Profiling Molecular Simulations of SARS-CoV-2 Main Protease (M^{PRO}) Binding to Repurposed Drugs using Neural Network Force Fields

Aayush Gupta*

Department of Chemistry, University of Illinois at Chicago, Chicago, IL 60607, USA

*Correspondence should be addressed to Aayush Gupta (e-mail: aayush.ict@gmail.com).

Abstract

With the current pandemic situation caused by a novel coronavirus disease (COVID-19), there is an urgent call to develop a working therapeutic against it. Efficient computations can minimize the efforts by identifying a subset of drugs that can potentially bind to the COVID-19 main protease or target protein (M^{PRO}). The results of computations are accompanied by an evaluation of their accuracy, which depends on the details described by the model used. Neural network models trained on millions of points and with unmatched accuracies are the best approach to employ in this process. In this work, I first identified and described the interaction sites of the M^{PRO} protein using a geometric deep learning model. Second, I conducted virtual screening (at one of the sites identified) on FDA-approved drugs and selected 91 drugs with the highest binding affinities (below -8.0 kcal/mol). Then, I conducted 10 ns of molecular dynamics (MD) simulations using classical force fields and classified 37 drugs to be binding (including Lopinavir, Saquinavir, and Indinavir) based on the RMSD between MD-binding trajectories. To drastically improve the dynamics profile of the 37 selected drugs, I used the highly accurate neural network force field (ANI) method trained on coupled-cluster method (CCSD(T)/CBS) data points and performed 1 ns of binding dynamics for each drug with the protein. Using this approach, 19 drugs were qualified based on their RMSD cutoffs, and based on free energy (ANI/MM/PBSA) computations, 7 of the drugs were rejected. The final selection of 12 drugs was validated based on an MD trajectory clustering approach where 11 of the 12 drugs (Targretin, Eltrombopag, Rifaximin, Deflazacort, Ergotamine, Doxazosin, Lastacaft, Rifampicin, Victrelis, Trajenta, Toposar, and Indinavir) were confirmed to exhibit binding. Further investigations were performed to study their interactions with the protein and an accurate 2D-interaction map was generated. These findings and mappings of drug-protein interactions are highly accurate and may be potentially used to guide rational drug discovery against COVID-19.

Keywords: COVID-19, molecular dynamics, neural network, CCSD(T)/CBS, ANI/MM/PBSA

Introduction

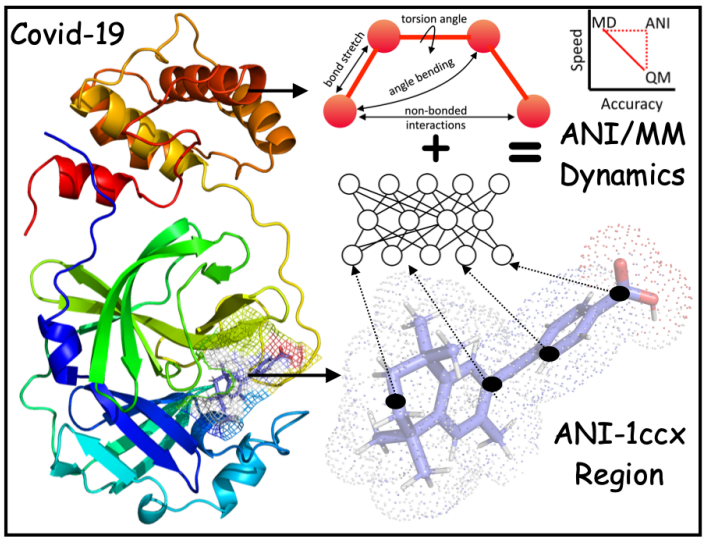
The whole world is experiencing an adverse phase due to the pandemic caused by a viral respiratory infection, COVID-19, from the family of coronaviruses. This deadly infection is caused by the novel severe acute respiratory syndrome virus known as SARS-CoV-2.[1] Since its inception in January 2019, more than 8.5 million cases have been reported with greater than 450,000 fatalities (5.3%) (Source: Worldometer, June 18th, 2020). SARS-CoV-2, a coronavirus with crown-like spikes on the surface, can affect mammals and can then easily mutate to enable transfer from animals to humans[2]. The contagious virus spreads mainly from human to human and is rapidly becoming the world's leading cause of death.[2] COVID-19 was declared a pandemic by the World Health Organization (WHO) in March 2020, and strict safety guidelines were enacted for the entire world. Currently, there are no targeted vaccines or treatments available for SARS-CoV-2, and there is an urgent need to develop them. The aim of the present study is to use conventional computational approaches combined with accurate methods to explore protein-drug interactions that can be useful in the fight against COVID-19.

The structure of the novel SARS-CoV-2 virus is spherical and is reported to have mushroom-shaped proteins on its surface (spikes) that resemble the shape of a crown. The main protease of SARS-CoV-2 was identified several weeks after the outbreak and was named the M^{PRO} protease[3]. The crystal structure of the target protein (M^{PRO}) was solved, and its coordinates were immediately deposited into a protein data bank (PDB ID: 6LU7, RCSB)[3]. Currently, there is a significant effort aimed at developing a drug that can inhibit M^{PRO}. However, since the outbreak occurred, we still lack potent inhibitors/drugs to treat COVID-19, as the identification of drugs is a tedious process ranging from screenings to clinical trials and could take years[4]. When a new target protein to the disease is identified, the first step involves repurposing drugs that are FDA-approved and are readily available in the market. To repurpose the drugs, the basic tools involve computational characterization of the interaction sites of the main protease and drug interactions with the target protein. The concept of drug repurposing has proven to be successful in the past and is the most convenient method for screening new drugs for any novel diseases. Computations based on drug repurposing have identified the HIV antivirals Lopinavir and Ritonavir to be effective against COVID-19[5,6]. Wang and other groups have conducted drug repurposing computational analyses to identify inhibitors to the SARS-CoV-2 main protease[6–8].

