https://doi.org/10.26565/2220-637X-2024-42-01 ISSN 2220-637X
УДК: 544.16+577.29

TOWARDS THE DISCOVERY OF MOLECULES WITH ANTI-COVID-19 ACTIVITY: RELATIONSHIPS BETWEEN SCREENING AND DOCKING RESULTS

D. O. Anokhin[*,a], S. M. Kovalenko[*,b], P. V. Trostianko[*,c], A. V. Kyrychenko[*,d], A. B. Zakharov[*,e], T. O. Zubatiuk[†,f], V. V. Ivanov[*,g], O. M. Kalugin[*,h]

[*]*V. N. Karazin Kharkiv National University, School of Chemistry, 4 Svobody sqr., Kharkiv, 61022 Ukraine*

[†]*Mellon College of Science, Carnegie Mellon University, Department of Chemistry, Pittsburgh, Pennsylvania 15213, USA*

a) dmitriy25102002@gmail.com https://orcid.org/0000-0002-4958-2692
b) kovalenko.sergiy.m@gmail.com https://orcid.org/0000-0003-2222-8180
c) trostianko.p.v@gmail.com https://orcid.org/0000-0002-1333-9375
d) a.v.kyrychenko@karazin.ua https://orcid.org/0000-0002-6223-0990
e) abzakharov@karazin.ua https://orcid.org/0000-0002-9120-8469
f) tetiana@zubatyuk.com https://orcid.org/0000-0002-2866-7849
g) vivanov@karazin.ua https://orcid.org/0000-0003-2297-9048
h) onkalugin@gmail.com https://orcid.org/0000-0003-3273-9259

The study presents the results of a combined approach to the theoretical description of potential antiviral activity against COVID-19. We found that pharmacophore screening based on limited experimental data on "protein-ligand" binding complexes might have low predictive ability. Therefore, in this study, we build a model based on the statistical description of QSAR for data obtained from docking which serves as a basis for adequate prediction of ligand activity. We use the logistic regression to construct the predictive model for the main protease M^{pro} inhibitors.

Keywords: *QSAR, Docking, Pharmacophore, Logistic Regression*

Introduction

The problem of drug discovery against COVID-19 disease still actual. As of 25.03.2024, there are 110 565 new infection cases per week and 1141 deaths worldwide. [1]. The experimental evaluation of therapeutic compounds for in vivo COVID-19 antiviral efficacy based ortant to achieve a high selectivity, which can be achieved by advanced data analysis and drug design techniques. Computational chemistry provides a set approaches implemented in corresponding programme code for this purpose. Among them, there are chemoinformatic methods in the general machine learning frameworks as well as molecular modelling approaches, which include molecular dynamic simulation and docking.

First of all, these methods can be applied to database of perspective compounds. Preliminary evaluation includes ligand and protein preparation, pharmacophore set generation which characterises essential features generation for protein-ligand binding site, and pharmacophore screening. of large database. Consequently, the selected molecules will be used for direct docking for evaluation of efficiency of ligand-protein interaction. Hit identification and the lead generation are the consequent stages of computer modelling during drug discovery.

Essential question arises at the stage of pharmacophore screening. Usually, information about possible ligands is restricted by available experimental data (X-Ray, NMR). It is why the structure of pharmacophore set, which is formed by restricted numbers of active ligands, cannot describe full possible interactions within the binding site. In the present article we examined correspondence of pharmacophore screening results and docking results. Essentially, we are interested in possible statistical qualitative model, which can give an additional information about prognostic abilities of pharmacophore model.

As the objects of our investigation, we used SARS-CoV-2 main protease (M^{pro}). SARS-CoV-2 M^{pro} is a key enzyme of coronaviruses which has a function of mediating DNA replication and

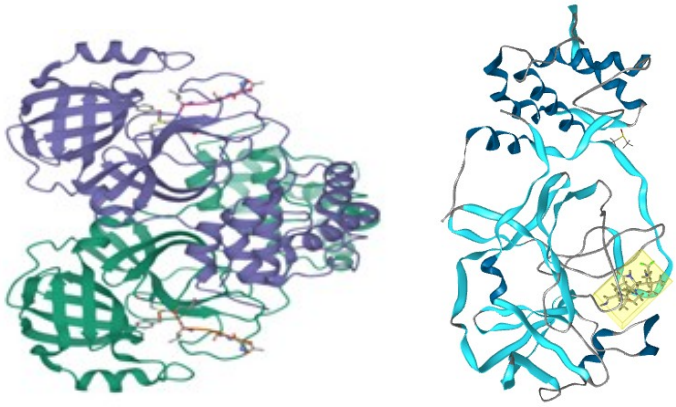
transcription. SARS-CoV-2 M^{pro} inhibitors are investigated in numerous articles (see for instance [2-7]). Because of importance in viral replication, this protein is a common target for drug discovery.

