ACADEMIA ROMÂNĂ
Revue Roumaine de Chimie
https://www.icf.ro/rrch/

Rev. Roum. Chim.,
2022, *67*(4-5), 321–328
DOI: 10.33224/rrch.2022.67.4-5.11

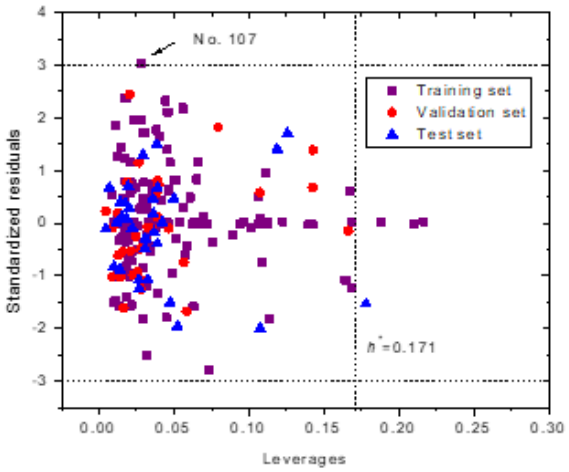
SUPPORT VECTOR MACHINE MODEL FOR PREDICTING ACTIVITY OF INHIBITORS AGAINST SARS-COV 3CLpro ENZYME**

Linhai HU,[a] Xianwei HUANG,[a,*] Qun ZENG,[b,*] Yuting WANG[a] and Shizhang LI[a]

[a] Hunan Provincial Key Laboratory of Environmental Catalysis & Waste Regeneration, College of Materials and Chemical Engineering, Hunan Institute of Engineering, Xiangtan, Hunan 411104, China
[b] Department of Neurosurgery, Central Hospital of Xiangtan, Xiangtan, Hunan 411100, China

Received December 25, 2021

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a beta coronavirus which led to coronavirus disease-2019 (COVID-19) and has threatened global public health and economy. Currently there is no specific medicine for COVID-19. So there is an urgent need to develop broad-spectrum anti-coronavirus drugs. The SARS-CoV 3-chymotrypsin-like protease (3CLpro) is highly conservative in beta-coronavirus and becomes viable target used for anti-SARS drugs. Support vector machine (SVM) algorithm was used to build quantitative structure–activity relationships (QSARs) for the activity (logIC_{50}) of 204 inhibitors for SARS-CoV 3CLpro enzyme. Seven molecular descriptors were selected for the optimal SVM model with parameters C = 250 and γ = 0.15, which has root-mean-square (*rms*) errors being 0.341 (training set), 0.337 (validation set) and 0.336 (test set). Comparison with other models in the literature shows that the SVM model was proved to be satisfactory although the SVM model in this paper has more samples. The investigation results provide a powerful tool for searching new 3CLpro enzyme inhibitors for SARS-CoV.



INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has threatened global public health and economy since the outbreak of an infectious disease in Wuhan, in December 2019.[1] As of 24 November 2021, 1:45 pm GMT+8, SARS-CoV-2 has caused 257,469,528 infected cases and 5,158,211 deaths (https://www.who.int/). In the 21st century, the two human coronavirus diseases, severe acute respiratory syndrome CoV (SARS-CoV) and Middle East respiratory syndrome CoV (MERS-CoV) were, respectively, discovered in 2002 and 2012. More than 8,422 and 1,700 persons were, respectively, subjected to infection, with the fatality rates of 10% for SARS-CoV and 36% for MERS-CoV.[2] Fortunately, SARS-CoV epidemic was successfully contained by July 2003, but the MERS-CoV still poses a threat to public health and global safety. SARS-CoV2 is the 3rd human coronavirus that can cause serious respiratory epidemics, named coronavirus disease-2019 (COVID-19). Currently there is no specific medicine for COVID-19.

* Corresponding author: Tel. +86 731 58680049; Fax: +86 731 58680125; hxw1o3o@126.com (X. Huang) or jiejie6l6l8@163.com (Q. Zeng)
** Supporting information on https: //www.icf.ro/rrch/ or https: //revroum.lew.ro

