

# FDA-approved Pralatrexate identified by virtual drug screening inhibits SARS-CoV-2 replication *in vitro*

**Haiping Zhang**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences <https://orcid.org/0000-0003-2133-1768>

**Yang Yang**

Second Hospital Affiliated to Southern University of Science and Technology

**Junxin Li**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

**Min Wang**

Institute of Microbiology, Chinese Academy of Sciences

**Konda Mani Saravanan**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences <https://orcid.org/0000-0002-5541-234X>

**Jinli Wei**

Second Hospital Affiliated to Southern University of Science and Technology

**Justin Tze-Yang Ng**

Nanyang Technological University

**Md. Tofazzal Hossain**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences and University of Chinese Academy of Sciences <https://orcid.org/0000-0002-8281-6059>

**Maoxuan Liu**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

**Huiling Zhang**

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

**Xiaohu Ren**

Shenzhen Center for Disease Control and Prevention

**Yi Pan**

Georgia State University

**Yin Peng**

Shenzhen University

**Yi Shi**

Institute of Microbiology, Chinese Academy of Sciences <https://orcid.org/0000-0002-3053-2687>

**Xiaochun Wan** ([✉ xc.wan@siat.ac.cn](mailto:xc.wan@siat.ac.cn))

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

**Yingxia Liu** (✉ [yingxiali@hotmai.com](mailto:yingxiali@hotmai.com))

Second Hospital Affiliated to Southern University of Science and Technology

**Yanjie Wei** (✉ [yj.wei@siat.ac.cn](mailto:yj.wei@siat.ac.cn))

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences

---

## Article

**Keywords:** SARS-CoV-2, RdRp, Virtual screening, Pralatrexate, Azithromycin

**Posted Date:** August 24th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-36439/v2>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

## Abstract

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic poses serious threats to the global public health and leads to an unprecedented worldwide crisis. Unfortunately, no effective drugs or vaccines are available till now. Since the RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 is a promising therapeutic target, a deep learning and molecular simulation based hybrid drug screening procedure was proposed and applied to identify potential drug candidates targeting RdRp from 1906 approved drugs. Among the four selected FDA-approved drug candidates, Pralatrexate and Azithromycin were confirmed to effectively inhibit SARS-CoV-2 replication *in vitro* with EC<sub>50</sub> values of 0.008 μM and 9.453 μM, respectively. For the first time, our study discovered that Pralatrexate is able to potently inhibit SARS-CoV-2 replication with a stronger inhibitory activity than Remdesivir within the same experimental conditions. The paper demonstrates the feasibility of accurate virtual drug screening for inhibitors of SARS-CoV-2 and provides potential therapeutic agents against COVID-19.

## Introduction

The Coronavirus Disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has developed into a global pandemic with millions of people infected and tens of thousands of lives being lost [1]. SARS-CoV-2 can be transmitted from person to person with stronger infection ability than SARS-CoV [2,3]. The presence of asymptomatic transmission and a long incubation period results in an extra challenge in preventing SARS-CoV-2 transmission and spread [4]. As of 11th July 2020, more than 12,556,000 human infections with more than 558,000 deaths have been reported. To date, there are no clinically approved drugs or vaccines available for the treatment or prevention of COVID-19. Therefore, identifying antiviral agents that can combat SARS-CoV-2 is of high significance.

De novo drug development process is time-consuming and costly, which cannot meet the urgent need to combat COVID-19. Given current emergencies, repurposing existing FDA-approved drugs for COVID-19 provides a shortcut [5]. Recently, some drugs under clinical trials such as Remdesivir were shown to inhibit the replication of SARS-CoV-2 *in vitro* [6–8]. The structural basis of the RNA-dependent RNA polymerase (RdRp) inhibited by Remdesivir is well illustrated in a recent work [9]. Some severe SARS-CoV-2 patients treated with compassionate-use Remdesivir targeting RdRp, have shown a significant clinical improvement [10]. As a core component of the RNA synthesis machinery, RdRp is believed to be one of the most promising therapeutic targets for COVID-19 [11,12]. Small compounds binding with the catalytic site of RdRp have potential to interfere with SARS-CoV-2 RNA synthesis[13].

Several computational drug screening methods relying on molecular docking, deep learning or Molecular Dynamics (MD) simulations have been applied in drug repositioning studies for COVID-19 [12]. However, most studies normally rely on a single technique or lack experimental validation. Each computational technique has different pros and cons, and a proper combination and modifications of such methods may provide a better solution. Previously, we have developed two deep learning-based models to estimate the protein-ligand interaction, the DFCNN [14] and DeepBindBC (<http://hpcc.siat.ac.cn/DeepBindBC/>).

Instead of using protein-ligand docking conformation, DFCNN uses molecular vectors of protein pocket and ligand to estimates the protein-ligand pair as binding or non-binding with a probability value between 0 and 1. On the other hand, DeepBindBC estimates the binding possibility from the interface atom contact information of the modelled 3D protein-ligand complex. The input of DeepBindBC contains spatial information of the protein-ligand interface, thus it is strongly complementary to DFCNN.

In the present work, we propose a deep learning and molecular simulation based hybrid screening method which consists of DFCNN [14], DeepBindBC, Autodock Vina [15], pocket molecular dynamics simulation, and metadynamics to explore the potential 1906 drugs from TargetMol-Approved\_Drug\_Library, a collected drug library by TargetMol. We systematically selected four FDA-approved drugs for experimental validation, Pralatrexate was found to efficiently inhibit the replication of SARS-CoV-2 *in vitro*.

## Results

The overall workflow of virtual drug screening against RdRp is illustrated in Fig. 1**a**. The interaction patterns (taken from the last frame of 100 ns MD simulation) between RdRp and Azithromycin, Pralatrexate are shown in Fig. 1**b(i)(ii)**, respectively. The interaction between Remdesivir in its monophosphate form and RdRp (PDB ID: 7BV2) is also given for comparison in Fig 1**b(iii)**. Azithromycin forms 2 hydrogen bonds with GLN573 and ILE494 through keto and hydroxyl groups respectively, and many hydrophobic related interactions with the RdRp binding site (e.g. LYS577, LEU576, ALA685) through alkyl groups, whereas Pralatrexate shows enhanced and more stable interactions with RdRp binding site, including 6 hydrogen bonds with GLN573, ARG569, ASN496, ASN497, LYS500 and GLY590. Pralatrexate also forms Alkyl or Pi-Alkyl interaction with LYS577 and LEU576, and salt bridges with ARG569 and LYS500. Azithromycin and Pralatrexate share 10 common neighbor residues (62.5%) of RdRp, as shown in Fig 1**b(i)** and **b(ii)** and Supplementary Table 1, indicating a similar binding cavity.

A recent study shows the Remdesivir in its monophosphate form interacts with the RdRp involving with partial double-stranded RNA template, and covalently incorporated into the primer strand at the +1 position [9]. Shown in Fig. 1**b(iii)** and Supplementary Table 1, the key interacting contacts between Remdesivir and RdRp include ARG555. Remdesivir in its monophosphate form only shares 2 common close contact residues with Pralatrexate, and no common close contact residue with Azithromycin, suggesting a different interaction pattern.

