ISSN 0027-1314, Moscow University Chemistry Bulletin, 2021, Vol. 76, No. 2, pp. 95–113. © Allerton Press, Inc., 2021.
Russian Text © The Author(s), 2021, published in Vestnik Moskovskogo Universiteta, Seriya 2: Khimiya, 2021, No. 3, pp. 192–212.

Virtual Screening and Molecular Design of Potential SARS-COV-2 Inhibitors

O. V. Tinkov[a, b, *], V. Yu. Grigorev[c], and L. D. Grigoreva[d]

[a] *Medical Department, Shevchenko Transnistria State University, Tiraspol, 3300 Moldova*
[b] *Military Institute of the Ministry of Defense, Tiraspol, 3300 Moldova*
[c] *Institute of Physiologically Active Compounds, Russian Academy of Science, Chernogolovka, Moscow oblast, 142432 Russia*
[d] *Department of Fundamental Physical-Chemical Engineering, Moscow State University, Moscow, 119991 Russia*
**e-mail: oleg.tinkov.chem@mail.ru*

Received January 11, 2021; revised January 14, 2021; accepted January 20, 2021

Abstract—According to recent studies, the main M^{pro} protease of the SARS-CoV-2 virus, which is the most important target in the development of promising drugs for the treatment of COVID-19, is evolutionarily conservative and has not undergone significant changes compared with the main M^{pro} protease of the SARS-CoV virus. Many researchers note the similarity between the binding sites of the main M^{pro} protease of SARS-CoV and SARS-CoV-2 viruses; thus, with the spreading epidemic, further studies on inhibitors of the main M^{pro} protease of the SARS-CoV virus to fight COVID-19 seems logical. In the course of the study, satisfactory QSAR models are built using simplex, fractal, and HYBOT descriptors; the Partial Least Squares (PLS), Random Forest (RF), Support Vectors, Gradient Boosting (GBM) methods; and the OCHEM Internet platform (https://ochem.eu), in which different types of molecular descriptors and machine learning methods are implemented. The structural interpretation, which allowed us to identify molecular fragments that increase and decrease the activity of SARS-CoV inhibitors, is performed for the obtained models. The results of the structural interpretation are used for the rational molecular design of potential SARS-CoV-2 inhibitors. The resulting QSAR models are used for the virtual screening of 2087 FDA-approved drugs.

Keywords: M^{pro} protease, QSAR, molecular descriptors, machine learning, structural interpretation

DOI: 10.3103/S0027131421020127

INTRODUCTION

In 2002, the global community was faced with the SARS-CoV coronavirus, with caused an epidemic of atypical pneumonia (severe acute respiratory syndrome). The first cases of infection with the SARS-CoV coronavirus were detected in Southern China, before the epidemic spread to 29 countries, as a result of which more than 8000 people were infected, while 916 persons died [1]. In 2012, the second epidemic, caused by the spread of a coronavirus, in this case the MERS-CoV virus, which is characterized by higher lethality, was recorded in Saudi Arabia. By the end of August 2015, 574 persons died among the total number of infected people (1511 patients); i.e., the lethality from the MERS-CoV coronavirus was over 37% in contrast to SARS-CoV, for which the lethality was estimated at 10%.

A number of experts forecast the emergence of a threat for the human race, which will be caused by a new type of coronavirus [3]. The authors of this study proposed that the new coronavirus, circulating in the Chinese populations of horseshoe bats, will bind to the human angiotensin converting enzyme II (ACE2) followed by the efficient replication in the cell of the respiratory system.

Unfortunately, the prognoses of the above-mentioned experts turned out to be right, and the Chinese public health authorities recorded the first case of infection with the new SARS-CoV-2 coronavirus on December 8, 2019 [4]. The level of the lethality from the new SARS-CoV-2 coronavirus is estimated to be lower (about 7%) than for SARS-CoV and MERS-CoV; however, it was shown that only 48 days are required for infection of the first 1000 patients with the SARS-CoV-2 coronavirus, while 130 days are required for infection with SARS-CoV, and two-and-a-half years for MERS-CoV [5].

The pandemic caused by the new SARS-CoV-2 coronavirus represents a serious medical and socio-economic problem for all mankind.

The drug Favipiravir recommended in the Russian Federation for the treatment of COVID-19 [6], according to a number of researchers, has a teratogenic effect [7]. Thus, the search and development of highly effective and safe drugs that can stop the spread of the COVID-19 pandemic is an urgent issue.

Chemoinformatics methods can provide significant assistance in reducing the time and financial costs in repositioning and developing new drugs [8–10]. Since the emergence of the new SARS-CoV-2 coronavirus, a number of studies using the methods of molecular docking, molecular dynamics, and pharmacophore analysis have been conducted in this direction [11–22].

Currently, the molecular structure of potential SARS-CoV-2 inhibitors is being considered from different points of view. The resulting diverse information is of significant interest to the world scientific and medical community. In the studies conducted, the most significant target for developing drugs is the main protease of the virus M^{pro}, also known as 3-chymotrypsin-like protease ($3CL^{pro}$), which plays a key role in replication of coronaviruses. It was found that this enzyme, being evolutionarily conservative, did not undergo significant changes in contrast to the main M^{pro} protease of the SARS-CoV virus, which caused an outbreak of acute respiratory syndrome in 2002–2003 [23–27]. Previous studies also indicate the conservatism of the M^{pro} sequences and spatial M^{pro} structures of different types of coronaviruses [28]. At the same time, close homologues of this enzyme have not been identified in the human body, which has a positive effect on the specificity and a decrease in the number of potential side effects of inhibitors of the main M^{pro} protease [29].

In the study [25], a virtual screening of the DrugBank library of chemical compounds was performed based on the similarity of the binding sites of the main M^{pro} protease of SARS-CoV and SARS-CoV-2 viruses using the method of molecular docking [30]. As a result, a list of ten potential inhibitors of the main M^{pro} protease was proposed, which, according to the authors [25], are the most promising for combating SARS-CoV-2.

In the study [31], compounds included in the list of drugs of traditional Chinese medicine were initially selected. For these compounds, an assessment of their pharmacokinetic characteristics such as adsorption, distribution, metabolism, and excretion is given. The most promising compounds were studied using molecular docking. The next step was the selection of medicinal herbs that contain at least two compounds proposed in the course of molecular docking. As a result of the study, the authors of [31] identified 26 medicinal herbs of Chinese medicine that are potentially promising for the treatment of the COVID-19 disease caused by the SARS-CoV-2 coronavirus.

The publication [23], which presents the results of the consensus in vitro and in silico screening, deserves special attention. The authors studied a database of more than 10000 compounds, for which the binding to M^{pro} of the SARS-CoV-2 coronavirus was experimentally measured by the method of fluorescent resonance energy transfer. The most promising seven compounds were further investigated for their ability to prevent the infection of cells with the SARS-CoV-2 virus. The in silico screening of these 10000 compounds was performed using the Glide v8.2 and Maestro software (Schrödinger). According to the consensus results, the most promising were ebselen (2-phenyl-1,2-benzoselenazol-3-one, CAS no. 60940-34-3), carmofur (1-hexylcarbamoyl-5-fluorouracil, CAS no. 61422-45-5), compound TDZD-8 (2-methyl-4-(phenylmethyl)-1,2-athiadiazolidine-3,5-dione, CAS no. 327036-89-5), and N3 peptidomimetic, the previously proposed SARS-CoV inhibitor, which covalently binds to M^{pro} according to the Michael reaction.

The authors of [32] conducted a virtual screening of 1.3 billion molecules in order to identify the most active inhibitors of the main M^{pro} protease of the SARS-CoV-2 virus. In this study, the recently developed Deep Docking algorithm, which integrated classical docking and the methodological foundations for constructing QSAR (Quantitative Structure–Activity Relationship) models, was used; this allowed them to increase the screening performance compared with traditional docking methods. The authors note that the scoring functions are determined by the used docking methods, and the QSAR models were used to optimize the virtual screening. Based on the results of the virtual screening, a hit list of 1000 compounds was proposed, which is available for free download at https://drive.google.com/drive/folders/1xgA8ScPRqIunxEAXFrUEkavS7y3tLIMN.

Using deep learning the authors performed the study [33] in which models that describe the structure of compounds using character strings composed according to the rules of SMILES [34] were developed. The principle of the used prediction method is based on a technology called natural language processing [35], used in the analysis of human speech by a computer, but in this case the language is a string of characters written according to the SMILES rules and the sequence of the target protein. In order to identify and analyze regularities, convolutional neural networks were used [36]. The study predicted the activity of inhibitors of the main M^{pro} protease, RNA replicase (RNA-dependent RNA-polymerase, RdRP), helicase, and a number of other enzymes of the SARS-CoV-2 virus. Molecular docking (in particular, the AutoDock Vina v.1.1.2 program) was used for the comparative study. As a result, the authors identified three drugs against HIV (ritonavir, atazanavir, efavirenz), as well as the antiviral agent ganciclovir.

The TMPRSS2 protease (Transmembrane protease, serine 2, membrane-bound serine protease) serves as another target for the fight against the coronavirus; its inhibitors can prevent the entry of the virion into the cell [37]. However, the number of studies devoted to the computer modeling of TMPRSS2 protease inhibitors is significantly less than that of the main M^{pro} protease. Thus, we can note the publication [38], in which a virtual screening of a database con-

taining more than 30000 natural compounds was performed using molecular docking and pharmacophore analysis. For the selected 12 compounds, the authors evaluated adsorption, distribution, metabolism, elimination, and toxicity. The small molecule compound geniposide (CAS no. 24512-63-8) turned out to be the most promising.

