DOI: 10.14529/jsfi260107

High-Throughput Computational Discovery of Anti-Coronavirus Agents in the COVID-19 Era: Crucial Insights for Combating Emerging Biogenic Threats

Dmitry S. Druzhilovskiy[1], *Dmitry A. Filimonov*[1], *Pavel V. Pogodin*[1], *Anastasia V. Rudik*[1], *Leonid A. Stolbov*[1], *Olga A. Tarasova*[1], *Alexander V. Veselovsky*[1], *Vladimir V. Poroikov*[1]

In May 2020, the Joint European Disruptive Initiative (JEDI) launched the "Billion Molecules against COVID 19" challenge – an extensive open science effort aimed at identifying small molecule inhibitors of SARS-CoV-2 and related human receptors. Our research group joined this initiative among 130 international teams, focusing on the in silico screening for potential anti coronavirus agents that target three viral proteins and one human receptor. The screening campaign covered more than one billion synthetically accessible structures, including approved pharmaceuticals. By July 17, 2020, our team submitted a subset of 10000 prioritized compounds to the organizers for expert evaluation. The results from our selection, together with those from 19 other participating teams, contributed to a pool of approximately 1000 molecules selected for chemical synthesis and bioactivity testing. In total, 878 compounds were successfully synthesized and evaluated for inhibitory activity against various SARS-CoV-2 targets as well as the human serine protease TMPRSS2. Ultimately, 27 compounds – including one proposed by our group – demonstrated measurable anti coronavirus activity. The collective outcomes of these collaborative efforts were reported in the "Molecular Informatics" journal in 2024. In the present study, we summarize our participation in the JEDI challenge and discuss broader methodological and organizational considerations critical for improving the efficiency of rapid scientific responses to future emerging biological threats.

Keywords: COVID-19, JEDI COVID-19 challenge, anti-coronavirus agents, in silico screening, future biogenic threats, effective response.

Introduction

In December 2019, physicians from several hospitals in Wuhan notified the local Center for Disease Control and Prevention of pneumonia cases of unknown etiology. Subsequently, Vision Medicals (Guangzhou) confirmed a novel coronavirus in specimens from Wuhan Central Hospital. In January 2020, the initial SARS-CoV-2 genome sequence was deposited in GISAID [33]. Concurrently, infections were identified beyond China, in Thailand, Japan, and the United States. On January 30, the World Health Organization (WHO) proclaimed a Public Health Emergency of International Concern (PHEIC), reporting 9800 cases and 213 fatalities. The disease received its official designation, "COVID-19", in February 2020; on March 11, WHO classified it as a pandemic, with over 118000 cases in 114 countries and 4291 deaths.

Mitigating the COVID-19 pandemic necessitated a coordinated international effort, comprising border closures and lockdowns, reconfiguration of healthcare systems, rigorous sanitary protocols, advancement of diagnostic tools and vaccines, and identification of new therapeutic modalities [64]. After release of the viral genome sequence, laboratories rapidly designed RT PCR assays targeting SARS-CoV-2, with WHO issuing technical guidance on detection, testing, and case management by January 10, 2020 [48]. National public health laboratories and manufacturers produced and scaled PCR based tests during January–February 2020; by March 2020, many

[1]Institute of Biomedical Chemistry (IBMC), Moscow, Russian Federation

countries had authorized emergency use diagnostic assays. Following viral genome publication, multiple groups initiated vaccine design programs using mRNA, viral vector, inactivated, and protein subunit platforms. In March 2020, at least four vaccine candidates entered first in human (phase I) clinical trials, marking an unprecedented acceleration compared to traditional vaccine timelines [48].

In early 2020, before the authorization of any pharmaceutical agents specifically indicated for COVID-19 treatment, several small-molecule drugs previously employed against other viral infections were proposed for repurposing against SARS-CoV-2. These included lopinavir & ritonavir (in combination), interferon (type I; mainly IFN α/β), ribavirin, chloroquine, hydroxychloroquine, favipiravir, remdesivir, and ivermectin. Except for remdesivir, most of these compounds were subsequently shown to lack clinical efficacy, and some (e.g., chloroquine and hydroxychloroquine) were associated with notable adverse effects. Nevertheless, during the initial stage of the pandemic, they comprised the primary therapeutic approaches available for managing SARS-CoV-2 infection [51].

It became evident that the discovery of new anti coronavirus drugs was urgently needed. In addition to extensive studies conducted by major pharmaceutical companies, numerous academic groups also sought to identify novel promising candidates [37].

In response to the unmet need for effective COVID-19 therapies, the Joint European Disruptive Initiative (JEDI) launched the "Billion Molecules against COVID-19 Grand Challenge" on April 23, 2020 [40]. Following an invitation from our European colleagues, including Prof. Dr. Alexandre Varnek (University of Strasbourg), we chose to participate in this collaborative project.

The terms and conditions of this study were defined as follows:

(1) to conduct virtual screening of more than one billion synthetically accessible compounds, including clinically approved drugs;

(2) to assess their potential interactions with at least three molecular targets implicated in antiviral activity against coronaviruses;

(3) to employ three independent computational methodologies; and

(4) to complete the study within the period of May–June 2020.

The set of targets available for subsequent experimental validation comprised the following proteins: 3C-like protease (3CLpro), papain-like protease (PLpro), transmembrane serine protease 2 (TMPRSS2), spike glycoprotein (S), nucleocapsid protein (N), and RNA-dependent RNA polymerase (RdRp).

Participants were required to submit compound lists containing 10000 molecules for a minimum of three protein targets (amounting to a total of 30000 compounds) using the designated *.csv template. Additionally, a detailed report describing the applied computational methods and performance metrics, selected targets, compound libraries, and obtained results was to be submitted in the *.docx template provided by the organizers.

The consolidated results obtained by participants of the JEDI COVID-19 Challenge are reported in the joint publication [67]. In summary, 31 research teams proposed a total of 639024 candidate compounds with putative activity against the aforementioned anti-coronavirus targets. Among these, 878 compounds were synthesized and subjected to experimental evaluation in the corresponding biological assays. As a result, 27 compounds demonstrating weak inhibitory or binding activity were identified through binding, cleavage, and/or viral suppression assays. It was concluded that the open-science framework adopted in the JEDI COVID-19

Challenge made a measurable contribution to the collective knowledge base, thereby facilitating future drug discovery initiatives aimed at the development of improved therapeutic agents against SARS-CoV-2 [67].

In this manuscript, we describe our methodology, developed within the framework of the JEDI COVID-19 Challenge, for the selection of putative anti coronavirus agents from a chemical space comprising billions of compounds, as well as the results obtained and the insights gained, which may contribute to future strategies for mitigating the emerging biogenic threats.

The article is organized as follows. The Introduction describes the context of the onset of the COVID-19 pandemic and the background of launching the open JEDI "Billion Molecules against COVID-19" initiative, which aimed to urgently search for new antiviral drugs. In Section 1 we detail the proposed computational workflow, which includes three sequential stages: initial hit selection based on structural similarity (MNA and QNA), filtering and ranking using machine learning algorithms (PASS and GUSAR software), and verification of the results via molecular docking. This section also describes the selection process for viral targets and the preparation of a consolidated library of over one billion synthetically accessible compounds. Section 2 provides an analysis of the data obtained at each stage of the virtual screening for potential inhibitors of the 3CLpro and PLpro proteases, RdRp polymerase, and the TMPRSS2 receptor. The authors also share key takeaways from their participation in the JEDI challenge, analyzing the reasons for the low efficiency of mass screening under conditions of initial scarcity and noisiness of the training data. Section 2.2 summarizes the overall results of the study, emphasizing that despite the limitations and inaccuracy of early data, the application of machine learning methods and molecular modeling drastically reduced the material, financial, and time costs of experimental validation, narrowing down the screening of billions of molecules to the synthesis of just 878 compounds.

