



ELSEVIER

journal homepage: www.elsevier.com/locate/csbj

Communications

Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model

Bo Ram Beck^a, Bonggun Shin^{a,b}, Yoonjung Choi^a, Sungsoo Park^{a,*}, Keunsoo Kang^{c,*}^aDeargen, Inc., Daejeon, Republic of Korea^bDepartment of Computer Science, Emory University, Atlanta, GA, United States^cDepartment of Microbiology, College of Science & Technology, Dankook University, Cheonan, Republic of Korea

ARTICLE INFO

Article history:

Received 22 February 2020

Received in revised form 23 March 2020

Accepted 25 March 2020

Available online 30 March 2020

Keywords:

COVID-19

SARS-CoV-2

Coronavirus

MT-DTI

Deep learning

Drug repurposing

Atazanavir

ABSTRACT

The infection of a novel coronavirus found in Wuhan of China (SARS-CoV-2) is rapidly spreading, and the incidence rate is increasing worldwide. Due to the lack of effective treatment options for SARS-CoV-2, various strategies are being tested in China, including drug repurposing. In this study, we used our pre-trained deep learning-based drug-target interaction model called Molecule Transformer-Drug Target Interaction (MT-DTI) to identify commercially available drugs that could act on viral proteins of SARS-CoV-2. The result showed that atazanavir, an antiretroviral medication used to treat and prevent the human immunodeficiency virus (HIV), is the best chemical compound, showing an inhibitory potency with K_d of 94.94 nM against the SARS-CoV-2 3C-like proteinase, followed by remdesivir (113.13 nM), efavirenz (199.17 nM), ritonavir (204.05 nM), and dolutegravir (336.91 nM). Interestingly, lopinavir, ritonavir, and darunavir are all designed to target viral proteinases. However, in our prediction, they may also bind to the replication complex components of SARS-CoV-2 with an inhibitory potency with $K_d < 1000$ nM. In addition, we also found that several antiviral agents, such as Kaletra (lopinavir/ritonavir), could be used for the treatment of SARS-CoV-2. Overall, we suggest that the list of antiviral drugs identified by the MT-DTI model should be considered, when establishing effective treatment strategies for SARS-CoV-2.

© 2020 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Coronaviruses (CoVs), belonging to the family *Coronaviridae*, are positive-sense enveloped RNA viruses and cause infections in birds, mammals, and humans [1–3]. The family includes four genera, such as *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus* [4]. Two infamous infectious coronaviruses in the genus *Betacoronavirus* are severe acute respiratory syndrome coronavirus (SARS-CoV) [5] and Middle East respiratory syndrome coronavirus (MERS-CoV) [6], which have infected more than 10,000 people around the world in the past two decades. Unfortunately, the incidence was accompanied by high mortality rates (9.6% for SARS-CoV and 34.4% for MERS-CoV), indicating that there is an urgent need for effective treatment at the beginning of the

outbreak to prevent the spread [7,8]. However, this cannot be achieved with current drug development or an application system, taking several years for newly developed drugs to come to the market. Unexpectedly, the world is facing the same situation as the previous outbreak due to a recent epidemic of atypical pneumonia (designated as coronavirus disease 2019; COVID-19) caused by a novel coronavirus (severe acute respiratory syndrome coronavirus 2; SARS-CoV-2) in Wuhan, China [5,9].

SARS-CoV-2, which belongs to *Betacoronavirus*, contains a positive-sense single-stranded RNA [(+)ssRNA] genome (29,903 bp) and contains genes encoding 3C-like proteinase, RNA-dependent RNA polymerase (RdRp), 2'-O-ribose methyltransferase, spike protein, envelope protein, nucleocapsid phosphoprotein, and several unknown proteins, according to the genome sequencing data of SARS-CoV-2 (<https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/>). Typical clinical symptoms of COVID-19 are fever, dry cough, and fatigue within 3–7 days of latency on average after infection. This is relatively slower than

* Corresponding authors.

E-mail addresses: sspark@deargen.me (S. Park), kang1204@dankook.ac.kr (K. Kang).

severe acute respiratory syndrome (SARS), which was caused by SARS-CoV [10]. During the life cycle of coronaviruses, the virus replicates via the following processes after entering the host cell: 1) translation of genomic RNA (gRNA), 2) proteolysis of the translated polyprotein with viral 3C-like proteinase, 3) replication of gRNA with the viral replication complex that consists of RNA-dependent RNA polymerase (RdRp), helicase, 3'-to-5' exonuclease, endoRNase, and 2'-O-ribose methyltransferase, and 4) assembly of viral components [11]. These replication-associated proteins are the primary targets of post-entry treatment drugs to suppress viral replication. Although much intensive effort is being made worldwide to develop drugs or vaccines for SARS-CoV-2, patients currently suffering from COVID-19 cannot expect benefits from them due to the slow development process of novel drugs or vaccines.