Thus, the majority of works was carried out using the method of molecular docking, which, like any method of research, has limitations. In particular, the methodological difficulties of docking are associated with considering conformations of the ligand, the choice of methods for constructing the scoring function, and the flexibility of receptors. The main methods of molecular docking and their inherent limitations are described in detail in the reviews [39–41].

QSAR is an alternative method of computer drug development, which has successfully solved various problems [42]. The literature contains information on the similarity of the binding sites of the main M^{pro} protease of SARS-CoV and SARS-CoV-2 viruses, which were confirmed in the course of independent studies [23–27]. In relation to this, we assumed that the assessment of potential SARS-CoV-2 inhibitors during drug development can be realized using QSAR models of SARS-CoV inhibitors.

The authors of [43] developed a QSAR model of SARS-CoV inhibitors using 3D-QSAR methods (CoMFA, CoMSIA), the limiting feature of which is the ambiguity of the three-dimensional alignment of the structures of the studied compounds [44]. In another study [45], 33 QSAR models of SARS-CoV inhibitors were developed, but the authors did not provide indicators of their predictive ability, assessed using compounds of the test sample.

In accordance with the fifth principle of QSAR modeling, developed by the OECD expert group [46], the interpretation of the obtained models is desirable. In the reviewed publications [43, 45], there is no structural interpretation of the QSAR models, which does not allow conducting molecular design and limits the use of simulation results for studying the mechanisms of biological reactions [47].

Recently, a study was published [48] in which acceptable QSAR models of inhibitors of the main M^{pro} protease of the SARS-CoV virus, which were developed using PaDEL and Dragon descriptors, and the method of multiple linear regression (MLR) were proposed in the search for effective drugs against COVID-19. Using the developed QSAR models, the authors conducted a virtual screening of more than 50 000 different compounds in order to identify the most active inhibitors of the main M^{pro} protease of the virus. Based on the proposed regression equations, namely, the contributions of some significant descriptors, the authors of [48] analyzed the effect of the structural features of the studied compounds on a change in the inhibitory activity.

In the study [49], adequate QSAR models of SARS-CoV inhibitors, in the course of the structural interpretation of which molecular fragments that decrease and increase this type of activity were identified, were developed for 54 peptidomimetics. To construct the models, the authors of [49] also used the MLR method.

The regression method of the MLR data analysis applied in [48, 49] can give adequate results only in the presence of a linear relationship between the structure and activity [50]. One of the ways to overcome this disadvantage can be the use of nonparametric methods, in particular, various methods of machine learning (ML).

The study [27], in the course of which high-quality classification-based QSAR models of SARS-CoV inhibitors were developed, deserves special attention. The reliability of the constructed QSAR models was confirmed by subsequent experimental studies, as a result of which some compounds demonstrated high activity and were recommended for further study. In parallel with the QSAR analysis, the authors of [27] carried out a study using methods of molecular docking, while the revealed unacceptably low level of predictive ability did not allow considering the results of studying SARS-CoV inhibitors by molecular docking methods. Undoubtedly, the study [27] is very successful, but it lacks the structural interpretation of QSAR models.

The present study consisted of the following stages:

(1) construction of QSAR models of the main M^{pro} protease inhibitors of SARS-CoV;

(2) performing a virtual screening of the most promising compounds of potential drugs for the treatment of COVID-19;

(3) structural interpretation of QSAR models and rational molecular design of the main M^{pro} protease inhibitors.

EXPERIMENTAL

The well-known CHEMBL database (ID: CHEMBL3927) [51] served as the source for the sample formation for QSAR modeling. Inorganic compounds, polymers, mixtures, and compounds in the salt form were removed from the obtained sample. The final set of inhibitors of the main M^{pro} protease of SARS-CoV contained 65 compounds.

The experimental values of the activity of inhibitors of the SARS-CoV main M^{pro} protease, expressed in terms of the half-maximal inhibitory concentration (IC_{50}, nM or μM), which were given in the primary sources, were converted (1) into the negative common logarithm of the pIC_{50} value, which is generally accepted in QSAR studies and is used in cases when a linear increase in the concentration causes an exponential increase in the effect:

$$pIC_{50} = -\log(IC_{50}). \quad (1)$$

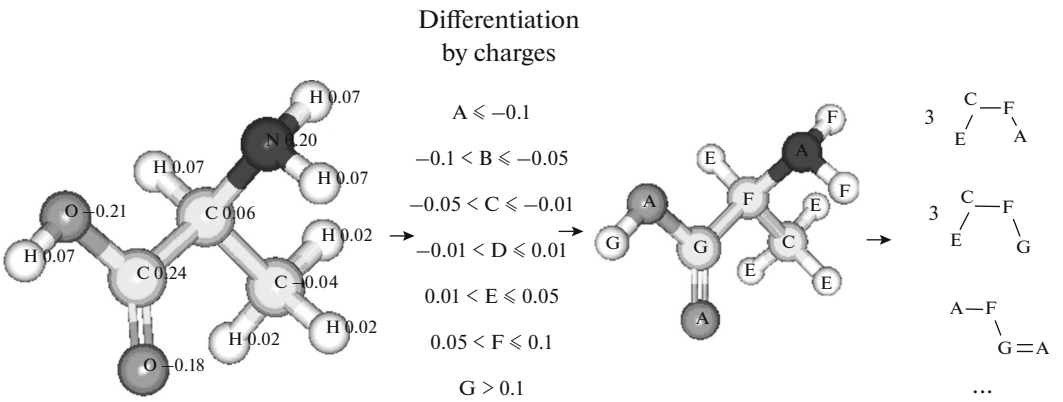
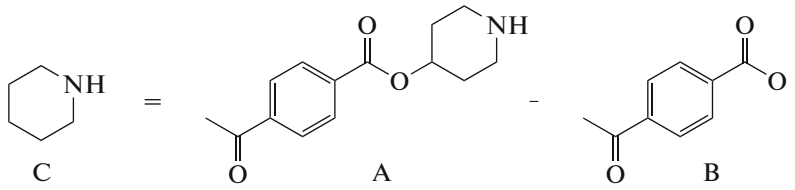


Fig. 1. An example of 2D generation of simplex descriptors for alanine at the 2D level using differentiation of atoms by their partial charges



$$W(C) = X(A) - X(B)$$

Fig. 2. The principle of structural interpretation used. W(C) is the contribution of the fragment (C); X(A) is the predicted activity value of the parent structure (A); X (B) is the predicted activity value for a hypothetical structure (B)

The exported experimental values and structures of compounds are given in the Appendix A (Table A1).

To describe the molecular structure, 2D simplex descriptors, which are calculated in the simplex representation of the molecular structure (SIRMS), were used [52]. Within the SIRMS, a molecule is considered as a system of various simplexes, tetraatomic molecular fragments of a fixed structure (Fig. 1).

The descriptor in this case is the number of simplices of a certain type. At the 2D level, atoms (vertices of a simplex) are differentiated not only by the nature "label" of the atom but also considering different physicochemical properties (partial charge on the atom, lipophilicity, refraction, and the ability to act as a donor or acceptor of hydrogen during the formation of a hydrogen bond).

Structural interpretation was performed in accordance with the approach [53], in which the contribution of the studied fragment (C) was calculated by the difference between the calculated values of the activity for the parent structure (A) and the hypothetical structure (B) obtained by removing the studied fragment (C) from the parent structure (A) (Fig. 2).

When simulating using simplex descriptors, we used the Scikit-learn package [54] for the Python programming language, which implements the methods of partial least squares (PLS), random forest (RF), support vector machine (SVM), and gradient boosting method (GBM).

Due to the small number of studied compounds and their structural diversity, a five-fold internal cross-validation (CV) was performed. For this, all compounds of the training sample are randomly divided into five parts. Then, a QSAR model is built (trained) on four pieces of data combined into a training sample, and the rest of the data is used as an external test sample; i.e., the predictive ability of the model is checked on the compounds of this group. This procedure is repeated 5 times; as a result, each of the five portions of the data is sequentially used for testing. Note that the studied compounds are never simultaneously used as a part of both the training and the external test set.

During QSAR modeling, the inclusion of the compounds in the applicability domain (AD) [55] was considerered for test samples, while if the value of at least one descriptor went beyond its minimum or maximum value for the training sample, then the compound of the test sample containing this descriptor was not included in the bounding box. This approach for QSAR modeling using simplex descriptors is implemented in the form of the SPCI software, which

Table 1. Statistical characteristics of QSAR models developed using the OCHEM internet resource

Method	Descriptor	R^2_{cv}	$RMSE_{cv}$
ASNN	ISIDA Fragments	0.67	0.50
	ALogPS, OEstate	0.68	0.49
	Dragon	0.66	0.50
	CDK	0.60	0.53
	alvaDesc	0.65	0.51
RF	StructuralAlerts	0.63	0.52
	ISIDA Fragments	0.67	0.49
Consensus model (https://ochem,eu/model/43078789)		0.70	0.47

is freely available at http://qsar4u.com/pages/sirms_qsar.php.

In addition, we used the OCHEM Internet platform ((https://ochem.eu) for QSAR analysis. The best modeling results were achieved using a number of descriptors (ALogPS, OEstate, Dragon, CDK, ISIDA Fragments, StructuralAlerts, alvaDesc) and the RF and associative neural networks (ASNN) methods. The consensus model was constructed by averaging the predictions of the best individual models. In this case, the applicability domain was assessed using the concept of the distance to the model (in particular, the CLASS-LAG approach). A brief description of the used methods and descriptors, as well as links to the original works are given in the OCHEM user manual [56].

The OCHEM internet resource implements the method of molecular pairs [57], which also allows us to interpret models constructed on any descriptors.

The assessment of the accuracy and predictive ability of the models proposed in this study and their comparison with other QSAR models was performed based on the following criteria.