Coronavirus (SARS-CoV) genome possesses 6-12 open reading frames (ORFs), which accounts for translating two polyproteins (pp1a and pp1ab).[3] The polyproteins can be incised and functionalized into 16 non-structural proteins (nsps) by 3-chymotrypsin-like protease (3CLpro), a cysteine protease composed of approximately 300 amino acids and three domains.[4,5] Furthermore, 3CLpro has the ability of cleaving the intracellular protein NEMO and of inhibiting the active state of interferon signaling pathway. Thus, SARS-CoV 3CLpro is instrumental in bringing about viral genome replication and transcription, and other important viral life processes, such as protein translation, cleavage, and modification and nucleic acid synthesis.[6] There is 96.1% sequence similarity in 3CLpros of SARS-CoV-2, SARS-CoV and MERS-CoV. [7-10] Further, their 3CLpros exhibit a high degree of structural conservatism.[11] Thus SARS-CoV 3CLpro is taken as a viable target for developing anti-SARS drug and 3CLpro inhibitors for SARS-CoV should be effective for SARS-CoV-2.[6-8]

Kumar and Roy introduced a model for inhibitory activity (logIC_{50}) based on quantitative structure–activity relationships (QSARs). Eight descriptors and 69 candidates for SARS-CoV 3CLpro enzyme were used to develop the model.[8] The training and test sets, respectively, have 56 and 13 molecules, possessing R^2 of 0.764 and 0.711, respectively. Recently, Yu correlated inhibitory constants (pKi) of molecules for SARS-CoV 3CLpro enzyme with molecular descriptors. The numbers of molecules and descriptors are, respectively 89 and 6. Both the training and test sets have R^2 above 0.7.[1] QSAR models reflect the correlations between the physicochemical or biological properties studied and molecular descriptors representing information related to the structure of compounds. These models can be used for drug identification and optimization through predicting the activity or property of candidate molecules, including that have not been synthesized. Thus this technology can save time and money and accelerate the efficiency of drug development.[12] The purpose of this paper is to establish a QSAR of 204 inhibitory activities (logIC_{50}) for SARS-CoV 3CLpro enzyme.[6,13] Support vector machine (SVM) algorithm, along with seven descriptors, was adopted to build the model that will be a powerful tool for searching novel 3CLpro enzyme inhibitors for SARS-CoV.

METHODS

There are 204 inhibitors and their activity (log*IC50*) for SARS-CoV 3CLpro enzyme listed in Table S1 (see Supplemental Information), that were taken from the literature.[6,13] The experimental IC_{50} values varied from 0.5 to 780 μM, and the logIC_{50} values were in the range of -0.301~2.892 by converting to logarithm of IC_{50}. A lower IC_{50} value indicates the stronger activity for the inhibitor. These experimental data were composed of peptidomimetic inhibitors and nonpeptidic small molecule inhibitors. The later species include decahydroisoquinoline, octahydro-isochromene, unsymmetrical aromatic disulphides, pyrazolone, pyrimidines, flavonoids, bioflavonoids, chalcones, isatin, terpenoid, triazole and piperidine derivatives. Generally, the training sets account for 2/3 ~ 4/5 of the total samples.[14,15] In this paper, inhibitors were randomly partitioned into a training set (Nos. 1-140 in Table S1), a validation set (Nos. 141-172) and a test set (Nos. 173-204). The training set was used for training SVM models, the validation set for adjusting SVM parameters, and the test set for evaluating models.

The molecular structures of inhibitors were constructed with ChemBioDraw Ultra 12.0, followed by optimization with the AM1 method in Gaussian 09. In the end, 4885 molecular descriptors were calculated for each inhibitor with Dragon 6.0.[16] After removing those molecular descriptors with partial correlation coefficient above 0.90 or near to a constant, 1036 descriptors were obtained for each molecule.

By using the principle of structural risk minimization, SVM algorithms have good prediction performance even if a small number of samples are used. As one of the most common application forms of SVM algorithms, support vector regression (SVR) deals with nonlinear problems via mapping training samples into a high dimensional feature space, followed by carrying out linear regression. The linear function in SVR to be dealt with is: [17-20]

$$f(x) = \sum_i^n \varphi(x_i)\omega + b \qquad (1)$$

when (non-negative) slack variables ξ_i and ξ_i^* are introduced, the coefficients w and b are calculated by:

$$\min_{\omega,b,\zeta,\zeta^*} J(\omega,\zeta,\zeta^*,b) = \tfrac{1}{2}\|\omega\|^2 + C\sum_i(\zeta_i + \zeta_i^*) \qquad (2)$$

subject to:

$$y_i - \varphi^T(x_i)\omega - b \le \varepsilon + \zeta_i \quad (3)$$

$$\varphi^T(x_i)\omega + b - y_i \le \varepsilon + \zeta_i^* \quad (4)$$

The penalty factor (C >0) is decided by the user. The ε-insensitive function is described by:

$$\left|f(x)-y\right|_\varepsilon = \begin{cases} \left|f(x)-y\right| - \varepsilon & \left|f(x)-y\right| \ge \varepsilon \\ 0 & \left|f(x)-y\right| < \varepsilon \end{cases} \quad (5)$$

