Predicting inhibitors for SARS-CoV-2 RNA-dependent RNA polymerase using machine learning and virtual screening

Nazim Medzhidov and Shinya Yuki
Elix Inc.

1. Introduction

The global spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused the COVID-19 disease pandemic. SARS-CoV-2 has a positive-sense single-strand RNA genome [(+)ssRNA] and belongs to the genus *Betacoronavirus* of the *Coronaviridae* family within the *Nidovirales* order of viruses [1]. The COVID-19 disease is a potentially fatal respiratory disease characterized by atypical pneumonia [2]. The unavailability of specific medications to treat COVID-19 has led to drug repositioning efforts using various approaches [3–5], including computational analyses. In this study, we trained several machine learning algorithms and used them together with a molecular docking approach to screen for antiviral and anti-inflammatory drugs with potential activity against SARS-CoV-2 RNA-dependent RNA polymerase (RdRp), a key component of the virus replication machinery. Based on the ligand information of RdRp inhibitors, the machine learning models used in this study were able to identify candidates such as remdesivir, a molecule with documented activity against RdRp of the novel coronavirus.

2. Coronaviruses and SARS-CoV-2

2.1 Overview

Apart from infecting different animals, coronaviruses can cause respiratory infections in humans ranging in scale from mild to severe. In the past two decades, three highly pathogenic coronaviruses with fatal outcomes emerged in humans, namely severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002, Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012 [6] and the novel coronavirus SARS-CoV-2 in 2019. The latter caused coronavirus disease 2019 (COVID-19) with the symptoms of unusual pneumonia. Due to its high transmissibility the novel coronavirus quickly spread around the globe and surpassed both SARS and MERS in terms of the number of infected patients and fatal outcomes.

2.2 Emergence and Spread of SARS-CoV-2

In late December 2019, clusters of patients with pneumonia of unknown cause were reported by several health facilities in Wuhan, China [7].These patients showed symptoms of viral pneumonia, including fever, cough and chest discomfort, and in severe cases dyspnea and bilateral lung infiltration, similar to patients with SARS and MERS [7,8]. Among the first documented hospitalized patients, most cases were epidemiologically linked to a wet market located in downtown Wuhan, the Huanan Seafood Wholesale Market [9,10]. On 31st of December, Wuhan Municipal Health Commission notified the public of a pneumonia outbreak of unidentified cause and informed the World Health Organization (WHO) [11].

2.3 Genomic Characterization of SARS-CoV-2

SARS-CoV-2 was reported to be 79% and 50% genetically identical with SARS-CoV and MERS-CoV, respectively.[12]. Genome organization is similar to the other betacoronaviruses, and SARS-CoV-2 has six open reading frames (ORFs). From 5′ to 3′, replicase (ORF1a/ORF1b) is followed by spike (S), envelope (E), membrane (M), and nucleocapsid (N) ORFs. Moreover, accessory proteins are encoded in seven putative ORFs that can be found between the structural genes [1]. The length for most of the SARS-CoV-2 proteins shows similarity to their corresponding counterparts in the SARS-CoV. With the exception for the S gene, other structural genes in SARS-CoV-2 and SARS-CoV share more than 90% amino acid sequence identity [12]. At the 5′ end, replicase gene is found that covers two thirds of the genome and encodes a large polyprotein (pp1ab). Pp1ab protein is subsequently cleaved into 16 non-structural proteins that play role in viral replication. These non-structural proteins share strong amino acid sequence identity of more than 85% with SARS-CoV [1].

2.4 Why targeting SARS-CoV-2 RdRp?

RdRps are multi-domain proteins able to catalyze the RNA synthesis from RNA templates, and are responsible for the viral genome replication and transcription processes [13]. Conserved nature and involvement in the essential role of viral replication makes RdRps attractive for antiviral drug development.

Nidovirales order, to which the Coronavirus genus belongs, are characterized by a complex machinery dedicated to RNA synthesis, that is operated by non-structural proteins (nsps), being produced as cleavage products of the ORF1a and ORF1b viral polyproteins to facilitate virus replication and transcription [14].