1. Materials and Methods

Considering the incomplete and occasionally contradictory information available on May 4, 2020, regarding the SARS-CoV-2 virus and its interactions with host cells, we developed a systematic approach for the virtual screening of potential anti-coronavirus hits from extensive chemical libraries.

The proposed workflow comprised three successive stages:

- Initial hit selection. Potential hits were identified among more than one billion compounds by assessing structural similarity to reference molecules with experimentally confirmed anti-coronavirus activity.
- Filtering and ranking. The selected compounds were further filtered and prioritized using machine learning algorithms implemented in the PASS and GUSAR (see Sections 1.2 and 1.3) software packages. The corresponding training sets were continuously expanded and refined throughout the project to improve predictive performance.
- Verification by molecular modeling. Representative compounds from the prioritized set were analyzed using molecular modeling methods to evaluate their potential interactions with SARS-CoV-2 targets. A schematic representation of the overall workflow for selecting compounds with potential anti-coronavirus activity is presented in Fig. 1.

The following sections describe four distinct methods applied across the three stages.

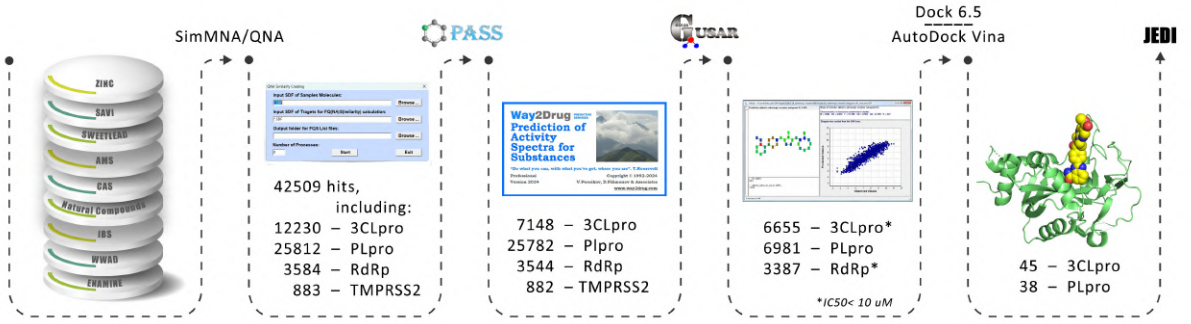


Figure 1. General workflow and results of selection of anti-coronavirus hits

1.1. Similarity Assessment

"Similar molecules tend to exhibit similar biological activities" [45]. Although this principle is occasionally violated in the presence of the so called activity cliffs [18], it remains a cornerstone of medicinal chemistry and is extensively employed in the design of structural analogues presumed to interact with the same target or to elicit comparable pharmacological responses [78]. Furthermore, it constitutes the method of choice in studies involving novel pharmacological targets, particularly when the number of known ligands is insufficient for the construction of a pharmacophore model or the development of a (Quantitative) Structure-Activity Relationship ((Q)SAR) regression or classification models.

At present, no universal method exists for assessing the similarity between molecules belonging to different chemical classes and exhibiting diverse biological activities [6]. Within the framework of the JEDI COVID-19 Challenge, molecular similarity was evaluated using our original molecular descriptors – Multilevel Neighborhoods of Atoms (MNA) [24] and Quantitative Neighborhoods of Atoms (QNA) [25]. These descriptors have demonstrated high efficacy in the analysis of structure-activity relationships for heterogeneous datasets, which is fully consistent with the objectives of the present study, namely the virtual screening of compounds with potential anti-coronavirus activity among more than one billion molecules. Earlier, we examined the applicability of these descriptors for activity prediction based on similarity across a dataset comprising 16770 inhibitors of HIV-1 protease, reverse transcriptase, and integrase [71], and identified both the advantages and limitations of this approach [20].

To conduct the similarity search and identify compounds exhibiting the desired biological activity among a library exceeding one billion molecules, we selected "reference substances", defined as the most potent inhibitors of the four investigated targets reported as of June 2020. These reference substances were subsequently employed as query molecules.

3CLpro. The five most active compounds were collected from various sources and evaluated using different experimental protocols. The activities of GC376, Tideglusib, 11b, and TZDZ-8 were obtained from the respective original publications [16, 41, 49], while MAT-POS-916a2c5a-1 was selected from the PostEra resource [59]. All five compounds were tested against the recombinant SARS-CoV-2 main protease (3CLpro) and exhibited low micromolar inhibitory activities.

PLpro. The PLpro inhibitors 6-thioguanine, GRL0617, 679818, and psoralidin were selected from the corresponding original reports [13, 34, 62] as the most potent inhibitors of the SARS-CoV papain-like protease.

RdRp. The selection of the most active compounds was performed using data from the Stanford Coronavirus Antiviral Research Database [70]. Three compounds were identified as

lead candidates: PubChem CID 44468216 (GS 441524), PubChem CID 121304016 (Remdesivir), and ChEMBL ID CHEMBL2178720 (β-D-N4-Hydroxycytidine). The antiviral activities of GS 441524 and Remdesivir were documented in multiple preprints [9, 17, 61, 63, 66, 77], whereas evidence for β-D-N4-Hydroxycytidine originated from a single study [22]. All three compounds exhibited submicromolar half maximal effective concentrations (EC50) in assays employing SARS-CoV-2 and human cell lines. Notably, both Remdesivir and GS 441524 were also reported to suppress viral RNA expression, corroborating their potent antiviral properties.

TMPRSS2. The selection of the most potent compounds targeting TMPRSS2 was conducted using data retrieved from the ChEMBL database [12]. Three chemical entities exhibiting submicromolar Ki values were identified: CHEMBL1809250, CHEMBL1229259, and CHEMBL1809251. According to the assay description provided in ChEMBL, these compounds were evaluated against the recombinant catalytic domain of TMPRSS2 expressed in Escherichia coli, employing D-cyclohexylalanine-Pro-Arg-AMC as a fluorogenic substrate and fluorescence plate reader analysis for activity quantification. The experimental results were originally reported in reference [68].

In addition to the reference compounds reported in available publications and databases, we also included ligands complexed with SARS-CoV-2 proteins from the Protein Data Bank (PDB) [60] in the similarity search. Ligands from the following six complexes were used: 6LU7 (PRD_002214), 7BRP (HU5), 7BRR (K36), 6Y2G (O6K), 6W63 (X77), and 7BV2 (F86). Ligand IDs are given in parentheses.

1.2. Machine Learning with PASS Computer Program

PASS (Prediction of Activity Spectra for Substances, version 2019) is a computational system developed to predict more than five thousand biological activities with an average Independent Accuracy of Prediction (IAP), quantified as the Receiver Operating Characteristic Area Under the Curve (ROC AUC), of approximately 0.97. These predictions are derived solely from the structural formula of a drug-like compound [57]. The development of PASS began in the late 1980s [10], and over the subsequent three decades, the training datasets have been progressively refined, the range of predictable biological activities expanded, and extensive benchmarking of chemical descriptors and machine learning algorithms conducted [27, 58].