Thus, a rapid drug application strategy that can be immediately applied to the patient is necessary. Currently, the only way to address this matter is to repurpose commercially available drugs for the pathogen in so-called “drug-repurposing”. However, in theory, artificial intelligence (AI)-based architectures must be taken into account in order to accurately predict drug-target interactions (DTIs). This is because of the enormous amount of complex information (e.g. hydrophobic interactions, ionic interactions, hydrogen bonding, and/or van der Waals forces) between molecules. To this end, we previously developed a deep learning-based drug-target interaction prediction model, called Molecule Transformer-Drug Target Interaction (MT-DTI) [12].

In this study, we applied our pre-trained MT-DTI model to identify commercially available antiviral drugs that could potentially disrupt SARS-CoV-2's viral components, such as proteinase, RNA-dependent RNA polymerase, and/or helicase. Since the model utilizes simplified molecular-input line-entry system (SMILES) strings and amino acid (AA) sequences, which are 1D string inputs, it is possible to quickly apply target proteins that do not have experimentally confirmed 3D crystal structures, such as viral proteins of SARS-CoV-2. We share a list of top commercially available antiviral drugs that could potentially hinder the multiplication cycle of SARS-CoV-2 with the hope that effective drugs can be developed based on these AI-proposed drug candidates and act against SARS-CoV-2.

2. Materials & methods

2.1. Amino acid sequences used in this study

Amino acid sequences of 3C-like proteinase (accession YP_009725301.1), RNA-dependent RNA polymerase (accession YP_009725307.1), helicase (accession YP_009725308.1), 3'-to-5' exonuclease (accession YP_009725309.1), endoRNase (accession YP_009725310.1), and 2'-O-ribose methyltransferase (accession YP_009725311.1) of the SARS-CoV-2 replication complex were extracted from the SARS-CoV-2 whole genome sequence (accession NC_045512.2), from the National Center for Biotechnology Information (NCBI) database.

2.2. Prediction of drug-target interactions using binding affinity scores

Molecule transformer-drug target interaction (MT-DTI) was used to predict binding affinity values between commercially available antiviral drugs and target proteins. MT-DTI is based on the self-attention mechanism that showed remarkable success in natural language process (NLP) literature. MT-DTI is inspired by the idea that for a chemist, understanding a molecule sequence is analogous to understanding a language. To apply the NLP model to drug-target interaction (DTI) tasks, MT-DTI is pre-trained with

‘chemical language’ (represented as SMILES) of approximately 1,000,000,000 compounds. Similar to the NLP model, which successfully extracts complex patterns from word sequences, MT-DTI successfully finds useful information in DTI tasks. Therefore, it shows the best performance and most robust results in diverse DTI datasets according to a previous study [12].

To train the model, the Drug Target Common (DTC) database [13] and BindingDB [14] database were manually curated and combined. Three types of efficacy value, K_i , K_d , and IC_{50} were integrated by a consistency-score-based averaging algorithm [15] to make the Pearson correlation score over 0.9 in terms of K_i , K_d , and IC_{50} . Since the BindingDB database includes a wide variety of species and target proteins, the MT-DTI model has the potential power to predict interactions between antiviral drugs and SARS-CoV-2 proteins.

After the MT-DTI prediction, the raw prediction results were screened for antiviral drugs that are FDA approved, target viral proteins, and have a K_d value < 1000 nM. SMILES containing salt forms were excluded from the final results as the prediction is focused to pairs of a single molecule and the target protein. In addition, remdesivir was also incorporated in the analysis as its therapeutic potential to COVID-19 is recently suggested by Wang et al. [16] and Gilead Sciences announcements (<https://www.gilead.com/purpose/advancing-global-health/covid-19>).

2.3. Prediction of drug-target interactions using AutoDock Vina

AutoDock Vina (version 1.1.2), which is a molecular docking and virtual screening application [17], was used to predict binding affinities (kcal/mol) between 3C-like proteinase of SARS-CoV-2 and 3,410 FDA-approved drugs. SMILES of 3,410 FDA-approved drugs were converted to the PDBQT format using Open Babel (version 2.3.2) [18] with the following options: --gen3d and -p 7.4. The hydrogens were added to the 3C-like proteinase model using MGLTools (version 1.5.6) [19]. Then, binding affinities between the protein and FDA-approved drugs were calculated using AutoDock Vina. The exhaustiveness parameter was set to 10.