Entrance of the virus into the host cells is crucial for the initiation of infection. SARS-CoV-2 utilizes the receptor-binding domain on the viral spike (S) protein to bind to the angiotensin converting enzyme 2 (ACE2) on host cells and enters the cells via endosomal pathway [15]. Once inside the cells, viral genomic RNA is released, serving as a template for translation of the viral proteins and for making copies of the viral genome. RNA dependent RNA polymerase (RdRp), also known as non-structural protein 12 (nsp12), is a crucial component of the viral replicase complex responsible for the production of genomic RNA for new virions [16].

Due to the central role of RdRps in the replication of RNA viruses, almost all RNA viruses have an RdRp encoded in their genome. Despite relatively poor sequence similarity, RdRps from different RNA viruses share structural similarity resembling a right hand with thumb, palm, and fingers subdomains, which is also similar to reverse transcriptases. Sequence analyses of RdRps from various viruses revealed conservation of key residues in the active sites as well as in the palm domains [13,17–19]. The structure of SARS-CoV-2 RdRp was resolved and reported. The general architecture of RdRps, presence of conserved motifs A-G, conservation of key catalytic amino acids as well as structural similarity to Hepatitis C virus (HCV) and poliovirus RdRp was confirmed [20]. RdRps are considered important therapeutic targets due to their crucial role in the viral replication cycle and absence of a counterpart in humans, which can reduce the risk of having undesired side effects during treatment.

3. Application of Artificial Intelligence in Drug Repositioning

3.1 Molecular Docking vs Machine Learning Approaches

The urgent need for effective therapeutic agents against SARS-CoV-2 has resulted in numerous studies focusing on identifying potential drug candidates. A significant amount of *in silico* drug discovery reports has been published[5,21–24] recently that propose various candidates for drug repurposing against different protein targets of SARS-CoV-2. Most of those studies utilized conventional molecular docking analyses for which the information on the 3D structure of the target and ligand is necessary. Once this information is available, docking simulation can be performed against a specified region on the target protein to predict binding energy between protein and the ligand. Obtaining a 3D structure for the target proteins is a challenging, expensive, and lengthy process. Moreover, good binding energy does not necessarily mean good inhibition by the ligand. To circumvent the challenges associated with the molecular docking approaches, use of machine learning algorithms has been increasingly gaining attention.

3.2 AI applications in SARS-CoV-2 drug repurposing

Recent technological developments in the applications of machine learning to drug discovery have shown that it is potentially possible to facilitate the conventional process and reduce the cost for the discovery of new drugs [25,26]. Several artificial intelligence approaches are recently being explored as an alternative that can help researchers find potential drug candidates in a relatively short period even in cases where the 3D structure information is not available.

UK-based BenevolentAI incorporated biomedical information obtained from the scientific literature into their AI knowledge graph in order to identify inhibitors of the host protein AAK1. As a result, they identified Baricitinib, which is used for treatment of rheumatoid arthritis. Baricitinib was predicted to inhibit viral infection by targeting clathrin-mediated endocytosis [4]. Moreover, in addition to its potential anti-viral activity, it is possible that the anti-inflammatory nature of bariticinib might be helpful for the inflammation observed in COVID-19 patients [27]. Beck et al. applied their Deep Learning-based Molecule Transformer-Drug Target Interaction (MT-DTI) model to predict SARS-CoV-2 protease and helicase enzyme inhibitors from the commercially available antiviral drugs. The model architecture uses simplified molecular-input line-entry system (SMILES) strings and amino acid sequences as input thus allowing to use target proteins for which confirmed 3D structure is not available [3]. Atomwise, which is based in US, aims to develop a new broad-spectrum antivirals by targeting highly conserved protein regions in SARS-CoV-2. Together with their 15 research partnerships, they screen millions of small molecules against these targets using their deep convolutional network AtomNet [28], to be further tested in the in vitro assays [29].

Structural similarities between RdRps of several viruses, conservation of key amino acids in the active site as well as identification of broad-spectrum anti-RdRp drug Remdesivir indicated to potential similarity patterns in the chemical structures of effective RdRp inhibitors. This led us to implement supervised machine learning algorithms for the identification of potential RdRp inhibitors.

