



# In silico approach for the development of novel antiviral compounds based on SARS-COV-2 protease inhibition

Vanja P. Ničković<sup>1</sup> · Gordana R. Nikolić<sup>2</sup> · Biserka M. Nedeljković<sup>2</sup> · Nebojša Mitić<sup>2</sup> · Snežana Filipović Danić<sup>2</sup> · Jadranka Mitić<sup>2</sup> · Zoran Marčetić<sup>2</sup> · Dušan Sokolović<sup>3</sup> · Aleksandar M. Veselinović<sup>4</sup>

Received: 29 May 2021 / Accepted: 5 March 2022 / Published online: 3 April 2022  
© Institute of Chemistry, Slovak Academy of Sciences 2022

## Abstract

The COVID-19 pandemic emerged in 2019, bringing with it the need for greater stores of effective antiviral drugs. This paper deals with the conformation-independent, QSAR model, developed by employing the Monte Carlo optimization method, as well as molecular graphs and the SMILES notation-based descriptors for the purpose of modeling the SARS-CoV-3CLpro enzyme inhibition. The main purpose was developing a reproducible model involving easy interpretation, utilized for a quick prediction of the inhibitory activity of SAR-CoV-3CLpro. The following statistical parameters were present in the best-developed QSAR model: (training set)  $R^2 = 0.9314$ ,  $Q^2 = 0.9271$ ; (test set)  $R^2 = 0.9243$ ,  $Q^2 = 0.8986$ . Molecular fragments, defined as SMILES notation descriptors, that have a positive and negative impact on 3CLpro inhibition were identified on the basis of the results obtained for structural indicators, and were applied to the computer-aided design of five new compounds with (4-methoxyphenyl)[2-(methylsulfanyl)-6,7-dihydro-1H-[1,4]dioxino[2,3-f]benzimidazol-1-yl]methanone as a template molecule. Molecular docking studies were used to examine the potential inhibition effect of designed molecules on SARS-CoV-3CLpro enzyme inhibition and obtained results have high correlation with the QSAR modeling results. In addition, the interactions between the designed molecules and amino acids from the 3CLpro active site were determined, and the energies they yield were calculated.

**Keywords** 3CLpro inhibitors · COVID-19 therapy · QSAR · Molecular modeling · Drug design

## Introduction

The period from the end of 2019 and the beginning of 2020 will be remembered as the onset of the most hazardous pandemic in modern history, i.e. the COVID-19 outbreak. COVID-19 is caused by coronavirus 2 (SARS-CoV-2), or by the positive-sense single-stranded RNA virus (Chen et al. 2020; Guo et al. 2020; Kumar et al. 2020; Rothan and Byrareddy 2020). The Center for Disease Control and

Prevention (CDC) first established the existence of this virus in Wuhan, China in the final days of 2019, and from then on, the virus has spread to other countries, posing a considerable threat to global health (Lai et al. 2020; Pedersen and Ho 2020). The droplets produced when those infected exhale, cough or sneeze represent the main means of SARS-CoV-2 distribution. The crucial factor in spreading the virus is the ease with which it is transmitted, since breathing in the proximity of someone who has COVID-19, or touching a virus contaminated surface before making contact with one's mouth, eyes, or nose, constitute the main routes of infection (Peng et al. 2020; Singhal 2020; Sohrabi et al. 2020). Conversely, the chief danger of contracting COVID-19 lies in the possibility of an asymptomatic person spreading the virus, since 10% of the infections originate from people who do not exhibit any symptoms whatsoever, in spite of the fact that COVID-19 is the most infectious once someone develops symptoms related to COVID-19 (Kampf et al. 2020; Xiao et al. 2020). The chief symptoms attributed to COVID-19 are the potential loss of smell, cough, or

✉ Aleksandar M. Veselinović  
aveselinovic@medfak.ni.ac.rs

<sup>1</sup> Clinical-Hospital Center, Priština, Gracanica, Serbia

<sup>2</sup> Medical Faculty, University of Priština, Kosovska Mitrovica, Serbia

<sup>3</sup> Department of Biochemistry, Faculty of Medicine, University of Niš, Niš, Serbia

<sup>4</sup> Department of Chemistry, Faculty of Medicine, University of Niš, Bulevar Dr Zorana Đinđića 81, 18000 Niš, Serbia

taste, as well as the occurrence of fever, or having trouble breathing (Shang et al. 2020; Wu et al. 2020). Fortunately, to date, proper care and treatment usually lead to the recovery of most people infected with COVID-19 (Jiang et al. 2020; Sanders et al. 2020; Vellingiri et al. 2020). The WHO (World Health Organization) has determined two groups which pose a greater risk of suffering from severe illness and even mortality caused by the SARS-CoV-2 infection—adults over 60 years of age, and people whose immune systems are compromised by various health issues (Xiao et al. 2020, Kampf et al. 2020). Since SARS-CoV-2 is a recently discovered virus, a lot of its characteristics are still unknown.

One promising approach for COVID-19 treatment is blocking the replication of the virus by inhibiting viral protease (Lin et al. 2005; Hoffmann et al. 2020; Jin et al. 2020). One of the key steps in viral replication, which also includes SARS-CoV-2, is the processing of replicase polypeptides 1a and 1ab into functional proteins, which is a process determined with 3C-like protease (3CLpro) (Chen et al. 2005; Deng et al. 2014; Ramajayam et al. 2010; Zhavoronkov et al. 2020). Two proteases, papain-like protease (PLpro) and main protease or 3-chymotrypsin-like protease (Mpro/3CLpro), are essential components for replication of SARS-CoV-2, with main role related to cleaving the two polyproteins, i.e., PP1A and PP1AB into several functional components. In this process, N-terminal domain is split to viral precursor protein at the three sites by PLpro, while the C-terminal domain is split to precursor protein at the 11 sites by 3CLpro (Akaji and Konno 2020). Targeting these enzymes had been used for the treatment of other pathogenic coronaviruses (i.e., MERS-CoV) (Amin and Jha 2020) and since the sequence homology of SARS-CoV-2 3CLpro is 96% structurally closer to the SARS-CoV 3CLpro targeting SARS-CoV-2 3CLpro as a potential therapeutic target is quite a feasible approach for anti-CoVID-19 drug development (Amin et al. 2021; Zhang et al. 2020).

