



An Efficient Modern Strategy to Screen Drug Candidates Targeting RdRp of SARS-CoV-2 With Potentially High Selectivity and Specificity

Haiping Zhang^{1*†}, Xiaohua Gong^{2†}, Yun Peng², Konda Mani Saravanan³, Hengwei Bian⁴, John Z. H. Zhang¹, Yanjie Wei⁵, Yi Pan⁵ and Yang Yang^{2*}

OPEN ACCESS

Edited by:

Simone Brogi,
University of Pisa, Italy

Reviewed by:

Giuseppe Felice Mangiatordi,
Institute of Crystallography, National
Research Council (IC-CNR), Italy
Wagdy Mohamed Eldehna,
Kafrelsheikh University, Egypt
Michele Saviano,
Institute of Crystallography, National
Research Council (IC-CNR), Italy

*Correspondence:

Yang Yang
yyszth2018@163.com
Haiping Zhang
hp.zhang@siat.ac.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Medicinal and Pharmaceutical
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 05 May 2022

Accepted: 06 June 2022

Published: 12 July 2022

Citation:

Zhang H, Gong X, Peng Y,
Saravanan KM, Bian H, Zhang JZH,
Wei Y, Pan Y and Yang Y (2022) An
Efficient Modern Strategy to Screen
Drug Candidates Targeting RdRp of
SARS-CoV-2 With Potentially High
Selectivity and Specificity.
Front. Chem. 10:933102.
doi: 10.3389/fchem.2022.933102

¹Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, ²Shenzhen Key Laboratory of Pathogen and Immunity, National Clinical Research Center for Infectious Disease, State Key Discipline of Infectious Disease, Shenzhen Third People's Hospital, Second Hospital Affiliated to Southern University of Science and Technology, Shenzhen, China, ³Department of Biotechnology, Bharath Institute of Higher Education and Research, Chennai, India, ⁴Shanghai Engineering Research Center of Molecular Therapeutics and New Drug Development, Shanghai Key Laboratory of Green Chemistry and Chemical Process, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai, China, ⁵Center for High Performance Computing, Joint Engineering Research Center for Health Big Data Intelligent Analysis Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Desired drug candidates should have both a high potential binding chance and high specificity. Recently, many drug screening strategies have been developed to screen compounds with high possible binding chances or high binding affinity. However, there is still no good solution to detect whether those selected compounds possess high specificity. Here, we developed a reverse DFCNN (Dense Fully Connected Neural Network) and a reverse docking protocol to check a given compound's ability to bind diversified targets and estimate its specificity with homemade formulas. We used the RNA-dependent RNA polymerase (RdRp) target as a proof-of-concept example to identify drug candidates with high selectivity and high specificity. We first used a previously developed hybrid screening method to find drug candidates from an 8888-size compound database. The hybrid screening method takes advantage of the deep learning-based method, traditional molecular docking, molecular dynamics simulation, and binding free energy calculated by metadynamics, which should be powerful in selecting high binding affinity candidates. Also, we integrated the reverse DFCNN and reversed docking against a diversified 102 proteins to the pipeline for assessing the specificity of those selected candidates, and finally got compounds that have both predicted selectivity and specificity. Among the eight selected candidates, Platycodin D and Tubeimoside III were confirmed to effectively inhibit SARS-CoV-2 replication *in vitro* with EC₅₀ values of 619.5 and 265.5 nM, respectively. Our study discovered that Tubeimoside III could inhibit SARS-CoV-2 replication potently for the first time. Furthermore, the underlying mechanisms of Platycodin D and Tubeimoside III inhibiting SARS-CoV-2 are highly possible by blocking the RdRp cavity according to our screening procedure. In addition, the careful analysis predicted common critical residues involved in the binding with active inhibitors Platycodin D and Tubeimoside III, Azithromycin, and Pralatrexate, which hopefully promote the development of non-covalent binding inhibitors against RdRp.

Keywords: antiviral agent, RdRp, compound specificity estimation, virtual screening, tubeimoside III

INTRODUCTION

RdRp Is an Important Therapeutic Target of SARS-CoV-2

The global pandemic caused by SARS-CoV-2 has continued for more than 2 years and caused a huge threat to public health and the global economy. Variants carrying numerous mutations in the spike protein of SARS-CoV-2, which result in higher transmissibility and immune evasion of the current vaccines, and therapeutic monoclonal antibodies emerged during the transmission (Tao et al., 2021; Flemming, 2022). Unlike S protein, the RdRp of SARS-CoV-2 is highly conserved and plays a critical role in the virus replications, making it a potential therapeutic target to combat SARS-CoV-2 (V'kovski et al., 2021). There are already remdesivir (Yin et al., 2020) and molnupiravir (Lee et al., 2021) drug that target RdRp, and our previous work also identify Pralatrexate and Azithromycin which may target RdRp and shows strongly block virus replication in cells (Zhang et al., 2020b).

Drug Virtual Screening Has Been Accelerated by Deep Learning

Drug virtual screening has a very long history, with many related techniques have been developed, such as docking, Quantitative Structure-Activity Relationship (QSAR), pharmacophore, and structure-based ligand similarity. As emerging of deep learning algorithms and the experimental protein-ligand interaction dataset accumulates, deep learning-based protein-ligand interaction will greatly promote virtual drug screening. Currently, there are graphic-based protein-ligand interaction models that use a graphic representation of ligand or protein and can grasp the spatial and physical-chemical feature concisely; also, the graphic convolution network was used as training architecture (Torng and Altman, 2019). Also, a 4D CNN-based protein-ligand interaction model directly uses the X, Y, and Z coordinates of a protein-ligand complex, plus the extra atom feature as input (Stepniewska-Dziubinska et al., 2018). There are many other types of deep learning-based protein-ligand interaction prediction models, such as DeepDTAF (Wang et al., 2021), DeepBindRG (Zhang et al., 2019b), and DeepAffinity (Karimi et al., 2019). We also developed several protein-ligand interaction models, and some of them have successfully been applied to virtual screening applications for targets such as RdRp (Zhang et al., 2020b), 3C proteases (Zhang et al., 2020a), and TIPE2 (Zhang et al., 2021). For those protein-ligand binding prediction models, some predict binding affinity, whereas some are binary predictions of the binding possibility. Since the number of unbinding compounds usually dominates over the binding compounds in the virtual screening scenarios, binary prediction models that have considered the unbinding data during the training are more suitable for virtual screening.

Current Virtual Screening Problem

A more effective integrated virtual screening strategy should be proposed. Deep learning-based protein-ligand prediction,

molecular docking, and molecular dynamics simulation can be used for virtual screening. Each has different pros and cons in terms of efficacy and accuracy. However, the virtual screening database can be very large since the potential druglike compound size can be large as 10^{60} . Hence, maintaining the efficiency and accuracy of the virtual screening is essential. The virtual drug screening can also be used for *de novo* compounds by integrating the deep learning-based *de novo* compounds generative model (Gupta et al., 2018). Many more advanced deep learning generative models based on RNN architecture have been gradually developed (Moret et al., 2021; Creanza et al., 2022). Among them, the beam search algorithm proposed by Michael Moret, et al., has great potential to generate more reasonable novel compounds (Moret et al., 2021). Beam search sampling overcomes the need for external scoring methods and extends the applicability of machine learning-driven molecular design. And they have tried to better adapt them to VS procedures and take into account some of VS limitations. Also, a new strategy called sampling with substitutions (SWS) usually generates molecules structurally similar to bioactive compounds or with given desired properties (Creanza et al., 2022). It generates molecules structurally similar to bioactive compounds or with given desired properties. It partly solves the problem that exploring a new chemical space could make the following synthesis difficult and expensive. Those newly developed methods also have the potential to preserve drug-likeness and synthetic accessibility of the generated *de novo* compounds. They can help discover *de novo* compounds with better binding affinity or other desired properties, such as better specificity or drug-likeness.

Furthermore, the current virtual screening often has not fully considered compounds' specificity. Without specificity, there is a risk of potential side effects and cause the drug failure in a later stage. Currently, most virtual screening procedures don't consider specificity, which can lead to obtaining pain compounds that have side effects and fail to pass the clinic test. Some researchers proposed a local-specific test strategy (Yan and Wang, 2012). However, estimating the given compound's binding affinity with many diversified targets is necessary to test the relative global specificity. Hence, assessing the specificity of the compound by computational methods efficiently requires high-speed protein-ligand prediction tools (such as DFCNN) in the first stage. If a compound can bind a target that has significantly high affinity compared to other proteins, then such binding can be defined as high specificity and high affinity. Contrary, if a compound can bind to many proteins tightly, then it has a risk of low specificity.