4. Discovery of SARS-CoV-2 RdRp Inhibitors

4.1 Dataset Curation and Model Training

We established a dataset containing small molecules with the experimentally confirmed activity values against RNA dependent RNA polymerases of Hepatitis C Virus (HCV), Dengue virus, Poliovirus, and Influenza virus. Dataset was obtained from PubChem [30] and ChEMBL [31] bioassays. Entries with known experimental activity values (IC50/EC50) were selected and assigned binary activity labels based on the activity values to train classification models. The cutoff threshold of activity for training was set at 5μM. The final dataset included 1356 (656 inactive, 700 active) compounds with activity labels. An equal number of active and inactive compounds, amounting to

20**%** of the whole dataset, were randomly selected and used as a validation set. The remaining 80**%** was used as a training set. Molecules exerted inhibitory activity against hepatitis C virus (HCV), poliovirus, dengue virus, and influenza virus RdRps. By using these inhibitors to train machine learning models we expected our model to learn the chemical features of effective RdRp inhibitors. We then evaluated the ability of our models in identifying known pre-clinical and clinical RdRp inhibitors. Finally, we used our models to screen FDA approved antiviral and anti-inflammatory drugs to identify potential candidates with inhibitory activity against RdRps. We also performed molecular docking analysis of the antiviral and anti-inflammatory drugs against the SARS-CoV-2 RdRp protein. For our machine learning models, compounds were converted to molecular fingerprints using RdKit [32], which in turn were used as input features. We experimented both with circular (Morgan fingerprints) [33] and topological fingerprints. The topological fingerprints were computed by extracting all subgraphs of a compound with a minimum of 1 and a maximum of 7 number of bonds. To implement the models, we mainly relied on the scikit-learn library [34].

Several models achieved an area under the receiver operating characteristic curve (ROC-AUC) score of over 0.8, namely, the graph convolutional network, the message passing network, the random forest classifier (both fingerprint types), the ridge classifier (with circular fingerprints), the lasso classifier (with topological fingerprints), the 3 layered multilayer perceptron (with circular fingerprints), and the XGBoost classifier (with topological fingerprints). One of them, the random forest classifier on circular fingerprints even surpassed 0.9 in terms of ROC-AUC. In terms of accuracy, the best result of 84**%** was also observed with the random forest classifier.

4.2 Model Evaluation on a test set of pre-clinical RdRp inhibitors

To validate the results, we used 3 best models (based on ROC-AUC score) to run inference on the test set of known pre-clinical RdRp inhibitors. In addition to the ROC-AUC and the accuracy scores, we also report the percentage of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) cases. The result is described in Table 1.1.

Table 1.1: Performance of the best 3 models on the test set.

Model	AUROC	ACC	Confidence Interval (alpha=0.05)	TP	TN	FP	FN
GraphConv	0.700	0.700	[0.558, 0.842]	0.65	0.75	0.25	0.35
RandomForest (C)	0.725	0.725	[0.587, 0.863]	0.50	0.95	0.05	0.50
3-layer MLP (C)	0.625	0.625	[0.475, 0.775]	0.50	0.75	0.25	0.50

Abbreviations: AUROC, area under the receiver operating characteristic curve; ACC, accuracy; TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

Models like the Random Forest classifier were very good at detecting negative examples (a true negative rate of 95%), however, the number of detected positive cases was also affected and the model was able to detect half of the active molecules (true positive rate of 50%). Other models, like the Graph Convolutional Models, can detect more active molecules (true positive rate of 65%), but the true negative rate drops to 75%, and more false positives are detected. Furthermore, the correlation between the outputs of different models was not very high, therefore, we thought that an ensemble model might improve the overall performance.

A plain Support Vector Machine with an "RBF" kernel worked the best in our experiments. The model used the outputs of the 10 best models as input features. We first trained the individual models on the original training set. Then, the validation set was split into 2 equal subsets; one of them was used to train the ensemble model, while the second one was set aside as the validation set. The ensemble model slightly outperformed all individual models on the test set (Table 1.2).

Table 1.2: Ensemble model results.

Dataset	AUROC	ACC	Confidence Interval (alpha=0.05)	TP	TN	FP	FN
Validation	0.875	0.875	[0.819, 0.931]	0.871	0.879	0.129	0.121
Test	0.750	0.750	[0.616, 0.884]	0.600	0.900	0.100	0.400

Abbreviations: ROC-AUC, area under the receiver operating characteristic curve; ACC, accuracy; TP, true positives; TN, true negatives; FP, false positives; FN, false negatives.