PASS 2019 utilizes structure-activity relationship (SAR) analysis for 1025468 biologically active compounds employing MNA descriptors in combination with a modified nave Bayes classifier [27]. This methodology enables accurate SAR characterization for compounds within the training set and demonstrates sufficient generalizability to provide reliable predictions of biological activity profiles for novel chemical entities, even in the presence of incomplete training data [58].

For each compound under prediction, PASS calculates two probabilities: Pa, representing the likelihood of belonging to the class of "actives", and Pi, the likelihood of belonging to the class of "inactives". By default, compounds for which Pa exceeds Pi are classified as "active".

The performance of PASS exceeds that of other established methods for predicting biological activity profiles, as demonstrated by comparative computational studies [3, 32, 52]. The Professional version of PASS provides functionality for constructing novel training sets, retraining the program to generate an updated SAR knowledge base, and assessing predictive accuracy and reliability through leave-one-out and 20-fold cross-validation procedures, respectively.

Within the framework of the present study, a specialized training set was developed by systematically compiling data from both freely accessible and commercial databases [15], as well as from a wide range of relevant scientific publications. Compounds exhibiting IC50 values below 10 μM were designated as active to serve as the selection threshold. To enhance the representativeness of the chemical space, all newly available information on the structures and activities of anti-coronaviral agents was incorporated into the PASS 2019 training set. Upon completion of the training and validation procedures, an updated SAR knowledge base with the following characteristics was obtained: 1025630 substances, 106828 unique MNA descriptors, 8 selected activities; average IAP equals to 0.9138.

Characteristics of SAR models for each particular activity are given in the Tab. 1. Here, N denotes the number of compounds in the training set that exhibit the given activity; *IAP* (Invariant Accuracy of Prediction) is the predictive performance estimated by leave-one-out cross-validation and is equivalent to the AUC ROC; *20-F IAP* denotes the IAP estimated using 20-fold cross-validation.

Table 1. Characteristics of SAR models for different anti-coronavirus activities

N	**IAP**	**20-F IAP**	**Activity Type**
62	0.9585	0.9587	3C-Like Protease (SARS-CoV) Inhibitors
18	0.9908	0.9909	3C-Like Protease (SARS-CoV-2) Inhibitors
6	0.8296	0.8320	Papain-Like Protease (SARS-CoV-2) Inhibitors
3	0.9970	0.9980	RNA-Directed RNA Polymerase (SARS-CoV-2) Inhibitors
808	0.7535	0.7535	SARS-CoV-2 infection reduction in cell-based assay
5	0.9678	0.9684	SARS-CoV-2 viral Entry Inhibitors
371	0.8129	0.8147	Spike Glycoprotein (S) (SARS-CoV-2)/ACE2 Interaction Inhibitors
3	1.0000	1.0000	Transmembrane Protease Serine 2 (TMPRSS2) Inhibitors

As evident from the data presented above, the accuracy (leave-one-out cross-validation) and predictive performance (20-fold cross-validation) of the developed specialized version of PASS are sufficiently high to support its practical application. This conclusion was also supported by the results of prediction for the reference substances (remdesivir, umifenovir, etc.).

1.3. Machine Learning with GUSAR Computer Program

GUSAR (General Unrestricted Structure-Activity Relationships) is a software for the quantitative structure-activity relationship (QSAR) analysis based on compound structural formulas and corresponding activity or property data, and for predicting activities or properties of novel compounds. It enables the development of (Q)SAR models for organic molecules from both homogeneous and heterogeneous chemical classes. GUSAR employs QNA descriptors, which

represent a molecule as a set of tuples of real values P, Q. The P and Q values are calculated for each atom in a molecule using the connectivity matrix together with values of the standard ionization potential and electron affinities of its constituent atoms.

The current version of GUSAR additionally incorporates selected physicochemical descriptors and biological descriptors derived from Pa-Pi predictions produced by the PASS algorithm. The underlying modeling procedure is based on the self-consistent regression (SCR) method [28], which is used in combination with nearest-neighbor evaluation and a radial basis function artificial neural network (RBF ANN) constructed from SCR outputs to obtain a multiple-model consensus [81]. A comparative study of the first GUSAR version with widely used methodologies such as CoMFA, CoMSIA, GOLPE/GRID, and HQSAR demonstrated clear advantages of this approach for QSAR model construction [25]. In the Collaborative Modeling Project for Androgen Receptor Activity (CoMPARA), GUSAR predictions were also found to be robust and accurate, further supporting the reliability of the method [50].

Within the framework of the JEDI COVID-19 Challenge, QSAR models were developed using the GUSAR software for three viral targets: 3CLpro, PLpro, and RdRp. For 3CLpro and RdRp, regression models were obtained with the following characteristics: 3CLpro, N = 45, $R^2 = 0.981$, $Q^2 = 0.836$, F = 7.220, SD = 0.410, V = 11; RdRp, N = 888, $R^2 = 0.875$, $Q^2 = 0.818$, F = 10.570, SD = 0.501, V = 213. Here, N is the number of compounds in the training set; R^2 is the determination coefficient; Q^2 is the leave-one-out cross-validated determination coefficient; SD is the standard deviation; and V is the number of variables.

Due to the small size of the training set, only a classification model was built for PLpro, with the following characteristics: N = 16, Sens = 1.000, Spec = 0.700, BA = 0.850, V = 3. Here, N is the number of compounds in the training set; Sens denotes sensitivity; Spec, specificity; BA, balanced accuracy; and V is the number of variables.

1.4. Molecular Modeling for Verification of Selection

The final validation of a selected subset of hits was conducted using molecular docking to predict ligand binding poses and estimate binding affinities based on docking scores. Docking calculations were performed with DOCK 6.5 [76] and AutoDock Vina [4]. The scoring function cutoffs for compound selection were set to −65 kcal/mol for DOCK 6.5 and −8.0 kcal/mol for AutoDock Vina, respectively. The resulting docking poses were visually inspected to evaluate their compatibility with the subpockets within the protease active sites and to analyze key binding interactions, including hydrogen bonding, steric complementarity, and electrostatic fit.

1.5. Targets Selection

Four out of the six targets proposed by the organizers of the JEDI COVID-19 Challenge against COVID-19 were selected based on the following criteria: (1) the critical role of the target in coronavirus entry into host cells or viral replication; (2) the availability of reference compounds enabling activity assessment by similarity; (3) the availability of data on drug-like compounds suitable for constructing (Q)SAR training sets; and (4) the presence of a resolved three-dimensional structure in the Protein Data Bank. Targets satisfying at least three of these four criteria were deemed suitable for subsequent analysis.

3-chymotrypsin-like protease (3CLpro/Mpro). The 3C-like protease (3CLpro), also referred to as nonstructural protein 5 (Nsp5), is first auto-catalytically cleaved from the viral

polyprotein to yield the mature enzyme. Subsequently, it mediates proteolytic processing of downstream nonstructural proteins at 11 distinct cleavage sites, thereby releasing Nsp4-Nsp16. Numerous three-dimensional structures of this protease are currently available in the Protein Data Bank (PDB). At the initial stage of this study, all available crystallographic structures of 3CLpro were retrieved from the RCSB PDB and analyzed to identify key structural features involved in inhibitor binding. For molecular docking investigations, the crystal structure with PDB ID 6LU7, complexed with the peptide-like inhibitor N3, was selected as the target. This structure was chosen because it contains one of the largest inhibitors, which closely mimics the natural substrate of the protease. Protein structure preparation was carried out using the SYBYL-X 8.1 software suite [73] and involved the following steps: (a) removal of the co-crystallized inhibitor, water molecules, and ions; (b) addition of hydrogen atoms; (c) assignment of atomic charges using the Gasteiger—Hückel method; and (d) energy minimization in vacuum employing the Tripos force field.