Our Current Work

This work identifies candidates from three TargetMol datasets (Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library) that can potentially bind to a larger range of RdRp cavity. The hybrid pipeline has effectively integrated deep learning-based method (DFCNN, DeepBindBC), docking (Autodock Vina), specificity checking (through reverse DFCNN, reverse docking against 102 representative proteins from DUDE), and force field-based screening (pocket MD simulation and

metadynamics simulation). In a step-by-step manner, the high binding affinity compounds are accumulated while the specificity is also considered. The proposed pipeline enables the screening of large compounds dataset with high efficiency while increasing the chance to obtain high affinity and high specificity compounds. With the proposed screening strategy, we obtained eight candidates of SARS-CoV-2 for experimental validation. Finally, we identified that Platycodin D and Tubeimoside III could inhibit SARS-CoV-2 replication *in vitro* with IC₅₀ values of 619.5 and 265.5 nM, respectively. Unlike our previous work, we used a large pocket definition and considered the specificity during the screening; the obtained active compounds should be more specific and more diversified.

METHODS

The Target and Compounds Database

Our previous work obtained the RdRp-ligand complex (Zhang et al., 2020b). In that work, the RdRp sequence and its modeled structure were obtained from <https://zhanglab.ccmb.med.umich.edu/C-I-TASSER/2019-nCov/>. The RdRp-ligand model was constructed by I-TASSER (Zhang, 2008), whose ligand was taken from the template protein (PDB ID: 3BR9) (Zhou et al., 2008) by COFACTOR algorithm (Roy et al., 2012) within I-TASSER using structure comparison and protein-protein networks. Unlike previous work, we extract the amino acids within 1.2 nm of the ligand as the binding pocket instead of 1 nm since the RdRp cavity is very large. The three TargetMol datasets (Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library) were used as virtual screening libraries. These datasets have diversified compounds including active compounds, natural compounds, and approved compounds. Moreover, most of the compounds can be easily pursued, we deposited the compound structures in GitHub (https://github.com/haiping1010/targetmol_datasets) for the convenience of other users.

Molecular Vector-Based and Structure-Based Drug Screening

Similar to our previous work (Zhang et al., 2020b; Zhang et al., 2021), we have used a hybrid screening strategy that first uses molecular vector-based as initial drug screening and structure-based screening for further selection. The molecular vector-based relies on our previously developed DFCNN (Zhang et al., 2019a). The parameter setting for the DFCNN is the same as our previous work, which is drug repurposing against RdRp of SARS-CoV-2 (Zhang et al., 2020b), except that this time we used a large pocket definition, which defines residues with 1.5 nm from the predicted ligand as a pocket. The structure-based screening depends on DeepBindBC and autodock vina (Trott and Olson, 2010). The DeepBindBC prediction relies on the interface information of docked protein-ligand interface. By incorporating the cross-docking (docking proteins and ligands from different experimental complexes) conformation as negative training

data, DeepBindBC can distinguish non-binders. The pocket is determined by the location of the ligand in the template protein. We set the cavity volume space with 3.5, 3.5, and 3.5 nm in *x*, *y*, and *z* dimensions from the pocket mass center. Other settings are the same as our previous work (Zhang et al., 2020b, 2021).

The Reverse DFCNN Prediction and Reverse Docking to Access the Specificity of the Selected Compounds

To fast estimate the specificity for large amounts of compounds, we first use the DFCNN to make the reverse prediction against 102 proteins from DUD.E. We have defined a function to estimate the DFCNN-based specificity. The formula is used as follows:

$$\text{specificity} = \log_{10}(103 / (N_{c1} + 1))$$

where N_{c1} is the counted number of compounds that have a DFCNN score larger than 0.99 during the reverse DFCNN prediction with 102 protein targets.

However, the DFCNN has not considered the spatial information; hence we carried a reverse docking by Autodock Vina for further relative specificity. The relative specificity is calculated by following formulas.

$$\text{Relative specificity} = \log_{10}(103 / (N_{c2} + 1))$$

where N_{c2} is the counted number of compounds that have an autodock vina score smaller than the mean autodock vina score of 102 protein-compound complexes during the reverse autodock vina docking with 102 protein targets. Notably, the smaller autodock vina score indicates better binding affinity.

Force Field-Based Screening

Further drug screening was carried out by pocket molecular dynamics (MD) simulations and metadynamics similar to our previous work (Zhang et al., 2020b; Zhang et al., 2021), except that we used a large pocket definition (residues with 1.5 nm from the docked compounds was kept as pocket). This study selected 26 compound binding complexes, which were predicted candidates by previous deep learning screening and specificity checking, for MD simulation. Metadynamics simulations can estimate binding free energy surfaces to explore whether protein-ligand prefers to bind in solution. Metadynamics simulations facilitate the sampling of the free energy landscape and a specific collective variable of interest by adding a history-dependent biasing potential (Laio and Gervasio, 2008; Saleh et al., 2017). The detailed procedure of pocket molecular dynamics and metadynamics simulation was described as follows.

The initial protein-compound complexes were from the top score conformation Autodock Vina docking, the ligand was edited by pymol software (DeLano, 2002) to make it in the correct protonation state at pH 7. In this study, we selected 26 compound binding complexes that have a high DFCNN score (larger or equal to 0.99), low Vina score (low than -9 kcal/mol), large DeepBindBC score (larger or equal than 0.99), and especial high reverse DFCNN based specificity (larger or equal than 1.3)

and reverse Autodock Vina based relative specificity (larger or equal than 0.6). We also refined a pocket molecular dynamics simulation (pocket MD, **Supplementary Figure 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 the addition of a bias potential to sample the free energy landscape along with a specific collective variable of interest (Laio and Gervasio, 2008; Saleh et al., 2017). Note that the binding free energy calculations from Metadynamics may only be suitable for detecting 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 (Zhang et al., 2020b), which is inspired by a previous dynamic undocking (DUck) method (Ruiz-Carmona et al., 2017). An in-house script was used to extract the pocket region of the protein (here, we used 1.2 nm within the binding ligand), the N terminal and C terminal ends were capped with the ACE and NHE terminals, respectively. We applied position restrains to the ACE and NHE terminals to maintain the relative conformation of the pocket. MD simulation was carried out by Gromacs with AMBER-99SB force field (Hornak and Simmerling, 2003; Hess et al., 2008). The topology of the ligand and the partial charges of ligand was generated by ACPYPE (Sousa Da Silva and Vranken, 2012), which relies on Antechamber (Wang et al., 2006). First, 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 (Jorgensen et al., 1983), the counter ions were added to neutralize the total charge using the Gromacs program tool (Van Der Spoel et al., 2005). The long-range electrostatic interactions under the periodic boundary conditions was calculated with Particle Mesh Ewald approach (Darden et al., 1993). 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 (Hess et al., 1997).

We performed the energy minimization steps with a step-size of 0.001 ns, 100 ps simulation with isothermal-isovolumetric ensemble (NVT), and 10 ns simulation with isothermal-isobaric ensemble (NPT) for water equilibrium. After that, a 100 ns NPT production run (step size 2 fs) was carried out. The Parrinello-Rahman barostat 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 a collective variable (CV). The protein-ligand interface coordination numbers correlate with the numbers of atom contact, and larger coordination number usually indicates that the protein-ligand is in a binding state.

The coordination number C is defined as follows by Plumed:

$$C = \sum_{i \in A} \sum_{j \in B} S_{ij} \quad (1)$$

and

$$S_{ij} = \frac{1 - \left(\frac{r_{ij}-d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij}-d_0}{r_0}\right)^m} \quad (2)$$

In the simulation, n was 6, m was 12, d_0 was 0 nm and r_0 was 0.5 nm. d_0 is a parameter of the switching function. r_{ij} is the distance between atom i and atom j . The degrees of contacts between two groups of atoms can be estimated by the above function (1) (Tribello et al., 2014). Metadynamics simulation for each protein-ligand system was performed for 100 ns (except protein-Azithromycin, which was extended to 300 ns in order to reach the 0 Coordination Number and achieve convergences). During the metadynamics simulation, Gaussian values were deposited every 1 ps with a height of 0.3 kJ/mol. The widths of the Gaussians were 5 for the coordination number. The free energy landscapes of the metadynamics simulations along the CV were generated by the Plumed program and plotted using Gnuplot (Williams et al., 2012).

