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 aid to minimize the efforts by identifying a subset of drugs that can potentially bind to COVID-19 main protease or target protein (M^{PRO}). The results of computations are always accompanied by 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 bet to employ in this process. With this work, I first identified and described the interaction sites of M^{PRO} protein using a geometric deep learning model. Secondly, I conducted virtual screening (at one of the sites identified) on FDA approved drugs and picked 91 drugs having the highest binding affinity (below -8.0 kcal/mol). Then, I carried out 10 ns of molecular dynamics (MD) simulations using classical force fields and classified 37 drugs to be binding (includes drugs like Lopinavir, Saquinavir, Indinavir etc.) based on RMSD between MD-binding trajectories. To drastically improve the dynamics profile of selected 37 drugs, I brought in the highly accurate neural network force field (ANI) trained on coupled-cluster methods (CCSD(T)/CBS) data points and performed 1 ns of binding dynamics of each drug with protein. With the accurate approach, 19 drugs were qualified based on their RMSD cutoffs, and again with their free energy (ANI/MM/PBSA) computations another 7 drugs were rejected. The final selection of 12 drugs was validated based on MD trajectory clustering approach where 11 of 12 drugs (Targretin, Eltrombopag, Rifaximin, Deflazacort, Ergotamine, Doxazosin, Lastacraft, Rifampicin, Victrelis, Trajenta, Toposar, and Indinavir) were confirmed to be binding. Further investigations were made to study their interactions with the protein and an accurate 2D-interaction map was generated. These findings and mapping of drug-protein interactions are highly accurate and could be potentially used to guide rational drug discovery against the COVID-19.

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

Introduction

The whole world is going through 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 known as SARS-CoV-2¹. Since its inception in Jan'19, more than ~ 5.9 millions of cases have been reported with more than ~360,000 fatalities (8%) (Source: Worldometer, May'2020). COVID-19, a coronavirus with crown-like spikes on the surface can affect mammals and then easily mutate to transfer from animals to humans². The virus being contagiousness is mainly spreading from human to human and is now rapidly becoming the world's leading cause of death². It was declared as a pandemic by world health organization (WHO) in March'20 and strict guidelines to stay safe were passed to the whole world. Currently, there are no targeted vaccines or treatments available for SARS-CoV-2 and there is an urgent need to develop it. The present study aimed to use conventional computational approaches combined with accurate methods to explore the protein-drugs interactions which can be useful to fight against COVID-19.

The structure of the novel SARS-CoV-2 is spherical and reported to have mushroom-shaped protein (spikes) which gives it a shape of a crown. Scientific groups were able to identify the main protease of SARS-CoV-2 few weeks of the outbreak and named it as M^{PRO} protease³. The crystal structure of the target protein (M^{PRO}) was solved, and its coordinates were immediately deposited into the protein data bank (PDB ID: 6LU7, RCSB)³. Now, the whole scientific community is working towards developing a drug that can inhibit the M^{PRO}. Few month have been passed since outbreak and we still lack potent inhibitor/drug to COVID-19 as identification of drugs is a tedious process ranging from several screenings to clinical trials and thus could take years to produce one⁴. When a new target protein to the disease is identified, the first step involves is to repurpose the drugs which are FDA approves and are readily available in the market. To repurpose the drugs, the basic tools involve computational characterization of the interaction sites of main protease and drugs interaction with the target protein. The concept of drug repurposing has been proved to be successful in the past and is the most convenient way to screen a new drug for any novel diseases. Junmei Wang and other groups have already carried out drug repurposing to identify an inhibitor to SARS-CoV-2 main protease⁵⁻⁹. In past, computations based on drug repurposing have identified HIV antivirals Lopinavir and Ritonavir to be effective against COVID-19^{5,10}.

