# Combined Deep Learning and Molecular Docking Simulations Approach Identifies Potentially Effective FDA Approved Drugs for Repurposing against SARS-CoV-2

*Muhammad U Anwar[1]‡, Farjad Adnan[2]‡, Asma Abro[3]‡, Rayyan A Khan[1]‡, Asad U Rehman[4], Muhammad Osama[4], Saad Javed[4], Ahmadullah Baig[4], Muhammad R Shabbir[4], Muhammad Z Assir[4,5*]*

Authors' affiliations:

[1]Department of Electrical and Computer Engineering, Technical University Munich, Germany

[2]Paderborn University, Paderborn, Germany

[3]Department of Biotechnology, Faculty of Life Sciences and Informatics, Balochistan University of Information Technology, Engineering and Management Sciences, Quetta, Pakistan

[4]Department of Medicine, Allama Iqbal Medical College, University of Health Sciences, Lahore, 54550, Pakistan

[5]Department of Molecular Biology, Shaheed Zulfiqar Ali Bhutto Medical University, Islamabad, 44000, Pakistan

‡These authors contributed equally.

[*]Corresponding Author:

Muhammad Zaman Assir

Assistant Professor of Medicine

Department of Medicine, Medical Unit 3

Allama Iqbal Medical College/ University of Health Sciences Lahore, Pakistan

dr.zamankhan@gmail.com

**Author Contributions**

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Conflict of Interest:

All authors declare no competing interest.

**Abstract:** (Word Count 280)

The ongoing pandemic of Coronavirus Disease 2019 (COVID-19), the disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has posed a serious threat to global public health. Currently no approved drug or vaccine exists against SARS-CoV-2. Drug repurposing, represented as an effective drug discovery strategy from existing drugs, is a time efficient approach to find effective drugs against SARS-CoV-2 in this emergency situation. Both experimental and computational approaches are being employed in drug repurposing with computational approaches becoming increasingly popular and efficient. In this study, we present a robust experimental design combining deep learning with molecular docking experiments to identify most promising candidates from the list of FDA approved drugs that can be repurposed to treat COVID-19. We have employed a deep learning based Drug Target Interaction (DTI) model, called DeepDTA, with few improvements to predict drug-protein binding affinities, represented as KIBA scores, for 2,440 FDA approved and 8,168 investigational drugs against 24 SARS-CoV-2 viral proteins. FDA approved drugs with the highest KIBA scores were selected for molecular docking simulations. We ran docking simulations for 168 selected drugs against 285 total predicted and/or experimentally proven active sites of all 24 SARS-CoV-2 viral proteins. We used a recently published open source AutoDock based high throughput screening platform virtualflow to reduce the time required to run around 50,000 docking simulations. A list of 49 most promising FDA approved drugs with best consensus KIBA scores and AutoDock vina binding affinity values against selected SARS-CoV-2 viral proteins is generated. Most importantly, anidulafungin, velpatasvir, glecaprevir, rifabutin, procaine penicillin G, tadalafil, riboflavin 5'-monophosphate, flavin adenine dinucleotide, terlipressin, desmopressin, elbasvir, oxatomide, enasidenib, edoxaban

and selinexor demonstrate highest predicted inhibitory potential against key SARS-CoV-2 viral proteins.

Key Words:

SARS-CoV-2, COVID 19, Deep Learning, DeepDTA, Docking, FDA approved Drugs, Drug Repurposing

# List of Abbreviations:

| | |
|---|---|
| SARS-CoV-2 | Severe acute respiratory syndrome coronavirus 2 |
| COVID19 | Novel 2019 Coronavirus Disease |
| SARS-CoV | Sever acute respiratory syndrome coronavirus |
| MERS-CoV | Middle East respiratory syndrome coronavirus |
| FDA | Food and Drug Administration |
| PDB | The Protein Data Bank |
| DAVIS | Densely Annotated VIdeo Segmentation |
| CNN | Convolutional neural network |
| DeepDTA | Deep Drug-Target Binding Affinity Prediction |
| CPU | Central Processing Unit |
| Ki | Inhibitory constant |
| Kd | Dissociation constant |
| IC50 | Half maximal inhibitory concentration |
| Bi-LSTM | Bi-directional long short term memory |
| BPE | Byte pair encoding |
| ReLU | Rectified Linear Unit |
| MSE | Mean Squared Error |
| CI | Concordance index |
| C-I-TASSER | Contact-guided Iterative Threading ASSEmbly Refinement |
| S | Spike |
| E | Envelope |
| M | Membrane |
| N | Nucleocapsid |
| RNA | Ribonucleic acid |
| DNA | Deoxyribonucleic acid |
| Nsp1 | Non-structural protein 1 |
| Nsp7 | Non-structural protein 7 |
| Nsp8 | Non-structural protein 8 |
| Nsp12 | Non-structural protein 12 |
| HCV | Hepatitis C virus |
| HBV | Hepatitis B virus |
| FAD | Flavin Adenine Dinucleotide |
| DPP-4 | Dipeptidyl peptidase 4 |
| ACE | Angiotensin converting enzyme |
| HMG-CoA | β-Hydroxy β-methylglutaryl-CoA |
| RMSD | Root-mean-square deviation |
| PLpro | Papain-like protease |
| RdRp | RNA-directed RNA polymerase |
| Hel | Helicase |
| CPK | Corey-Pauling-Koltun |
| RCSB | Research Collaboratory for Structural Bioinformatics |
| XPO1 | Exportin-1 |
| FMN | Flavin mononucleotide |

## Introduction

Severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) are three highly pathogenic human coronaviruses (CoVs) that can cause severe disease in humans[1] [2]. SARS-CoV resulted in an outbreak in 2003 while MERS-CoV was reported in Saudi Arabia in June 2012. SARS-CoV-2 was first reported in the city of Wuhan in central China and the disease caused by this virus has been termed as Coronavirus Disease 2019 (COVID-19). COVID-19 has now become a global pandemic with severe effects on global public health. As of 28th April, 2020, approximately 3 million confirmed cases of COVID-19 have been reported with an estimated 211,147 deaths globally[3]. Currently no approved drug or vaccine is available for COVID-19. Therefore, there is a pressing urgency to make an expedited discovery of effective therapeutics for COVID-19.

*De-novo* drug development is a time consuming, complex, and expensive process that typically costs 2.8 billion dollars[4]. How to decrease the costs and speed up new drug discovery has become a challenging and urgent question in the industry. Drug repurposing, a process of investigating approved or investigational drugs for new therapeutic purposes, offers a cost and time effective alternative[5]. Drug repurposing is based on a paradigm shift in our understanding that many effective drugs act via modulation of multiple proteins rather than single targets[6] [7]. Both experimental and computational approaches are being employed in drug repurposing with computational approaches becoming increasingly popular, robust and efficient[5].

Recent advancements in the field of deep learning have significantly improved the computational approaches for drug repurposing. Conventional machine learning methods are limited by the lack of ability to process the data in the raw form and hence depend on feature

engineering[8] for machine learning and pattern recognition[9]. Deep learning, a novel machine learning approach utilizing deep neural networks, based techniques overcome the issue of manual feature engineering and processes the crude multi-dimensional data (images, text sequences etc) in their layers for algorithm training[10]. The architecture of deep learning model is a multi-layer cascade of several modules, mapping a nonlinear relationship between input and output employing a back propagation method to fine-tune the corresponding weights[11]. Deep learning is also being used to study Drug-Target interaction[12].

The approach of supervised deep learning can be used to predict drug-target binding affinities. In this stratagem, a deep learning model is trained on the experimentally available binding affinities of several protein-ligand complexes. Several available benchmark protein-ligand complex datasets, providing the experimental binding affinities, are PDB[13], DAVIS[14] and KIBA[15 16] datasets. These have been used in several studies to predict the binding affinities of the complexes. The training of the deep learning model is realized using physical or structural features of protein ligand complexes on the training dataset. A properly trained model can be employed to predict the binding affinities of unseen protein ligand complexes. A binary prediction classifier has already been proposed[17 18] which takes into account the several input representations of the protein-ligand complex. In several other deep learning network studies, a protein-ligand interaction scoring has been predicted, training the convolutional neural network (CNN) using the three dimensional structure of the associated complexes[19 20]. The main source of these complexes is PDB[13] ; however small number of interactions is a limitation as only 25000 protein-ligand complexes have been documented. DeepDTA[21] is another novel approach which uses SMILES and FASTA sequences of ligands and proteins, respectively. In the DeepDTA approach, two separate CNN blocks have

been employed to train the protein and ligand sequences and eventually combined in a fully connected layer called DeepDTA. In this study, we have employed DeepDTA with few improvements.

Another computational approach that is employed to study drug target interaction is virtual screening. Virtual screening can be performed by using *molecular docking* - a technique that samples the ensemble of binding modes to predict preferred pose(s) in which the ligand can bind with the receptor at a certain location, known as active / binding site, in order to form a stable complex[22] [23]. A binding mode refers to a unique conformation along with orientation and translation of the ligand. Ranking of the preferred binding modes / poses is carried out by evaluating a mathematical function, known as scoring function, that quantifies the stability of the complex formed by a particular pose of ligand with the receptor[24] [25,]. In our work, docking was performed using actual simulation as it is more realistic compared to its alternatives[25] [26,]. We have used the results of docking simulations to get docking scores of potential candidate drugs with COVID19 viral proteins. In order to reduce the time required for around 50,000 docking simulations, we made use of a recently published open source high throughput screening platform virtualflow[27] to parallelize the docking scenarios across multiple machines and CPU cores.