With the current advances in computational techniques in combination with physical chemistry methods that utilize data learning algorithms, we are witnessing numerous impressive predictions in the field of drug discovery. These computations are used for screening and prediction of binding affinities and to generate fingerprint interactions with the protein[9]. The protein-drug binding simulations reveal interaction fingerprints that can be used efficiently to design other potent drugs[10]. Computer-aided drug design (CADD) makes use of extensive docking and molecular dynamics (MD) simulations to generate these interaction fingerprints. The accuracy of MD code relies on the accuracy of the potentials (forcefields), which depends on how they are parametrized[11]. Several classical forcefields, such as CHARMM, AMBER and OPLS, have been optimized to model the biomolecular system, bulk water accurately and can provide 99% of the biochemical information. However, with most of the chemistry (electronic, nuclear interactions) providing 1% of the information, classical forcefields are not adequately trained to adjust to these phenomena and are lacking in their knowledge base. Quantum chemical forces, density functional theory (DFT), and coupled cluster (CCSD(T)/CBS) methods are known to approximate the Schrödinger equation accurately (solving electronic and nuclear interactions), but at the same time, they are too complex to be solvable. To bridge the gap, machine learning and neural networks have been applied, and with recent augmentations in data, these approaches have become powerful approaches to accurately approximate any model[12,13]. Recently, the Roitberg group developed an ANI forcefield method that uses neural network-based models that are publicly available as ANI-1x and ANI-1ccx[14,15]. The ANI-1x model was trained on millions of small molecules with their DFT energies labeled, and the ANI-1ccx model was trained on highly accurate CCSD(T)/CBS data points. The accuracy of the ANI-1x model has been shown to be similar to DFT and found to be 106x faster than DFT calculations (a speed that matches classical force field methods). A similar case using ANI-1ccx that was trained on 500k CCSD(T)/CBS points was a billion times faster. While DFT is limited to 500 atoms and CCSD(T)/CBS to 10 atoms, ANI can utilize ~10,000 atoms with a speed comparable to classical forcefields[14–16]

This work for the first time reports interaction maps of repurposed drugs with the M^{PRO} protein using ANI potentials. This work also focuses on shortlisting the drugs by several levels of computations ranging from virtual screening (using docking) to the accurate methods. To bypass

the problem of parametrization associated with classical force fields with drugs, ANI-1ccx forcefields were used in combination with classical forcefields (CHARMM36) to simulate protein-drug binding. The main focus of this work is to analyze the binding profile of ANI/MM MD simulations and to investigate the key interactions that can guide future drug discovery. MM/PBSA calculations performed on trajectories produced by ANI/MM MD (termed ANI/MM/PBSA) validates the binding and filters out the decoys. Using trajectory clustering, the relationship between the number of clusters and the binding affinity of the drug with protein was explored. With an overall workflow that combines accurate techniques 12 drugs were identified that may be potent inhibitors of the M^{PRO} protein. The selected drugs were characterized to obtain a better understanding of the chemistry of target inhibition and to assist in the fight against COVID-19.



Graphical Abstract

Computational Methodologies:

Identification of interaction sites. To identify the protein-protein interaction (PPI) sites in M^{PRO} protein, I used MaSIF, which was developed by Gainza *et al* (2020)[17]. The MaSIF method makes use of geometric deep learning to explore patterns on the surface of proteins that are important for interactions with other proteins and a variety of drug molecules. Using this method, the protein surface can be defined as an array of geometrical and chemical features. A descriptor is then computed for each surface patch in the protein that contains information regarding all the features. Further, the descriptor is pushed to the layers of the deep learning model where they are classified based on the trained data. MaSIF was trained on data from more than 30,000 proteins. The docker

implementation of MaSIF was used in this work to generate the interaction sites of PDB ID (6LU7).

Molecular Docking. First, the crystal structure of M^{PRO} (PDB ID:6LU7), which was deposited by Jin *et al*, [3] was obtained from the RCSB PDB databank. To prepare the M^{PRO} protein for docking, I used AutoDock Tools (ADT) to assign Kollman charges and atom/bond types[18]. For drug repurposing I chose the database of drugs that are FDA approved and readily available in the market. I obtained FDA-approved docking ready drugs from ZINC15 and extensively used open babel codes to perform the conversion between SDF and PDBQT file formats[2]. Screening against M^{PRO} was performed using Autodock Vina code, which defined a grid box to sufficiently cover the active site of proteins [19]. The box size was defined as 28 × 28 × 28 $Å^3$ with the center at (x = 6.194, y = 1.944 and z = 24.778) using ADT. Using docking methods, I obtained the positions of inhibitors in the binding site of the protein, which was used as an initial pose for validation of the dynamics.

Classical Molecular Dynamics Simulations. All molecular dynamics (MD) simulations were conducted using NAMD[20]. I used CHARMM-GUI webserver to generate the CHARMM36m parameters and topology files for the protein and the SwissParam server to generate topology and parameters for the drugs (small molecules)[21,22]. All complexes were enclosed in triclinic boxes. The TIP3P water model was used to solvate all complexes. Ions (Na^+ and Cl^-) were added at a concentration of 0.15 M to neutralize the systems. After neutralization, the systems were submitted to 5000 steps of steepest descent minimization. Minimized systems were further equilibrated under both NVT and NPT conditions for 1 and 2 ns, respectively. During equilibration, position restraints were applied to both protein and drug molecules. Temperature (303 K) and pressure (1 atm) were controlled by the Langevin piston method. The particle mesh Ewald (PME) method was used to calculate long-range interactions[23]. The equilibrated systems were finally submitted to a 10 ns NPT production run without position restraints. To compute the time-dependent RMSD for the drug molecule with respect to the initial pose (drug and protein's secondary structure aligned), 800 snapshots from 2 to 10ns were averaged.

ANI Potential and Dynamics. In this work, I combined accurate ANI forcefields for drugs with CHARMM36m forcefields for proteins and ions along with TIP3 for water to run the drug binding dynamics with M^{PRO} (Figure 1). I performed ANI/MM calculations per the hybrid model previously combined by the Rowley group[24]. In hybrid ANI/MM calculations, the total potential energy (v) of the system was defined as the sum of the v of the ANI region (i.e., the intramolecular interactions of the drug molecule), the v of the surroundings (protein, water, and ions), and the v of the interactions between the drug and its surrounding (Equation (1))[24].

$$v(r) = v_{ANI}(r_{ANI}) + v_{MM}(r_{MM}) + v_{ANI/MM}(r_{ANI}) \qquad (1)$$

where r represents coordinates of all the atoms, r_{MM} is the coordinates of the atoms surrounding the drug molecule, and r is the coordinates of the drug molecule. The MM region is described by the CHARMM36m force field. For noncovalent protein–drug binding, the $v_{ANI/MM}$ term is the conventional MM nonbonded interactions between the protein and the drug, which is simply the sum of Lennard-Jones and pairwise coulombic interactions between the ANI atoms and MM atoms (Equation (2))[24].