With the current advances in computations techniques and a combination of physical chemistry methods with data learning algorithms, we are witnessing tons of impressive predictions in the field of drug discovery. These computations are used to screen, binding affinity predictions, and generate fingerprint interactions with the protein¹¹. The protein-drug binding simulations reveal clues of interaction fingerprints which can be efficiently used to design other potent drugs¹². Computational aided drug design (CADD) makes use of extensive docking and molecular dynamics (MD) simulations to generate these interactions fingerprints. Accuracy of MD code relies on the accuracy of potentials (forcefields) which depends on the way they are parametrized¹³. Several classical forcefields like CHARMM, AMBER, and OPLS have been optimized to accurately model the biomolecular system, bulk water and can provide 99% of the information. But with most of the chemistry (electronic, nuclei interactions) happening at 1%, classical forcefields are not trained enough to adjust to these phenomena, thus lacking their knowledge base. Quantum chemical forces, Density functional theory (DFT), and Coupled cluster (CCSD(T)/CBS) methods are known to accurately approximate Schrödinger equation (solving electronic and nuclei interactions), but at the same time, they are too complex to be soluble. To bridge the gap, machine learning and neural networks come to rescue and with the recent augmentations in data, these approaches have become powerful to accurately approximate any model^{14,15}. Recently developed the ANI forcefield by the Roitberg group are neural network based models and they are publicly available as ANI-1x and ANI-1ccx^{16,17}. The model ANI-1x is trained on millions of small molecules with their DFT energies labeled which ANI-1ccx model is trained on highly accurate CCSD(T)/CBS data points. The accuracy of ANI-1x model has been shown similar to DFT and found to be 106x faster than DFT calculations (speed matches classical force fields). Similar case with ANI-1ccx, trained on 500k CCSD(T)/CBS points and billion times faster. While DFT is just limited to 500 atoms and CCSD(T)/CBS to 10 atoms, ANI can go beyond ~10,000 of atoms with a speed compared to classical forcefields¹⁶⁻¹⁸

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 of drugs, I used ANI potentials

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 which 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, I explored the relation of the number of clusters with the binding affinity of the drug with protein. With the overall workflow which combines accurate techniques, I identified 12 drugs which can be a potent inhibitor to M^{PRO} protein. Using the selected drugs, I work on characterizing them to fetch a better understanding of the chemistry of inhibiting the target and fight against COVID-19.

Graphical Abstract

Computational Methodologies:

Interaction sites identification. To identify the protein-protein interaction (PPI) sites in M^{PRO} protein, I used MaSIF developed by Gainza et. al (2020)¹⁹. MaSIF method demonstrate the use of geometric deep learning to explore the patterns in the surface of protein which are important for interactions with other protein and a variety of drug molecules. Using this method, the protein surface can be defined as an array of geometrical and chemical features. Then a descriptor is computed for each surface patch in protein which contains information of 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 is trained on data of more than 30,000 proteins. I used the docker implementation of MaSIF to generate the interaction sites of PDB ID (6LU7).

Molecular Docking. First, I obtained the crystal structure of M^{PRO} (PDB ID:6LU7) deposited by Jin et al from RCSB PDB databank²⁰. To prepare the M^{PRO} protein for docking, I used AutoDock Tools (ADT) to assign Kollman charges and atom/bond types²¹. 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 carry out the conversion between SDF and PDBQT file formats². Screening against M^{PRO} was performed using autodock vina code and defined a grid box to sufficiently cover the active site of protein²². Box size was defined as $28 \times 28 \times 28 \text{ \AA}^3$ with the center at (x = 6.194, y = 1.944, z = 24.778). Using docking, I obtained inhibitors positions in the binding site of the protein which was used as an initial pose for dynamics validation.

Classical Molecular Dynamics Simulations. All molecular dynamics (MD) simulations were carried out using NAMD²³. I used CHARMM-GUI webserver to generate the CHARMM36m parameters and topology files for the protein and SwissParam server to generate topology and parameters for drugs (small molecules)^{24,25}. All complexes were enclosed in triclinic boxes. The TIP3P water model was used to solvate all complexes. Ions (Na^+ and Cl^-) were added at 0.15 M concentration 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. Particle mesh Ewald (PME) scheme was used to calculate long-range interactions²⁶. The equilibrated systems were finally submitted for a 10 ns NPT production run without position restraints. To compute the time dependent RMSD of drug molecule, with respect to the docking pose (drug and protein's secondary structure aligned), I averaged over 800 snapshots from 2ns to 10ns.