In this study we have employed a multidisciplinary, multimodal approach combining deep machine learning and large scale molecular docking experiments in a sequential manner to identify FDA approved drugs that can be used as effective treatment against SARS-CoV-2.

**Methods:**

Deep Learning Model Overview

In our work, we have employed a deep learning based DTI-model, called DeepDTA[20], with few improvements. Our deep learning model automatically incorporates useful and required features

from raw ligands and proteins into the model to predict the drug and protein interactions. We have utilized 1-D sequences of ligands (SMILES)[28] and proteins (FASTA)[29] to train our model. SMILES (Simplified Molecular Input Line Entry System) representation of molecules have been exploited rather than physical and chemical properties associated with the ligands. We have employed the Bi-LSTM[30] blocks of neural networks instead of CNN used by DeepDTA[20] to learn the SMILE representation of molecules whereas fully connected CNN have been engaged to learn the FASTA representation of proteins.

Dataset

We have used a benchmark KIBA dataset for training our model and prediction evaluation of binding affinities, which has been tapped previously in a handful of studies. The KIBA dataset comprises selectivity assays of the kinase proteins family and the associated inhibitors[15]. It predominantly embodies the corresponding KIBA scores. The KIBA values have been contrived by combining $K_i$, $K_d$ and $IC_{50}$ values obtained from several sources. This Dataset has been created from original 52,498 drugs[14], which has been filtered to 2111 unique drugs. The pool of all shortlisted drugs has at least 10 measured interactions, yielding a total of 229 proteins out of all 467 targets and a total of 118,254 interactions[20]. In our preliminary work we have employed the same KIBA dataset used by DeepDTA[20].

Input Representation

In our approach, we have employed byte pair encoding[30] to represent the SMILES of ligands and FASTA sequences of the proteins. SMILES have a varied length based upon the number of atoms and type of bonds in the drug. Since we used Bi-LSTM[31] to learn the representation of drugs, we used an end token to indicate the end of the SMILES. We utilized CNN for learning protein

representation, so we set the maximum length of the FASTA sequence to 1000 characters. We truncated larger sequences and padded zeros to smaller sequences.

Proposed Model

Our prediction model is based upon two sub-models. One part takes SMILES as input and the second part takes FAST as input. SMILES are encoded via byte pair encoding (BPE) and then passed to a novel Bi-LSTM[31] learning approach for sequence analysis. It has been successfully used in the recent studies for text sequence analysis. This part yields an effective representation of drugs. FASTA sequences of proteins are also encoded with BPE and then trained with Convolutional neural networks (CNN). This part learns the representation of the proteins. These separate representations are then concatenated and then passed through three fully connected layers. Our CNN model greatly relies on the number and size of filters to define dependencies of the sequences and we have chosen appropriate values fit to the scenario. As the size of filters and the number of filters are varied, performance of CNN also varies with it. All layers of networks are shown in the block diagram (Figure 1).

The interaction pairs of the KIBA dataset have been split into training and testing datasets. As an activation function, we used Rectified Linear Unit (ReLU)[32] mathematically represented as $g(z) = \max\{0, z\}$, which has been widely used in deep learning studies[8]. A learning model is trained to minimize the difference between the expected and actual value. We have formulated it as a regression task, mean squared error (MSE) has been chosen as an appropriate loss function. The learning has been accomplished with the 100 epochs and mini-batch size of 512 is chosen to update the weights and hyperparameters of our networks. The chosen optimization algorithm to train the networks is ADAM[33] which had the default learning rate of 0.001.

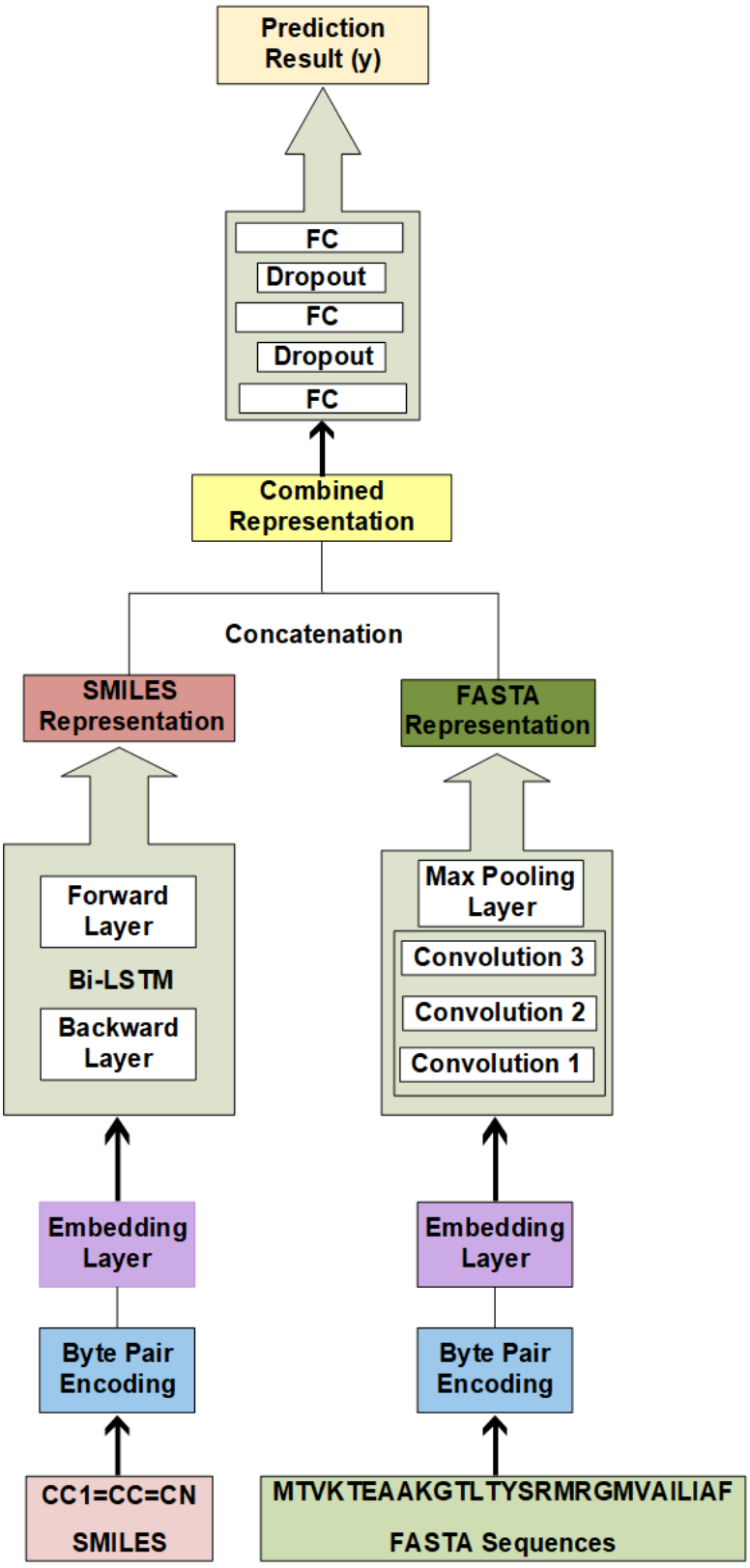

Prediction Result (y)
FC
Dropout
FC
Dropout
FC
Combined Representation
Concatenation
SMILES Representation
FASTA Representation
Forward Layer
Bi-LSTM
Backward Layer
Max Pooling Layer
Convolution 3
Convolution 2
Convolution 1
Embedding Layer
Embedding Layer
Byte Pair Encoding
Byte Pair Encoding
CC1=CC=CN
SMILES
MTVKTEAAKGTLTYSRMRGMVAILIAF
FASTA Sequences

**Figure 1:** Block Diagram of Modified DeepDTA model with Bi-LSTM and CNN blocks to learn from sequences - The proposed Bi-LSTM and CNN methodology is represented. It consists of two separate BI-LSTM and CNN blocks, for training the representations of ligands and proteins respectively. The output representations from both blocks have been concatenated and fed to the fully connected layers, which eventually predict the drug.-protein binding affinities at their output.

Prediction of KIBA Scores for FDA approved and investigational drugs

After training of the model, we have predicted the KIBA scores of FDA approved and investigational drugs. The FDA approved and investigational drugs have been retrieved from Canadian DrugBank[34]. The DrugBank database is a handy pool of drugs which includes detailed drug information and their corresponding interactions. The DrugBank is composed of approved 3,546 drugs, including 2,630 small molecule drugs and 1,372 approved biologics (proteins, peptides, vaccines, and allergenics) and over 9,000 drugs which are either under investigation or experimentation[34]. From DrugBank, we have retrieved SMILE sequences for both FDA approved and investigational drugs. The SMILE sequences of various drugs were not available on the DrugBank and ultimately based upon available SMILE sequences we have incorporated 2,440 FDA approved drugs whereas 8,168 investigational drugs have been analyzed. The FASTA sequences of 24 viral proteins have been acquired from the published genome of SARS-CoV-2 available at C-I-TASSER[28]. Both of them, in combination, have been supplied to the trained deep learning model, which predicted all corresponding KIBA scores.

Selection of FDA approved drugs for Virtual Screening

FDA approved drugs were ranked according to their KIBA scores for each viral protein. Two mutually non-exclusive subsets of drugs were prepared. Subset A comprised of drugs making it to top 50 drugs with highest KIBA score for any of 24 viral proteins. Subset B comprised of drugs

with KIBA scores of greater than or equal to 11.5 for all 24 viral proteins. Two subsets were then combined, duplicates were removed and drugs with significant interaction with less than 3 viral proteins were removed. Resulting set of FDA approved drugs were then subjected to Virtual Screening.