Tools Used in the Analysis

The USCF Chimera, VMD, ICM-browserPro, and Discovery Studio Visualizer 2019 were used to generate the structure and visualize the 2D protein-ligand interactions (Humphrey et al., 1996; Pettersen et al., 2004; BIOVIA, 2005; icm_browser_pro, 2020).

Viral Stock Titration by 50% Tissue Culture Infective Dose

TCID₅₀ was measured as previously reported (Yang et al., 2022). Briefly, Vero cells in 96-well plates were grown to 80% confluence and infected with 10-fold serial dilutions of the stock SARS-CoV-2 (hCoV-19/China/SZTH-025/2021, GISAID No. EPI_ISL_11799984) for 1 h at 37°C. Then the inoculum was removed, and cells were overlaid with fresh DMEM plus 2% FBS. Plates were assessed for the lowest dilution in which 50% of the wells exhibited cytopathic effects on the fifth-day post-infection (d.p.i). The values of TCID₅₀ were calculated according to the Reed-Muench method (Reed and Muench, 1938).

Evaluating the Antiviral Activities of the Candidate Drugs in Vero Cells

The antiviral activities of the drugs were evaluated as previously reported with some modifications (Zhang et al., 2020b). Vero cells were seeded at 4×10^4 cells per well in 24-well plates and allowed to adhere for 24 h, the virus (MOI \approx 0.02) and different doses of the indicated drugs were added to allow infection for 1 h at 37°C. Then viral inoculum was removed, and cells were washed 2 times with PBS. Then the cells were further cultured with fresh DMEM with 2% FBS and the indicated concentrations of drugs. At 48 h 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

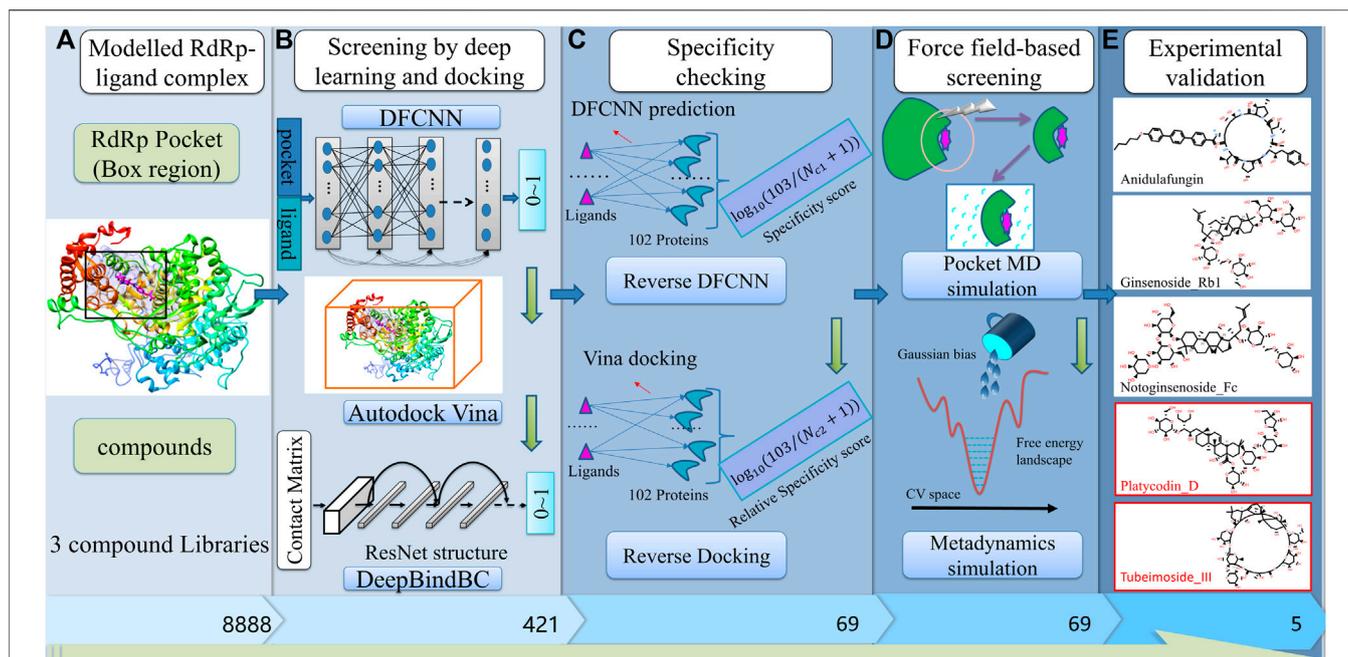


FIGURE 1 | The pipeline of the virtual screening procedure against RdRp. **(A)** build RdRp-ligand model, and prepare three compounds libraries (Targetmol-Approved_Drug_Library, Targetmol-Natural_Compound_Library, and Targetmol-Bioactive_Compound_Library); **(B)** Deep learning and docking-based screening by DFCNN, Autodock Vina and DeepBindBC; **(C)** Specificity checking by reverse DFCNN and reverse docking against 102 diversified targets from DUD.E dataset; **(D)** Fine screening by pocket MD simulation and Metadynamics simulation; **(E)** Experimental validation of final selected candidates.

using quantitative reverse transcription-polymerase chain reaction (qRT-PCR) was performed using a commercial kit (Mabsky Biotech Co., Ltd.). All the experiments involving infectious SARS-CoV-2 were handled in BSL-3 facilities at the Shenzhen Third People's Hospital. The dose-response curves were plotted from viral RNA copies versus the drug concentrations using GraphPad Prism 8 software.

Cytotoxicity Testing Assay

The cytotoxicity of Platycodin D and Tubeimoside III on Vero cells was evaluated by the Cell Counting Kit-8 (CCK-8, Beyotime Biotech) according to the manufacturer's protocols. In brief, Vero cells grown in 96-well plates were treated with various concentrations of compounds or mock-treated for 24 h, followed by 4 h incubation in the media supplemented with 10% CCK-8. Viable cells were counted according to the absorbance at 450 nm.

RESULTS

The overall workflow is illustrated in **Figure 1**, which is composed of several stages, including pocket determination and compound libraries preparation. (**Figure 1A**); preliminary screening by deep learning and docking (**Figure 1B**); specificity checking by Reverse DFCNN and Reverse Docking; Fine screening by pocket MD and metadynamics simulation; most important, we experimentally validate the candidates in the final step. Notably, among eight compounds that were finally obtained from our hybrid screening

strategy, two compounds (Platycodin D and Tubeimoside III) are active in inhibiting the SARS virus with IC₅₀ values of 619.5 nM and 265.5 nM, respectively. Overall, the hybrid virtual screening procedure is designed for compounds with a high chance of binding and high specificity against a given target.

The obtained candidates from preliminary screening are shown in **Table 1**. Since the pipeline methods have different advantages and are often complementary to each other, for instance, the DFCNN achieves extremely high efficiency by only considering the physical-chemical feature of pocket and ligand through vector representation. At the same time, the Autodock Vina and DeepBindBC have also considered spatial information. Hence, we selected those compounds based on their high DFCNN score (larger or equal to 0.99), low Vina score (low than -9 kcal/mol), large DeepBindBC score (larger or equal than 0.99), and especial high reverse DFCNN based specificity (larger or equal than 1.3) and reverse Autodock Vina based relative specificity (larger or equal than 0.6). The specificity check aims to eliminate those compounds that may bind to multiple off-targets, hence guaranteeing the relative safety of the candidates in the late stage of drug development.

The Final Candidates From MD Simulation and Metadynamics Simulation

The 26 candidates from the preliminary screening were considered for 100 ns pocket MD simulations. Among them, we focused on the eight candidates that showed stability in the last 50 ns simulation, which fulfilled

TABLE 1 | The result list with corresponding prediction scores after gradually screening by DFCNN, Autodock Vina, DeepBindBC, DFCNN specificity checking, and Autodock Vina relative specificity checking.