Papain-like proteinase (PLpro). PLpro cleaves the N terminal region of the replicase polyprotein to release Nsp1, Nsp2, and Nsp3, an essential step in assembling the viral replicase complex and enabling efficient viral replication. The crystal structure 6WUU was selected as the target for molecular docking, and it was prepared using the same protocol previously applied for 3CLpro.

RNA-dependent RNA polymerase (RdRp). Nsp12, a highly conserved protein among coronaviruses, serves as the RNA-dependent RNA polymerase (RdRp) and constitutes the central catalytic component of the viral replication-transcription complex.

Transmembrane peptidase serine 2 (TMPRSS2). TMPRSS2 mediates proteolytic cleavage of the SARS-CoV-2 spike protein, thereby enhancing viral infectivity. However, the three-dimensional structure of TMPRSS2 has not yet been experimentally determined.

1.6. Libraries

Nine libraries were used for preparation of "billion compounds" set for virtual screening.

Library 1. ZINC [82] included 920839556 structures. Over 750 million compounds were potentially purchasable.

Library 2. SAVI (Synthetically Accessible Virtual Inventory) [65] included about 1.75 billion proposed products structures with reactions generated in the first full enumeration of the SAVI project. Number of the synthesizable compounds is about 976 million (621 million without stereoisomers).

Library 3. SWEETLEAD [72] included 9127 structures (7636 without stereoisomers).

Library 4. AMS (Aldrich Market Select) [1] included 4787319 structures, with samples available in stock of Merck KGaA collected in the framework of the program "Antimicrobial Stewardship".

Library 5. Antiviral CAS dataset [2] included 49408 structures of antiviral compounds and their analogs collected by Chemical Abstracts Services.

Library 6. Natural Compounds Set included 118894 structures of natural compounds collected by our team from several publicly available databases: ChEBI [11], NANP DB [55], NPASS [54], NuBBE DB [56], UNPD [75].

Library 7. IBS Natural Compounds Set [38] included 69034 structures of natural compounds, their analogs and derivatives, which samples could be purchased from InterBioScreen Ltd.

Library 8. WWAD (World Wide Approved Drugs) [79] included 4108 structures of the launched drugs prepared by our team in the framework of our project dedicated to drug repurposing.

Library 9. ENAMINE in-stock compounds [23] included 1.94 million structures that could be obtained from Enamine Ltd.

All data were subjected to pre-processing and standardization procedures using ChemAxon JChem Instant software and the in-house developed program ClearSDF, in full accordance with current methodological recommendations [29–31]. Following the data curation and prioritization of compounds with a higher probability of experimental availability or synthetic feasibility, a final library comprising 1082000000 structures was prepared and subsequently analyzed using the computational approaches described above.

A unified dataset of curated chemical compound structures was constructed by integrating data from nine independent sources, followed by the removal of duplicate entries. Establishing structural uniqueness requires evaluation of molecular graph isomorphism, a NP-complete problem, making pairwise comparisons among approximately two billion structures computationally infeasible. To overcome this, QNA descriptors were calculated for each structure, and a single real-valued parameter – the Q-index (the sum of atomic Q values) – was assigned to each compound. Using the quicksort algorithm, compounds were ordered by increasing Q-index. Only those compounds with Q-index differences below 10^{-9} were subsequently tested for isomorphism. This approach reduced the number of required graph isomorphism checks from N(N-1)/2N(N-1)/2 to a near-linear-scale computation.

2. Results and Discussion

2.1. Selection of Potential Anti-coronavirus Agents

As illustrated in Fig. 1, based on the evaluation of MNA and QNA similarity for the reference compounds described in Section 1, a total of 42509 hits were identified. These included 12230 putative 3CLpro inhibitors, 25812 putative PLpro inhibitors, 3584 putative RdRp inhibitors, and 883 putative TMPRSS2 inhibitors. Subsequent selection was performed using PASS predictions, yielding 7148 potential 3CLpro inhibitors, 25782 potential PLpro inhibitors, 3544 potential RdRp inhibitors, and 882 potential TMPRSS2 inhibitors.

Owing to the absence of a resolved spatial structure for transmembrane peptidase serine 2 (TMPRSS2), and the inability to construct both regression and classification models for its inhibitors using the GUSAR platform, this stage of selection was considered final for that molecular target.

As the probability of TMPRSS2 inhibitory activity predicted by PASS is below 0.4, it may be inferred that the likelihood of detecting this activity experimentally is relatively low. Nevertheless, should the prediction be experimentally confirmed, the identified compound could serve as a lead structure representing a novel chemical class associated with the investigated biological activity (New Chemical Entity) [26].

Subsequent analyses were performed for the remaining three targets employing both regression and classification (Q)SAR models developed using the GUSAR platform. As a result, 6655 potential 3CLpro inhibitors and 3387 potential RdRp inhibitors with estimated IC_{50} values below 10 μM were identified. In the case of PLpro, the classification models yielded 6981 hits predicted to belong to the “active” class.

For RdRp, this stage represented the final step of the selection process. The top five predicted RdRp inhibitors demonstrated markedly high Pa-Pi values, suggesting a high probability of experimental confirmation. Nevertheless, these compounds exhibited strong structural similarity to approved antiviral agents, with four of the five being listed in the CAS antiviral database.

The compounds included in the RdRp training set were nucleotide analogues. Their proposed mechanism of inhibition involves incorporation into the growing RNA chain, thereby terminating RNA elongation by preventing the subsequent addition of nucleotides. The molecular docking programs employed in this study are not appropriate for accurately predicting binding poses or estimating the binding affinities of inhibitors that act through such a mechanism. Consequently, the molecular docking approach was not utilized at the final stage of compound selection for RdRp.

For compounds predicted to possess 3CLpro and PLpro inhibitory activity, additional molecular docking studies were conducted as described above. This analysis resulted in the identification of 45 potential 3CLpro inhibitors and 38 potential PLpro inhibitors. Notably, for these compounds, the computational predictions derived from similarity assessment and Tab. 2 summarizes the three compounds predicted to be the most probable inhibitors for each target.

In consideration of the requirements of the JEDI COVID-19 Challenge (10000 hits per target), the highest-scoring compounds described above were supplemented with additional compounds of lower scores, where such data were available.

2.2. Lessons Learned from Our Participation in the JEDI COVID-19 Challenge

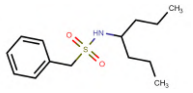
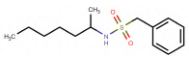
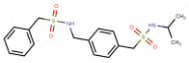
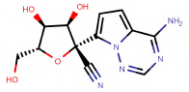
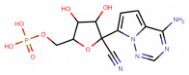
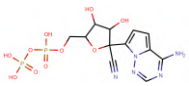
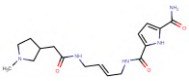
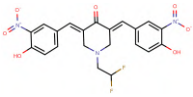
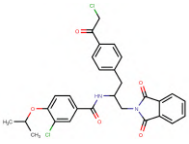
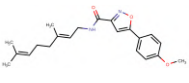
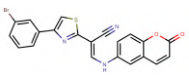
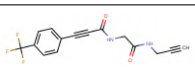
According to the expert evaluation conducted by the JEDI COVID-19 Challenge organizers, thirty-six compounds proposed by our research group were synthesized and assessed in anti-coronavirus bioassays. Among these, one compound exhibited inhibitory activity against PLpro, with an IC_{50} value in the micromolar range [67].

