

Cite as: P. K. Panda *et al.*, *Sci. Adv.*
10.1126/sciadv.abb8097 (2020).

Structure-based drug designing and immunoinformatics approach for SARS-CoV-2

Pritam Kumar Panda^{1*}, Murugan Natarajan Arul², Paritosh Patel³, Suresh K Verma³, Wei Luo¹, Horst-Günter Rubahn⁴, Yogendra Kumar Mishra⁵, Mrutyunjay Suar³, Rajeev Ahuja^{1,6*}

¹Condensed Matter Theory Group, Materials Theory Division, Department of Physics and Astronomy, Uppsala University, Box 516, SE-75120, Uppsala, Sweden. ²Department of Theoretical Chemistry and Biology, Royal Institute of Technology (KTH), AlbaNova University Center, Stockholm 10691, Sweden. ³School of Biotechnology, KIIT University, Bhubaneswar 751024, India. ⁴Syddansk Universitet, Alision 2, DK-6400, Sønderborg, Denmark. ⁵Mads Clausen Institute, NanoSYD, University of Southern Denmark, Alision 2, DK-6400, Sønderborg, Denmark. ⁶Applied Materials Physics, Department of Materials Science and Engineering, Royal Institute of Technology (KTH) SE-10044 Stockholm, Sweden.

*Corresponding author. Email: rajeev.ahuja@physics.uu.se; pritam.panda@physics.uu.se

The prevalence of respiratory illness caused by the novel SARS-CoV-2 associated with multiple organ failures is spreading rapidly due to its contagious human-to-human transmission and inadequate global healthcare systems. Pharmaceutical re-use, reflecting an effective drug development technique using existing drugs, could shorten the time and reduce the costs relative to de novo drug discovery. We have performed virtual screening of antiviral compounds targeting the spike glycoprotein (S), main protease (M^{Pro}), and the SARS-CoV-2 RBD-ACE2 complex of SARS-CoV-2. PC786, an antiviral polymerase inhibitor, showed improved binding affinity toward all the targets. Furthermore, the post-fusion conformation of the trimeric S protein RBD domain with ACE2 revealed conformational changes associated with the PC786 drug binding. The proposed T cell and B cell epitope identification using the immunoinformatics approach could direct the experimental study with a higher probability of discovering appropriate vaccine candidates with fewer experiments and higher reliability.

Introduction

A new coronavirus disease previously known as 2019-nCoV but later known as SARS-CoV-2 has recently emerged from China with a total of >4M confirmed cases and >300K deaths worldwide (1). Similar to SARS-CoV, SARS-CoV-2 tends to transfer rapidly from human to human, distributed across multiple continents (2). Epidemiological studies help determine the health status of a nation and enable a better distribution of economic resources. An epidemiological data source of Respiratory Viral Infections from Yale and BioRender scientific team shows that COVID-19 has a high rate of hospitalization due to the high mortality rate as well as the declaration of SARS-CoV-2 as a pandemic by WHO. The community attack rate is way much higher in comparison to other respiratory viral infections, as shown in Fig. 1A. To mitigate this challenge, several researchers from all over the world try to develop novel or repurpose antiviral drugs through experimental and computational methods to diminish the fear of this pandemic outbreak.

The first reported genome sequencing of SARS-CoV-2 lead the researchers to determine a key target, the SARS-CoV-2 spike (S) glycoprotein for therapeutic and diagnostics (3). The S protein of SARS-CoV-2 (PDB ID: 6VSB) contains a receptor-binding domain (RBD), which interacts with the peptidase domain of the Angiotensin-converting enzyme 2 (ACE2) (4), thus mediates receptor recognition and membrane fusion (5).

Research into finding appropriate drug compounds targeting the S protein RBD-ACE2 complex facilitated through virtual screening of drug compounds computationally is in pursuit to ensure potential treatments (6). Another vital enzyme, i.e., SARS-CoV-2 main protease (PDB ID: 6LU7) (7) that intercedes viral replication and transcription together with the absence of closely related homologs protein in humans can be a potential target for therapeutics development. In the mission to stop the outbreak, developing medical countermeasures using molecular modeling, virtual screening of drug candidates along with receptor-drug molecular dynamics simulation can facilitate the ease of finding antiviral drugs for SARS-CoV-2. The plan to re-use old drugs introduced for past outbreaks, e.g., MERS-CoV, SARS-CoV, Ebola, and HIV could, therefore, accelerate the discovery process (8).