$$v_{ANI/MM}(r_{ANI}) = \sum_{i}^{MM}\sum_{j}^{ANI} \frac{q_i q_j}{4\pi\varepsilon r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (2)$$

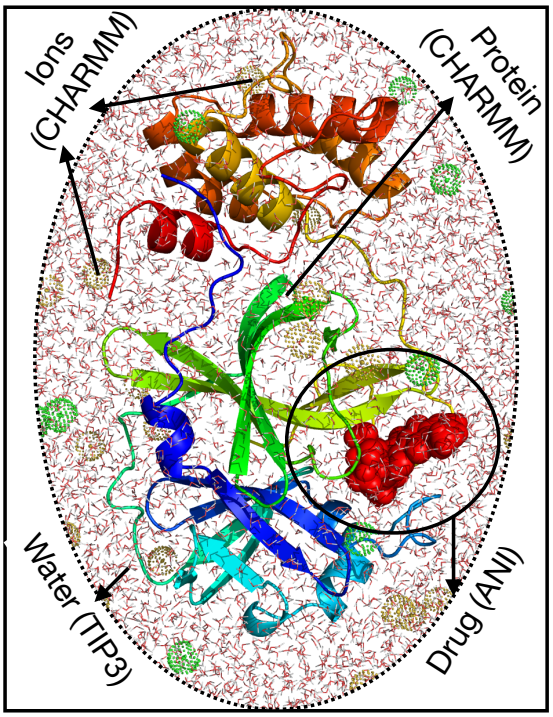


Figure 1. Initial system for MD simulation. The initial system consists of a docking pose conformation for a drug (in red surface) combined with rigid protein conformation (cartoon representation) used for docking. The drug-protein complex is solvated with TIP3 water molecules in a box shown with line representation. Neutralizing ions Na^+ and Cl^- (0.15 M) are represented by green and brown mesh bubbles.

ANI/MM/PBSA free energy calculations. To compute the ANI/MM/PBSA free energy of protein-drug binding, I used CaFe (open source code for calculation of free energy), which was developed by Liu *et al*. (2016)[25]. CaFe is an automated tool for predicting binding free energies using endpoint methods such as the molecular mechanics Poisson-Boltzmann surface area (MM/PBSA) method. This tool is specially designed to incorporate CHARMM/NAMD file formats (such as PSF, DCD, etc.)[25]. In this work, I used trajectory files obtained from hybrid ANI/MM calculations from 0.2 to 1 ns (80 snapshots in total). As the trajectory files involve drug molecules defined using an ANI forcefield, the free energy was computed using $\Delta E_{ANI/MM}$ (similar to the $v_{ANI/MM}$ term in equation 2) instead of ΔE_{MM}. Here, in this ANI/MM/PBSA computation, the ANI/MM term was computed for a gas-phase. For PB calculations, I used APBS programs (interfaced to CaFe) where boundary conditions were set to multiple Debye Huckel values and charges were mapped to grids using a cubic B spline method. To perform the SA calculations, the surface tension value was defined as 0.00542 with a surface offset of 0.92. Finally, the binding energy was summed and averaged over an ensemble of conformations (80 snapshots) as follows.

$$\Delta G_{bind} = \Delta H - T\Delta S \approx \langle \Delta E_{ANI/MM} + \Delta G_{sol}^{polar} + \Delta G_{sol}^{nonpolar} - T\,\Delta S \rangle \qquad (3)$$

Pose Clustering & 2D-interaction map. To cluster the trajectories (drugs poses) obtained by hybrid ANI/MM trajectories, I used TTclust, a trajectory clustering program developed by Tubiana *et al* (2018)[26]. The academically free Maestro software from Schrödinger (GUI) was used to generate the 2D-interaction map of the drug combined with the protein. [27]. To map 3D-interactions of the drugs with the residues in the binding site, I used PLIP code (fully automated protein-ligand interaction profiler)[28]. By combining this code with PLIP, I report the number of hydrophobic interactions and hydrogen bonds between drugs and protein and extract common interacting residues.

Results

I extracted the target protein, a crystal structure of M^{PRO}, the main protease of COVID-19, that was solved and deposited by Jin *et al*. (6LU7 chain A) from the RCSB PDB databank for all the analyses performed in this work. I used the MaSIF method to predict PPI sites on the protein surface that are defined as surface patches on a protein that are more likely to interact with other proteins and with potential drug molecules. The MaSIF method predicted five interaction sites on the M^{PRO} surface with the highest score; these sites were labeled A, B, C, D, and E (Figure 2). Site 'A' consists of residues 140-145, 163-168, and 187-192; site 'B' consists of residues 112-116, 124-128, and 137-140; site 'C' covers residues in regions 192-199 and 235-239; site 'D' covers residues 292-301 and site 'E' consists of residues 65-68 and 19-22. Furthermore, I also found that site 'A' matches with the surface patch where an N3 inhibitor was identified in a bound form with the crystal structure of the protein. For drug screening, only site 'A' was used, while other sites reported here can also be potentially used to further screen a database of protein or drugs against M^{PRO}.