For the building of corresponding chemoinformatic models we used pharmacophore screening and docking by AutoDock 4.2 and AutoDock Vina 1.1 programs. All of mentioned programs are integrated in LigandScout software suite [8]. All proteins, complexes and pharmacophore structures illustrations were done within LigandScout. BIOVIA Draw 2018 program was employed for the ligands formulas representation [9].

All calculations were performed for rather small dan large-scale screening in vitro. To maximize the likelihood of successful screening, it is impotabase composed of 424 5-(phenylsulfonyl)-4-pyrimidone derivatives. Pharmacophore screening and followed by docking respective to all three investigated complexes has been performed.

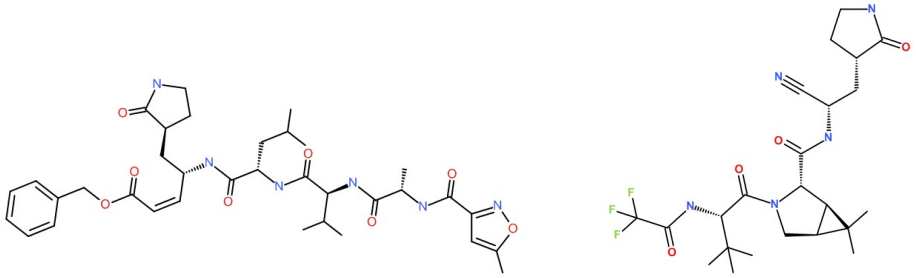
Target proteins and inhibitors

The structures of protein-inhibitor complexes SARS-CoV-2 main protease M^{pro} (PDB code of complexes 6lu7 and 7vh8) are presented in Fig. 1 A and B. The structure of corresponding inhibitors N3/PRD_002214 (6lu7), PF-07321332/nirmatrelvir (7vh8) presented in Fig. 2. The inhibitors of main proteases are oligopeptides. The PDB structures of protein complexes were taken from the RCSB database [10].



a b

Figure 1. A schematic structure of SARS-CoV-2 main protease 6lu7 (a) and 7vh8 (b).



a b

Figure 2. Structural formulas for inhibitors: N3 (a) and PF-07321332 (b).

Ligand library

The library of 424 molecules, which are 2-(5-arylsulfonyl-4-oxo-3,4-dihydro-2-pyrimidinethio)acetamide derivatives (see Fig. 3) was used. Here Ar corresponds to aryl substituent containing alkyl-, halogen-, methoxygroups, *etc*. The R substituent can be a simple aryl, benzyl, N-aryl-N`-piperidyl, *etc*. A variety of polar and non-polar groups, hydrogen bond donors and acceptors, hydrophobic sites, and a number of rotatable bonds make these molecules potentially biologically active.

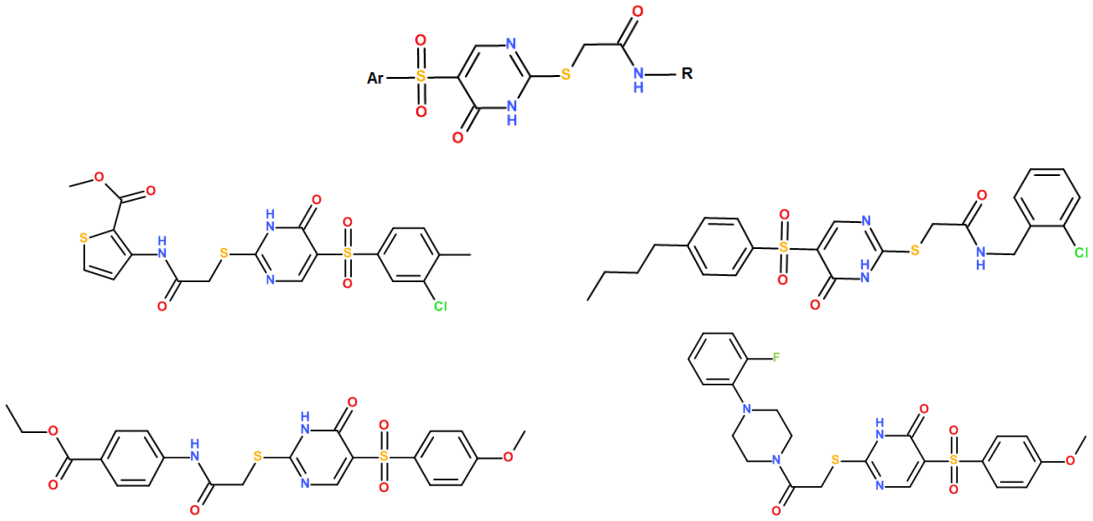


Figure 3. A general structural formula for investigated ligands examples of typical systems.