ANI Potential and Dynamics. Here in this work, I combined accurate ANI forcefields for drugs with CHARMM36m forcefields of protein and ions alongwith TIP3 for water to run the drug binding dynamics against M^{PRO} (Figure 1). I performed ANI/MM calculations as per the hybrid model previously combined by Rowley group²⁷. 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 surroundings (protein, water, and ions), and v of the interactions between the drug and its surrounding (Equation (1))²⁷.

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

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 CHARMM36m force field. For non-covalent protein-drug binding, the $v_{ANI/MM}$ term is the conventional MM non-bonded 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))²⁷.

$$v_{ANI/MM}(r_{ANI}) = \sum_i^{MM} \sum_j^{ANI} \frac{q_i q_j}{4\pi\epsilon r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2)$$

Figure 1. Initial system for MD simulation. The initial system consists of docking pose conformation of 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 represented with lines representation. Neutralizing ions Na^+ and Cl^- (concentration 0.15 M) are represented with a green and brown mesh bubble.

ANI/MM/PBSA free energy. To compute the ANI/MM/PBSA free energy of protein-drug binding, I used CaFe (open source code for calculation of free energy) developed by Liu et al (2016)²⁸. CaFe is an automated tool to predict binding free energy using endpoint methods like

molecular mechanics Poisson-Boltzmann surface area (MM/PBSA). This tool is specially designed to work to directly take CHARMM/NAMD file formats (like PSF, DCD, etc)²⁸. In this work, I used trajectory files obtained from hybrid ANI/MM calculations from 0.2 ns to 1 ns (80 snapshots in total). As the trajectory files involve drug molecule defined using ANI forcefield, free energy will be computed with $\Delta E_{ANI/MM}$ (similar to $v_{ANI/MM}$ term in equation 2) instead of ΔE_{MM} . Here in this ANI/MM/PBSA computations, ANI/MM term was computed for gas-phase. For PB calculation, I used APBS programs (interface to CaFe) where boundary conditions were set to multiple Debye Huckel and charges were mapped to grids using cubic B spline method. To perform the SA calculations, surface tension value was defined as 0.00542 with a surface offset of 0.92. At last the binding energy is summed and averaged over an ensemble of conformations (80 snapshots) as follows.

$$\Delta G_{bind,aq} = \Delta H - T\Delta S \approx \Delta E_{ANI/MM} + \Delta G_{bind,solv} - T\Delta S \quad (3)$$

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)²⁹. To generate the 2D-interaction map of drug with the protein, I used academically free Maestro software from Schrödinger (GUI)³⁰. To map the 3D - interaction of drugs with the residues in the binding site, I used the PLIP code (fully automated protein-ligand interaction profiler)³¹. With the combination of my code with PLIP, I report the count 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} (main protease to COVID-19) solved and deposited by Jin et al (6LU7 chain A) from the RCSB PDB databank for all the analyses carried out in this work. I used MaSIF method to predict PPI sites on the protein surface which are defined as the surface patches on a protein that are more likely to interact with other proteins and potential drug molecules. The MaSIF method predicted five interaction sites on M^{PRO} surface having the highest score, I termed these sites as A, B, C, D, and E (Figure 2). Site ‘A’ consists of residues 140-145, 163-168, 187-192, site ‘B’ consists of residues 112-116, 124-128, 137-140, site ‘C’ covers residues in regions 192-199, 235-239, site ‘D’ with residues 292-301 and site ‘E’ consisting

of residues 65-68, 19-22. Furthermore, I also found that the site ‘A’ matches with the surface patch where N3 inhibitor was identified in a bound form with the crystal structure of protein. For the drug screening part performed in this work, I only work with site ‘A’ while other sites reported here can also be potentially used to further screen a database of protein or drugs against M^{PRO}.