Models were evaluated based on performance on the validation set and a test set of known pre-clinical RdRp inhibitors. Three best performing separate models and the best ensemble model were chosen for the inference analyses. Since only ligand information was used to train the models and no 3D structure of the target SARS-CoV-2 RdRp was used, we wanted to compare our approach to the conventional molecular docking approach, which is based on the 3D structure information of protein target and ligands. We performed virtual screening of the antiviral and anti-inflammatory datasets against the active site of SARS-CoV-2 RdRp (PDB ID: 6m71) using AutoDock Vina [35].

5. Combining machine learning and docking simulation results

5.1 Results on antiviral dataset

Our analyses identified several candidates from both antiviral and anti-inflammatory datasets. From the antiviral dataset our models were able to identify Remdesivir, a nucleoside analog confirmed to target SARS-CoV-2 RdRp that was approved by the US FDA for treatment of COVID-19 patients. Remdesivir was included in the test set as a positive control. Interestingly, baloxavir marboxil, TMC-310911 (ASC09), and umifenovir (Arbidol) identified by our models have been investigated in clinical trials for COVID-19. Clinical trial registration and identification numbers are ChiCTR2000029544 for baloxavir marboxil, NCT04261907 for ASC09, and NCT04350684 for umifenovir. Baloxavir marboxil acts on RdRp of Influenza virus [36], while TMC-310911 is a protease inhibitor developed against HIV-1 [37], and umifenovir is an anti-influenza drug that perturbs virus entry into the cells by targeting hemagglutinin (HA) glycoprotein [38].

In addition to these drug candidates, all four of our best performing models identified beclabuvir and asunaprevir as potential RdRp inhibitors from the *antiviral* dataset. Beclabuvir is a non-nucleoside inhibitor of HCV RdRp (NS5B) [39]. Similar to SARS-CoV-2, HCV is also a single-stranded enveloped positive-sense RNA virus. The active site of both SARS-CoV-2 and HCV RdRp show a degree of structural similarity and they both share the same conserved amino acids in the catalytic site [20]. Binding energy calculations of beclabuvir (-9.2 kcal/mol) towards SARS-CoV-2 RdRp in our experiments also suggest the inhibitory potential of this candidate. Asunaprevir, another anti-HCV drug, is known to target the protease of HCV. Interestingly, all of our best models identified asunaprevir as a potential anti-RdRp candidate, but the binding energy calculation using AutoDock Vina was -7.5 kcal/mol. Among other candidates both being predicted by at least two of our best

models and having relatively low binding energy towards SARS-CoV-2 RdRp were paritaprevir, faldaprevir, simeprevir, vedroprevir (HCV protease inhibitors), ledipasvir, odalasvir, and velpatasvir (HCV NS5A inhibitors) Please refer to the Table 1.3 for the complete list. Our models were trained only on RdRp inhibitors, however, several anti-HCV drugs targeting either protease or NS5A protein of the HCV were classified as potential RdRp inhibitors. Interestingly, those candidate molecules also had good binding energy predictions towards SARS-CoV-2 RdRp based on molecular docking analysis. Experimental validation is necessary to confirm whether these molecules act on RdRp or not.

Table 1.3: Antiviral drugs predicted to act on RdRps along with the binding energy values against SARS-CoV-2 RdRp (PDB ID 6m71) calculated using AutoDock Vina.

Compound	Predicted by # of models	Binding energy to SARS-CoV-2 RdRp (kcal/mol)
Beclabuvir	4	-9.2
Asunaprevir	4	-7.5
Paritaprevir	3	-10.5
Faldaprevir	3	-9.6
Odalasvir	3	-8.8
Simeprevir	3	-8.7
Vedroprevir	3	-8.6
Velpatasvir	3	-8.6
Telaprevir	3	-8.3
Dolutegravir	3	-8.0
Sofosbuvir	3	-6.9
Uprifosbuvir	3	-6.8
Entecavir	3	-6.6
Lobucavir	3	-6.6
Trifluridine	3	-6.3
Nevirapine	3	-6.1
Ledipasvir	2	-9.2
Ruzasvir	2	-8.1
Baloxavir marboxil	2	-8.0
TMC-310911(ASC09)	2	-7.9
Adafosbuvir	2	-7.8
Remdesivir	2	-7.5
Saquinavir	2	-7.2
Abacavir	2	-7.1
Maribavir	2	-7.1
Elvitegravir	2	-6.6
Vidarabine	2	-6.5
Efavirenz	2	-6.3
Valganciclovir	2	-6.2
Valomaciclovir	2	-6.2
Sorivudine	2	-6.1