In accordance with the initial guidelines provided by the JEDI COVID-19 Challenge organizers, several approved drugs were also included among the compounds proposed as potential anti-coronavirus agents. However, their inhibitory activity was not examined within the framework of the JEDI COVID-19 Challenge, as multiple studies had already reported the results of screening approved drug libraries against SARS-CoV-2 targets [39, 41, 63, 74].

Several of our predictions concerning potential anti-coronavirus agents among approved drugs were subsequently corroborated by independent investigations. Specifically, inhibition of 3CLpro was demonstrated for narlaprevir, boceprevir, telaprevir [5, 44], carmofur, and disulfiram [46]; inhibition of RdRp for sofosbuvir and gemcitabine [80]; and inhibition of PLpro for dihydroquercetin and hesperetin [42].

The requirement established by the JEDI COVID-19 Challenge organizers to submit 10000 compounds predicted to interact with each of the three SARS-CoV-2 viral targets appears to have been overly stringent. Despite computational screening of approximately one billion synthesizable structures and the application of three independent in silico approaches, our study did not yield that number of hits supported by sufficient evidence. If other teams in the JEDI COVID-19 Challenge faced similar difficulties, it is plausible that the organizers received a large volume of low-confidence data, necessitating substantial additional effort for its further evaluation.

Table 2. The most probable hits for the analyzed targets

Transmembrane peptidase serine 2 (TMPRSS2)		
Name (Database)	Structure	Pa-Pi
ZINC001252905755 (ZINC)		0.315
Z355234742 (Enamine)		0.286
Z198103156 (Enamine)		0.280
RNA-dependent RNA polymerase		
Name (Database)	**Structure**	**Pa-Pi**
BRDWIEOJOWJCLU-LTGWCKQJSA-N (AMS)		0.977
1911578-74-9 (CAS antiviral DB)		0.977
1911578-77-2 (CAS antiviral DB)		0.973
3-chymotrypsin-like protease		
Name (Database)	Structure	Scoring Function
CHUUJOGSXZEWIU-NSCUHMNNSA-N (AMS)		−66.2 (Dock 6.5)
SPSIFTRUXBQBRF-YOENDLTHSA-N (AMS)		−8.4 (AutoDock Vina)
SXCFTBTXHZXEIN-NRFANRHFSA-N (AMS)		−8.6 (AutoDock Vina)
Papain-like proteinase		
Name (Database)	Structure	Scoring Function
NIKRPEWINGWQFH-FOWTUZBSSA-N (AMS)		−8.2 (AutoDock Vina)
DUJJXYLPLPJJQH-RVDMUPIBSA-N (AMS)		−9.7 (AutoDock Vina)
ORPOQLQFKDBKIH-UHFFFAOYSA-N (AMS)		−8.2 (AutoDock Vina)

Another plausible inference from this observation is that the chemical spaces of known antiviral agents and those of currently available synthesizable compounds differ substantially. This finding aligns with earlier work published in 2016 [43], which showed that antiviral compounds from ChEMBL cluster within specific regions of chemical space, with distinct antiviral classes occupying "privileged" zones on GTM (Generative Topographic Mapping) maps that diverge

from the more general drug-like chemical space. Subsequent studies similarly demonstrated that a curated and similarity expanded set of SARS-CoV-2 active compounds occupies a region of chemical space that extends well beyond that of a large commercially available coronavirus focused library (over 20000 molecules) and exhibits distinct scaffold distributions [7].

Currently, the clinically validated direct-acting small-molecule antivirals approved for the treatment of SARS-CoV-2 infection (as opposed to in vitro activity only) include remdesivir, molnupiravir, and the combination nirmatrelvir/ritonavir (Paxlovid). Remdesivir (GS 5734), originally developed by Gilead Sciences, is a nucleotide analogue initially designed for the treatment of severe RNA virus infections with pandemic potential, with an early focus on the Ebola virus [21]. Molnupiravir (EIDD 2801, MK 4482) was originally conceived as an orally bioavailable, broad-spectrum nucleoside analogue for the treatment of alphavirus and influenza virus infections, reflecting its early development as an anti influenza and anti respiratory RNA virus candidate [19]. The combination of nirmatrelvir with ritonavir (Paxlovid) was developed and authorized by Pfizer in record time; notably, nirmatrelvir was not synthesized de novo but rather derived from the earlier optimized SARS-CoV protease inhibitor PF 00835231 [35].

Therefore, it is not surprising that, despite the substantial efforts of 130 research teams comprising approximately 600 experts from leading institutions worldwide (about 45% from Europe, 30% from the Americas, 20% from Asia, and 5% from Africa) participating in the JEDI COVID-19 Challenge and screening billions of synthesizable compounds, only 27 weak inhibitors of SARS-CoV-2 targets were ultimately identified [36]. Clearly, the ambitious goal announced at the outset of the JEDI COVID-19 Challenge "To screen billions of molecules with blocking interactions relevant to SARS-CoV-2, and fast-track the route to a therapeutic treatment" was not achieved and, in all likelihood, could not have been achieved using this approach.

The initial plan was to commence virtual screening on May 4, 2020, and complete it by June 6, 2020. However, this deadline was extended several times, with the final submission date for reports set for July 17, 2020. Aggregation and evaluation of the submitted results were completed by November 18, 2020. In total, 1200 compounds were included in the final list, of which 1000 were selected for synthesis. By April 21, 2021, 878 compounds were synthesized and were tested in anti-SARS-CoV-2 assays by May 24, 2021. At that stage, it became evident that none of the identified inhibitors exhibited activity with an IC50 value of 100 nanomolar or better.

On July 7, 2022, Prof. Thomas Hermans, program manager of the "JEDI Billion Molecules Against COVID 19 Grand Challenge," submitted his Letter of Resignation, outlining the organizational challenges encountered during the project. It was subsequently decided to prepare a joint manuscript representing the collective efforts of the participating community. The manuscript was submitted to the Journal of the American Chemical Society on March 3, 2023, and rejected on April 11, 2023. The revised version was later accepted for publication in Molecular Informatics on October 13, 2023, and published in January 2024 [53].

Just as "One woman can give birth to a child in nine months, but nine women cannot give birth to a child in one month", the discovery of an effective anti-coronavirus drug cannot be hastened merely by increasing the number of researchers, even if they possess substantial expertise in the field.

The testing results for compounds selected by thirty research teams using three independent computational approaches indicated that the accuracy of virtual screening remained low – only a few percent ($27/878 \approx 0.03$). This outcome can be attributed primarily to the pronounced scarcity of reliable data at the initial stage of the project. At that time, all available compounds

with reported ("measured") activity were incorporated into the training sets for the development of (Q)SAR models. Subsequent analyses revealed, however, that some of these compounds lacked genuine biological activity. The inclusion of low-confidence data introduced substantial noise into the training sets, thereby diminishing the predictive performance of the models – a clear case of purgamentum init, exit purgamentum ("garbage in, garbage out").