Our drug screening protocol relies on non-covalent bond interactions and assumes RdRp is in apo form (without RNA primer), Fig. 1**b(i)(ii)** shows Azithromycin and Pralatrexate interacts with the region of RdRp which is consisted of residues such as GLN573, ARG569, ALA685. Comparing our modeled structure with Remdesivir in its monophosphate-RdRp complex, it is possible Azithromycin and Pralatrexate occupies part of the RdRp cavity with the non-covalent binding which may interfere the entry of the RNA primer strand to the cavity.

### Deep learning and docking based virtual screening

Molecular vector-based deep learning method (DFCNN), structure feature-based deep learning method (DeepBindBC) and molecular docking (Autodock Vina) are complementary, which can be seen from Supplementary Table 2 that three methods predict different drug compounds as top hits. For instance, Amenamevir and Azithromycin have better Autodock Vina scores whereas Odanacatib and Nitisinone are found to have high DFCNN scores.

We first reject those drugs that have poor prediction by any of the methods. The 22 drugs with DFCNN score above 0.9 and docking score below -7 kcal/mol were firstly selected and presented in Supplementary Table 2. Among the 22 drugs, we further excluded the drugs with a DeepBindBC score below 0.7, resulting in 14 drugs shown in bold in Supplementary Table 2. Overall there are 5 nucleoside analogues, 2 antibiotic drugs, 3 antivirus drugs, 2 anticancer drugs and 3 other drugs (Supplementary Fig. 1) selected by the molecular vector-based and structure-based screening process. Sofosbuvir is both a nucleoside analogue and an antivirus drug. These 14 drugs are subject to force field-based screening in the next stage.

Among 2 antibiotic drugs, Azithromycin, a drug used to treat a variety of bacterial infections, showed top Autodock Vina score of -8.6 kcal/mol, good DFCNN score and DeepBindBC score (0.9093 and 0.8589), respectively. Gautret, P. *et al* claimed that combined with Hydroxychloroquine, Azithromycin can have good efficiency in treating COVID-19 with significant viral load reduction [16]. However it should be noticed that currently there is no evidence of the effectiveness of Azithromycin in the treatment of COVID-19 and have many debates about effective of Azithromycin on COVID-19 [17].

The top two predicted molecules by DeepBindBC are nucleotide analogues. Sofosbuvir is a nucleotide analogue inhibitor of hepatitis C virus (HCV) NS5B polymerase to treat infectious liver disease, [18] whereas Clofarabine is a purine nucleoside antimetabolite used for treating refractory acute lymphoblastic leukaemia [19]. More nucleotides analogues in the candidate list were selected by our method, such as Adenosine, Vidarabine, and Gemcitabine, indicating some RdRp-nucleotides interaction patterns have been implicitly recognized by the proposed hybrid drug screening method.

### Force field-based virtual screening

To further screen the 14 selected drugs (2D structures shown in Supplementary Fig. 1) and understand their interactions and stability, we have performed MD simulations on RdRp-drug complexes. The structure stability is estimated by Root Mean Square Deviation (RMSD) over 100ns time scales, shown in Supplementary Fig. 2. The drugs such as Azithromycin, Pralatrexate, Romidepsin, Teriflunomide and Vidarabine are found stable indicated by the minimum RMSD fluctuations. On the other hand, the drugs such as Adenosine, Amenamevir, Fipronil, Gemcitabine and Sofosbuvir have high RMSD fluctuations. The number of hydrogen bonds formed between RdRp and the drug in the 100 ns MD simulation was also computed and shown in Supplementary Fig. 3. Among the 14 RdRp-drug complexes, Pralatrexate clearly showed more hydrogen bonds with the RdRp than others.

The binding free energy vs coordination number (CV: collective variable) from metadynamics simulations is shown in Supplementary Fig. 4a. The lowest energy conformations of protein-drug complexes for Amoxicillin, Azithromycin, Pralatrexate and Sofosbuvir showed more contacts in the interface region, as indicated by the high coordination numbers (Supplementary Fig. 4a (iii)), while most other compounds favor smaller coordination number (close to zero) indicating no or weak interactions (Supplementary Fig. 4a (i)).

Detailed interaction patterns between RdRp and the four most optimal compounds (Azithromycin, Pralatrexate, Amoxicillin and Sofosbuvir) are shown in Fig. 1b(i)(ii) and Supplementary Fig. 4b (i)(ii), whose structures are taken from the last frame of the 100 ns MD simulation. Azithromycin and Pralatrexate interact with 16 amino acid residues of RdRp to form a stable complex. The RdRp-Azithromycin complex is mainly dominated by van der waals interactions, whereas Pralatrexate involves more polar and charge interactions. According to the calculated free energy difference ( $\Delta G$ ) values from the metadyanmics simulations between the unbound state and the binding state for Amoxicillin, Azithromycin, Pralatrexate and Sofosbuvir (Supplementary Table 3), Azithromycin and Pralatrexate (-305.76 kJ/mol, -128.58 kJ/mol) show more favorable binding energy than Amoxicillin and Sofosbuvir (-67.33 kJ/mol and -89.88kJ/mol).

It is noted that all the nucleoside analogues highly recommended by deep learning-based screening methods were excluded from the force field based screening process. The possible explanation is that our protein-drug systems do not contain the RNA primers during the MD simulation, and covalent bond formation, such as the Remdesivir in its monophosphate form, could not be estimated by traditional MD simulation.

### Pralatrexate and Azithromycin inhibit the replication of SARS-CoV-2 *in vitro*.

To further confirm the efficiency of the hits from the virtual screening, we tested the antiviral activity of the Azithromycin, Pralatrexate, Amoxicillin and Sofosbuvir *in vitro*. Vero cells were infected with SARS-CoV-2 (BetaCoV/Shenzhen/SZTH-003/2020, GISAID No. EPI\_ISL\_406594) at a MOI of 0.02 (the cytopathic effect was mild at 48 hours post-infection with this MOI) in the presence of varying concentrations of the tested drugs, and the inhibition rates were evaluated by quantification of viral copy numbers in the cell supernatant via quantitative reverse transcription polymerase chain reaction (qRT-PCR) and confirmed with immunofluorescence assay (Fig. 2). The results showed that Pralatrexate and Azithromycin could efficiently inhibit the replication of SARS-CoV-2, with half-maximal effective concentration ( $EC_{50}$ ) values of 0.008 and 9.453  $\mu$ M (Fig. 2a), whereas Remdesivir achieved an inhibitory activity with  $EC_{50}$  value of 8.777  $\mu$ M within the same experimental system (Supplementary Fig. 5). IFA showed similar results with qRT-PCR assay (Fig. 2b). CCK-8 assay of the two drugs showed that the half-cytotoxic concentration ( $CC_{50}$ ) values of Pralatrexate and Azithromycin on Vero cells were 0.167  $\mu$ M and > 100  $\mu$ M, respectively, and the calculated the selectivity indexes (SI) of Pralatrexate and Azithromycin were 20.878 and >10.579, respectively. Whether the two drugs worked at the stage of viral entry or post entry was analyzed using time-of-addition assay as previously reported[8]. The results showed that

Pralatrexate functioned at a stage post virus entry, while Azithromycin functioned at both entry and post-entry stages of the SARS-CoV-2 infection in Vero cells (Fig. 2c). Furthermore, surface plasmon resonance (SPR) experiments were performed to test the *in vitro* binding of Pralatrexate and Azithromycin with immobilized RdRp protein of SARS-CoV-2. Both drugs showed obvious binding response in Supplementary Fig. 6.