Ibacitabine	2	-6.1
Idoxuridine	2	-5.9
Fialuridine	2	-5.9
Didanosine	2	-5.8
Umifenovir	2	-5.8

5.2 Results on anti-inflammatory dataset

Effective antiviral drugs can help reduce the viral load in the patients, however, they do not address virus-induced pneumonia directly. This pneumonia is a result of inflammation in the lungs caused by SARS-CoV-2 [40]. Thus COVID-19 patients who have developed pneumonia might need additional therapeutic intervention to suppress the inflammation in the lungs. The ability of our models to identify RdRp inhibitors from the antiviral set motivated us to run the inference analyses on a set of anti-inflammatory drugs. As in the analysis of the antiviral set, we focused on both RdRp inhibitor signature and binding energy predictions against SARS-CoV-2 RdRp. Analysis of the anti-inflammatory dataset revealed that all of our best models predicted betulinic acid and lupeol, both natural products, to possess anti-RdRp activity. Lifitegrast, antrafenine, ursolic acid, dexamethasone acetate, prednisolone phosphate were other candidates predicted by at least two of our models and are also predicted to bind to the active site of SARS-CoV-2 RdRp with binding energy in the range between -7.5 to -9.5 kcal/mol (Table 1.4). Interestingly, both betulinic acid and ursolic acid are pentacyclic triterpenoids with documented antiviral activity against HIV [41].

Table 1.4: Anti-inflammatory drugs predicted to act on RdRps along with the binding energy values against SARS-CoV-2 RdRp (PDB ID 6m71) calculated using AutoDock Vina.

Compound	**Predicted by # of models**	**Binding affinity to SARS-CoV-2 RdRp (kcal/mol)**
Betulinic Acid	4	-7.4
Lupeol	4	-7.2
Lifitegrast	3	-9.5
Antrafenine	3	-8.7
Ursolic acid	3	-8.0
Floctafenine	3	-7.1
Cimicoxib	3	-7.0
Acemetacin	3	-6.8
Morniflumate	3	-6.8
Loteprednol	3	-6.8
Polmacoxib	3	-6.8
Andrographolide	3	-6.7
Dexamethasone acetate	2	-7.6

Prednisolone phosphate	2	-7.5
Cortisone acetate	2	-7.3
Mometasone furoate	2	-7.3
Prednicarbate	2	-7.1
Deflazacort	2	-7.1
Clobetasone	2	-6.8
Rimexolone	2	-6.8
Robenacoxib	2	-6.8
Hydrocortisone probutate	2	-6.8
Mometasone	2	-6.6
Diflunisal	2	-6.5
Lumiracoxib	2	-6.5
Etoricoxib	2	-6.5
Clobetasol	2	-6.5
Apremilast	2	-6.5
Bisindolylmaleimide I	2	-6.5
Talniflumate	2	-6.3
NS-398	2	-6.2
Firocoxib	2	-5.6
Dimethyl sulfone	2	-3.0

6. Concluding Remarks

The social and economic effects of the global COVID-19 pandemic continue to severely affect the lives of millions of people worldwide. The presence of animal reservoirs for various potentially deadly viruses in the wildlife and the ease of transportation in the current modern world can facilitate the global transmission of such infectious agents and make us even more susceptible to the emergence of new pandemics in the future. Despite the enormous amount of effort put into discovering effective therapeutic agents, so far there is still a lack of specific drugs that could help to treat COVID-19 patients and thousands of deaths continue to be reported daily due to this disease. This indicates that the current methods of drug discovery need to be carefully re-evaluated for the possibilities of speeding up the overall process. In this study, we implemented machine learning algorithms, which enabled us to rediscover remdesivir, known to have inhibitory activity against SARS-CoV-2 RdRp, as well as to identify potential drug candidates that can be evaluated for this problem. We are optimistic that further advancements in the application of AI in drug discovery will enable us to facilitate the drug discovery process. This is of outmost importance considering the cases like the current pandemic, where urgent solutions are necessary and traditional drug discovery process timeline is suboptimal, to say the least.

