

**Computational search for potential COVID-19 drugs from FDA-approved drugs and small molecules
of natural origin identifies several anti-virals and plant products**

Abhishek Sharma⁺, Vikas Tiwari⁺ and Ramanathan Sowdhamini^{*}

National Centre for Biological Sciences

GKVK Campus

Bellary Road

Bangalore 560065

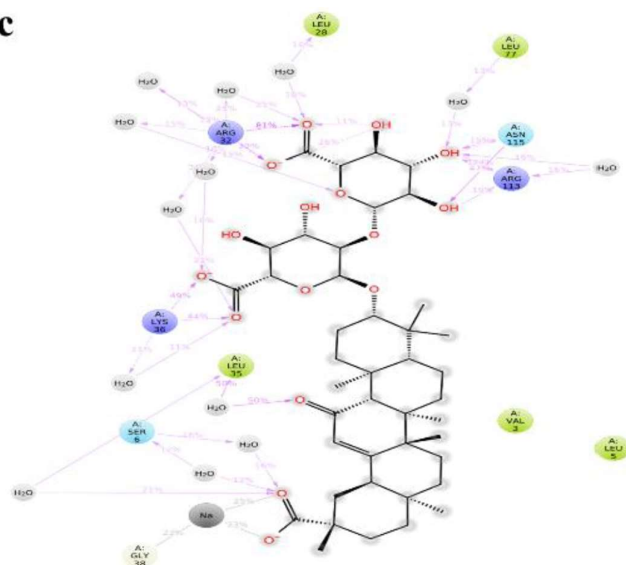
INDIA

*mini@ncbs.res.in

⁺ Joint first authors

Abstract

C



The world is facing COVID-19 pandemic at the present time, for which mild symptoms include fever and dry cough. In severe cases, it could lead to pneumonia and ultimately death in some instances. Moreover, the causative pathogen is highly contagious and there are no drugs or vaccines for it yet. The pathogen, SARS-CoV-2, is one of the human coronaviruses which was identified to infect humans first in December 2019. SARS-CoV-2 shares evolutionary relationship to other highly pathogenic viruses such as Severe Acute Respiratory Syndrome (*SARS*) and Middle East

respiratory syndrome (*MERS*). We have exploited this similarity to model a target non-structural protein, NSP1, since it is implicated in regulation of host gene expression by the virus and hijacking of host machinery. We next interrogated the capacity to repurpose around 2300 FDA-approved drugs and more than 3,00,000 small molecules of natural origin towards drug identification through virtual screening and molecular dynamics. Interestingly, we observed simple molecules like lactose, previously known anti-virals and few secondary metabolites of plants as promising hits. Disclaimer: we would **not like to recommend uptake of these small molecules** for suspect COVID patients **until it is approved by competent national or international authorities**.

Introduction

Coronavirus (CoV) belongs to the family *Coronaviridae* and the order *Nidovirales* (sharing with *Arteriviridae* and *Roniviridae*). Coronaviruses are enveloped, long positive sense single-stranded RNA viruses, which are best known for causing mild to severe respiratory and enteric infection among vast range of hosts [1]. These are further divided into 4 groups/genera named as Alphacoronavirus, Betacoronavirus, Gammacoronavirus and Deltacoronavirus (α -CoV, β -CoV, γ -CoV and δ -CoV), based on sequence similarities and antigenic cross-reactivity. Human-CoV belongs to group I and group II of Betacoronavirus. HCoV-OC43, HCoV-229E, SARS and MERS are some examples of Human-CoV, out of which SARS and MERS are highly pathogenic in nature [1][2]. Recently, a new pathogenic Human-CoV strain known as SARS-CoV2, spreading the COVID-19 infection, has emerged by December 2019 with Wuhan of Hubei province in China as the epicentre [3]. Origin of this virus is still under investigation but has been speculated as zoonotic shift from bat to human. Outbreak of COVID-19 has spread across the globe and has taken the shape of a pandemic (Novel Coronavirus

(2019-nCoV) situation reports - World Health Organization). Italy, America, China, Spain, France and Iran are among worst-hit countries. As of 3 April 2020, this has infected 1,079,978 individuals and has caused more than 58,110 fatalities across the globe (Johns Hopkins Coronavirus Resource Center; Worldmeter. <https://www.worldometers.info/coronavirus/>; Novel Coronavirus (2019-nCoV) situation reports - World Health Organization). At present, there are no drugs or vaccine available against this and patients are treated according to symptoms shown by them. Remdisivir (drug originally designed to treat Ebola), Chloroquine and Hydroxychloroquine (an antimalarial drug), Kevsara (an arthritis drug) and few other anti-viral drugs are being considered for treatment. But they do not directly make use of the virome of SARS-CoV2 [4]

(<https://www.nasdaq.com/articles/8-experimental-coronavirus-treatments-to-watch-2020-03-31>).

First genome of COVID-19 strain was sequenced by Wu *et. Al* [3] from 41-year-old man and was found to be closely similar to SARS-CoV. The structural component of virus consists of 4 proteins: Spike (S), Membrane (M), Envelope (E) and Nucleocapsid (N) protein respectively. S protein is critical for viral infection as it enable host-pathogen interaction and mediated viral entry into host cell. M protein is a multipass transmembrane protein, major constituent of virion envelope and known to provide its shape. E protein, unlike the name suggest it is minor constituent of envelope and 80-120 aa in length. N protein as the name suggests forms helical nucleocapsid of virion [1].

The 5'-end of the genome encode two open reading frames - ORF1a and ORF1b respectively, which code for all non-structural proteins (nsp1-16) [1][2]. These proteins are essential for viral replication as well as infection, whereas function of some is yet to be identified. Among these non-structural proteins within CoV family, some are conserved in sequence whereas some are highly diverged in nature. Nsp1 (non-structure protein 1) is one of such diverged proteins which is encoded by ORF1a and varied in amino-acid length among CoV-groups [2]. COVID-19 Nsp1 consists of 180 and shows sequence similarity with SARS protein [3][5]. In spite of differences in amino-acid sequence and length, it has shown to be functionally conserved [2][6]. SARS-CoV Nsp1 is most well studied among coronavirus family. Nsp1 has shown to act as virulence factor [2][7][8] and mutation in this protein results in production of attenuated virus *in vitro* and *in vivo* [7]. Nsp1 deploys two strategies to inhibit host cell expression a). Inhibition of host translation b). Induction of host mRNA degradation. It inhibits host translation by forming complex with 40s ribosome subunit which prevent formation of active polysome. Complex formation with 40s subunit also shown to inhibit its translational ability [9]. It further effects host cell gene expression by inducing host mRNA degradation in template-specific manner. Term template-specific does not imply its association with protein sequence but relates to ability to specifically degrade capped host mRNA [8-12] as compared to SARS-mRNA [10]. mRNA is hypothesized to be cleaved by unknown host endonuclease as Nsp1 does not possess any endonucleolytic activity. Other than these, Nsp1 shown to cause chemokine dysregulation which correlates with high inflammation in severe

patients [13-15]. It suppresses innate immune response by degrading IFN-beta mRNA [12] and affecting anti-viral signaling [16]. Yeast-two hybrid assays have shown Nsp1 to interact with multiple host proteins [17].