QSAR modelling: pharmacophore screening and docking

A pharmacophore screening was performed onto 424 molecules by LigandScout 4.4, the program suite with inbuilt pharmacophore creating and matching programs. The pharmacophore screening was performed by "first fitting" mode, i.e. calculation stops after the first fitting, a further geometry modification is not carried out. This mode allows fast screening a large number of molecules. According to LigandScout suite for screened molecules pharmacophore-fit score (PFS) is calculated:

$$PFS = 10n + (9 - 3\min(\mathrm{r}, 3)) \tag{1}$$

where n is a number of matched pharmacophore features, $r = \sqrt{\frac{1}{N}\sum_i r_i^2}$ is a root mean square deviation (RMSD) of pharmacophores and corresponding ligand, r_i – Euclidian distances between matched pair of pharmacophores and ligand features.

After virtual screening we also performed docking procedure for whole library against 6lu7 and 7vh8 protein structures using AutoDock 4.2 and AutoDock Vina 1.1, incorporated in LigandScout. According to AutoDock Vina ideology, there is an exhaustiveness parameter specified as the number of binding modes for one ligand. Exhaustiveness defines a number of parallel searching runs. This parameter usually has been set to 8 and 9 conformations.

When analysing the docking results, we examine the Binding Affinity Score (BAS) and the binding affinity (kcal/mol). The last parameter is a target parameter for docking optimisation procedure. Due to stochastic nature of search algorithm of docking, the results obtained from each run are random and have a different energy. In order to calculate probability of definite random state with defined energy (p_i), the Boltzmann weight factor is calculated

$$p_i = \frac{\exp\left(-\frac{E_i}{kT}\right)}{\sum_j \exp\left(-\frac{E_j}{kT}\right)} \tag{2}$$

Therefore, the total binding energy of the ligand can be estimated as a weighted sum of the obtained binding energies of the modes.

$$E = \sum_i p_i E_i \tag{3}$$

We will denote the corresponding approach as Boltz.

Since pharmacophore matching is not an absolute criterion for activity, we also consider additional QSAR (Quantitative Structure-Activity Relationship) modelling to accurately predict the biological activity of a compound. For all the molecules we have calculated 1974 2D and 3D molecular descriptors by using PaDEL-Descriptor program [11]. The logistic classification regression model [12,13] has been used for description of activity against M^{pro}. In logistic regression (strictly

speaking, it is a non-linear least squares method, NLS, for a logistic function), the dependent variable has only two outcomes (true/false, active/inactive, etc.). This binary variable can be represented as '0' or '1'. To solve the NLS equations for the logistic function, we use the Levenberg-Marquardt approach to Newton's method, which makes it possible to solve the NLS problem even when the matrix is close to degeneracy [14]. The logistic function has the following form

$$Y = \frac{1}{1+\exp(-t)}, \quad t = a_0 + a_1 X_1 + a_2 X_2 + ... \tag{4}$$

Where the Y is the classification response (inactive-active: $0 \le Y \le 1$) and $X_1, X_2, ...$ are descriptors of molecular systems. Corresponding computer program has been implemented in the FORTRAN language.

Results of calculations

We employed a LigandScout suite to create the pharmacophore structures for 6lu7 and 7vh8 protein shown in Fig. 4. Here, the red arrows correspond to H-bond acceptor, green arrows are H-bond donors and yellow regions are hydrophobic fragments. The amino acids which give dominant contributions to ligand-protein interactions designated along with numerations in corresponding protein link.

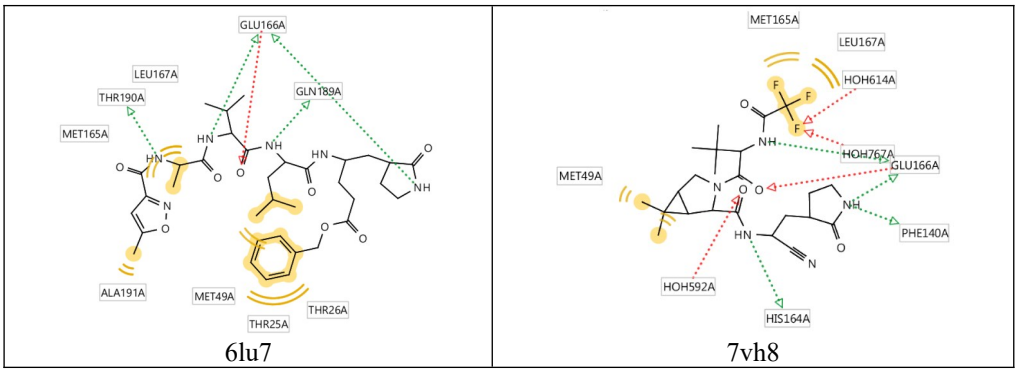


Figure 4. 2D structure of pharmacophores.