Then, Equation (1) becomes:

$$f(x) = \sum_i^n (a_i - a_i^*)\varphi(x_i)\varphi(x) + b \quad (6)$$

After introducing the kernel function for the dot product in the D-dimensional feature space, the minimizing function is:

$$f(x) = \sum_i^s (a_i - a_i^*)K(x,y) + b \quad (7)$$

with $a_i a_i^* = 0, a_i, a_i^* \ge 0$. The coefficients a_i, a_i^* can be obtained by maximizing the following form:

$$\text{Max}: R(a^*, a) = -\tfrac{1}{2}\sum_{i,j=1}^n (a_i^* - a_i)(a_j^* - a_j)K(x_i, x_j) - \varepsilon\sum_{i=1}^n (a_i^* + a_i) + \sum_{i=1}^n y_i(a_i^* - a_i) \quad (8)$$

$$.S.T.\ \sum_{i=1}^n (a_i^* - a_i) = 0, \quad 0 \le a_i^*, a_i \le C \quad (9)$$

Gaussian radial basis function can be used:

$$K(x_i, x_j) = exp(-\gamma\left\|x_i - x_j\right\|^2) \quad (10)$$

where the parameter γ is the kernel width. Both C and γ values need to be optimized carefully because SVR models may produce over-fitting or under-fitting if the parameters C and γ are too large or too small.[17-20]

RESULTS AND DISCUSSION

Stepwise multiple linear regression (MLR) in IBM SPSS statistical 19 was carried out, through using 1036 descriptors as independent variables and 204 $\log IC_{50}$ in Table S1 (see Supplemental Information) as dependent variables. New variable was introduced when its increment of determination coefficient ΔR^2 above 0.01. The seven descriptors, *VE1_B(m)*, *VE1_B(s)*, *SM5_X*, *GATS7m*, *JGI9*, *nRNHR* and *F07[N-O]*, were entered QSAR models. Table 1 shows the descriptor definitions.

The 2D matrix-based descriptors *VE1_B(w)* are molecular descriptors derived from the sum of the coefficients of the eigenvector associated with the last (largest negative) eigenvalue of Burden matrix on a H-depleted molecular graph.[16] *VE1_B(w)* are calculated with:

$$VE1\_B(w) = \sum_{i=1}^{nsk} |l_i| \quad (11)$$

here l_i is the ith coefficient of the last eigenvector of Burden matrix, and *nSK* means the number of graph vertices. w denotes the vertex weighting scheme, i.e., atomic properties used as the atomic weightings for descriptor calculation, including atomic mass (m) and intrinsic state (s), etc. Similarly, the 2D matrix-based descriptor, spectral moment of order 5 from chi martix (*SM5_X*) is derived from:

$$SM5\_X = \text{sgn}(\sum_i^{nsk} \ln(1 + \left|\sum_i^{nsk} \lambda_i^5\right|)) \quad (12)$$

where X denotes Chi matrix (X) and λ is the eigenvalue of chi matrix.[16]

The descriptor *GATS7m* (Geary autocorrelations of lag 7 weighted by mass) belongs to Geary autocorrelations (*GATSkw*) that are calculated with:

$$GATSkw = \frac{\frac{1}{2\Delta}\sum_{i=1}^{nAT}\sum_{j=1}^{nAT}\delta_{ij}(w_i - w_j)^2}{\frac{1}{nAT}\sum_{i=1}^{nSk}(w_i - \overline{w})^2} \quad (13)$$

The parameters k, w, nAT, δ_{ij} and Δ denote the lag value, weighting method by atomic mass, the total number atoms in a molecule, the Kronecker delta, and the sum of the Kronecker deltas, respectively.[16] *GATSkw* encodes the information on molecular size, geometry and symmetry. [16,21,22]

A mean topological charge index (*JGIk*) (k being an integer between one and ten) is related to the unsymmetrical matrix CT. The element CT_{ij} is equal to δ_i if $i = j$, and (m_{ij} - m_{ji}), otherwise.[16] The descriptor *JGI9* is correlated with charge distribution along the topological structure in a molecule.