Cheminformatics and computational chemistry have developed methodologies which have proven to be of great importance when designing potential therapeutics. When it comes to drug development research, the most significant contributions made by these *in silico* methods are the search for new leading compounds, or the optimization of therapeutic activities (or pharmacokinetic properties) of the series of chemical compounds whose biological activities have already been determined (Ekins et al. 2007; Tabeshpour et al. 2018). Among the most used methodologies are the quantitative structure–activity relationship (QSAR) and molecular docking. The mathematical equation linking the biological activities of the studied molecules with their chemical characteristics, defined as molecular descriptors, represents one of the most widely used practices for QSAR model representation (Cherkasov et al. 2014). In current practice, the applied representation of molecule structure is

used for the calculation of a considerable number of descriptors, calculated in such fashion that could be further utilized for the development of a relevant QSAR model (Liu and Long 2009; Pérez González et al. 2008). A novel approach involves the use of conformation-independent optimal descriptors which are based on the constitutional and topological molecular features, and the descriptors based on the Simplified Molecular Input Line Entry System (SMILES) notation for QSAR model development, where the Monte Carlo optimization approach is used for this purpose. One of the recently suggested approaches to the manner of overcoming issues related to mechanistic interpretation is the application of SMILES descriptors, since these have a physical meaning and can be associated with molecular fragments (Toropova et al. 2016; Veselinović et al. 2015; Zivkovic et al. 2020), which is why this type of QSAR modeling is conducted under the OECD (Organization for Economic Cooperation and Development) guidelines for QSAR model development.

Several *in silico* methods were administered for the purpose of revealing new compounds having a potential 3C-like protease (3CLpro) inhibition activity in the presented research. Since molecular descriptors based on the SMILES notation were used, along with local graph invariants, the developed QSAR models were conformation-independent. One of the chief aims of this study was to determine the structural requirements or molecular fragments responsible for the 3CLpro inhibition effect. Moreover, this research defined the structural attributes found in small molecules concerning ligand-receptor interactions, which could be used in designing and developing SARS-CoV infection treatment therapeutics. Since the SMILES notation is used for chemical representation by most of the currently utilized chemical databases, this type of modeling could be employed by other scientists for screening purposes, as the results of the developed QSAR models have displayed considerably good values for various validation metrics, both internal and external ones.

## Experimental

### Methods

#### Dataset selection

A group of 84 molecules encompassing a wide range of heterocyclic compound classes, with an established 3C-like protease inhibiting activity against the infectious bronchitis virus (SARS-CoV virus), were selected from the binding database, according to “The Scripps Research Institute Molecular Screening Center” (Gilson et al. 2015), and used for QSAR model development. What needs to be specially

emphasized is that the 3CLpro protease of SARS-CoV-2 has a sequence similarity greater than 95% when compared with that of the bronchitis virus (SARS-CoV virus), which was the main reason for the use of the surrogate enzyme for developing inhibitors against SAR-CoV-2 3CLpro (Liu et al. 2020). All the used molecules have an activity in  $IC_{50}$  (nM) values, which was determined using the same bioassay protocol (QFRET-based dose–response biochemical high throughput screening assay) (Jacobs et al. 2013). The  $IC_{50}$  values were converted to  $pIC_{50}$  ( $pIC_{50} = -\log IC_{50}$ ) values as a studied activity. Afterwards, Open Babel was utilized to canonize the SMILES notations obtained from the above-stated database. Table S1 in the Supplementary Material

molecular graph-based descriptors do not have this feature, and the SMILES notation was used as a highly convenient alternative to them. One of the key features of the SMILES notation-based descriptors is the possibility of relating them to the appropriate molecular fragment. Data concerning the detailed description of the SMILES notation, as well as its application for defining molecule structure can be found in literature (Toropova et al. 2016; Veselinović et al. 2015). The descriptors based on the SMILES notation need a numerical value that is to be used in further QSAR model development, as is the case with all molecular descriptors. The numerical value used for this purpose is defined as correlation weight. Its mathematical representation is the sum of all the defined SMILES descriptors, according to Eq. 1.

$$DCW(T, Nepoch) = zCW(ATOMPAIR) + xCW(NOSP) + yCW(BOND) + tCW(HALO) + rCW(HARD) + \alpha \Sigma CW(S_k) + \beta \Sigma CW(SS_k) + \gamma \Sigma CW(SSS_k) \quad (1)$$

cites the SMILES notations of all the used molecules, with their  $pIC_{50}$ . Furthermore, the obtained data set was divided randomly into the training set (63 compounds, 75%) and the test set (21 compounds, 25%) for three random splits. In addition, the published method was used to verify the normality of the activity distribution (Ojha and Roy 2011).

### Descriptor calculation and QSAR model development

The Monte Carlo optimization method was applied as the main algorithm for the conformation-independent QSAR models development. A hybrid approach was employed for the selection of adequate molecular descriptors as this approach represents the combination of the best features both from the molecular graph and from the descriptors based on the SMILES notation.

Molecular graph-based descriptors were derived from local graph invariants, and the most elementary theoretical graph concepts, such as paths and walks, were used in this research. Literature contains detailed mathematical definitions and applications in QSAR model development (Stoičkov et al. 2018). The following local graph invariants were applied for the purpose of defining the optimal descriptors used in conformation-independent QSAR models: Morgan extended connectivity indices of increasing orders (EC0), path numbers with the length of 2 and 3 ( $p_2$ ,  $p_3$ ), valence shells in the range of 2 and 3 ( $s_2$ ,  $s_3$ ), as well as the number of carbon atom neighbors (*NumberOfCarbon*), and the non-carbon atom neighbors (*NumberOfNonCarbon*).

Medicinal chemists find descriptors with mechanistic interpretations very appealing, in addition to those that could be correlated with molecular fragments. Unfortunately, most

$Z$ ,  $x$ ,  $y$ ,  $t$ ,  $\alpha$ ,  $\beta$  and  $\gamma$  can have the values of 1 (yes) or 0 (no). These values represent coefficients based on which these descriptors are used in QSAR modeling, which is done by following a simple rule—if the coefficient is 1, the descriptor is used, and if it is 0, then it is discarded. In Eq. 1, the  $S_k$  symbol is related to the local descriptors which are associated with one SMILES notation symbol (or two that cannot be separated)—SMILES atoms. The new optimal descriptors, represented with the  $SS_k$  and  $SSS_k$  symbols, respectively, are defined by the linear combinations of two and three SMILES atoms. With the exception of the local ones used in the QSAR modeling, the global SMILES notation descriptors could be used. When compared to the local ones, the global descriptors differ since they are related to the global features of a studied molecule, not only its structure. The following global SMILES notation-based descriptors were used in the presented research: NOSP, HALO, BOND, and ATOMPAIR. These are defined on the basis of the presence or absence of the following chemical elements: nitrogen, oxygen, sulfur and phosphorus (NOSP); fluorine, chlorine and bromine (HALO); double, triple, or stereochemical bonds (BOND), as well as the presence of seven chemical elements: F, Cl, Br, N, O, S, and P (ATOMPAIR). The presence/absence of eight chemical elements (fluorine, chlorine, bromine, iodine, nitrogen, oxygen, sulfur and phosphorus) and different types of chemical bonds (stereo chemical bond, double bond and triple bond) is defined as with the HARD-index a global SMILES notation-based descriptor and represented as a line of eleven symbols. A combination of both SMILES notation (both local and global) and local graph invariant descriptors is utilized in the presented research for the QSAR model deDCW, and is defined as the sum of all the used optimal descriptors' correlation weights, according to Eq. 2.