3. Results

To identify potent FDA-approved drugs that may inhibit the functions of SARS-CoV-2's core proteins, we used the MT-DTI deep learning-based model, which can accurately predict binding affinities based on chemical sequences (SMILES) and amino acid sequences (FASTA) of a target protein, without their structural information [12]. This deep learning-based approach is particularly useful, since it does not require protein structural information, which can be a bottleneck for identifying drugs targeted for uncharacterized proteins with traditional three-dimensional (3D) structure-based docking approaches [20]. Nevertheless, MT-DTI showed the best performance [12] when compared to a deep learning-based (DeepDTA) approach [21] and two traditional machine learning-based algorithms SimBoost [22], and KronRLS [23], with the KIBA [24] and DAVIS [25] data sets. Taking advantage of this sequence-based drug-target affinity prediction approach, binding affinities of 3,410 FDA-approved drugs against 3C-like proteinase, RdRp, helicase, 3'-to-5' exonuclease, endoRNase, and 2'-O-ribose methyltransferase of SARS-CoV-2 were predicted. To confirm the performance of MT-DTI at least *in silico*, we compared the binding affinities of 3,410 FDA-approved drugs predicted by MT-DTI to those estimated by AutoDock Vina (a widely used 3D structure-based docking algorithm). It was possible since the 3D structure of the 3C-like proteinase protein was recently unveiled by the X-ray crystallography (PDBID 6LU7) [26]. Significant negative correlations, meaning that the results of both algorithms

showed moderate similarities (higher is better for MT-DTI, whereas lower is better for AutoDock Vina) were observed in both the antiviral drug dataset ($R = -0.34$, and p -value = 0.0071) and the FDA-approved drug dataset ($R = -0.32$, and p -value < $2.2e-16$) (Fig. 1). While it is not possible to determine which algorithm is more reliable without various experimental evaluations, a previous study showed that the MT-DTI model is one of the best deep learning-based models that can predict the binding affinity between a given protein and compound [12]. Therefore, we further applied the MT-DTI model to repurpose those FDA-approved drugs that have the potential to inhibit key proteins of SARS-CoV-2.

The SARS-CoV-2 3C-like proteinase was predicted to bind with atazanavir (K_d 94.94 nM), followed by remdesivir, efavirenz, ritonavir, and other antiviral drugs that have a predicted affinity of $K_d > 100$ nM potency (Table 1). No other protease inhibitor antiviral drug was found in the $K_d < 1000$ nM range. Although there is no real-world evidence about whether these drugs will act as predicted against COVID-19 yet, some case studies have been identified. For example, a docking study of lopinavir along with other HIV proteinase inhibitors of the CoV proteinase (PDBID 1UK3) suggests atazanavir and ritonavir, which are listed in the present prediction results, may inhibit the CoV proteinase in line with the inhibitory potency of lopinavir [27]. According to the prediction, viral proteinase-targeting drugs were predicted to act more favorably on the viral replication process than viral proteinase through

the DTI model (Tables 2–6). The results include antiviral drugs other than proteinase inhibitors, such as guanosine analogues (e.g., acyclovir, ganciclovir, and penciclovir), reverse transcriptase inhibitors, and integrase inhibitors.

Among the prediction results, atazanavir was predicted to have a potential binding affinity to bind to RNA-dependent RNA polymerase (K_d 21.83 nM), helicase (K_d 25.92 nM), 3'-to-5' exonuclease (K_d 82.36 nM), 2'-O-ribose methyltransferase (K_d of 390.67 nM), and endoRNase (K_d 50.32 nM), which suggests that all subunits of the COVID-19 replication complex may be inhibited simultaneously by atazanavir (Tables 2–6). Also, ganciclovir was predicted to bind to three subunits of the replication complex of the COVID-19: RNA-dependent RNA polymerase (K_d 11.91 nM), 3'-to-5' exonuclease (K_d 56.29 nM), and RNA helicase (K_d 108.21 nM). Lopinavir and ritonavir, active materials of AbbVie's Kaletra, both were predicted to have a potential affinity to COVID-19 helicase (Table 3) and are suggested as potential MERS therapeutics [28]. Recently, approximately \$2 million worth of Kaletra doses were donated to China [29], and a previous clinical study of SARS by Chu et al. [30] may support this decision [30]. Another anti-HIV drug, Prezcoibix of Johnson & Johnson, which consists of darunavir and cobicistat, was to be sent to China [29], and darunavir is also predicted to have a K_d of 90.38 nM against COVID-19's helicase (Table 3). However, there was no current supporting literature found for darunavir to be used as a CoV therapeutic. Although remdesivir is not a