Table 1

The descriptor definitions and blocks

Descriptor	Definition	Block
VE1_B(m)	Coefficient sum of the last eigenvector from Burden matrix weighted by mass	2D matrix-based descriptors
VE1_B(s)	Coefficient sum of the last eigenvector from Burden matrix weighted by I-state	2D matrix-based descriptors
SM5_X	Spectral moment of order 5 from chi martix	2D matrix-based descriptors
GATS7m	Geary autocorrelations of lag 7 weighted by mass	2D autocorrelations
JGI9	Mean topological charge index of order 9	2D autocorrelations
nRNHR	Number of secondary amines (aliphatic)	Functional group counts
F07[N-O]	Frequency of N-O at topological distance 7	2D Atom Pairs

Table 2

Descriptor characteristics for the MLR model

Descriptor	Standardized Coefficient	*t*-test	*Sig.*	*VIF*
VE1_B(m)	0.519	6.989	0.001	2.301
VE1_B(s)	-0.242	-3.162	0.001	2.442
SM5_X	-0.180	-3.274	0.000	1.257
GATS7m	0.191	3.775	0.000	1.070
JGI9	-0.176	-3.421	0.000	1.104
nRNHR	0.539	10.604	0.000	1.076
F07[N-O]	-0.198	-3.770	0.000	1.151

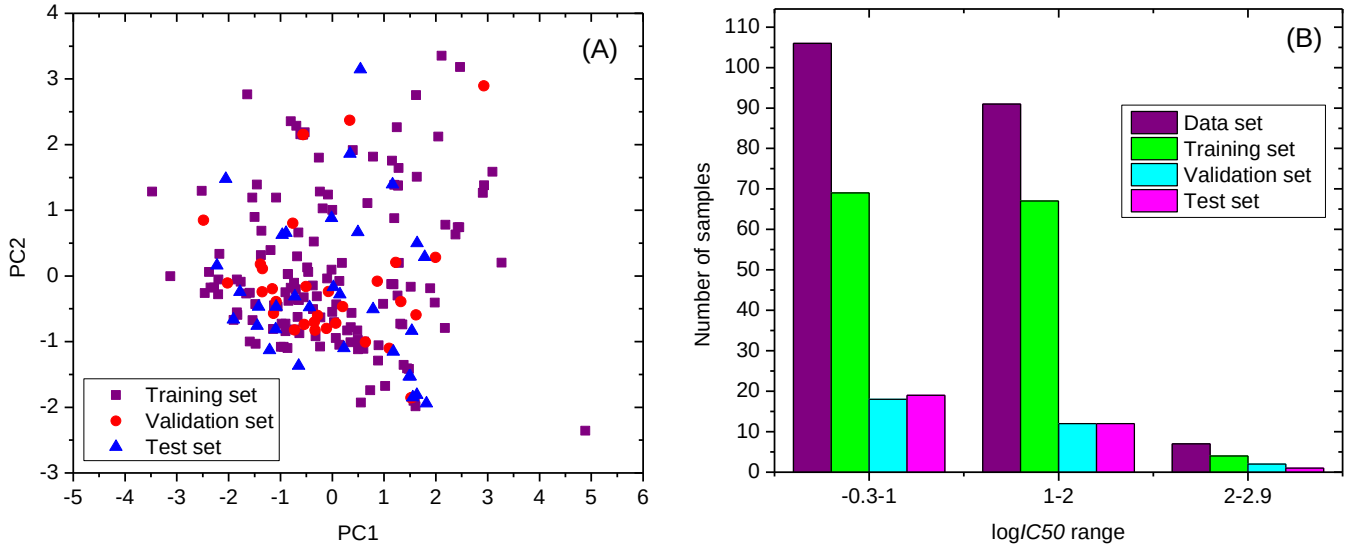


Fig. 1 – Structural and activity diversity maps. (A) Chemical space defined by first two principal components. (B) Distribution comparison of logIC_{50} values in different data sets.

In addition, the descriptor *nRNHR* (R: aliphatic group linked through C, not C = O) is the number of secondary amines (aliphatic). A molecule with functional group *nRNHR* is inclined to form hydrogen bond with targets. The descriptor *F07[N-O]* means the number of N-O groups at topological distance 7.

The characteristics of seven descriptors, *VE1_B(m)*, *VE1_B(s)*, *SM5_X*, *GATS7m*, *JGI9*, *nRNHR* and *F07[N-O]* selected for QSARs are shown in Table 2. As can been seen from Table 2, each of the descriptors possesses low *sig.*-value (< 0.05), which suggest these descriptors being important variables and related to the inhibitory activity logIC_{50}. The variance inflation factor (*VIF*) of each descriptor is less than 10, suggesting that they do not have serious multicollinearity problem among them. According to the *t*-test, their *t*-test absolute values decrease in the order: *nRNHR*, *VE1_B(m)*, *GATS7m*, *F07[N-O]*, *JGI9*, *SM5_X*, and *VE1_B(s)*, and their importance affecting the inhibitory activity logIC_{50} decreases with the same order. In addition, it should be pointed out that the descriptors, *VE1_B(m)*, *GATS7m*, *JGI9*, *SM5_X*

and *VE1_B(s)*, have no obviously physical meaning, resulting in difficulties in mechanism interpretation for QSAR models. Despite this fact, QSARs can be used to predict the activities since QSAR studies are independent of mechanism information.[23]