We have chosen COVID-19 Nsp1 as our target and hypothesized that its inhibition, can potentially attenuate the virus and suppress adverse immune-pathology caused by it. We have generated a homology model of the Nsp1 and used this model to carry out virtual screening to identify potential inhibitors and lead compounds. Our searches are directed within database of FDA approved drugs (DRUGBANK) and those which are derived from molecules of natural origin (SUPERNATURALDB). Finally, we performed MD simulation to ensure that there are indeed stable protein-ligand interactions when the system is visualized to undergo limited conformational freedom. We find several anti-viral compounds, few secondary metabolites of plant origin and simple compounds (like lactose) to retain high potential to act as NSP1 inhibitors.

Methods:

- 1. Sequence retrieval and analysis:** Full repository of COVID-19 protein sequences was downloaded from NCBI [18][19]. Wuhan-Hu-1 strain [Accession number: NC_045512] was among the first to be sequenced from Wuhan of Hubei province. Hence this is considered as the 'reference genome' in this study. Nsp1 protein sequences were extracted, incomplete sequences were removed and curated sequences were then passed to SNP analyzer (utility in ViPR Database) to understand variation among the Nsp1 sequences [20]. Similar analysis had been done for the only two samples taken from patients within India at present. [Accession number: QHS34545.1 and QIA98582.1]. A set of key residues, important in suppressing host gene expression and anti-viral signaling were identified. Mutation study done by Jauregui et. al [16] used as a reference for this.
- 2. Homology Modelling:** Nsp1 protein sequence [Accession number: YP_009725297] was retrieved from NCBI for Homology modelling. Blastp [21] was used to search for nearest structural homologue in Protein Data Bank (PDB) [22] to serve as template for modelling. Segments of Nsp1 sequence, where the association with the template was unknown, were removed. Modeller 9.12 [23] was used for Homology modelling. Homology models were first filtered according to DOPE score. Top 3 predicted models were then subjected to structure validations (by using SAVES 5.0 [24] (<https://servicesn.mbi.ucla.edu/SAVES/>) and ProSA server [25]). Based on DOPE score, Ramachandran plot and ProSA profile, the best predicted model was selected for virtual screening.
- 3. Virtual Screening of inhibitors:** FDA approved drugs and Super Natural II database (database of natural products) were used for docking purpose. FDA approved drugs were downloaded in SDF format (Standard Delay Format) from Drug-bank [26] whereas supernatural compounds were taken from supernatural database [27].

- a. Ligand and protein preparation: Downloaded compounds were prepared for screening using Ligprep module in Schrodinger (Schrödinger Release 2019-4: LigPrep, Schrödinger, LLC, New York, NY, 2019). For FDA approved drugs OPLS3e force field, targeted pH 7.4 +/-0.0, retain specified chiralities and 1 structure per ligand was specified during ligand preparation. For supernatural database we had specified pH range from 6.0 to 8.0 with maximum 32 structure per ligand. This was performed to scan and produce broad chemical and structural diversity from each molecule.
- b. Protein was prepared for docking by using Protein-preparation wizard [28] in Maestro Schrodinger (Schrödinger Release 2019-4: Maestro, Schrödinger, LLC, New York, NY, 2019.).
- c. Docking site prediction: SiteMap [29][30] was used to predict the potential drug-able deep and shallow sites on target protein. Site with high S-score, as well as D-score, was selected for ligand docking.
- d. Receptor grid Generation: Receptor-grid around docking region on protein was generated using receptor-grid generation module in Glide, Residues from top predicted deep and shallow site were specified and rotatable bonds across the site (if any) were checked during grid generation.
- e. Protein-Ligand docking: Using glide docking module, library of prepared ligands and protein with prepared receptor binding grid were docked. First, High-throughput virtual screening (HTVS) was performed. This narrowed down the list of potential ligands and top 10 percent from this were then screened with Standard Precision (SP) mode. Finally, 10 percent of hits obtained from SP were passed to Extra precision (XP). Selection of top 10 percent compounds were done based on top dock score and binding energy [31]-33] (Schrödinger Release 2019-4: Glide, Schrödinger, LLC, New York, NY, 2019.)
- f. Binding energy was calculated using MM-GBSA (Molecular Mechanics energies combined with Generalized Born and Surface Area continuum solvation) tool of Schrodinger.

4. **MD simulation:** The conformer of protein-ligand complex, emerging from XP docking, was assembled using system builder and subject to Molecular Dynamics using Desmond package of Schrodinger [34]. For water, TIP4P model was specified and orthorhombic box shape was used having buffer distance of 10 Å. Box volume was minimized. The system was neutralized and 150 mM salt (NaCl) was added. The output of system builder was used for MD. The default relaxation protocol was used to relax the solvated system followed by production MD run for 20 nanoseconds (ns). The relaxation protocol involves energy minimization steps using steepest descent method with maximum 2000 steps. The energy minimization was done with solute being restrained using 50 kcal/mol/Å force constant on all solute atoms and without restraints. Energy minimization was followed by short MD simulation steps which involve 1) Simulation for 12 picoseconds at 10K in NVT ensemble using Berendsen thermostat with restrained non-hydrogen

solute atoms 2) Simulation for 12 picoseconds at 10K and 1 atmospheric pressure in NPT ensemble using Berendsen thermostat and Berendsen barostat with restrained non-hydrogen solute atoms 3) Simulation for 24 picoseconds at 300K and 1 atmospheric pressure in NPT ensemble using Berendsen thermostat and Berendsen barostat with restrained non-hydrogen solute atoms 4) Simulation for 24 picoseconds at 300K and 1 atmospheric pressure in NPT ensemble using Berendsen thermostat and Berendsen barostat without restraints. After relaxation, production MD was run in NPT ensemble using OPLS 2003 force field [35]. For simulations, default parameters of RESPA integrator [36] (2 femtoseconds time step for bonded and near non-bonded interactions while 6 femtoseconds for far non-bonded interactions) were used. The temperature and pressure were kept at 300K and 1 bar using Nose-Hoover chain method [37] and Martyna-Tobias-Klein method [38] respectively. The production MD was run for 20 nanoseconds.

5. Simulation analysis: MD simulation analysis was performed using Simulation interaction diagram (SID) module of Desmond package. The entire range of simulation time was considered for all analyses. RMSD is calculated for each frame by aligning the structure of that frame with the reference frame structure (structure at the start of the simulation). Lig fit protein RMSD indicates the movement of ligand with respect to protein. It measures ligand heavy atom RMSD after aligning the complex to protein backbone of the reference frame. Significantly higher values of "Lig fit Prot" than protein RMSD signifies the diffusion of ligand away from its initial binding site. Lig fit lig RMSD is calculated by aligning the ligand on the reference ligand conformation and it indicates the internal fluctuation of ligand. Along with RMSD, the RMSF (Root Mean Square Fluctuation) was also assessed for each MD run. Protein RMSF shows the fluctuation of protein residues, highlights secondary structure (Pink: α helix; Blue: β strand) and ligand interacting residues marked by green vertical lines. Protein-ligand interactions were also monitored throughout the simulation time. Different type of protein-ligand interactions measured. are H-bond, Hydrophobic interaction, ionic interaction and water bridges. Hydrophobic interaction also includes π -Cation and π - π interactions. The normalized stacked bar charts suggest the fraction of simulation time for which an interaction is maintained over the course of simulation trajectory: for example, a value of 0.6 implies that a specific interaction is maintained for 60% of the simulation time. If a protein residue makes multiple interaction of the same type with ligand then values more than 1.0 are possible.