Nevertheless, even under such conditions, despite the highly limited and noisy training data, the (Q)SAR models developed using machine learning and molecular modeling approaches enabled a substantial reduction in the human, temporal, material, and financial resources required for experimental investigations. Instead of synthesizing and biologically testing billions of molecules, only 878 compounds were synthesized and evaluated in biological assays, leading to the identification of 27 novel anti-coronavirus agents.

Conclusion

On May 5, 2023, the World Health Organization lifted the Public Health Emergency of International Concern (PHEIC) designation, signifying the transition from the acute phase of the COVID-19 pandemic to its long-term endemic management. Nevertheless, the consequences of SARS-CoV-2 infection – particularly long COVID – remain a significant health burden for many patients [69].

The urgent need for a prompt and coordinated response to the COVID-19 pandemic has driven an unprecedented acceleration of scientific and clinical research. The viral genome has been sequenced; diagnostic assays based on PCR and ELISA methodologies have been developed; fundamental mechanisms underlying viral pathogenesis have been elucidated; putative molecular targets have been identified; and experimental models have been established for the in vitro evaluation of potential antiviral agents. Ongoing clinical studies are focused on repurposing existing pharmacotherapies, assessing the safety and efficacy of candidate vaccines, and characterizing the distinctive features of patients responses to infection and therapeutic interventions. A substantial portion of these findings is disseminated almost immediately through online platforms of scientific journals and various specialized research databases.

Certain inconsistencies have been noted between experimental findings and clinical outcomes, highlighting the need for further methodological refinement and validation. Moreover, some studies disseminated through accelerated publication venues lack sufficient methodological rigor and therefore warrant more comprehensive evaluation.

The distinctiveness of the JEDI COVID-19 Challenge lies in its requirement to conduct integrative analyses of experimental and clinical data on SARS-CoV-2/COVID-19 in near real time, in parallel with the rapid emergence of new scientific evidence. Beginning with data on several dozen compounds that demonstrated inhibitory activity against viral infection in cell-based assays, we subsequently assembled training datasets and developed classification and regression (Q)SAR models. These models were then applied for large-scale virtual screening to identify compounds with prospective anti-coronavirus activity among more than one billion drug-like molecules.

Experimental validation of these selected hits is expected to significantly enrich the current knowledge base in this domain, considering that conventional training sets employed in (Q)SAR model development typically comprise only several thousand compounds, whereas the estimated size of the drug-like chemical space approaches approximately 10^{60} molecules [53]. We anticipate that the continued accumulation of experimental data and systematic exploration of the chemical

space will accelerate the discovery of novel therapeutic agents with enhanced safety and efficacy profiles for the treatment of COVID-19.

According to existing estimates, in addition to viruses, there exists a vast array of other potential biogenic threats worldwide, including up to 10^{12} distinct microorganisms, 66 fungal strains, and approximately 391000 plant species. Consequently, new diseases threatening human health may emerge as a result of hypersensitivity reactions to substances secreted by these biological species, as well as from infections or toxic effects they induce [69]. Furthermore, the potential for such biogenic threats has been convincingly demonstrated by Chinese researchers, who conducted systematic studies and identified eight novel pathogenic viruses in rodents inhabiting the tropical island of Hainan [47]. These viruses are considered highly likely to infect humans should they succeed in crossing the species barrier.

As evidenced by the COVID-19 pandemic, the only rapid response to an emerging biogenic threat is the application of the already available drugs – that is, their repurposing for new therapeutic indications. The consequences of previous biogenic threats, such as SARS, MERS, and others, were fortunately less severe; however, this circumstance led to a rapid decline in investments in related research. Consequently, humanity was insufficiently prepared for the COVID-19 pandemic. Considering the lessons learned from this crisis, it is essential to maintain and further develop the research initiatives established during this period, particularly within existing consortia, to continue the discovery of new antiviral agents and to establish robust theoretical, practical, and methodological frameworks for an effective response to future biogenic threats [8].

In view of these considerations, we are in full agreement with the assertion of former NIH Director Francis Collins, who stated: "Perhaps the most valuable lesson that COVID-19 has taught the research community – and hopefully society more broadly – is the importance of collective effort and continuous investment in basic and applied research." [14].

Acknowledgments

The study was conducted in the framework of the Program for Basic Research in the Russian Federation for a long-term period (20212030) (No. 122030100170-5).

References

1. Aldrich Market Select (AMS). `https://www.sigmaaldrich.com/chemistry/chemistry-services/aldrich-market-select.html`, accessed: 2020-07-15

2. Antiviral CAS dataset. `https://www.cas.org/covid-19-antiviral-compounds-dataset`, accessed: 2020-07-15

3. Anusevicius, K., Mickevicius, V., Stasevych, M., *et al.*: Design, synthesis, in vitro antimicrobial activity evaluation and computational studies of new N-(4-iodophenyl)-β-alanine derivatives. Res. Chem. Intermed. 41(10), 7517–7540 (2014). `https://doi.org/10.1007/s11164-014-1841-0`

4. AutoDock Vina. `http://vina.scripps.edu/`, accessed: 2020-07-15

5. Baker, J.D., Uhrich, R.L., Kraemer, G.C., *et al.*: A drug repurposing screen identifies hepatitis C antivirals as inhibitors of the SARS-CoV2 main protease. PLoS One 16(2), e0245962 (2021). https://doi.org/10.1371/journal.pone.0245962

6. Bender, A.: How similar are those molecules after all? Use two descriptors and you will have three different answers. Expert Opin. Drug Discov. 5(12), 1141–1151 (2010). https://doi.org/10.1517/17460441.2010.517832

7. Betow, J.Y., Turon, G., Metuge, C.S., *et al.*: The chemical space spanned by manually curated datasets of natural and synthetic compounds with activities against SARS-CoV-2. Mol. Inform. 44(1), e202400293 (2025). https://doi.org/10.1002/minf.202400293

8. Bobrowski, T., Melo-Filho, C.C., Korn, D., *et al.*: Learning from history: do not flatten the curve of antiviral research! Drug Discov. Today 25(9), 1604–1613 (2020). https://doi.org/10.1016/j.drudis.2020.07.008

9. Bojkova, D., McGreig, J.E., McLaughlin, K.M., *et al.*: SARS-CoV-2 and SARS-CoV differ in their cell tropism and drug sensitivity profiles. bioRxiv preprint (2020). https://doi.org/10.1101/2020.04.03.024257

10. Burov, Yu.V., Poroikov, V.V., Korolchenko, L.V.: National system for registration and biological testing of chemical compounds: facilities for new drugs search. Bulletin of the National Center for Biologically Active Compounds 1, 4–25 (1990).