Molecular Docking Simulations

We made use of recently published structures of 24 viral proteins for COVID19[28] and the three dimensional structures of these proteins were retrieved. We performed extensive literature search for the availability of binding site data for these viral proteins. If the structures are available and their binding site data is elucidated, it is made use in this research work. In the other case, the predicted binding sites for all the proteins are provided as determined by several binding site prediction algorithms e.g. COACH, S-Site, FINDSITE, ConCavity etc[29]. In our simulations for the sake of uniformity, we made use of the predicted binding sites given by C-I-TASSER[28] from which we have obtained our protein structures. For every site, we computed the search box mean by averaging the coordinates of residues forming the site. The size of the search box was set to be the difference of maximum and minimum along all three dimensions. Afterwards we discarded the search boxes that were contained in other search boxes. This resulted in 284 potential sites to be tested in total. In case of published active site data, we relied on that data completely. Autodock tool was used to preprocess docking files for Autodock Vina docking algorithm[30]. The structure files for ligands were obtained from PubChem[31] and converted to pdbqt. Various tools like Chimera[32] and ChemOffice[33] were used for the refinement and proper energy minimization of the structures of ligands. The refined and best stereochemical quality structures having suitable number of minimization steps, were then docked into the active site of the target proteins using AutoDock and AutoDock Vina. The simulations were executed in parallel on eight compute nodes

with eight CPU cores each. To parallelize the whole procedure, we made use of virtualflow[27], an open source platform that automates docking simulations across multiple machines in a scalable manner. Every docking scenario was run once on one CPU core with exhaustiveness value fixed to 8 for all scenarios. VMD[34] (Visual Molecular Dynamics) and Chimera were used for the analysis of best docked conformations and interaction of drugs with active residues.

**Results and Discussion**

Modified DeepDTA identified key FDA approved drugs targeting SARS-CoV-2 viral proteins

Using the DeepDTA based model, we predicted drug-protein binding affinities, represented as KIBA scores, for 2,440 FDA approved and 8,168 investigational drugs against 24 SARS-CoV-2 viral proteins, yielding KIBA scores for a total of 254,592 interactions (Supplementary Figure 1 and Supplementary Table 1). These proteins include four structural proteins; spike (S), envelope (E), membrane (M), and nucleocapsid (N)[2]. N protein forms the capsid that protects the viral RNA while E, M and S proteins make the outer coat of the virus that surrounds the capsid[2 28]. Spike protein projects from the surface of the virus and plays a crucial role in viral attachment, entry and fusion into the target host cell[35-37]. Two essential proteins constituting the viral replication-transcription complex are helicase and non-structural protein 12 (nsp12)[38 39]. Nsp12 is an RNA-dependent RNA polymerase[39] that binds with nsp7 and nsp8 to make a multi-subunit complex essential for viral replication[40]. Helicase (nsp13) assists in viral replication by unwinding the duplex viral RNA. Main protease ($M^{pro}$, also called $3CL^{pro}$)[41 42] is another essential protein that works in conjunction with papain-like protease(s) to process the huge polyproteins encoded by the SARS-CoV-2 genome. These proteins are key targets for an effective antiviral therapy[38 39 41-44]. Predicted KIBA scores for all these interactions are provided in the supplementary table S1.

By applying our predefined filtering strategy (See Section 2.6 of Methods for details), 184 out of 2,440 FDA approved drugs (top 7.5%) with high predicted binding affinity scores were identified. Sixteen drugs with either severe toxicity or topical use only (eyes and skin) were excluded. Remaining 168 drugs included 48 antimicrobial, 18 antineoplastic, 22 central nervous system acting, 35 hormonal, 9 vitamin derivatives and 36 other agents from miscellaneous classes of drugs. (Supplementary table S2). Antimicrobial agents included antibiotics, antiviral, antifungal and antimycobacterial drugs. Two most frequently observed classes of antibiotics were beta lactam agents (including penicillin derivatives and cephalosporins) and quinolones. Antiviral drugs included 3 anti-retroviral drugs (didanosine, nelfinavir and cobicistat), 5 anti-HCV drugs (sofosbuvir, elbasvir, pibrentasvir, glecaprevir and velpatasvir), 2 neuraminidase inhibitor (peramivir, oseltamivir) and 1 anti-HBV drug (adefovir dipivoxil). Three highest predicted KIBA score antifungal agents included anidulafungin, isavuconazonium and natamycin. Antineoplastic drugs included tyrosine and BRAF kinase inhibitors, anthracyclins and growth factor inhibitors amongst others. Vitamin derivatives included vitamin D derivatives and analogues, flavin derivatives (Riboflavin monophosphate and FAD) and biotin. Hormonal drugs included glucocorticoids, androgen antagonists and analogues of estrogen, progesterone, oxytocin, vasopressin and somatostatin. Drugs acting on central nervous system included dopamine agonists, selective serotonin reuptake inhibitors and antipsychotic agents. Miscellaneous group of drugs included dipeptidyl peptidase 4 (DPP-4) inhibitors (alogliptin and linagliptin), anticoagulants (edoxaban, ticagrelor and dabigatran etexilate), calcium channel blockers (manidipine and diltiazem), angiotensin-converting enzyme (ACE) inhibitors (cilazapril, perindopril, trandolapril, enalaprilat and reserpinine), angiotensin II receptor blocker (eprosartan) and HMG-CoA reductase

inhibitors (rosuvastatin and cerivastatin). (For a complete list of drugs with the predicted KIBA scores against each selected SARS-CoV-2 protein see supplementary table 2)

Molecular docking simulations identified most promising inhibitors of selected SARS-CoV-2 proteins

We obtained the protein structures for docking from Zhang lab server (https://zhanglab.ccmb.med.umich.edu/COVID-19/). In order to verify the accuracy of these structures, we performed a comparison with available crystal structures. We found that the root-mean-square deviation (RMSD) of backbone atomic positions was in the range of 0.5 to 1.9 Å, establishing the reliability of the Zhang lab structures. It is worth mentioning that these structures were determined experimentally[35-37 39 41] and being published when the manuscript was in the process of compilation.

This server has reported 24 proteins or peptides encompassing the complete genome of SARS-CoV-2. Our aim is to disrupt the pathways where these proteins are involved in order to inhibit normal functioning and replication cycle of the virus. We ran docking simulations for 168 selected drugs with high predicted KIBA scores against 285 total predicted and/or experimentally proven active sites of all 24 SARS-CoV-2 viral proteins[28 36 37 39 41]. This yielded binding affinity values (kcal/mol) for 47,880 docking simulations (For details of all docking sites and docking energies, see supplementary table S3). AutoDock vina binding affinity values were plotted against KIBA scores of these drugs and drugs with best *consensus* KIBA scores and binding affinity values were selected. (For visualization of KIBA scores and docking binding affinity values of shortlisted drugs, see our interactive plot in Supplementary Figure S2 available online). This provided a list of top 49 FDA approved drugs that are predicted to effectively inhibit selected SARS-CoV-2 viral proteins through 134 key drug-protein interactions with high degree of confidence. These include

antimicrobials (n=12), hormonal agents (n=14), anti-neoplastic drugs (n=11), vitamin derivatives (n=3) and miscellaneous drugs (n=9) (Table 1). Details of these drugs with the predicted KIBA scores and docking affinity values are provided in table 1.

We have shortlisted eighteen docked complexes after extensive analysis and interaction mapping based on their significance in the viral pathways. Complete results for docking energies and active site details are provided in Supplementary Table S3. The shortlisted complexes include proteins host translation inhibitor (NSP1), papain-like protease (PLpro), proteinase 3CL-pro, RNA-directed RNA polymerase (RdRp), helicase (Hel), surface glycoprotein (S) and nucleocapsid phosphoprotein (N) with key drugs anidulafungin, velpatasvir, glecaprevir, rifabutin, procaine penicillin G, tadalafil, riboflavin 5'-monophosphate, flavin adenine dinucleotide, terlipressin, desmopressin, elbasvir, oxatomide, enasidenib, edoxaban and selinexor. The energy values and other information for the selected complexes are provided in the Table 1.

In a broader context, PLpro, 3CL-pro, RdRp and Hel are involved in virus RNA synthesis and replication. Henceforth, more research is being carried out on these targets due to their biological significance. Their structures are mostly available and well elucidated. The fifth protein, NSP1, is the virulence factor that is related to assisting the virus in gene expression and interfering host immune response. The remaining two proteins, that is, S and N are the structural proteins of the virus assisting it in attachment and binding to the host. Results of molecular docking and screening experiments clearly showed that certain drugs have higher affinities for a particular protein target. Here we briefly discuss the results of docking experiments for these drug-protein complexes.

Active site of PLpro is deduced through superimposition with a crystal structure having PDB code 6W9C.A. The RMSD value for backbone atoms came out to be 0.76 Å. There are 4 domains in the monomer of PLpro enzyme, i) an extended ubiquitin-like domain, ii) thumb domain, ii) palm

domain and iv) fingers domain. The binding site for inhibitors is present in between the thumb and palm domains. Three key drugs namely riboflavin 5'-monophosphate, oxatomide and selinexor have shown better binding affinity values than others. Active pocket of the enzyme majorly consists of hydrophilic residues as Tyr857, Asp909, Arg911, Glu912, Tyr1009, Asn1012, Tyr1013, Gln1014, Cys1015, His1017 and Tyr1018. Other residues present on the interface of thumb and palm domains are: Leu907, Gly908, Val910, Thr913, Met953, Pro992, Pro993, Gly1016, and Thr1046. It is noted that riboflavin 5'-monophosphate binds to more hydrophilic residues as compared to other two drugs. However, the best binding affinity is shown by oxatomide having both hydrophilic and hydrophobic interactions in a better fit (Figure 2).