$$\begin{aligned} DCW(T, N_{\text{epoch}}) = & \Sigma CW(S_k) + \Sigma CW(SS_k) + \Sigma CW(SSS_k) + \Sigma CW(EC0_k) + \Sigma CW(PT2_k) \\ & + \Sigma CW(PT3_k) + \Sigma CW(VS2_k) + \Sigma CW(VS3_k) + \Sigma CW(NNC_k) \end{aligned} \quad (2)$$

where  $S_k$  is one SMILES symbol, defined as the SMILES atom,  $SS_k$  and  $SSS_k$  are the linear combinations of two or three neighbor SMILES atoms,  $EC0_k$  is the Morgan connectivity index of the zero order (a hydrogen suppressed graph was used in this research),  $PT2_k$  and  $PT3_k$  are paths in the length of 2 and 3,  $VS2_k$  and  $VS3_k$  are valence shell 2 and 3, and  $NNC_k$  is Nearest Neighbors (Toropov et al. 2003).

The calculations of all the above mentioned molecular descriptors and the QSAR model development were conducted with the use of the Monte Carlo optimization method and its algorithm, implemented in CORAL (CORrelation and Logic) (<http://www.insilico.eu/coral>) software. The Monte Carlo optimization method depends on two essential parameters for the QSAR model development—the threshold (T), used for eliminating rare molecule features, and the number of epochs ( $N_{\text{epoch}}$ ), which represents the number of iterative processes of the algorithm for the purpose of reaching the top correlation coefficient value. Within the QSAR model developing process, the Monte Carlo optimization method matches molecular descriptors with their numerical values (CW) until CWs are determined. In this manner, the Least Squares method can calculate the DCW ( $T, N_{\text{epoch}}$ ) for the training and test set compounds, as defined in Eq. 3. A systematic review of the applied method is described in the literature (Toropova et al. 2016; Veselinović et al. 2015). The search for the best combination of T and  $N_{\text{epoch}}$  in the presented research was performed within the values of 0–10 for T, and 0–70 for  $N_{\text{epoch}}$ .

$$Ac = C_0 + C_1 \times DCW(T, N_{\text{epoch}}) \quad (3)$$

The quality of the developed QSAR model needs to be evaluated by employing different validation methods. The process should provide information on the benefits of the developed model and whether it could be used for future predictions of the studied activities. The methodology from the published paper was used entirely for this purpose (Veselinović et al. 2015; Ojha et al. 2011; Roy et al. 2008), which involved the calculation of various statistical parameter values, such as the regular and cross-validated correlation coefficient, standard estimation error, mean absolute error (MAE), the Fischer ratio, root-mean-square error,  $R_m^2$ , and MAE-based metrics. What is more, the validation of the developed QSAR models in this research was achieved through data randomization (Y-scrambling test) and with the determination of concordance correlation coefficient (CCC), as well as the novel parameter, dubbed the index of ideality

of correlation (IIC) (Stoičkov et al. 2018; Toropov and Toropova 2017; Veselinović et al. 2018).

## Molecular docking

For molecular docking studies were applied the Molegro Virtual Docker (MVD) software. The studied molecules were drawn using the Marvin sketch (Marvin 6.1.0, 2013, ChemAxon), whereas the MMFF94 force field was utilized to gain their optimal 3D geometry. The protein databank provided the structure of the studied enzyme with the PDB id: 6lu7 (the crystal structure of COVID-19 main protease in a complex with an inhibitor N3) (Jin et al. 2020). The purpose of MVD was obtaining an adequate geometrical orientation of the flexible ligand within the active site of the studied enzyme, surrounded by rigid amino acids, as well as identifying the hydrogen bonds and hydrophobic interactions between them. Finally, MVD was also used to calculate the relevant binding energies, also defined as “scoring” functions (Thomsen and Christensen 2006). The use of these functions is assessing the studied molecules’ inhibitory effect, and the following “scoring” functions were calculated in this research—Hbond, NoHbond, VdW, Steric, Pose energy, MolDock, and Rerank Score. The energy from hydrogen and no hydrogen bond interactions were calculated with HBond and NoHBond90, respectively; the energies from the Van der Waals and Steric interactions were calculated with VdW and Steric “scoring” functions; the overall energy of the best-calculated pose was calculated with the Pose energy. MolDock Score and Rerank Score were calculated as the final estimators of ligand and amino acids from the enzyme’s active site interaction energies, and the whole molecular docking protocol was validated in accordance with the published methodology (Amin et al. 2018, 2019; Jain et al. 2020; Zivkovic et al. 2020). All the crystallized water molecules were removed before the docking. The binding site was computed within spacing in such a way that it was well sampled with the grid resolution of 0.3 Å. The MolDock SE was used as a search algorithm, with the set number of runs being up to 100. The docking procedure parameters were: population size – 50; maximum number of iterations – 1.500; energy threshold – 100.00, and the maximum number of steps – 300. The maximum number of poses to generate was increased to 10 from a default value of 5. Discovery Studio Client v20.1.0.19 was used for showing two-dimensional representations of the interactions between the studied molecules and the amino acids 3CLpro active site.



## Results and discussion

To consider any QSAR model for prediction purposes, its applicability domain (AD) should be defined prior to use (Gadaleta et al. 2016; Gramatica 2007). The methodology described in the literature was used to define the AD for all the developed QSAR models in this study (Tropova et al. 2016; Veselinović et al. 2015). Based on the results obtained, all the molecules fall within the defined AD, and there were no outliers. The numerical values for all the metrics used for the determination of the developed QSAR model's quality are cited in Table 1. On the basis of the results presented, the Monte Carlo optimization method was able to produce a QSAR model which exhibited good reproducibility, and a high predictability potential. When it comes to the 3CLpro inhibitory activity, the second split yielded the best QSAR model, developed with the T value of 1, whereas the best  $N_{\text{epoch}}$  value was 20. Figure 1 presents the graphical representations of the best Monte Carlo optimization runs (the highest value for  $r^2$ ) for the developed QSAR models (all the three splits). Equation 4 contains the mathematical representation of the best-developed model. The reproducibility concordance correlation coefficient (CCC) was used for the assessment of the developed QSAR models, and the obtained numerical values for the CCC indicate that the presented QSAR models show high reproducibility values. The metrics results based on the MAE indicate a GOOD model, thus classifying the developed QSAR model as valid. The numerical values of the index of ideality of correlation (IIC) were calculated for the purpose of making a final estimation regarding the quality of the developed QSAR models, and the obtained results exhibit a high predictive potential for the developed QSAR models. The numerical values calculated for Y-randomization, in which the Y values were scrambled in 1000 trials in ten separate runs, are presented in Table 2, and these were used to evaluate the sturdiness of the developed QSAR models. The presented results indicate that the developed QSAR models were free from correlation by chance because the numerical value for  ${}^cR_p^2$  was higher than 0.5.