Figure 1 shows the distribution maps for the structural and activity diversity covered by 204 inhibitors. Figure 1A is the chemical space expressed by first two principal components that were calculated with the principal component analysis (PCA) from seven molecular descriptors used. The smaller distance among the points, the higher is the structural similarity. Figure 1A, B shows that structurally distinct chemotypes cover a wide range of activities logIC_{50} and appear in three data sets, which indicates that it is feasible to predict the activities logIC_{50} for the test set, by applying the QSARs based on the training set.[24,25]

The seven descriptors, *nRNHR*, *VE1_B(m)*, *GATS7m*, *F07[N-O]*, *JGI9*, *SM5_X*, and *VE1_B(s)*, were used for developing SVM model for inhibitory activity logIC_{50}, through applying MATLAB R2014a and LibSVM.[21] The grid search method was selected to train the SVM parameters C and γ. The *rms* errors of 32 compounds from the validation set were used to estimate the performance of SVM model when the SVM parameter C varied from 50 to 500 with the step of C = 50. Figure 2 shows their *rms* errors versus C parameter with γ = 0.15. As is shown in Figure 2, the *rms* error reaches to the minimum value of 0.337 when C is 250. Similarly, the parameter γ varied from 0.05 to 0.5 with the step of γ= 0.05. As can be seen from Figure 3, the curve falls to the lowest level when *rms* error is 0.337 and γ is 0.15. Therefore, the optimal SVM parameters C and γ are 250 and 0.15, respectively. The optimal SVM model was tested with the 32 compounds in the test set. Figure 4 shows the relationship between experimental and calculated log IC_{50} from the SVM model. Their *rms* errors of the training, validation and test sets are, respectively, 0.341, 0.337 and 0.336. Their determinant coefficients are 0.642, 0.619, and 0.603, respectively, which are less than the results (R^2 of 0.764 and 0.711) of Kumar and Roy.[8] However, the results in this paper still could be acceptable because their model was based on only 69 molecules and 16 compounds were removed from the data set. Further, they used more descriptors (n = 8) for QSAR models.[8] Furthermore, the *rms* errors in this paper are more less than that (the training set, *rms* = 0.435; the test set, *rms* = 0.525) by Yu,[1] although this paper dealt with more samples.

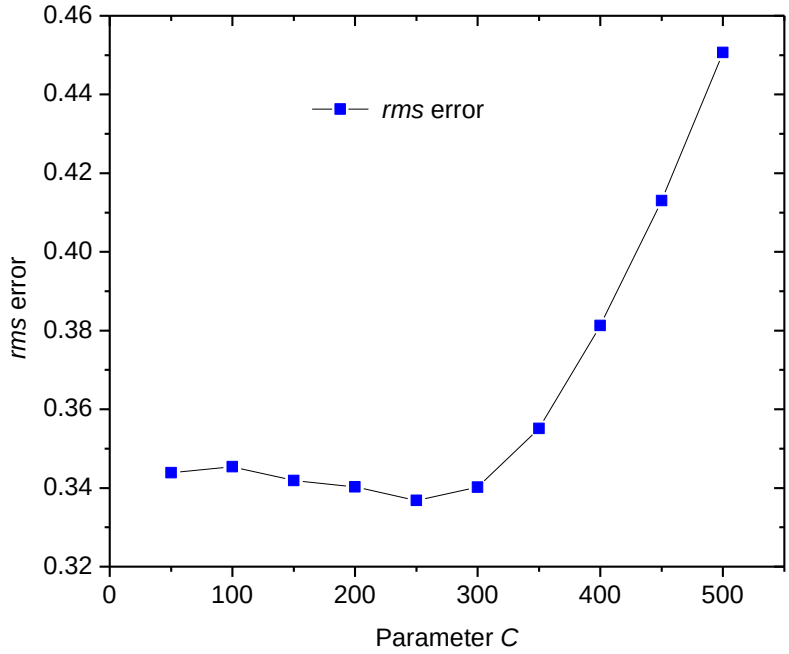


Fig. 2 – The *rms* error versus C parameter with γ = 0.15.