## Discussion

To perform the drug screening process efficiently and accurately is still a challenge for computer-aided drug design. Though a recent deep learning-based approaches has demonstrated its potential to be efficient/accurate by learning from a sufficient amount of training data, problems such as overfitting, and the discrepancy between training data and real-world data remain [20]. The proposed deep learning and molecular simulation based drug screening method was able to select 4 FDA-approved drug candidates targeting RdRp from 1906 drugs, and 2 out of 4 (Pralatrexate and Azithromycin) can effectively inhibit SARS-CoV-2 replication *in vitro* with EC<sub>50</sub> values of 0.008μM and 9.453 μM. The molecular vector-based deep learning method and the structure-based deep learning method are complementary to each other in the sense that high efficiency and accuracy are both achieved.

For the first time, Pralatrexate is found to potently inhibit SARS-CoV-2 replication *in vitro* with a stronger inhibitory activity (EC<sub>50</sub> value  $\leq$  0.008μM) than Remdesivir ( $P < 0.0001$ ) within the same experimental system. Compared with the GHDDI drug list (The Global Health Drug Discovery Institute: <https://ghddi-ailab.github.io/Targeting2019-nCoV/preclinical/>) that inhibit SARS-CoV-2 *in vitro*, Pralatrexate showed the smallest EC<sub>50</sub>. Among the 154 current reported drugs by GHDDI, NSC319726 have top inhibitory activity over SARS-CoV-2 (EC<sub>50</sub> value  $< 0.02\mu\text{M}$ ) [21]. Pralatrexate is a folate analogue metabolic inhibitor, which was approved by FDA in 2009 for the treatment of patients with relapsed or refractory peripheral T cell lymphoma (PTCL). Pralatrexate inhibits the folate metabolism pathway through inhibition of dihydrofolate reductase (DHFR) [22]. The peak concentration in plasma (Cmax) can achieve 10.5 μM from a standard dosing regimen [23]. Its Cmax is around 800-fold higher than the EC<sub>90</sub> of antiviral activity, suggesting a great potential for clinical implications.

Pralatrexate was selected by the virtual screening pipeline based on its potential acts of inhibiting the RNA dependent RNA protease (RdRp) enzyme, whereas, it's extremely low EC50 for the virus replication compared to Remdesivir (RdRp inhibitor) may have multiply mechanism of action. Pralatrexate is known to be an antifolate that efficiently prevents synthesis of DNA and presumably also RNA [24], which may explain inhibition of SARS-CoV-2 replication. Pralatrexate was approved by FDA in spite of its toxicity, therefore, we should be aware that FDA approval does not guarantee the possibility of immediate use of the drug against COVID-19.

Though both Pralatrexate and Azithromycin inhibit SARS-CoV-2 replication *in vitro*, the time-of-addition experiment showed that they functioned at different stages of SARS-CoV-2 infection. Similar to Remdesivir, Pralatrexate mainly inhibited the replication of SARS-CoV-2 at the stages of post-entry. On the

other hand, Azithromycin inhibited the replication of SARS-CoV-2 at both entry and post-entry stages like chloroquine [8]. This indicates the Azithromycin may also have multiple mechanism of action.

Out of the 4 selected drug candidates targeting RdRp, Amoxicillin and Sofasbuvir have failed to inhibit SARS-CoV-2 replication *in vitro*. Molecular dynamic simulations show they have deviated from its initial binding position (Supplementary Fig. 4 **a(i)** **a(ii)**) with their ligand RMSD > 1.5 nm for most of the simulation time and large fluctuation was observed (Supplementary Fig. 2). The calculated free energies difference between binding state and unbound state (coordination number around 0) also indicates better binding for Pralatrexate and Azithromycin than Amoxicillin and Sofasbuvir, shown in Supplementary Fig. 4 **a(iii)** **a(iv)** and Supplementary Table 3.

To examine why Sofosbuvir can efficiently inhibit RdRp of hepatitis C virus (HCV)[18] while not RdRp of SARS-CoV-2, we have carried a sequence and structural comparison between RdRp of HCV and RdRp of SARS-CoV-2 virus (Supplementary Fig. 7). In addition to the low sequence identity (23.75%) between RdRp of HCV and RdRp of SARS-CoV-2 virus, binding pockets of both complexes showed a quite different composition. For instance, there are 5 vs 3 ASPs, 2 vs 1 LYSs, 1 vs 3 GLUs, 0 vs 6 ARGs in RdRp pocket of SARS-CoV-2 and RdRp pocket of HCV, respectively. The RdRp pocket of SARS-CoV-2 is more negatively charged, while the RdRp pocket of HCV is more positively charged.

Full system protein-ligand MD simulations for RdRp-Pralatrexate, RdRp-Azithromycin were performed to validate the robustness of the pocket MD method. Compared to pocket MD simulation, similar hydrogen bond numbers as well as similar low RMSD fluctuations in full MD simulation were observed according to Supplementary Fig. 3 and Supplementary Fig. 8**a(i)(ii)**. Some key neighbor residues in pocket MD simulation for Azithromycin and Pralatrexate were also kept during the full system MD simulation, according to Fig. 1**b(i)(ii)** and Supplementary Fig. 8**b(i)(ii)**. For instance, LEU576, ILE589, ALA580 and ALA685 have formed alkyl related hydrophobic interaction with Azithromycin in the last frames of both simulations, and ARG569, ASN496 and LYS500 of RdRp have formed salt bridge or hydrogen bonds with Pralatrexate in the last frames of both simulations.

The efficiency and effectiveness of the DFCNN method have been examined by screening about 10 million drugs targeting 8 representative protein targets taken from the DUD.E diverse data set. DFCNN was able to screen the 10 million drugs within 5 hours using a workstation with 80 Intel CPU cores (2.00 GHz ) and 60 GB RAM. The effectiveness is evaluated by the prediction-random ratio ( $\text{Ratio}_{0.9}$ ), shown in Supplementary Table 4. For 6 out of 8 protein targets,  $\text{Ratio}_{0.9}$  is greater than 1.4, indicating DFCNN is able to enrich the active compounds in ten million compound pools. Among the 8 test cases, the DFCNN achieved best performance on HIVPR (Human immunodeficiency virus type 1 protease) with  $\text{Ratio}_{0.9}$  of ~860 ( about 860 times better than random guess in selecting active compounds in terms of TPR). DFCNN performed worse for GPCR proteins (such as CXCR4) and protein with small inner pocket (such as AKT1). The possible reason is that GPCRs have limited number of reliable structure of protein-ligand complexes in our training dataset and membrane proteins may have very different binding mechanism compared to other type proteins. The poor performance for proteins with small inner pocket is likely due

to the special physical-chemical and spatial features. As an enzyme, RdRp has large ligand binding cavity and should be suitable for virtual screening by DFCNN.