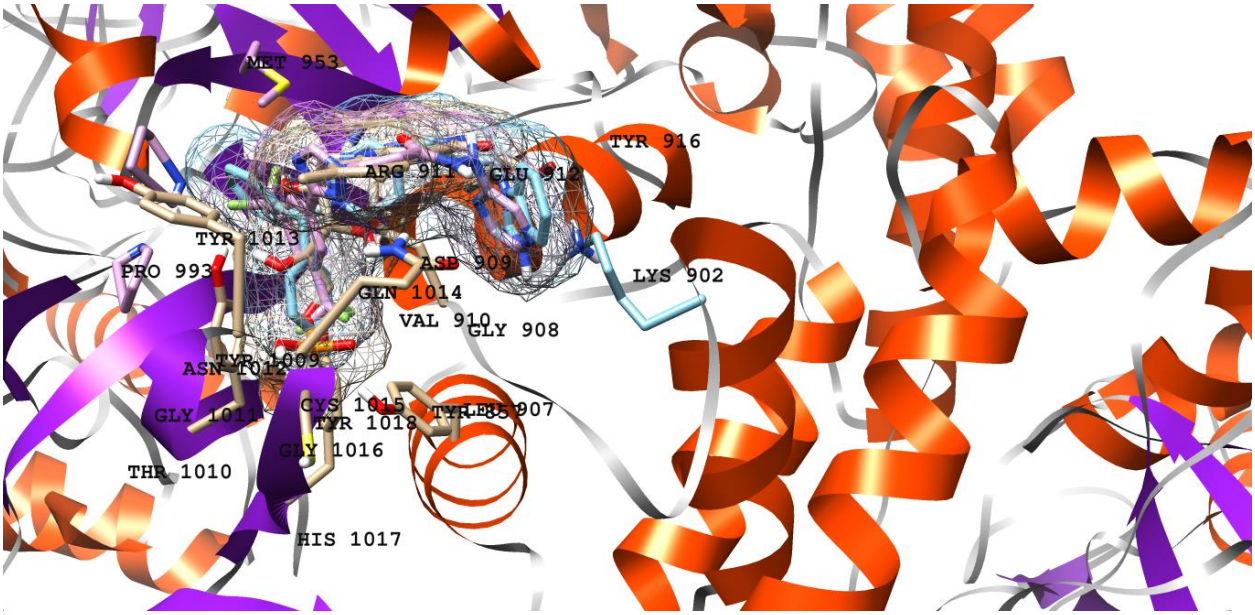



**Figure 2.** 3D representation of PLpro active residues and three selected drugs riboflavin 5'-monophosphate, oxatomide and selinexor in active pocket

3CL-pro contains two chymotrypsin like β-domains and an α-helical domain[41]. Domains I and II have the substrate-binding domain in between them and it is represented in figure (Figure 3a). Residues spanning the active site around the drugs procaine penicillin G, enasidenib and edoxaban

are Thr24, Thr25, Thr26, Leu27, His41, Cys44, Thr45, Ser46, Glu47, Met49, Leu50, Phe140, Leu141, Asn142, Gly143, Ser144, Cys145, His163, Met165, Glu166, Leu167, Pro168 and Gln189 (Figure 3b).

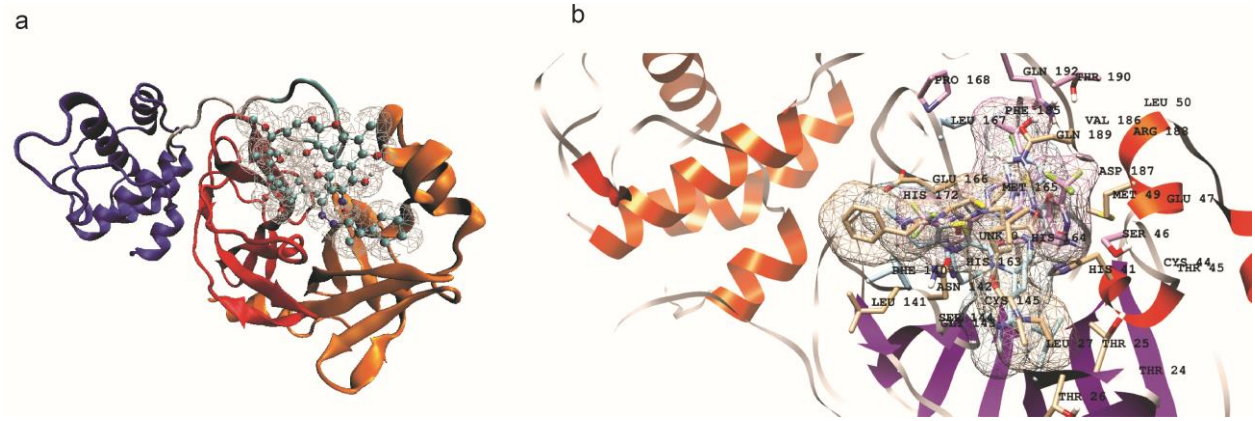



**Figure 3.** a) 3D structure of 3CLpro highlighting three domains, the two β-domains are shown in orange and red colors respectively, blue shows α-helical domain and drug is shown with CPK representation b) 3D representation of active site residues of 3CLpro surrounding the active drugs

Residues His41, Ser46, Met49, leu141, Asn142, Glu166, Pro168 and Gln189 are involved in hydrogen bonding with the inhibitors while Asn142 in all three drugs is forming a salt-bridge interaction as well.

RdRp contains a RdRp domain, a nidovirus N-terminal extension domain, both connected by an interface domain[39] (Fig. 4a). The inhibitor binding site is present on the RdRp domain for RdRp inhibition.

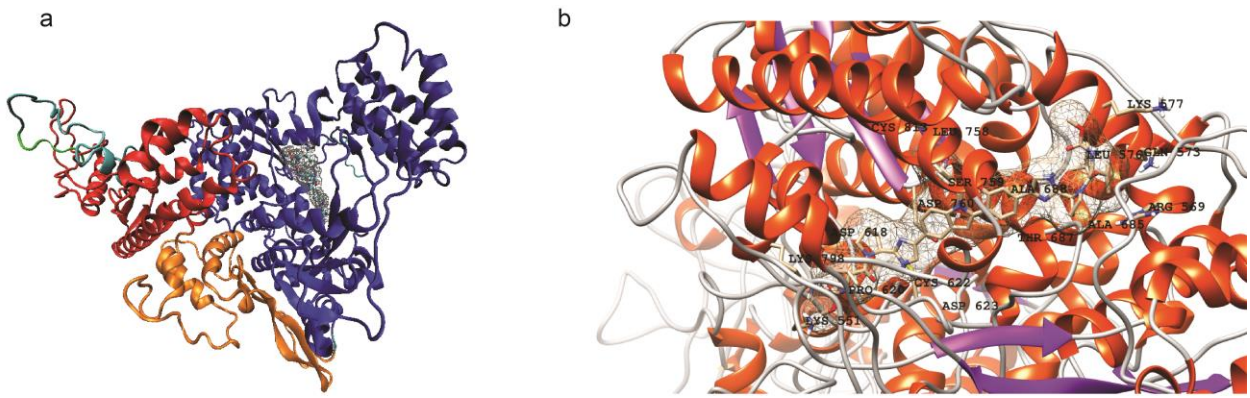



**Figure 4:** a) Structure of RdRp enzyme highlighting different domains: N-terminal domain (red), RdRp domain (blue), interface domain (orange) an additional β-hairpin (green). b) 3D representation of active site residues surrounding the drug elbasvir.

Elbasvir, that is a direct antiviral medication, shows good binding with RdRp (Figure 4b). A number of hydrophobic residues are involved in binding of elbasvir and anidulafungin into the binding site of RdRp. Major contributing residues in drug binding are: Ile548, Ser549, Arg569, Ile589, Ala685, Ser759, Leu758, Ala688, Gln573, Leu576, Asp760, Asp761 and Cys622. Hydrogen bonding interactions are formed by Ile548, Ser549, Ser759, Cys622, Asp760, Ala550, Lys551, Tyr689, Lys798, Lys577, Cys813 and Ser814.