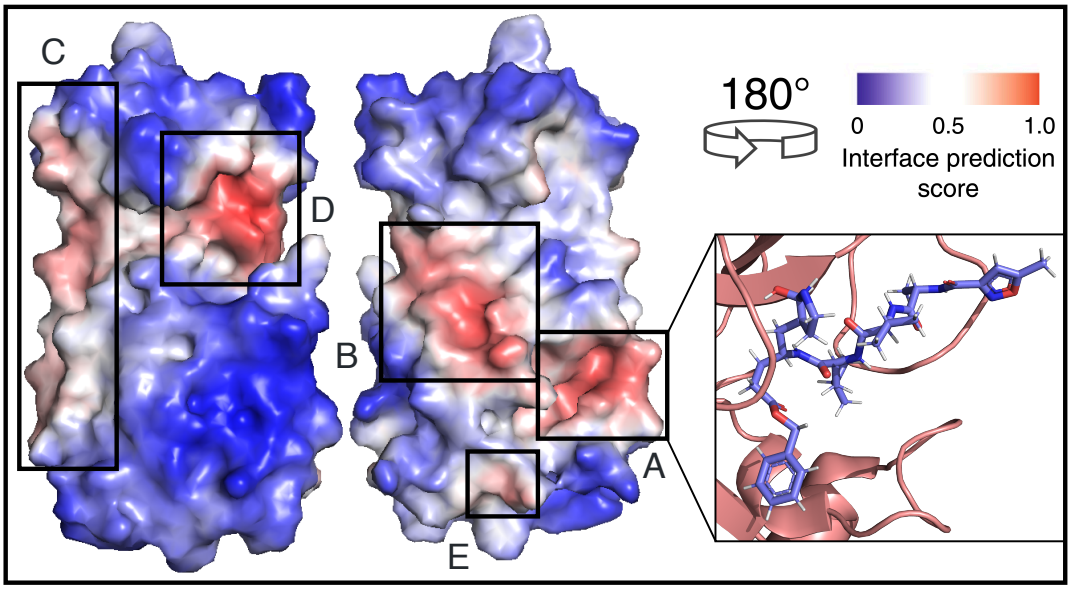


Figure 2. Interaction sites on the COVID-19 target protein. The surface representation of the main protease of COVID-19 (PDBID: 6LU7) generated using geometric deep learning and visualized using Pymol. The red region represents the best interaction sites and are labeled A, B, C, D, and E. The left and right representation is for the same

structure but with a 180° rotation. The rightmost panel shows the N3 inhibitor that was crystallized with M^{PRO}. The N3 inhibitor binds to site A of the protein.

The next step involved virtual screening of FDA-approved drugs or repurposing drugs against M^{PRO}. To perform this task, I first obtained the FDA-approved drug (dock ready) database from ZINC15. To select the drugs that can fit these criteria, I filtered out compounds that contained atoms other than C, H, N, and O and all drugs that contained charged atoms (to permit use of the ANI forcefield), thereby reducing the number of drugs in the database to 1188. Next, I performed molecular docking-based screening of the 1188 drugs using AutoDock Vina and ranked each drug according to their decreasing binding affinity. From this ranking list of drugs and their docking scores, I selected 91 compounds with binding affinity scores equal to or less than -8.0 kcal/mol. From this group of 91 compounds, I found that 10% had a score below -9.0 kcal/mol, 20% were between -8.9 kcal/mol and -8.5 kcal/mol, 30% ranged between -8.4 and -8.1 kcal/mol (both included), and the remaining 40% had a score of -8.0 kcal/mol (Figure 3). Because docking programs have a tendency to generate false poses, I performed classical MD simulations to validate these scores.

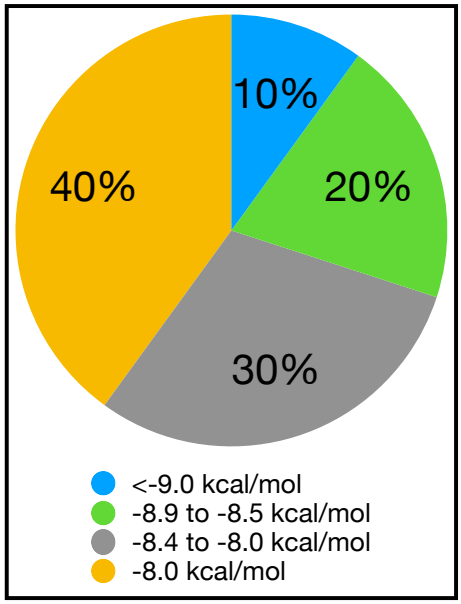


Figure 3. Distribution of 91 drugs selected based on docking-based screening. Here, 10% of the most effective drugs had scores below -9.0 kcal/mol, 20% had scores

between -8.9 and -8.5 kcal/mol, 30% had scores between -8.4 and -8.0 kcal/mol, and the remainder of the drugs had scores of exactly -8.0 kcal/mol.

MD simulations with classical forcefields (e.g., CHARMM36m) can improve and validate the pose obtained by a docking program. MD samples a conformation space of protein-drug binding conformations starting with the initial conformation extracted from docking experiments. I carried out 10 ns MD simulations for 82 of the 91 drugs (using parameters that are available in SwissParam) in complex with protein in a TIP3-solvated box and with Na^+ Cl^- as neutralizing ions. From the trajectories obtained, I classified the drugs that bound. The drugs that left the active site of the protein were classified, and the average RMSD values were computed for each drug with the secondary structure of the protein and drug aligned to the initial conformation. With the drugs arranged according to their decreasing RMSD values, drugs with the highest RMSD (of ~10 Å) are those that do not bind to the protein, while drugs with the lowest RMSD (~1.5 Å) exhibit favorable binding. In this manner, and with a cutoff of 4 Å, I classified 37 drugs as binding, and the remainder as leaving drugs, which were disqualified from proceeding to the next level (Figure 4). The RMSDs of selected drugs along with their ZINC-ID are shown in Supplementary Table 1.

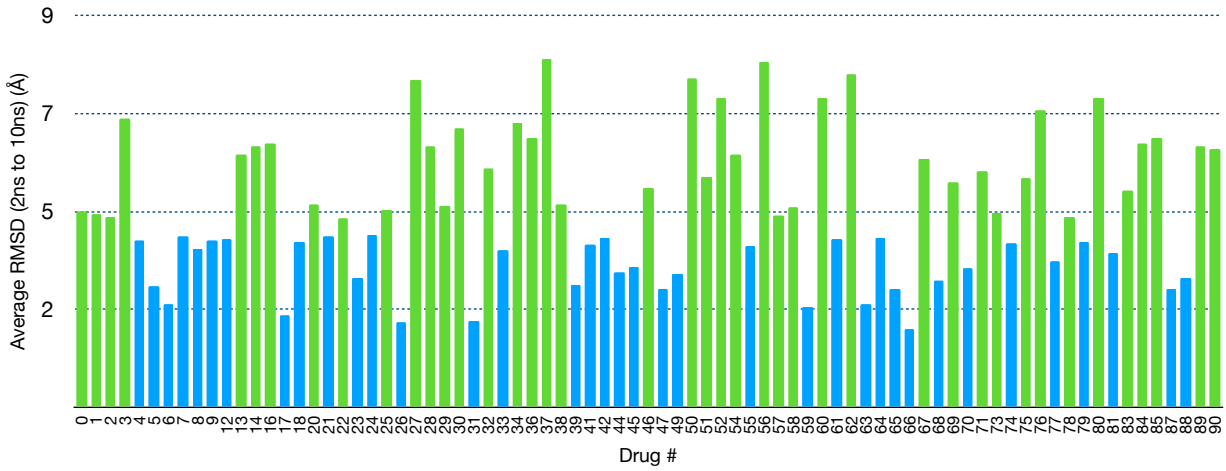


Figure 4. Average RMSD values for 91 drugs. The averaged RMSD values for each drug with respect to the initial conformation. The blue bars represent binding drugs with an average drug RMSD below 4 Å and green bars represent leaving drugs with an average ligand RMSD above 4 Å. Indexing is based on the docking score from highest to lowest.