Results:

Sequence Analysis identified COVID-19 Nsp1 to be conserved:

60 nsp1 sequences of SARS-CoV2 were available in the public domain (NCBI as of 3 March 2020) and downloaded. Out of these, 28 were sequences deposited from China, 19 from USA and two from India. However, NSP1 displays only two amino acid changes within this dataset (D75E and A117T; data not shown). 100% sequence identity is noticed between the Wuhan-Hu-1 Chinese sample and the two Indian samples within NSP1.

As COVID-19 Nsp1 has not been studied, we took SARS Nsp1 as a template for modeling. Major reason for this assumption is 100% query coverage and 84.44% sequence similarity of COVID-19 Nsp1 with SARS Nsp1 (**Figure 1**). We performed extensive literature survey to identify a set of key residues, important in suppressing host gene expression and anti-viral signaling which are shown in **Figure 1**. Most of residues among this set are found to be conserved between COVID and SARS Nsp1.

Model for virtual screening generated by Homology Modelling:

Blastp search of COVID-19 Nsp1 sequence with PDB database enabled to identify 2hsx as best template (with 68% query coverage and 86% identity). N- and C-terminal overhangs in SARS CoV-2 NSP1 have not been considered for modelling. Amino acid variations and key residues, important for function, are marked on the alignment (**Figure 1**). Predicted models, derived using Modeller9.22, were sorted according to DOPE score and top three models were validated using ProSA and SAVES5.0 sever. The best model from the above was chosen for virtual screening (**Figure 2a**).

List of potent inhibitors are identified by In-silico screening of FDA approved drugs and Supernatural Database compounds:

Three deep and five shallow ligand binding sites could be recognized on the surface of COVID19—Nsp1 protein. Sites were ranked according to their ability to bind various ligands which was depicted as SITEMAP site score and D-Score (please see Methods). We selected Site 1 with a site score of 0.927 and D-score 0.791 among deep sites and site 3 with site score 0.883 and D-score 1.012 among shallow sites for ligand docking (**Figure 2a**). These sites also contain functionally important residues (Figure 1) which showed their biological importance. Selected sites were then used to generate a receptor grid for molecular docking. Molecular docking for each site was carried out using glide dock program with generated libraries of 2413 FDA approved drugs and 3,25,287 natural compounds, respectively. The top hits from FDA approved drug library were ranked according to their XP and MMGBSA scores. We have also considered ligands with well-known anti-viral and anti-inflammatory properties, along with topranked ones (entries 15-17 in **Table 1**). Final list of compounds was taken further for MD simulation run (Table 1). The top hits from Supernatural Database compounds were ranked according to their MMGBSA score and were further selected for MD simulation runs. List of top hits, selected on the basis of either binding energy or mode of action, for both deep and shallow binding sites are shown in Table 1.

MD simulation of protein-ligand complexes:

Best compounds from docking analysis were further subjected to 20 ns of MD simulation to assess the stability of protein-ligand complex. The interactions between protein and ligand were designated as stable, if there less structural variations and high percentage of hydrogen bonds or Hydrophobic interactions with various residues of the protein at the docked site throughout the course of simulations. Among the FDA approved drugs docked at deep site, Esculin is an example of stable complexes, while Zinc-gluconate is an example of unstable complex. **Figure 3a** and **Figure S1a** show the interaction of Esculin with Nsp1 in the docking pose, where Esculin interacts mainly with Arg62, Ser63, Ala68 and His72. MD simulation of Nsp1-deep-Esculin complex for 20 ns revealed the stability of the complex as assessed by the RMSD (root mean square deviation) plot. Residues in the secondary structure are expected to have less fluctuations than residues in the loop regions and the trend is followed for Nsp1 which shows high RMSF between residues 62-76 which form a loop and also interact with Esculin (**Figure S1b**). Arg62, Ser63, Ala68 and His72 (major interacting residues in the docking pose) interact mainly through H-bond interactions with Esculin. Met74 was also found to interact with Esculin mainly through H-bond (**Figure 3b**). Few other residues interact with Esculin, but with less amount of simulation time. Further details of these interactions are provided in **Figures S1c and S1d**.

Results of similar detailed analysis for all the ligands, as in Table 1, are provided in Supplementary Figures S1-S35. Other promising lead compounds, among the FDA approved ligands docked at deep site of Nsp1, are Cidofovir (**Figure S2**), Remdesivir (drug under investigatory group; **Figure S16**), Brivudine (**Figure S15**) and Edoxudine (**Figure S13**). In case of FDA approved drugs docked at shallow site, acarbose was found to be the most stable ligand. It interacts mainly with Arg32, Leu77 and Asn115 through H-bond and water bridge interactions (**Figure S30**).

Amongst compounds from SuperNatural database docked at deep binding site of Nsp1, SN00003849 interacts mainly with Arg62, Arg66, Ala68, Gly71, His72 and Met74 (**Figure 4a**). Further, the Nsp1-SN00003849 complex was found to be stable, as suggested by RMSD plot of 20ns MD simulation (**Figure S24a**). Residues interacting with SN00003849 are similar to that of Esculin (**Figure S24b**). These include Arg62, Arg66, Gly71, His72 and Met74 interacting mainly through H-bond and water bridge interactions (**Figure 4b**). Arg61, Gly71 and Met74 interact with the same atom of SN00003849 for more than 80% of simulation time (**Figure S24c**). At any point during the simulation minimum number of contacts between SN00003849 and Nsp1 is more than four, suggesting the strong interaction at the binding site (**Figure S24d**). SN00003849 also has the highest binding energy as per MM-GBSA calculation (**Table 1**). SN00003849 is structurally similar to terpene/steroid and can be classified as proto and pseudo alkaloids. SN00003832 and SN00216190 also form stable complex with Nsp1 at deep site (**Figure S28 and S29**). For shallow binding site, none of the SuperNatural compounds form complex that are as stable as that for deep binding site. Natural compounds (entries 18, 19, 31 and 32) are derived from herbal plants well-known to treat coughs and viral fevers. Another

FDA approved compound is Glycyrrhizic acid which was ranked bit lower for deep site, as well as shallow binding site during docking. The MD simulation was run for deep as well as shallow site complex of Nsp1 with Glycyrrhizic acid. Glycyrrhizic acid bound at the shallow site interacts mainly with Arg32, Lys36, Arg113 and Asn115 in the docked pose (**Figure 5a**). The MD simulation of 20 ns suggested that the complex is stable as per RMSD plot (**Figure S23a**). Nsp1-rmsf plot indicates that the few residues of α helix along with residues at the N and C-termini are also involved in interaction with Glycyrrhizic acid (**Figure S23b**). Major interacting residues of Nsp1 are same as that in the docking pose (**Figure 5b**). Atom wise interactions of Glycyrrhizic acid with Nsp1 has been shown in **Figure 23c**. Similar to SN00003849, Glycyrrhizic acid also maintains at least 4 contacts with Nsp1 over the entire course of simulation time (**Figure 23d**). Glycyrrhizic acid bound at the deep site is not stable (**Figure S22**). Interestingly, Glycyrrhizic acid is from the plant Mulethi or Liquorice (also referred as Yashtimadhura (*Glycyrrhiza glabra*), which is a natural herb for cough and has expectorant properties. It can also reduce infection of the upper respiratory tract. It reduces throat irritation and helps cases of a chronic cough.

