Repurposing Peptidomimetic Drugs as Potential Inhibitors of Novel SARS-CoV-2 Main Protease Using Network Theory: A Protein-Ligand Binding Site Similarity Approach

Sagar Bhayye
Shiv Nadar University - Campus Delhi NCR: Shiv Nadar University
Nagamani Sukumar (n.sukumar@snu.edu.in)
Shiv Nadar University https://orcid.org/0000-0002-2724-9944

Research Article

Keywords: Protein-ligand binding sites, Network, SARS-CoV-2, Main protease, Drug repurposing

Posted Date: March 3rd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1353780/v1

Abstract

A new strain of coronavirus known as severe acute respiratory syndrome (SARS) coronavirus-2 (SARS-CoV-2) is responsible for the current COVID-19 pandemic, which has not only affected the health of millions of individuals, but also caused severe socio-economic disruption. To curb the spread of the virus, various strategies have been employed to develop new drug therapy. Considering the pandemic condition, targeting the substrate binding site of main protease enzyme of SARS-CoV-2, which plays an important role in replication of the coronavirus, will be beneficial. Its high similarity with its predecessor virus's main protease and dissimilarity with human protease makes it a promising target and will also facilitate drug repurposing. In this study, we have used Protein Encoded Shape Distributions (PESD) to calculate similarity between the ligand binding site of the SARS-CoV-2 main protease and other proteins. Similarity networks were constructed between these proteins using different distance metrics with edge density < 0.1 percent. Construction of low edge density protein-ligand binding site similarity networks helped in rational identification of the most similar ligand binding sites of proteins with the SARS-CoV-2 main protease. Based on this knowledge, a dataset of FDA approved drug molecules as well as experimental drugs was collected from the literature. These molecules were subjected to virtual screening through molecular docking against the SARS-CoV-2 main protease, followed by conventional molecular dynamics simulation, replica exchange molecular dynamics and binding free energy calculations. Based on these studies Q27458218 have been found to be suitable for repurposing as a SARS-CoV-2 main protease inhibitor.

Introduction

The COVID-19 pandemic the world is currently facing, is caused by novel SARS-CoV-2, a new strain of coronavirus, which belongs to the virus family known as Coronaviridae [1, 2]. The genome of coronaviruses are the largest known viral RNA genomes with sizes ranging from 26 to 32 kilobases. The viral genome is an enveloped, unsegmented, positive-sense single-stranded RNA. This single-stranded RNA along with phosphorylated nucleocapsid protein is covered by phospholipid bilayers, which are further wrapped by layers of two different types of spike proteins forming a corona known as the spike glycoprotein trimmer that can be found in all kinds of corona viruses. In some corona viruses these spike proteins exist as hemagglutinin-esterase [3], and help the virus attach itself to a complementary host cell receptor protein [4]. Several coronaviruses cause diseases in mammals and birds. In humans they mainly infect the respiratory tract and may cause Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS) or the more recent pneumonia-like illness COVID-19 [5].

The SARS-CoV-2 strain is related to the beta-coronavirus first identified in 2002 after the first global outbreak of coronaviruses was traced to bats [6]. COVID-19 has now affected 227 countries and territories with a frightening rate of transmission, resulting in disruption of social life and downturn in the world economy. At the present time over 409 million cases and 5.8 million deaths have been recorded. Since the outbreak of the COVID-19 pandemic, there has been increasing urgency to develop a therapy to treat novel SARS-CoV-2, which includes vaccines and antiviral agents. Despite this, few treatments are currently available, such as *Molnupiravir*, a small-molecule ribonucleoside prodrug of N-hydroxycytidine, and dexamethasone, a corticosteroid used for its anti-inflammatory and immunosuppressant effects. Some anti-retroviral drugs such as remdesivir, lopinavir, ritonavir, etc. along with anti-inflammatory drugs which have shown some positive results in COVID-19 patients, have sometimes been prescribed for treatment [7]. However, the response to this treatment varies from individual to individual. Therefore, rapid identification of new drug candidates for better treatment is urgently needed. As some antiviral and anti-inflammatory drugs have shown promising results in COVID-19 treatment, therefore screening, identifying and repurposing already FDA-approved drugs and drug candidates in the investigational phase or experimental phase, can be an effective approach. Rapid virtual screening of drug molecule libraries gives an edge by saving time and cost. However, rationalizing the target protein and strategies for virtual screening of drug databases are very important when working with limited resources and time.

Viral spike proteins, envelope protein, protease, nucleocapsid protein, 2'-O-ribose methyltransferase and 3CL hydrolase of novel SARS-CoV-2 are rapidly emerging as targets to prevent its replication [8]. Some studies already have been reported on virtual screening of FDA approved drugs and various molecular databases against these targets [9–13]. The viral genome encodes two proteases (PL^{pro} and $3CL^{pro}$) that are vital for virus replication; both cleave the two translated polyproteins (PP1A and PP1AB) into individual functional components [14]. The 3-chymotrypsin-like protease ($3CL^{pro}$) also known as main protease (Mpro) is considered to be a promising drug target because as it is dissimilar to human proteases, and because the sequence and structure of the main protease are closely linked to other beta-coronaviruses, aiding drug discovery endeavors based on earlier lead compounds [15].