Helicase contains five domains namely N-terminal zinc-binding domain, stalk domain, 1A, 2A and 1B with inhibitor binding site between domain 1A and 1B (Figure 5a). The drugs flavin adenine dinucleotide, desmopressin, glecaprevir and rifabutin inhibit viral helicase protein with lowest binding affinities. These drugs mainly form hydrogen bonding and electrostatic interactions due to the presence of majority hydrophilic amino acid residues such as: Asp260, Glu261, Asn265, His290, Glu319, Lys320, Arg442, Arg443, His464, Lys465, Ser539, Glu540, Asp542, Arg567 and Lys569. Both Arg442 and Arg443 form salt-bridge interactions with the drugs contributing to tight binding. Figure 5b shows a snapshot of all the residues involved in binding

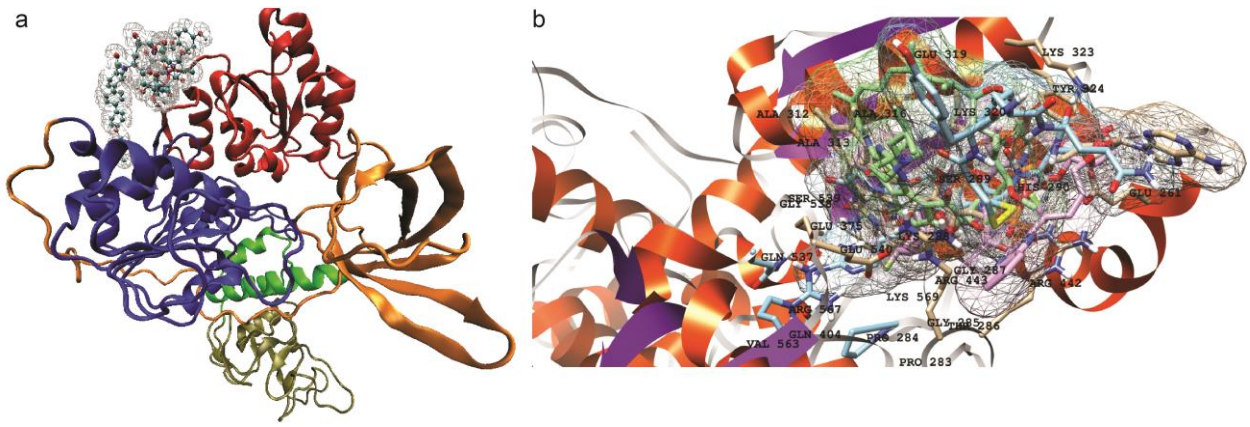



**Figure 5:** a) 3D representation of helicase structure highlighting N-terminal zinc-binding domain (tan), stalk domain (green), 1A (blue), 2A (red) and 1B (orange) with inhibitor binding site between domain 1A and 1B b) 3D representation of active site residues surrounding the drugs

In a second category of targets, Host translation inhibitor or NSP1 is selected for detailed analysis. Tadalafil is a drug that improves exercise capacity by relaxing muscles of the blood vessels and increase of blood flow. Tadalafil and enasidenib show promising results and best energy values in our docking experiments with NSP1. The binding site contains hydrophobic residues such as Leu39, Leu88, Val89, Leu123, Phe143 and Pro153. Several hydrogen bonds are formed by amino acids: Arg43, Lys72, Glu87, Lys125, Asp144, Tyr145 and Asp144.

Our third category of targets include S and N proteins. S protein consists of two functional subunits (S1 and S2) (Fig 6a). it is reported that S protein binds to human ACE2 to enter the cell correlating with its speedy dissemination[36]. A structure for spike protein is reported[35] and deposited in RCSB protein databank (PDB) having PDB ID 6VXX. It was matched with the structure from Zhang lab server and RMSD value for backbone atoms came out to be 1.88 Å. We have observed three different drugs to show strong potency for S protein target: anidulafungin, velpatasvir and terlipressin. The key residues that are present in the active binding pocket include:

Ser45, Ser46, Leu48, Glu281, Leu303, Ser305, Phe306, Thr307, Glu309, Lys310, Arg815, Phe823. Asn824, Thr827, Leu828, Ala829, Ala831, Ala846, Arg847, Leu849, Lys854, Leu945, Gly946, Leu1203, Lys1205, Tyr1209, Ile1210, Pro1213, Ile1216, and Trp1217. Terlipressin is a large compound covering a wide area over the protein surface. It is involved in majority hydrogen bonds and electrostatic interactions.

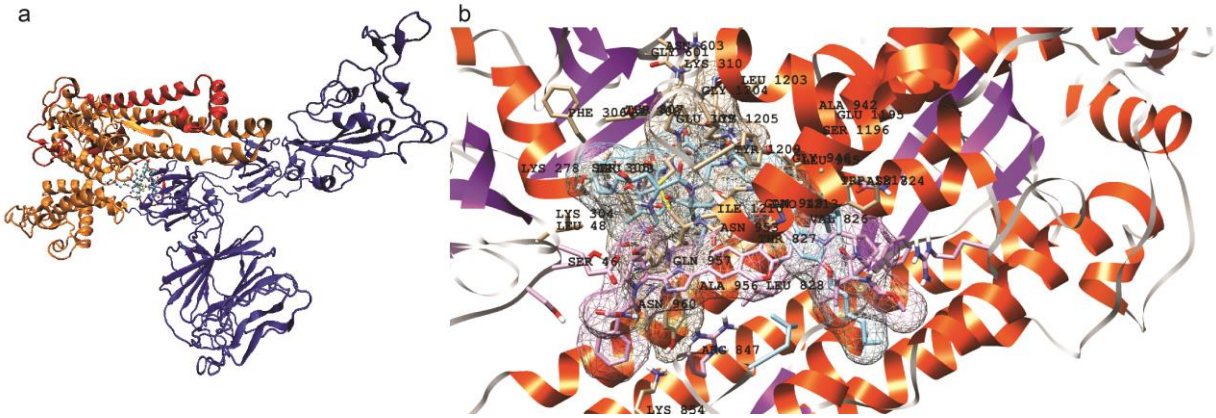



**Figure 6:** a) 3D representation of S protein structure highlighting S1 (blue), S2 (red) and S2' (orange) with inhibitor binding site b) 3D representation of active site residues surrounding the three selected drugs for S protein anidulafungin, velpatasvir and terlipressin

The model of the SARS-CoV-2 nucleocapsid phosphoprotein was superimposed with the NMR structure of RNA-binding domain of this protein[45] (PDB ID 6YI3) and RMSD value was 1.14 Å. It is observed that the drug anidulafungin binds at the interface of the protein and interacts with many polar residues naming a few: Arg92, Tyr109, Asn150, Arg259, Gln272, Gln283 and Asp399.

Anidulafugin, a member of the echinocandin class of antifungals, appears to be a promising candidate against SARS-CoV-2 and has high predicted KIBA scores and low docking energies against key viral proteins including RdRp, helicase, exonuclease, S and N. Anidulafungin is used for the treatment of mucosal and invasive fungal infections[46]. Isavuconazonium, another triazole antifungal approved for the treatment of invasive aspergillosis and invasive mucormycosis,

showed high binding affinities for nsp2 and N proteins (Table 1). Interestingly, Anidulafungin was shown in a recent study to exhibit strong in-vitro antiviral activity against SARS-CoV-2 virus with an IC50 value of 4.64 μmol[47]. Although rare, cases of invasive fungal infections have been reported in literature in association with severe influenza[48] and severe acute respiratory syndrome (SARS) virus infection[49]. Whether COVID-19 patients are also at increased risk of invasive fungal infections is yet to be investigated. Anidulafungin with its potential antiviral activity against SARS-CoV2 and proven antifungal activity against invasive fungal infections appears to be a promising candidate.

Our study highlights the potential of HCV protease inhibitors (elbasvir, velpatasvir, glecraprevir and pibrentasvir) in inhibiting SARS-CoV-2 proteins. Elbesvir, an HCV NS5A inhibitor, demonstrated best docking energies with RdRp, a key viral enzyme and an important therapeutic target for COVID-19. In other studies, elbasvir[50] and velpatasvir[51] have been reported to dock well with 3CLpro. Glecraprevir, an HCV NS3/4A protease inhibitor, is often given in combination with pibrentasvir, an NS5A inhibitor, for treatment experienced cases of HCV infection. While safety profile of this combination has been well-established, the potential of this combination in treating COVID-19 has not been examined so far. Rifabutin and rifapentine belong to the rifamycin group of antibiotics. Both rifabutin and rifapentin inhibits mycobacterial DNA-dependent RNA polymerase, thereby suppressing the initiation of RNA formation and are used in combination with other drugs for the treatment of tuberculosis. With the potential of inhibiting RdRp and helicase amongst other SARS-CoV-2 viral proteins, rifabutin can be a promising therapeutic option that needs further investigation.

Selinexor, an FDA approved drug for the treatment of relapsed or refractory multiple myeloma, is predicted to inhibit SARS-CoV-2 PLpro in our study. Selinexor is an inhibitor of Exportin-1

(XPO1) which is an important protein involved in the transport of multiple proteins across nuclear envelope. In addition to its role in cancer, XPO1 also plays an important role in facilitating transport of viral proteins across the host cell nuclear envelope[52]. Studies have shown that XPO1 plays a critical role in SARS-CoV viral replication by controlling the export of certain SARS-CoV proteins out of the nucleus[53]. We hypothesize that selinexor has both direct antiviral activity through inhibiting PLpro as well as indirect activity through modulation of host target proteins. A multinational clinical trial has recently been launched to study the efficacy of selinexor in patients with severe COVID-19 disease. Enasidenib, another anti—neoplastic drug, inhibits mutant forms of *isocitrate dehydrogenase* 2 (IDH2) and is approved for the treatment of refractory form of acute myeloid leukemia (AML). Here, we have shown that enasidenib demonstrates high binding affinity with two key SARS-CoV-2 viral proteins: 3CLpro and nsp1.

Edoxaban is a rapidly acting selective factor Xa inhibitor and belongs to Novel Oral Anti-Coagulant (NOACs) class of drugs. In our study, edoxaban has demonstrated best binding affinity with 3CLpro that is a key enzyme involved in SARS-CoV-2 viral replication and an emerging drug target. Recently, a wealth of clinical data has suggested that COVID-19 is a hypercoagulable state associated with increased incidence of thrombosis in critically ill patients. Therefore, anticoagulation is being recommended for prophylactic and therapeutic[54]. Given the potential to inhibit 3CLpro, ease of oral administration and anticoagulant activity, edoxaban appears to be a promising candidate drug for treating COVID-19. Further in vitro and clinical studies are warranted.