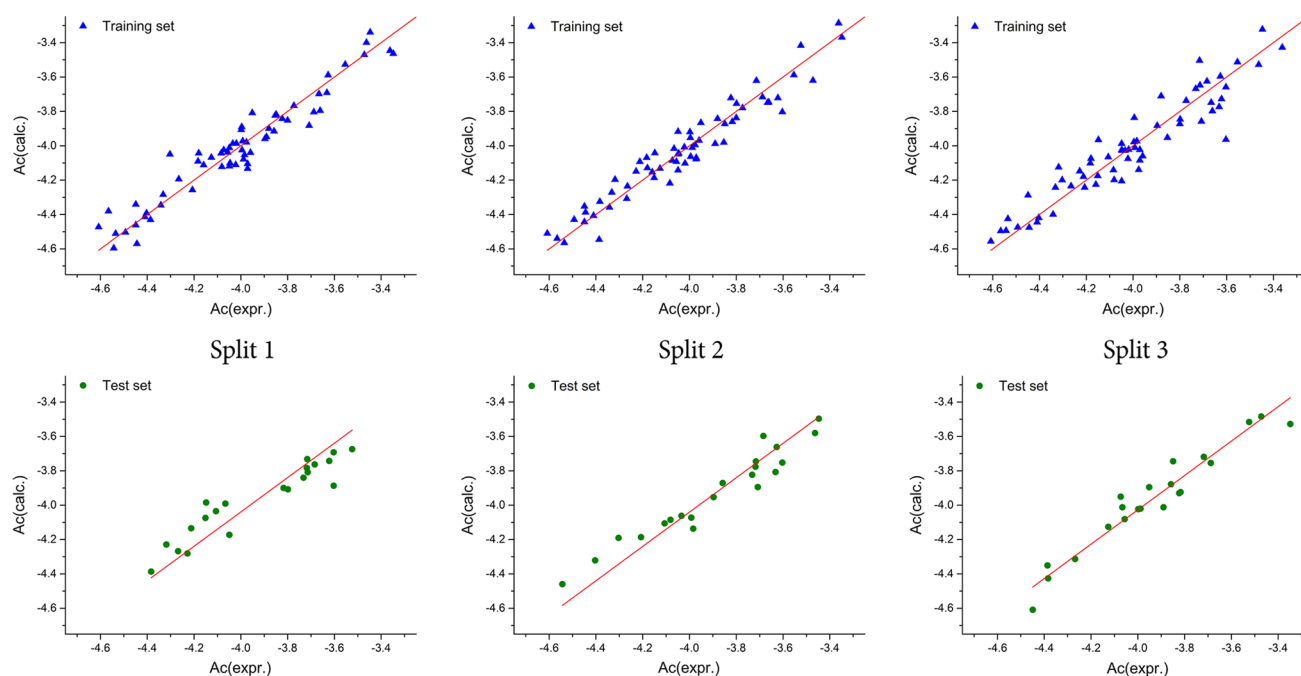
$$\text{pIC}_{50} = -4.9496(\pm 0.0040) + 0.0096 \times \text{DCW}(1, 20) \quad (4)$$

A QSAR model was developed by Kumar and Roy for the same activity, with the use of 2D molecular descriptors and the multiple linear regression (MLR) technique (Kumar and Roy 2020). When compared with this research, where the authors used 69 molecules (16 molecules with outlier behavior were omitted from the original database), all the molecules were used in the presented research as they all fall into the defined AD. The following validation parameters

**Table 1** The statistical quality of the developed QSAR models for 3CLpro inhibitory activity

Run	Training set					Test set											
	$R^2$	CCC	IIC	$Q^2$	$s$	MAE	F	$R^2$	CCC	IIC	$Q^2$	$Q^2_{F1}$	$Q^2_{F2}$	$Q^2_{F3}$	$s$	MAE	F
Split 1	1	0.9393	0.9687	0.8812	0.9353	0.082	0.066	944	0.8723	0.9151	0.9339	0.8623	0.8523	0.9014	0.107	0.073	130
	2	0.9294	0.9634	0.7712	0.9236	0.089	0.072	803	0.8841	0.8897	0.9402	0.8384	0.8268	0.8843	0.115	0.086	145
	3	0.9351	0.9665	0.9367	0.9311	0.085	0.067	880	0.8862	0.8990	0.9413	0.8676	0.8366	0.8909	0.112	0.092	148
Split 2	Av	0.9346	0.9662	0.863	0.9300	0.085	0.068	876	0.8809	0.9013	0.9385	0.8494	0.8386	0.8992	0.111	0.084	141
	1	0.9326	0.9651	0.7242	0.9285	0.084	0.064	844	0.9042	0.9294	0.9456	0.8913	0.8738	0.8923	0.108	0.087	161
	2	0.9314	0.9645	0.9350	0.9271	0.085	0.066	828	0.9243	0.9414	0.9561	0.9102	0.8958	0.9111	0.098	0.075	203
Split 3	3	0.9407	0.9695	0.9396	0.9368	0.079	0.064	968	0.9317	0.9417	0.9653	0.9129	0.8989	0.9137	0.096	0.077	259
	Av	0.9349	0.9664	0.8663	0.9308	0.083	0.065	880	0.9201	0.9375	0.9557	0.9048	0.8895	0.9057	0.107	0.080	208
	1	0.8685	0.9296	0.8472	0.8610	0.120	0.091	403	0.9077	0.9463	0.9527	0.8688	0.9019	0.9235	0.093	0.073	187
Split 3	2	0.8853	0.9391	0.8025	0.8792	0.112	0.086	471	0.9140	0.9172	0.9560	0.8691	0.8642	0.8941	0.109	0.080	202
	3	0.9036	0.9493	0.8641	0.8981	0.102	0.080	572	0.9298	0.9600	0.9642	0.9100	0.9228	0.920	0.084	0.063	252
	Av	0.8858	0.9393	0.8379	0.8794	0.111	0.086	482	0.9172	0.9412	0.9576	0.8891	0.8954	0.9184	0.095	0.072	214

$R^2$  correlation coefficient,  $Q^2$  cross-validated correlation coefficient,  $s$  standard error of estimation, MAE mean absolute error, F Fischer ratio, Av average value for statistical parameters obtained from three independent Monte Carlo optimization runs