11. ChEBI. https://www.ebi.ac.uk/chebi/, accessed: 2020-07-15

12. ChEMBL database. https://www.ebi.ac.uk/chembl, accessed: 2026-02-15

13. Chou, C.Y., Chien, C.H., Han, Y.S., *et al.*: Thiopurine analogues inhibit papain-like protease of severe acute respiratory syndrome coronavirus. Biochem. Pharmacol. 75(8), 1601–1609 (2008). https://doi.org/10.1016/j.bcp.2008.01.005

14. Collins, F., Adam, S., Colvis, C., *et al.*: The NIH-led research response to COVID-19. Science 379(6631), 441–444 (2023). https://doi.org/10.1126/science.adf5167

15. Cortellis Drug Discovery Intelligence. https://www.cortellis.com/drugdiscovery/, accessed: 2020-07-15

16. Dai, W., Zhang, B., Jiang, X.M., *et al.*: Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science 368(6497), 1331–1335 (2020). https://doi.org/10.1126/science.abb4489

17. De Meyer, S., Bojkova, D., Cinatl, J., *et al.*: Lack of antiviral activity of darunavir against SARS-CoV-2. Int. J. Infect. Dis. 97, 7–10 (2020). https://doi.org/10.1016/j.ijid.2020.05.085

18. Dimova, D., Bajorath, J.: Advances in activity cliff research. Mol. Inform. 35(5), 181–191 (2016). https://doi.org/10.1002/minf.201600023

19. Do, T.N.D., Abdelnabi, R., Boda, B., *et al.*: Remdesivir: The triple combination of REmdesivir (GS-441524), molnupiravir and ribavirin is highly efficient in inhibiting coronavirus replication in human nasal airway epithelial cell cultures and in a hamster infection model. Antiviral Res. 231, 105994 (2024). https://doi.org/10.1016/j.antiviral.2024

20. Druzhilovskiy, D.S., Stolbov, L.A., Savosina, P.I., *et al.*: Computational approaches to identify a hidden pharmacological potential in large chemical libraries. Supercomputing Frontiers and Innovations 7(3), 57–76 (2020). https://doi.org/10.14529/jsfi200306

21. Eastman, R.T., Roth, J.S., Brimacombe, K.R., *et al.*: Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of COVID-19. ACS Cent. Sci. 6(5), 672–683 (2020). https://doi.org/10.1021/acscentsci.0c00489

22. Ellinger, B., Bojkova, D., Zaliani, A., *et al.*: A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. Sci. Data 8(1), 70 (2021). https://doi.org/10.1038/s41597-021-00848-4

23. ENAMINE in-stock compounds. https://enamine.net/, accessed: 2026-02-15

24. Filimonov, D., Poroikov, V., Borodina, Yu., Gloriozova, T.: Chemical similarity assessment through multilevel neighborhoods of atoms: definition and comparison with the other descriptors. J. Chem. Inf. Comput. Sci. 39(4), 666–670 (1999). https://doi.org/10.1021/ci980335o

25. Filimonov, D.A., Zakharov, A.V., Lagunin, A.A., Poroikov, V.V.: QNA-based 'Star Track' QSAR approach. SAR QSAR Environ. Res. 20(7–8), 679–709 (2009). https://doi.org/10.1080/10629360903438370

26. Filimonov, D.A., Lagunin, A.A., Gloriozova, T.A., *et al.*: Prediction of the biological activity spectra of organic compounds using the PASS online web resource. Chem. Heterocycl. Comp. 50(3), 444–457 (2016). https://doi.org/10.1007/s10593-014-1496-1

27. Filimonov, D.A., Druzhilovskiy, D.S., Lagunin, A.A., *et al.*: Computer-aided prediction of biological activity spectra for chemical compounds: opportunities and limitations. Biomedical Chemistry: Research and Methods 1(1), e00004 (2018). https://doi.org/10.18097/bmcrm00004

28. Filimonov, D.A., Akimov, D.V., Poroikov, V.V.: Method of self-consistent regression in analysis of quantitative structure-property relationships of chemical compounds. Pharm. Chem. J. 38(1), 21–24 (2004). https://doi.org/10.1023/B:PHAC.0000027639.17115.5d

29. Fourches, D., Muratov, E., Tropsha, A.: Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. J. Chem. Inf. Model 50(7), 1189–1204 (2010). https://doi.org/10.1021/ci100176x

30. Fourches, D., Muratov, E., Tropsha, A.: Curation of chemogenomics data. Nat. Chem. Biol. 11(8), 535 (2015). https://doi.org/10.1038/nchembio.1881

31. Fourches, D., Muratov, E., Tropsha, A.: Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. J. Chem. Inf. Model 56(7), 1243–1252 (2016). https://doi.org/10.1021/acs.jcim.6b00129

32. Geronikaki, A., Druzhilovskiy, D., Zakharov, A., Poroikov, V.: Computer-aided predictions for medicinal chemistry via Internet. SAR QSAR Environ. Res. 19(1-2), 27–38 (2008). https://doi.org/10.1080/10629360701843649

33. GIAID. https://gisaid.org/, accessed: 2026-02-15

34. Ghosh, A.K., Takayama, J., Aubin, Y., *et al.*: Structure-based design, synthesis, and biological evaluation of a series of novel and reversible inhibitors for the severe acute respiratory syndrome-coronavirus papain-like protease. J. Med. Chem. 52(16), 5228–5340 (2009). https://doi.org/10.1021/jm900611t

35. Halford, B.: The path to Paxlovid. ACS Cent. Sci. 8(4), 405–407 (2022). https://doi.org/10.1021/acscentsci.2c00369

36. Hodgson, C.L., Broadley, T.: Long COVID – unravelling a complex condition. Lancet Respir. Med. 11(8), 667–668 (2023). https://doi.org/10.1016/S2213-2600(23)00232-1

37. Huang, L., Chen, Y., Xiao, J., *et al.*: Progress in the research and development of anti-COVID-19 drugs. Front. Public Health 8, 365 (2020). https://doi.org/10.3389/fpubh.2020.00365

38. IBS Natural Compounds Set. https://www.ibscreen.com/, accessed: 2020-07-15

39. Jeon, S., Ko, M., Lee, J., *et al.*: Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. Antimicrob. Agents Chemother. 64(7), e00819–20 (2020). https://doi.org/10.1128/AAC.00819-20

40. JEDI billion molecules against COVID-19 grand challenge. https://www.jedi.foundation/covid19challenge, accessed: 2026-02-15

41. Jin, Z., Du, X., Xu, Y., *et al.*: Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. Nature 582(7811), 289–293 (2020). https://doi.org/10.1038/s41586-020-2223-y

42. Kakhar Umar, A., Zothantluanga, J.H., Luckanagul, J.A., *et al.*: Structure-based computational screening of 470 natural quercetin derivatives for identification of SARS-CoV-2 Mpro inhibitor. PeerJ. 11, e14915 (2023). https://doi.org/10.7717/peerj

43. Klimenko, K., Marcou, G., Horvath, D., Varnek, A.: Chemical space mapping and structure-activity analysis of the ChEMBL antiviral compound set. J. Chem. Inf. Model 56(8), 1438–1454 (2016). https://doi.org/10.1021/acs.jcim.6b00192

44. Kneller, D.W., Galanie, S., Phillips, G., *et al.*: Malleability of the SARS-CoV-2 3CL Mpro active-site cavity facilitates binding of clinical antivirals. Structure 28(12), 1313–1320 (2020). https://doi.org/10.1016/j.str.2020.10.007

45. Kubinyi, H.: Chemical similarity and biological activities. J. Braz. Chem. Soc. 13(6), 717–726 (2002). https://doi.org/10.1590/S0103-50532002000600002

46. Kuzikov, M., Costanzi, E., Reinshagen, J., *et al.*: Identification of inhibitors of SARS-CoV-2 3CL-pro enzymatic activity using a small molecule in vitro repurposing screen. ACS Pharmacol. Transl. Sci. 4(3), 1096–1110 (2021). https://doi.org/10.1021/acsptsci.0c00216