Flavin mononucleotide (FMN; also known as riboflavin-5'-phosphate) and flavin adenine dinucleotide (FAD) are two coenzymes produced from riboflavin (vitamin B2) and function as prosthetic group of various oxidoreductases. FMN is predicted to have high binding affinity with

PLpro while FAD has high predicted affinities for RdRp, helicase, S and N amongst other SARS-CoV-2 viral proteins. FAD is a redox-active cofactor that is essential for the functioning of flavoenzymes that play critical role in many biochemical processes such as oxidative metabolism of macromolecules and electron transport chain[55]. Studies have shown that the intracellular redox state may play an important role in inhibiting viral replication[56]. In one study FAD was shown to enhance the antiviral activity of interferons against herpes virus-1 and influenza virus type A[56]. FAD can also increase intracellular activity of glutathione and nitric oxide synthase, both of which may play important roles in inhibiting viral replication. FAD has shown binding affinity with spike protein of SARS-CoV-2 in another docking based study (vina score -7.3)[57]. In another study using molecular docking, riboflavin was found to interact with Papain-like proteinase (PLpro) and flavin mononucleotide (FMN) interacted with 3C-like main protease (3CLpro) of SARS- CoV-2 virus[43]. Other studies have shown that FAD can decrease lung injury in influenza A H5N1 infected mice by altering the levels of immune response related genes[58]. In conclusion, the results of our study coupled with evidence from literature dictate that FAD may play an important role against SARS-CoV-2 virus by directly targeting the virus as well as host response to the viral replication. However, further evidence from in-vitro and in-vivo studies is required.

**Conclusion**

In conclusion, we have combined deep learning and molecular docking simulations to identify most promising candidates from the list of FDA approved drugs that can be repurposed to treat COVID-19. These drugs include anidulafungin, velpatasvir, glecaprevir, rifabutin, procaine penicillin G, tadalafil, riboflavin 5'-monophosphate, flavin adenine dinucleotide, terlipressin, desmopressin, elbasvir, oxatomide, enasidenib, edoxaban and selinexor amongst others. Further *in vitro* studies are indicated to investigate antiviral potential of some of

these drugs. For drugs with proven *in vitro* antiviral activity against SARS-CoV-2, clinical trials are warranted.

## Description of Supplemental Data

Supplemental data include 2 figures and 3 tables.

## Declaration of interest

All authors declare no competing interest.

## Acknowledgements

We are very thankful to Riazuddin S for his guidance and review of the manuscript.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Authors' Contribution (CRediT author statement):

Muhammad Umer Anwar: Methodology, Software, Formal analysis, Resources Farjad Adnan: Methodology, Software, Resources, Data Curation, Writing - Original Draft Asma Abro: Methodology, Software, Validation, Formal analysis, Writing - Original Draft Rayyan Ahmad Khan: Methodology, Software, Formal analysis, Resources Asad ur Rehman: Formal analysis, Resources, Data Curation Muhammad Osama: Formal analysis, Data Curation Saad Javed: Formal

analysis, Data Curation Ahmadullah Baig: Data Curation, Visualizations  Muhammad Raffey Shabbir: Formal analysis, Data Curation Muhammad Zaman Assir: Conceptualization, Methodology, Validation, Writing - Original Draft, Supervision

**Online Resources:**

C-I-TASSER: https://zhanglab.ccmb.med.umich.edu/COVID-19/

PubChem**:** https://pubchem.ncbi.nlm.nih.gov/

DrugBank: https://www.drugbank.ca/

RCSB PDB: https://www.rcsb.org/

References

1. de Wit E, van Doremalen N, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. *Nature Reviews Microbiology* 2016;14(8):523-34.

2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270-73.

3. https://www.worldometers.info/coronavirus/ last visited on April 28, 2020

4. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics* 2016;47:20-33.

5. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. *Nature Reviews Drug Discovery* 2019;18(1):41-58.

6. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nature Chemical Biology* 2008;4(11):682-90.

7. Medina-Franco JL, Giulianotti MA, Welmaker GS, Houghten RA. Shifting from the single to the multitarget paradigm in drug discovery. *Drug Discovery Today* 2013;18(9):495-501.

8. Feature engineering in Context-Dependent Deep Neural Networks for conversational speech transcription. 2011 IEEE Workshop on Automatic Speech Recognition & Understanding; 2011 11-15 Dec. 2011.

9. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436-44.

10. Goodfellow I, Bengio Y, Courville A. *Deep learning*: MIT press, 2016.

11. Lillicrap TP, Cownden D, Tweed DB, Akerman CJ. Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications* 2016;7(1):1-10.

12. Pahikkala T, Airola A, Pietilä S, Shakyawar S, Szwajda A, Tang J, et al. Toward more realistic drug–target interaction predictions. *Briefings in bioinformatics* 2015;16(2):325-37.

13. Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, et al. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic acids research* 2016;45(D1):D271-D81.

14. Davis MI, Hunt JP, Herrgard S, Ciceri P, Wodicka LM, Pallares G, et al. Comprehensive analysis of kinase inhibitor selectivity. *Nature biotechnology* 2011;29(11):1046-51.

15. He T, Heidemeyer M, Ban F, Cherkasov A, Ester M. SimBoost: a read-across approach for predicting drug-target binding affinities using gradient boosting machines. *J Cheminform* 2017;9(1):24-24.

16. Tang J, Szwajda A, Shakyawar S, Xu T, Hintsanen P, Wennerberg K, et al. Making Sense of Large-Scale Kinase Inhibitor Bioactivity Data Sets: A Comparative and Integrative Analysis. *Journal of Chemical Information and Modeling* 2014;54(3):735-43.

17. Tian K, Shao M, Wang Y, Guan J, Zhou S. Boosting compound-protein interaction prediction by deep learning. *Methods* 2016;110:64-72.

18. Large-scale prediction of drug-target interactions from deep representations. 2016 International Joint Conference on Neural Networks (IJCNN); 2016. IEEE.

19. Wen M, Zhang Z, Niu S, Sha H, Yang R, Yun Y, et al. Deep-learning-based drug–target interaction prediction. *Journal of proteome research* 2017;16(4):1401-09.

20. Gomes J, Ramsundar B, Feinberg EN, Pande VS. Atomic convolutional networks for predicting protein-ligand binding affinity. *arXiv preprint arXiv:1703.10603* 2017.

21. Öztürk H, Özgür A, Ozkirimli E. DeepDTA: deep drug–target binding affinity prediction. *Bioinformatics* 2018;34(17):i821-i29.

22. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Annual review of biophysics and biomolecular structure* 2003;32(1):335-73.

23. Lyne PD. Structure-based virtual screening: an overview. *Drug Discovery Today* 2002;7(20):1047-55.

24. Jain AN. Scoring functions for protein-ligand docking. *Current Protein and Peptide Science* 2006;7(5):407-20.

25. Huang S-Y, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics* 2010;12(40):12899-908.

26. Carneiro J, Stewart J. Rethinking" Shape Space": Evidence from Simulated Docking Suggests that Steric Shape Complementarity is not Limiting for Antibody--Antigen Recognition and Idiotypic Interactions. *Journal of theoretical biology* 1994;169(4):391-402.

27. Gorgulla C, Boeszoermenyi A, Wang Z-F, Fischer PD, Coote P, Das KMP, et al. An open-source drug discovery platform enables ultra-large virtual screens. *Nature* 2020:1-8.

28. Zhang C, Zheng W, Huang X, Bell EW, Zhou X, Zhang Y. Protein structure and sequence re-analysis of 2019-nCoV genome refutes snakes as its intermediate host or the unique similarity between its spike protein insertions and HIV-1. *Journal of proteome research* 2020.

29. Yang J, Roy A, Zhang Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics* 2013;29(20):2588-95.

30. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry* 2010;31(2):455-61.

31. PubChem." https://pubchem.ncbi.nlm.nih.gov/. Accessed 6 Apr. 2020.

32. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 2004;25(13):1605-12.

33. Cousins KR. ChemDraw Ultra 9.0. CambridgeSoft, 100 CambridgePark Drive, Cambridge, MA 02140. www. cambridgesoft.com. See Web site for pricing options. *Journal of the American Chemical Society* 2005;127(11):4115-16.

34. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics* 1996;14(1):33-38.

35. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 2020.

36. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;367(6485):1444-48.

37. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 2020;367(6483):1260-63.

38. Shannon A, Tuyet Le NT, Selisko B, Eydoux C, Alvarez K, Guillemot J-C, et al. Remdesivir and SARS-CoV-2: structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *Antiviral Research* 2020:104793.

39. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 2020:eabb7498.

40. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications* 2019;10(1):2342.

41. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, et al. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* 2020:eabb3405.

42. Jin Z, Du X, Xu Y, Deng Y, Liu M, Zhao Y, et al. Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 2020.

43. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B* 2020.

44. Sheahan TP, Sims AC, Zhou S, Graham RL, Pruijssers AJ, Agostini ML, et al. An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Science Translational Medicine* 2020:eabb5883.

45. Sarma P, Sekhar N, Prajapat M, Avti P, Kaur H, Kumar S, et al. In-silico homology assisted identification of inhibitor of RNA binding against 2019-nCoV N-protein (N terminal domain). *Journal of Biomolecular Structure and Dynamics* 2020(just-accepted):1-11.

46. Vazquez JA. Anidulafungin: a new echinocandin with a novel profile. *Clinical therapeutics* 2005;27(6):657-73.

47. Jeon S, Ko M, Lee J, Choi I, Byun SY, Park S, et al. Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *BioRxiv* 2020.

48. Alshabani K, Haq A, Miyakawa R, Palla M, Soubani AO. Invasive pulmonary aspergillosis in patients with influenza infection: report of two cases and systematic review of the literature. *Expert review of respiratory medicine* 2015;9(1):89-96.

49. Wang H, Ding Y, Li X, Yang L, Zhang W, Kang W. Fatal aspergillosis in a patient with SARS who was treated with corticosteroids. *New England Journal of Medicine* 2003;349(5):507-08.