Fig. 5 shows the pharmacophore screening results for our molecular library vs the obtained pharmacophore for all two complexes. According to the diagrams, one can divide molecules onto low-active ($PFS < 39$, *i.e.* 3 or less pharmacophore matches); medium-active ($40 < PFS < 49$, *i.e.* the molecules have 4 pharmacophore matches); and high-active ($PFS > 50$, the molecules with 5 or more pharmacophore matches). On the other hand, one can consider molecule with $PFS > 50$ as similar to reference only conditionally since a reference molecule contains ten pharmacophore features. All the presented results based on the structures shown in Fig. 4.

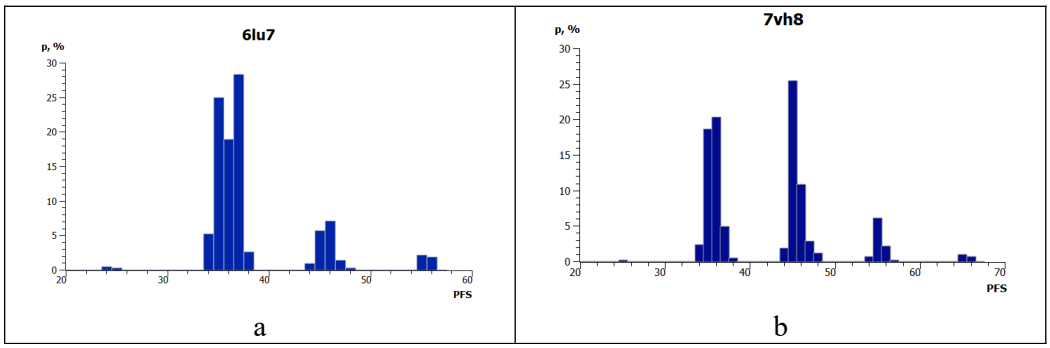


Figure 5. PFS distribution for three complexes under investigation. *p* is a probability of certain *PFS* ($p_i=(N_i/N_{total})\cdot 100\%$)

Some examples of pharmacophore alignment of our sample are presented in Fig. 6. The first case (Fig. 6, a) corresponds to a value of $PFS = 56.59$, *i.e.* 5 pharmacophore matches, such as two hydrogen bond donors, one hydrogen bond acceptor and two hydrophobic interactions. The second case (Fig. 6, b)

corresponds to *PFS* = 24.91. In this case only two features matched: H-bond acceptor and hydrophobicity. This molecule is expected to show lower activity compared to the first one.

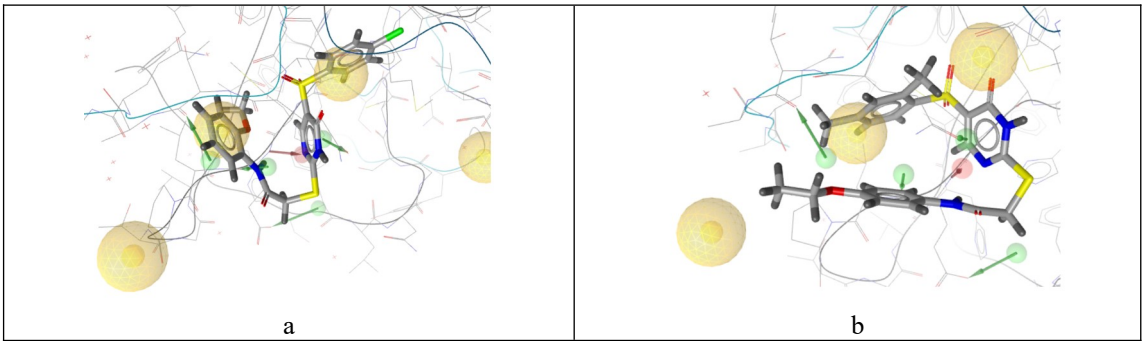


Figure 6. Pharmacophore alignment (6lu7 complex) with a *PFS* = *56.6* (a) and *24.9* (b). Yellow spheres designate hydrophobic features, green arrows designate H-bond donors and red arrows – H-bond acceptors.

Docking of the reference ligand PRD_002214 (or N3 in another designation) with an exhaustiveness parameter equal to 8 (ex = 8) gave a weighted-mean binding affinity score -25.97. Nine ligands with greater activity than N3 have been identified (Fig. 7). We refer these molecules as active.