To accurately describe the inhibition profile for the selected drugs, ANI/MM dynamic runs were conducted using NAMD. To obtain the initial conformation for the ANI/MM setup, the final snapshot was extracted from the 10 ns MD production run which was performed using a classical forcefield. To setup the ANI/MM system, I defined the ANI region to contain the atoms of drug molecules and combined the ANI region with the region defined by a classical forcefield (CHARMM36m) for protein, water, and ions. Short 1 ns snapshots of ANI/MM MD at constant pressure and temperatures were then conducted. The RMSD computations for 1 ns of trajectory (100 snapshots) produced by ANI/MM dynamics were conducted again to classify binding drugs from leaving drugs. Using a cutoff of 4 Å, 18 drugs were eliminated and 19 drugs were selected (Figure 5). To further investigate the set of 19 selected drugs, I examined the binding affinity of each with protein by computing binding free energies using the MM/PBSA method. Using ANI/MM trajectories as an input, these calculations were labeled ANI/MM/PBSA, replacing the MM term with ANI/MM. Based on the energies obtained using ANI/MM/PBSA, 12 of the 19 drugs were determined to bind spontaneously, having free energy values that were negative (Figure 5). In addition, two drugs with the highest binding affinities, Targretin and Eltrombopag, have reported ANI/MM/PBSA free energies of -11.74 and -11.16 kcal/mol, respectively. The MM/PBSA data for the binding of 12 selected drugs with protein (energy differences of terms Elec, vdW, Pol, Sol, PB, SA, Gas and Npol) are reported in Supplementary Tables S1 and S2.

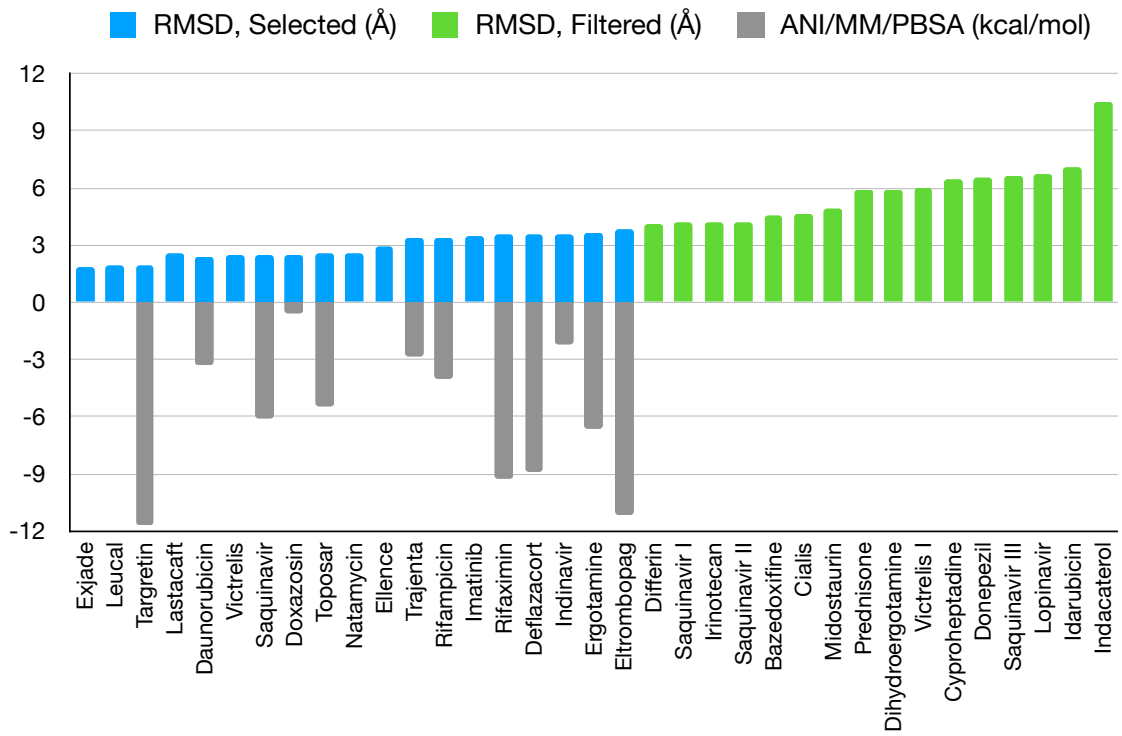


Figure 5. RMSD and ANI/MM/PBSA values from ANI/MM MD. Averaged RMSD values for the trajectories generated from ANI/MM simulations. Blue bars represent binding drugs (below 4 Å) and green bars represent weaker binding drugs (above 4 Å). Gray bars represent ANI/MM/PBSA free energy for binding drugs; only drugs with a negative free energy of binding are shown.

In the final selection, 12 drugs were chosen to characterize the drug-protein interactions from the trajectories produced by the ANI/MM calculations. Using TTclust, I clustered snapshots from ANI/MM trajectories based on their similarities The trajectories of 8 drugs, Lifaximin, Doxazosin, Deflazacort, Lastacaft, Rifampicin, Victrelis, Topovar and Indinavir binding to protein were clustered in as few as 3 clusters, Targretin, Eltrombopag, and Trajenta binding trajectories were clustered in 4 clusters, and Ergotamine binding trajectories were clustered in 5 clusters (Figure 6). Trajectory profiles are important tools for analyzing drug-binding protein interactions, and they provide more detail than can be obtained by computing RMSDs. Based on this description, 11 of the 12 drugs should be subjected to further simulations. The 2D interaction maps from the last trajectory of ANI/MM calculations were analyzed using PLIP and combined with the 2D drug/ligand interaction images generated from Maestro. Based on these results, GLU166, GLN189

and THR190 were determined to be key residues in binding site 'A' that interacted with most of the selected drugs. The other key residues involved in binding of drugs to the binding site of the protein can be easily identified, i.e., the residues of the protein that are common to the 12 selected drugs (Figure 7).