Figure 2. Interactions sites of COVID-19 target protein. The surface representation of the main protease of COVID-19 (PDBID: 6LU7) generated using MaSIF code and visualized in Pymol. The red region represents the best interaction sites also labeled as A, B, C, D, and E. The left and right representation is for the same structure but with a 180 degrees rotation. The rightmost panel shows N3 inhibitor which is included with the 6LU7 crystal structure. The N3 inhibitor bounds to site A of the protein.

The next step involved in this work is to carry out virtual screening of FDA approved drugs or repurposing drugs against M^{PRO}. To perform the task, I first obtained the FDA approved drugs (dock ready) database from ZINC15. To pick drugs that can fit my criterion, I filtered out compounds that have atoms other than C, H, N, and O and removed out drugs with any charged atoms (so that ANI forcefield can be used), reducing the size of the drugs in the database to 1188 drugs. Next, I performed molecular docking-based screening of 1188 drugs using AutoDock Vina and ranked each drug with their decreasing binding affinity. With the ranking list of drugs and

their docking scores, I selected 91 compounds with the binding affinity score equals or below -8.0 kcal/mol. From the distribution of 91 compounds, I found that 10% of compounds have a score below -9.0 kcal/mol, 20% of compounds were found in between -8.9 kcal/mol and -8.5 kcal/mol, 30% of compounds with a range of -8.4 kcal/mol and -8.1 kcal/mol (both included) and remaining 40% have a score of -8.0 kcal/mol (Figure 3). As docking programs have the tendency to generate false poses, so to validate these scores, I performed classical MD simulations.

Figure 3. Distribution of 91 drugs selected based on docking screening. 10% of the best drugs were found to have score below -9.0 kcal/mol, 20% of the drugs have score between -8.9 kcal/mol and -8.5 kcal/mol, 30% with scores between -8.4 kcal/mol and -8.0 kcal/mol and rest of the drugs have scored exactly -8.0 kcal/mol.

MD simulations with classical forcefields (like CHARMM36m) are well known to 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. I carried out 10 ns MD simulations of each 82 drugs out of 91 drugs (with available parameters on SwissParam) in complex with protein in a TIP3 solvated box and with Na⁺ Cl⁻ as neutralizing ions. With the trajectories obtained, I attempt to classify the drugs binding with the drugs leaving from the active site of the protein. To numerically deal with this, I computed the average RMSD value of each drug with the secondary structure of protein aligned. With the drugs arranged in their decreasing RMSD values, I obtained the highest RMSD of ~10 Å which strongly indicates that drugs do not bind to the protein and the lowest RMSD of ~1.5 Å which implies favorable drugs.

In this manner, and with a cutoff of 4 Å, I classified 37 drugs as binding, and the rest of the drugs as leaving and leaving drugs were disqualified to step into the next level (Figure 4). RMSDs of selected drugs along with their ZINC-ID is shown in Supplementary Table 1.

Figure 4. Average RMSD values of 91 drugs. The average root mean square deviation (RMSD) value of trajectory of drug in complex with protein with respect to the docking pose with ligand and secondary structure of protein aligned to the docking pose (initial frame). Each bar represents average RMSD values from 2ns snapshot to 10ns for each drug. The blue bars represent binding drugs with average drug RMSD below 4 Å and green bar represents leaving drugs with average ligand RMSD above 4 Å.

To accurately describe the inhibition profile of above-selected drugs, I extensively carried out ANI/MM dynamics using NAMD. To obtain the initial conformation for ANI/MM setup, I extracted the last snapshot from the 10 ns MD production run which was performed with the classical forcefield. To setup the ANI/MM system, I described the ANI region to contain the atoms of drug molecule and combined the ANI region with the region defined by classical forcefield (CHARMM36m) for the residues of the protein, water, and ions. Then, I carried out short 1 ns of ANI/MM MD at constant pressure and temperatures. The RMSD computations on 1 ns of trajectory (100 snapshots) produced by ANI/MM dynamics were carried out with respect to the 10 nsth frame (from classical MD) to again classify binding drugs from the leaving drugs. With a similar methodology of calculating average RMSD (for whole 1 ns) and a cutoff of 4 Å, I again filtered out 18 drugs and selected 19 drugs (Figure 5). To further investigate the set of selected 19