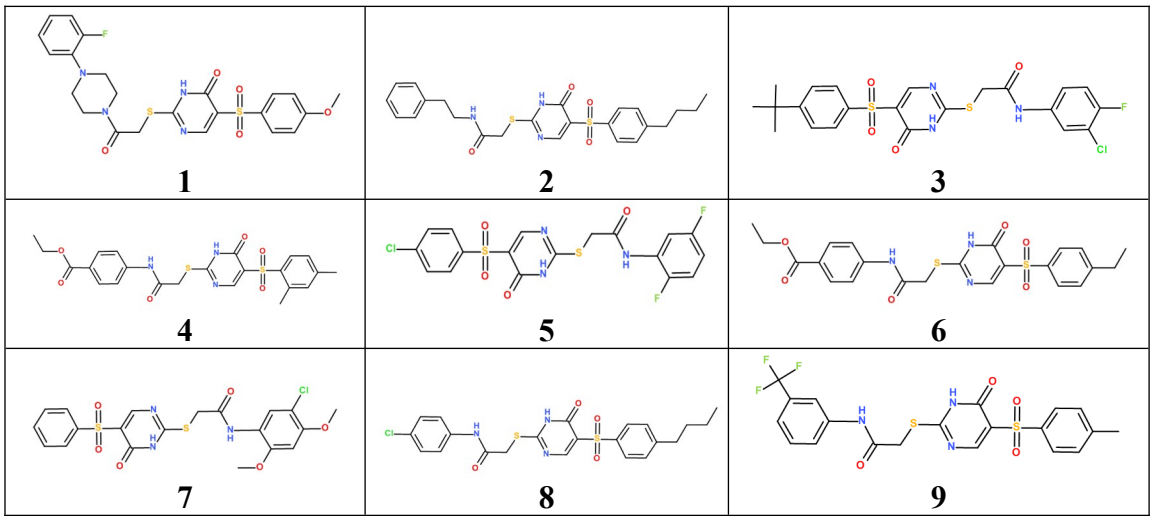


Figure 7. Structure formulas of ligands with higher activity than reference N3 ligands against SARS-CoV-2 M^{pro} (6lu7).

These nine molecules were re-docked by AutoDock Vina with exhaustiveness parameter (ex) equal to 32 and by AutoDock 4. Weighted sums of conformation binding affinities calculated with Eq. (2) are shown in Table 1. In this table, values in the Mean column are Boltzmann-weighted average of the binding affinities according to Eq. (2).

It should be noted that no clear relationship or correlation was found between certain structural features of the molecules and their activity. This can also be confirmed by the lack of significant correlation between molecular descriptors obtained with PaDEL-Descriptor and the molecular activity. Indeed, from the Table 1 one can see that there are no noticeable correlations between PFS and different variant BAS ($R^2 \sim 0.02$)!

Table 1. PFS and Binding Affinity Score of active molecules

Molecule	PFS	Binding Affinity Score (BAS)			
		Vina, ex = 8	Vina, ex = 32	AutoDock 4	Boltz
1	36.55	-31.01	-25.90	-30.27	-30.34
2	35.96	-29.14	-19.76	-19.27	-28.70
3	35.84	-28.94	-21.56	-19.69	-28.32
4	37.22	-28.21	-20.56	-18.24	-27.65
5	45.59	-28.11	-17.87	-21.47	-27.47
6	35.63	-27.97	-8.33	-16.22	-27.83
7	44.70	-27.29	-17.43	-19.05	-26.77
8	36.53	-26.85	-20.23	-19.27	-26.04
9	35.68	-26.15	-18.43	-18.47	-25.44
N3	–	***-25.97***	***-27.47***	***-19.43***	***-26.72***

Table 1 shows good correlation between different BASs, despite the stochastic nature of the algorithms. The best correspondence is obtained between mean value (2) and Vina, ex = 8. The determination coefficient, which can be calculated from the corresponding columns, is equal to $R^2 = 0.99$. The results obtained with the Vina program (ex=8) also showed good agreement with the results obtained with the AutoDock program.

From the Table 1, one can consider the first molecule, as active because the BAS values are high. Therefore, an investigation of molecule 1 (Fig 7.) as an alternative to N3 might could be perspective.

In connection with the results collected in the Table 1 we have analysed the distribution and clustering of the docking and pharmacophore screening results. Corresponding graphs presented in the Fig. 8. The distribution of the dependence of BAS vs PFS (Figure 8a) and Affinity vs PFS (Figure 8b) are similar. It is evident that there is not only a lack of noticeable correlation between docking parameters and PFS, but also that the data for the four groups are clearly clustered. "Each cluster contains definite number of active and inactive molecules. Furthermore, according to our calculations there are *seven active molecules* with *PFS* within range $30 < PFS < 39$, two active molecules with $40 < PFS < 49$ and neither active molecule with PFS>50. Consequently, if we select ligands only according to the pharmacophore matching parameter, we will have to deal with low- and moderately-active ligands. In this case, ligands with the highest binding affinity will be excluded.