Discussion:

COVID-19 outbreak has turned into a pandemic, which makes identification of new target molecule, repurposing of drugs and designing vaccine an imminent necessity. Since the outbreak, many studies have been conducted along these lines [39][46-48]. We used Nsp1 protein as our target protein. It shows 86 % identity with SARS Nsp1. A model of COVID-19 Nsp1 was made using SARS Nsp1 as a template. Virtual screening, against Nsp1 protein suggests a list of FDA-approved drugs and natural compounds against deep and shallow binding site on Nsp1. Deep and shallow binding site include functionally important residues such as H81, H83, R124 and R43, K47, E91, R124, K125 respectively [16]. R124 has shown to be important for Nsp1 to interact with viral mRNA 5'-UTR region which prevent viral mRNA from Nsp1 mediated mRNA degradation [10] (Note: Residue number in modelled structure starts with 12th residue of the sequence). Docking and MMGBSA scores suggest the binding potential of these compounds towards Nsp1. Further, MD simulation of the selected compounds in complex with Nsp1 ensures that some of these hits form stable interactions with Nsp1.

Esculin, Cidofovir, Edoxudine, Brivudine and Remdesivir were found to form stable complex with Nsp1, among FDA approved drugs binding at deep site of Nsp1. Esculin is a glucoside and naturally occur in barley, horse chestnut etc. It is given to improve capillary permeability and fragility and has been reported to inhibit collagenase and hyaluronidase enzymes. This molecule has been shown to have antioxidant and anti-inflammatory activity [26]. This suggests the ability of esculin to not only inhibit Nsp1 activity but also being effective against secondary symptoms such as inflammation. Cidofovir is a known anti-viral agent against CMV infection and acts via inhibition of CMV DNA Polymerase. Edoxudine is a deoxy-thymidine analog

shown to be effective against herpes simplex virus type 1 and type 2. It acts as competitive inhibitor of viral DNA polymerase in its phosphorylated form. Edoxudine is initially phosphorylated by viral thymidine kinase and it is specifically incorporated in viral DNA. Edoxudine has been discontinued. Brivudine is an organic compound and a pyrimidine 2'-deoxyribonucleosides analog. This is used in treatment of herpes zoster, results from reactivation of varicella-zoster virus. Remdesivir is proposed as a potential anti-viral drug against Ebola [26]. However, this molecule appears within the Investigational group of DRUGBANK. It is an adenosine-triphosphate analog and has shown effectivity against coronaviruses. Recent publication on COVID-19 treatment, shows it to be a potential drug along with chloroquinone [4]. Remdesivir is an RNA polymerase inhibitor. Hence our study suggests an additional mechanism of action for this drug. An interesting and unexpected molecule among this list is lactose. Lactose is a disaccharide of glucose and galactose and used as nutrient supplement. Derivatives of lactose, 3'-sialyllactose have been shown to have broad-spectrum neutralization activity against avian influenza viruses in chickens [40]. Further investigation is necessary to check the anti-viral property of lactose against coronavirus.

Acarbose, Iopromide and Glycyrrhizic Acid form stable interactions with shallow binding site of Nsp1. Acarbose is an alpha glucosidase inhibitor and administered to patients with non-insulin dependent diabetes mellitus [26]. As death rate among COVID-19 patients with diabetes is high, hence anti-diabetic nature of acarbose can be highly useful in treatment regime. Iopromide is contrast agent, used in radiographic studies. Glycyrrhizic acid is plant product obtained from Mulethi or Liquorice (also referred as Yashtimadhura (*Glycyrrhiza glabra*)). It has been shown to have anti-inflammatory, anti-diabetic, anti-oxidant, anti-tumor and anti-viral properties [41]. These properties suggest Glycyrrhizic acid to be of high importance in COVID-19 treatment.

We next pursued virtual screening against supernaturaldb – a database of 3,25,287 natural small molecules (giving rise to 5,03,604 conformations). Virtual screening for shallow site also predicted natural products with high medicinal value such as Gingerenone A (SN00156190) and Shogaol (SN00002189), but with lower docking score (please see Table 1). Gingerenone A has anti-obesity, anti-inflammatory and antibiotic properties [42][43] whereas Shogaol is anticancer, anti-oxidant, antimicrobial, anti-inflammatory anti-allergic and antibiotic in nature [43][44]. MD simulation was not performed for above two because of their lower rank, but can be tested further. Molecules like Galangin, Gingerenone and Shogaol are reported in high quantities in the medicinal plant, Sitharathai (*Alpinia Officinarum*; a form of ginger, also referred as 'Kulanjan' [49]) which has been used for bronchial infections, as a carminative and recently recognized for its anti-viral properties [50]. These herbal plants are already practised in Ayurveda over centuries to treat respiratory problems and inflammation. Extracts from herbal plants provide a host of secondary metabolites which could have a combinatorial effect to reduce the viral load, once consumed in the proper manner.

Other hits from supernaturaldb include compounds SN00003849, SN00003832 and SN00216190, which were found to have stable interactions with deep binding site of Nsp1 as suggested by docking and MD simulation.

Therefore, along with FDA approved drugs which will constitute the treatment by repurposing, these new natural compounds can also be tested for their activity against COVID-19.

Conclusion:

Virtual screening helps in identification of novel drug candidates and repurposing of known drugs. Current pandemic caused by SARS-Cov2. In order to assist in the development of a cure, we have targeted Nsp1 protein of this virus and screened known drugs and natural compounds against it. In this effort, we have identified known anti-viral compounds like Remdesivir and Edoxudine. Other drugs, like Esculin and Acarbose which are not anti-viral, but are used as anti-inflammatory and antidiabetic (respectively) were also identified. These FDA approved drugs can be considered as potential candidates for drug repurposing. Natural compounds like Glycyrrhizic acid (entry 19 in Table 1) from Liquorice and Galangan, Gingeronone and Shogaol (entries 18, 31 and 32 in Table 1) from Sitharathai, were also found to be interacting with Nsp1. These compounds can be considered as novel drug candidates against COVID-19. We find these results to be encouraging and hopefully useful immediately to the community and follow-up validation by other researchers.