50. Sekhar T. Virtual Screening based prediction of potential drugs for COVID-19: Preprints, 2020.

51. Chen YW, Yiu C-PB, Wong K-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. *F1000Research* 2020;9.

52. Widman DG, Gornisiewicz S, Shacham S, Tamir S. In vitro toxicity and efficacy of verdinexor, an exportin 1 inhibitor, on opportunistic viruses affecting immunocompromised individuals. *PloS one* 2018;13(10).

53. Freundt EC, Yu L, Park E, Lenardo MJ, Xu X-N. Molecular determinants for subcellular localization of the severe acute respiratory syndrome coronavirus open reading frame 3b protein. *Journal of virology* 2009;83(13):6631-40.

54. Bikdeli B, Madhavan MV, Jimenez D, Chuich T, Dreyfus I, Driggin E, et al. COVID-19 and Thrombotic or Thromboembolic Disease: Implications for Prevention, Antithrombotic Therapy, and Follow-up. *Journal of the American College of Cardiology* 2020.

55. Giancaspero TA, Busco G, Panebianco C, Carmone C, Miccolis A, Liuzzi GM, et al. FAD synthesis and degradation in the nucleus create a local flavin cofactor pool. *Journal of Biological Chemistry* 2013;288(40):29069-80.

56. Yokochi S, Ishiwata Y, Saito H, Ebinuma H, Tsuchiya M, Ishii H. Stimulation of antiviral activities of interferon by a liver extract preparation. *Arzneimittel-Forschung* 1997;47(8):968-74.

57. Zhang J, Wang G, Li H, Wei T, Wang H, Wu X, et al. In Silico Screening of Potential Spike Glycoprotein Inhibitors of SARS-CoV-2 with Drug Repurposing Strategy. 2020.

58. Huang F, Zhang C, Liu Q, Zhao Y, Zhang Y, Qin Y, et al. Identification of amitriptyline HCl, flavin adenine dinucleotide, azacitidine and calcitriol as repurposing drugs for influenza A H5N1 virus-induced lung injury. *PLoS Pathogens* 2020;16(3):e1008341.

Table 1: FDA approved drugs with best predicted KIBA and AutoDock vina binding affinity values against selected SARS-CoV-2 Viral Proteins

| Sr # | Drug Name | PubChem Drug ID | Drug Action | Protein Name | KIBA Score | Docking Binding Affinity Values (kcal/mol) |
|---|---|---|---|---|---|---|
| A. | Antimicrobials | | | | | |
| 1 | Anidulafungin | 166548 | Antifungal | RNA-directed RNA polymerase (RdRp) | 11.7602 | -13.4 |
| | | | | Helicase (Hel) | 11.7768 | -9.0 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.7008 | -13.2 |
| | | | | Uridylate-specific endoribonuclease | 11.7605 | -12.3 |
| | | | | Surface glycoprotein (S) | 11.7623 | -15.3 |
| | | | | N | 11.6237 | -14.3 |
| 2 | Isavuconazonium | 6918606 | Antifungal | Non-structural protein 2 (nsp2) | 11.5459 | -11.1 |
| | | | | N | 11.6357 | -10.9 |
| 3 | Procaine Penicillin G | 5903 | Antibiotic | Proteinase 3CL-PRO | 11.2148 | -9.3 |
| 4 | Quinupristin | 5388937 | Antibiotic | RNA-directed RNA polymerase (RdRp) | 11.7965 | -12.2 |
| | | | | Surface glycoprotein (S) | 11.5520 | -12.1 |
| | | | | N | 11.6461 | -11.8 |
| 5 | Rifapentine | 135659016 | Antibiotic | Non-structural protein 9 (nsp9) | 11.6792 | -8.6 |
| | | | | Surface glycoprotein (S) | 11.6241 | -11.4 |
| | | | | N | 11.7584 | -11.6 |
| 6 | Rifabutin | 135415564 | Antibiotic | Non-structural protein 2 (nsp2) | 11.8004 | -11 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.8209 | -11.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Helicase (Hel) | 11.8829 | -10.5 |
| | | | | Surface glycoprotein (S) | 11.7498 | -12.7 |
| | | | | N | 11.8253 | -11.7 |
| 7 | Polymyxin B | 49800004 | Antibiotic | RNA-directed RNA polymerase (RdRp) | 11.8949 | -14.2 |
| | | | | Surface glycoprotein (S) | 11.8941 | -12.6 |
| | | | | N | 11.7840 | -12.7 |
| 8 | Cobicistat | 25151504 | Antiviral | Non-structural protein 2 (nsp2) | 11.8630 | -10.7 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.9491 | -11.4 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.8669 | -14.1 |
| 9 | Elbasvir | 71661251 | Antiviral | RNA-directed RNA polymerase (RdRp) | 11.9396 | -13.6 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.8797 | -13.5 |
| | | | | Uridylate-specific endoribonuclease | 11.9112 | -11.7 |
| | | | | Surface glycoprotein (S) | 11.8991 | -13.7 |
| | | | | N | 11.8131 | -13.8 |
| 10 | Velpatasvir | 67683363 | Antiviral | Non-structural protein 2 (nsp2) | 11.4887 | -10.8 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.5925 | -11.9 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.3040 | -13.5 |
| | | | | Uridylate-specific endoribonuclease | 11.5538 | -11.2 |
| | | | | Surface glycoprotein (S) | 11.7373 | -14.5 |
| | | | | N | 11.5965 | -11.7 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 11 | Pibrentasvir | 58031952 | Antiviral | RNA-directed RNA polymerase (RdRp) | 11.8812 | -12.7 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.7442 | -12.3 |
| | | | | Uridylate-specific endoribonuclease | 11.8894 | -11.4 |
| | | | | Surface glycoprotein (S) | 11.8692 | -13.1 |
| | | | | N | 11.7755 | -12.9 |
| 12 | Glecaprevir | 66828839 | Antiviral | Non-structural protein 2 (nsp2) | 11.8902 | -11.3 |
| | | | | Helicase (Hel) | 11.9120 | -10.5 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.9551 | -11.7 |
| | | | | 2'-O-methyltransferase (2'-O-MT) | 11.8422 | -11.8 |
| | | | | Surface glycoprotein (S) | 11.7496 | -11.8 |
| | | | | N | 12.0036 | -11.9 |
| B. Hormonal | | | | | | |
| 13 | Abiraterone | 132971 | Antiandrogen | M | 11.5199 | -7.7 |
| 14 | Amcinonide | 443958 | Corticosteroid | Helicase (Hel) | 11.8256 | -9.1 |
| | | | | M | 11.6823 | -7.5 |
| 15 | Atosiban | 5311010 | Tocolytic | RNA-directed RNA polymerase (RdRp) | 11.8447 | -12.5 |
| | | | | Surface glycoprotein (S) | 11.9298 | -11.7 |
| 16 | Carbetocin | 16681432 | Uterotonic | Helicase (Hel) | 11.6491 | -9.8 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.6623 | -13.8 |
| | | | | N | 11.5698 | -12.3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 17 | Cortisone | 5745 | Corticosteroid | M | 11.6792 | -7.3 |
| 18 | Danazol | 28417 | Androgen | M | 11.8513 | -7.2 |
| 19 | Deoxycorticosterone | 5952 | Corticosteroid | M | 11.6867 | -7.6 |
| 20 | Desmopressin | 5311065 | ADH Analog | RNA-directed RNA polymerase (RdRp) | 11.9439 | -13.6 |
| | | | | Helicase (Hel) | 11.9303 | -10.6 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.9756 | -12.3 |
| | | | | Uridylate-specific endoribonuclease | 11.9761 | -11 |
| | | | | Surface glycoprotein (S) | 12.0064 | -13.1 |
| | | | | N | 11.8459 | -12.4 |
| 21 | Ethynodiol diacetate | 9270 | Progesterone Receptor Agonist | M | 11.7367 | -7.5 |
| 22 | Pentetreotide | 72128 | Octreotide Analog | RNA-directed RNA polymerase (RdRp) | 12.0372 | -11.9 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.9211 | -12.9 |
| | | | | Uridylate-specific endoribonuclease | 12.0956 | -12.7 |
| | | | | Surface glycoprotein (S) | 11.8687 | -14.5 |
| | | | | N | 11.8935 | -12.4 |
| 23 | Somatostatin | 16129706 | Octreotide Analog | RNA-directed RNA polymerase (RdRp) | 11.8702 | -15.3 |
| | | | | Helicase (Hel) | 11.9113 | -9.7 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.7610 | -12 |
| | | | | Uridylate-specific endoribonuclease | 11.8051 | -12 |