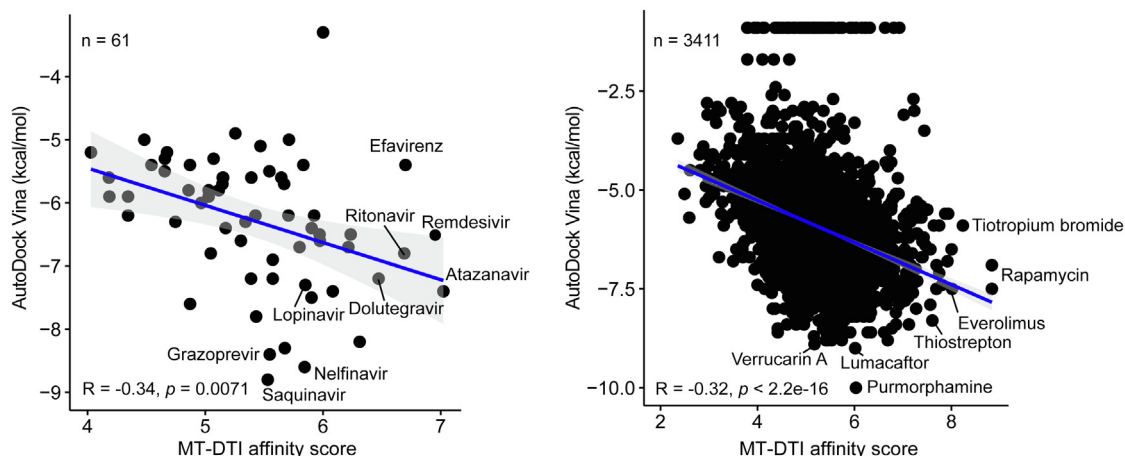


Fig. 1. Comparison of MT-DTI and AutoDock Vina results. 60 known FDA-approved antiviral drugs (left) and 3410 FDA-approved drugs (right) were evaluated by means of the MT-DTI deep learning-based affinity score (higher is better), and AutoDock Vina docking score (lower is better). Remdesivir, which is not an FDA-approved drug, but regarded as a promising antiviral drug for SARS-CoV-2, was included in this analysis.

Table 1
Drug-target interaction (DTI) prediction results of FDA approved antiviral drugs available on markets against a novel coronavirus (SARS-CoV-2, NCBI reference sequence NC_045512.2) 3C-like proteinase (accession YP_009725301.1). * indicates isomeric form SMILES.

Small molecules	SMILES used in the DTI prediction	K_d in nM	MT-DTI Rank out of 3411
Atazanavir	<chem>COC(=O)NC(C(=O)NC(Cc1cccc1)C(O)CN(Cc1ccc(-c2cccn2)cc1)NC(=O)C(NC(=O)OC)C(C)C(C)C(C)C</chem>	94.94	66
Remdesivir*	<chem>CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@H]1[C@H]([C@H]([C@H](O1)C#N)C2 = CC = C3N2N = CN = C3N)O)O</chem> <chem>OC4 = CC = CC = C4</chem>	113.13	70
Efavirenz*	<chem>O = C1Nc2ccc(Cl)cc2[C@@](C#CC2CC2)(C(F)(F)F)O1</chem>	199.17	116
Ritonavir	<chem>CC(C)c1nc(CN(C)C(=O)NC(C(=O)NC(Cc2ccccc2)CC(O)C(Cc2ccccc2)NC(=O)OCc2cncs2)C(C)C)cs1</chem>	204.05	119
Dolutegravir	<chem>CC1CCOC2Cn3cc(C(=O)NCc4ccc(F)cc4F)c(=O)c(O)c3C(=O)N12</chem>	336.91	162
Asunaprevir	<chem>C = CC1CC1(NC(=O)J1CC(Oc2ncc(OC)c3ccc(Cl)cc23)CN1C(=O)C(NC(=O)OC(C)C(C)C(C)C(=O)NS(=O)(=O)C1CC1</chem>	581.77	270
Ritonavir*	<chem>CC(C)c1nc(CN(C)C(=O)N[C@H](C(=O)N)[C@H](Cc2ccccc2)C[C@H](O)[C@H](Cc2ccccc2)NC(=O)OCc2cncs2)C(C)C)cs1</chem>	609.02	283
Simeprevir*	<chem>COc1ccc2c(O)[C@H]3CC4C(=O)N(C)CCCC/C = C\1[C@H]5[C@H]5(C(=O)NS(=O)(=O)C5CC5)NC(=O)[C@H]4C3)cc(-c3nc(C(C)C)cs3)nc2c1C</chem>	826.24	356

Table 2

Drug-target interaction (DTI) prediction results of antiviral drugs available on markets against a novel coronavirus (SARS-CoV-2, NCBI reference sequence NC_045512.2) RNA-dependent RNA polymerase (accession YP_009725307.1). * indicates isomeric form SMILES.