drugs, I examined the binding affinity of each one of them with protein by computing binding free energy using the MM/PBSA method. With ANI/MM trajectories as an input, I termed these calculations as ANI/MM/PBSA replacing MM term with ANI/MM. From the energies obtained from ANI/MM/PBSA, I found 12 out of 19 drugs to be binding spontaneously, having their free energy to be negative (Figure 5). I also found two of the drugs, Targretin and Eltrombopag with the highest binding affinity, ANI/MM/PBSA free energy reported to be -11.74 and -11.16 kcal/mol respectively. The MM/PBSA data of selected 12 drugs binding with protein (energy differences of terms Elec, vdW, Pol, Sol, PB, SA, Gas and Npol) is reported in Supplementary Tables S1 and S2.

Figure 5. RMSD and ANI/MM/PBSA values from ANI/MM MD. Average RMSD values (from 0.2 ns to 1 ns) of the trajectories generated from ANI/MM simulations. Blue bars represent binding drugs (below 4 Å) and green bars represent (above 4 Å). Grey bars represent ANI/MM/PBSA free energy for binding drugs, only showing drugs with the negative free energy of binding.

With my final selection of 12 drugs, I further take a step forward to characterize the drug-protein interaction from the trajectories produced by ANI/MM calculations. Using *ttclut*, I clustered snapshots from ANI/MM trajectories based on their similarity. I found that trajectories of 4 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). The trajectory profiles are an important tool to analyze the drug-binding protein interaction and provides much more details then computing RMSDs. With this description, I further comment that 11 out of 12 drugs I selected will tend to stay in longer simulations. With these details, I further work on analyzing the 2D interaction maps from the last trajectory of ANI/MM calculation and combine plip with the 2D drug/ligand interaction images generated from Maestro. From the data, I found GLU166, GLN189, THR190 to be the key residues in the binding site which were common to interact with most of the selected drugs. The other key residues of protein in binding the drugs to the binding site of protein can be easily identified i.e. the residues of protein which are common to 12 selected drugs (Figure 7).

Figure 6. Clustering of ANI/MM MD trajectories: Clustering of trajectories generated using ANI/MM dynamics. 12 drugs selected based on the ANI/MM/PBSA are shown. Different colors represent different clusters. Total of 100 snapshots were clustered for each drug-protein binding simulation.

Figure 6. 2D interaction map of selected drugs with M^{PRO}. 2D - Structures of 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 getting affected and fatalities, I present another investigation on potential inhibitors against it by drug repurposing approach. The whole scientific community is putting a tremendous amount of effort in characterizing potential drugs inhibiting this virus and still a vast amount of information is needed until a unique treatment is approved³². Careful studies on the interaction of the protein with available drugs that are FDA approved are crucial to discover binding behavior of protein and aids in accelerating experimental biochemical assays³³. Molecular dynamics with classical forcefield can describe many important interactions

for drug binding to protein but might tend to miss little details at electronic and nuclei level³⁴. These missed details can be recovered if 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 description of electronic and nucleic phenomenon of small drug molecules. In this work, using neural network force fields, I explored the interactions of repurposed drug molecules binding to M^{PRO} protein. The findings of this study, interaction sites and structure of selected drugs might be important to design a novel drug which can fight against COVID-19.