1. The coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} (y_i - y_{i\text{mean}})^2}, \quad (2)$$

where $\hat{y}_i$ is the calculated value of the property for the ith molecule, y_i is the observed (experimental) value of the property for the ith molecule, m is the number of molecules in the sample, and $y_{i\ \text{mean}}$ is the mean value of the observed property.

2. Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{m}}. \quad (3)$$

Considering the fact that the QSAR modeling mentioned above requires the use of a large number (hundreds and thousands) of descriptors, for comparison, regression models containing a small number of HYBOT variables [59] with the addition of a number of topological and fractal characteristics [60] were constructed using the RF method [58]. In this case, AD was assessed using the interval method.

For the virtual screening, we used the DrugBank database [30], represented by FDA-approved drugs. Inorganic compounds, polymers, mixtures, and compounds in the salt form were removed from the exported DrugBank database. The final sample for the virtual screening contained 2087 FDA-approved drugs.

RESULTS AND DISCUSSION

The results of the QSAR modeling are shown in Tables 1–3. The consensus model is freely available at the link given in Table 1. All the constructed models have satisfactory statistical characteristics and possess comparable predictive power.

For the consensus QSAR models obtained (Tables 1, 2), a structural interpretation was performed. When interpreting the consensus model built using simplex descriptors (Table 2), the contributions of molecular fragments to the activity of inhibitors of the SARS-CoV main M^{pro} protease were determined (Fig. 3).

This set of molecular fragments is formed from the standard functional groups (51 fragments) and six molecular fragments obtained during the automatic fragmentation of compounds of the training set using the SPCI program using the SMART template [#6+0;!$(*=,#[!#6])]!@!=!#[*], which encodes breakable bonds [53]. Only those molecular fragments that were found in three or more compounds were subjected to interpretation, which, from our point of view, allowed us to focus on the fragments that stably affect the inhibitory activity and to avoid, to some extent, the influence of random factors, for example errors in the experimental data or predicted values of

Table 2. Statistical characteristics of QSAR models developed using 2D simplex descriptors

Method	Descriptor	R^2_{cv}	$RMSE_{cv}$
GBM	SIRMS	0.57	0.57
RF		0.65	0.51
SVM		0.52	0.60
PLS		0.64	0.52
Consensus model		0.64	0.51

Table 3. Statistical characteristics of QSAR model developed using HYBOT, topological, and fractal descriptors

Method	Descriptor	R^2_{cv}	$RMSE_{cv}$
RF	$MaxE_a$; $MaxC_a$; Nv2; IC0; D^*_{unb}	0.62	0.53

$MaxE_a$ is the maximum H-acceptor enthalpy descriptor; $MaxC_a$ is the maximum H-acceptor free energy descriptor; Nv2 is the number of vertices with the degree 2; IC0 is the mean informational content of the 0-th order; D^*_{unb} is the fractal density of unbound atoms. The used descriptors are considered in works [59, 60].

the activity and contributions of fragments. The complete list of identified molecular fragments in the form of SMARTS with the calculated average contributions to the activity is given (Appendix A, Table A2).

The interpretation allowed us to quantitatively describe and rank the effect of molecular fragments on the change in the activity of SARS-CoV M^{pro} inhibitors and detail the molecular environment of the known functional groups, highlighting derivative fragments that increase and decrease contributions to the indicated type of the activity. For example, when detailing pyrimidine, the 2-sulfanylpyrimidin-4-ol molecular fragment (**f9** in Fig. 3), which significantly reduces the activity of SARS-CoV M^{pro} inhibitors, was isolated. In this case, carboxyl derivatives of furan and pyridine (fragments **f1** and **f2** in Fig. 3), on the contrary, increase the activity of SARS-CoV M^{pro} inhibitors.

The interpretation was also performed for the consensus model (Table 1), built using the OCHEM internet resource. Table 4 shows the results of the interpretation, according to which the inhibitory activity increases under the substitution of hydrogen atoms with chlorine or methyl group. An increase in the activity of SARS-CoV M^{pro} inhibitors is also observed when phenyl and *n*-propyl radicals are replaced by naphthyl radicals. The results of the interpretations for the consensus models described above consistently indicate an increase in the activity of SARS-CoV M^{pro} inhibitors under the substitution of fragments containing iodine (**f5**) by the carbamoyl group (**f4**).

Considering the trends in the effects of the structure of compounds on the change in the activity revealed during the interpretation, we carried out a rational molecular design and proposed a number of promising agents against COVID-19. In this case, molecular fragments that reduce the activity were replaced by the fragments which increase the activity of M^{pro} inhibitors according to the interpretation results. As a result, hypothetical compounds (Table 5, substances **2**, **4**, **6**), which possess a significant calculated inhibitory activity and fall into the applicability domain of the consensus QSAR model, developed using OCHEM, were proposed. For example, when the residue of 6-methyl-2-sulfanylpyridine-4-ol (compound 1, Table 5) is substituted with carboxyl derivative of pyridine (compound 2, Table 5), a significant increase in the activity of the SARS-CoV main M^{pro} protease inhibitors is noted. Also an increase in the activity is characteristic for the substitution of a fragment, containing nitrile (compound 3), with trifluoromethyl (compound 4) or of the 4-(1,3-thiazol-4-yl)pyrimidine-2-thiol residue (compound 5, Table 5) with the above-mentioned carboxyl pyridine derivative (compound 6). It should be noted that during molecular design in these examples the results of the interpretation of the QSAR model built using simplex descriptors were considered, while the prediction of the inhibitor activity was performed using the QSAR model built by the OCHEM internet resource.

When determining the strategies of synthesis and testing, it is important to evaluate various types of toxicity and lipophilicity in addition to the target property (activity), which are important factors when deciding whether to recommend the use of a compound as an active substance of the drug. For this purpose, the acute toxicity (LD_{50}) after oral administration to rats and the probability of mutagenicity (the Ames test) was assessed for compounds **1**–**9** using the T.E.S.T. v.4.2. program, developed by experts from the Environmental Protection Agency of the United States [61]. Also, using the swissADME Internet platform of the Swiss Bioinformatics Institute (http://www.swissadme.ch/) [62], the lipophilicity (log Po/w), compliance with Lipinski's rules [63], the presence of PAINS fragments [64], the synthetic availability on a ten-point scale (0 is the maximum degree of synthetic availability, 10 is the minimum degree of synthetic availability) [65] were assessed for these compounds, which is extremely important for the proposed, but not yet synthesized, compounds. The prediction results are shown in Table 6, from which it can be seen that compounds **2**, **4**, **6**, **8**, and **9** proposed in the course of molecular design have comparable synthetic availability in comparison with the synthesized substances **1**, **3**, **5**, and **7**. All substances satisfy Lipinski's rules of five, except for compound **4**, and do not contain PAINS

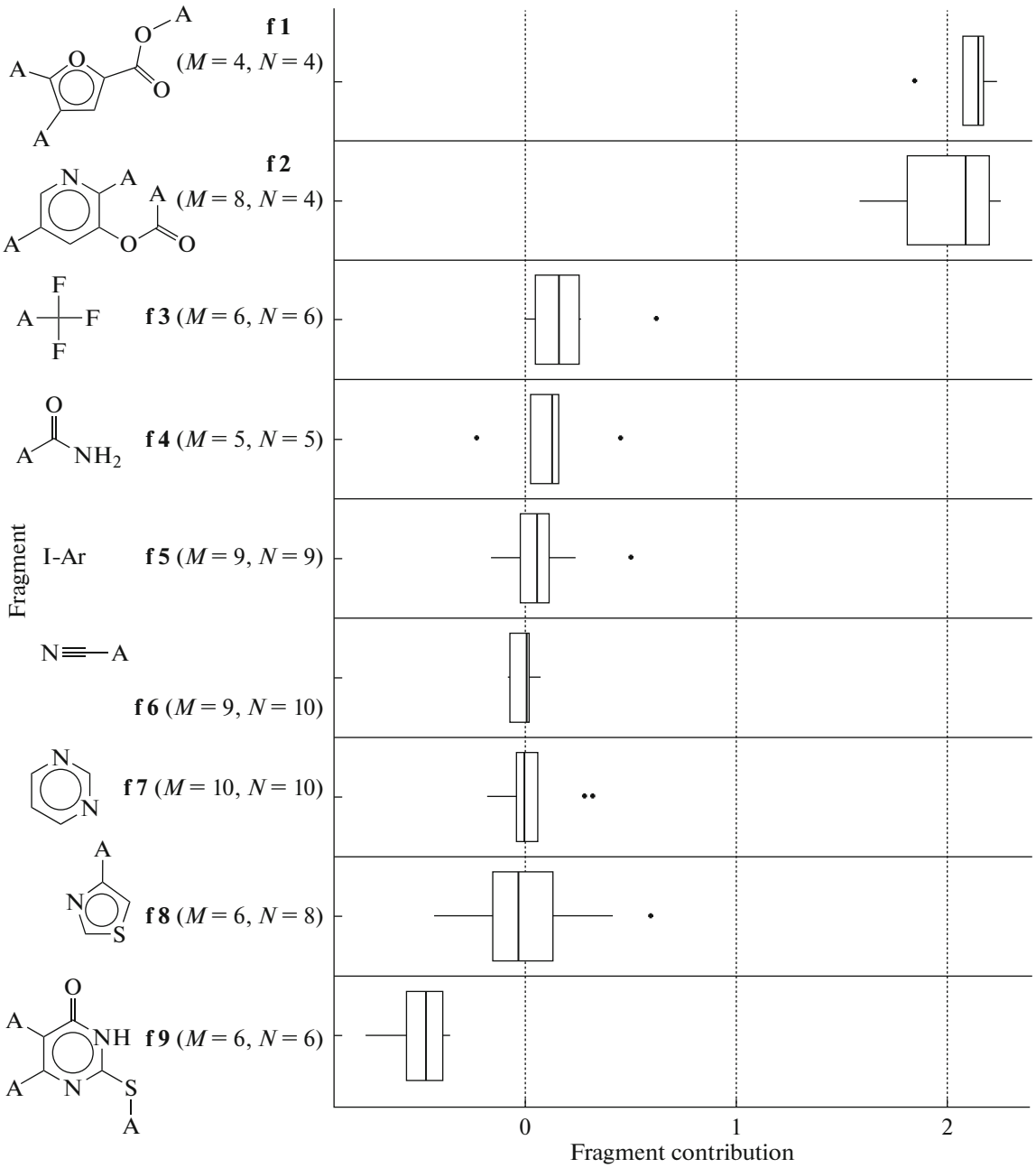


Fig. 3. Contributions of molecular fragments to the ability of compounds to inhibit the SARS-CoV main M^{pro} protease. A, the place of the fragment's attachment to the other part of the molecule; f, the order number of the fragment; *M*, the number of compounds containing the given fragment; *N*, the number of detections of the corresponding fragment in the sample.