**Fig. 1** Graphical representation of the best Monte Carlo optimization runs (the highest value for  $r^2$ ) for the developed QSAR models

**Table 2** Y-randomization of the best QSAR model (best optimization run) for three independent splits

	Split 1		Split 2		Split 3	
	Training	Test	Training	Test	Training	Test
1	0.9351	0.8862	0.9407	0.9317	0.9036	0.9298
2	0.0064	0.0340	0	0.0197	0	0.0726
3	0.0009	0.0019	0.0090	0.0202	0.0235	0.0823
4	0.0935	0.0152	0.0196	0.0042	0.0151	0.0708
5	0.0087	0.0193	0.0309	0.0281	0.0422	0.1384
6	0.0248	0.0008	0.0853	0.0098	0.0061	0.0009
7	0.0035	0.0028	0.0007	0	0.0207	0.0129
8	0.0051	0.0149	0.0004	0.0215	0.0041	0.0002
9	0.0020	0.0096	0.0006	0.0368	0.0145	0.0237
10	0.0061	0.0410	0.0015	0.0006	0.0103	0.0134
$R_r^2$	0.0040	0.1760	0.0246	0.0965	0.0684	0.0780
$C R_p^2$	0.0155	0.0316	0.0173	0.0237	0.0205	0.0493
	0.9274	0.8703	0.932	0.9198	0.8933	0.9048

$$C R_p^2 = R \times (R^2 - R_r^2)^{1/2} \text{ should be } > 0.5 \text{ (Ojha and Roy 2011)}$$

were present in the QSAR model reported by Kumar and Roy: Internal validation)  $R^2 = 0.764$ ,  $Q^2 = 0.627$ ; External)  $Q^2_{F1} = 0.727$ ,  $Q^2_{F2} = 0.652$ ,  $Avg r_{m2} = 0.610$ ,  $\Delta r_{m2} = 0.110$ ,  $MAE = 0.127$ . In addition to the validation parameters cited in Table 1, the QSAR models developed with the application of the Monte Carlo optimization method had  $Avg r_{m2}$  and  $\Delta r_{m2}$  values of 0.8966–0.6097 and 0.1885–0.0191, respectively. When comparing the two approaches, by applying solely the reported statistical parameters, in terms of predictability, the

QSAR model developed with the application of the Monte Carlo optimization method is superior.

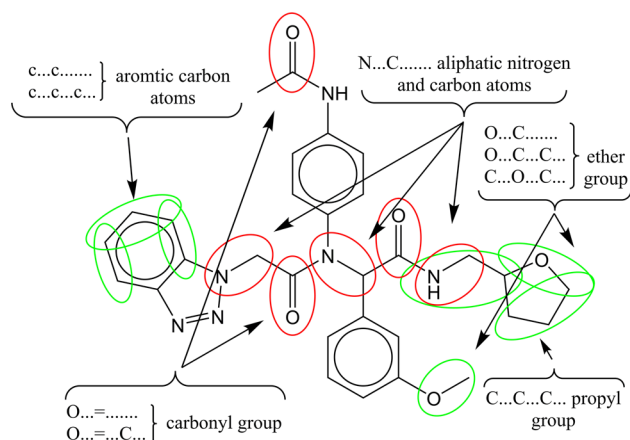
The determination of molecular fragments, defined as the SMILES notation optimal descriptors having a positive and negative impact on the studied activity, was one of the chief aims of this research (Halder 2018; Kumar et al. 2019; Manisha et al. 2019; Toropov et al. 2019). The full list of molecular descriptors, both based on the molecular graph and the SMILES notation, are shown in Table S2 (Supplementary material), while the selected

**Table 3** Selected promoters of inhibition activity increase/decrease

Promoters of inhibition activity increase	Promoters of inhibition activity decrease
(...(...	(...Br.(...
(...N...(...	(...Cl.(...
(...O...(...	=...1.....
+++CL-N===	C...2...(...
++++N-O===	N.....
++++O-B2==	N...2.....
C.../.....	N...=.....
C.../...C...	N...=...C...
C...C...C...	N...C...(...
C...O...C...	N...C.....
BOND10000000	N...1.....
N...C...C...	O...=.....
O...1.....	O...=...C...
O...C.....	O...C...N...
O...C...C...	n...n.....
s...c.....	s...n.....
s.....	s...n...n...
c...c.....	C...(.....
c...c...c...	N...(.....

ones are detailed in Table 3. The calculation example is presented in Table S2, where molecular graph-based descriptors were omitted with the aim of achieving an easier interpretation.

The selected molecular fragments, defined as the SMILES notation descriptors which have a positive influence on the studied activity (a lower concentration of the compound is required to achieve IC<sub>50</sub>) are as follows: “(... (...”, “(...N...(...” and “(...O...(...”—additional branching in the molecule, and additional branching in the molecule realized by adding aliphatic nitrogen and oxygen atom; “++++CL-N===”, “++++N-O===” and “++++O-B2==”—the presence of chlorine and nitrogen atoms in the molecule, the presence of oxygen and nitrogen atoms in the molecule, and the presence of oxygen and a double bond in the molecule; “BOND10000000”, “C.../.....”, and “C.../...C...”—double bond between two carbon atoms; “C...C...C...”—propyl group; “N...C...C...”—ethyl amine group; “O...1.....” oxygen atom inside the ring; “O...C.....”, “O...C...C...” and “C...O...C...”—methoxy, ethoxy and dimethoxy groups, respectively; “s...c.....” and “s.....”—aromatic sulfur atom; “c...c.....” and “c...c...c...”—aromatic carbon atoms. The same analysis may be performed for molecular fragments, defined as the SMILES notation descriptors with a negative impact on the studied activity. The selected molecular fragments are: “C...(.....”, “N...C...(...” and

**Fig. 2** Identified molecular fragments with their impact on 3CLpro inhibitory activity

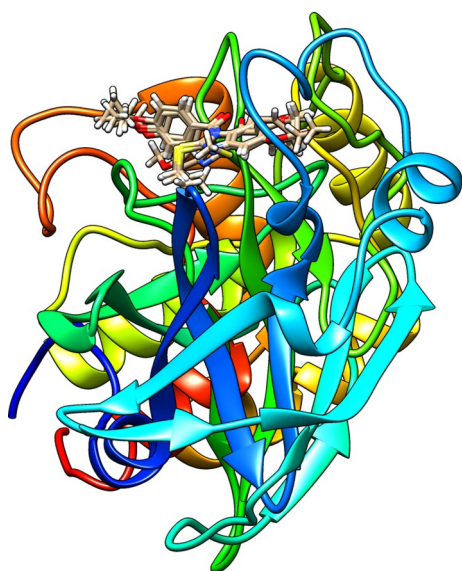
“N...(.....”—simple molecular branching on the carbon and nitrogen atom; “(...Br.(...” and “(...Cl.(...”—additional molecular branching with bromine and chlorine atoms; “C...2...(...” and “N...2.....”—second ring inside the molecule, and second ring inside the molecule with a nitrogen atom; “=...1.....”—double bond within the ring; “N.....” and “- aliphatic nitrogen; “N...=.....”, “N...=...C...”, “O...=.....” and “O...=...C...”—the double bond between the nitrogen or oxygen atom and the carbon atom; “n...n.....”, “s...n.....” and “s...n...n...”—aromatic nitrogen atoms and the presence of both aromatic nitrogen and sulfur atoms. The graphical representation of the impact of molecular fragments on the selected molecule activity is displayed in Fig. 2. The green color in Fig. 2 indicates a positive influence, while red indicates a negative impact on the studied activity.