Name	DeepBindBC	Vina (kcal/mol)	DFCNN	Sp1*	C1*	Sp2*	C2*
Tubeimoside_I	1.000	-12.8	0.996	1.708	1	2.013	0
Tubeimoside_III	1.000	-14.4	0.991	1.708	1	2.013	0
Anidulafungin	0.999	-10.8	0.990	1.531	2	1.708	1
Ciwujianoside-B	1.000	-12.3	0.994	1.531	2	1.230	5
Platycodin_D	1.000	-12.2	0.997	1.398	3	1.146	6
Chikusetsusaponin_IV	1.000	-11.5	0.998	1.398	3	1.041	8
Polygalasaponin_V	1.000	-12.4	0.998	1.531	2	1.041	8
Officinalisinin_I	1.000	-11.3	0.993	1.708	1	1.000	9
Ginsenoside_Rb1	1.000	-10.6	0.990	1.708	1	0.954	10
Protodioscin	1.000	-11.4	0.995	1.398	3	0.954	10
Dipsacoside_B	1.000	-11.6	0.997	1.531	2	0.903	11
Deapi-platycodin_D	1.000	-11.6	0.996	1.398	3	0.845	13
Ginsenoside_Ro	1.000	-11.7	0.997	1.531	2	0.845	12
Pulsatilla_saponin_D	1.000	-12	0.998	1.398	3	0.845	12
anemarsaponin_B	1.000	-10.9	0.993	1.531	2	0.778	16
Esculentoside_H	1.000	-10.8	0.997	1.531	2	0.778	14
Timosaponin_BII	1.000	-10.7	0.993	1.708	1	0.778	14
Astragaloside_IV	1.000	-10.9	0.997	1.398	3	0.699	19
Eleutheroside_E	1.000	-9.2	0.992	1.531	2	0.699	19
Ginsenoside_Rd	1.000	-10.3	0.993	1.398	3	0.699	18
Asiaticoside	1.000	-11.1	0.998	1.398	3	0.602	22
Ginsenoside_Rc	1.000	-10.4	0.992	1.708	1	0.602	20
Hederacoside_D	1.000	-11.2	0.997	1.531	2	0.602	22
Notoginsenoside_Fc	1.000	-10.1	0.990	1.708	1	0.602	22
Rebaudioside_A	1.000	-10.9	0.995	1.398	3	0.602	24
Syringaresinol-di-O-glucoside	1.000	-9.2	0.992	1.531	2	0.602	24

Sp1: DFCNN specificity.

C1: the number of DFCNN scores in the predictions against 102 protein targets larger or equal than 0.99.

Sp2: Vina relative specificity.

C2: the number of vina scores in the docking against 102 protein targets is larger or equal to the known compound-RdRp vina score.

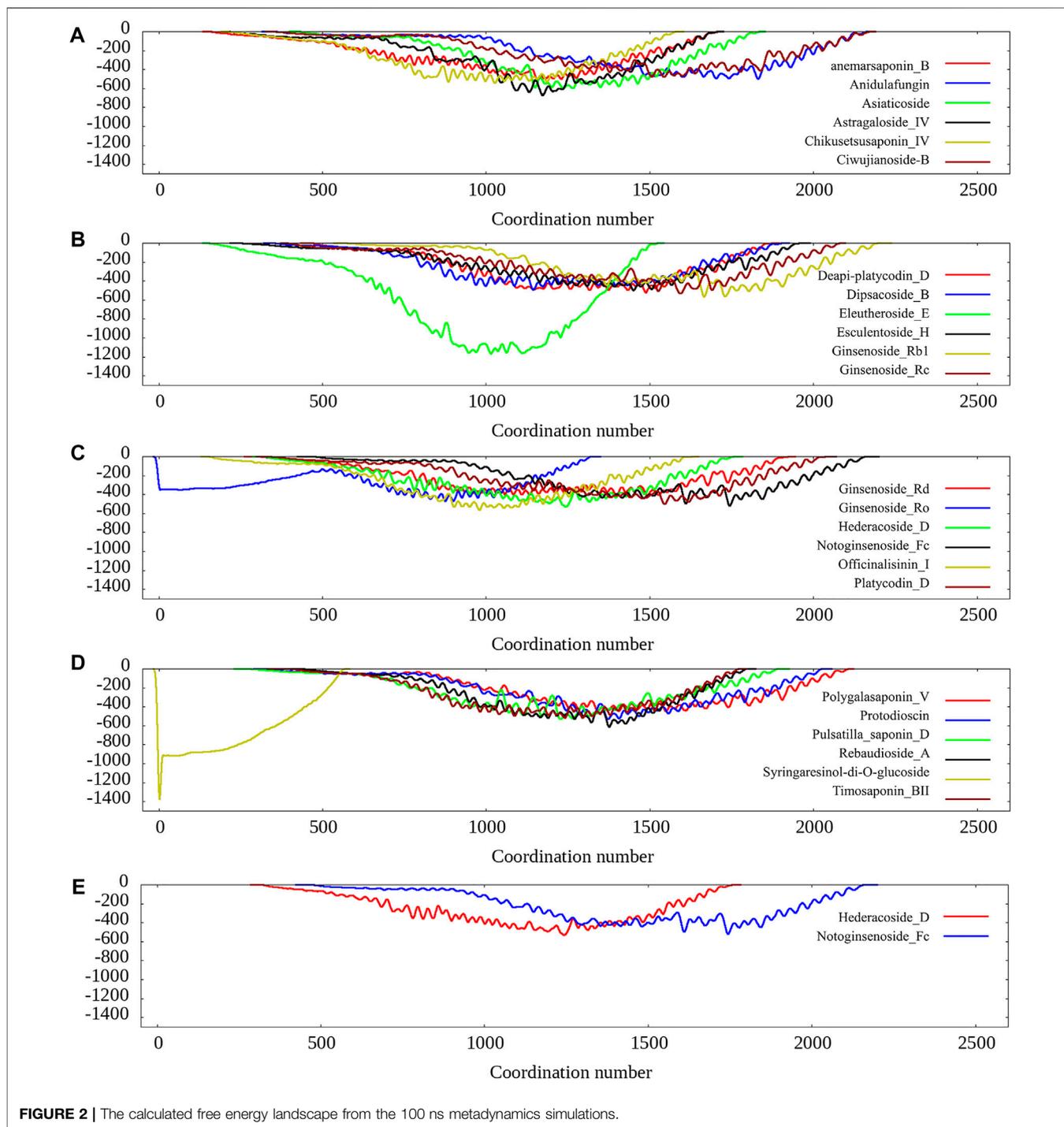
conditions of average RMSD < 0.5 nm, standard deviation < 0.1 nm or RMSD < 0.6 nm, and standard deviation < 0.5 nm. The RMSD values of these eight pocket-compound complexes are shown in **Supplementary Figure S1A**. Furthermore, we find many hydrogen bonds formed between the pocket and ligand of these eight candidates, which show in **Supplementary Figure S1B**. Among the eight compounds, the Anidulafungin is an approved drug, and recent reports have shown that Anidulafungin has anti-SARS-nCov-2 active with IC₅₀ of 4.64 μM (Jeon et al., 2020), and other researcher's docking results support that it may bind to NSP12 (RdRp) (Dey et al., 2021). The other seven compounds are mostly from TCM herbal ingredients; for instance, the Ginsenoside_Rb1 and Ginsenoside_Rd can be isolated from ginseng (Lu et al., 2009), and Notoginsenoside_Fc is an ingredient from the leaves of Panax notoginseng (Yang et al., 1983).

We also carried 100 ns metadynamics for the 26 pocket-ligand complexes to further evaluate the ligand-pocket binding free energy landscape. The results show that most of the protein-ligands prefer to bind except for a few cases such as Syringaresinol-di-O-glucoside, Ginsenoside_Ro, as shown in **Figure 2**. Since the eight compounds selected in the previous stage all show preferring to bind according to the free energy landscape, we choose them as the final candidates for experimental validation.

To explore the detailed interaction between pocket and ligand, we have analyzed their interaction in atomic details. We noticed that Anidulafungin and Tubeimoside_III are compounds that contain a large cyclic substructure, which is consistent with our previous work that Azithromycin is an active inhibitor against SARS-nCov-2 that contain cyclic structure (Zhang et al., 2020b), and highly possible through binding to RdRp through our screening prediction. The interaction patterns of these two candidates from the last frame of MD simulation are shown in **Figure 3**.

Other three candidates with ligand average RMSD < 0.5 nm and standard deviation < 0.1 nm during the last 50 ns MD simulation are Ginsenoside_Rb1, Notoginsenoside_Fc, and Platycodin_D. Their interaction with the RdRp pocket from the last frame conformation of the 100 ns MD simulation is shown in **Supplementary Figure S2**. We observed that these three candidates contain glucan substructures and form many hydrogen bonds with the RdRp. Compared with our previous screening work, it is the first time we discovered that glucan substructure might be able to play an essential role in interacting with RdRp large pocket.