fragments. The studied compounds are characterized by a wide range of lipophilicity values, which should be considered when studying pharmacokinetics and choosing dosage forms. According to the calculations performed, compound **7** has a nonzero probability of mutagenicity, which can reduce its attractiveness as a lead compound, even though it has the maximum experimentally measured inhibitory activity (Table 5) among the compounds of the exported sample from the CHEMBL database (ID: CHEMBL3927). When modifying compound **7**, a hypothetical compound **9**, which does not have the probability of mutagenicity according to the calculated data, while the indicators of LD_{50} and the inhibitory activity are comparable with the initial compound **7**, was proposed. In addition, the modification of compound **1** into compound **2** allowed us not only to increase the inhibitory activity by almost two orders, reaching comparable values with the most active substances in the exported sample but also to reduce the toxicity (LD_{50}) by a factor of almost 2.7.

Table 4. The interpretation results for the consensus model developed using the OCHEM internet resource

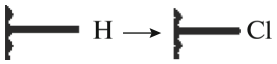
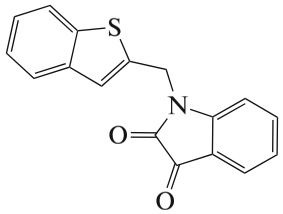
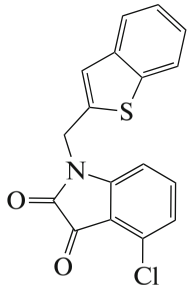
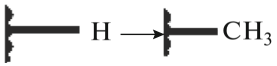
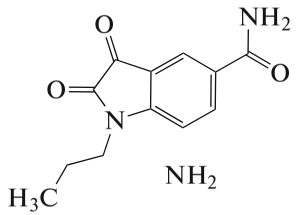
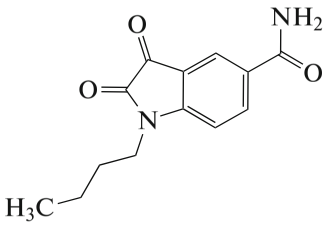
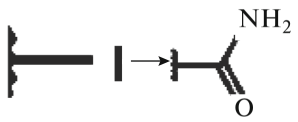
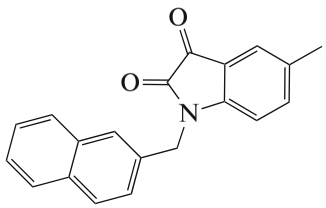
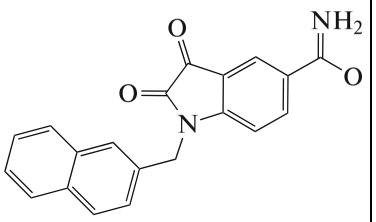
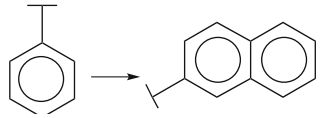
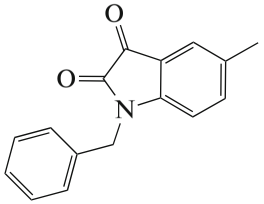
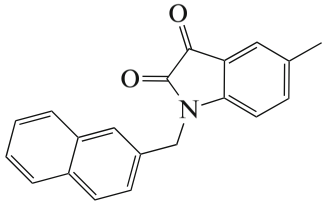
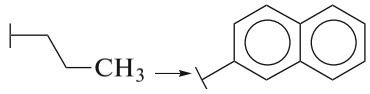
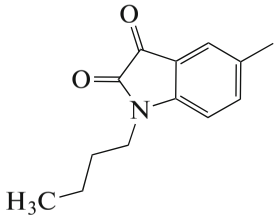
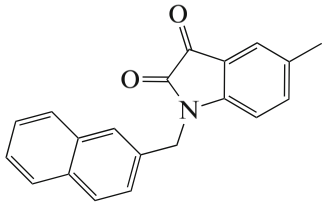
Molecular transformation	Pairs of compounds and their experimental characteristics	*N*	Δmean	SMIRKS
	 –log (IC_{50}) = 4.88  –log (IC_{50}) = 4.95	2	0.035 ± 0.049	*[H] → *Cl
	 –log (IC_{50}) = 4.60  –log (IC_{50}) = 4.72	2	0.06 ± 0.085	*[H] → C*
	 –log(IC_{50}) = 5.96  –log(IC_{50}) = 6.43	3	0.54 ± 0.065	*I → NC(*) = 0

Table 4. (Contd.)

Molecular transformation	Pairs of compounds and their experimental characteristics	*N*	Δmean	SMIRKS
	 $-\log(IC_{50}) = 4.30$  $-\log(IC_{50}) = 5.96$	2	1.6 ± 0.092	*clcccccl → *clccccccccc2cl
	 $-\log(IC_{50}) = 4.18$  $-\log(IC_{50}) = 5.96$	2	1.7 ± 0.049	CCC* → *clcc2cccc2cl

N is the number of molecular pairs, which meet the molecular transformation; Δmean is the mean difference of the values $-\log(IC_{50})$ when performing the molecular transformation; SMIRKS is the recording format for the molecular transformation (https://www.daylight.com/dayhtml/doc/theory/theory.smirks.html).

Table 5. The results of molecular design

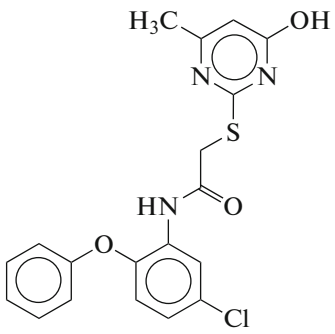
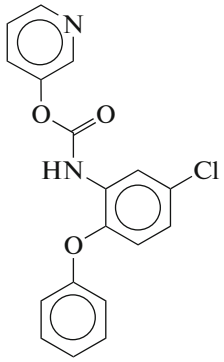
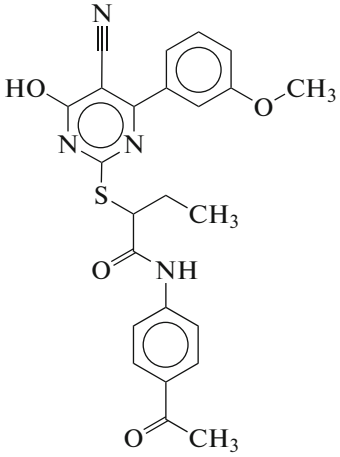
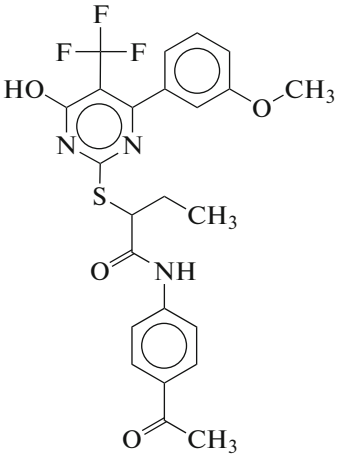
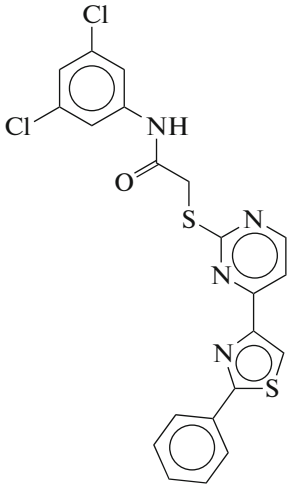
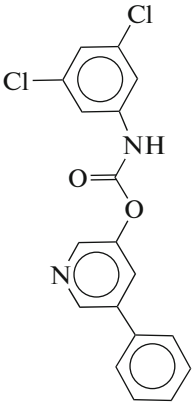
Sample compounds	Experimental values, $-\log(IC_{50})$	Hypothetical compounds	Predicted values $-\log(IC_{50})$
 **1**	4.00	 **2**	5.93
 **3**	4.22	 **4**	4.34
 **5**	5.52	 **6**	6.19

Table 5. (Contd.)