SARS-CoV-2 main protease, consists of three functional domains, domain I (residues 10–99) and domain II (residues 100–182) consisting of 2-β sheet barrel fold, while domain III (residues 198–303) forms a compact α-helical domain connected to domain II by a long linker loop, see Fig. 1 [16]. The substrate binding active site targeted in this study is located in a cleft present in-between domains I and II, which also holds a histidine/cysteine catalytic dyad. Both histidine 41/cysteine 145 catalytic dyad and N-terminus residue 1–7 play a vital role in proteolytic activity. The substrate binding active cavity consists of four sub pockets – S1`, S1, S2, S4. The aim of this study is to identify drug candidates using a combined approach of network theory, virtual drug screening, molecular docking and molecular dynamics. For this purpose, Property-Encoded Shape Distributions (PESD) are employed to construct a similarity network of protein-ligand binding site surfaces [17, 18]. Proteins with very low sequence conservation in the binding site region, but with similar electrostatics and lipophilicity, have been seen to bind similar ligands. PESD provides a "ligand's eye" view of a protein's ligand binding site by encoding the distributions of electrostatic and lipophilic potentials on the binding site surface. A similarity network of the novel SARS-CoV-2 main protease, constructed with PESD signatures, identifies proteins having similar ligand binding sites to the main protease. Ligands that are known to bind to these similar protein binding sites are thus likely to also bind to the SARS-CoV-2 main protease. Accordingly, a virtual library of drug candidates in clinical, investigational and experimental phases which bind to these proteins, is screened against the novel SARS-CoV-2 main protease, thereby providing potential lead drug candidates in the fight against COVID-19.

Materials And Methods

Protein-ligand binding site similarity network

In this study, a set of 6916 protein crystal structures belonging to different classes and species were randomly selected and used with 97 SARS-CoV-2 main protease crystal structures to construct a similarity network based on protein-ligand binding site properties. All crystal structures were prepared using Chimera molecular modelling tool [19]. Preparation of the dataset involved addition of hydrogens, other missing heavy atoms and side chains of amino acids, followed by addition of charges on amino acid residues using AMBER (ff14SB) force field [20]. Charges on ligands were added using AM1-BCC method. Finally, the complexes were energy minimized using 100 steepest descent steps, followed by 10 conjugate gradient steps. The Property-Encoded Shape Distributions (PESD) method was utilized to calculate similarities between ligand binding sites of the selected set of proteins. The PESD server [21, 22] calculates surface complementary descriptors which encode the distribution of polar and nonpolar regions, electrostatic potential as well as lipophilicity on the protein-ligand interaction surface between protein and ligand surfaces and uses metrics such as Euclidean, Chi-Square and Manhattan distances [23–25] to calculate similarities between ligand binding sites of the protein crystal structures present in its database. The output consists of crystal structures ranked according to binding site similarity with the input crystal structure. A similarity threshold is selected such that the edge density of each network remains less than 0.1%. Then adjacency matrices were prepared and used for construction of similarity networks, which were then analyzed for properties such as vertex degree, average path length, degree distribution, average clustering coefficient, degree assortativity, modularity and different centrality measures [26].

Molecular docking

Based on binding-site similarity networks between SARS-CoV-2 main protease and other proteins, information on 1305 molecules binding to proteins/enzymes which showed high binding-site similarity to SARS-CoV-2 main protease were collected from the published literature and their 3D structures downloaded from PUBCHEM [27] (Listed in the Supplementary Material). This dataset contains molecules which are in experimental or investigational phase, and in clinical use for different therapeutic conditions. To perform molecular docking to SARS-CoV-2 main protease, the co-crystal structure with the best resolution (PDB ID: 7BRR) was retrieved from the protein data bank [28]. The crystal structure was prepared by using the Protein Preparation Wizard of Schrödinger Suite [29]. Protein preparation involved addition of missing hydrogens, optimization of the hydrogen bond network of amino acid residues, prediction of the ionization state of histamine residues, followed by energy minimization of the crystal structure using OPLS 2005 force field to minimize steric clashes [30]. Co-crystalized water molecules beyond 5 Å of the inhibitor binding site were deleted during the protein preparation process. The grid was defined and generated with a default rectangular box centered on the inhibitor binding site residues of SARS-CoV-2 main protease. The dataset of collected molecules was prepared for docking studies using the LigPrep module of Schrödinger suite [31]. During ligand preparation, different ionization and tautomer states of the molecules were generated, followed by geometry optimization of each conformer using OPLS 2005 force field. Finally, molecular docking was performed using the Glide module with extra precision (XP) mode [32].

Molecular Dynamics and binding free energy calculations

Based on interaction and docking score, 19 protein-ligand complexes were selected for molecular dynamics (MD) studies. To study the stability of protein-ligand interactions of the complexes obtained from molecular docking studies, MD simulations were performed in Gromacs software 2018.1 [33]. The CHARMM36 (March 2019) force field was used to prepare the protein topology and the CHARMM General Force Field (CGenFF) server used to generate the ligand topology [34, 35]. Explicit water molecules of the TIP3P water model were added by keeping the protein-ligand complex at the center of a dodecahedral box at 10 Å distance from the edges [36]. The charge of each system was neutralized by adding the appropriate number of sodium ions. Then a steepest descent minimization algorithm was used to minimize the energy of each system to optimize the hydrogen bond networks [37]. Equilibration of each system was performed with NVT and NPT ensembles respectively for 100 ps to avoid distortion of the protein-ligand complex. Exploratory MD simulations were performed for 15 ns at 300 K. Periodic boundary conditions were applied during the MD simulation to avoid boundary effects. The Particle Mesh Ewald (PME) method was used to estimate long-range electrostatic interactions [38]. Energy and co-ordinates of the system were recorded every 10 ps. Hydrogen bond interaction analysis between protein and ligand during MD was performed in visual molecular dynamics (VMD) tool by keeping the cut off distance 3 Å and hydrogen bond angle 20° [39]. The binding free energy of each system was determined by MM-PBSA method [40, 41]. Compounds which show stable interactions with SARS-CoV-2 main protease and average binding free energy less (more negative) than − 50 kcal/mol during 15 ns MD were selected and the MD simulation extended for 100 ns. MD simulation trajectories were again analyzed, and the binding free energies of the compounds calculated.

Replica Exchange Molecular Dynamics

To further assess the binding affinity between Q27458218 in complex with SARS-CoV-2 main protease, replica exchange molecular dynamics (REMD) was performed using multiple replicas in Gromacs software 2018.1 [33]. Each system was equilibrated at a target temperature of 300 K, 303 K, 306 K and 309 K, thus creating four replicas for each system. Production MD of each replica was performed for 100 ns with time step of 2fs. Exchanges between replicas were attempted every 100 steps for all simulations. After completion of MD, REMD trajectories were de-multiplexed using demux.pl script provided in Gromacs to get continuous MD trajectories. All the MD trajectories of four replicas for each system were then subjected to structural, protein-ligand interaction and MM-PBSA analysis.