With different selecting criteria, we may obtain some other potential drug candidates as well. Here, we tried the criteria of ligand average RMSD < 0.6 nm and standard deviation < 0.5 nm in the last 50 ns MD simulation and obtained the other three candidates, which are shown in **Supplementary Figure S3**.

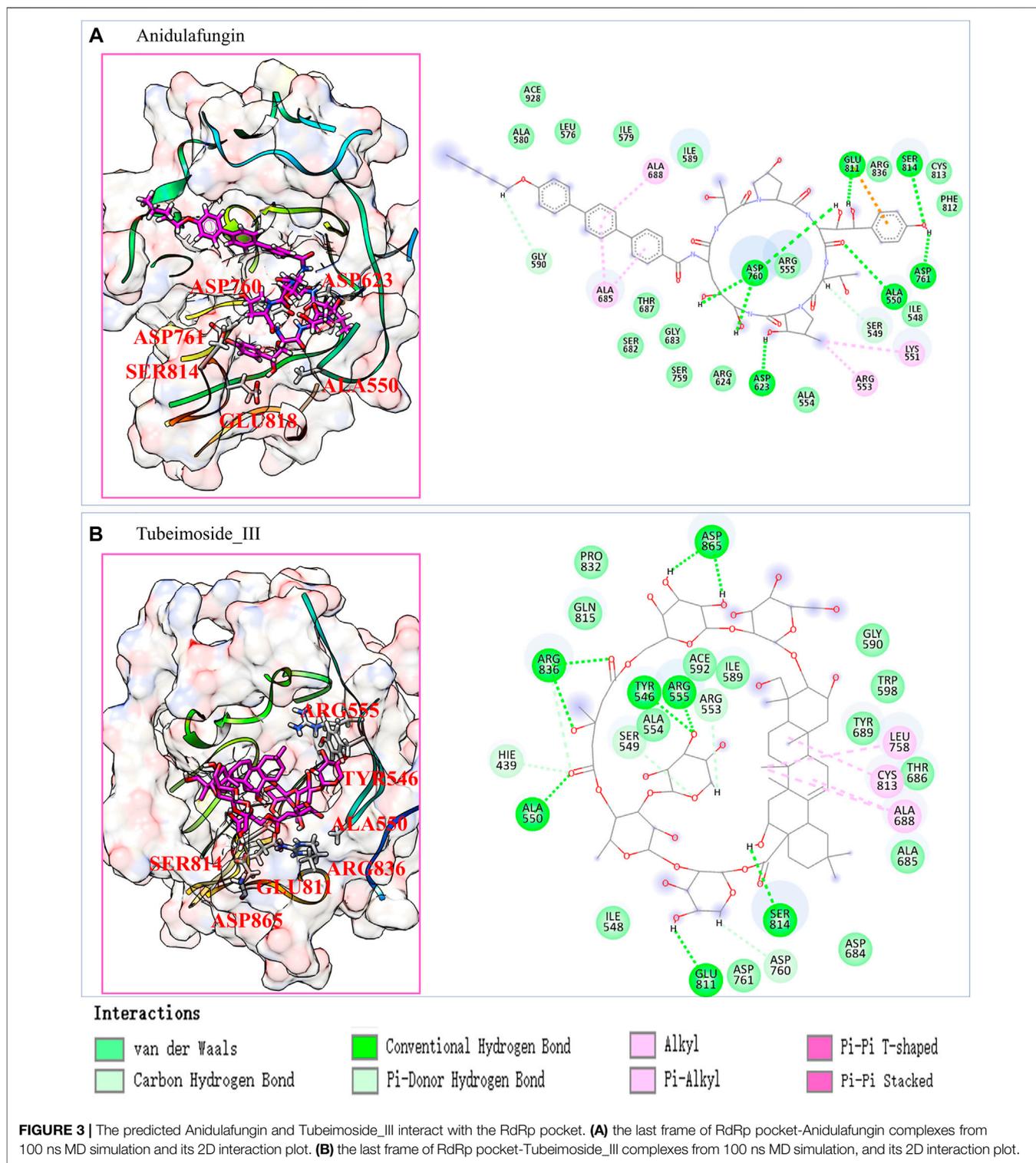


Similarly, they also contain a glucan substructure, and all form many hydrogen bonds with the RdRp pocket.

Platycodin_D and Tubeimoside_III Inhibit the Replication of SARS-CoV-2 *In Vitro*

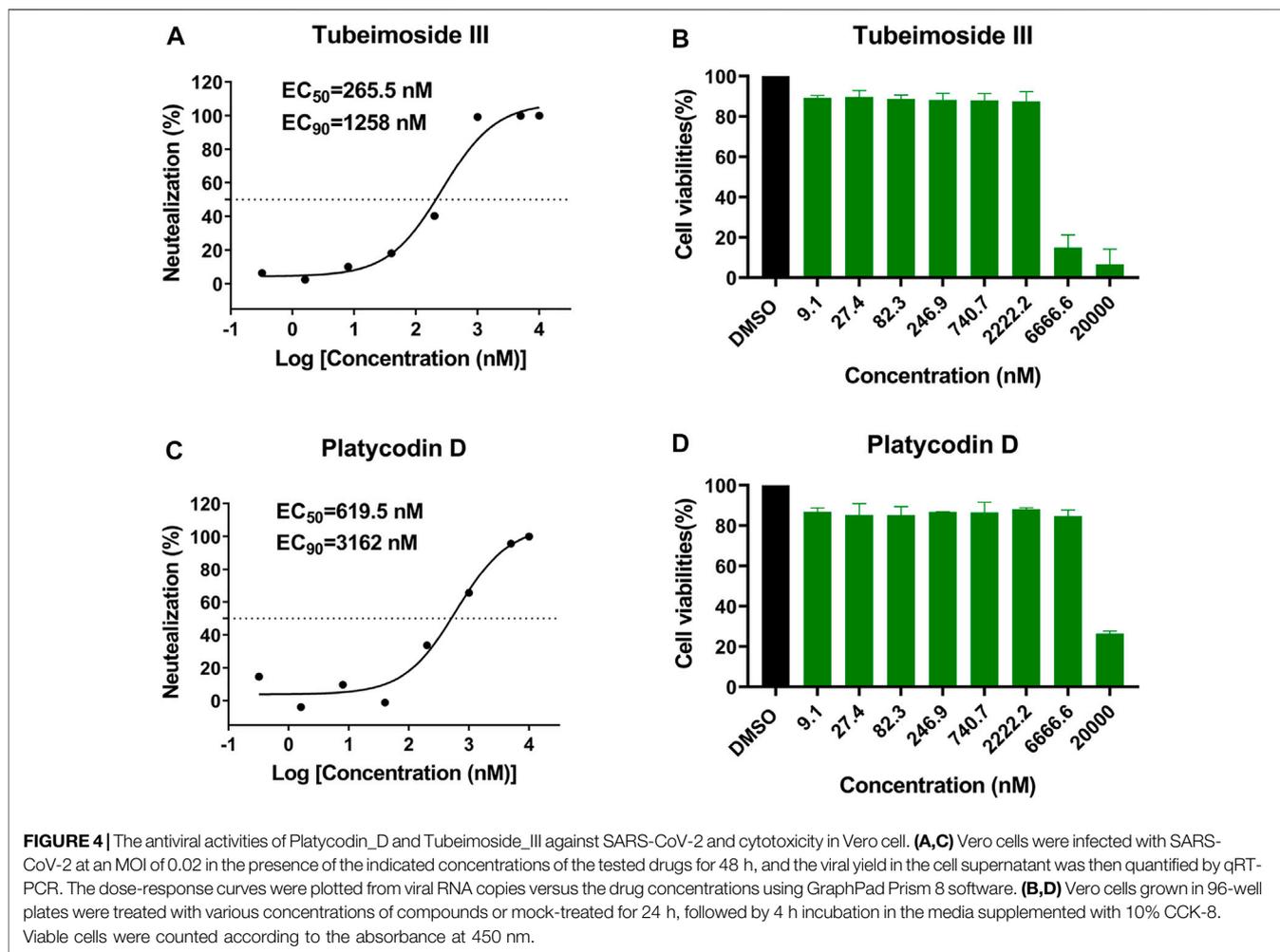
To further confirm the efficiency of virtual screening, we tested the antiviral activity of the eight candidate drugs

in vitro, including Anidulafungin, Ginsenoside_Rb1, Notoginsenoside_Fc, Platycodin_D, Tubeimoside_III, Ginsenoside_Rd, Polygalasaponin_V, and Rebaudioside_A. Similar to our previous study (Zhang et al., 2020b), Vero cells were infected with SARS-CoV-2 at an MOI of 0.02 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 qRT-PCR (Figure 4). The



results showed that Platycodin_D and Tubeimoside_III could efficiently inhibit the replication of SARS-CoV-2, with half-maximal effective concentration (EC_{50}) values of 619.5 and 265.5 nM, and EC_{90} of 3,162 and 1,258 nM, respectively.