Small molecules	SMILES used in the DTI prediction	K_d in nM	MT-DTI Rank out of 3411
Grazoprevir*	<chem>C = C[C@H]1C[C@]1(NC(=O)[C@@H]1C[C@@H]2CN1C(=O)[C@H](C(C)(C)NC(=O)O[C@@H]1C[C@H]1CCCCc1nc3ccc(OC)cc3nc1O2)C(=O)NS(=O)(=O)C1CC1</chem>	8.69	40
Ganciclovir	<chem>Nc1nc(=O)c2ncn(COC(CO)CO)c2[nH]1</chem>	11.91	58
Remdesivir*	<chem>CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@@H]1[C@H]([C@H]([C@](O1)(C#N)C2 = CC = C3N2N = CN = C3N)O)O)OC4 = CC = CC = C4</chem>	20.17	93
Atazanavir	<chem>COC(=O)NC(C(=O)NC(Cc1cccc1)C(O)CN(Cc1ccc(-c2cccn2)cc1)NC(=O)C(NC(=O)OC)C(C)C(C)C(C)C</chem>	21.83	100
Daclatasvir	<chem>COC(=O)NC(C(=O)N1CCCC1c1ncc(-c2ccc(-c3ccc(-c4cnc(C5CCCN5C(=O)C(NC(=O)OC)C(C)C)[nH]4)cc3)cc2)[nH]1)C(C)C</chem>	23.31	107
Acyclovir	<chem>Nc1nc(=O)c2ncn(COCCO)c2[nH]1</chem>	26.66	122
Etravirine	<chem>Cc1cc(C#N)cc(C)c1Oc1nc(Nc2ccc(C#N)cc2)nc(N)c1Br</chem>	33.09	151
Entecavir*	<chem>C = C1[C@H](n2cnc3c(=O)nc(N)[nH]c32)C[C@H](O)[C@H]1CO</chem>	52.83	215
Efavirenz*	<chem>O = C1Nc2ccc(Cl)cc2[C@@](C#CC2CC2)(C(F)F)O1</chem>	76.70	287
Asunaprevir	<chem>C = CC1CC1(NC(=O)C1CC(Oc2ncc(OC)c3ccc(Cl)cc23)CN1C(=O)C(NC(=O)OC)C(C)C(C)C(=O)NS(=O)(=O)C1CC1</chem>	78.36	291
Abacavir	<chem>Nc1nc(NC2CC2)c2ncn(C3C = CC(CO)C3)c2n1</chem>	131.51	417
Darunavir*	<chem>CC(C)CN(C[C@@H](O)[C@H](Cc1cccc1)N = C(\O)O[C@H]1CO[C@H]2OCC[C@@H]12)S(=O)(=O)c1ccc(N)cc1</chem>	148.74	464
Dolutegravir	<chem>CC1CCOC2Cn3cc(C(=O)NC4ccc(F)cc4F)c(=O)c(O)c3C(=O)N12</chem>	150.15	471
Rilpivirine	<chem>Cc1cc(C = CC#N)cc(C)c1Nc1ccnc(Nc2ccc(C#N)cc2)n1</chem>	228.94	628
Lomibuvir	<chem>CC1CCC(C(=O)N(c2cc(C#C(C)C)C)sc2C(=O)O)C2CCC(O)CC2)CC1</chem>	280.96	703
Penciclovir	<chem>Nc1nc(=O)c2ncn(CCC(CO)CO)c2[nH]1</chem>	312.93	740
Triflurdine	<chem>O = c1[nH]c(=O)n(C2CC(O)C(CO)O2)cc1C(F)(F)F</chem>	315.79	744
Nevirapine	<chem>Cc1ccnc2c1NC(=O)c1cccnc1N2C1CC1</chem>	367.50	812
Danoprevir*	<chem>CC(C)C(OC(=O)N[C@H]1CCCC/C = C/[C@@H]2[C@@]2(CNS(=O)(=O)C2CC2)NC(=O)[C@@H]2[C@@H](OC(=O)N3C4cccc(F)c4C3)CN2C1 = O</chem>	405.66	860
Efavirenz	<chem>O = C1Nc2ccc(Cl)cc2C(C#CC2CC2)(C(F)F)O1</chem>	406.33	862
Lamivudine	<chem>Nc1ccn(C2CSC(CO)O2)c(=O)n1</chem>	511.88	981
Ritonavir	<chem>CC(C)c1nc(CN(C)C(=O)NC(C(=O)NC(Cc2ccccc2)CC(O)C(Cc2ccccc2)NC(=O)OCc2cncs2)C(C)C)cs1</chem>	624.30	1077
Saquinavir	<chem>CC(C)C(NC(=O)[C@@H]1C[C@@H]2CCCC[C@H]2CN1C[C@H](O)[C@H](Cc1cccc1)NC(=O)[C@H](CC(N) = O)NC(=O)c1ccc2ccccc2n1</chem>	704.86	1146
Ganciclovir*	<chem>Nc1nc2c(ncn2COC(CO)CO)c(=O)[nH]1</chem>	723.55	1163
Raltegravir	<chem>Cc1nnc(C(=O)NC(C)C)c2nc(C(=O)NCc3ccc(F)cc3)c(O)c(=O)n2Co1</chem>	832.25	1241
Lamivudine*	<chem>Nc1ccn([C@@H]2CS[C@H](CO)O2)c(=O)n1</chem>	999.92	1330

Table 3

Drug-target interaction (DTI) prediction results of antiviral drugs available on markets against a novel coronavirus (SARS-CoV-2, NCBI reference sequence NC_045512.2) helicase (accession YP_009725308.1). * indicates isomeric form SMILES.