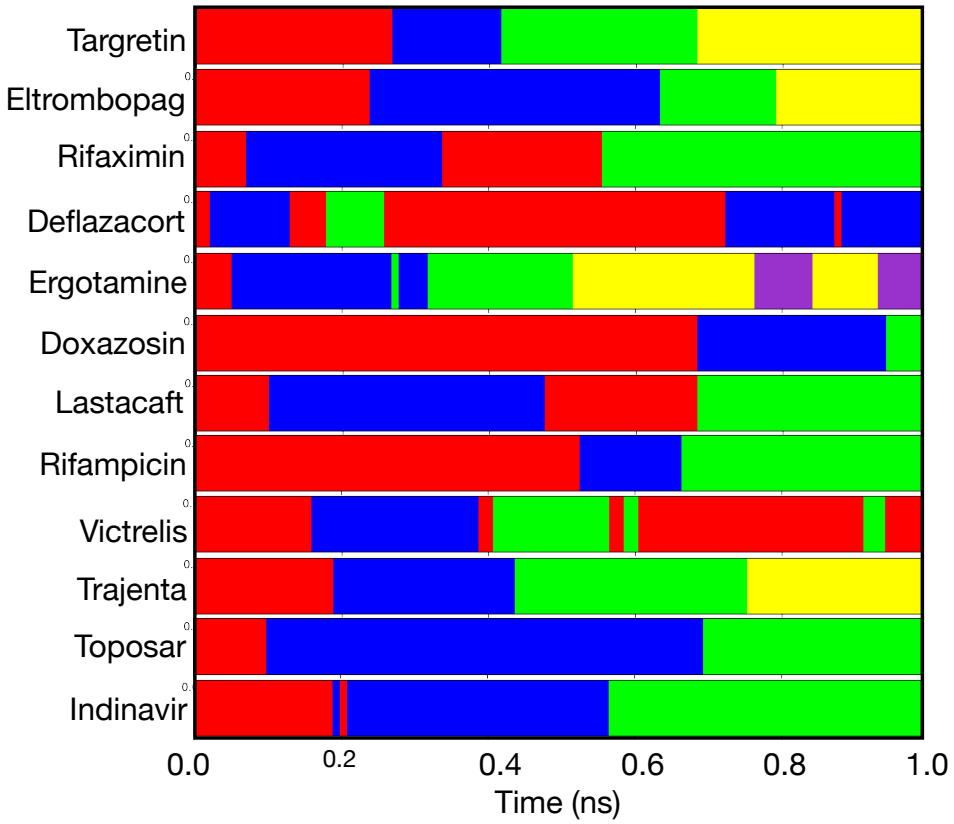


Figure 6. Clustering of ANI/MM MD trajectories: Clustering of trajectories generated using ANI/MM dynamics. Twelve drugs that were selected based on the ANI/MM/PBSA are shown. Different colors represent different clusters. A total of 100 ANI/MM snapshots were clustered for each drug-protein binding simulation.

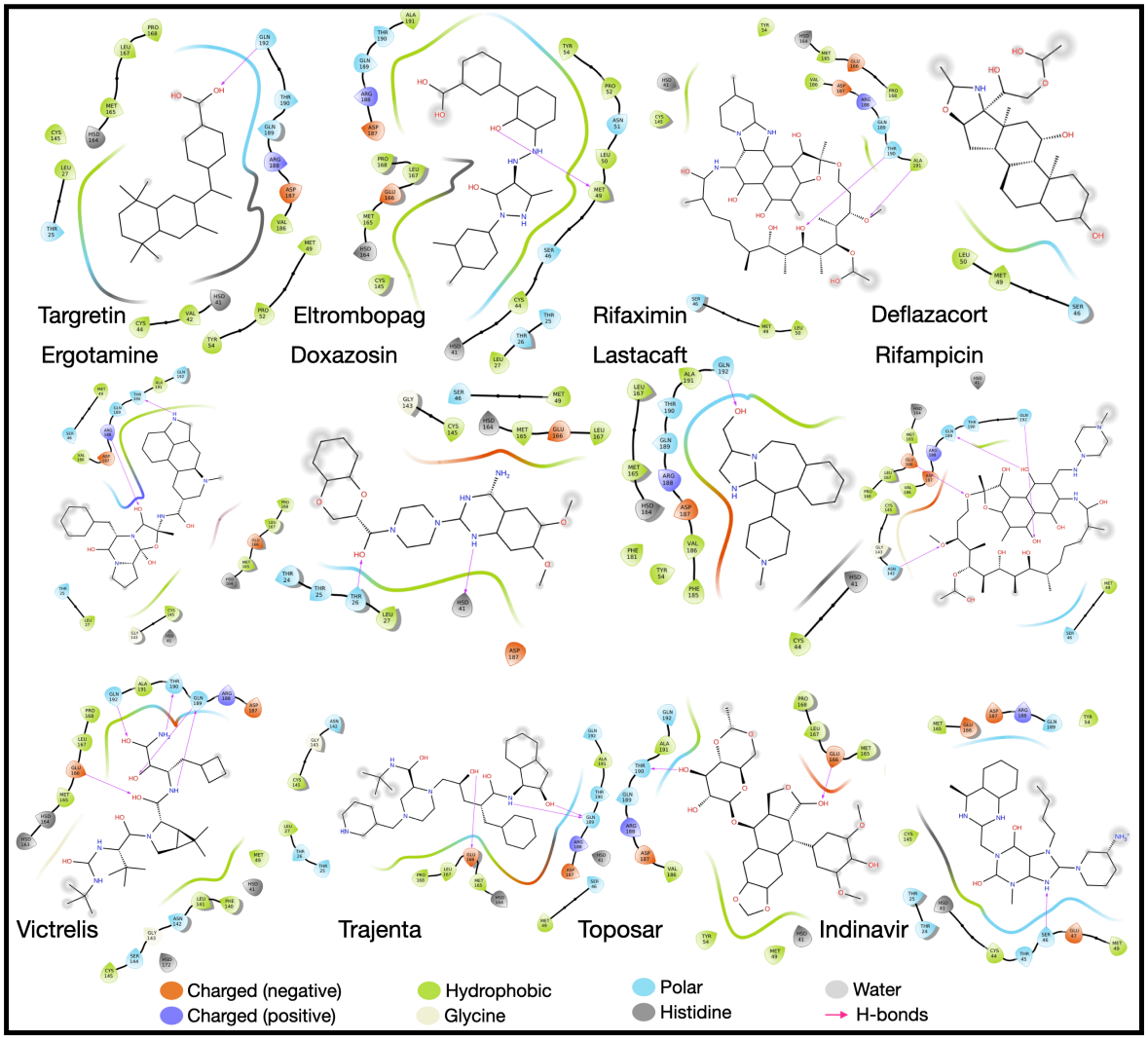


Figure 7. 2D interaction map of selected drugs with M^{PRO}. 2D - Structures for each selected drug is shown. Colored drops represent different properties of interacting residues of the protein (representation is shown with the figure).