To study how the molecular vector-based deep learning screening method selects the 139 candidate drugs from 1906 drugs, 1906 drugs were clustered into 20 groups (Supplementary Fig. 9). Group 20 has the highest ratio of drugs being selected (31/89 drugs, Supplementary Fig. 9a). The drugs in the Groups 19, Groups 20, Group 17 and Group 15 with high selection ratio tend to contain many electrical donors and electrical acceptors, likely due to the RdRp pocket containing many charged groups (Supplementary Fig. 10), including 5 ASPs, 2 LYSSs, and 1 GLU. The percentage of charge and polar residues in the RdRp pocket reaches 54.35% (Supplementary Fig. 10), which explains why DFCNN prefers to select polar and charged drugs for the RdRp. The structure-based screening (Autodock Vina plus DeepBindBC) selected 14 drugs from 139 drugs, 6 drugs belong to Group 15 (Supplementary Table 5), including nucleotide analogues as well as Pralatrexate. Groups 17, 18, 19 all have 2 drugs selected after structure-based screening. Belonging to different clusters, Pralatrexate has many hydrogen donors and acceptors while Azithromycin contains a macrocycle, which tends to form hydrogen bond (or salt bridge) and macrocyclic hydrophobic interactions, respectively.

## Conclusion

Identifying effective drugs that can treat COVID-19 is important and urgent, especially the approved drugs that can be immediately tested in clinical trials. In this work, we have developed a hybrid protocol of combining deep learning methods with molecular simulations to search for potential drug candidates against RdRp that can inhibit the replication of SARS-CoV-2. Four potential drugs were systematically selected for experimental validation, and Pralatrexate and Azithromycin showed an inhibiting effect with EC<sub>50</sub> values of 0.008μM and 9.453 μM, respectively. Experimental results from qRT-PCR, CCK-8 assay, indirect immunofluorescence assay (IFA), Time-of-addition and Surface plasmon resonance (SPR) assay show the proposed screening protocol successfully identified a new therapeutic agent Pralatrexate against COVID-19 by targeting RdRp. The hybrid strategy of combining deep learning, molecular docking, MD simulation in a virtual screening pipeline can effectively help with drug repurposing application and facilitate virtual drug screening against other targets in SARS-CoV-2.

## Materials And Methods

In this paper, a deep learning and molecular simulation based hybrid strategy is proposed for virtual drug screening against RdRp over the TargetMol-Approved-Drug-Library, an approved drug library with 1906 compounds collected by TargetMol, resulting in four candidates (Pralatrexate, Azithromycin, Sofosbuvir, Amoxicillin) for drug repurposing. qRT-PCR assay, indirect immunofluorescence assay (IFA) and CCK-8 assay were carried out to validate the efficacy for Pralatrexate, Azithromycin which inhibit SARS-CoV-2 replication *in vitro*. Surface plasmon resonance (SPR) assay was used to evaluate the RdRp-drug binding affinity.

## Structural modeling of RdRp and drug compound dataset

The RdRp sequence and its modelled structure were obtained from <https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/>. The RdRp-ligand model was constructed by I-TASSER[25]. The ligand was taken from the template protein (PDB ID: 3BR9)[26] by COFACTOR algorithm[27] within the I-TASSER using structure comparison and protein-protein networks. We extract the amino acids within 1 nm of the ligand as the binding pocket. RMSD between the modeled structure and the recent experimental RdRp structure (PDB ID 6M71) is calculated (~0.516 Å), shown in Supplementary Fig. 11a)[13]. RNA polymerase superfamily region is also very similar between these two structures (RMSD=0.456 Å, Supplementary Fig. 11a).

TargetMol-Approved\_Drug\_Library, which contains 1906 compounds, was used as virtual screening library. These 1906 compounds collected by TargetMol are drugs approved by Food and Drug Administration (FDA), the European Medicine Agency (EMA), or China Food and Drug Administration (CFDA), or included in the US Pharmacopeia (USP) Dictionary, the British Pharmacopoeia (BP), the European Pharmacopoeia (EP), the Japanese Pharmacopoeia (JP), or Chinese Pharmacopoeia (CP) Dictionary.

## Molecular vector-based drug screening

A deep learning-based method, DFCNN (Dense fully Connected Neural Network), has been developed for predicting protein-drug binding probability [14] and used in this paper for the initial drug screening (Fig. 1a). DFCNN utilizes the concatenated molecular vector of protein pocket and ligand as input representation, and the molecular vector are generated by Mol2vec[28] which is inspired by the word2vec model in natural language processing. DFCNN model was trained on a dataset extracted from PDBbind database [29]. Negative data samples in the dataset were generated by cross-combination of proteins and ligands from PDBbind database and positive data samples were taken from protein-ligand pairs in experimental structure. The details of the method were described in our previous paper [14], and DFCNN achieved an AUC value around 0.9 for the independent testing set [14]. The model is about ~100,000 times faster than Autodock Vina in predicting protein-ligand binding probability (range 0~1), because it does not rely on the protein-drug complex conformation.

To examine the efficiency and effectiveness of the DFCNN method, we screen a large scale chemical compound dataset (about 10 million compounds) targeting 8 representative protein targets taken from the DUD.E diverse data set. For each target, the corresponding dataset contains some active compounds (between 40 and 536) in the DUD.E dataset and 10,402,895 drug-like compounds from ZINC database. The effectiveness is measured by the prediction-random ratio ( $\text{Ratio}_{0.9}$ ), defined as  $\text{TPR}_{0.9}/\text{Random}_{0.9}$ , where  $\text{TPR}_{0.9}$  indicates the ratio ( $N_{0.9}/\text{Active\_num}$ ) between the number of active compounds with a DFCNN score larger than 0.9 ( $N_{0.9}$ ) and the active number of compounds (Active\_num). The total number of the compounds (Total\_num) with score above 0.9 is defined as NN. The random selection rate

(Random<sub>0.9</sub>) is defined as NN/Total\_num. Using cutoff score of 0.9, the prediction-random ratio measures the ratio of predicted TPR and random selection TPR.

### Structure-based drug screening

DeepBindBC, an in-house deep learning-based software, is used for structure based drug screening. Unlike the DFCNN, the input of DeepBindBC includes both the physical-chemical information and spatial information between the protein-ligand interfaces (Fig. 1a), hence DeepBindBC is able to achieve higher accuracy, but requires protein-drug complex structure information as input generated by Autodock Vina.

Autodock Vina is used to dock the target with the potential ligands [15]. The pocket is determined by the location of ligand in the template protein (PDB ID: 3BR9) [26]. We set the cavity volume space with 3.5 nm, 3.5 nm and 3.5 nm in x, y, z dimensions from the pocket mass center. AutoDock Tools were used to convert PBD file format to PDBQT file format [30]. The exhaustiveness was set to 8; the num\_modes was set to 20, and energy\_range was set to 3. The scoring function and optimization algorithm of Autodock Vina have been well discussed in a previous article[15]. In this study, we selected the most likely targets for further validation by setting a binding energy threshold value of -7 kcal/mol.