We have devised a similar strategy using *In-Silico* approaches that include (i) inhibition of the S protein, M^{Pro} and RBD-ACE2 complex interaction using small molecules/antiviral drugs (ii) immunoinformatics approach for designing specific epitopes of MHC class-I antigens for adaptive immunity using S protein and M^{Pro} (Fig. 1B). Both the computational strategies could pave a path for the experimentalists and pharmaceuticals companies to design drugs and vaccines for this SARS-CoV-2 in a short period. Repurposing of potential drug candidates having a broad-spectrum antiviral activity targeting the viral entry mechanism could be beneficial

for clinical use. The current study deals with a similar strategy to find potential drug or vaccine candidates suitable for possible experimental studies targeting the infection pathway of SARS-CoV-2.

The key mechanism underlying the SARS-CoV-2 replication and maturation of the virion mediated through host cell attachment is illustrated in Fig. 1C. The interaction of spike protein (RBD domain) initiates the coronavirus attachment to the host cell to the host receptor through membrane fusion and endocytosis (Fig. 1C). Followed by the receptor binding, the release of the viral genome is accomplished by acid-dependent proteolytic cleavage of S protein by protease enzyme. The translation of the viral polymerase protein starts after utilizing RNA pseudoknot and a slippery sequence (5'-UUUAAAC-3') that causes ribosomal frameshifting. The assembly of the viral replicases complexes follows the translation and assembly by producing genomic and sub-genomic RNA's. The genomic and sub-genomic RNA's were produced through negative-strand intermediates and from which the sub-genomic RNA's goes through nested transcription. The Spike (S), Membrane (M), and Envelope (E) protein then undergoes translation or assembly through insertion to Endoplasmic Reticulum (ER) that moves along a secretory pathway into the endoplasmic-reticulum-Golgi intermediate compartment (ERGIC). There, viral genomes encapsidated by (Nucleocapsid) N protein bud into membranes of the ERGIC containing viral structural proteins, forming mature virions. Upon assembly, vesicles transport virions to the cell surface and induce exocytosis (9).

A variable number (6 to 11) of Open reading frames (ORFs) included in the coronavirus genome, which has a size from about 26,000 to 32,000 bases. The first ORF codes 16 non-structural proteins (NSPs), covering almost 67% of the entire genome. The other ORFs include accessory proteins and structural proteins. The 5' gene contains more than two-thirds of the orflab encoding orflab polyproteins. The 3' gene comprises structural-protein encoding (S), envelope (E), membrane (M), and nucleocapsid (N) proteins (10). (Fig. 1D) The cryo-electron microscopy structure of CoV trimeric spike (S) glycoprotein (PDB ID-6VSB) determined by Wrapp *et al.* (3), is considered to be a key target in therapeutics and diagnostics of this pandemic coronavirus spread. The spike protein is in a state of trimeric form with three receptor binding domains and in metastable prefusion conformation that undergoes structural rearrangements to fuse the viral membrane with the host human cell membrane. The receptor-binding domains undergo hinge-like conformational movements when attaches with the angiotensin-converting enzyme 2 (ACE2) host receptor. The crystal structure of SARS-CoV-2 main protease (PDB ID-6LU7) consists of three domains, i.e., Domain I and II have a chymotrypsin-like and two- β -barrel fold conformations, and Domain III consists of

five helices that adopt a globular structure (11). The coronavirus main protease (M^{pro}) enzyme is essential for proteolytic maturation of the virus. It is a promising target for the discovery of small-molecule drugs that would inhibit cleavage of the viral polyprotein and prevent the spread of the infection.