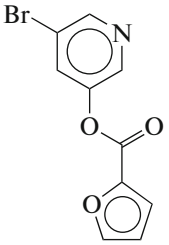
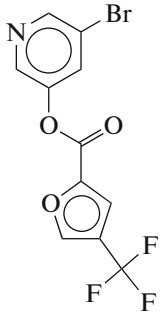
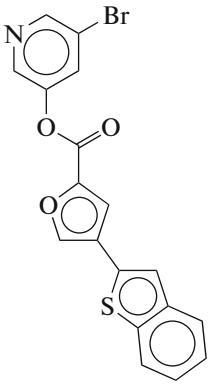
Sample compounds	Experimental values, $-\log(IC_{50})$	Hypothetical compounds	Predicted values $-\log(IC_{50})$
 **7**	7.3	 **8**	6.90
		 **9**	6.94

Table 6. The assessment of some physicochemical properties, types of toxicity and synthetic availability of compounds studied within molecular design (the structures of compounds are presented in Table 5)

Compound	log *P*o/w	Number of inconsistencies with Lipinski's rules	Number of PAINS fragments	Synthetic availability	Rat LD50 after oral administration, mg/kg	Probability of mutagenicity**
1	3.75	0	0	2.86	1382.28	0
2*	3.71	0	0	2.94	3788.68	0
3	3.35	0	0	3.76	956.51	0
4*	4.45	1 (Molecular weight over 500)	0	3.89	–	0
5	4.98	0	0	3.33	663.22	0
6*	4.33	0	0	2.78	772.67	0
7	2.18	0	0	2.5	581.59	1
8*	3.28	0	0	2.81	193.04	0
9*	4.58	0	0	3.26	561.22	0

* Hypothetical compounds proposed as a result of molecular design; ** 0, negative; 1, positive; – compound is out of the applicability domain of the QSAR model.

Thus, compound **2** can be recommended for synthesis and further testing.

Since the synthesis of new compounds and their clinical trials take a long time, the most important means of combating a new, rapidly spreading pandemic is the repositioning of approved drugs that have passed all the necessary clinical studies. In order to identify promising inhibitors of the SARS-CoV-2 main M^{pro} protease, 2087 FDA-approved drugs were screened. The consensus model built by the OCHEM

Table 7. The most promising FDA-approved drugs for inhibiting SARS-CoV-2 replication according to the results of virtual screening

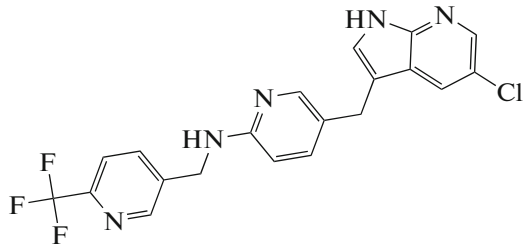
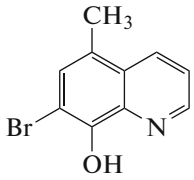
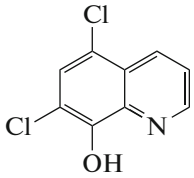
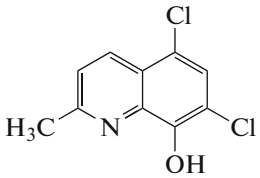
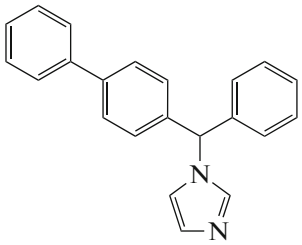
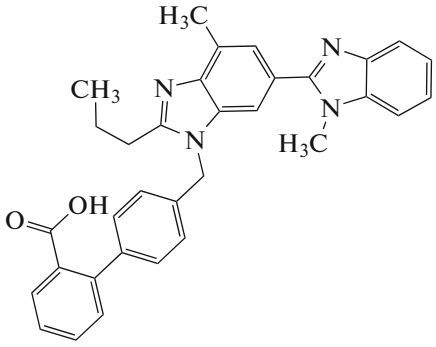
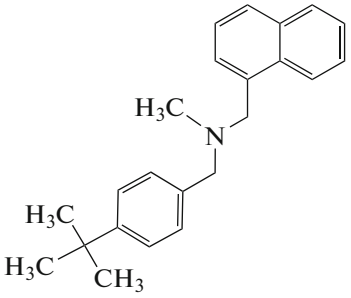
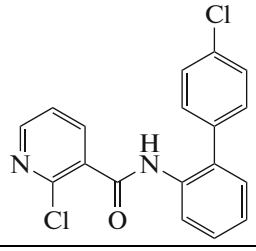
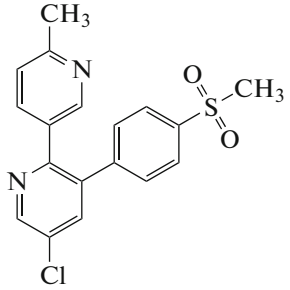
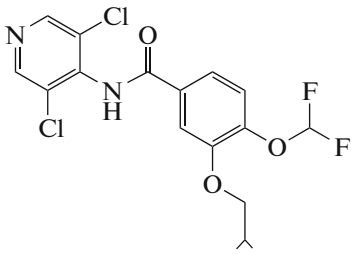
$-\log(IC_{50})$	Name	Chemical structure	Description
6.09	Pexidartinib		Antitumor agent, tyrosine kinase inhibitor
5.66	Tilbroquinol		Antiprotozoal agent effective against amebiasis; the drug was also used against *Vibrio cholerae*
5.61	Chloroxine		Drugs with bacteriostatic, fungistatic, and antiprotozoal properties
5.6	Chlorquinaldol		
5.6	Bifonazole		Antifungal drug
5.59	Telmisartan		Antihypertensive agent, angiotensin II receptor antagonist

Table 7. (Contd.)

$-\log(IC_{50})$	Name	Chemical structure	Description
5.57	Butenafine		Synthetic antifungal benzylamine
5.54	Boscalid		Has fungicidal properties
5.53	Etoricoxib		Anti-inflammatory, analgesic agent; selective cyclooxygenase-2 inhibitor
5.51	Roflumilast		Anti-inflammatory agent representing phosphodiesterase-4 (PDE4) inhibitor

expert system was used for screening, since it has better statistical characteristics and can be used by all interested persons for the virtual screening of their own sets of compounds. The QSAR model obtained using simplex descriptors was not used due to the peculiarities of the method described above for determining the applicability domain, which severely limits the structural space of the model.

In the course of the virtual screening, ten compounds that are within the applicability domain of the consensus QSAR model developed using OCHEM and have the highest calculated inhibitory activity were proposed (Table 7). Antitumor, antiprotozoal, antifungal, antibacterial, antihypertensive, and anti-inflammatory drugs are among these compounds. Tilbroquinol, Chloroxine, and Chlorquinaldol, which are halogenated quinoline derivatives as are the well-known chloroquine and hydroxychloroquine used in the treatment of COVID-19, are of particular interest [66].

According to the data of the virtual screening, the highest inhibitory activity among the FDA-approved drugs is possessed by Pexidartinib, which is an anticancer agent, a tyrosine kinase inhibitor. In the study [12], conducted using molecular docking and molecular dynamics, the antitumor agent Neratinib, which blocks the functioning of receptor tyrosine kinases, was also proposed as a promising inhibitor of the SARS-CoV-2 main M^{pro} protease. The conclusions of the authors [12] are based on the assumption of a similar binding of this antitumor agent to the cysteine residue in the active centers of the kinase domains of receptor tyrosine kinases and the SARS-CoV-2 main

M^{pro} protease. Another antitumor agent, carmofur, was also isolated as a promising inhibitor of the SARS-CoV-2 main M^{pro} protease according to the results of high-throughput screening in the above-mentioned study [23]. Recent additional studies using X-ray diffraction analysis [67] describe the mechanism of the inhibition of the SARS-CoV-2 main M^{pro} protease by carmofur through covalent binding to the cysteine residue Cys145 in the active center. Based on the foregoing, the proposal to repurpose pexidartinib, identified during the virtual screening, for the treatment of COVID-19 seems logical. It should be noted that the confirmation of the effectiveness of the drugs proposed for repurposing in the fight against COVID-19 requires significant additional experimental research. Drugs should be taken only according to the medical prescription by the physician.

Thus, in the course of computational experiments using conceptually different descriptors and machine learning methods, acceptable QSAR models of the main M^{pro} protease inhibitors were developed.

The structural interpretation of the QSAR models allowed us to reveal the common regularities in the effect of the structure of chemical compounds on their inhibitory activity by isolating molecular fragments and transformations that increase and decrease the activity of SARS-CoV inhibitors. The results of the structural interpretation were used to perform rational molecular design, in the course of which a number of promising compounds for combating COVID-19 were proposed.

The virtual screening of FDA-approved drugs identified ten substances that can be recommended for repurposing as drugs against the new coronavirus infection.

The results of this study can help to reduce financial, time, and labor costs when determining the strategy for the development of new drugs and repositioning existing drugs that are SARS-CoV-2 inhibitors.

FUNDING

Part of the study was carried out within a state task for 2020 (topic no. 0090-2020-0004) of the Institute of Physiologically Active Compounds, Russian Academy of Sciences.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

SUPPLEMENTARY MATERIALS

Additional materials are available.