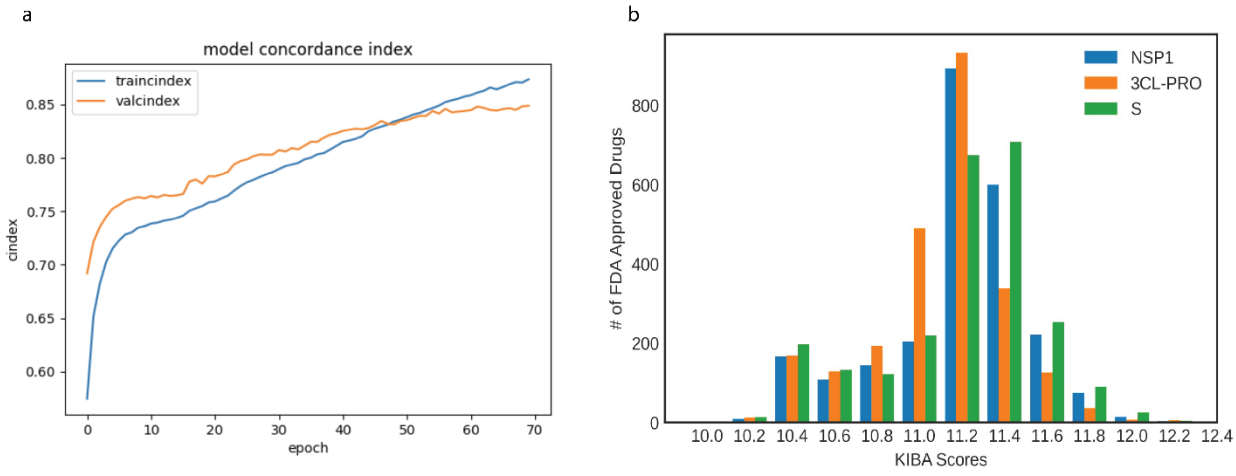
| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Surface glycoprotein (S) | 11.8407 | -14.4 |
| | | | | N | 11.7700 | -12.5 |
| 24 | Progesterone | 5994 | Sex Steroid | M | 11.6636 | -7.3 |
| 25 | Prednisone | 5865 | Corticosteroid | M | 11.7184 | -7.3 |
| 26 | Terlipressin | 72081 | Vasopressin Analog | RNA-directed RNA polymerase (RdRp) | 11.5954 | -13.8 |
| | | | | Helicase (Hel) | 11.6692 | -9.8 |
| | | | | Uridylate-specific endoribonuclease | 11.7842 | -11.9 |
| | | | | Surface glycoprotein (S) | 11.5950 | -14.5 |
| | | | | N | 11.5939 | -12.4 |
| C. Anti-neoplastic | | | | | | |
| 27 | Vindesina | 11643449 | | Guanine-N7 methyltransferase (ExoN) | 11.4910 | -12.8 |
| | | | | Surface glycoprotein (S) | 11.7376 | -11.6 |
| | | | | N | 11.7234 | -10.9 |
| 28 | Nilotinib | 644241 | Tysrosine-Kinase Inhibitor | Host translation inhibitor nsp1 | 11.7249 | -8 |
| | | | | Non-structural protein 2 (nsp2) | 11.7242 | -12.2 |
| | | | | Papain-like proteinase | 11.7398 | -6.7 |
| | | | | Proteinase 3CL-PRO | 11.4559 | -8.5 |
| | | | | E | 11.4128 | -7.5 |
| 29 | Exemestane | 60198 | Aromatase Inhibitor | M | 11.7052 | -7.5 |
| 30 | Etoposide | 36462 | Topoisomerase Inhibitor | Proteinase 3CL-PRO | 11.7101 | -8.6 |
| 31 | Epirubicin | 41867 | Anthracycline | Proteinase 3CL-PRO | 11.5659 | -8.5 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | antineoplastic antibiotic | | | |
| 32 | Enasidenib | 89683805 | Isocitrate Dehydrogenase Inhibitor | Host translation inhibitor nsp1 | 11.9167 | -8.5 |
| | | | | Non-structural protein 2 (nsp2) | 11.9669 | -10.7 |
| | | | | Proteinase 3CL-PRO | 11.8091 | -8.7 |
| | | | | Papain-like proteinase | 11.8572 | -6.9 |
| 33 | Daunorubicin | 30323 | Anthracycline antineoplastic antibiotic | Proteinase 3CL-PRO | 11.5225 | -8.5 |
| 34 | Cabazitaxel | 9854073 | Microtubule Inhibitors | RNA-directed RNA polymerase (RdRp) | 11.9236 | -12.3 |
| | | | | Uridylate-specific endoribonuclease | 12.0602 | -11.7 |
| | | | | N | 11.9556 | -11.3 |
| 35 | Docetaxel | 148124 | Microtubule Inhibitors | Non-structural protein 10 (nsp10) | 11.7887 | -10.1 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.8086 | -11.8 |
| | | | | N | 11.7737 | -11.4 |
| 36 | Brigatinib | 68165256 | Tyrosine Kinase Inhibitor | Non-structural protein 2 (nsp2) | 11.7242 | -10.7 |
| 37 | Selinexor | 71481097 | Antineoplastic | Papain-like proteinase | 11.5764 | -7.1 |
| D. Vitamins | | | | | | |
| 38 | Cholecalciferol | 5280795 | Vitamin D- Steroid | M | 11.7619 | -7.2 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 39 | Flavin adenine dinucleotide | 643975 | Coenzyme | Non-structural protein 2 (nsp2) | 11.7630 | -11.8 |
| | | | | RNA-directed RNA polymerase (RdRp) | 11.6978 | -12.8 |
| | | | | Helicase (Hel) | 11.8304 | -11.2 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.3726 | -13.6 |
| | | | | Uridylate-specific endoribonuclease | 11.8032 | -12.3 |
| | | | | Surface glycoprotein (S) | 12.2218 | -12.9 |
| | | | | N | 11.6575 | -13.3 |
| 40 | Riboflavin 5'-monophosphate | 643976 | Vitamin B2 derivative | Papain-like proteinase | 11.8168 | -7 |
| E. Miscellaneous | | | | | | |
| 41 | Dabigatran etexilate | 135565674 | Anticoagulant | Helicase (Hel) | 11.8329 | -9.7 |
| | | | | Guanine-N7 methyltransferase (ExoN) | 11.7765 | -12.5 |
| | | | | Surface glycoprotein (S) | 11.6961 | -12.7 |
| | | | | N | 11.7036 | -11.2 |
| 42 | Dihydroergotamine | 10531 | Ergot Derivative | Host translation inhibitor nsp1 | 11.9048 | -8.2 |
| | | | | Helicase (Hel) | 11.9629 | -9.2 |
| 43 | Edoxaban | 10280735 | Anticoagulant | Proteinase 3CL-PRO | 11.5352 | -8.7 |
| 44 | Elexacaftor | 134587348 | Corrector of the CFTR protein | Non-structural protein 2 (nsp2) | 11.8181 | -11.1 |
| 45 | Ergotamine | 8223 | Ergot alkaloid with vasoconstr | Host translation inhibitor nsp1 | 11.9048 | -8 |
| | | | | Proteinase 3CL-PRO | 11.3610 | -8.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | ictor and analgesic property. | Helicase (Hel) | 11.9629 | -9.3 |
| 46 | Manidipine | 4008 | Calcium Channel Blocker | Non-structural protein 2 (nsp2) | 12.0029 | -11.1 |
| 47 | Mivacurium | 5281042 | Neuromuscular Blocker | Guanine-N7 methyltransferase (ExoN) | 11.7171 | -12.1 |
| | | | | Surface glycoprotein (S) | 11.5691 | -13.2 |
| 48 | Tadalafil | 110635 | PDE-5 Inhibitor | Host translation inhibitor nsp1 | 11.7874 | -8.5 |
| | | | | Non-structural protein 2 (nsp2) | 11.6870 | -11.2 |
| 49 | Oxatomide | 4615 | First-generation H1-antihistamine | Papain-like proteinase | 12.1627 | -7.3 |

Supplementary Figures:




Supplementary Figure 1: Prediction of KIBA scores using modified DeepDTA

Performance of modified DeepDTA model demonstrating a high concordance index of 0.899.

b) Frequency distribution of DeepDTA for FDA approved drugs against three selected SARS-CoV-2 viral proteins nsp1, 3CLpro and S. Drugs with the highest KIBA scores were shortlisted for docking simulations.

Supplementary Figure 2: Interactive plot of selected FDA approved drugs with KIBA scores and docking binding affinity values against key SARS-CoV-2 viral proteins

Supplementary Table 1: Predicted KIBA scores of all FDA approved and experimental drugs in DrugBank against all 24 SARS-CoV-2 viral proteins

Supplementary Table 2: KIBA scores of selected 168 FDA approved drugs against SARS-CoV-2 viral proteins

Supplementary Table 3: Docking binding affinity values for 168 drugs against all predicted and/or experimentally proven active sites of SARS-CoV-2 proteins

AUTHOR INFORMATION

### Corresponding Author

*(Word Style “FA_Corresponding_Author_Footnote”). * (Word Style “FA_Corresponding_Author_Footnote”). Give contact information for the author(s) to whom correspondence should be addressed.

### Present Addresses

†If an author’s address is different than the one given in the affiliation line, this information may be included here.

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally. (match statement to author names with a symbol)

### Funding Sources

Any funds used to support the research of the manuscript should be placed here (per journal style).

### Notes

Any additional relevant notes should be placed here.

ACKNOWLEDGMENT

(Word Style “TD_Acknowledgments”). Generally the last paragraph of the paper is the place to acknowledge people, organizations, and financing (you may state grant numbers and sponsors here). Follow the journal’s guidelines on what to include in the Acknowledgments section.

ABBREVIATIONS

CCR2, CC chemokine receptor 2; CCL2, CC chemokine ligand 2; CCR5, CC chemokine receptor 5; TLC, thin layer chromatography.

REFERENCES

(Word Style "TF_References_Section"). References are placed at the end of the manuscript. Authors are responsible for the accuracy and completeness of all references. Examples of the recommended format for the various reference types can be found at http://pubs.acs.org/page/4authors/index.html. Detailed information on reference style can be found in *The ACS Style Guide,* available from Oxford Press.

BRIEFS (Word Style "BH_Briefs"). If you are submitting your paper to a journal that requires a brief, provide a one-sentence synopsis for inclusion in the Table of Contents.

SYNOPSIS (Word Style "SN_Synopsis_TOC"). If you are submitting your paper to a journal that requires a synopsis, see the journal's Instructions for Authors for details.