Discussion and Conclusion

With the rapidly growing concern over coronavirus (COVID-19), which is leading to an exponential increase in the number of people becoming affected and fatalities, I present an investigation on the development of potential inhibitors against the virus using a drug repurposing approach. The scientific community is devoting a tremendous amount of effort to characterizing potential drugs to inhibit this virus, yet much more information and effort is required before a unique treatment can be approved[29]. Careful studies on the interaction of viral proteins with

available drugs that are FDA-approved are crucial for understanding the binding behavior of these proteins and can aid in accelerating the development of biochemical assays[30]. Molecular dynamics with classical forcefields can describe many important drug-protein binding interactions, but they tend to miss smaller details at the electronic and nuclear levels[31]. These missed details can be recovered when MD classical forcefields can be combined with methods such as DFT and CCSD(T)/CBS that are most effectively used to provide the most accurate descriptions of electronic and nuclear phenomenon for small drug molecules. In this work, neural network force fields were explored to study binding interactions of repurposed drug molecules to the M^{PRO} protein. The findings from this study, the interaction sites and the structure of selected drugs may be useful for designing novel drugs that can be used against COVID-19.

To initiate the conventional process of drug discovery, identification of drug binding sites is extremely important and several methods for identifying binding sites on the protein have already been proposed and can perform an accurate search of the protein surface. Neural network methods such as MaSIF take advantage of accurate methods for training and mimic performance compared to physical chemical methods that are based on prediction accuracy[17]. The MaSIF method also provides other approaches, one of them can be used to screen other proteins that can interact with the protein of interest based on the descriptors that are generated for site identification. This work is limited to identifying sites for PPI using a single site (site 'A') for drug screening and binding dynamics. PPI sites are reported to be important as interaction sites for proteins as well as drugs[32]. Other sites reported in this work ('B', 'C', 'D' and 'E') can also be used to screen other types of interactions.

To perform the conventional drug screening work, drugs were selected from an FDA-approved drugs database based on the atoms they contain and charges they possess. Using molecular docking techniques, the number of drugs was reduced to 91 by selecting specific drugs, i.e., those with the best docking scores. To sample the conformational space of drug-protein complexes, molecular dynamics simulations were conducted using classical molecular dynamics, which were then combined with an ANI/MM approach to provide an accurate description of the interaction profile of drugs with the protein. The steps involving virtual screening and classical MD are traditionally used in any drug discovery process, and this process has been performed by several groups[6–8]. In

this work, classical MD methods were combined with ANI/MM dynamics to investigate other sets of interactions that may be ignored by classical MD simulations but could prove to be useful for experimental drug design. Additionally, combining ANI/MM trajectories with endpoint MM/PBSA computations assists in obtaining physically important information related to binding. The ANI/MM/PBSA calculations reported in this work require further development in order to handle the ANI/MM interactions and will be the focus of future research. As ANI forcefields are still in development, and the analysis tools are not as developed compared to classical MD methods, other basic tools were used to analyze the ANI/MM simulations. Cluster analysis of ANI/MM trajectories was performed to further validate the outcome and 100 snapshots were observed to cluster in as few as 3 or 4 clusters (Figure 6). This result may also imply that drug binding to protein does not induce drastic conformational changes when the number of clusters are limited to 3 or 4 (for 100 snapshots). I found that one of the drugs had binding trajectories that were clustered in 5 clusters, implying rapid conformational changes, and thus could be eliminated, retaining 11 of the 12 drugs. This workflow with the trajectory validation can be useful for analyzing other simulations that involve drug interactions with M^{PRO} or other proteins in the future.

Based on these findings, the structures of drugs such as Targretin and Eltrombopag and antiviral drugs such as Indinavir and Victrelis are potent drugs that have the most inhibitory action against the M^{PRO} protein. Ten other drugs were also identified that might be repurposed against M^{PRO}. By characterizing the output from each level, 91 FDA-approved drugs were selected from virtual screenings, 37 drugs from MD simulations profile, and 19 drugs from ANI/MM Trajectories. An attempt was made to provide the maximum amount of information related to each of the drugs and to cover all of the interaction aspects of these drugs with the M^{PRO} protein in this study. The structure of the shortlisted drugs and their detailed interaction fingerprints with the protein are discussed in great detail, which may be useful for discovering other small molecules. This study performs only computational studies, and the drugs reported in this work are only characterized by computations. No experiments have been performed related to these drugs binding with the protein.

References.

(1) Xu, Z.; Shi, L.; Wang, Y.; Zhang, J.; Huang, L.; Zhang, C.; Liu, S.; Zhao, P.; Liu, H.;

Zhu, L.; Tai, Y.; Bai, C.; Gao, T.; Song, J.; Xia, P.; Dong, J.; Zhao, J.; Wang, F. S. Pathological Findings of COVID-19 Associated with Acute Respiratory Distress Syndrome. *Lancet Respir. Med.* **2020**, *8* (4), 420–422. https://doi.org/10.1016/S2213-2600(20)30076-X.

(2) Huynh, T.; Wang, H.; Luan, B. In Silico Exploration of Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease. *J. Phys. Chem. Lett.* **2020**, 4413–4420. https://doi.org/10.1021/acs.jpclett.0c00994.

(3) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X. X.; You, T.; Liu, X. X.; Yang, X. X.; Bai, F.; Liu, H.; Liu, X. X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of Mpro from COVID-19 Virus and Discovery of Its Inhibitors. *Nature* **2020**, 1–5. https://doi.org/10.1038/s41586-020-2223-y.

(4) Hughes, J. P.; Rees, S. S.; Kalindjian, S. B.; Philpott, K. L. Principles of Early Drug Discovery. *British Journal of Pharmacology*. John Wiley & Sons, Ltd March 1, 2011, pp 1239–1249. https://doi.org/10.1111/j.1476-5381.2010.01127.x.

(5) Nutho, B.; Mahalapbutr, P.; Hengphasatporn, K.; Pattaranggoon, N. C.; Simanon, N.; Shigeta, Y.; Hannongbua, S.; Rungrotmongkol, T. Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochemistry* **2020**, *59* (18), 1769–1779. https://doi.org/10.1021/acs.biochem.0c00160.