To initiate the conventional process of drug discovery, identification of drug binding sites is extremely important and several methods to identify binding sites of the protein have already been proposed and are proved to perform an accurate search on the protein surface. Neural network methods like MaSIF takes advantage of accurate methods for training and mimic the performance as compared to the physical chemistry method based on prediction accuracy¹⁹. MaSIF method also provides other pipelines, one of them can be used to screen other proteins that can interact with our protein of interest based on the descriptors generated for site identification. I limit this work to just identifying sites for PPI and using a single site (site 'A') for drug screening and binding dynamics. PPI sites are reported to be important in suggesting interaction sites for protein as well as drugs³⁵. 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, I started filtering drugs from FDA approved drugs database based on the atoms and type of charges. With the molecular docking, I reduced the number of the drugs to only 91 by selecting specific drugs i.e. with the best docking score. To sample the conformational space of drugs protein complexes, I performed molecular dynamics simulations using classical molecular dynamics and then used ANI/MM approach for providing an accurate description of the interaction profile of drugs with protein. The steps involving virtual screening and classical MD is a traditional pipeline for any drug discovery process and the workflow has already been performed by several groups^{5-9,19}. In this work, I further combine classical MD with ANI/MM dynamics to crucially investigate the other set of interactions which might be ignored by classical MD simulations and might be useful for experimental drug design. Also, combining ANI/MM trajectories with endpoint MM/PBSA computations assists in fetching

physically important information related to binding. ANI/MM/PBSA calculations reported in this work still need major improvement to handle the ANI/MM interactions and will be another research area to explore in the future. As ANI forcefields are still in development nature and analysis tools are not as developed as compared to the classical MD, I used some basic tools to analyze the ANI/MM simulations. I clustered ANI/MM trajectories to further validate my outcome and found that 100 snapshots were clustered in as few as 3 or 4 clusters (Figure. 6). This result can also be implied in a way that drug binding to protein doesn't induce drastic conformational changes when number of clusters are limited to 3 or 4 (for 100 snapshots). I found one of the drugs where binding trajectories were clustered in 5 clusters implying rapid conformational changes and thus can be filtered out keeping only 11 out of 12 drugs. This workflow with the trajectory validation can be useful for analyzing other simulation which involves drug interaction with M^{PRO} or other protein in the future.

With my findings, I report structures of the drug like Targretin and Eltrombopag to be the potent drugs with the most inhibitory action against M^{PRO} protein. I also identified the other 10 drugs that might be repurposed against M^{PRO}. On breaking the output from each level, I report 91 FDA approved drugs from virtual screening, 37 drugs from MD simulations profile, 19 drugs from ANI/MM Trajectories. I attempted to provide the maximum information related to each of the drugs and tried to cover all the interaction aspects of these drugs with the M^{PRO} in this study. The structure of shortlisted drugs and their detailed interaction fingerprints with the protein are discussed in great detail which can be used to discover other small molecules. This study performs only computational studies and drugs reported in this work are only characterized by computations. No experiments have been performed related to these drugs binding with the protein.

Author Information

*Corresponding Author

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

Acknowledgments

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

Author declares no competing financial interest.

Supplementary Information

Supplementary Table S1. Table containing ZINC IDs, Vina Score, Marketed name of the drug, ANI/MM/PBSA energy and the number of hydrophobic and hydrogen bond interaction obtained from PLIP tools.

ZINC15-ID	Vina Score (kcal/mol)	Marketed Name of the Drug molecule	ANI/MM/PB SA Energy (kcal/mol)	No. of Hydrophobic Bonds b/n Drug-Protein (PLIP)	No. Of Hydrogen-Bonds b/n Drug-Protein (PLIP)
ZINC000001539579	-8.2	Targretin	-11.7393	1	1
ZINC000011679756	-9.0	Eltrombopag	-11.1614	5	3
ZINC000169621200	-8.1	Rifaximin	-9.2305	2	4
ZINC000004212809	-8.0	Deflazacort	-8.8929	1	0
ZINC000052955754	-9.1	Ergotamine	-6.6227	2	2
ZINC000094566092	-8.4	Doxazosin	-6.1367	0	4
ZINC000011726211	-8.0	Lastacaft	-5.4417	1	1
ZINC000169621223	-8.1	Rifampicin	-3.9846	0	6
ZINC000014210455	-8.1	Victrelis	-3.2906	0	9
ZINC000003820029	-9.3	Trajenta	-2.8430	2	1
ZINC000022448696	-8.2	Indinavir	-2.2321	2	4
ZINC000003938684	-8.9	Toposar	-0.6248	1	4
ZINC000001481815	-8.1	Exjade	0.9994	3	3
ZINC000003938704	-9.7	Ellence	1.3162	1	2
ZINC000019632618	-8.4	Imatinib	1.9175	1	4
ZINC000026985532	-8.7	Saquinavir	2.0107	2	2
ZINC000009212428	-8.0	Leucal	3.4004	0	9
ZINC000003917708	-8.9	Daunorubicin	11.7926	3	8
ZINC000008220909	-8.0	Natamycin	14.8378	2	8

Supplementary Table 2. ANI/MM/PBSA energy data for selected 12 drug molecules.

