanding the roles of the "Two QSARs"*, J. Chem. Inf. Model. 56 (2016), pp. 269–274. doi: 10.1021/acs.jcim.5b00229

[24] T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, and 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), pp. 6595–6628. doi: 10.1021/acs.jmedchem.5b01461.

[25] P. Thanigaimalai, S. Konno, T. Yamamoto, Y. Koiwai, A. Taguchi, K. Takayama, F. Yakushiji, K. Akaji, Y. Kiso, Y. Kawasaki, S.-E. Chen, A. Naser-Tavakolian, A. Schön, E. Freire, and 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 (2013), pp. 436–447. doi: 10.1016/j.ejmech.2013.05.005

[26] P. Thanigaimalai, S. Konno, T. Yamamoto, Y. Koiwai, A. Taguchi, K. Takayama, F. Yakushiji, K. Akaji, S.-E. Chen, A. Naser-Tavakolian, A. Schön, E. Freire, and 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 (2013), pp. 372–384. doi: 10.1016/j.ejmech.2013.07.037.

[27] 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, and Y. Hayashi, *Design and synthesis of new tripeptide-type SARS-CoV 3CL protease inhibitors containing an electrophilic arylketone moiety*, Bioorg. Med. Chem. 21 (2013), pp. 412–424. doi: 10.1016/j.bmc.2012.11.017.

[28] L. Zhang, D. Lin, X. Sun, U. Curth, C. Drosten, L. Sauerhering, S. Becker, K. Rox, and R. Hilgenfeld, *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors*, Science 368 (2020), pp. 409–412. doi: 10.1126/science.abb3405.

[29] L. 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, and R. Hilgenfeld, *α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: Structure-based design, synthesis, and activity assessment*, J. Med. Chem. 63 (2020), pp. 4562–4578. doi: 10.1021/acs.jmedchem.9b01828.

[30] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, and G.R. Hutchison, *Open Babel: An open chemical toolbox*, J. Cheminf. 3 (2011). doi: 10.1186/1758-2946-3-33

[31] J.A. Rackers, Z. Wang, C. Lu, M.L. Laury, L. Lagardère, M.J. Schnieders, J.-P. Piquemal, P. Ren, and J.W. Ponder, *Tinker 8: Software tools for molecular design*, J. Chem. Theory Comput. 14 (2018), pp. 5273–5289. doi: 10.1021/acs.jctc.8b00529

[32] P. Tosco, T. Balle, and F. Shiri, *Open3DALIGN: An open-source software aimed at unsupervised ligand alignment*, J. Comput. Aid. Mol. Des. 25 (2011), pp. 777–783. doi: 10.1007/s10822-011-9462-9.

[33] V.H. Masand and V. Rastija, *PyDescriptor: A new PyMOL plugin for calculating thousands of easily understandable molecular descriptors*, Chemom. Intell. Lab. Syst. 169 (2017), pp. 12–18. doi: 10.1016/j.chemolab.2017.08.003.

[34] C.W. Yap, *PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints*, J. Comput. Chem. 32 (2011), pp. 1466–1474. doi: 10.1002/jcc.21707

[35] V.H. Masand, D.T. Mahajan, A.M. Alafeefy, S.N. Bukhari, and N.N. Elsayed, *Optimization of antiproliferative activity of substituted phenyl 4-(2-oximidazolidin-1-yl) benzenesulfonates: QSAR and CoMFA analyses*, Eur. J. Pharm. Sci. 77 (2015), pp. 230–237. doi: 10.1016/j.ejps.2015.06.001

[36] P. Gramatica, S. Cassani, and A. Sangion, *PBT assessment and prioritization by PBT Index and consensus modeling: Comparison of screening results from structural models*, Environ. Int. 77 (2015), pp. 25–34. doi: 10.1016/j.envint.2014.12.012

[37] P. Gramatica, S. Cassani, and N. Chirico, *QSARINS-chem: Insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS*, J. Comput. Chem. 35 (2014), pp. 1036–1044. doi: 10.1002/jcc.23576.

[38] P. Gramatica, N. Chirico, E. Papa, S. Cassani, and S. Kovarich, *QSARINS: A new software for the development, analysis, and validation of QSAR MLR models*, J. Comput. Chem. 34 (2013), pp. 2121–2132. doi: 10.1002/jcc.23361

[39] V.H. Masand, D.T. Mahajan, G.M. Nazeruddin, T.B. Hadda, V. Rastija, and 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 (2015), pp. 1241–1264. doi: 10.1007/s00044-014-1193-8

[40] A. Cherkasov, E.N. Muratov, D. Fourches, A. Varnek, I.I. Baskin, 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, and A. Tropsha, *QSAR modeling: Where have you been? Where are you going to?* J. Med. Chem. 57 (2014), pp. 4977–5010. doi: 10.1021/jm4004285

[41] T.M. Martin, P. Harten, D.M. Young, E.N. Muratov, A. Golbraikh, H. Zhu, and A. Tropsha, *Does rational selection of training and test sets improve the outcome of QSAR modeling?* J. Chem. Inf. Model. 52 (2012), pp. 2570–2578. doi: 10.1021/ci300338w.

[42] N. Chirico and 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), pp. 2044–2058. doi: 10.1021/ci300084j.

[43] V. Consonni, R. Todeschini, D. Ballabio, and F. Grisoni, *On the misleading use of QF32 for QSAR model comparison*, Mol. Inform. 38 (2019), pp. 1800029. doi: 10.1002/minf.201800029

[44] V. Consonni, D. Ballabio, and R. Todeschini, *Comments on the definition of the Q2 parameter for QSAR validation*, J. Chem. Inf. Model. 49 (2009), pp. 1669–1678. doi: 10.1021/ci900115y.

[45] P. Gramatica, *External evaluation of QSAR models, in addition to cross-validation: Verification of predictive capability on totally new chemicals*, Mol. Inform. 33 (2014), pp. 311–314. doi: 10.1002/minf.201400030

[46] P. Gramatica, *On the development and validation of QSAR models*, Meth. Mol. Biol. 930 (2013), pp. 499–526. doi: 10.1007/978-1-62703-059-5_21.

[47] P. Gramatica, S. Cassani, P.P. Roy, S. Kovarich, C.W. Yap, and E. Papa, *QSAR modeling is not push a button and find a correlation: A case study of toxicity of (benzo-)triazoles on algae*, Mol. Inform. 31 (2012), pp. 817–835. https://doi.org/10.1002/minf.201200075 doi: 10.1002/minf.201200075.

[48] N. Chirico and 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), pp. 2320–2335. doi: 10.1021/ci200211n.