The DeepBindBC is a ResNet model trained over the PDBbind database. In DeepBindBC, the protein-ligand interaction interface information will be converted into figure-like metric, similar to DeepBindRG [20]. By incorporating the cross-docking (docking proteins and ligands from different experimental complexes) conformation as negative training data, DeepBindBC is highly possible to distinguish non-binders. Since DeepBindBC relies on docking conformation and DFCNN only uses molecular vector information, these two methods are complementary to each other and DeepBindBC takes much more time than DFCNN.

### Force field-based screening

Further drug screening was carried out by force field based molecular dynamic (MD) simulations. In this study, we selected 14 drug binding complexes for MD simulation, including Adenosine, Amenavir, Amoxicillin, Azithromycin, Clofarabine, Fipronil, Gemcitabine, Nitisinone, Pralatrexate, Raltegravir, Romidepsin, Sofosbuvir, Teriflunomide and Vidarabine, respectively.

We also proposed a pocket molecular dynamics simulation (pocket MD, Supplementary Fig. 11b) to facilitate the simulation process by only keeping the binding pocket region for simulation. Binding free energy calculation can be estimated by metadynamics simulations to explore whether protein-ligand will bind in solution. Metadynamics relies on addition of a bias potential to sample the free energy landscape along a specific collective variable of interest [31][32]. Noted that the binding free energy calculations from Metadynamics may only be suitable for detect the general trend of binding in virtual screening.

The pocket MD is same as the classical MD simulation, except that we only using the pocket region to reduce system size for simulation (Supplementary Fig. 11b), which is inspired by a previous dynamic undocking (DUck) method [33]. An in-house script was used to extract the pocket region of the protein

(1nm within the binding ligand), the N terminal and C terminal ends were capped with the ACE and NHE, respectively. Terminals will be applied a position restrain to maintain the relative conformation of the pocket. MD simulation was carried out by Gromacs with AMBER-99SB force field [34,35]. The topology of ligand and the partial charges of ligand was generated by ACPYPE [36], which relies on Antechamber [37]. Firstly, we created a dodecahedron box and put the target-ligand complex at the center. A minimum distance from the protein to box edge was set to 1 nm. We filled the dodecahedron box with TIP3P water molecules [38], the counter ions was added to neutralize the total charge using the Gromacs program tool [39]. The long-range electrostatic interactions under the periodic boundary conditions was calculated with Particle Mesh Ewald approach [40]. A cutoff of 14 Å was used for van der Waals non-bonded interactions. Covalent bonds involving hydrogen atoms were constrained by applying the LINCS algorithm [41].

We performed the energy minimization steps with a step-size of 0.001ns, 100 ps simulation with isothermal-isovolumetric ensemble (NVT), and 10ns simulation with isothermal-isobaric ensemble (NPT) for water equilibrium. After that, a 100ns NPT production run (step size 2 fs) was carried out. The Parrinello-Rahmanbarostat and the modified Berendsen thermostat were used for simulation with a fixed temperature of 308 K and a pressure of 1 atm. RMSD and hydrogen bond number of the trajectory were calculated using Gromacs tools.

The simulation was continued using the metadynamics approach for exploring the free energy landscape. The interface coordination number of atoms of protein ligand complex was used as collective variable (CV). The protein-ligand interface coordination numbers correlate with the numbers of atom contact, and larger coordination number usually indicates protein-ligand is in binding state.

The coordination number  $C$  is defined as follows by Plumed: **see formulas 1 and 2 in the supplementary files.**

## Tools used in analysis

The UCSF Chimera, VMD, ICM-browserPro and Discovery Studio Visualizer 2019 were used to generate the structure and to visualize the 2D protein-ligand interactions [44–47]. Clusfps (<https://github.com/kaiwang0112006/clusfps>) which depends on RDKit[48] was used to cluster the drugs in the dataset. The drug fingerprint was used as inputs with algorithm of Murtagh[49] being used for clustering 1906 drugs into 20 groups.

## Cell line and drugs

Vero cell (ATCC, CCL-81) was cultured at 37°C in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% fetal bovine serum (FBS, Gibco) in the atmosphere with 5% CO<sub>2</sub>. Cells were seeded in 96-well plates and cultured overnight with a density of  $5 \times 10^4$  cells/well prior infection or drug feeding. Remdesivir, Azithromycin, Pralatrexat, Sofosbuvir and Amoxicillin were obtained from Selleck Chemicals. All drugs were dissolved in DMSO to prepare 50 mM stock solutions, and stored at -20°C. DMSO was used in the controls.

## **Viral stock titration by 50% tissue culture infective dose (TCID<sub>50</sub>)**

TCID<sub>50</sub> was measured as previously reported [50]. In brief, Vero cells in 96-well plates were grown to 80% confluence and infected with 10-fold serial dilutions of the stock SARS-CoV-2 (BetaCoV/Shenzhen/SZTH-003/2020, GISAID No. EPI\_ISL\_406594) for 1 h at 37°C. The inoculum was removed, and cells were overlaid with fresh DMEM plus 2% FBS. At 5 days post infection (d.p.i), plates were assessed for the lowest dilution in which 50% of the wells exhibited cytopathic effects. The values of TCID<sub>50</sub> were calculated according to the Reed-Muench method[51].

## **Evaluation of antiviral activities of the drugs in Vero cells**

Firstly, the cytotoxicity of the five drugs on Vero Cells were determined by CCK8 assays (Sangon). Then the antiviral activities of the drugs were evaluated as previously reported with some modification[8]. Vero cells seeded in 96-well plates were pre-treated with the different doses of the indicated drugs for 1 h, and then virus was subsequently added at multiplicity of infection (MOI) of 0.02 to allow infection for 2 h. Then, the virus-drug mixture was removed and cells were further cultured with fresh DMEM with 2% FBS and the indicated concentrations of drugs. At 48 hours post infection (h.p.i), the cell supernatant was collected and viral RNAs were extracted using the QIAamp RNA Viral Kit (Qiagen, Heiden, Germany) for further quantification analysis. The cells were collected for indirect immunofluorescence assay (IFA). All the experiments involving infectious SARS-CoV-2 were handled in BSL-3 facilities at the Shenzhen Third People's Hospital.

## **Quantitative reverse transcription polymerase chain reaction**

This assay was carried out as described previously [52]. Viral RNAs were extracted from the samples using the QIAamp RNA Viral Kit (Qiagen, Heiden, Germany), and quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed using a commercial kit (Genrui-bio) targeting the S and N genes. The specimens were considered positive if the Ct value was  $\leq 38.0$ , and negative if the results were undetermined. Specimens with a Ct higher than 38 were repeated. The specimen was considered positive if the repeat results were the same as the initial result and between 38 and 40. If the repeat Ct was undetectable, the specimen was considered negative.