An additional and alternative approach to build a predictive model of activity can be based on regression analysis of the functional dependence of docking results on a set of possible molecular parameters (descriptors). However, as our analysis has shown, it is impossible to construct a simple multiple linear regression for this sample even using the partial least squares method. Therefore, we focused our choice on a qualitative logistic regression (4). The results of our calculations for logistic functions based at Vina and Boltz collected in the Table 2 – 3. Using these equations one can classify all molecules onto two categories: active and inactive. As a measure of activity, we were using the affinity of reference ligand for each complex.

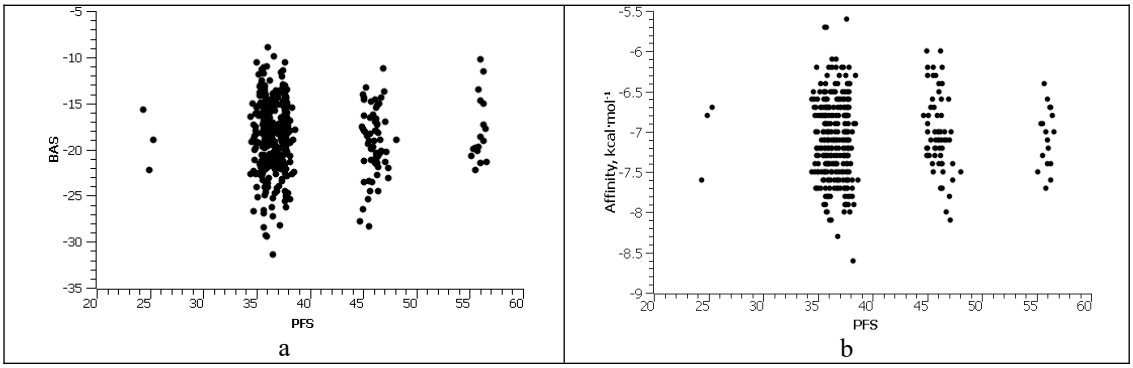


Figure 8. A dependence docking results (BAS, Affinity) *vs* PFS for 6lu7

We selected four parameters as descriptors, which are briefly described in the Table 2. Among the descriptors, the Broto-Moreau autocorrelation function, also known as the autocorrelation of a topological structure, is presented here [15]. The autocorrelation descriptors (AATSC3v and AATS7s) describe how a property is distributed along the topological structure. For the indices AATSC3v and AATS7s the lag parameter is the topological distance between a pair of atoms (3 and 7 respectively in our equations).

Table 2. Parameters of Logistic regression (eq. 4) for sample under consideration (protein 6lu7)

Params	Vina, ex=8	Boltz, eq. (2)	Description
a_0	164.9	157.8	–
ASP-1	-374.8	-352.7	Average simple path, order 1
nRotB	-0.1975	-0.5768	Number of rotatable bonds, excluding terminal bonds
AATSC3v	0.0764	0.0799	Average centered Broto-Moreau autocorrelation – lag 3 / weighted by van der Waals volumes
AATS7s	0.7734	0.9828	Average centered Broto-Moreau autocorrelation – lag 7 / weighted by Sanderson electronegativity
η_{calc}/η_{LOO} (%)	75.9/75.7	79.2/78.3	Percentage of correctly classified molecules
TA/TI	63/259	124/212	Confusion matrix. True active / True inactive
FA/FI	71/31	48/40	False Active / False Inactive
C(A)/C(I)	0.13/-1.68	0.73/-1.67	Centroids: Active / Inactive

More information about the descriptors used can be found in the PaDEL-Descriptor manual [11] and in [16,17]. In the table, one can note a fairly good separation of active/inactive molecules both when using the derived logistic equations η_{calc} and the Leave-One-Out procedure [18] η_{LOO}. A slightly better result was shown by the Boltz approach ~ $\eta_{LOO} \approx 73.8\%$. More detailed information about the accuracy of the calculated logistic equations can be obtained from the confusion matrix (TA/TI, FA/FI in the Table 2). The confusion matrix [19] (sometimes called the error matrix) provides additional information about the accuracy of the classification function. Namely, this 2x2 matrix contains information about the number of active systems (molecules) recognized as active (true active, TA). In the same way, true inactive systems (TI) were determined. Calculations of the number of false classifications of molecules as active (FA) and inactive (FI) give information about the total false recognition. It can be noted that the recognition of active compounds in Boltz is noticeably more accurate (TA=124) than in Vina, ex=8 (TA=63). Also, the distance between the centroids is larger for the Boltz approach, which indicates that the method more clearly separates active and inactive molecules.