Although both Platycodin_D and Tubeimoside_III showed high cytotoxicity in the CCK8 assay (**Figure 4**), the two drugs showed little cytotoxicity even at the concentrations of EC_{90} .

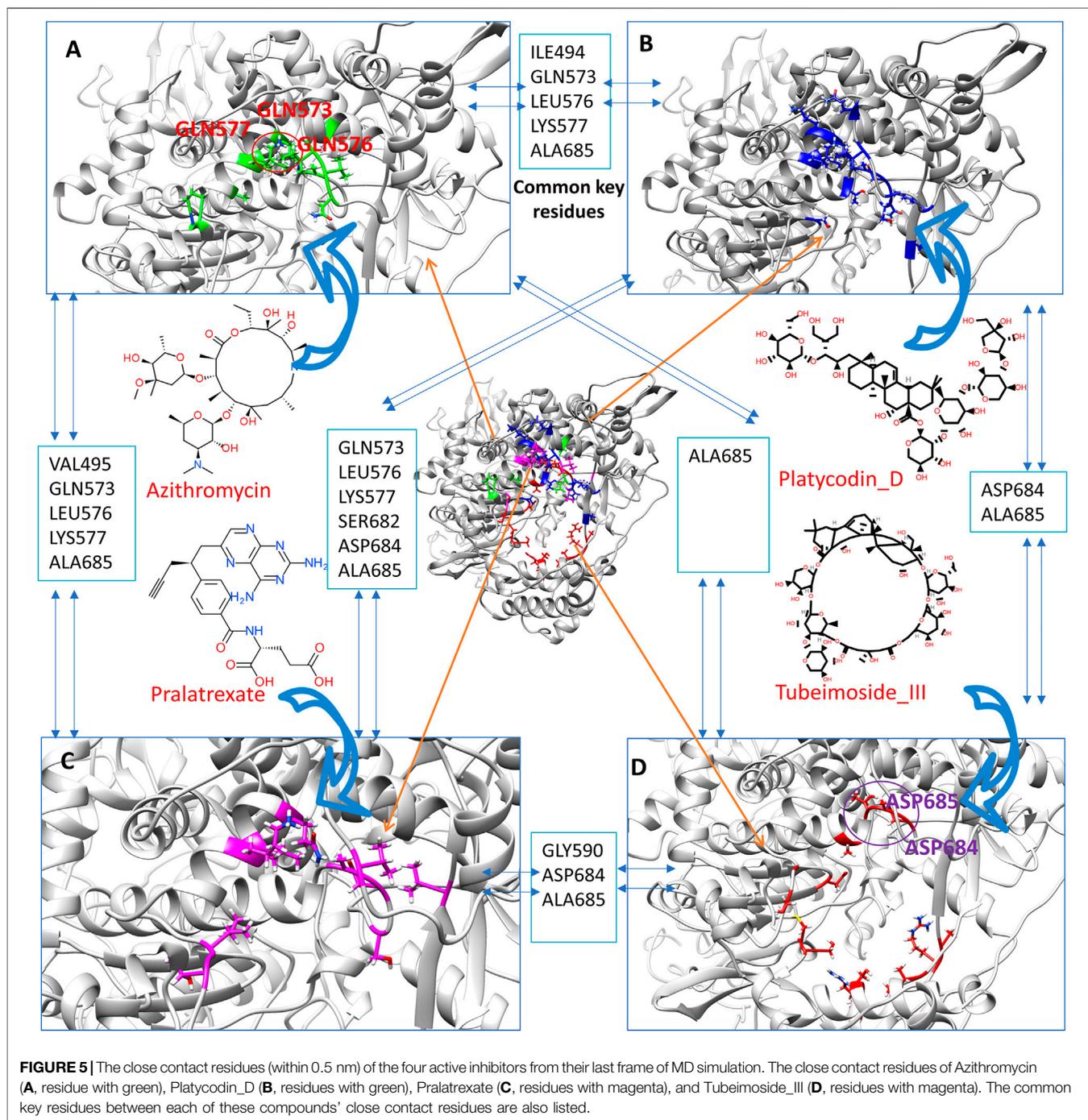


DISCUSSION

Compound specificity is extremely important for drug development. In contrast, our first-stage drug screening has seriously ignored this or lacks an effective way to assess this. Due to low specificity would not cause the failure of direct later experimental validation of protein-ligand binding affinity and cell activity. However, the ultimate goal of drug screening is to find compounds that can pass the preclinical stage and have good potential to pass the clinical stage and finally benefit patients through providing the on-market available drugs. Experimentally evaluating the compound specificity is expensive. This work demonstrates that with the increased efficiency and accuracy of protein-ligand prediction models, it is possible to evaluate the specificity by reverse compound-target assessment in the early stage of drug development. Dual or multi-targeting inhibitors are important for some complicated diseases (Ramsay et al., 2018), such as cancer, some research pointed out that next-generation anticancer agents may be dual or multi-targeting inhibitors (Raghavendra et al., 2018). A similar strategy as the reverse target searching in specificity check may provide new insight for screening and designing dual or multi-target compounds.

Obtaining compounds for one given target is already a challenging task, to obtain dual or multi-target compounds would require a much more efficient new method and strategy, our proposed DFCNN and autodock vina-based evaluation multi-ligand VS multi-target binding relationship may provide some clues for future development of dual or multi-target compounds.

To assess whether any specific active binding sites in the large cavity of RdRp are more likely to bind active inhibitors, we compare the binding site of Platycodin_D and Tubeimoside_III and our previous obtained Azithromycin and Pralatrexate from their last frame of MD simulation. **Figure 5** shows that many common key residues (close contact residues within 0.5 nm of the binding compound) are among the four active compounds. For instance, there are three common key residues (GLN573, LEU576, LYS577) among Azithromycin, Platycodin_D, and Pralatrexate, indicating the region occupied by these three residues is critical for active ligand binding. Unlike the Azithromycin, Platycodin_D, and Pralatrexate, which bind mainly in a small cavity region, the Tubeimoside_III has interacted with more key residues around the whole cavity, which forms a cycle. Tubeimoside_III also has some common



key residues with the other three active compounds. Altogether, the four compounds can potentially bind to the large cavity of RdRp, which is a site for DNA binding. This analysis strongly supports our hypothesis that ligands that can strongly bind to the cavity of RdRp can block the DNA primer entry into the cavity, hence potentially stopping the viral replications. Notably, a recent paper underlined that phospholipidosis is a shared mechanism underlying the antiviral activity of many repurposed drugs (Tummino et al., 2021). In other words, the antiviral activity

found for many compounds could be an artifact. Since we first obtain candidates through screening over RdRp, it should have a much low chance to obtain compounds that inhibit viral through phospholipidosis compared to those directly screening through cell experiment. We carefully examine the four found active compounds with those compounds with phospholipidosis mentioned in the paper. Also, we have checked the literature report about whether each compound involves phospholipidosis. We find that azithromycin can induce phospholipidosis

according to the reports (Liu et al., 2015). Hence, azithromycin's anti-SARS-CoV-2 effect in the cell may not only through binding with the RdRp. For the other three compounds (Platycodin_D, Tubeimoside_III, and Pralatrexate), no literature report show they can induce phospholipidosis. Furthermore, their physicochemical feature is not like those compounds that can induce phospholipidosis, which indicates compounds selected by our method more tends to exert inhibiting effect by binding with the protein target.

To assess the general features of those low specificity compounds that have high potential affinity but low DFCNN specificity (**Supplementary Figure S4A**) or autodock vina-based relative specificity (**Supplementary Figure S4B**) and select those compounds to have Vina relative specificity larger than 1 from **Table 1**. For better comparison, we also list cases that have high calculated specificity cases in **Supplementary Figure S4C**. We observed that high specificity compounds are usually composed of more unique groups. However, it is not easy to distinguish high specificity from abundant compounds by direct visual observation; hence, specificity calculation is necessary to determine the compound's specificity quickly. In this work, two specificity scores were used, the first one using the DFCNN, which mainly considers binding through the pocket and ligand physicochemical property, and the second one which use autodock vina docking, which can provide structure insight about specificity. In terms of selectivity, we have used MD simulation and metadynamics simulation, which provide atomic details and dynamics information about structural interaction between compound and target.