Identified molecular fragments were used for the computer-aided design of novel 3CLpro inhibitors. The molecule with the highest inhibitory effect (compound 1 from the dataset) was used as a starting/template molecule for the purpose of producing molecules with a higher inhibition potential. Table 4 indicates the SMILES notation of all the designed molecules, along with their calculated values for the -pIC<sub>50</sub>, utilizing the best-developed QSAR model, while their chemical structures are presented in Fig. 3.

Added molecular fragments were identified as structural alerts that have a positive effect on the studied activity in all the designed molecules (A1-A5). Molecules A1-A3 had the following molecular fragments added—the isopropyl and cyclopropyl group on the different parts of template molecule A. The groups are defined as the following SMILES notation descriptors, all promoters of the studied activity increase—“(... (...”, “1.....”, “C.....”, “C...C.....”, “C...C...C...”. Molecules A4 and A5 have added methyl ether group. They all have a positive







**Fig. 4** The best-calculated poses for designed molecules (A1-A5) and template molecule A inside the active site of 3CLpro

therapeutics. Compounds should possess physical and chemical parameters enabling them to be considered as potential medication. The application of Lipinski's "Rule of Five" represents one of the extensively used approaches. These rules determine the absorption/permeation of the compound, and if the compound possesses a molecular weight higher than 500, logP over 5, and the number of hydrogen bond donors and acceptors is higher than 5 and 10, respectively, it will most likely have poor absorption/permeation. In addition, the number of rotatable bonds, associated with conformational molecular flexibility, is used to assess the efficacy of binding to receptors/channels, as well as their bioavailability. The molecule should have a satisfactory number of rotatable bonds, i.e. 10 or fewer (Lipinski et al. 2001). In addition to Lipinski's "Rule of Five", Veber proposed having an additional set of rules for a molecule's drug-likeness assessment (Veber et al. 2002), and they can be summarized as acceptable when the oral bioavailability is more

likely to occur when the molecule has 10 or fewer rotatable bonds, a polar surface area equal to or less than  $140 \text{ \AA}^2$ , and 12 or fewer hydrogen donors and acceptors. For template molecule A and all the designed molecules, the physical and chemical parameters defined above are calculated using the Molinspiration Cheminformatics software ([www.molinspiration.com](http://www.molinspiration.com)), and presented in Table 6. In accordance with the obtained results, all the molecules were in accordance with Lipinski's "Rule of Five", and the Veber rules, which may be associated with their potentially good solubility and permeability through biological membranes, thus leading to satisfactory bioavailability, which is why they may be considered as potential therapeutics.

## Conclusion

The present research demonstrates the development of new robust and reliable QSAR models for 3CLpro inhibition based on the Monte Carlo optimization method as the main model, and developed on the basis of optimal descriptors derived both from a local graph and the SMILES notation invariants. The assessment of the developed QSAR model's robustness and its predictive potential was achieved by applying various statistical techniques and obtained numerical values, used to validate the developed QSAR model, indicating that it has high applicability. The applied methodology was able to determine the molecular fragments, used as the SMILES notation fragments in QSAR modeling, which have a positive and negative impact on 3CLpro inhibition. They were used for the computer-aided design of novel 3CLpro inhibitors with a potentially higher activity when compared to template molecule A. To assess the developed QSAR models' predictability all designed molecules and template molecule A were subjected to molecular docking studies with 3CLpro. Their binding potential was assessed according to the comparison of numerical values for all calculated "scoring" functions, related to energies from the interactions between the molecules and the amino

**Table 6** Calculated physico-chemical properties of the designed molecules used for drug-likeness determination

Molecule	miLogP	TPSA	natoms	MW	nON	nOHNH	nviolations	nrotb	volume
A	3.43	62.6	25	356.4	6	0	0	3	300.61
A1	4.17	62.6	27	384.46	6	0	0	4	333.99
A2	4.3	62.6	28	396.47	6	0	0	5	340.44
A3	4.17	62.6	27	384.46	6	0	0	4	333.99
A4	3.4	71.83	27	386.43	7	0	0	5	326.39
A5	3.4	71.83	27	386.43	7	0	0	5	326.39

*miLogP* octanol–water partition coefficient, *TPSA* molecular polar surface area, *natoms* number of atoms in molecule, *MW* molecular weight, *nON* number of nitrogen and oxygen atoms in molecule, *nOHNH* number of amino and hydroxyl groups, *nviolations* number of "Rule of five" violations, *nrotb* number of rotatable bonds, *volume* volume of molecule

acids within the active site of the 3CLpro, and calculated pIC<sub>50</sub> from the best QSAR model developed with the Monte Carlo optimization method. Values for all defined “scoring” functions calculated for each designed molecule and calculated values for pIC<sub>50</sub> show high inter-correlation. Also, all designed molecules have potentially good pharmacokinetic properties and possess drug-likeness. The methodology presented in this research paper can be applied in the search for novel therapeutics for the treatment of COVID-19 based on 3CLpro inhibition.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11696-022-02170-8>.

**Acknowledgements** This work is supported by the Ministry of Education and Science, the Republic of Serbia and the Faculty of Medicine, University of Niš, Republic of Serbia (project No. 70). The authors would like to thank the Ministry of Education, Science and Technological Development of Republic of Serbia (Grant No: 451-03-9/2021-14/200113) for financial support.