Drug Name	PB	Pol	SA	Sol	Npol	Elec	Gas	vdW	Total
-----------	----	-----	----	-----	------	------	-----	-----	-------

Targretin	34.9433	31.3743	-4.8578	30.0854	-43.1136	-3.5689	-41.8247	-38.2558	-11.7393
Eltrombopag	51.9919	41.6501	-5.6052	46.3867	-52.8115	-10.3417	-57.5481	-47.2063	-11.1614
Rifaximin	57.8197	38.5798	-5.486	52.3337	-47.8102	-19.24	-61.5642	-42.3242	-9.2305
Deflazacort	41.0969	33.1243	-4.5814	36.5155	-42.0171	-7.9727	-45.4084	-37.4357	-8.8929
Ergotamine	57.1271	39.8994	-5.415	51.7121	-46.5222	-17.2277	-58.3349	-41.1072	-6.6227
Doxazosin	67.1361	42.1584	-5.4402	61.6959	-48.2951	-24.9777	-67.8326	-42.8549	-6.1367
Lastacaft	48.508	33.5135	-4.3079	44.2	-38.9552	-14.9945	-49.6418	-34.6473	-5.4417
Rifampicin	80.4137	58.2085	-6.1423	74.2715	-62.1931	-22.2052	-78.256	-56.0508	-3.9846
Victrelis	56.7291	31.4692	-4.2642	52.4648	-34.7598	-25.2598	-55.7555	-30.4956	-3.2906
Trajenta	62.9573	19.9315	-3.1942	59.7631	-22.7744	-43.0258	-62.606	-19.5802	-2.8430
Indinavir	58.1583	41.1596	-5.2445	52.9138	-43.3917	-16.9987	-55.1459	-38.1472	-2.2321
Toposar	71.1821	54.2041	-5.8966	65.2855	-54.8288	-16.978	-65.9103	-48.9323	-0.6248
Exjade	68.8271	36.6762	-4.557	64.2701	-35.6768	-32.1509	-63.2707	-31.1198	0.9994
Ellence	144.1884	46.1297	-4.7783	139.4101	-44.8135	-98.0586	-138.0938	-40.0352	1.3162
Imatinib	146.8817	48.0101	-5.5096	141.3722	-46.0926	-98.8716	-139.4547	-40.5831	1.9175
Saquinavir	93.6254	65.0741	-6.8566	86.7688	-63.0633	-28.5514	-84.7581	-56.2067	2.0107
Leucal	105.2573	53.5946	-5.5217	99.7356	-50.1942	-51.6627	-96.3352	-44.6725	3.4004
Daunorubicin	242.8116	60.3981	-5.8227	236.9889	-48.6055	-182.4135	-225.1963	-42.7829	11.7926
Natamycin	262.0643	65.1207	-6.6677	255.3966	-50.2829	-196.9435	-240.5588	-43.6153	14.8378

References.