For the 7vh8 protein complex, the corresponding logistic parameters shown in Table 3 demonstrated more accurate selection of inactive molecules than active ones (η_{calc}/η_{LOO} (%) = 96.9/96.5 for the Vina, ex=8 calculations).

Table 3. Parameters of Logistic regression (eq. 4) for sample under consideration (protein 7vh8)

Params	Vina, ex=8	Boltz, eq. (2)	Description
a_0	276.3	215.5	–
ASP-1	-636.2	-492.3	Average simple path, order 1
nRotB	0.2621	-0.5857	Number of rotatable bonds, excluding terminal bonds
AATSC3v	0.2306	0.1389	Average centered Broto-Moreau autocorrelation – lag 3 / weighted by van der Waals volumes
AATS7s	0.4514	0.4091	Average centered Broto-Moreau autocorrelation – lag 7 / weighted by Sanderson electronegativity
η_{calc}/η_{LOO} (%)	96.9/96.5	94.8/93.6	Percentage of correctly classified molecules
TA/TI	12/399	13/389	Confusion matrix. True active / True inactive
FA/FI	10/3	20/2	False Active / False Inactive
C(A)/C(I)	-0.74/-4.71	-1.00/-3.75	Centroids: Active / Inactive

Conclusion

In this article, we investigated the predictive ability of the pharmacophore concept on the activity of a given library of 424 derivatives of 2-(5-(arylsulfonyl)-4-oxo-3,4-dihydro-2-pyrimidinethio)acetamide against COVID-19. As target receptors, we used the SARS-CoV-2 main protease. We utilized pharmacophore screening and docking were employed for the selection of active molecular systems. Pharmacophore screening provides a similarity measure between an analyzed ligand and a reference one. Since a reference ligand is selected to have proven activity, one might consider such similarity as a promising indicator of biological activity. However, an objective (more physical) measure of biological activity is a binding affinity energy obtained by docking. For a given library, we performed docking against main protease (Mpro) using a structure of its complex with the oligopeptide inhibitor N3 by AutoDock Vina program. It identified nine ligands with higher affinity than the reference ligand. These molecules were re-docked by Vina with a higher exhaustiveness parameter and then by AutoDock 4. The ligand N-(2-(5-(4-metoxyphenylsulfonyl)-4-oxo-3,4-dihydro-2-pyrimidinesulfo)acetyl-N'-(2-fluorophenyl)piperazine exhibited the highest binding affinity and can be considered as an alternative to N3. We compared pharmacophore screening results and docking results. No correlation between these two values was found. The most active molecules by binding affinity criteria have not so high pharmacophore-fit scores. Therefore, pharmacophore screening is not always an effective method for drug discovery.

QSAR modeling presents an alternative to pharmacophore screening. We constructed a qualitative logistic regression model using a sample of 424 molecules, capable of predicting whether a molecule is active or not. In our approach, we utilized the AutoDock Vina program, which generates multiple conformations for each molecule. We employed the average affinity weighted by the Boltzmann probability factor for regression analysis. This methodology has demonstrated a high level of accuracy in predicting molecular activity.

Acknowledgement

The work was performed as part of a research project Grant 42/0062 (2021.01/0062) "*Molecular design, synthesis and screening of new potential antiviral pharmaceutical ingredients for the treatment of infectious diseases COVID-19*" from the National Research Foundation of Ukraine.

We thank Prof. T. Langer for the opportunity to work with the LigandScout suite.

References

1. https://www.worldometers.info/coronavirus (Last updated: April 13, 2024, 01:00 GMT)
2. Jin Z et al. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature*. **2020**, *582*(7811), 289–293. https://doi.org/10.1038/s41586-020-2223-y
3. Zhao Y et al. Crystal structure of SARS-CoV-2 main protease in complex with protease inhibitor PF-07321332. *Protein Cell*, **2022**, *13*(9), 689–693. https://doi.org/10.1007/s13238-021-00883-2
4. Yevsieieva, L. V.; Lohachova, K. O.; Kyrychenko, A. V.; Kovalenko, S. M.; Ivanov, V. V.; Kalugin O.N. Main and papain-like proteases as prospective targets for pharmacological treatment of coronavirus SARS-CoV-2. *RSC Advances*. **2023**, *13*, 35500-35524. https://doi.org/10.1039/D3RA06479D
5. Citarella, A.; Dimasi, A.; Moi, D.; Passarella, D.; Scala, A.; Piperno, A.; Micale, N. Recent Advances in SARS-CoV-2 Main Protease Inhibitors: From Nirmatrelvir to Future Perspectives. *Biomolecules* **2023**, *13*, 1339. https://doi.org/10.3390/biom13091339
6. Huang, C.; Shuai, H.; Qiao, J. *et al.* A new generation M^{pro} inhibitor with potent activity against SARS-CoV-2 Omicron variants. *Sig Transduct Target Ther* **2023**, *8*, 128. https://doi.org/10.1038/s41392-023-01392-w
7. Lohachova, K. O.; Sviatenko, A. S; Kyrychenko, A; Ivanov, V. V.; Langer, T.; Kovalenko, S. M.; Kalugin, O. N. Computer-aided drug design of novel nirmatrelvir analogs inhibiting main protease of Coronavirus SARS-CoV-2. *Journal of Applied Pharmaceutical Science*, **2024**, *14(05)*, 232-239. http://doi.org/10.7324/JAPS.2024.158114
8. Pojtanadithee, P; Isswanich K; Buaban, K; Chamni, S, *et al*. A combination of structure-based virtual screening and experimental strategies to identify the potency of caffeic acid ester