47. Li, Y., Tang, C., Zhang, Y., *et al.*: Diversity and independent evolutionary profiling of rodent-borne viruses in Hainan, a tropical island of China. Virol. Sin. 8(5), 651–662 (2023). https://doi.org/10.1016/j.virs.2023.08.003

48. Listings of WHOs response to COVID-19. https://www.who.int/news/item/29-06-2020-covidtimeline, accessed: 2026-02-15

49. Ma, C., Hurst, B., Hu, Y., *et al.*: Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease. Cell Res. 30(8), 678–692 (2020). https://doi.org/10.1038/s41422-020-0356-z

50. Mansouri, K., Kleinstreuer, N., Abdelaziz, A.M., *et al.*: CoMPARA: Collaborative modeling project for androgen receptor activity. Environ Health Perspect. 128(2), 27002 (2020). https://doi.org/10.1289/EHP5580

51. Martinez, M.A.: Lack of effectiveness of repurposed drugs for COVID-19 treatment. Front. Immunol. 12, 635371 (2021). https://doi.org/10.3389/fimmu.2021

52. Murtazalieva, K.A., Druzhilovskiy, D.S., Goel, R.K., *et al.*: How good are publicly available web services that predict bioactivity profiles for drug repurposing? SAR QSAR Environ. Res. 28 (10), 843–862 (2017). https://doi.org/10.1080/1062936X.2017.1399448

53. Muratov, E.N., Bajorath, J., Sheridan, R.P., *et al.*: QSAR Without Borders. Chem. Soc. Rev. 49(11), 3525–3564 (2020). https://doi.org/10.1039/d0cs00098a

54. Natural Product Activity and Species Source (NPASS). https://bidd2.nus.edu.sg/NPASS/, accessed: 2020-07-15

55. Natural Products from Northern African Sources (NANPDB). https://african-compounds.org/nanpdb/, accessed: 2020-07-15

56. Nuclei of Bioassays, Ecophysiology and Biosynthesis of Natural Products Database (NuBBE). https://nubbe.iq.unesp.br/portal/nubbedb.html, accessed: 2020-07-15

57. Poroikov, V.V., Filimonov, D.A., Gloriozova, T.A., *et al.*: Computer-aided prediction of biological activity spectra for organic compounds: the possibilities and limitations. Russ. Chem. Bull. 68(12), 2143–2154 (2019). https://doi.org/10.1007/s11172-019-2683

58. Poroikov, V.V., Filimonov, D.A., Borodina, Yu.V., *et al.*: Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds. J. Chem. Inf. Comput. Sci. 40(6), 1349-01355 (2000). https://doi.org/10.1021/ci000383k

59. PostERA activity data. https://postera.ai/covid/activity_data, accessed: 2020-07-15

60. Protein Data Bank. https://www.rcsb.org/, accessed: 2026-02-15

61. Pruijssers, A.J., George, A.S., Schäfer, A., *et al.*: Remdesivir potently inhibits SARS-CoV-2 in human lung cells and chimeric SARS-CoV expressing the SARS-CoV-2 RNA polymerase in mice. Cell Rep. 32(3), 107940 (2020). https://doi.org/10.1016/j.celrep.2020.107940

62. Ratia, K., Pegan, S., Takayama, J., *et al.*: A noncovalent class of papain-like protease/deubiquitinase inhibitors blocks SARS virus replication. PNAS 105(42), 16119–16124 (2008). https://doi.org/10.1073/pnas.0805240105

63. Riva, L., Yuan, S., Yin, X., *et al.*: A large scale drug repositioning survey for SARS-CoV-2 antivirals. Nature 586(7827), 113–119 (2020). https://doi.org/10.1038/s41586-020-2577-1

64. Sachs, J.D., Karim, S.S.A., Aknin, L., *et al.*: The Lancet Commission on lessons for the future from the COVID-19 pandemic. Lancet 400(10359), 1224–1280 (2022). https://doi.org/10.1016/S0140-6736(22)01585-9

65. SAVI. https://cactus.nci.nih.gov/download/savi_download/, accessed: 2020-07-15

66. Sheahan, T.P., Sims, A.C., Zhou, S., *et al.*: An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 and multiple endemic, epidemic and bat coronavirus. bioRxiv preprint (2020). https://doi.org/10.1101/2020.03.19.997890

67. Schimunek, J., Seidl, P., Elez, K., *et al.*: A community effort in SARS-CoV-2 drug discovery. Mol. Inform. 43, e202300262 (2024). https://doi.org/10.1002/minf.202300262

68. Sielaff, F., Böttcher-Friebertshäuser, E., Meyer, D., *et al.*: Development of substrate analogue inhibitors for the human airway trypsin-like protease HAT. Bioorg. Med. Chem. Lett. 21(16), 4860–4864 (2011). https://doi.org/10.1016/j.bmcl.2011.06.03

69. Smith, C.I.E., Bergman, P., Hagey, D.W.: Estimating the number of diseases – the concept of rare, ultra-rare, and hyper-rare. iScience 25(8), 104698 (2022). https://doi.org/10.1016/j.isci.2022.104698

70. Stanford Coronavirus Antiviral Research Database. https://covdb.stanford.edu/, accessed: 2020-07-15

71. Stolbov, L.A., Druzhilovskiy, D.S., Filimonov, D.A., *et al.*: (Q)SAR models of HIV-1 proteins inhibition by drug-like compounds. Molecules 25(1), 87 (2019). https://doi.org/10.3390/molecules25010087

72. SWEETLEAD. https://simtk.org/projects/sweetlead, accessed: 2020-07-15

73. SYBYL-X Suite. https://www.g6g-softwaredirectory.com/bio/proteomics/structure-modeling/20710-Tripos-SYBYL-X-Suite.php, accessed: 2020-07-15

74. Touret, F., Gilles, M., Barral, K., *et al.*: In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. Sci. Rep. 10(1), 13093 (2020). https://doi.org/10.1038/s41598-020-70143-6

75. Universal Natural Products Database (UNPD). http://pkuxxj.pku.edu.cn, accessed: 2020-07-15

76. UCSF Dock. http://dock.compbio.ucsf.edu/, accessed: 2020-07-15

77. Wang, M., Cao, R., Zhang, L., *et al.*: Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 30(3), 269–271 (2020). https://doi.org/10.1038/s41422-020-0282-0

78. Wermuth, C.G.: Similarity in drugs: reflections on analogue design. Drug Discov. Today 11(7-8), 348–354 (2006). https://doi.org/10.1016/j.drudis.2006.02.006

79. World Wide Approved Drugs (WWAD) database. https://way2drug.com/wwad/, accessed: 2026-02-15

80. Yuan, C., Goonetilleke, E.C., Unarta, I.C., Huang, X.: Incorporation efficiency and inhibition mechanism of 2'-substituted nucleotide analogs against SARS-CoV-2 RNA-dependent RNA polymerase. Phys. Chem. Chem. Phys. 23(36), 20117–20128 (2021). https://doi.org/10.1039/d1cp03049c

81. Zakharov, A.V., Peach, M.L., Sitzmann, M., Nicklaus, M.C.: A new approach to radial basis function approximation and its application to QSAR. J. Chem. Inf. Model 54(3), 713-719 (2014). https://doi.org/10.1021/ci400704f

82. ZINC. https://zinc.docking.org/, accessed: 2020-07-15