Small molecules	SMILES used in the DTI prediction	K_d in nM	MT-DTI Rank out of 3411
Remdesivir*	<chem>CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@@H]1[C@H]([C@H]([C@](O1)(C#N)C2 = CC = C3N2N = CN = C3N)O)O)OC4 = CC = CC = C4</chem>	6.48	32
Daclatasvir	<chem>COC(=O)NC(C(=O)N1CCCC1c1ncc(-c2ccc(-c3ccc(-c4cnc(C5CCCN5C(=O)C(NC(=O)OC)C(C)C)[nH]4)cc3)cc2)[nH]1)C(C)C</chem>	10.95	66
Danoprevir*	<chem>CC(C)C(OC(=O)N[C@H]1CCCC/C = C/[C@@H]2[C@@]2(CNS(=O)(=O)C2CC2)NC(=O)[C@@H]2[C@@H](OC(=O)N3C4cccc(F)c4C3)CN2C1 = O</chem>	19.83	113
Simeprevir*	<chem>COc1ccc2c(O)[C@H]3CC4C(=O)N(C)CCCC/C = C/[C@H]5[C@@]5(C(=O)NS(=O)(=O)C5CC5)NC(=O)[C@@H]4C3)cc(-c3nc(C(C)C)es3)nc2c1C</chem>	23.34	122
Atazanavir	<chem>COC(=O)NC(C(=O)NC(Cc1cccc1)C(O)CN(Cc1ccc(-c2cccn2)cc1)NC(=O)C(NC(=O)OC)C(C)C(C)C(C)C</chem>	25.92	136
Grazoprevir*	<chem>C = C[C@H]1C[C@]1(NC(=O)[C@@H]1C[C@@H]2CN1C(=O)[C@H](C(C)(C)NC(=O)O[C@@H]1C[C@H]1CCCCc1nc3ccc(OC)cc3nc1O2)C(=O)NS(=O)(=O)C1CC1</chem>	26.28	139
Asunaprevir	<chem>C = CC1CC1(NC(=O)C1CC(Oc2ncc(OC)c3ccc(Cl)cc23)CN1C(=O)C(NC(=O)OC)C(C)C(C)C(=O)NS(=O)(=O)C1CC1</chem>	28.20	151
Telaprevir	<chem>CCCC(NC(=O)C1C2CCCC2CN1C(=O)C(NC(=O)C(NC(=O)c1cncn1)C1CCCC1)C(C)C(C)C(=O)C(=O)NC1CC1</chem>	40.75	194
Ritonavir	<chem>CC(C)c1nc(CN(C)C(=O)NC(C(=O)NC(Cc2ccccc2)CC(O)C(Cc2ccccc2)NC(=O)OCc2cncs2)C(C)C)cs1</chem>	41.60	196
Lopinavir*	<chem>Cc1cccc(C)c1OCC(=O)N[C@@H](Cc1cccc1)[C@@H](O)[C@H](Cc1cccc1)NC(=O)[C@H](C(C)C)N1CCNC1 = O</chem>	78.49	295
Darunavir*	<chem>CC(C)CN(C[C@@H](O)[C@H](Cc1cccc1)N = C(\O)O[C@H]1CO[C@H]2OCC[C@@H]12)S(=O)(=O)c1ccc(N)cc1</chem>	90.38	328
Ganciclovir	<chem>Nc1nc(=O)c2ncn(COC(CO)CO)c2[nH]1</chem>	108.21	365
Efavirenz*	<chem>O = C1Nc2ccc(Cl)cc2[C@@](C#CC2CC2)(C(F)F)O1</chem>	112.69	376
Penciclovir	<chem>Nc1nc(=O)c2ncn(CCC(CO)CO)c2[nH]1</chem>	129.41	420
Etravirine	<chem>Cc1cc(C#N)cc(C)c1Oc1nc(Nc2ccc(C#N)cc2)nc(N)c1Br</chem>	175.50	522
Raltegravir	<chem>Cc1nnc(C(=O)NC(C)C)c2nc(C(=O)NCc3ccc(F)cc3)c(O)c(=O)n2Co1</chem>	299.81	727
Dolutegravir	<chem>CC1CCOC2Cn3cc(C(=O)NC4ccc(F)cc4F)c(=O)c(O)c3C(=O)N12</chem>	333.32	780
Nelfinavir	<chem>Cc1c(O)cccc1C(=O)NC(CSc1cccc1)C(O)CN1CC2CCCC2CC1C(=O)NC(C)C</chem>	365.96	823
Indinavir*	<chem>CC(C)C(NC(=O)[C@@H]1CN(Cc2cccn2)CCN1C[C@@H](O)[C@H](Cc1cccc1)C(=O)N[C@H]1c2cccc2[C@H]1O</chem>	401.78	866
Efavirenz	<chem>O = C1Nc2ccc(Cl)cc2C(C#CC2CC2)(C(F)F)O1</chem>	412.86	880
Entecavir*	<chem>C = C1[C@H](n2cnc3c(=O)nc(N)[nH]c32)C[C@H](O)[C@H]1CO</chem>	452.78	943
Ritonavir*	<chem>CC(C)c1nc(CN(C)C(=O)N[C@H](C(=O)N[C@H](Cc2ccccc2)NC(=O)OCc2cncs2)C(C)C)es1</chem>	462.20	961
Boceprevir	<chem>CC(C)C(NC(=O)NC(C(=O)N1CC2C(C1C(=O)NC(C1CCC1)C(=O)C(N) = O)C2(C)C(C)C(C)C</chem>	510.35	1011
Lomibuvir	<chem>CC1CCC(C(=O)N(c2cc(C#C(C)C)C)sc2C(=O)O)C2CCC(O)CC2)CC1</chem>	543.41	1049
Acyclovir	<chem>Nc1nc(=O)c2ncn(COCCO)c2[nH]1</chem>	661.76	1162
Nevirapine	<chem>Cc1ccnc2c1NC(=O)c1cccnc1N2C1CC1</chem>	864.38	1373