The RdRp (nsp12) is highly conserved among the subtypes, and it only has one mutation (P323L) in Omicron (Ilmjärvi et al., 2021). To explore whether the mutation can influence the large cavity, we have marked P323 residue with yellow in **Supplementary Figure S5** from the nsp12-nsp7-nsp8 complex with PDB ID 7ED5 (Shannon et al., 2022). It should be noted that it is far away from the Template-primer RNA binding cavity. This strongly supports our assumption that active compounds binding to RdRp of original SARS-CoV-2 should also become active against Omicron. It is also interesting to note that the ligand AT9 is covalent binding to the Template-primer RNA to stop its elongation. At the same time, our hit compounds aim to prevent the Template-primer RNA from entering the cavity.

Since we are exploring compounds inhibitory RdRp, the most possible off-target would be some other polymerases that similar to RdRp in humans. We have checked the 102 targets and found there is one poly (ADP-ribose) polymerase (PARP) (PDB ID: 3l3m). It would be much more helpful if it contains more other polymerases, especially DNA polymerase in this study, since the compounds that prefer binding with RdRp may also prefer binding with human DNA polymerase, and lead to unwanted side effects. We suspect this is the reason why the selected compounds still have slight cell toxicity. And strongly suggest that in some specific cases, the user should add more relevant protein to the target pool, in order to obtain compounds with high specificity. For instance, in the future, it should use similar method to check whether the compounds will bind to human DNA polymerase when designing RdRp inhibitors, in order to reduce the undesired side effects.

The compound ADME is very important. We have calculated the ADME property of the final selected inhibitors by the online SwissADME (<http://www.swissadme.ch/>). Most calculations suggest that our selected compounds have too larger a molecular weight. Hence, it is very important to analyze structural features essential to simplify those inhibitors while keeping their activities. Also, the topological polar surface area (TPSA) of a molecule is also too large. Take Platycodin_D as an example, removing some of its glycan groups may improve its MADE property. But we should note that some on-market drugs also break the Lipinski rule of five. Breaking one or more of Lipinski's rules, does not mean a drug candidate cannot be effective. Moreover, compounds beyond the rule of five are often better suited for investigation when targeting large and flat binding sites (Egbert et al., 2019). Interestingly, the RdRp pocket is actually very large.

CONCLUSION

In the present work, we identify highly potential RdRp inhibitors that may bind at a different site of a large RdRp cavity by screening more compounds. Also, we selected more specific candidates by integrating the specificity checking procedures into the screening pipeline. Moreover, we have discovered that Platycodin_D and Tubeimoside_III have a strong inhibiting effect against SARS-CoV-2 with IC₅₀ values of 619.5 and 265.5 nM, respectively. Notably, our methods strongly indicate similarities to Azithromycin and Pralatrexate. Their inhibitory effect is highly likely through competitive prevention of the RNA template-primer RNA entering into its RdRp binding cavity. In addition, we have explored the detailed binding sites of these RdRp nonbonded inhibitors (Azithromycin, Platycodin_D, Pralatrexate, and Tubeimoside_III) and illustrated their binding mechanism by comparing some of the common key residues. Altogether, we again show the workability and power of the hybrid screening strategy, bringing in specificity checking methods, which can be applied to a wide range of drug screening applications, and illustrated the binding mechanism of RdRp nonbonded inhibitors. Also, the screening pipeline will hopefully become a generalized drug screening pipeline against many other therapeutic protein targets in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, and further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

HZ and YY designed the study. HZ, KMS, and HB performed computations and data analyses. YY and XG performed TCID₅₀, antiviral activities, and cytotoxicity testing assay experiments. All authors contributed to writing the manuscript. HZ, YY, JZ, YP,

and YW supervised the study. All authors read and approved the final manuscript.

FUNDING

This study was supported in part by the National Science Foundation of China [Grant No. 62106253 (HZ), 21933010 (JZ)]. This work was supported in part by the Shenzhen KQTD Project (No. KQTD20200820113106007) (YP), the

National Key Research and Development Program of China under grant No. 2018YFB0204403 (YW), and the Strategic Priority CAS Project XDB38000000 (WY).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2022.933102/full#supplementary-material>

REFERENCES

- Biovia, D. S. (2005). *Discovery Studio Visualizer v21.1.0.20298*. Dassault Systèmes: BIOVIA.
- Creanza, T. M., Lamanna, G., Delre, P., Contino, M., Corriero, N., Saviano, M., et al. (2022). DeLA-Drug: A Deep Learning Algorithm for Automated Design of Druglike Analogues. *J. Chem. Inf. Model.* 62, 1411–1424. doi:10.1021/acs.jcim.2c02025
- Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald: AnN-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98, 10089–10092. doi:10.1063/1.464397
- DeLano, W. L. (2002). Pymol: An Open-Source Molecular Graphics Tool. *CCP4 Newsl. Protein Crystallogr.*
- Dey, S. K., Saini, M., Dhembla, C., Bhatt, S., Rajesh, A. S., Anand, V., et al. (2021). Suramin, Penciclovir, and Anidulafungin Exhibit Potential in the Treatment of COVID-19 via Binding to Nsp12 of SARS-CoV-2. *J. Biomol. Struct. Dyn.*, 1–17. doi:10.1080/07391102.2021.2000498
- Egbert, M., Whitty, A., Keserü, G. M., and Vajda, S. (2019). Why Some Targets Benefit from beyond Rule of Five Drugs. *J. Med. Chem.* 62, 10005–10025. doi:10.1021/acs.jmedchem.8b01732
- Flemming, A. (2022). Omicron, the Great Escape Artist. *Nat. Rev. Immunol.* 22, 75. doi:10.1038/s41577-022-00676-6
- Gupta, A., Müller, A. T., Huisman, B. J. H., Fuchs, J. A., Schneider, P., and Schneider, G. (2018). Generative Recurrent Networks for De Novo Drug Design. *Mol. Inf.* 37, 1700111. doi:10.1002/minf.201700111
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINC: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* 18, 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463:aid-jcc4>3.0.co;2-h
- Hess, B., Kutzner, C., and De Van Der, S. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* 4, 435–447. doi:10.1021/ct700301q
- Hornak, V., and Simmerling, C. (2003). Generation of Accurate Protein Loop Conformations through Low-Barrier Molecular Dynamics. *Proteins* 51, 577–590. doi:10.1002/prot.10363
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual Molecular Dynamics. *J. Mol. Graph.* 14, 33–38. doi:10.1016/0263-7855(96)00018-5
- icm_browser_pro (2020). *Icm_Browser_Pro*.
- Ilmjärvi, S., Abdul, F., Acosta-Gutiérrez, S., Estarellas, C., Galdadas, I., Casimir, M., et al. (2021). Concurrent Mutations in RNA-dependent RNA Polymerase and Spike Protein Emerged as the Epidemiologically Most Successful SARS-CoV-2 Variant. *Sci. Rep.* 11, 13705. doi:10.1038/s41598-021-91662-w
- Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S. Y., Park, S., et al. (2020). Identification of Antiviral Drug Candidates against SARS-CoV-2 from FDA-Approved Drugs. *Antimicrob. Agents Chemother.* 64. doi:10.1128/AAC.00819-20
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79, 926–935. doi:10.1063/1.445869
- Karimi, M., Wu, D., Wang, Z., and Shen, Y. (2019). DeepAffinity: Interpretable Deep Learning of Compound-Protein Affinity through Unified Recurrent and Convolutional Neural Networks. *Bioinformatics* 35, 3329–3338. doi:10.1093/bioinformatics/btz111
- Laio, A., and Gervasio, F. L. (2008). Metadynamics: A Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* 71, 126601. doi:10.1088/0034-4885/71/12/126601
- Lee, C.-C., Hsieh, C.-C., and Ko, W.-C. (2021). Molnupiravir-A Novel Oral Anti-SARS-CoV-2 Agent. *Antibiotics* 10, 1294. doi:10.3390/antibiotics10111294
- Liu, Y., Kam, W. R., Ding, J., and Sullivan, D. A. (2014). One Man's Poison Is Another Man's Meat: Using Azithromycin-Induced Phospholipidosis to Promote Ocular Surface Health. *Toxicology* 320, 1–5. doi:10.1016/j.tox.2014.02.014
- Lu, J.-M., Yao, Q., and Chen, C. (2009). Ginseng Compounds: An Update on Their Molecular Mechanisms and Medical Applications. *Cvp* 7, 293–302. doi:10.2174/157016109788340767
- Moret, M., Helmstädter, M., Grisoni, F., Schneider, G., and Merk, D. (2021). Beam Search for Automated Design and Scoring of Novel ROR Ligands with Machine Intelligence. *Angew. Chem. Intl Ed.* 60, 19477–19482. doi:10.1002/anie.202104405
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., et al. (2004). UCSF Chimera?A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* 25, 1605–1612. doi:10.1002/jcc.20084
- Raghavendra, N. M., Pingili, D., Kadasi, S., Mettu, A., and Prasad, S. V. U. M. (2018). Dual or Multi-Targeting Inhibitors: The Next Generation Anticancer Agents. *Eur. J. Med. Chem.* 143, 1277–1300. doi:10.1016/j.ejmech.2017.10.021
- Ramsay, R. R., Popovic Nikolic, M. R., Nikolic, K., Uliassi, E., and Bolognesi, M. L. (2018). A Perspective on Multi-target Drug Discovery and Design for Complex Diseases. *Clin. Transl. Med.* 7, 3. doi:10.1186/s40169-017-0181-2
- Reed, L. J., and Muench, H. (1938). A Simple Method of Estimating Fifty Per Cent Endpoints. *Am. J. Epidemiol.* 27, 493–497. doi:10.1093/oxfordjournals.aje.a118408
- Roy, A., Yang, J., and Zhang, Y. (2012). COFACTOR: An Accurate Comparative Algorithm for Structure-Based Protein Function Annotation. *Nucleic Acids Res.* 40, W471–W477. doi:10.1093/nar/gks372
- Ruiz-Carmona, S., Schmidtke, P., Luque, F. J., Baker, L., Matassova, N., Davis, B., et al. (2017). Dynamic Undocking and the Quasi-Bound State as Tools for Drug Discovery. *Nat. Chem.* 9, 201–206. doi:10.1038/nchem.2660
- Saleh, N., Ibrahim, P., Saladino, G., Gervasio, F. L., and Clark, T. (2017). 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.* 57, 1210–1217. doi:10.1021/acs.jcim.6b00772
- Shannon, A., Fattorini, V., Sama, B., Selisko, B., Feracci, M., Falcou, C., et al. (2022). A Dual Mechanism of Action of AT-527 against SARS-CoV-2 Polymerase. *Nat. Commun.* 13, 621. doi:10.1038/s41467-022-28113-1
- Sousa Da Silva, A. W., and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfAcE. *BMC Res. Notes.* 5, 367. doi:10.1186/1756-0500-5-367
- Stepniewska-Dziubinska, M. M., Zielenkiewicz, P., and Siedlecki, P. (2018). Development and Evaluation of a Deep Learning Model for Protein-Ligand Binding Affinity Prediction. *Bioinformatics* 34, 3666–3674. doi:10.1093/bioinformatics/bty374
- Tao, K., Tzou, P. L., Nohuin, J., Gupta, R. K., de Oliveira, T., Kosakovsky Pond, S. L., et al. (2021). The Biological and Clinical Significance of Emerging SARS-CoV-2 Variants. *Nat. Rev. Genet.* 22, 757–773. doi:10.1038/s41576-021-00408-x
- Torng, W., and Altman, R. B. (2019). Graph Convolutional Neural Networks for Predicting Drug-Target Interactions. *J. Chem. Inf. Model.* 59, 4131–4149. doi:10.1021/acs.jcim.9b00628