Results And Discussion

Protein-ligand binding-site similarity network

The networks constructed using the different similarity metrics are depicted in Fig. 2. Nodes represent ligand binding sites of proteins and edges connect similar binding sites. The binding sites of SARS-CoV-2 main protease are shown in red color, while other protein binding sites similar to the main protease are shown in blue. In the layout employed, hubs connecting various similar nodes and clusters are concentrated at the center, while disconnected communities of nodes migrate to the periphery. Properties of the three similarity networks were compared with the Erdös-Renyi (ER) random network (Table 1) at similar edge density [42, 43]. The protein binding-site similarity networks have high clustering coefficients and lower average path length than the random network, suggesting that these networks show small world behavior. This means all three networks have small diameters and two nodes are separated by ≤ 6 steps irrespective of their position in the graphs. The algebraic connectivity or Fiedler eigenvalue of a graph is the second-smallest eigenvalue of its Laplacian

matrix (defined as L = D − A, where D is the diagonal matrix of vertex degrees, and A is the adjacency matrix); its magnitude is a measure of the overall connectivity of the graph [44, 45]. Low average path length as well as high value of the second-smallest eigenvalue of the Laplacian matrix in the Manhattan network suggest highly similar protein-ligand binding sites. The Euclidean network shows positive coefficient of degree assortativity, suggesting that high degree nodes preferentially connect to high degree nodes. Negative degree assortativity coefficient of the other networks suggest high degree nodes preferentially connect to low degree nodes in the Chi-Square and Manhattan networks. The group of similar protein-ligand binding sites is divided into different communities, as indicated by the modularity. The high modularity value in the Chi-Square and Manhattan networks suggest these networks cluster into different communities. A total of 153 unique protein/enzymes showing binding-site similarity to the SARS-CoV-2 main protease (Fig. 3) are listed in supplementary material (Supplementary file.xlsx). These proteins belong to different classes of proteins and found in different species. These group of proteins were then used to identify the ligands which bind to them.

Table 1
Different properties calculated from the three networks at edge density < 0.1%. ER = Erdős–Rényi model. Average clustering coefficient very much greater than that of the ER network and path length less than that of the ER network at similar edge density indicates the small world property.

Parameters	Euclidean Network	Chi-Square Network	Manhattan Network
Distance cutoff	2200	10200	64000
Number of nodes	7013	7013	7013
Number of edges	10562	10455	9203
Average degree	3.03	3.00	2.64
Average clustering coefficient	0.0096	0.022	0.015
Degree assortativity coefficient	0.015	-0.141	-0.30
Transitivity	0.042	0.013	0.0041
Edge density	0.00043	0.00043	0.00037
Average path length	5.23	4.01	3.75
Modularity	0.47	0.64	0.52
2nd smallest Eigenvalue of Laplacian matrix	0.014	0.044	0.052
Average clustering coefficient (ER)	0.00029	0.00027	0.00034
Average path length (ER)	8.01	8.00	9.04
Transitivity (ER)	0.00047	0.00038	0.00025
Small world property	Yes	Yes	Yes

Molecular Docking

Based on interaction and binding affinity given as the Glide XP score, the top 9 molecules that are in clinical use and 10 molecules that are in investigational and experimental phase, were selected, and are shown in Fig. 4, and listed in Table 2. Hit consists of structurally and functionally diverse compounds such as Doxycycline and Minocycline tetracycline-based antibiotics, Neomycin and Tobramycin aminoglycoside-based antibiotics, and polypeptides such as PRD_000771, PRD_000772, Q27458218 and Miraziridine A. While Rutin, Isoquercitrin, Troxerutin and Hyperoside belong to the class of flavonoid-3-o-glycosides. Indinavir which is an HIV protease inhibitor also shows very good affinity towards SARS-CoV-2 main protease. SB-219383 which is a bicyclic compound inhibitor of Tyrosyl-TRNA Synthetase also shows good affinity. The drug molecules Neomycin, Tobramycin, Rutin, Indinavir, Doxycycline, Minocycline, Vorinostat, Alendronate, and Enalapril can be investigated for repurposing as SARS-CoV-2 main protease inhibitors. All the molecules show hydrogen bond interaction with residues such as PHE 140, ASN 142, GLU 166, CYS 145, GLN 189, etc. as depicted in **Figure S1** and **S2** of the supplementary material, while the residue HIE 41 shows π-π interaction with molecules having an aromatic ring. However, not all the sub pockets of the substrate binding site of SARS-CoV-2 main protease are occupied by compounds, so it is necessary to study their behavior using molecular dynamics simulation technique for their interaction stability and sub pocket occupancy.

Table 2
Molecules which show good interaction and binding affinity towards SARS-CoV-2 main protease with their Glide XP scores.

Sr. No	Compound	Glide XP score (kcal/mol)	Indication	Status
1	Rutin	-12.75	Nutraceutical [48]	Clinical Use
2	Indinavir	-10.51	HIV Protease Inhibitor [56]	Clinical Use
3	Neomycin	-9.95	Antibiotics [57]	Clinical Use
4	Tobramycin	-9.90	Antibiotics [57]	Clinical Use
5	Doxycycline	-9.45	Antibiotics [58]	Clinical Use
6	Minocycline	-9.21	Antibiotics [58]	Clinical Use
7	Vorinostat	-9.19	ADAM-10 Inhibitor [59]	Clinical Use
8	Alendronate	-8.88	Matrix Metalloproteinase Inhibitor [60]	Clinical Use
9	Enalapril	-8.80	Liver Carboxylesterase Inhibitor [46]	Clinical Use
10	Isoquercitrin	-12.76	3C-Like Proteinase Inhibitor [47]	Investigational
11	Troxerutin	-12.30	Nutraceutical [48]	Investigational
12	Hyperoside	-12.73	3C-Like Proteinase Inhibitor [49]	Experimental
13	PRD_000772	-12.33	3C-Like Proteinase Inhibitor [50]	Experimental
14	Monoxerutin	-12.13	Nutraceutical [51]	Experimental
15	Q27458218	-11.93	Prostasin Inhibitor [52]	Experimental
16	PRD_000771	-11.48	3C-Like Proteinase Inhibitor [50]	Experimental
17	ARC-1034	-11.10	Protein Kinase A Inhibitor [53]	Experimental
18	Miraziridine A	-11.06	Cathepsin Inhibitor [54]	Experimental
19	SB-219383	-11.01	Tyrosyl-TRNA Synthetase Inhibitor [55]	Experimental