APPENDIX A

Table A1. Structures and activities of studied compounds

Number	SMILES	Molecule ChEMBL ID	IC_{50}, nM	pChEMBL Value	Document ChEMBL ID
1	Brc1cncc(OC(=O)c2ccco2)c1	CHEMBL427404	50	7.3	CHEMBL1144475
2	Clc1cncc(OC(=O)c2ccco2)c1	CHEMBL426898	60	7.22	CHEMBL1144475
3	Clc1ccc(cc1)-c1ccc(o1)C(=O)Oc1cncc(Cl)c1	CHEMBL426082	63	7.2	CHEMBL1144475
4	Clc1cncc(OC(=O)c2cc3ccccc3[nH]2)c1	CHEMBL384739	65	7.19	CHEMBL1144475
5	Clc1cncc(OC(=O)c2cc3ccccc3s2)c1	CHEMBL383725	95	7.02	CHEMBL1144475
6	Clc1cncc(OC(=O)c2cc3ccccc3o2)c1	CHEMBL380470	170	6.77	CHEMBL1144475
7	Clc1cncc(OC(=O)c2cscn2)c1	CHEMBL380403	270	6.57	CHEMBL1144475
8	Cc1cc(c(Cl)cc1Cl)S(=O)(=O)c1c(cc(cc1[N+]([O–])=O)C(F)(F)F)[N+]([O–])=O	CHEMBL379727	300	6.52	CHEMBL1145342
9	COc1cccc(c1)C(=O)Oc1cncc(Cl)c1	CHEMBL379642	340	6.47	CHEMBL1144475
10	NC(=O)c1ccc2N(Cc3ccc4ccccc4c3)C(=O)C(=O)c2c1	CHEMBL378700	370	6.43	CHEMBL1148529
11	ClC(Cl)=C(Cl)C(=O)Oc1ccc(cc1)S(=O)(=O)c1ccc(OC(=O)C(Cl)=C(Cl)Cl)cc1	CHEMBL378674	900	6.05	CHEMBL1145342
12	Ic1ccc2N(Cc3cc4ccccc4s3)C(=O)C(=O)c2c1	CHEMBL378342	950	6.02	CHEMBL1139624
13	Brc1cccc2C(=O)C(=O)N(Cc3cc4ccccc4s3)c12	CHEMBL377253	980	6.01	CHEMBL1139624

Table A1. (Contd.)

Number	SMILES	Molecule ChEMBL ID	IC_{50}, nM	pChEMBL Value	Document ChEMBL ID
14	Ic1ccc2N(Cc3ccc4ccccc4c3)C(=O)C(=O)c2c1	CHEMBL377150	1100	5.96	CHEMBL1148529
15	[O–][N+](=O)c1cccc2C(=O)C(=O)N(Cc3cc4ccccc4s3)c12	CHEMBL375130	2000	5.7	CHEMBL1139624
16	FC(F)(F)c1nnc(SC(=O)c2ccc(o2)C#Cc2ccccc2)[nH]1	CHEMBL370923	3000	5.52	CHEMBL1145342
17	Clc1cc(Cl)cc(NC(=O)CSc2nccc(n2)-c2csc(n2)-c2ccccc2)c1	CHEMBL365469	3000	5.52	CHEMBL1148632
18	Fc1ccc2N(Cc3cc4ccccc4s3)C(=O)C(=O)c2c1	CHEMBL365134	4820	5.32	CHEMBL1139624
19	Cc1noc(NC(=O)c2ccc(s2)-c2cc(nn2C)C(F)(F)F)c1[N+]([O–])=O	CHEMBL358279	5000	5.3	CHEMBL1145342
20	Nc1ncc(c(N)n1)S(=O)(=O)c1ccc(Cl)cc1	CHEMBL348660	6000	5.22	CHEMBL1145342
21	Cc1noc(C)c1CN1C(=O)C(=O)c2cc(ccc12)C#N	CHEMBL225515	7200	5.14	CHEMBL1139624
22	Fc1ccc(CN2C(=O)C(=O)c3cc(I)ccc23)c(Cl)c1	CHEMBL222893	9400	5.03	CHEMBL1139624
23	Cn1nc(cc1C(F)(F)F)-c1ccc(s1)-c1ccnc(SCC(=O)Nc2ccc(Cl)cc2)n1	CHEMBL222840	10000	5	CHEMBL1148632
24	Cc1oc(cc1-c1cc(N S(=O)(=O)c2cccs2)[nH]n1)C(C)(C)C	CHEMBL222769	10000	5	CHEMBL1145342
25	CSc1sc(c(C)c1-c1ccnc(SCC(=O)Nc2ccc(Cl)cc2)n1)-c1nc(C)cs1	CHEMBL222735	11000	4.96	CHEMBL1148632
26	Clc1cccc2N (Cc3cc4ccccc4s3)C(=O)C(=O)c12	CHEMBL222628	11200	4.95	CHEMBL1139624
27	[O–][N+](=O)c1cc(ccc1S(=O)(=O)c1ccc(Cl)cc1)C(F)(F)F	CHEMBL222234	12000	4.92	CHEMBL1145342
28	CSc1sc(c(C)c1-c1ccnc(SCC(=O)Nc2ccccc2Cl)n1)-c1nc(C)cs1	CHEMBL215732	12000	4.92	CHEMBL1148632
29	NC(=O)c1ccc2N(Cc3ccccc3)C(=O)C(=O)c2c1	CHEMBL215397	12500	4.9	CHEMBL1148529
30	Clc1ccc(NC(=O)c2ccc(CN3C(=O)C(=O)c4cc(I)ccc34)s2)cc1	CHEMBL214372	12570	4.9	CHEMBL1139624
31	Cc1nc(c(C#N)c(C)c1[N+]([O–])=O)S(=O)(=O)c1ccccc1	CHEMBL213581	13000	4.89	CHEMBL1145342
32	Cc1ccc(cc1)S(=O)(=O)c1nc(C)c(c(C)c1C#N)[N+]([O–])=O	CHEMBL212504	13000	4.89	CHEMBL1145342
33	O=C1N(Cc2cc3ccccc3s2)c2ccccc2C1=O	CHEMBL212454	13110	4.88	CHEMBL1139624
34	Ic1ccc2N(CC3COc4ccccc4O3)C(=O)C(=O)c2c1	CHEMBL212399	13500	4.87	CHEMBL1139624
35	Cc1nc(cs1)-c1nc(cs1)-c1ccnc(SCC(=O)Nc2ccc(Cl)cc2)n1	CHEMBL212240	14000	4.85	CHEMBL1148632
36	[O–][N+](=O)c1ccc([n+]([O-])c1)S(=O)(=O)c1ccc(Cl)cc1	CHEMBL212218	15000	4.82	CHEMBL1145342
37	CSc1[nH]nc(NC(=O)c2cccs2)c1S(=O)(=O)c1ccccc1	CHEMBL212190	15000	4.82	CHEMBL1145342
38	FC(F)(F)c1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccc(Cl)cc2Cl)cc1	CHEMBL212019	15000	4.82	CHEMBL1148632
39	Clc1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccccc2Cl)cc1	CHEMBL211969	15000	4.82	CHEMBL1148632
40	Clc1ccc(NC(=O)CSc2nccc(n2)-c2cc(no2)-c2ccccc2)cc1	CHEMBL210632	15000	4.82	CHEMBL1148632
41	CC(=O)c1ccccc1S(=O)(=O)c1ccccc1C(O)=O	CHEMBL210612	16000	4.8	CHEMBL1145342
42	CCO\C(O)=C(\C=N\c1ccc(cc1)S(=O)(=O)c1ccc(N\C=C(/C#N)C(=O)OCC)cc1)/C#N	CHEMBL210525	16000	4.8	CHEMBL1145342
43	OC(=O)c1ccc(cc1)S(=O)(=O)c1cc(Br)c(O)c(Br)c1	CHEMBL210497	16000	4.8	CHEMBL1145342

Table A1. (Contd.)

Number	SMILES	Molecule ChEMBL ID	IC_{50}, nM	pChEMBL Value	Document ChEMBL ID
44	CC1(C)Cc2c(sc(NCc3ccco3)c2C(=O)C1)C#N	CHEMBL210487	16000	4.8	CHEMBL1145342
45	Ic1ccc2N(Cc3ccc(s3)C(=O)N3CCCCC3)C(=O)C(=O)c2c1	CHEMBL210097	17500	4.76	CHEMBL1139624
46	CSc1nn(c(-c2cccs2)c1C#N)-c1c(c(C)nn1C)[N+]([O–])=O	CHEMBL210092	18000	4.75	CHEMBL1145342
47	CCCCN 1C(=O)C(=O)c2cc(ccc12)C(N)=O	CHEMBL209667	19000	4.72	CHEMBL1148529
48	Cc1nn(C)c(NCc2ccc(s2)-c2cccs2)c1[N+]([O–])=O	CHEMBL209287	20000	4.7	CHEMBL1145342
49	Ic1ccc2N(C\C=C\c3cc4ccccc4s3)C(=O)C(=O)c2c1	CHEMBL209227	23500	4.63	CHEMBL1139624
50	[O–][N+](=O)c1ccc(cc1)S(=O)(=O)c1ccc(cc1)[N+]([O–])=O	CHEMBL208763	25000	4.6	CHEMBL1145342
51	CCCN 1C(=O)C(=O)c2cc(ccc12)C(N)=O	CHEMBL208732	25000	4.6	CHEMBL1148529
52	CCCc1cc(O)nc(SCC(=O)Nc2ccc(Cl)cc2)n1	CHEMBL208584	30000	4.52	CHEMBL1148632
53	CCO\C(O)=C(\C=N\c1ccc(cc1)S(=O)(=O)c1ccc(NC=C(C(=O)OCC)C(=O)OCC)cc1)/C(=O)OCC	CHEMBL207207	32000	4.5	CHEMBL1145342
54	O=C(Cc1nccs1)c1nccs1	CHEMBL196635	40000	4.4	CHEMBL1145342
55	CC(C)c1ccc(NC(=O)CSc2nccc(n2)-c2cccs2)cc1	CHEMBL194398	40000	4.4	CHEMBL1148632
56	COc1cccc(c1)-c1nc(SCC(=O)Nc2ccc(cc2)S(N)(=O)=O)nc(O)c1C#N	CHEMBL191575	40000	4.4	CHEMBL1148632
57	COc1ccc(NC(=O)CSc2nc(O)cc(n2)-c2ccccc2)cc1OC	CHEMBL190743	45000	4.35	CHEMBL1148632
58	CCOC(=O)\C=C\[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@@H](CC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C)Cc1ccccc1	CHEMBL188983	45000	4.35	CHEMBL1141032
59	Ic1ccc2N(Cc3ccccc3)C(=O)C(=O)c2c1	CHEMBL188487	50000	4.3	CHEMBL1148529
60	COc1cccc(c1)-c1nc(SCC(=O)Nc2ccc(cc2)C(C)=O)nc(O)c1C#N	CHEMBL187717	60000	4.22	CHEMBL1148632
61	CCC(Sc1nc(O)c(C#N)c(n1)-c1cccc(OC)c1)C(=O)Nc1ccc(cc1)C(C)=O	CHEMBL187598	60000	4.22	CHEMBL1148632
62	CCCCN 1C(=O)C(=O)c2cc(I)ccc12	CHEMBL187579	66000	4.18	CHEMBL1148529
63	CCOC(=O)\C=C\[C@H](C[C@@H]1CCNC1=O)NC(=O)[C@H](CC=C(C)C)CC(=O)[C@@H](NC(=O)c1cc(C)on1)C(C)C	CHEMBL185698	70000	4.16	CHEMBL1141032
64	CN1C(=O)C(=O)c2cc(ccc12)C(N)=O	CHEMBL148483	71000	4.15	CHEMBL1148529
65	Cc1cc(O)nc(SCC(=O)Nc2cc(Cl)ccc2Oc2ccccc2)n1	CHEMBL118596	100000	4	CHEMBL1148632