## Declarations

**Conflict of interest** We have no conflicts of interest to disclose.

## References

- Akaji K, Konno H (2020) Design and evaluation of anti-SARS-coronavirus agents based on molecular interactions with the viral protease. *Molecules* 25:3920. <https://doi.org/10.3390/molecules25173920>
- Amin SA, Jha T (2020) Fight against novel coronavirus: a perspective of medicinal chemists. *Eur J Med Chem*. <https://doi.org/10.1016/j.ejmech.2020.112559>
- Amin SA, Adhikari N, Gayen S, Jha T (2018) Exploring pyrazolo[3,4-d]pyrimidine phosphodiesterase 1 (PDE1) inhibitors: a predictive approach combining comparative validated multiple molecular modelling techniques. *J Biomol Struct Dyn* 36:590–608. <https://doi.org/10.1080/07391102.2017.1288659>
- Amin SA, Adhikari N, Gayen S, Jha T (2019) Reliable structural information for rational design of benzoxazole type potential cholesteryl ester transfer protein (CETP) inhibitors through multiple validated modelling techniques. *J Biomol Struct Dyn* 37:4528–4541. <https://doi.org/10.1080/07391102.2018.1552895>
- Amin SA, Banerjee S, Gayen S, Jha T (2021) Protease targeted COVID-19 drug discovery: what we have learned from the past SARS-CoV inhibitors? *Eur J Med Chem*: <https://doi.org/10.1016/j.ejmech.2021.113294>
- Chen S, Chen LL, Luo HB, Sun T, Chen J, Ye F, Cai JH, Shen JK, Shen X, Jiang HL (2005) Enzymatic activity characterization of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer technique I. *Acta Pharmacol Sin* 26:99–106. <https://doi.org/10.1111/j.1745-7254.2005.00010.x>
- Chen Y, Liu Q, Guo D (2020) Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol* 92:418–423. <https://doi.org/10.1002/jmv.25681>
- Cherkasov A, Muratov EN, Fourches D, Varnek D, Baskin A, Cronin II, Dearden MJ, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, 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. <https://doi.org/10.1021/jm4004285>
- Deng X, St John SE, Osswald HL, O'Brien A, Banach BS, Sleeman K, Ghosh AK, Mesecar AD, Baker SC (2014) Coronaviruses resistant to a 3C-like protease inhibitor are attenuated for replication and pathogenesis, revealing a low genetic barrier but high fitness cost of resistance. *Int J Virol* 20:11886–11898. <https://doi.org/10.1128/JVI.01528-14>
- Ekins S, Mestres J, Testa B (2007) In silico pharmacology for drug discovery: applications to targets and beyond. *Br J Pharmacol* 152:9–20. <https://doi.org/10.1038/sj.bjp.0707306>
- Gadaleta D, Mangiatordi GF, Catto M, Carotti A, Nicolotti O (2016) Applicability domain for QSAR models: where theory meets reality. *IJQSPR* 1:45–63. <https://doi.org/10.4018/IJQSPR.2016010102>
- Gilson MK, Liu T, Baitaluk MG, Nicola G, Hwang L, Chong J (2015) BindingDB: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res* 19:1045–1053. <https://doi.org/10.1093/nar/gkv1072>
- Gramatica P (2007) Principles of QSAR models validation: internal and external. *QSAR Comb Sci* 26:694–701. <https://doi.org/10.1002/qsar.200610151>
- Guo Y-R, Cao Q-D, Hong Z-S, Tan Y-Y, Chen S-D, Jin H-J, Tan K-S, Wang D-Y, Yan Y (2020) The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—An update on the status. *Mil Med Res*. <https://doi.org/10.1186/s40779-020-00240-0>
- Halder A (2018) Finding the structural requirements of diverse HIV-1 protease inhibitors using multiple QSAR modelling for lead identification. *SAR QSAR Environ Res* 29:911–933. <https://doi.org/10.1080/1062936X.2018.1529702>
- Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, Müller MA, Drosten C, Pöhlmann S (2020) SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* 181:271–280. <https://doi.org/10.1016/j.cell.2020.02.052>
- Jacobs J, Zhou S, Dawson E, Daniels JS, Hodder P, Tokars V, Mesecar A, Lindsley CW, Stauffer SR (2013) Discovery of non-covalent inhibitors of the SARS main proteinase 3CLpro, In Probe Reports from the NIH Molecular Libraries Program [Internet], National Center Biotechnol Inform (US), Available at <https://www.ncbi.nlm.nih.gov/books/NBK133447>
- Jain S, Amin SA, Adhikari N, Jha T, Gayen S (2020) Chemical-informatics approach to COVID-19 drug discovery: Monte Carlo based QSAR, virtual screening and molecular docking study of some in-house molecules as papain-like protease (PLpro) inhibitors. *J Biomol Struct Dyn* 38:66–77. <https://doi.org/10.1080/07391102.2020.1780946>
- Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z (2020) Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *J Gen Inter Med* 35:1545–1549. <https://doi.org/10.1007/s11606-020-05762-w>
- Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, Zhang B, Li X, Zhang L, Peng C, Duan Y, Yu J, Wang L, Yang K, Liu F, Jiang R, Yang X, You T, Liu X, Yang X, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H (2020) Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582:289–293. <https://doi.org/10.1038/s41586-020-2223-y>
- Kampf G, Todt D, Pfaender S, Steinmann E (2020) Persistence of coronaviruses on inanimate surfaces and its inactivation with biocidal agents. *J Hosp Infect* 104:246–251. <https://doi.org/10.1016/j.jhin.2020.01.022>