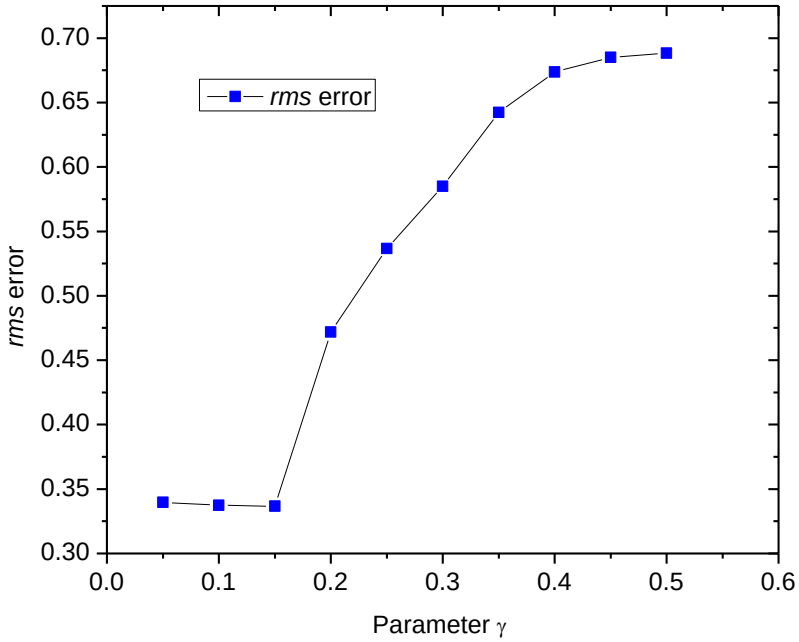


Fig. 3 – *rms* error versus γ parameter with C = 250.

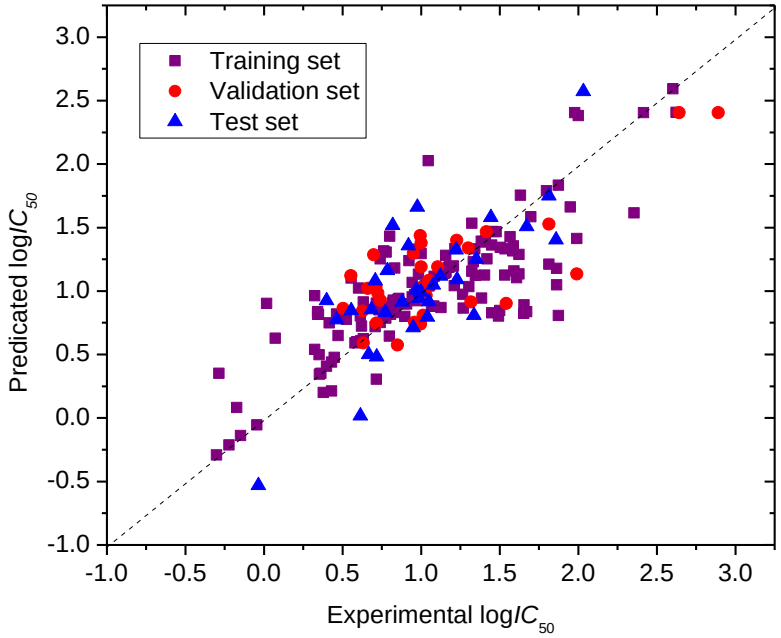


Fig. 4 – Experimental versus calculated logIC_{50} from SVM model.

Figure 5 is the Williams plot, describing the standardized residuals (σ) calculated with the optimal SVM model against the leverages (h). The prediction results are reliable when the sample points lie in the applicability domain.[26,27] Figure 6 shows that there is only one sample (No. 107, N-(4-(tert-butyl)phenyl)-N-(2-(tert-butylamino)-2-oxo-1-(pyridin-3-yl)ethyl)-1-methyl-1H-imidazole-4-carboxamide) with $|\sigma| > 3$. This phenomenon indicates that the compound of No. 107 may have great experimental error for the measured activity value. Moreover, the warning leverage h^* was calculated with the expression: $3\times(p+1)/n = 3\times(7+1)/140 = 0.171$, here p being the numbers of descriptors and n being the number of molecules in training set). There are four compounds (Nos. 8, 44, 136 and 175) possessing larger leverage values (> 0.171) and lower standardized residuals (σ) (< 3), which indicates the optimal SVM model has good generalizability. Note that, however, QSAR models with different performances can provided higher and lower prediction accuracies for

molecules with similar structures or with same groups.[28] Figure 6 shows three molecular structures (Nos. 63, 101, and 123) from the training set, that possess the lowest absolute errors. The prediction accuracies of molecules in the test set depend on their similarity to the molecule in the training set, e.g. the molecules of in Fig. 6.[24,25,28]