## **Indirect immunofluorescence assay (IFA)**

IFA was carried out as previously reported [53,54]. Vero cells were fixed in 4% formaldehyde at 48 hours post infection. Then cells were permeabilized in 0.5% Triton X-100, blocked in 5% BSA in PBS, and then probed with the plasma of this patient or healthy control at a dilution of 1:500 for 1 h at room temperature. The cells were washed three times with PBS and then incubated with either goat anti-human IgG conjugated with Alexa fluor 488 at a dilution of 1:500 for 1 h (Invitrogen). The cells were then washed and stained with hoechest-33342 (Invitrogen) to detect nuclei. Fluorescence images were obtained and analyzed using EVOS FL Auto Imaging System (Invitrogen).

## **Protein expression and purification**

The genes for nsp12 of SARS-CoV-2 isolate BetaCov/Wuhan/WH01/2019 (EPI\_ISL\_406798) was chemically synthesized with codon optimization for insect cells (*Spodoptera frugiperda*) by Synbio Technologies. The sequence was fused with a C-terminal thrombin cleavage site, a 6×His-tag and a 2×Strep-tag, and incorporated into pFastbac-1 plasmid. Recombinant protein was expressed with Hi5 cells at 27°C. Cells were harvested at 48 hpi (hour post infection) and resuspended in 25 mM HEPES pH 7.4, 1 M NaCl, 1 mM MgCl<sub>2</sub> and 2mM TCEP. An equal volume of the same buffer supplemented with 0.2% (v/v) Igepal CA-630 (Anatrace) was added and incubated at 4 °C for 10 min. Cells were lysed by sonication and the lysate was clarified by ultracentrifugation. Cleared lysates were passed through a 0.22-μm filter film before further purification. The protein was purified by tandem affinity chromatography and SEC.

## **Surface plasmon resonance (SPR) assay**

The affinities between nsp12 and drugs were measured at room temperature (r.t.) using a Biacore 8K system with CM5 chips (GE Healthcare). The nsp12 protein was immobilized on the chip with a concentration of 100 μg/mL (diluted by 0.1mM NaAc, PH 4.0).

Drug samples were prepared according to procedure 29264621AA of GE Healthcare Life Sciences. 1×PBS solution plus 5% DMSO and 0.005% p20 was used for running and diluting drugs. A blank channel of the chip was used as the negative control. Serial diluted drugs were then flowed through the chip surface. The LMW multi-cycle kinetics was analyzed with the Biacore 8K Evaluation Software (version 1.1.1.7442) and fitted with a 1:1 binding model.

## **Statistical analysis**

Data are presented as the mean ± SD (Standard Deviation). All analyses were performed using GraphPad Prism version 7.0 for Windows (GraphPad Software, San Diego California, USA). Data were subjected to statistical analysis by two-way ANOVA or two-tailed Student's t-test. The P values less than 0.05 were considered statistically significant.

## **DATA AVAILABILITY**

The structure for SARS-CoV-2 RdRp has been deposited in the Protein Data Bank with accession number 6M71, 6WTT. The template structure we used to construct RdRP-ligand complexes are deposited in the Protein Data Bank with accession number 3BR9).

## **Declarations**

### **Acknowledgement**

Thanks for providing the approved drug library by TargetMol. This work was partly supported by Strategic Priority CAS Project XDB38000000, the National Key Research and Development Program of China under Grant No. 2018YFB0204403, 2016YFB0201305 and 2019YFA0906100, National Science and Technology Major Project under Grant No. 2018ZX10101004, Shenzhen Basic Research Plan under Grant No. JCYJ20190807170801656, National Science Foundation of China under Grant no. U1813203 and 61433012; the National Natural Youth Science Foundation of China (Grant no. 31601028), the Shenzhen Basic Research Fund under Grant no. JCYJ20180507182818013, JCYJ20170413093358429 and GGF2017073114031767, the China Postdoctoral Science Foundation under Grant no. 2019M653132, CAS Key Lab under Grant no. 2011DP173015, and the SIAT Innovation Program for Excellent Young Researchers. We would also like to thank the funding support by the Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence, Youth Innovation Promotion Association, CAS to Yanjie Wei.

**Author contributions:** H.Z. conceived the original idea, designed and implemented the hybrid virtual screening method, performed the drug screening step, analyzed the results, and participated in manuscript preparation. Y.Y. and J.W. tested inhibitory activity with qRT-PCR, CCK-8 assay, indirect immunofluorescence assay (IFA), Time-of-addition experiment and contributed to the analysis of the results and to the writing of the manuscript. J.L. participated in manuscript preparation and supervised the experimental process. M.W. helped with the manuscript writing and tested the drug-RdRP binding with Surface plasmon resonance (SPR). S.K., N.J., M.H., M.L., H.Z., Y.P., X.R. and Y.P. helped with manuscript preparation. Y.W., Y.S., X.W., and Y.L. supervised the project, and corrected the manuscript.

**Competing interests:** The authors declare no competing interests.

## References

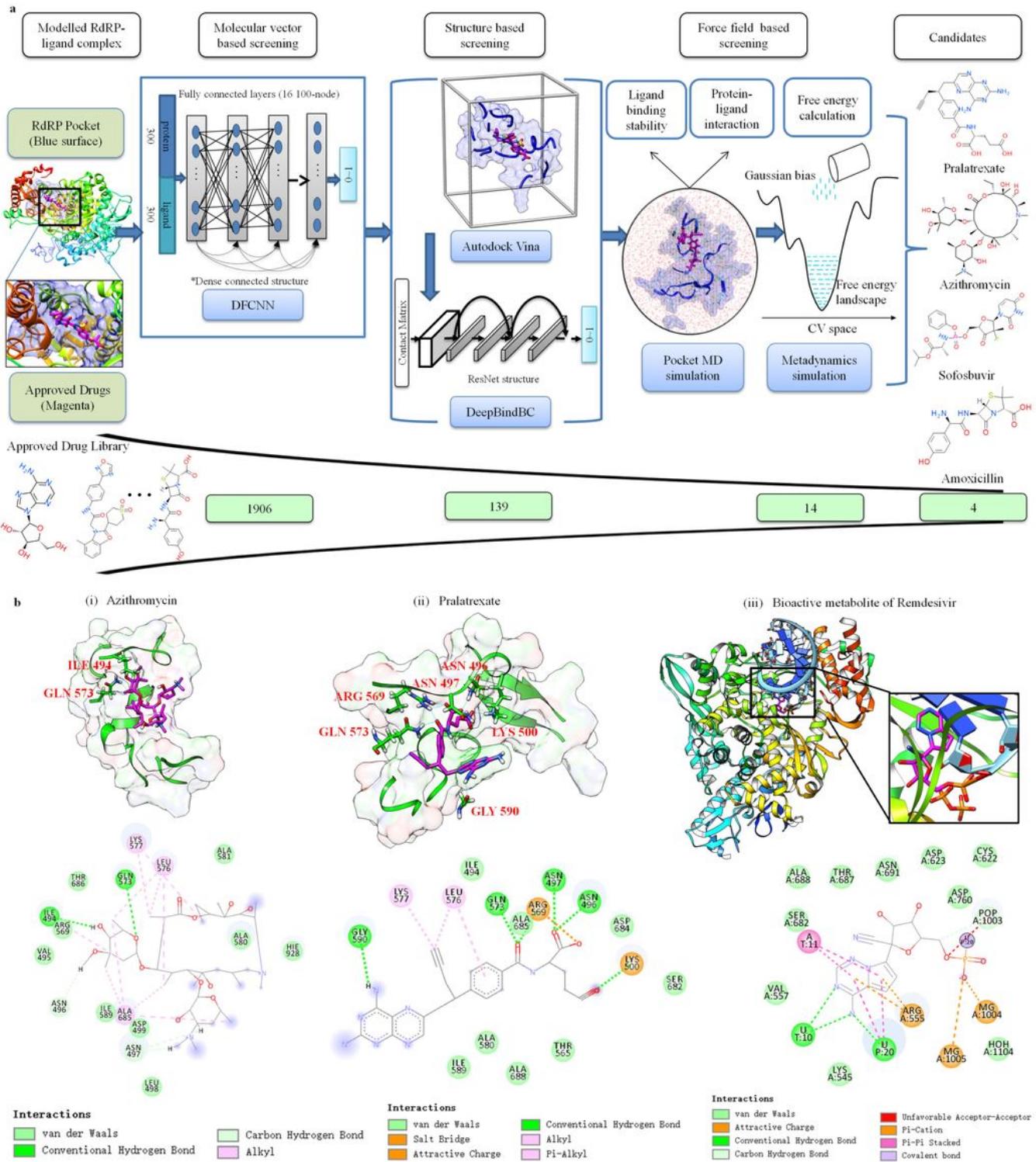
1. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020.
3. Xu J, Zhao S, Teng T, et al. Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV. *Viruses* 2020.
4. Qian G, Yang N, Ma AHY, et al. A COVID-19 Transmission within a family cluster by presymptomatic infectors in China. *Clin Infect Dis* 2020.
5. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* 2020.
6. Caly L, Druce JD, Catton MG, Jans DA, Wagstaff KM. The FDA-approved Drug Ivermectin inhibits the replication of SARS-CoV-2 in vitro. *Antiviral Res* 2020;104787.
7. Choy K-T, Wong AY-L, Kaewpreedee P, et al. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Res* 2020; **178**:104786.