References:

- 1 Masters, P.S. (2006) The Molecular Biology of Coronaviruses. *Adv. Virus Res.* 65, 193–292
- 2 Narayanan, K. *et al.* (2015) Coronavirus nonstructural protein 1: Common and distinct functions in the regulation of host and viral gene expression. *Virus Res.* DOI: 10.1016/j.virusres.2014.11.019.
- 3 Wu, F. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269
- 4 Wang, M. *et al.* (2020) Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* 30, 269–271
- 5 Elbe, S. and Buckland-Merrett, G. (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Challenges* DOI: 10.1002/gch2.1018
- 6 Shen, Z. *et al.* (2019) A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. *J. Biol. Chem.* 294, 13606–13618
- 7 Züst, R. *et al.* (2007) Coronavirus non-structural protein 1 is a major pathogenicity factor: Implications for the rational design of coronavirus vaccines. *PLoS Pathog.* DOI: 10.1371/journal.ppat.0030109

- 8 Huang, C. *et al.* (2011) SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: Viral mRNAs are resistant to nsp1-induced RNA cleavage. *PLoS Pathog.* 7,
- 9 Kamitani, W. *et al.* (2009) A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein. *Nat. Struct. Mol. Biol.* 16, 1134–1140
- 10 Kamitani, W. *et al.* (2006) Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc. Natl. Acad. Sci. U. S. A.* 103, 12885–12890
- 11 Narayanan, K. *et al.* (2008) Severe Acute Respiratory Syndrome Coronavirus nsp1 Suppresses Host Gene Expression, Including That of Type I Interferon, in Infected Cells. *J. Virol.* 82, 4471–4479
- 12 Tanaka, T. *et al.* (2012) Severe Acute Respiratory Syndrome Coronavirus nsp1 Facilitates Efficient Propagation in Cells through a Specific Translational Shutoff of Host mRNA. *J. Virol.* 86, 11128–11137
- 13 Wong, C.K. *et al.* (2004) Plasma inflammatory cytokines and chemokines in severe acute respiratory syndrome. *Clin. Exp. Immunol.* DOI: 10.1111/j.1365-2249.2004.02415.x
- 14 R., C. and S., P. (2017) Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin. Immunopathol.* DOI: 10.1007/s00281-017-0629-x.
- 15 Law, A.H.Y. *et al.* (2007) Role for Nonstructural Protein 1 of Severe Acute Respiratory Syndrome Coronavirus in Chemokine Dysregulation. *J. Virol.* DOI: 10.1128/jvi.02336-05
- 16 Jauregui, A.R. *et al.* (2013) Identification of Residues of SARS-CoV nsp1 That Differentially Affect Inhibition of Gene Expression and Antiviral Signaling. *PLoS One* 8, 1–11
- 17 Pfefferle, S. *et al.* (2011) The SARS-Coronavirus-host interactome: Identification of cyclophilins as target for pan-Coronavirus inhibitors. *PLoS Pathog.* 7,
- 18 Brister, J.R. *et al.* (2015) NCBI viral Genomes resource. *Nucleic Acids Res.* DOI: 10.1093/nar/gku1207
- 19 Hatcher, E.L. *et al.* (2017) Virus Variation Resource-improved response to emergent viral outbreaks. *Nucleic Acids Res.* DOI: 10.1093/nar/gkw1065
- 20 Pickett, B.E. *et al.* (2012) ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* DOI: 10.1093/nar/gkr859
- 21 Camacho, C. *et al.* (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* DOI: 10.1186/1471-2105-10-421
- 22 Berman, H. *et al.* Announcing the worldwide Protein Data Bank. , *Nature Structural Biology.* (2003)
- 23 Eswar, N. *et al.* (2006) Comparative Protein Structure Modeling Using Modeller. *Curr. Protoc. Bioinforma.* DOI: 10.1002/0471250953.bi0506s15
- 24 Laskowski, R.A. *et al.* (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* DOI: 10.1107/s0021889892009944

- 25 Wiederstein, M. and Sippl, M.J. (2007) ProSA-web: Interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* DOI: 10.1093/nar/gkm290
- 26 Wishart, D.S. *et al.* (2018) DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* DOI: 10.1093/nar/gkx1037
- 27 Banerjee, P. *et al.* (2015) Super Natural II-a database of natural products. *Nucleic Acids Res.* DOI: 10.1093/nar/gku886
- 28 Madhavi Sastry, G. *et al.* (2013) Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided. Mol. Des.* DOI: 10.1007/s10822-013-9644-8
- 29 Halgren, T.A. (2009) Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* DOI: 10.1021/ci800324m
- 30 Halgren, T. (2007) New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* DOI: 10.1111/j.1747-0285.2007.00483.x
- 31 Friesner, R.A. *et al.* (2006) Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* DOI: 10.1021/jm051256o
- 32 Halgren, T.A. *et al.* (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* DOI: 10.1021/jm030644s
- 33 Friesner, R.A. *et al.* (2004) Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* DOI: 10.1021/jm030643o
- 34 Bowers, K.J. *et al.* (2006) , Scalable algorithms for molecular dynamics simulations on commodity clusters. , in *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06*
- 35 Harder, E. *et al.* (2016) OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* DOI: 10.1021/acs.jctc.5b00864
- 36 Humphreys, D.D. *et al.* (1994) A multiple-time-step Molecular Dynamics algorithm for macromolecules. *J. Phys. Chem.* DOI: 10.1021/j100078a035
- 37 Martyna, G.J. *et al.* (1992) Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* DOI: 10.1063/1.463940
- 38 Martyna, G.J. *et al.* (1994) Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* DOI: 10.1063/1.467468
- 39 Wu, C. *et al.* (2020) Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* DOI: 10.1016/j.apsb.2020.02.008
- 40 Pandey, R.P. *et al.* (2018) Broad-spectrum neutralization of avian influenza viruses by sialylated human milk oligosaccharides: In vivo assessment of 3'-sialyllactose against H9N2 in chickens. *Sci. Rep.* DOI: 10.1038/s41598-018-20955-4
- 41 Ming, L.J. and Yin, A.C.Y. Therapeutic effects of glycyrrhizic acid. , *Natural Product Communications.* (2013) .