Table A2. Complete list of the identified molecular fragments, written in the form of SMARTS

SMARTS	*M*	*N*	Average contribution of fragment
O=C(O[*])c1cc([*])c([*])o1	3	3	2.08424025
O=C(Oc1cc([*])cnc1[*])[*]	8	8	2.0206205
Clc1c([*])ncc([*])c1[*]	7	7	1.8665205
O=C(O[*])[*]	9	9	1.59494225
O=C1C(=O)N(C([*])[*])c2c1cccc2[*]	3	3	0.6641145
c1c([*])cc2cc([*])sc2c1[*]	8	8	0.652451375
Clc1cc(Cl)c([*])c([*])c1[*]	3	3	0.30955275
FC(F)(F)[*]	6	6	0.146696125
NC(=O)[*]	5	5	0.13180325
O=S(=O)(c1cc([*])c(Cl)cc1[*])[*]	4	4	0.1251
c1cc(C[*])c([*])cc1[*]	4	4	0.12120975
c1c([*])sc(C[*])c1[*]	3	3	0.09563475
Cl[*]	25	30	0.04179175
Cn1nc([*])c([*])c1[*]	4	4	0.030965625
F[*]	8	8	0.027254125
N#C[*]	7	7	0.0052065
O=C1C(=O)N([*])c2c([*])cc(I)c([*])c21	9	9	0.00389875
c1cc(-c2cc([*])on2)c([*])cc1[*]	3	3	0.002501
SMARTS	*M*	*N*	Average contribution of fragment
c1nc([*])nc([*])c1[*]	10	10	0.001548625
O=[N+]([O–])[*]	10	10	0
I[*]	9	9	–0.000999
n1c([*])sc([*])c1[*]	6	8	–0.025514375
O=C(Nc1ccc([*])cc1[*])[*]	12	12	–0.0292055
O=C(N[*])[*]	17	17	–0.1267215
O=C(CS[*])N[*]	14	14	–0.17362825
NC(=O)c1c([*])cc2c(c1[*])C(=O)C(=O)N2[*]	4	4	–0.178003875
O=c1[nH]c(S[*])nc([*])c1[*]	5	5	–0.3823725

Designations: *M* is the number of compounds containing the given fragment; *N* is the number of detections of the corresponding fragment in the sample.

REFERENCES

1. Enserink, M., *Science*, 2013, vol. 339, no. 6125, p. 1266. https://doi.org/10.1126/science.339.6125.1266
2. WHO. Middle East respiratory syndrome coronavirus (MERS-CoV)—Republic of Korea. Global Alert and Response (GAR). www.who.int/csr/don/01-june-2015-mers-korea/en/. Accessed September 28, 2020.
3. Menachery, V.D., Yount, B.L., Jr., Debbink, K., Agnihothram, S., Gralinski, L.E., Plante, J.A., Graham, R.L., Scobey, T., Ge, X.Y., Donaldson, E.F., Randell, S.H., Lanzavecchia, A., Marasco, W.A., Shi, Z.L., and Baric, R.S., *Nat. Med.*, 2015, vol. 21, no. 12, p. 1508. https://doi.org/10.1038/nm.3985
4. Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Xia, J., Yu, T.,

Zhang, X., and Zhang, L., *Lancet*, 2020, vol. 395, no. 10223, p. 507. https://doi.org/10.1016/S0140-6736(20)30211-7

5. Comparing the Wuhan coronavirus outbreak with SARS and MERS. https://graphics.reuters.com/CHINA-HEALTH-VIRUSCOMPARISON/0100B5BY3CY/index.html. Accessed September 28, 2020.
6. The Russian Ministry of Health registered the first preparation for coronavirus, June 1, 2020. www.rosminzdrav.ru/news/2020/06/01/14086-minzdrav-rossii-zaregistriroval-pervyy-preparat-ot-koronavirusa. Accessed September 28, 2020.
7. Shiraki, K. and Daikoku, T., *Pharm. Ther.*, 2020, vol. 209, 107512. https://doi.org/10.1016/j.pharmthera.2020.107512
8. Xu, J. and Hagler, A., *Molecules*, 2002, vol. 7, no. 8, p. 566. https://doi.org/10.3390/70800566
9. Lo, Y.C., Rensi, S.E., Torng, W., and Altman, R.B., *Drug Discovery Today*, 2018, vol. 23, no. 8, p. 1538. https://doi.org/10.1016/j.drudis.2018.05.010
10. Yu, W. and Mackerell, A.D., Jr., *Methods Mol. Biol.*, 2017, vol. 1520, p. 85. https://doi.org/10.1007/978-1-4939-6634-9_5
11. Pant, S., Singh, M., Ravichandiran, V., Murty, U., and Srivastava, H.K., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/07391102.2020.1757510
12. Skvortsov, V.S., Druzhilovskiy, D.S., and Veselovsky, A.V., *Biomed. Chem.: Res. Methods*, 2020, vol. 3, no. 1, e00124. https://doi.org/10.18097/BMCRM00124
13. Wang, J., *J. Chem. Inf. Model.*, 2020, vol. 60, no. 6, p. 3277. https://doi.org/10.1021/acs.jcim.0c00179
14. Mittal, L., Kumari, A., Srivastava, M., Singh, M., and Asthana, S., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/07391102.2020.1768151
15. Gyebi, G.A., Ogunro, O.B., Adegunloye, A.P., Ogunyemi, O.M., and Afolabi, S.O., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/0739110.2020.1764868
16. Enmozhi, S.K., Raja, K., Sebastine, I., and Joseph, J., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/07391102.2020.1760136
17. Elmezayen, A.D., Al-Obaidi, A., Sahin, A.T., and Yelekci, K., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/07391102.2020.1758791
18. Mahanta, S., Chowdhury, P., Gogoi, N., Goswami, N., Borah, D., Kumar, R., Chetia, D., Borah, P., Buragohain, A.K., and Gogoi, B., *J. Biomol. Struct. Dyn.*, 2020. https://doi.org/10.1080/07391102.2020.1768902
19. Kandeel, M. and Al-Nazawi, M., *Life Sci.*, 2020, vol. 251, 117627. https://doi.org/10.1016/j.lfs.2020.117627
20. Gentile, D., Patamia, V., Scala, A., Sciortino, M.T., Piperno, A., and Rescifina, A., *Mar. Drugs*, 2020, vol. 18, no. 4, p. 225. https://doi.org/10.3390/md18040225
21. Sepay, N., Sepay, N., Al Hoque, A., Mondal, R., Halder, U.C., and Muddassir, M., *Struct. Chem.*, 2020, vol. 31, p. 1831. https://doi.org/10.1007/s11224-020-01537-5
22. Shamsi, A., Mohammad, T., Anwar, S., AlAjmi, M.F., Hussain, A., Rehman, M.T., Islam, A., and Hassan, M.I., *Biosci. Rep.*, 2020, vol. 40, no. 6, BSR20201256. https://doi.org/10.1042/BSR20201256
23. 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., and Yang, H., *Nature*, 2020, vol. 582, no. 7811, p. 289. https://doi.org/10.1038/s41586-020-2223-y
24. Chen, Y.W., Yiu, C.B., and Wong, K.Y., *F1000Research*, 2020, vol. 9, p.129. https://doi.org/10.12688/f1000research.22457.2
25. Liu, X. and Wang, X.J., *J Genet Genomics*, 2020, vol. 47, no. 2, p. 119. https://doi.org/10.1016/j.jgg.2020.02.001
26. Ul, QamarM.T., Alqahtani, S.M., Alamri, M.A., and Chen, L.L., *J. Pharm. Anal.*, 2020, vol. 10, no. 4, p. 313. https://doi.org/10.1016/j.jpha.2020.03.009
27. Alves, V.M., Bobrowski, T., Melo-Filho, C.C., Korn, D.L., Auerbach, S., Schmitt, C., Muratov, E.N., and Tropsha, A., *Mol. Inf.*, 2020. https://doi.org/10.1002/minf.202000113
28. Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X.C., Liao, M., Bartlam, M., and Rao, Z., *J. Virol.*, 2008, vol. 82, no. 5, p. 2515. https://doi.org/10.1128/JVI.02114-07
29. Pillaiyar, T., Manickam, M., Namasivayam, V., Hayashi, Y., and Jung, S.H., *J. Med. Chem.*, 2016, vol. 59, no. 14, p. 6595. https://doi.org/10.1021/acs.jmedchem.5b01461
30. The DrugBank database. http://www.drugbank.ca/. Accessed September 28, 2020.
31. Zhang, D.H., Wu, K.L., Zhang, X., Deng, S.Q., and Peng, B., *J. Integr. Med.*, 2020, vol. 18, no. 2, p. 152. https://doi.org/10.1016/j.joim.2020.02.005
32. Ton, A.T., Gentile, F., Hsing, M., Ban, F., and Cherkasov, A., *Mol. Inf.*, 2020, vol. 39, no. 8, 2000028. https://doi.org/10.1002/minf.202000028
33. Beck, B.R., Shin, B., Choi, Y., Park, S., and Kang, K., *Comput. Struct. Biotechnol. J.*, 2020, vol. 18, p. 784. https://doi.org/10.1016/j.csbj.2020.03.025
34. Weininger, D., *J. Chem. Inf. Comput. Sci.*, 1988, vol. 28, no. 1, p. 31. https://doi.org/10.1021/ci00057a005
35. Kreimeyer, K., Foster, M., Pandey, A., Arya, N., Halford, G., Jones, S.F., Forshee, R., Walderhaug, M., and Botsis, T., *J. Biomed. Inf.*, 2017, vol. 73, p. 14. https://doi.org/10.1016/j.jbi.2017.07.012
36. Valueva, M.V., Nagornov, N.N., Lyakhov, P.A., Valuev, G.V., and Chervyakov, N.I., *Math. Comput. Simul.*, 2020, vol. 177, p. 232. https://doi.org/10.1016/j.matcom.2020.04.031
37. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Müller, M.A., Drosten, C., and Pöhlmann, S., *Cell*, 2020, vol. 181, no. 2, p. 271. https://doi.org/10.1016/j.cell.2020.02.052
38. Rahman, N., Basharat, Z., Yousuf, M., Castaldo, G., Rasrelli, L., and Khan, H., *Molecules*, 2020, vol. 25,