15 ns Molecular Dynamics

MD simulation for a protein-ligand complex is a computationally expensive and time-consuming process. To eliminate compounds which show unstable binding or low binding affinity towards SARS-CoV-2 main protease, initially 15 ns MD simulation was performed for each ligand-protein complex. Snapshots of trajectories at various stages of MD simulation showing movement of the ligand in the SARS-CoV-2 main protease binding site are shown in **Figures S3** and **S4** of the supplementary material. The root mean square deviation (RMSD) of the C-alpha atom of the protein and the complexed ligand is depicted in **Figure S5** of the supplementary material. During 15 ns of simulation it is observed that compounds Neomycin, Tobramycin, Vorinostat, Troxerutin and SB-219383 did not form stable complexes. These break away from the SARS-CoV-2 main protease after 5 to 10 ns of simulation time, except for SB-219383 which is displaced from the binding site after 10 ns. The RMSD of the C-alpha atom of the SARS-CoV-2 main protease shows stable trajectories with all complexed ligands except Indinavir and Isoquercitrin. Compounds Doxycycline, Minocycline, Alendronate, Isoquercitrin and Hyperoside show very low and stable RMSD compared to the other compounds. The RMSD trajectory of compound Q27458218 shows initial increase, then remains stable, while the RMSD trajectory of Monoxerutin after stabilization decreases at 9 ns, then again increases after 13 ns. Visual analysis of MD trajectories suggests that all compounds try to orient and adjust in the best possible pose into the binding site of the SARS-CoV-2 main protease. The root mean square fluctuation of the SARS-CoV-2 main protease is depicted in **Figure S6** of the supplementary material. Atoms from loop residues regions SER 1 – PRO 9, LEU 50 – ASN 53, ASP 153 – ASP 155, PRO 168 – THR 199, GLY 215 – THR 225 and GLY 302 – VAL 3063 show high root mean square fluctuation (RMSF). The number of hydrogen bonds formed by the ligand with the binding site residue in each frame of the trajectory is shown in **Figure S7** of supplementary material. Alendronate forms a large number of hydrogen bonds during the simulation. The number of hydrogen bonds formed by Indinavir and Enalapril with the SARS-CoV-2 main protease decrease as the simulation progresses; however, hydrogen bond formation by ARC-1034 and Miraziridine A are consistent throughout the simulations. The percentage hydrogen bond interaction by protein residue with ligand is listed in **Table S1** of supplementary material. Amino acid residues THR 24, THR 25, THR 26, HIS 41, SER 46, ASN 142, GLY 143, SER 144, HIS 164, GLU 166, ASP 187 and GLN 189 are mainly involved in hydrogen bond formation with different ligands during MD simulation. The average binding free energies (ΔG_{Bind}) of Neomycin, Tobramycin, Vorinostat, Troxerutin and SB-219383 to the SARS-CoV-2 main protease during 15 ns MD simulation, along with their various components, such as van der Waals (vdW), Electrostatic, Polar solvation and solvent accessible surface area (SASA) energy, are listed in **Table S2** of supplementary material. Compounds Q27458218 and Indinavir show high average binding affinity (ΔG_{Bind}) of -194.09 and − 177.13 KJ/mol, respectively, due to high contributions of van der Waals and electrostatic energies, followed by ARC-1034. Alendronate shows the lowest average binding free energy because of high contribution of polar solvation energy. The per residue energy contribution during 15 ns of MD simulation is shown in **Figures S8** and **S9** of supplementary material. The residues GLU 47, LEU 41, MET 49, CYC 145, MET 165, GLU 166, LEU 167, and PRO 168 have positive contributions, while residues ARG 45, PRO 39, SER 147, GLU 166, ASP 187, and HIS 164 negative contributions to the binding free energy of the protein-ligand complex.

100 ns Molecular Dynamics

MD simulations of the top 8 compounds Q27458218, Indinavir, ARC-1034, Monoxerutin, Minocycline, Miraziridine A, Doxycycline and Enalapril, which showed high average binding free energy (< -50 KJ/mol) during 15 ns MD runs, were extended to 100 ns. During 100 ns of MD simulations, it is observed that binding of compounds Q27458218, Doxycycline and Minocycline to the SARS-CoV-2 main protease is compact and stable throughout the simulation, as shown in Fig. 5. However, only Q27458218 occupies all the sub pockets of the substrate binding site throughout the simulation. The RMSD of the C-alpha atoms of the SARS-CoV-2 main protease remain stable during the simulation with all compounds except with Indinavir and Enalapril (Fig. 6A). Figure 6B shows the stable RMSD trajectories of these compounds. Monoxerutin is also able to bind in the binding site of the SARS-CoV-2 main protease but keeps readjusting. Similarly, Miraziridine A also tries to adjust in the binding site until 75 ns; thereafter it forms a stable complex with the protein. ARC-1034 gets displaced from the binding site after 20 ns to the edge of the site and is partially bound to it until 97 ns. Thereafter, ARC-1034 leaves the binding site completely and binds near it. Indinavir forms a stable interaction with the SARS-CoV-2 main protease until 50 ns of simulation; however, it leaves the binding site thereafter and again comes back to bind to the original site at 90 ns. Enalapril also leaves the binding site after 50 ns of simulation and binds to a different part of the protein as the simulation progresses.