- 42 Suk, S. *et al.* (2017) Gingerenone A, a polyphenol present in ginger, suppresses obesity and adipose tissue inflammation in high-fat diet-fed mice. *Mol. Nutr. Food Res.* DOI: 10.1002/mnfr.201700139
- 43 Rampogu, S. *et al.* (2018) Ginger (*Zingiber officinale*) phytochemicals-gingerenone-A and shogaol inhibit SaHPPK: Molecular docking, molecular dynamics simulations and in vitro approaches. *Ann. Clin. Microbiol. Antimicrob.* DOI: 10.1186/s12941-018-0266-9
- 44 Semwal, R.B. *et al.* (2015) Gingerols and shogaols: Important nutraceutical principles from ginger. *Phytochemistry* 117, 554–568
- 45 Chien, S.T. *et al.* (2015) Galangin, a novel dietary flavonoid, attenuates metastatic feature via PKC/ERK signaling pathway in TPA-treated liver cancer HepG2 cells. *Cancer Cell Int.* DOI: 10.1186/s12935-015-0168-2.
46. Chakraborti, Sohini; Srinivasan, Narayanaswamy (2020): Drug Repurposing Approach Targeted Against Main Protease of SARS-CoV-2 Exploiting ‘Neighbourhood Behaviour’ in 3D Protein Structural Space and 2D Chemical Space of Small Molecules. ChemRxiv. Preprint.
<https://doi.org/10.26434/chemrxiv.12057846.v1>.
47. Narayanan, N.; Nair, D.T. Vitamin B12 May Inhibit RNA-Dependent-RNA Polymerase Activity of nsp12 from the SARS-CoV-2 Virus. *Preprints* **2020**, 2020030347 .doi: 10.20944/preprints202003.0347.v1.
48. David Gordon *et. al.* (2020). A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. DOI:10.1101/2020.03.22.002386.
49. Y. Chen *et. al.* (2019). Secondary Metabolites from the Rhizomes of *Alpinia officinarum*. *Chemistry of Natural Compounds* volume 55, pages1176–1178.
50. Manoharan K. Pillai *et. al.* (2018). Therapeutic Potential of *Alpinia officinarum*. *Mini-Reviews in Medicinal Chemistry* Volume 18 , Issue 14 ,. DOI : 10.2174/13895557517666171002154123.

Figures:

Figure 1: Sequence analysis COVID-19 (Wuhan-Hu-1) Nsp1. Represents alignment between Wuhan-Hu-1 Nsp1 and SARS Nsp1 protein sequence. Red highlights consensus sequences whereas Blue highlights difference in amino acid sequence. Important residues, shown to play role in affecting host gene expression and anti-viral signaling, are highlighted in green and pink colour. Green highlighting similar residues whereas Pink highlighting residues which are different in COVID-19.

Figure 2: Model of COVID-19 (Wuhan-Hu-1) Nsp1 with Deep and shallow binding site predicted by SiteMap a). Represent COVID-19 Nsp1 model derived using Modeller 9.22, using 2hsx as a template. Red dot represents Shallow binding site consisting region of alpha-helix and beta-sheets. Blue dots represent deep binding site present in mostly loop region. b). Represents residues present in deep and shallow binding site respectively.

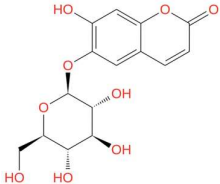
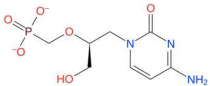
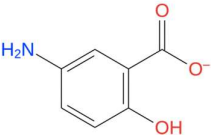
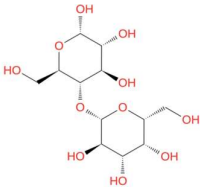
Figure 3: Docking and MD simulation results for Nsp1-deep-Esculin. a. Esculin-Nsp1 interactions after XP docking b. Interaction types and Interacting residues of Nsp1 with Esculin over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (Esculin).

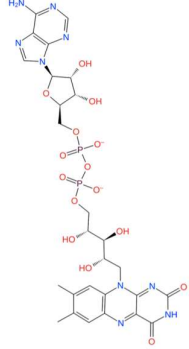
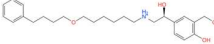
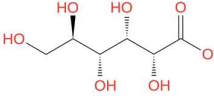
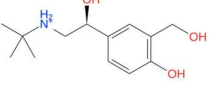
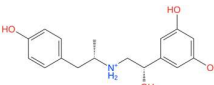
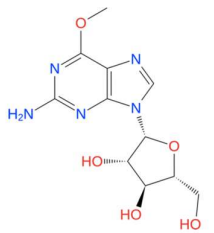
Figure 4: Docking and MD simulation results for Nsp1-deep-SN00003849. a. SN00003849-Nsp1 interactions after XP docking b. Interaction types and Interacting residues of Nsp1 with SN00003849 over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (SN00003849).

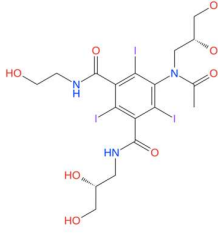
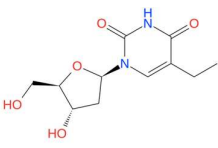
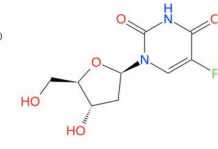
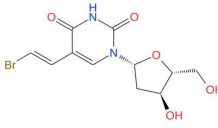
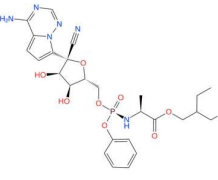
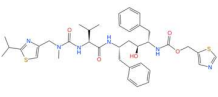

Figure 5: Docking and MD simulation results for Nsp1-shallow-Glycyrrhizic acid. a. Glycyrrhizic acid-Nsp1 interactions after XP docking. b. Interaction types and Interacting residues of Nsp1 with Glycyrrhizic acid over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (Glycyrrhizic acid).

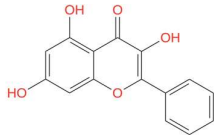
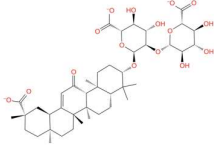
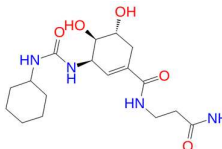
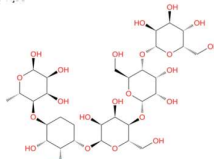
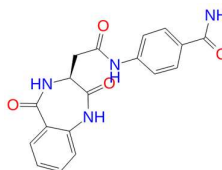
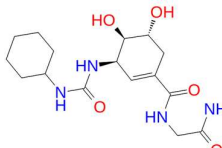
Acknowledgements: We would like to thank NCBS (TIFR) for infrastructural facilities. The authors thank Dr. Radhika Venkatesan for useful discussions. RS would like to acknowledge her JC Bose fellowship (JC Bose fellowship (SB/S2/JC-071/2015) from Science and Engineering Research Board, India.