- Kumar V, Roy K (2020) Development of a simple, interpretable and easily transferable QSAR model for quick screening antiviral databases in search of novel 3CLike protease (3CLpro) enzyme inhibitors against SARS-CoV diseases. *SAR QSAR Environ Res* 31:511–526. <https://doi.org/10.1080/1062936X.2020.1776388>
- Kumar P, Kumar A, Sindhu J (2019) In silico design of diacylglycerol acyltransferase-1 (DGAT1) inhibitors based on SMILES descriptors using Monte-Carlo method. *SAR QSAR Environ Res* 30:525–541. <https://doi.org/10.1080/1062936X.2019.1629998>
- Kumar D, Malviya R, Sharma PK (2020) Corona virus: A review of COVID-19. *EJMO* 4:8–25. <https://doi.org/10.14744/ejmo.2020.51418>
- Lai C-C, Shih T-P, Ko W-C, Tang H-J, Hsueh P-R (2020) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Ag.* <https://doi.org/10.1016/j.ijantimicag.2020.105924>
- Lin CW, Tsai FJ, Tsai CH, Lai CC, Wan L, Ho TY, Hsieh CC, Lee Chao PD (2005) Anti-SARS coronavirus 3C-like protease effects of Isatis indigotica root and plant-derived phenolic compounds. *Antivir Res* 68:36–42. <https://doi.org/10.1016/j.antiviral.2005.07.002>
- Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliver Rev* 46:3–26. [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0)
- Liu P, Long W (2009) Current mathematical methods used in QSAR/QSPR studies. *Int J Mol Sci* 10:1978–1998. <https://doi.org/10.3390/ijms10051978>
- Liu C, Zhou Q, Li Y, Garner LV, Watkins SP, Carter LJ, Smoot J, Albaiu D (2020) Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. *ACS Cent Sci* 6:315–331. <https://doi.org/10.1021/acscentsci.0c00272>
- Manisha CS, Kumar P, Kumar A (2019) Development of prediction model for fructose-1,6-bisphosphatase inhibitors using the Monte Carlo method. *SAR QSAR Environ Res* 30:145–159. <https://doi.org/10.1080/1062936X.2019.1568299>
- Ojha PK, Roy K (2011) Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection. *Chemometr Intell Lab* 109:146–161. <https://doi.org/10.1016/j.chemolab.2011.08.007>
- Ojha PK, Mitra I, Das RN, Roy K (2011) Further exploring rm2 metrics for validation of QSPR models. *Chemometr Intell Lab* 107:194–205. <https://doi.org/10.1016/j.chemolab.2011.03.011>
- Pedersen SF, Ho Y-C (2020) SARS-CoV-2: a storm is raging. *J Clin Invest* 130:2202–2205. <https://doi.org/10.1172/JCI137647>
- Peng X, Xu X, Li Y, Cheng L, Zhou X, Ren B (2020) Transmission routes of 2019-nCoV and controls in dental practice. *Int J Oral Sci.* <https://doi.org/10.1038/s41368-020-0075-9>
- Pérez González M, Terán C, Sañaz-Urra L, Teijeira M (2008) Variables selection methods in QSAR: an overview. *Curr Top Med Chem* 8:1606–1627. <https://doi.org/10.2174/156802608786786552>
- Ramajayam R, Tan KP, Liu HG, Liang PH (2010) Synthesis and evaluation of pyrazolone compounds as SARS-coronavirus 3C-like protease inhibitors. *Bioorg Med Chem* 18:7849–7854. <https://doi.org/10.1016/j.bmc.2010.09.050>
- Rothan HA, Byrareddy SN (2020) The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J Autoimm.* <https://doi.org/10.1016/j.jaut.2020.102433>
- Roy PP, Leonard JT, Roy K (2008) Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemometr Intell Lab* 90:31–42. <https://doi.org/10.1016/j.chemolab.2007.07.004>
- Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB (2020) Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. *JAMA J Am Med Assoc.* <https://doi.org/10.1001/jama.2020.6019>
- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F (2020) Structural basis of receptor recognition by SARS-CoV-2. *Nature* 30:1–8. <https://doi.org/10.1038/s41586-020-2179-y>
- Singhal T (2020) A review of coronavirus disease-2019 (COVID-19). *Indian J Pediatr* 87:281–286. <https://doi.org/10.1007/s12098-020-03263-6>
- Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R (2020) World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 76:71–76. <https://doi.org/10.1016/j.ijsu.2020.02.034>
- Stoičkov V, Stojanović D, Tasić I, Šarić S, Radenković D, Babović P, Sokolović D, Veselinović AM (2018) Development of non-peptide ACE inhibitors as novel and potent cardiovascular therapeutics: an in silico modelling approach. *Struct Chem* 29:441–449. <https://doi.org/10.1080/1062936X.2018.1485737>
- Tabeshpour J, Sahebkar A, Zirak MR, Zeinali M, Hashemzaei M, Rakhshani S, Rakhshani S (2018) Computer-aided drug design and drug pharmacokinetic prediction: a mini-review. *Curr Pharm Design* 24:3014–3019. <https://doi.org/10.2174/1381612824666180903123423>
- Thomsen R, Christensen MH (2006) MolDock: a new technique for high-accuracy molecular docking. *J Med Chem* 49:3315–3321. <https://doi.org/10.1021/jm051197e>
- Toropov AA, Toropova AP (2017) The index of ideality of correlation: a criterion of predictive potential of QSPR/QSAR models? *Mutat Res-Gen Tox En* 819:31–37. <https://doi.org/10.1016/j.mrgentox.2017.05.008>
- Toropov AA, Duchowicz P, Castro EA (2003) Structure-toxicity relationships for aliphatic compounds based on correlation weighting of local graph invariants. *Int J Mol Sci* 4:272–283. <https://doi.org/10.3390/i4050272>
- Toropov AA, Toropova AP, Selvestrel G, Benfenati E (2019) Idealization of correlations between optimal simplified molecular input-line entry system-based descriptors and skin sensitization. *SAR QSAR Environ Res* 30:447–455. <https://doi.org/10.1080/1062936X.2019.1615547>
- Toropova MA, Raška I Jr, Toropov AA, Raškova M (2016) The utilization of the Monte Carlo technique for rational drug discovery. *Comb Chem High T Sc* 19:676–687. <https://doi.org/10.2174/1386207319666160725145852>
- Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. *J Med Chem* 45:2615–2623. <https://doi.org/10.1021/jm020017n>
- Vellingiri B, Jayaramayya K, Iyer M, Narayanasamy A, Govindasamy V, Giridharan B, Ganesan S, Venugopal A, Venkatesan D, Ganesan H, Rajagopalan K, Rahman PKSM, Cho S-G, Kumar NS, Subramaniam MD (2020) COVID-19: a promising cure for the global panic. *Sci Total Environ.* <https://doi.org/10.1016/j.scitotenv.2020.138277>
- Veselinović AM, Veselinović JB, Živković JV, Nikolić GM (2015) Application of smiles notation based optimal descriptors in drug discovery and design. *Curr Top Med Chem* 15:1768–1779. <https://doi.org/10.2174/1568026615666150506151533>
- Veselinović AM, Toropov A, Toropova A, Stanković-Dordević D, Veselinović JB (2018) Design and development of novel antibiotics based on FtsZ inhibition—in silico studies. *New J Chem* 42:10976–10982. <https://doi.org/10.1039/C8NJ01034J>
- Wu D, Wu T, Liu Q, Yang Z (2020) The SARS-CoV-2 outbreak: what we know. *Int J Infect Dis* 94:44–48. <https://doi.org/10.1016/j.ijid.2020.03.004>

- Xiao WX, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, Li SB (2020) Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ* 368:606. <https://doi.org/10.1136/bmj.m606>
- Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved  $\alpha$ -ketoamide inhibitors. *Science* 368:409–412. <https://doi.org/10.1126/science.abb3405>
- Zhavoronkov A, Aladinskiy V, Zhebrak A, Zagribelnyy B, Terentiev V, Bezrukov DS, Polykovskiy D (2020) Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches. *Insilico Med Hong Kong Ltd A* 307:E1. <https://doi.org/10.26434/chemrxiv.11829102.v2>
- Zivkovic M, Zlatanovic M, Zlatanovic N, Golubović M, Veselinović AM (2020) The application of the combination of Monte Carlo optimization method based QSAR modeling and molecular docking in drug design and development. *Mini-Rev Med Chem* 20:1389–1402. <https://doi.org/10.2174/1389557520666200212111428>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.