The RMSF of protein backbone atoms during the 100 ns runs are depicted in Fig. 7. The loop region of the protein mentioned in the previous section shows high fluctuation. In the case of Indinavir, the fluctuation of the protein residues is highest, followed by Enalapril. The ligand binding site of the SARS-CoV-2 main protease is surrounded by loop regions which give some flexibility to it. Compounds which show stable binding cause less fluctuation in loop regions of the protein. Stability of compounds binding to the protein is also reflected in the number of hydrogen bonds formed by them during the simulation, as shown in Fig. 8. Compound Q27458218 consistently forms a large number of hydrogen bonds throughout the simulation. The number of hydrogen bonds formed by Miraziridine A increases after 75 ns of simulation time. For Indinavir, Enalapril, ARC-1034 and Monoxerutin, the number of hydrogen bonds formed decreases as the simulation progresses. Doxycycline and Minocycline, unlike other compounds, rarely form multiple hydrogen bonds during simulation. Residues which are involved in hydrogen bond formation with ligands during simulation are listed in Table 3. Hydrogen bonding between ligand and protein helps the ligand to bind in a specific orientation to initiate biochemical reaction. Amino acid residue GLU 166 seems to be important for ligand binding and forms a high percentage of the hydrogen bonds with every compound except Indinavir and Enalapril. Compound Monoxerutin forms hydrogen bonds with TYR 54, SER 144, ASP 187, ARG 188, THR 190 and GLN 192. Compound Q27458218 forms hydrogen bonds with multiple amino acid residues such as PHE 140, GLY 143 and HIS 164, signifying their important role in its stable binding to the SARS-CoV-2 main protease. Compounds ARC-1034 and Miraziridine A predominantly form hydrogen bonds with THR 26, HIS 41, SER 46 and GLN 189. Both Doxycycline and Minocycline predominantly form hydrogen bonds with HIS 41 and GLU 166, respectively.

The average binding free energies of all the ligands against the SARS-CoV-2 main protease are listed in Table 4. Compounds Q27458218, Doxycycline and Minocycline show increase in average ΔG_{Bind} to -202.61, -68.29 and − 70.21 KJ/mol, respectively as the simulation progresses after 15 ns of simulation time. Low standard deviation from average ΔG_{Bind} also signifies the strong and stable interaction between SARS-CoV-2 main protease and these three compounds. For the other compounds, not only is there a decrease in average ΔG_{Bind} but the standard deviation is also significant, suggesting that the interaction between these compounds and SARS main protease is not strong and stable. Both vdW and electrostatic interactions make important contributions to the free energy of binding of the ligand with the protein. The per residue energy contribution to the binding free energy is depicted in **Figure S10**. Residues contributing positively as well negatively to the binding free energy are as described in the previous section. However, the extent of contribution changes during 100 ns of simulation.

Replica Exchange Molecular Dynamics

To further ascertain the performance of MD simulations studies, REMD was performed on each system with four replicas. The overall replica exchange probability of each system during the simulation found to be 0.30. Q27458218 shows stable binding with SARS-CoV-2 main protease throughout the simulation in different replicas. The RMSD of C-alpha atoms of protein and ligand are depicted in **Figure S11** and **S12**, respectively. The average RMSD of C-alpha protein in all replicas of the three systems was observed to be between 2 to 3 Angstrom.

Table 3
Percentage hydrogen bond interaction (> 1%) shown by ligands during MD simulation with amino acid residues of SARS COV-2 main protease.

Compound	Percentage H-bond Occupancy																
	THR 26	HIS 41	GLU 47	SER 46	TYR 54	PHE 140	ASN 142	GLY 143	SER 144	CYS 145	HIS 163	HIS 164	GLU 166	ASP 187	ARG 188	GLN 189	TH 19
Monoxerutin	-	-	-	-	5.36	-	1.34	-	8.80	-	4.48	-	25.88	4.53	15.52	2.26	8.7
Q27458218	-	-	2.86	-	-	25.57	-	10.90	-	1.56	-	55.25	58.76	-	-	-	-
ARC-1034	43.21	2.18	26.38	5.79	-	-	1.64	-	-	-	-	-	21.02	-	-	2.08	-
Miraziridine A	8.53	2.10	-	1.87	-	-	-	4.32	12.44	-	-	-	23.38	-	-	11.41	-
Indinavir	3.47	-	-	-	5.80	-	-	-	-	-	-	1.40	-	-	-	-	-
Doxycycline		16.66		1.61	-	-	-	-	-	-	-	-	1.76	-	-	-	4.4
Minocycline	-	-	-	1.41	-	-	-	-	-	-	-	-	24.99	-	-	-	-
Enalapril	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.58	-

Table 4
Average binding free energy of ligand (In KJ/mol) obtained from 100 ns MD simulation.

Compound	Avg. van der Waals Energy (ΔG_{vdW})	Avg. Electrostatic energy (ΔG_{Elect})	Avg. Polar solvation energy (ΔG_{Polar})	Avg. SASA energy (ΔG_{SASA})	Avg. Binding free energy (ΔG_{Bind})
Monoxerutin	-143.57 +/- 19.91	-90.62 +/- 30.82	206.75 +/- 44.81	-16.97 +/- 1.95	-44.42 +/- 20.37
Q27458218	-227.97 +/- 22.96	-389.02 +/- 34.99	442.21 +/- 40.20	-27.83 +/- 2.30	-202.61 +/- 23.84
ARC-1034	-98.47 +/- 24.93	-272.95 +/- 57.74	302.22 +/- 83.84	-15.55 +/- 2.96	-84.75 +/- 43.80
Miraziridine A	-165.62 +/- 19.59	-49.91 +/-21.30	179.51 +/- 44.49	-20.36 +/- 2.18	-56.38 +/- 35.96
Indinavir	-70.56 +/- 58.24	-129.07 +/- 61.36	77.79 +/- 82.36	-9.53 +/- 7.76	-131.36 +/- 69.70
Doxycycline	-131.40 +/- 16.64	-22.92 +/- 10.24	101.25 +/- 17.89	-15.22 +/- 1.51	-68.29 +/- 15.42
Minocycline	-136.25 +/-18.93	-36.20 +/- 24.71	117.86 +/- 32.87	-15.62 +/- 1.59	-70.21 +/- 14.78
Enalapril	-85.80 +/- 38.14	-11.08 +/- 11.25	69.56 +/- 40.82	-11.39 +/- 4.51	-38.72 +/- 36.87

Table 5
Average binding free energy of **Q27458218** (In KJ/mol) obtained from REMD simulation.