derivatives as SARS-CoV-2 3CLpro inhibitor from an in-house database. *Biophysical Chemistry*, **2024,** *304*, 107125. https://doi.org/10.1016/j.bpc.2023.107125
9. BIOVIA, Dassault Systemes, BIOVIA Draw 2018, San Diego: Dassault Systemes, 2018.
10. Berman, H.M.; Westbrook, J; Feng, Z; Gilliland, G.; Bhat, T.N.; Weissig H., Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Research.* **2000**, *28*, 235–242. doi: 10.1093/nar/28.1.235. URL: http://www.rcsb.org/
11. Yap, C.W. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem.* **2011**, *32*(7), 1466–1474. http://doi.org/10.1002/jcc.21707
12. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning. Springer Series in Statistics.* **2009**, 119–122. http://doi.org/10.1007/978-0-387-84858-7
13. Hosmer Jr, D. W.; Lemeshow, S.; Sturdivant, R. X. *Applied Logistic Regression*; John Wiley & Sons, **2013**; Vol. 398.
14. Bilski, J.; Smol, J.; Kowalczyk, B.; Grzanek, K.; Izonin, I. Fast Computational Approach to the Levenberg-Marquardt Algorithm for Training Feedforward Neural Networks. *Journal of Artificial Intelligence and Soft Computing Research.* **2023**, *12*(2), 45-61. https://doi.org/10.2478/jaiscr-2023-0006
15. Moreau, G; Broto, P. Autocorrelation of a topological structure: A new molecular descriptor. *Nouv. J. Chim*. **1980**, *4*, 359–360.
16. Kier, L. B.; Hall, L. H. Molecular connectivity in chemistry and drug research, *New York: Academic Press.* **1976**. https://doi.org/10.1016/b978-0-12-406560-4.x5001-6
17. Todeschini, R.; Consonni, V. Molecular descriptors for chemoinformatics, *Weinheim: Wiley VCH.* **2009**, 27–37. https://doi.org/10.1007/b94608
18. James, G.; Witten, D.; Hastie, T.; Tibshiranu, R.; Taylor, J. Resampling Methods. in *An Introduction to Statistical Learning*, Springer Texts in Statistics, **2023,** 201–226. https://doi.org/10.1007/978-3-031-38747-0_5
19. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognition Letters. **2006**, 27(8),* 861-874. https://doi.org/10.1016/j.patrec.2005.10.010

Received 28.04.2024 *Accepted 07.06.2024*

Д. О. Анохін[*], С. М. Коваленко[*], П. В. Тростянко[*], А. В. Кириченко[*], А. Б. Захаров[*], Т. О. Зубатюк[†], В. В. Іванов[*], О. М. Калугін[*]. Відкриття молекул з анти-COVID-19 активністю: зв'язок між результатами скринінгу та докінгу.

[*]*Харківський національний університет імені В.Н. Каразіна, хімічний факультет, майдан Свободи, 4, Харків, 61022, Україна*

[†]*Меллонський науковий коледж, Університет Карнегі-Меллон, хімічний факультет, Піттсбург, Пенсильванія 15213, США*

Дослідження представляє результати комбінованого підходу до теоретичного опису потенційної противірусної активності проти COVID-19. Виявлено, що фармакофорний скринінг, заснований на обмежених експериментальних даних щодо комплексів «білок-ліганд», може мати погану передбачувану здатність. Разом із тим побудова моделі на основі статистичного опису QSAR для даних, отриманих в результаті докінгу, може служити основою для адекватного прогнозу. Використання логістичної регресії, як варіанту класифікаційної функції, дозволило побудувати прогностичну модель для основної протеази Mpro.

Ключові слова:.*QSAR, докінг, фармакофор, логістична регресія.*

Надіслано до редакції 28.04.2024 *Прийнято до друку 07.06.2024*