(6) Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease -19 (COVID-19) Through Computational Drug Repurposing Study. *J. Chem. Inf. Model.* **2020**. https://doi.org/10.1021/acs.jcim.0c00179.

(7) Elmezayen, A. D.; Al-Obaidi, A.; Şahin, A. T.; Yelekçi, K. Drug Repurposing for Coronavirus (COVID-19): In Silico Screening of Known Drugs against Coronavirus 3CL Hydrolase and Protease Enzymes. *J. Biomol. Struct. Dyn.* **2020**, 1–12. https://doi.org/10.1080/07391102.2020.1758791.

(8) Muralidharan, N.; Sakthivel, R.; Velmurugan, D.; Gromiha, M. M. Computational Studies of Drug Repurposing and Synergism of Lopinavir, Oseltamivir and Ritonavir Binding with SARS-CoV-2 Protease against COVID-19. *Journal of Biomolecular Structure and Dynamics*. Taylor and Francis Ltd. 2020.

https://doi.org/10.1080/07391102.2020.1752802.

(9) Vamathevan, J.; Clark, D.; Czodrowski, P.; Dunham, I.; Ferran, E.; Lee, G.; Li, B.; Madabhushi, A.; Shah, P.; Spitzer, M.; Zhao, S. Applications of Machine Learning in Drug Discovery and Development. *Nature Reviews Drug Discovery*. Nature Publishing Group June 1, 2019, pp 463–477. https://doi.org/10.1038/s41573-019-0024-5.

(10) Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacological Reviews*. American Society for Pharmacology and Experimental Therapeutics January 1, 2014, pp 334–395. https://doi.org/10.1124/pr.112.007336.

(11) Wang, L. P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1885–1891. https://doi.org/10.1021/jz500737m.

(12) Balabin, R. M.; Lomakina, E. I. Neural Network Approach to Quantum-Chemistry Data: Accurate Prediction of Density Functional Theory Energies. *J. Chem. Phys.* **2009**, *131* (7), 074104. https://doi.org/10.1063/1.3206326.

(13) Morawietz, T.; Behler, J. A Density-Functional Theory-Based Neural Network Potential for Water Clusters Including van Der Waals Corrections. *J. Phys. Chem. A* **2013**, *117* (32), 7356–7366. https://doi.org/10.1021/jp401225b.

(14) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.

(15) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching Coupled Cluster Accuracy with a General-Purpose Neural Network Potential through Transfer Learning. *Nat. Commun.* **2019**, *10* (1), 1–8. https://doi.org/10.1038/s41467-019-10827-4.

(16) Smith, J. S.; Roitberg, A. E.; Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Medicinal Chemistry Letters*. American Chemical Society November 8, 2018, pp 1065–1069. https://doi.org/10.1021/acsmedchemlett.8b00437.

(17) Gainza, P.; Sverrisson, F.; Monti, F.; Rodolà, E.; Boscaini, D.; Bronstein, M. M.; Correia, B. E. Deciphering Interaction Fingerprints from Protein Molecular Surfaces Using

Geometric Deep Learning. *Nat. Methods* **2020**, *17* (2), 184–192. https://doi.org/10.1038/s41592-019-0666-6.

(18) Forli, S.; Huey, R.; Pique, M. E.; Sanner, M. F.; Goodsell, D. S.; Olson, A. J. Computational Protein-Ligand Docking and Virtual Drug Screening with the AutoDock Suite. *Nat. Protoc.* **2016**, *11* (5), 905–919. https://doi.org/10.1038/nprot.2016.051.

(19) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. John Wiley and Sons Inc. 2005, pp 1781–1802. https://doi.org/10.1002/jcc.20289.

(21) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. SwissParam: A Fast Force Field Generation Tool for Small Organic Molecules. *J. Comput. Chem.* **2011**, *32* (11), 2359–2368. https://doi.org/10.1002/jcc.21816.

(22) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(23) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.

(24) Lahey, S. L. J.; Rowley, C. N. Simulating Protein-Ligand Binding with Neural Network Potentials. *Chem. Sci.* **2020**, *11* (9), 2362–2368. https://doi.org/10.1039/c9sc06017k.

(25) Liu, H.; Hou, T. CaFE: A Tool for Binding Affinity Prediction Using End-Point Free Energy Methods. *Bioinformatics* **2016**, *32* (14), 2216–2218. https://doi.org/10.1093/bioinformatics/btw215.

(26) Tubiana, T.; Carvaillo, J. C.; Boulard, Y.; Bressanelli, S. TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries. *J. Chem. Inf. Model.* **2018**, *58* (11), 2178–2182. https://doi.org/10.1021/acs.jcim.8b00512.

(27) Schrödinger, S. R.-; LLC, U.; York, N.; NY, U.; 2017, U. 1: Maestro.

(28) Salentin, S.; Schreiber, S.; Haupt, V. J.; Adasme, M. F.; Schroeder, M. PLIP: Fully

Automated Protein-Ligand Interaction Profiler. *Nucleic Acids Res.* **2015**, *43*, 443–447. https://doi.org/10.1093/nar/gkv315.

(29) Kupferschmidt, K.; Cohen, J. Race to Find COVID-19 Treatments Accelerates. *Science*. 2020, pp 1412–1413. https://doi.org/10.1126/science.367.6485.1412.

(30) Rismanbaf, A. Potential Treatments for COVID-19; a Narrative Literature Review. *Arch. Acad. Emerg. Med.* **2020**, *8* (1), e29. https://doi.org/10.22037/aaem.v8i1.596.

(31) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *J. Chem. Phys.* **1999**, *110* (21), 10452–10467. https://doi.org/10.1063/1.479049.

(32) Fuller, J. C.; Burgoyne, N. J.; Jackson, R. M. Predicting Druggable Binding Sites at the Protein-Protein Interface. *Drug Discovery Today*. Elsevier Current Trends February 1, 2009, pp 155–161. https://doi.org/10.1016/j.drudis.2008.10.009.

Author Information

*Corresponding Author

Phone: (773) 658-2050. Email: aayush.ict@gmail.com

Acknowledgments

The author acknowledges University of Illinois at Chicago (UIC) – ACER and EXTREME HPC (https://acer.uic.edu/) to provide computational resources and power for running extensive calculations.

Competing financial interests

The author declares no competing financial interest.