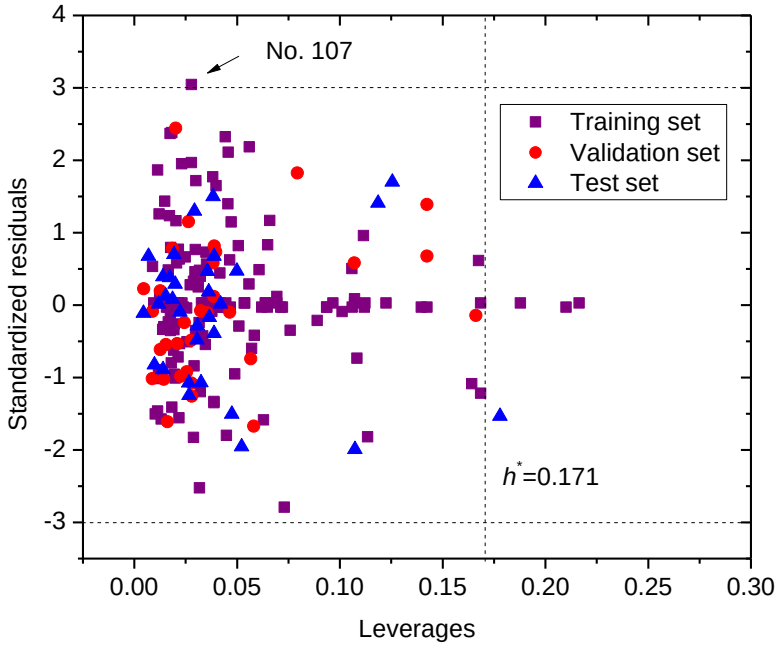


Fig. 5 – Plot of standardized residuals against the leverages (h^*= 0.171).

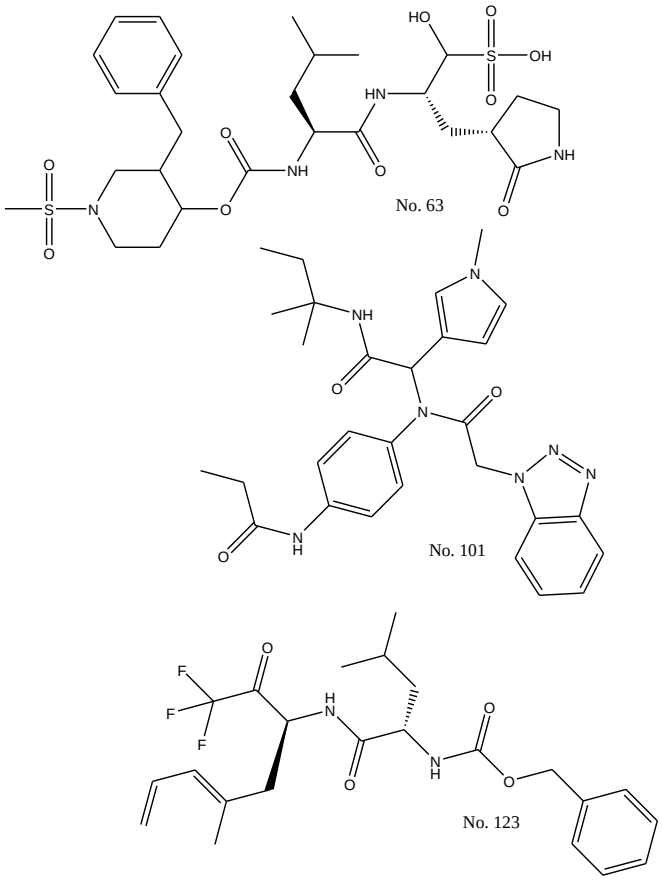


Fig. 6 – Molecular structures of the compounds of Nos. 63, 101, and 123.

CONCLUSIONS

Although many factors affect the inhibitory activities (logIC_{50}) against SARS-CoV 3CLpro enzyme, 2D matrix-based descriptors (*VE1_B(m)*, *VE1_B(s)* and *SM5_X)*, 2D autocorrelations (*GATS7m* and *JGI9*), functional group count (*nRNHR*) and 2D atom pair (*F07[N-O]*) reflect the structural information relating to logIC_{50}. The SVM model obtained in this paper has parameters of C = 250 and γ = 0.15, resulting in determinant coefficients of 0.642 (training set), 0.619 (validation set), and 0.603 (test set). The SVM model in this paper can provide a powerful tool for searching new 3CLpro enzyme inhibitors for SARS-CoV.