7. Acknowledgments

Authors would like to thank all healthcare workers, who risk their lives taking care of COVID-19 patients.

References

[1] Chan, J. F.-W. et al.: Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerging Microbes & Infections*, **9**, 221–236 (2020). http://doi:10.1080/22221751.2020.1719902.

[2] Zhou, P. et al.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579**, 270–273 (2020). http://doi:10.1038/s41586-020-2012-7.

[3] Beck, B. R. et al.: Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computational and Structural Biotechnology Journal*, **18**, 784–790 (2020). http://doi:10.1016/j.csbj.2020.03.025.

[4] Richardson, P. et al.: Baricitinib as potential treatment for 2019-nCoV acute respiratory disease. *The Lancet*, **395**, e30–e31 (2020). http://doi:10.1016/s0140-6736(20)30304-4.

[5] Kandeel, M.; Al-Nazawi, M.: Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sciences*, **251**, 117627 (2020). http://doi:10.1016/j.lfs.2020.117627.

[6] Cui, J. et al.: Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, **17**, 181–192 (2019). http://doi:10.1038/s41579-018-0118-9.

[7] Zhu, N. et al.: A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, **382**, 727–733 (2020). http://doi:10.1056/nejmoa2001017.

[8] Gralinski, L. E.; Menachery, V. D.: Return of the coronavirus: 2019-nCoV. *Viruses*, **12**, 135 (2020). http://doi:10.3390/v12020135.

[9] Deng, S.-Q.; Peng, H.-J.: Characteristics of and public health responses to the coronavirus disease 2019 outbreak in China. *Journal of Clinical Medicine*, **9**, 575 (2020). http://doi:10.3390/jcm9020575.

[10] Jiang, S. et al.: An emerging coronavirus causing pneumonia outbreak in Wuhan, China: calling for developing therapeutic and prophylactic strategies. *Emerging Microbes & Infections*, **9**, 275–277 (2020). http://doi:10.1080/22221751.2020.1723441.

[11] Wu, Z.; McGoogan, J. M.: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China. *JAMA*, **323**, 1239–1242 (2020). http://doi:10.1001/jama.2020.2648.

[12] Lu, R. et al.: Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, **395**, 565–574 (2020). http://doi:10.1016/s0140-6736(20)30251-8.

[13] Venkataraman, S. et al.: RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses*, **10**, 76 (2018). http://doi:10.3390/v10020076.

[14] Snijder, E. J. et al.: The nonstructural proteins directing coronavirus RNA synthesis and processing. *Advances in Virus Research*, **96**, 59–126 (2016). http://doi:10.1016/bs.aivir.2016.08.008.

[15] Lan, J. et al.: Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, **581**, 215–220 (2020). http://doi:10.1038/s41586-020-2180-5.

[16] Masters, P. S.: The molecular biology of coronaviruses. *Advances in Virus Research*, **66**, 193–292 (2006). http://doi:10.1016/s0065-3527(06)66005-3.

[17] Bruenn, J. A.: A structural and primary sequence comparison of the viral RNA-dependent RNA polymerases. *Nucleic Acids Research*, **31**, 1821–1829 (2003). http://doi:10.1093/nar/gkg277.

[18] Xu, X. et al.: Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Research*, **31**, 7117–7130 (2003). http://doi:10.1093/nar/gkg916.

[19] Černý, J. et al.: Evolution of tertiary structure of viral RNA dependent polymerases. *PLoS ONE*, **9**, e96070 (2014). http://doi:10.1371/journal.pone.0096070.

[20] Gao, Y. et al.: Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*, **368**, 779–782 (2020). http://doi:10.1126/science.abb7498.

[21] Hage-Melim, L. I. da S. et al.: Virtual screening, ADME/Tox predictions and the drug repurposing concept for future use of old drugs against the COVID-19. *Life Sciences*, **256**, 117963 (2020). http://doi:10.1016/j.lfs.2020.117963.