Table 6

Drug-target interaction (DTI) prediction results of antiviral drugs available on markets against a novel coronavirus (SARS-CoV-2, NCBI reference sequence NC_045512.2) 2'-O-ribose methyltransferase (accession YP_009725311.1). * indicates isomeric form SMILES.

Small molecules	SMILES used in the DTI prediction	K_d in nM	MT-DTI Rank out of 3411
Remdesivir*	<chem>CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@H]1[C@H]([C@H]([C@H]([C@H](O1)(C#N)C2 = CC = C3N2N = CN = C3N)O)O)OC4 = CC = CC = C4</chem>	134.39	40
Dolutegravir	<chem>CC1CCOC2Cn3cc(C(=O)NCc4ccc(F)cc4F)c(=O)c(O)c3C(=O)N12</chem>	153.73	46
Atazanavir	<chem>COC(=O)NC(C(=O)NC(Cc1cccc1)C(O)CN(Cc1ccc(-c2cccn2)cc1)NC(=O)C(NC(=O)OC)C(C)(C)C(C)(C)C</chem>	390.67	90
Efavirenz*	<chem>O = C1Nc2ccc(Cl)cc2[C@H](C#CC2CC2)(C(F)(F)F)O1</chem>	423.00	99
Boceprevir	<chem>CC(C)(C)NC(=O)NC(C(=O)N1CC2(C1C(=O)NC(CC1CCC1)C(=O)C(N) = O)C2(C)C(C)(C)C</chem>	433.93	101

to be more potent than chloroquine for inhibiting SARS-CoV-2 [32]. Remdesivir and lopinavir/ritonavir (Kaletra) also reduced pneumonia-associated symptoms of some COVID-19 patients [33,34]. However, these studies are based on previous knowledge that these drugs showed some inhibitory effects on similar coronaviruses such as SARS-CoV and/or MERS-CoV. In contrast, our approach was truly based on a pre-trained MT-DTI deep-learning model that understands drug-target interactions without domain knowledge [12]. In fact, MT-DTI successfully identified the epidermal growth factor receptor (EGFR)-targeted drugs that are used in clinics (in top-30 predicted candidates) among 1794 chemical compounds registered in the DrugBank database in a previous study [12], suggesting that 3D structural information of proteins and/or molecules is not necessarily required to predict drug-target interactions.

Our results showed the following intriguing findings that need to be tested experimentally and clinically in the near future. First, MT-DTI generally showed similar results overall compared to the conventional 3D structure-based prediction model, AutoDock Vina, but some differences were observed. For example, atazanavir, remdesivir, and efavirenz were the top three predicted drugs that may bind to the 3C-like proteinase of SARS-CoV-2. This is while saquinavir, nelfinavir, and grazoprevir were the top three drugs identified by AutoDock Vina (Fig. 1). Secondly, when the search space was expanded to all FDA-approved drugs, some immunosuppressant drugs (rapamycin and everolimus) and a drug (tiotropium bromide) for asthma and chronic obstructive pulmonary disease (COPD) were identified as promising candidates by MT-DTI. In contrast, AutoDock Vina predicted purmorphamine, lumacaftor, and verrucaric acid were the top three drugs that could bind to the 3C-like proteinase of SARS-CoV-2. However, there is currently no supporting evidence that these drugs may be effective in inhibiting SARS-CoV-2. Lastly, atazanavir appears to be effective in the treatment of COVID-19 by showing overall high binding affinities among tested antivirals for six proteins of SARS-CoV-2 including 3C-like proteinase and the replication complex components (Tables 1–6 and S1–6). But, this prediction also needs to be validated *in vitro*, *in vivo*, and in a wide range of clinical trials for efficacy and safety.