Replica	Avg. van der Waals Energy (ΔG_{vdW})	Avg. Electrostatic energy (ΔG_{Elect})	Avg. Polar solvation energy (ΔG_{Polar})	Avg. SASA energy (ΔG_{SASA})	Avg. Binding free energy (ΔG_{Bind})
1	-159.61 +/- 21.72	-306.99 +/- 45.56	340.77 +/- 74.15	-19.71 +/- 2.15	-145.55 +/- 58.31
2	-145.42 +/- 35.62	-266.29 +/- 69.69	272.50 +/- 107.63	-19.73 +/- 4.44	-158.94 +/- 40.49
3	-179.15 +/- 24.12	-197.34 +/- 30.47	194.37 +/- 45.63	-20.89 +/- 2.23	-203.02 +/- 27.60
4	-171.66 +/- 19.13	-346.86 +/- 37.58	361.63 +/- 49.53	-21.98 +/- 2.10	-178.87 +/- 26.16

Being a peptide, the RMSD of Q27458218 in all replicas is higher during the simulation and ranges between 3 to 5 Angstrom. Throughout the simulation, ligands kept interacting with SARS-CoV-2 main protease through hydrogen bonding as depicted in **Figure S13**. Q27458218 shows hydrogen bonding with PHE 140, GLY 143, HIS 164 and GLU 166 in all replicas.

The average binding free energy of ligands for all replicas of each system calculated is listed in Table 5. Q27458218 showed the highest binding affinity towards SARS-CoV-2 in all replicas and is estimated to be about − 171 KJ/mol. In all simulations, van der Waals and electrostatic interaction energy contributed positively while polar solvation energy contributed negatively towards binding free energy of ligands. This REMD study suggests Q27458218 binds to SARS-CoV-2 main protease effectively in different conditions.

Out of the dataset of 1305 compound Q27458218 performed well in molecular simulation studies (Fig. 9) and based on this evidence, should be considered for chemoprophylaxis against SARS-CoV-2 by inhibiting its main protease enzyme subsequent to replication in the host. Compound Q27458218 is an experimental peptidomimetic designed as a human prostasin inhibitor (PDB ID: 3E16) [52]. It has shown the highest affinity towards the main protease of SARS-CoV-2 and can be considered for pre-clinical and clinical trials considering no established therapy or inhibitor is available for the main protease.

Conclusion

Networks of similar protein-ligand binding site surface properties were constructed using Protein Encoded Shape Distributions (PESD) and different distance metrics. The networks constructed from three different metrics helped to identify a diverse set of proteins having high similarity to the SARS-CoV-2 main protease. Molecular docking studies were carried out to narrow down the diverse set of compounds that bind to proteins similar to the SARS-CoV-2 main protease, to a few potential hits. The potential hit compounds identified from molecular docking studies were further tested for their ability to bind stably to the main protease of SARS-CoV-2 by performing molecular dynamics simulations. Although the compounds Doxycycline and Minocycline show stable binding with the SARS-CoV-2 main protease, they fail to occupy all the sub pockets. In contrasts Compound Q27458218 not only binds to SARS-CoV-2 main protease throughout the simulations but also occupies all the sub pockets of the enzyme with consistent affinity, which is validated by binding free energy calculations. Compound Q27458218 has been tested *in vitro* as a prostasin inhibitor; and can be clinically explore as a SARS-CoV-2 main protease inhibitor and a possible potential therapy for SARS-CoV-2 infection.

Declarations

ACKNOWLEDGMENTS

The authors acknowledge financial support from the Science and Engineering Research Board (SERB), India, under grant EMR/2016/002141. We also acknowledge helpful discussions with Dr. Michael Krein during the initial stages of this work. Protein Binding Sitesimilarity calculations were performed on the PESD server at http://reccr.chem.rpi.edu/.

Funding

This work was supportedby the Science and Engineering Research Board (SERB), India, under grant EMR/2016/002141.

Conflict of Interest

None

Code availability

All software and code used were available on the internet, as cited in the references.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Sagar Bhayye. The first draft of the manuscript was written by Sagar Bhayye. All authors read and approved the final manuscript.

Data Availability

All data are provided as Supplementary files:

References

1. Susan Payne (2017) Family Coronaviridae. In: Viruses From Understanding to Investigation. pp 149–158
2. Spinelli a., Pellino G (2020) COVID-19 pandemic: perspectives on an unfolding crisis. Br J Surg 107:785–787. https://doi.org/10.1002/bjs.11627
3. Li G, Fan Y, Lai Y, et al (2020) Coronavirus infections and immune responses. J Med Virol 92:424–432. https://doi.org/10.1002/jmv.25685
4. Daga MK, Kumar N, Aarthi J, et al (2019) From SARS-CoV to Coronavirus Disease 2019 (COVID-19)-A Brief Review. J Adv Res Med (E-ISSN 2349-7181 P-ISSN 2394-7047) 6:1–9
5. McIntosh K, Perlman S (2015) Coronaviruses, including severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS). Mand Douglas, Bennett's Princ Pract Infect Dis 1928
6. Yang Y, Peng F, Wang R, et al (2020) The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. J Autoimmun 109:102434. https://doi.org/doi.org/10.1016/j.jaut.2020.102434
7. Cascella M, Rajnik M, Cuomo A, et al (2020) Features, evaluation and treatment coronavirus (COVID-19). In: Statpearls [internet]. StatPearls Publishing
8. Senger MR, Evangelista TCS, Dantas RF, et al (2020) COVID-19: molecular targets, drug repurposing and new avenues for drug discovery
9. Arun KG, Sharanya CS, Abhithaj J, et al (2020) Drug repurposing against SARS-CoV-2 using E-pharmacophore based virtual screening, molecular docking and molecular dynamics with main protease as the target. J Biomol Struct Dyn 1–12
10. Elfiky AA (2020) SARS-CoV-2 RNA dependent RNA polymerase (RdRp) targeting: An in silico perspective. J Biomol Struct Dyn 1–9
11. Gahlawat A, Kumar N, Kumar R, et al (2020) Structure-Based Virtual Screening to Discover Potential Lead Molecules for the SARS-CoV-2 Main Protease. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.0c00546
12. Gentile F, Cherkasov A (2020) Rapid Identification of Potential Inhibitors of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds Computational drug discovery of DNA repair inhibitors View project Molecular modeling of small molecule interactions with the TLR7 receptor V. JidcOrg. https://doi.org/10.26434/chemrxiv.11860077.v1
13. Pandey P, Rane JS, Chatterjee A, et al (2020) Targeting SARS-CoV-2 spike protein of COVID-19 with naturally occurring phytochemicals: an in silico study for drug development. J Biomol Struct Dyn 1–11
14. Liu W, Morse JS, Lalonde T, Xu S (2020) Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. Chembiochem 21:730–738
15. Goyal B, Goyal D (2020) Targeting the dimerization of main protease of coronaviruses: A potential broad-spectrum therapeutic strategy. ACS Comb Sci 22:297–305
16. Suárez D, Díaz N (2020) SARS-CoV-2 Main Protease: A Molecular Dynamics Study. J Chem Inf Model 60:5815–5831. https://doi.org/10.1021/acs.jcim.0c00575
17. Das S, Kokardekar A and, Breneman CM (2008) Property-Encoded Shape Distributions (PESD) for Binding Site Comparison in Proteins. Struct Bioinforma 2–14
18. Das S, Kokardekar A, Breneman CM (2009) Rapid Comparison of Protein Binding Site Surfaces with Property Encoded Shape. 2863–2872
19. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612
20. Maier JA, Martinez C, Kasavajhala K, et al (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J Chem Theory Comput 11:3696–3713
21. Das S, Krein MP, Breneman CM (2010) PESDserv: a server for high-throughput comparison of protein binding site surfaces. Bioinformatics 26:1913–1914. https://doi.org/https://doi.org/10.1093/bioinformatics/btq288
22. PESDserv. Rensselaer Exploratory Center for Cheminformatics Research. http://reccr.chem.rpi.edu/Software/pesdserv/23. Dokmanic I, Parhizkar R, Ranieri J, Vetterli M (2015) Euclidean distance matrices: essential theory, algorithms, and applications. IEEE Signal Process Mag 32:12–30
23. Craw S (2017) Manhattan distance. Encycl Mach Learn Data Min 790–791
24. Emran SM, Ye N (2002) Robustness of Chi-square and Canberra distance metrics for computer intrusion detection. Qual Reliab Eng Int 18:19–28