[22] Quimque, M. T. J. et al.: Virtual screening-driven drug discovery of SARS-CoV2 enzyme inhibitors targeting viral attachment, replication, post-translational modification and host immunity evasion infection mechanisms. *Journal of Biomolecular Structure and Dynamics*, **0**, 1–18 (2020). http://doi:10.1080/07391102.2020.1776639.

[23] Naik, B. et al.: High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID-19 treatment. *International Journal of Biological Macromolecules*, **160**, 1–17 (2020). http://doi:10.1016/j.ijbiomac.2020.05.184.

[24] Jiménez-Alberto, A. et al.: Virtual screening of approved drugs as potential SARS-CoV-2 main protease inhibitors. *Computational Biology and Chemistry*, **88**, 107325 (2020). http://doi:10.1016/j.compbiolchem.2020.107325.

[25] Zhavoronkov, A. et al.: Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nature Biotechnology*, **37**, 1038–1040 (2019). http://doi:10.1038/s41587-019-0224-x.

[26] Fleming, N.: How artificial intelligence is changing drug discovery. *Nature*, **557**, S55–S57 (2018). http://doi:10.1038/d41586-018-05267-x.

[27] Stebbing, J. et al.: COVID-19: combining antiviral and anti-inflammatory treatments. *The Lancet Infectious Diseases*, **20**, 400–402 (2020). http://doi:10.1016/s1473-3099(20)30132-8.

[28] Wallach, I. et al.: AtomNet: A deep convolutional neural network for bioactivity prediction in structure-based drug discovery. *arXiv*, (2015).

[29] https://www.businesswire.com/news/home/20200521005238/en/Atomwise-Partners-with-Global-Research-Teams-to-Pursue-Broad-Spectrum-Treatments-Against-COVID-19-and-Future-Coronavirus-Outbreaks

[30] Kim, S. et al.: PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research*, **47**, D1102–D1109 (2018). http://doi:10.1093/nar/gky1033.

[31] Gaulton, A. et al.: ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, **40**, D1100–D1107 (2012). http://doi:10.1093/nar/gkr777.

[32] Landrum, G.: RDKit: Open-source cheminformatics software. *null*, (2016).

[33] Rogers, D.; Hahn, M.: Extended-connectivity fingerprints. *Journal of Chemical Information and Modeling*, **50**, 742–754 (2010). http://doi:10.1021/ci100050t.

[34] Pedregosa, F. et al.: Scikit-learn: machine learning in python. *J. Mach. Learn. Res.*, **12**, 2825–2830 (2011).

[35] Trott, O.; Olson, A. J.: AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, **31**, 455–461 (2010). http://doi:10.1002/jcc.21334.

[36] Giacchello, I. et al.: Insights into RNA-dependent RNA polymerase inhibitors as anti-influenza virus agents. *Current Medicinal Chemistry*, **27** (2020). http://doi:10.2174/0929867327666200114115632.

[37] Dierynck, I. et al.: TMC310911, A novel human immunodeficiency virus type 1 protease inhibitor, shows in vitro an improved resistance profile and higher genetic barrier to resistance compared with current protease inhibitors. *Antimicrobial Agents and Chemotherapy*, **55**, 5723–5731 (2011). http://doi:10.1128/aac.00748-11.

[38] Gasparini, R. et al.: Compounds with anti-influenza activity: present and future of strategies for the optimal treatment and management of influenza. Part I: Influenza life-cycle and currently available drugs. *Journal of preventive medicine and hygiene*, **55**, 69–85 (2014).

[39] Gentles, R. G.: Discovery of Beclabuvir: A potent allosteric inhibitor of the hepatitis C virus polymerase. *Topics in Medicinal Chemistry*, , 193–228 (2019). http://doi:10.1007/7355_2018_38.

[40] Tay, M. Z. et al.: The trinity of COVID-19: immunity, inflammation and intervention. *Nature Reviews Immunology*, **20**, 363–374 (2020). http://doi:10.1038/s41577-020-0311-8.

[41] Xiao, S. et al.: Recent progress in the antiviral activity and mechanism study of pentacyclic triterpenoids and their derivatives. *Medicinal Research Reviews*, **38**, 951–976 (2018). http://doi:10.1002/med.21484.