1. Xu, Z. *et al.* Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **8**, 420–422 (2020).
2. Huynh, T., Wang, H. & Luan, B. *In Silico* Exploration of the Molecular Mechanism of Clinically Oriented Drugs for Possibly Inhibiting SARS-CoV-2's Main Protease. *J. Phys. Chem. Lett.* 4413–4420 (2020) doi:10.1021/acs.jpcclett.0c00994.
3. Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* 1–5 (2020) doi: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* vol. 162 1239–1249 (2011).
5. Wang, J. Fast Identification of Possible Drug Treatment of Coronavirus Disease -19 (COVID-19) Through Computational Drug Repurposing Study. *J. Chem. Inf. Model.* (2020) doi:10.1021/acs.jcim.0c00179.
6. 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* (2020) doi:10.1080/07391102.2020.1752802.
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.* 1–12 (2020) doi:10.1080/07391102.2020.1758791.
 8. Muralidharan, N., ... R. S.-J. of & 2020, undefined. Computational studies of drug repurposing and synergism of lopinavir, oseltamivir and ritonavir binding with SARS-CoV-2 Protease against COVID-19. *Taylor Fr.*
 9. Elmezayen, A., Al-Obaidi, A., ... A. Ş.-J. of B. & 2020, undefined. Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *Taylor Fr.*
 10. Nutho, B. *et al.* Why Are Lopinavir and Ritonavir Effective against the Newly Emerged Coronavirus 2019? Atomistic Insights into the Inhibitory Mechanisms. *Biochemistry* (2020) doi:10.1021/acs.biochem.0c00160.
 11. Vamathevan, J. *et al.* Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery* vol. 18 463–477 (2019).
 12. Sliwoski, G., Kothiwale, S., Meiler, J. & Lowe, E. W. Computational methods in drug discovery. *Pharmacological Reviews* vol. 66 334–395 (2014).
 13. Wang, L. P., Martinez, T. J. & Pande, V. S. Building force fields: An automatic, systematic, and reproducible approach. *J. Phys. Chem. Lett.* **5**, 1885–1891 (2014).
 14. Balabin, R. M. & Lomakina, E. I. Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies. *J. Chem. Phys.* **131**, 074104 (2009).
 15. Morawietz, T. & Behler, J. A density-functional theory-based neural network potential for water clusters including van der waals corrections. *J. Phys. Chem. A* **117**, 7356–7366 (2013).
 16. 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.* **8**, 3192–3203 (2017).
 17. Smith, J. S. *et al.* Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **10**, 1–8 (2019).

18. Smith, J. S., Roitberg, A. E. & Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Medicinal Chemistry Letters* vol. 9 1065–1069 (2018).
19. Gainza, P. *et al.* Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
20. Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* (2020) doi:10.1038/s41586-020-2223-y.
21. Forli, S. *et al.* Computational protein-ligand docking and virtual drug screening with the AutoDock suite. *Nat. Protoc.* **11**, 905–919 (2016).
22. 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.* **31**, NA-NA (2009).
23. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* vol. 26 1781–1802 (2005).
24. Zoete, V., Cuendet, M. A., Grosdidier, A. & Michielin, O. SwissParam: A fast force field generation tool for small organic molecules. *J. Comput. Chem.* **32**, 2359–2368 (2011).
25. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865 (2008).
26. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593 (1995).
27. Lahey, S. L. J. & Rowley, C. N. Simulating protein-ligand binding with neural network potentials. *Chem. Sci.* **11**, 2362–2368 (2020).
28. Liu, H. & Hou, T. CaFE: a tool for binding affinity prediction using end-point free energy methods. doi:10.1093/bioinformatics/btw215.
29. Tubiana, T., Carvaille, J. C., Boulard, Y. & Bressanelli, S. TTClust: A Versatile Molecular Simulation Trajectory Clustering Program with Graphical Summaries. *J. Chem. Inf. Model.* **58**, 2178–2182 (2018).
30. Schrödinger, S. R.-, LLC, undefined, York, N., NY, undefined & 2017, undefined. 1: Maestro.
31. Salentin, S., Schreiber, S., Haupt, V. J., Adasme, M. F. & Schroeder, M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res.* **43**, 443–447 (2015).

32. Association for the Advancement of Science, A. Race to find COVID-19 treatments accelerates. (2020) doi:10.1126/science.367.6485.1412.
33. Rismanbaf, A. Potential Treatments for COVID-19; a Narrative Literature Review. *Arch. Acad. Emerg. Med.* **8**, e29 (2020).
34. 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.* **110**, 10452–10467 (1999).
35. Fuller, J. C., Burgoyne, N. J. & Jackson, R. M. Predicting druggable binding sites at the protein-protein interface. *Drug Discovery Today* vol. 14 155–161 (2009).