REFERENCES

1. X. Yu, *J. Saudi Chem. Soc*., **2021**, *25*, 101262.
2. Y. A. Helmy, M. Fawzy, A. Elaswad, A. Sobieh, S. P. Kenney and A. A. Shehata, *J. Clin. Med.*, **2020**, *9*, 1225.
3. S. Hussain, J. Pan, Y. Chen, Y. Yang, J. Xu, Y. Peng, Y. Wu, Z. Li, Y. Zhu, P. Tien and D. Guo, *J. Virol.*, **2005**, *79*, 5288–5295.
4. X. Liu and X. Wang, *J. Genet. Genomics*, **2020**, *47*, 119-121.
5. U. Bacha, J. Barrila, A. Velazquez-Campoy, S. A. Leavitt and E. Freire, *Biochemistry*, **2004**, *43*, 4906-4912.
6. Y. Liu, C. Liang, L. Xin, X. Ren, L. Tian, X. Ju, H. Li, Y. Wang, Q. Zhao, H. Liu, W. Cao, X. Xie, D. Zhang, Y. Wang and Y. Jian, *Eur. J. Med. Chem.*, **2020**, *206*, 112711.
7. V. Masand, A. Gandhi V. Rastija and M. K. Patil, *Chemometr. Intell. Lab. Syst.*, **2020**, *206*, 104172.
8. V. Kumar and K. Roy, *SAR QSAR Environ. Res*., **2020**, *31*, 511-526.
9. Q.-X. Long, X.-J. Tang, Q.-L. Shi, Q. Li, H.-J. Deng, J. Yuan, J.-L. Hu, W. Xu, Y. Zhang, F.-J. Lv, K. Su, F. Zhang, J. Gong, B. Wu, X.-M. Liu, J.-J. Li, J.-F. Qiu, J.n Chen and A.-L. Huang, *Nat. Med.*, **2020**, *26*, 1200-0124.
10. R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi and W. Tan, *Lancet.*, **2020**, *395*, 565-574.
11. F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J .J. Zheng, L. Xu, E. C. Holmes and Y. Z. Zhang, *Nature*, **2020**, *579*, 265-269.
12. I. Ponzoni, V. Sebastián-Pérez, C. Requena-Triguero, C. Roca, M. J. Martínez1, F. Cravero, M. F. Díaz, J. A. Páez, R. G. Arrayás, J. Adrio and N. E. Campillo, *Sci. Rep.*, **2017**, *7*, 2403.
13. PubChem. PC QFRET-based dose response biochemical high throughput screening assay to identify inhibitors of the SARS coronavirus 3C-like Protease (3CLPro) PubChem Bioassay, 2009.
14. A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn and D. A. Dobchev, *Chem. Rev.*, **2010**, *110*, 5714-5789.
15. K. Khan, E. Benfenati and K. Roy, *Ecotox. Environ. Safe.*, **2019**, *168*, 287-297.
16. Talete srl, "DRAGON for Widows," Version 6.0, Milan, Italy, 2012.
17. X. Yu, R. Zhan, J. Deng and X. Huang, *J. Theor. Comput. Chem.*, **2017**, *16*, 1750014.
18. X. Yu and L. Huang, *J. Therm. Anal. Calorim.*, **2017**, *130*, 943-947.
19. C. C. Chang and C. J. Lin, *Acm. T. Intel. Syst. Tec.*, **2011**, *2*, 27.
20. X. Yu, *Aquat. Toxicol.*, **2020**, *224*, 105496.
21. E. Pourbasheer, R. Aalizadeh and M. R. Ganjali, *Arab. J. Chem.*, **2019**, *12*, 2141-2149.
22. E. Pourbasheer, A. Beheshti, H. Khajehsharifi, M. R. Ganjali and P. Norouzi, *Med. Chem. Res.*, **2013**, *22*, 4047-4058.
23. A. F. Zahrt, S. V. Athavale and S. E. Denmark, *Chem. Rev.*, **2020**, *120*, 1620-1689.
24. A. S. de Souza, L. L. G. Ferreira, A. S. de Oliveira and A. D. Andricopulo, *Int. J. Mol. Sci.*, **2019**, *20*, 2801
25. T. Lei, Y. Li, Y. Song, D. Li, H. Sun and T. Hou, *J. Cheminform.*, **2016**, *8*, 6.
26. R. Darnag, B. Minaoui and M. Fakir, *Arab. J. Chem.*, **2017**, *10*, S600-S608.
27. M. Shahlaei and A. Fassihi, *Med. Chem. Res.*, **2013**, *22*, 4384-4400.
28. I. V. Tetko, I. Sushko, A. K. Pandey, H. Zhu, A. Tropsha, E. Papa, T. Öberg, R. Todeschini, D. Fourches and A. Varnek, *J. Chem. Inf. Model.*, **2008**, *48*, 1733-1746.