We hope our prediction results may support experimental therapeutic options for China and other countries suffering from the SARS-CoV-2 pandemic and align with recent clinical trials [35].

CRedit authorship contribution statement

Bo Ram Beck: Data curation, Writing - original draft. **Bonggun Shin:** Conceptualization, Methodology, Software. **Yoonjung Choi:** Methodology, Writing - review & editing. **Sungsoo Park:** Conceptualization, Methodology, Data curation, Software. **Keunsoo Kang:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

Beck B.R., Choi Y., and Park S. are employed by company Deargen Inc. Shin B. is employed by Deargen Inc as a part-time advisor. Kang K. is one of the co-founders of, and a shareholder in, Deargen Inc.

Acknowledgements

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health & Welfare, Republic of Korea (1720100).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2020.03.025>.

References

- Cavanagh D. Coronavirus avian infectious bronchitis virus. *Vet Res* 2007;38:281–97.
- Lim YX, Ng YL, Tam JP, Liu DX. Human coronaviruses: a review of virus-host interactions. *Diseases* 2016;4(3).
- Weiss SR, Leibowitz JL. Coronavirus pathogenesis. *Adv Virus Res* 2011;81:85–164.
- Woo PC, Huang Y, Lau SK, Yuen KY. Coronavirus genomics and bioinformatics analysis. *Viruses* 2010;2:1804–20.
- Kuiken T, Fouchier RA, Schutten M, Rimmelzwaan GF, van Amerongen G, van Riel D, et al. Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 2003;362:263–70.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 2012;367:1814–20.
- WHO. Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003. Dec 31, 2003. (https://www.who.int/csr/sars/country/table2004_04_21/en/). 2020.
- WHO. Middle East respiratory syndrome coronavirus (MERS-CoV). November, 2019. (<http://www.who.int/emergencies/mers-cov/en/>). 2020.
- Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020.
- Xu J, Zhao S, Teng T, Abdalla AE, Zhu W, Xie L, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* 2020;12(2):244.
- Fung TS, Liu DX. Human coronavirus: host-pathogen interaction. *Annu Rev Microbiol* 2019;73:529–57.
- Shin B, Park S, Kang K, Ho JC. Self-attention based molecule representation for predicting drug-target interaction. *Proc Mach Learn Res* 2019;106:230–48.
- Tanoli Z, Alam Z, Vaha-Koskela M, Ravikumar B, Malyutina A, Jaiswal A, et al. Drug Target Commons 2.0: a community platform for systematic analysis of drug-target interaction profiles. *Database (Oxford)* 2018;2018:1–13.
- Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucl Acids Res* 2007;35:D198–201 (Database issue).
- Paolini GV, Shapland RH, van Hoorn WP, Mason JS, Hopkins AL. Global mapping of pharmacological space. *Nat Biotechnol* 2006;24(7):805–15.
- Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*. *Cell Res* 2020.
- Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455–61.
- O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform* 2011;3:33.

- [19] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;30(16):2785–91.
- [20] Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev* 2017;9(2):91–102.
- [21] Ozturk H, Ozgur A, Ozkirimli E. DeepDTA: deep drug-target binding affinity prediction. *Bioinformatics* 2018;34(17):i821–9.
- [22] He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines. *J Cheminform* 2017;9(1):24.
- [23] Pahikkala T, Airola A, Pietila S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug-target interaction predictions. *Brief Bioinf* 2015;16(2):325–37.
- [24] Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis. *J Chem Inf Model* 2014;54(3):735–43.
- [25] Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nat Biotechnol* 2011;29(11):1046–51.
- [26] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Z. Zhao, et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *bioRxiv*. 2020.
- [27] Dayer MR, Taleb-Gassabi S, Dayer MS. Lopinavir; a potent drug against coronavirus infection: insight from molecular docking study. *Arch Clin Infect Dis* 2017;12:e13823.
- [28] de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nat Rev Microbiol* 2016;14:523–34.
- [29] Hopkins JSUS. drugmakers ship therapies to China, seeking to treat coronavirus. *Wall Street J* 2020.
- [30] Chu CM, Cheng VC, Hung IF, Wong MM, Chan KH, Chan KS, et al. Role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* 2004;59(3):252–6.
- [31] Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* 2020;30(3):269–71.
- [32] Yao X, Ye F, Zhang M, Cui C, Huang B, Niu P, et al. In vitro antiviral activity and projection of optimized dosing design of hydroxychloroquine for the treatment of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clin Infect Dis* 2020.
- [33] Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020;382(10):929–36.
- [34] Wang Z, Chen X, Lu Y, Chen F, Zhang W. Clinical characteristics and therapeutic procedure for four cases with 2019 novel coronavirus pneumonia receiving combined Chinese and Western medicine treatment. *Biosci Trends* 2020;14(1):64–8.
- [35] Maxmen A. More than 80 clinical trials launch to test coronavirus treatments. *Nature* 2020;578:347–8.