8. Wang M, Cao R, Zhang L, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res* 2020.
9. Yin W, Mao C, Luan X, et al. Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science (80- )* 2020.
10. Grein J, Ohmagari N, Shin D, et al. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *N Engl J Med* 2020.
11. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 2019.
12. Wu C, Liu Y, Yang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B* 2020.
13. Gao Y, Yan L, Huang Y, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science (80- )* 2020:eabb7498.
14. Zhang H, Liao L, Cai Y, Hu Y, Wang H. IVS2vec: A tool of Inverse Virtual Screening based on word2vec and deep learning techniques. *Methods* 2019.
15. Brooks BR, III CLB, A. D. Mackerell J, et al. Autodock Vina. *J Comput Chem* 2009.
16. Gautret P, Lagier J-C, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents* 2020.
17. Gbinigie K, Frie K. Should azithromycin be used to treat COVID-19? A rapid review. *BJGP Open* 2020.
18. Bhatia HK, Singh H, Grewal N, Natt NK. Sofosbuvir: A novel treatment option for chronic hepatitis C infection. *J Pharmacol Pharmacother* 2014.
19. Faderl S, Ferrajoli A, Wierda W, et al. Clofarabine combinations as acute myeloid leukemia salvage therapy. *Cancer* 2008.
20. Zhang H, Liao L, Saravanan KM, Yin P, Wei Y. DeepBindRG: A deep learning based method for estimating effective protein-ligand affinity. *PeerJ* 2019.
21. Bernhard Ellinger, Denisa Bojkova, Andrea Zaliani, Jindrich Cinatl, Carsten Claussen, Sandra Westhaus, Jeanette Reinshagen, Maria Kuzikov, Markus Wolf, Gerd Geisslinger, Philip Gibbon SC. Identification of inhibitors of SARS-CoV-2 in-vitro cellular toxicity in human ( Caco-2 ) cells using a large scale drug repurposing collection CURRENT STATUS : UNDER REVIEW. :1–19.
22. Ito Y, Makita S, Tobinai K. Development of new agents for peripheral T-cell lymphoma. *Expert Opin Biol Ther* 2019.
23. Foss FM. Evaluation of the pharmacokinetics, preclinical and clinical efficacy of pralatrexate for the treatment of T-cell lymphoma. *Expert Opin Drug Metab Toxicol* 2011.
24. Kompis IM, Islam K, Then RL. DNA and RNA synthesis: Antifolates. *Chem Rev* 2005.
25. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008.
26. Zhou Y, Webber SE, Murphy DE, et al. Novel HCV NS5B polymerase inhibitors derived from 4-(1',1'-dioxo-1',4'-dihydro-1'λ6-benzo[1',2',4']thiadiazin-3'-yl)-5-hydroxy-2H-pyridazin-3-ones. Part 1: Exploration of 7'-substitution of benzothiadiazine. *Bioorganic Med Chem Lett* 2008; **18**:1413–1418.

27. Roy A, Yang J, Zhang Y. COFACTOR: An accurate comparative algorithm for structure-based protein function annotation. *Nucleic Acids Res* 2012.
28. Jaeger S, Fulle S, Turk S. Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition. *J Chem Inf Model* 2018.
29. Liu Z, Li Y, Han L, et al. PDB-wide collection of binding data: Current status of the PDBbind database. *Bioinformatics* 2015.
30. Morris GM, Ruth H, Lindstrom W, et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 2009.
31. Laio A, Gervasio FL. Metadynamics: A method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports Prog Phys* 2008.
32. Saleh N, Ibrahim P, Saladino G, Gervasio FL, Clark T. An Efficient Metadynamics-Based Protocol to Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *J Chem Inf Model* 2017.
33. Ruiz-Carmona S, Schmidtke P, Luque FJ, et al. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat Chem* 2017.
34. Hess B, Kutzner C, Spoel D Van Der. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem* 2008.
35. Hornak V, Simmerling C. Generation of accurate protein loop conformations through low-barrier molecular dynamics. *Proteins Struct Funct Genet* 2003.
36. Sousa Da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res Notes* 2012.
37. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 2006.
38. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983.
39. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem* 2005.
40. Darden T, York D, Pedersen L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J Chem Phys* 1993.
41. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A Linear Constraint Solver for molecular simulations. *J Comput Chem* 1997.
42. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. *Comput Phys Commun* 2014.
43. Williams T, Kelley C, Campbell J, et al. Gnuplot 4.6. *Softw Man* 2012.
44. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J Comput Chem* 2004.
45. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph* 1996.