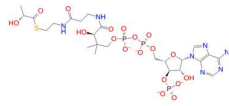
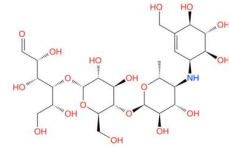
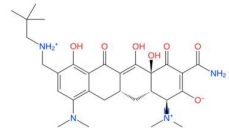
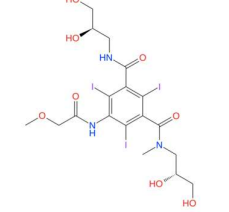
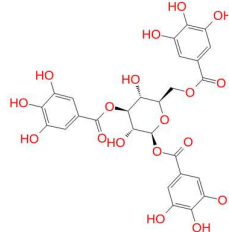
Table 1: Top-ranking hits identified by the virtual screening and other promising small molecules

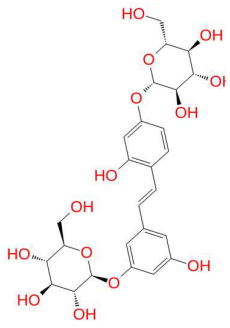
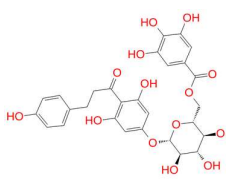
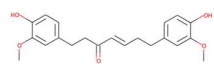
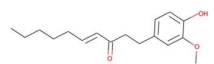
No	Compound	Structure	Source	Deep binding site			Shallow binding site			Comment
				XP Score	MMGBSA score	MD	XP Score	MMGBS A score	MD	
1	Esculin		Drugbank	-6.883	-29.24	Stable				Vast-Protective agent, antioxidant and anti-inflammatory [26]
2	Cidofovir		Drugbank	-5.776	-16.53	Stable				Anti-viral [26]
3	Mesalazine		Drugbank	-5.146	-20.74	Not stable				Anti-inflammatory agent [26]
4	Lactose		Drugbank	-11.16	-40.72	Not stable				It is a disaccharide of glucose and galactose. Used as a nutrient supplement. [26]

5	FAD		Drugbank	-7.895	-52.47	Not stable	-5.643	-40.25	Not stable	Used in ophthalmic treatment for vitamin B2 deficiency. [26]
6	Salmeterol		Drugbank	-5.69	-47.8	Not stable				Beta-2 adrenergic receptor agonist. Used in treatment of asthma and COPD. [26]
7	Zinc gluconate		Drugbank	-8.525	-18.19	Not stable				Treating diarrheal episodes in children and reduced duration of common cold. [26]
8	Salbutamol		Drugbank	-5.097	-40.86	Not stable				a short-acting, beta-2 adrenergic receptor agonist. [26]
9	Fenoterol		Drugbank	-6.543	-45.18	Not stable				Adrenergic beta-2 receptor agonist. [26]
10	Nelarabine		Drugbank	-6.307	-36.86	Not stable				Anti-neoplastic agent. [26]

11	Ioxilan		Drugbank	-8.451	-35.28	Not stable			Tri-iodinated diagnostic contrast agent. [26]
12	Edoxudine		Drugbank	-5.941	-33.63	Stable			deoxy-thymidine analog and Anti-viral agent. [26]
13	Floxuridine		Drugbank	-5.125	-33.27	Not stable			Anti-neoplastic and antimetabolite agent. [26]
14	Brivudine		Drugbank	-5.8	-30.61	Stable			Anti-viral [26]
15	Remdesivir		Drugbank	-5.795	-40.01	Stable			Anti-viral agent, a potential treatment for Ebola and shown to be effective against COVID-19. [26][4]
16	Ritonavir		Drugbank	-2.778	-27.75	Not stable	-2.25	-55.25	Not stable Anti-viral agent. [26]
17	Brincidofovir		Drugbank	-2.024	-27.75	Not stable			Anti-viral agent. [26]

18	Galangin		Natural product	-3.532	-17.69	Not stable	-2.278	-22.99	Not stable	dietary flavonoid having anticancer properties [45]
19	Glycyrrhizic Acid		Drugbank	-4.609	-23.59	Not stable	-3.643	-27.06	stable	anti-inflammatory, anti-diabetic, anti-oxidant, anti-tumor and anti-viral properties [41]
20	SN00003849		Super natural database	-6.295	-51.08	Stable				Plant product
21	SN00220639		Super natural database	-12.25	-50.79	Not stable	-10.59	-67.94	Not stable	Plant product
22	SN00103215		Super natural database	-5.264	-47.66	Not stable				Plant product
23	SN00003832		Super natural database	-6.308	-47.03	Stable				Plant product

24	SN00216190		Super natural database	-7.78	-46.11	Stable		Plant product
25	Acarbose		Drugbank		-8.365	-28.35	stable	Anti-diabetic. [26]
26	Omadacycline		Drugbank		-6.025	-35.52	Not stable	Antibiotic [26]
27	Iopromide		Drugbank		-4.73	-31.58	Stable	Used as a contrast agent [26]
28	SN00037405		Super natural database	-9.177	-61.45	Not stable		Plant product

29	SN00161170		Super natural database				-6.676	-57.94	Not stable	Plant product
30	SN00038342		Super natural database				-6.066	-51	Not stable	Plant product
31	SN00156190 (Gingerenone)		Super natural database	-4.39	-22.27	not done	-2.578	-36.38	not done	Anti-obesity , Anti-inflammatory and antibiotic [42][43]
32	SN00002189 (Shogaol)		Super natural database	-2.64	-36.18	not done				anticancer, anti-oxidant, antimicrobial , anti-inflammatory anti-allergic and antibiotic [43][44]

Molecules marked in **bold** have been discussed in the text.

Wuhan-Hu-1	MESLVPGFNEKTHVQLSLPVLQVRDVLVRGFGDSVEEVLSEARQHLKDGTCGLVEVEKGV	60
SARS	MESLVLGVNEKTHVQLSLPVLQVRDVLVRGFGDSVEEALSEAREHLKNGTCGLVELEKGV	60
Wuhan-Hu-1	LPQLEQPYVFIKRSDARTAPHGHVMVELVAELEGIQYGRSGETLGVLVPHVGEIPVAYRK	120
SARS	LPQLEQPYVFIKRSDALSTNHGHKVELVAEMDGIQYGRSGITLGVLVPHVGETPIAYRN	120
Wuhan-Hu-1	VLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQENWNTKHSSGVTRELMRELNGG	180
SARS	VLLRKNGNKGAGGHSYGIDLKSYDLGDELGTDPIEDYEQNWNTKHGSGALRELTRELNGG	180

Figure 1: Sequence analysis COVID-19 (Wuhan-Hu-1) Nsp1. Represents alignment between Wuhan-Hu-1 Nsp1 and SARS Nsp1 protein sequence. Red highlights consensus sequences whereas Blue highlights difference in amino-acid sequence. Important residues shown to play role in affecting host gene expression and anti-viral signaling are highlighted in green and pink color. Green highlighting similar residues whereas Pink highlighting residues which are different in COVID-19.