no. 10, E2271.
https://doi.org/10.3390/molecules25102271

39. Zsoldos, Z., Reid, D., Simon, A., Sadjad, S.B., and Johnson, A.P., *J. Mol. Graphics Modell.*, 2007, vol. 26, no. 1, p. 198.
https://doi.org/10.1016/j.jmgm.2006.06.002
40. Pyrkov, T.V., Ozerov, I.V., Balitskaya, E.D., and Efremov, R.G., *Russ. J. Bioorg. Chem.*, 2010, vol. 36, p. 446.
https://doi.org/10.1134/S1068162010040023
41. Śledź, P. and Caflisch, A., *Curr. Opin. Struct. Biol.*, 2018, vol. 48, p. 93.
https://doi.org/10.1016/j.sbi.2017.10.010
42. Muratov, E.N., Bajorath, J., Sheridan, R.P., Tetko, I.V., Filimonov, D., Poroikov, V., Oprea, T.I., Baskin, I.I., Varnek, A., Roitberg, A., Isayev, O., Curtalolo, S., Fourches, D., Cohen, Y., Aspuru-Guzik, A., Winkler, D.A., Agrafiotis, D., Cherkasov, A., and Tropsha, A., *Chem. Soc. Rev.*, 2020, vol. 49, no. 11, p. 3525.
https://doi.org/10.1039/d0cs00098a
43. Tsai, K.C., Chen, S.Y., Liang, P.H., Lu, I.L., Mahindroo, N., Hsieh, H.P., Chao, Y.S., Liu, L., Liu, D., Lien, W., Lin, T.H., and Wu, S.Y., *J. Med. Chem.*, 2006, vol. 49, no. 12, p. 3485.
https://doi.org/10.1021/jm050852f
44. Todeschini, R., Consonni, V., Ballabio, D., and Grisoni, F., in *Comprehensive Chemometrics*, Brown, S., Tauler, R., and Walczak, B., Eds., Amsterdam: Elsevier, 2020.
https://doi.org/10.1016/B978-0-12-409547-2.14703-1
45. Adhikari, N., Baidya, S.K., Saha, A., and Jha, T., in *Viral Proteases and Their Inhibitors*, Gupta, S.P., Ed., 2017.
https://doi.org/10.1016/B978-0-12-809712-0.00011-3
46. *OECD Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models*, OECD Series on Testing and Assessment, no. 69, Paris: OECD, 2014.
https://doi.org/10.1787/9789264085442-en
47. Polishchuk, P., *J. Chem. Inf. Model.*, 2017, vol. 57, no. 11, p. 2618.
https://doi.org/10.1021/acs.jcim.7b00274
48. Kumar, V. and Roy, K., *SAR QSAR Environ. Res.*, 2020, vol. 31, no. 7, p. 511.
https://doi.org/10.1080/1062936X.2020.1776388
49. Masand, V.H., Rastija, V., Patil, M.K., Gandhi, A., and Chapolikar, A., *SAR QSAR Environ. Res.*, 2020, vol. 31, no. 9, p. 643.
https://doi.org/10.1080/1062936X.2020.1784271
50. Flom, P., The disadvantages of linear regression, 2018. https://sciencing.com/disadvantages-linear-regression-8562780.html. Accessed September 28, 2020.
51. ChEMBL Database. http://www.ebi.ac.uk/chembl/g/#browse/activities/filter/target_chembl_id%3ACHEMBL3927%20AND%20standard_type%3A(%22IC50%22). Accessed September 28, 2020.
52. Simplex representation of molecular structure: A chemoinformatic tool for calculation of simplex (fragment) descriptors. https://github.com/DrrDom/sirms. Accessed September 28, 2020.
53. Polishchuk, P., Tinkov, O., Khristova, T., Ognichenko, L., Kosinskaya, A., Varnek, A., and Kuz'min, V., *J. Chem. Inf. Model.*, 2016, vol. 56, no. 8, p. 1455.
https://doi.org/10.1021/acs.jcim.6b00371
54. Scikit-learn. Free software machine learning library for the Python programming language. https://scikit-learn.org/stable/. Accessed September 28, 2020.
55. Jaworska, J., Nikolova-Jeliazkova, N., and Aldenberg, T., *ATLA, Altern. Lab. Anim.*, 2005, vol. 33, no. 5, p. 445.
https://doi.org/10.1177/026119290503300508
56. Web-based platform OCHEM. OCHEM user's manual. http://docs.ochem.eu/display/MAN/OCHEM+Introduction. Accessed September 28, 2020.
57. Sushko, Y., Novotarskyi, S., Korner, R., Vogt, J., Abdelaziz, A., and Tetko, I., *J. Cheminf.*, 2014, vol. 6, no. 1, p. 48.
https://doi.org/10.1186/s13321-014-0048-0
58. Breiman, L., RRforest software. http://www.stat.berkeley.edu/~breiman/RandomForests/reg_examples/RFR. Accessed September 28, 2020.
59. Raevsky, O.A., Grigorev, V.Yu., Kireev, D.B., and Zefirov, N.S., *Quant. Struct.-Act. Relat.*, 1992, vol. 11, no. 1, p. 49.
https://doi.org/10.1002/qsar.19920110109
60. Grigorev, V.Yu. and Grigoreva, L.D., *Moscow Univ. Chem. Bull. (Engl. Transl.)*, 2016, vol. 71, no. 3, p. 199.
https://doi.org/10.3103/S0027131416030056
61. Martin, T., Harten, P., and Young, D., TEST (Toxicity Estimation Software Tool), ver. 4.1, Washington DC: US EPA, 2012. www.epa.gov/chemical-research/toxicity-estimation-software-tool-test. Accessed September 28, 2020.
62. Daina, A., Michielin, O., and Zoete, V., *Sci. Rep.*, 2017, vol. 7, p. 42717.
https://doi.org/10.1038/srep42717
63. Lipinski, C.A., Lombardo, F., Dominy, B.W., and Feeney, P.J., *Adv. Drug Delivery Rev.*, 2001, vol. 46, p. 3.
https://doi.org/10.1016/S0169-409X(00)00129-0
64. Dahlin, J.L., Nissink, J.W., Strasser, J.M., Francis, S., Higgins, L., Zhou, H., Zhang, Z., and Walters, M.A., *J. Med. Chem.*, 2015, vol. 58, no. 5, p. 2091.
https://doi.org/10.1021/jm5019093
65. Ertl, P. and Schuffenhauer, A., *J. Cheminf.*, 2009, vol. 1, no. 1, p. 8.
https://doi.org/10.1186/1758-2946-1-8
66. Pastick, K.A., Okafor, E.C., Wang, F., Lofgren, S.M., Skipper, C.P., Nicol, M.R., Pullen, M.F., Rajasingham, R., McDonald, E.G., Lee, T.C., Schwartz, I.S., Kelly, L.E., Lother, S.A., Mitja, O., Letang, E., Abassi, M., and Boulware, D.R., *Open Forum Infect. Dis.*, 2020, vol. 7, no. 4, ofaa130.
https://doi.org/10.1093/ofid/ofaa130
67. Jin, Z., Zhao, Y., Sun, Y., Zhang, B., Wang, H., Wu, Y., Zhu, Y., Zhu, C., Hu, T., Du, X., Duan, Y., Yu, J., Yang, X., Yang, K., Liu, X., Guddat, L.W., Xiao, G., Zhang, L., Yang, H., and Rao, Z., *Nat. Struct. Mol. Biol.*, 2020, vol. 27, no. 6, p. 529.
https://doi.org/10.1038/s41594-020-0440-6

Translated by D. Novikova