46. Visualizer DS. v4. 0.100. 13345. *Accelrys Softw Inc* 2005.
47. Anon. icm\_browser\_pro. 2020.
48. Landrum G. RDKit: Open-source Cheminformatics. <Http://WwwRdkitOrg/> 2006.
49. Murtagh F, Contreras P. Algorithms for hierarchical clustering: An overview. *Wiley Interdiscip Rev Data Min Knowl Discov* 2012.
50. Yang Y, Wong G, Ye B, *et al*. Development of a reverse transcription quantitative polymerase chain reaction-based assay for broad coverage detection of African and Asian Zika virus lineages. *Virol Sin* 2017.
51. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. *Am J Epidemiol* 1938.
52. Zhao F, Yang Y, Wang Z, Li L, Liu L, Liu Y. The Time Sequences of Oral and Fecal Viral Shedding of Coronavirus Disease 2019 (COVID-19) Patients. *Gastroenterology* 2020.
53. Yang Y, Zhang L, Geng H, *et al*. The structural and accessory proteins M, ORF 4a, ORF 4b, and ORF 5 of Middle East respiratory syndrome coronavirus (MERS-CoV) are potent interferon antagonists. *Protein Cell* 2013.
54. Yang Y, Ye F, Zhu N, *et al*. Middle East respiratory syndrome coronavirus ORF4b protein inhibits type i interferon production through both cytoplasmic and nuclear targets. *Sci Rep* 2015.
55. Hunter S, Apweiler R, Attwood TK, *et al*. InterPro: The integrative protein signature database. *Nucleic Acids Res* 2009.

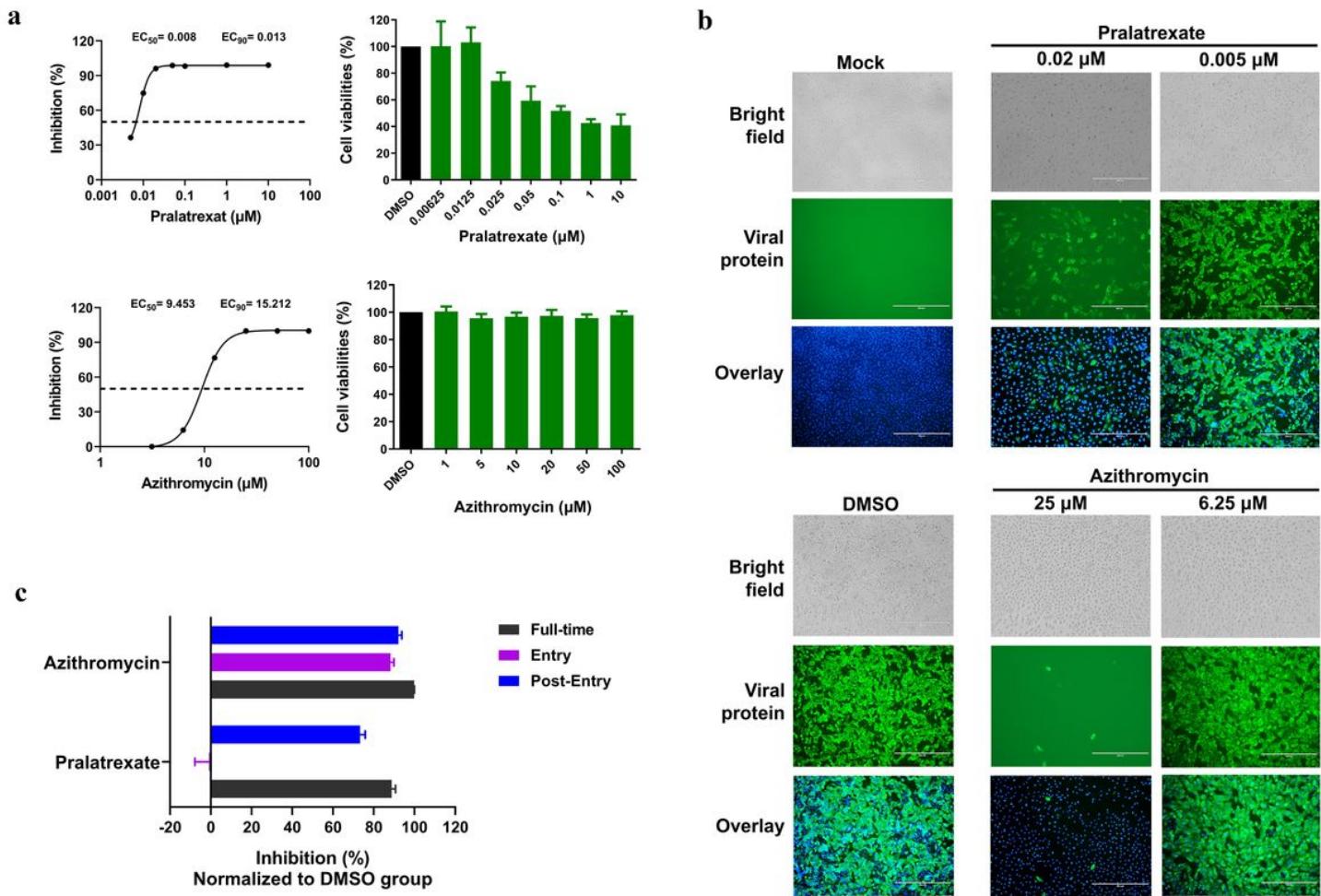
## Figures



**Figure 1**

Drug repurposing against RdRp for COVID-19 using a hybrid deep learning and molecular simulation strategy. a, 1906 approved drugs were subject to the proposed screening process which consists of molecular vector-based screening, structure-based screening and force field-based screening. DFCNN and DeepBindBC are both deep learning-based methods. 4 candidate drugs were selected by the proposed method, including Pralatrexate, Azithromycin, Sofosbuvir, Amoxicillin. b, Key interactions between the

studied drugs and RdRp from the last frame of MD simulation, for (i) Azithromycin and (ii) Pralatrexate. RdRp binding pocket is shown in green with surface representation and the corresponding drugs are shown in magenta. The 2D Schematic diagram of drug-RdRp interaction is given bottom, and neighbor residues (within 4 Å of the drug) are shown. b(iii), the experimental structure of Remdesivir in its monophosphate form with RdRp (PDB ID 7BV2), the 2D Schematic diagram of the interaction was also shown.



**Figure 2**

The antiviral activities of the test drugs against SARS-CoV-2 in vitro. a, Vero cells were infected with SARS-CoV-2 at an MOI of 0.02 in the presence of the indicated concentrations of the tested drugs for 48 hours. The viral yield in the cell supernatant was then quantified by qRT-PCR. Meanwhile, cytotoxicity of these drugs to Vero cells was measured by CCK-8 assay. b, Immunofluorescence microscopy of virus infection upon treatment of Pralatrexate and Azithromycin at the indicated concentrations. IFA was performed at 48 hours post-infection. Scale bar, 100 μm. Cells were immunostained for the Viral protein (green) and DNA (blue). c, Time-of-addition experiment of Pralatrexate and Azithromycin. Cells were infected with SARS-CoV-2 at an MOI of 0.02 with different treatment, and virus yield in the infected cell supernatants was quantified by qRT-PCR. For the group of "Entry", the drugs were added to the cells for 1 hour before viral attachment, and at 2 hours post-infection, the virus-drug mixture was replaced with fresh

culture medium. For the group of “Post-entry”, drugs were added at 2 hours post viral infection, and maintained until the end of the experiment. For the group of “Full-time”, Vero cells were pre-treated with the drugs for 1 hour, and the virus was then added to allow attachment for 2 hour. Afterwards, the virus-drug mixture was removed, and the cells were cultured with drug-containing medium until the end of the experiment.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryinformation7132020.pdf](#)
- [formulas.docx](#)