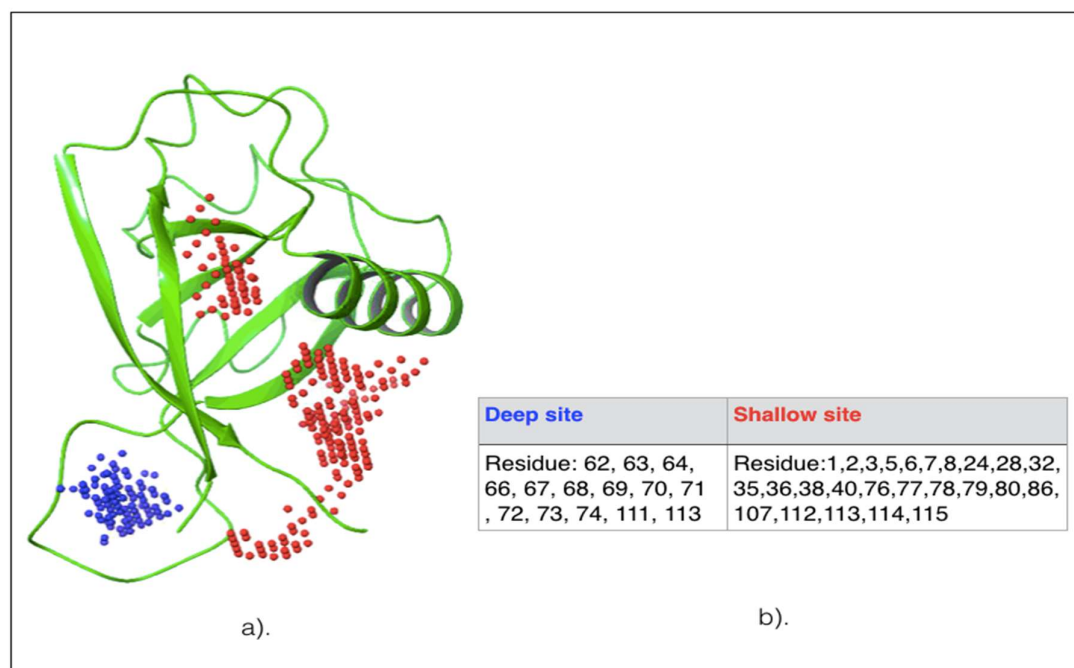


Figure 2: Model of COVID-19 (Wuhan-Hu-1) Nsp1 with Deep and shallow binding site predicted by SiteMap

- Represent COVID-19 Nsp1 model derived using Modeller 9.22, using 2hsx as a template. Red dot represents Shallow binding site consisting region of alpha-helix and beta-sheets. Blue dots represent deep binding site present in mostly loop region.
- Represents residues present in deep and shallow binding site respectively. Residue numbers are as per the structural model (Residue 1 of structure is residue 12 in the sequence)

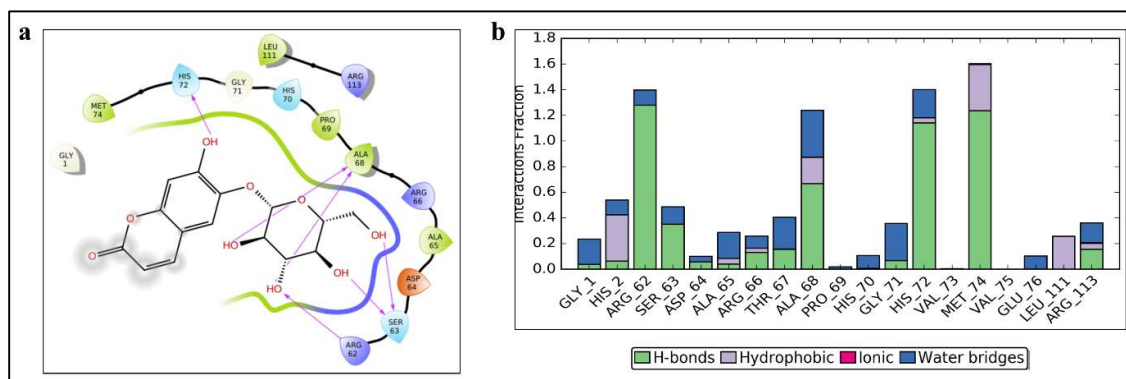


Figure 3: Docking and MD simulation results for Nsp1-deep-Esculin. **a.** Esculin-Nsp1 interactions after XP docking **b.** Interaction types and Interacting residues of Nsp1 with Esculin over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (Esculin) .

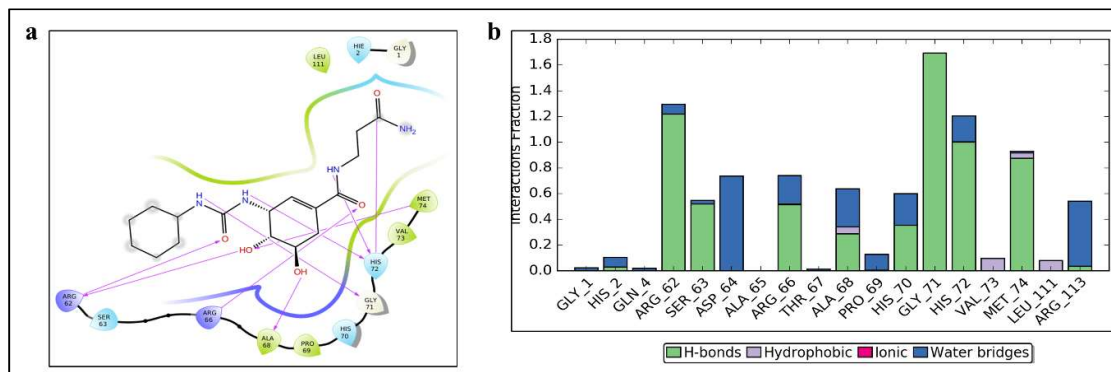


Figure 4: Docking and MD simulation results for Nsp1-deep-SN00003849. **a.** SN00003849-Nsp1 interactions after XP docking **b.** Interaction types and Interacting residues of Nsp1 with SN00003849 over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (SN00003849) .

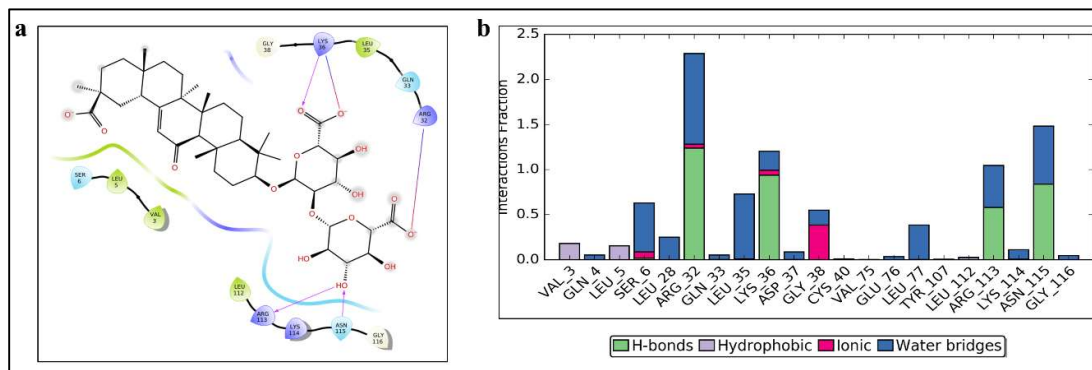


Figure 5: Docking and MD simulation results for Nsp1-shallow-Glycyrrhizic acid. **a.** Glycyrrhizic acid-Nsp1 interactions after XP docking **b.** Interaction types and Interacting residues of Nsp1 with Glycyrrhizic acid over simulation time. Normalized stacked bars indicate the fraction of simulation time for which a particular type of interaction was maintained. Values more than 1.0 suggests that the residue forms multiple interactions of same subtype with ligand (Glycyrrhizic acid) .