25. Assenov Y, Ramírez F, Schelhorn S-E, et al (2007) Computing topological parameters of biological networks. Bioinformatics 24:282–284
26. Kim S, Thiessen PA, Bolton EE, et al (2016) PubChem substance and compound databases. Nucleic Acids Res 44:D1202–D1213
27. Berman H, Henrick K, Nakamura H (2003) Announcing the worldwide protein data bank. Nat Struct Mol Biol 10:980
28. Schrödinger LLC (2016) Maestro
29. Schrödinger LLC (2016) Protein Preparation Wizard, Epik
30. Schrödinger LLC (2016) Ligprep
31. Friesner RA, Murphy RB, Repasky MP, et al (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem 49:6177–6196. https://doi.org/10.1021/jm051256o
32. Abraham MJ, Murtola T, Schulz R, et al (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2:19–25. https://doi.org/10.1016/j.softx.2015.06.001
33. Vanommeslaeghe K, Hatcher E, Acharya C, et al (2010) CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 31:671–690
34. Vanommeslaeghe K, MacKerell Jr AD (2012) Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. J Chem Inf Model 52:3144–3154
35. Mark P, Nilsson L (2001) Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J Phys Chem A 105:9954–9960
36. Fliege J, Svaiter BF (2000) Steepest descent methods for multicriteria optimization. Math Methods Oper Res 51:479–494
37. Essmann U, Perera L, Berkowitz ML, et al (1995) A smooth particle mesh Ewald method. J Chem Phys 103:8577–8593. https://doi.org/10.1063/1.470117
38. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
39. Genheden S, Ryde U (2015) The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 10:449–461. https://doi.org/10.1517/17460441.2015.1032936
40. Kumari R, Kumar R, Lynn A (2014) G-mmpbsa -A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 54:1951–1962. https://doi.org/10.1021/ci500020m
41. Erdős P, Rényi A (1964) On the strength of connectedness of a random graph. Acta Math Acad Sci Hungarica 12:261–267
42. Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–60
43. Fiedler M (1973) Algebraic connectivity of graphs. Czechoslov Math J 23:298–305
44. Fiedler M (1989) Laplacian of graphs and algebraic connectivity. Banach Cent Publ 25:57–70
45. Zou L-W, Jin Q, Wang D-D, et al (2018) Carboxylesterase inhibitors: an update. Curr Med Chem 25:1627–1649. https://doi.org/10.2174/0929867325666171204155558
46. Jo S, Kim H, Kim S, et al (2019) Characteristics of flavonoids as potent MERS-CoV 3C-like protease inhibitors. Chem Biol Drug Des 94:2023–2030
47. Arshad MS, Khan U, Sadiq A, et al (2020) Coronavirus disease (COVID-19) and immunity booster green foods: A mini review. Food Sci Nutr 8:3971–3976
48. Abad MJ, Bedoya LM, Apaza L, Bermejo P (2012) Anti-infective flavonoids: an overview. In: Bioactive Natural Products: Opportunities and Challenges in Medicinal Chemistry. World Scientific, pp 443–473
49. Zhu L, George S, Schmidt MF, et al (2011) Peptide aldehyde inhibitors challenge the substrate specificity of the SARS-coronavirus main protease. Antiviral Res 92:204–212. https://doi.org/10.1016/j.antiviral.2011.08.001
50. Arshad MS, Khan U, Sadiq A, et al (2020) Coronavirus disease (COVID-19) and immunity booster green foods: A mini review. Food Sci Nutr 8:3971–3976. https://doi.org/10.1002/fsn3.1719
51. Tully DC, Vidal A, Chatterjee AK, et al (2008) Discovery of inhibitors of the channel-activating protease prostasin (CAP1/PRSS8) utilizing structure-based design. Bioorg Med Chem Lett 18:5895–5899. https://doi.org/10.1016/j.bmcl.2008.08.029
52. Pflug A, Rogozina J, Lavogina D, et al (2010) Diversity of bisubstrate binding modes of adenosine analogue–oligoarginine conjugates in protein kinase a and implications for protein substrate interactions. J Mol Biol 403:66–77
53. Kos J, Mitrović A, Mirković B (2014) The current stage of cathepsin B inhibitors as potential anticancer agents. Future Med Chem 6:1355–1371. https://doi.org/10.4155/fmc.14.73
54. Cochrane RVK, Norquay AK, Vederas JC (2016) Natural products and their derivatives as tRNA synthetase inhibitors and antimicrobial agents. Medchemcomm 7:1535–1545
55. Lv Z, Chu Y, Wang Y (2015) HIV protease inhibitors: A review of molecular selectivity and toxicity. HIV/AIDS - Res Palliat Care 7:95–104. https://doi.org/10.2147/HIV.S79956
56. Cochrane RVK, Norquay AK, Vederas JC (2016) Natural products and their derivatives as tRNA synthetase inhibitors and antimicrobial agents. Medchemcomm 7:1535–1545. https://doi.org/10.1039/C6MD00274A
57. Mondal S, Adhikari N, Banerjee S, et al (2020) Matrix metalloproteinase-9 (MMP-9) and its inhibitors in cancer: A minireview. Eur J Med Chem 112260
58. Camodeca C, Nuti E, Tepshi L, et al (2016) Discovery of a new selective inhibitor of A Disintegrin And Metalloprotease 10 (ADAM-10) able to reduce the shedding of NKG2D ligands in Hodgkin's lymphoma cell models. Eur J Med Chem 111:193–201
59. Zipfel P, Rochais C, Baranger K, et al (2020) Matrix Metalloproteinases as New Targets in Alzheimer's Disease: Opportunities and Challenges. J Med Chem 0.: https://doi.org/10.1021/acs.jmedchem.0c00352

Figures

Figure 1

Structure of SARS-CoV-2 main proteases (Mpro). It consists of three distinct domains. Substrate binding site present between domain I and II and contains HIS 41 (yellow) and cysteine 145 (cyan) catalytic dyad.

Figure 2

Constructed networks based on **(A)** Euclidean **(B)** Chi-Square and **(C)** Manhattan distances. SARS-CoV-2 main protease is shown in red and proteins showing high similarity to it are shown in blue color.

Figure 3

Binding site similarity networks of the crystal structures of SARS-CoV-2 main protease (Red) and proteins (Blue) having similar binding sites. (A) Euclidean (B) Chi-Square and (C) Manhattan distances network.

Figure 4

Hit molecules selected from molecular docking studies.

Figure 5

Movement of compounds in the binding site of the SARS-CoV-2 main protease during 100 ns of MD simulation. Different atoms such as carbon (cyan), nitrogen (blue), oxygen (red) and hydrogen (white) of the ligand are shown as vdW spheres.

Figure 6

Root mean square deviation of **(A)** Protein C-alpha atoms and **(B)** Ligands during 100 ns MD simulation.

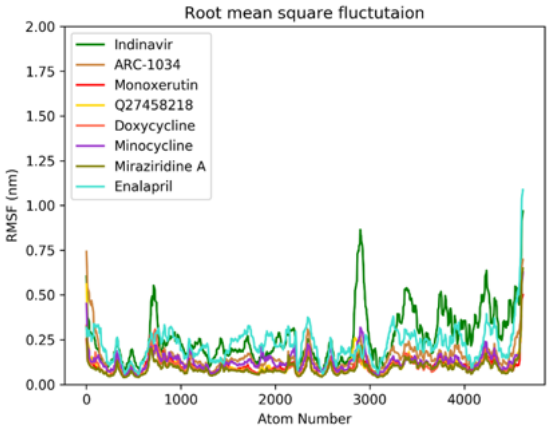

Figure 7

Root mean square fluctuation of protein backbone atoms during 100 ns of MD simulation when complexed with different compounds.

Figure 8

Number of hydrogen bonds formed by ligands with the protein during MD simulation.

Figure 9

Superimposed view of SARS-CoV-2 main protease before (red) and after (blue) 100 ns of MD simulation complexed with Q27458218. Green and yellow colored carbon models of the ligand represent the position of the ligand before and after MD simulation, respectively.

Supplementary Files

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

- ARC1034video.mp4
- Doxycyclinevideo.mp4
- Enalaprilvideo.mp4
- Indinavirvideo.mp4
- Minocyclinevideo.mp4
- MiraziridineAvideo.mp4
- Monoxerutinvideo.mp4
- Q27458218video.mp4
- SupplementaryFile21122.docx
- SuppplementryFileTableS3.xlsx