- Trott, O., and Olson, A. J. (2010). AutoDock Vina: Improving the Speed and Accuracy of Docking With a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* 31 2, 455–461. doi:10.1002/jcc.21334
- Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* 185, 604–613. doi:10.1016/j.cpc.2013.09.018
- Tummino, T. A., Rezelj, V. V., Fischer, B., Fischer, A., O'Meara, M. J., Monel, B., et al. (2021). Drug-induced Phospholipidosis Confounds Drug Repurposing for SARS-CoV-2. *Science* 373, 541–547. doi:10.1126/science.abc14708
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* 26, 1701–1718. doi:10.1002/jcc.20291
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus Biology and Replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* 19, 155–170. doi:10.1038/s41579-020-00468-6
- Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* 25, 247–260. doi:10.1016/j.jmgm.2005.12.005
- Wang, K., Zhou, R., Li, Y., and Li, M. (2021). DeepDTAF: A Deep Learning Method to Predict Protein-Ligand Binding Affinity. *Brief. Bioinform.* 22, bbab072. doi:10.1093/bib/bbab072
- Williams, T., Kelley, C., Campbell, J., Cunningham, R., Denholm, D., Elber, G., et al. (2012). *Gnuplot 4.6*. Softw. Man.
- Yan, Z., and Wang, J. (2012). Specificity Quantification of Biomolecular Recognition and its Implication for Drug Discovery. *Sci. Rep.* 2, 309. doi:10.1038/srep00309
- Yang, T. R., Kasai, R., Zhou, J., and Tanaka, O. (1983). Dammarane Saponins of Leaves and Seeds of Panax Notoginseng. *Phytochemistry* 22, 1473–1478. doi:10.1016/S0031-9422(00)84039-X
- Yang, Y., Yang, M., Peng, Y., Liang, Y., Wei, J., Xing, L., et al. (2022). Longitudinal Analysis of Antibody Dynamics in COVID-19 Convalescents Reveals Neutralizing Responses up to 16 Months after Infection. *Nat. Microbiol.* 7, 423–433. doi:10.1038/s41564-021-01051-2
- Yin, W., Mao, C., Luan, X., Shen, D.-D., Shen, Q., Su, H., et al. (2020). Structural Basis for Inhibition of the RNA-dependent RNA Polymerase from SARS-CoV-2 by Remdesivir. *Science* 368, 1499–1504. doi:10.1126/science.abc1560
- Zhang, H., Li, J., Saravanan, K. M., Wu, H., Wang, Z., Wu, D., et al. (2021). An Integrated Deep Learning and Molecular Dynamics Simulation-Based Screening Pipeline Identifies Inhibitors of a New Cancer Drug Target TIPE2. *Front. Pharmacol.* 12, 3297. doi:10.3389/fphar.2021.772296
- Zhang, H., Liao, L., Cai, Y., Hu, Y., and Wang, H. (2019a). IVS2vec: A Tool of Inverse Virtual Screening Based on Word2vec and Deep Learning Techniques. *Methods* 166, 57–65. doi:10.1016/j.jymeth.2019.03.012
- Zhang, H., Liao, L., Saravanan, K. M., Yin, P., and Wei, Y. (2019b). DeepBindRG: a Deep Learning Based Method for Estimating Effective Protein-Ligand Affinity. *PeerJ* 7, e7362. doi:10.7717/peerj.7362
- Zhang, H., Saravanan, K. M., Yang, Y., Hossain, M. T., Li, J., Ren, X., et al. (2020a). Deep Learning Based Drug Screening for Novel Coronavirus 2019-nCoV. *Interdiscip. Sci. Comput. Life Sci.* 12, 368–376. doi:10.1007/s12539-020-00376-6
- Zhang, H., Yang, Y., Li, J., Wang, M., Saravanan, K. M., Wei, J., et al. (2020b). A Novel Virtual Screening Procedure Identifies Pralatrexate as Inhibitor of SARS-CoV-2 RdRp and it Reduces Viral Replication *In Vitro*. *PLoS Comput. Biol.* 16, e1008489. doi:10.1371/journal.pcbi.1008489
- Zhang, Y. (2008). I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinforma.* 9, 40. doi:10.1186/1471-2105-9-40
- Zhou, Y., Webber, S. E., Murphy, D. E., Li, L.-S., Dragovich, P. S., Tran, C. V., et al. (2008). Novel HCV NS5B Polymerase Inhibitors Derived from 4-(1',1'-Dioxo-1',4'-Dihydro-1'\lambda6-Benzo[1',2',4']thiadiazin-3'-yl)-5-Hydroxy-2H-Pyridazin-3-Ones. Part 1: Exploration of 7'-substitution of Benzothiadiazine. *Bioorg. Med. Chem. Lett.* 18, 1413–1418. doi:10.1016/j.bmcl.2008.01.007

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Gong, Peng, Saravanan, Bian, Zhang, Wei, Pan and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.