There are currently no scientifically appropriate vaccinations or unique antiviral therapy for the prevention or treatment of COVID-19. The α -interferon mixture and anti-HIV Lopinavir / Ritonavir (Kaletra[®]) (12) medications have been used, but there is still a minimal curative benefit, and toxic side effects can occur (13). Remdesivir is also being explored for the treatment of COVID-19, a broad-spectrum antiviral developed by Gilead Sciences, Inc., but validation from clinical studies are needed to demonstrate its effectiveness (14). Based on the recent research report, we computationally screened 640 antiviral compounds from ChEMBL (15) database against the S protein and M^{pro} using Autodock Vina (16) (Fig. 1E). We have used UCSF Chimera (17) and Discovery Studio Visualizer (18) for the post-docking analyses. An antiviral polymerase inhibitor PC786 bearing ChEMBL ID 4291143 proved to be the best among all the antiviral drugs against both the target receptors. Apart from PC786, several other antiviral drugs, i.e., Lorecivivint, Tegavivint, Dolutegravir, also have better binding affinities toward S and M^{pro} .

Furthermore, we have also compared the binding affinities of the FDA approved drugs against these two target receptors, along with the RBD-ACE2 complex. The screened drugs were proven more effective than the FDA approved medications in terms of binding affinities. We also combined the molecular dynamics simulation with a virtual screening strategy to validate the binding strength of the PC786 drug in comparison to the FDA approved drugs, i.e., Zanamivir and Lopinavir. Moreover, upon interaction with the PC786 drug, the trimeric S protein RBD domain in complex with ACE2 seems to change its conformation. ACE2 directly interacts with the RBD domain in a close conformation with high binding affinity in its native state, whereas, in the case of the PC786 drug bound to trimeric S protein, the RBD domain seems to interact with ACE2 in an open conformation.

In addition to the drug screening method, the application of immunoinformatics using a bioinformatics approach to the design of different vaccine candidates for SARS-CoV-2 serves as an alternative and promising approach (Fig. 1F). Active counteractions to the recent appearance and rapid expansion of the SARS-CoV-2 entail the creation of data and resources to identify and track its spread and immune response. As of 27 January 2020, the IEDB curated 581 linear and 81 discontinuous B cell epitopes along with 320 T cell epitopes for SARS-CoV-2 (19). Similarly, we have also predicted T cell and B cell epitopes using structural sequences of S and M^{pro} proteins of SARS-CoV-2 using IEDB resources (20). Independent detection of the epitopes using the vital proteins

illustrates the high likelihood of identifying vaccine candidates for the immune responses to SARS-CoV-2. Such forecasts will promote the successful design of vaccinations against this high-priority virus.

Results

Structure-based drug-design approach: Screening of ChEMBL antiviral compounds

We have screened the ChEMBL database for antiviral drugs that have passed the Lipinski rule of 5 (RO5) for drug-likeness (Fig. 1E). The structure-based drug design approach was taken into consideration using both the S protein and the main protease (M^{pro}) of SARS-CoV-2. The antiviral drugs obtained from the ChEMBL database have been listed in supplementary data file 1. High-throughput virtual screening of the antiviral drugs using molecular docking approach resulted in a broad range of binding affinity toward both the receptors (Fig. 2, A and B), typically ranging from -1.5 kcal/mol to -11.5 kcal/mol. Among all the antiviral drugs, PC786 has been considered to have the highest binding affinity toward both the target receptors (Fig. 2C). The PC786 drug bearing ChEMBL ID 4291143 has already passed Phase I and Phase II clinical trials. It is a non-nucleotide inhibitor of RSV polymerase that inhibits replication of both the A and B subtypes of acute respiratory syncytial virus (RSV), thus interrupting the spread of infection within the respiratory tract. The inhalation administration of PC786 leads to high local airway concentration at a viral replication site with low systemic toxicity and, therefore, a low risk of systemic side effects. Via phase I research, PC786 shows excellent safety and tolerability. Pre-clinical and clinical pharmacokinetic data show low systemic concentrations with prolonged lung retention (21). The molecular docking analysis also revealed that PC786 has the best binding affinity of -11.3 kcal/mol with S protein (Fig. S1A) and -9.3 kcal/mol with the M^{pro} of SARS-CoV-2 (Fig. S1K and Fig. 2C).

The findings are the basis for the repurpose of the approved/investigational small molecules against SARS-CoV-2 infection. We have also compared the relative binding affinities of screened antiviral compounds with the FDA approved drugs using the two target receptors. The results showed superior binding affinities of computationally screened viral compounds i.e., PC786 when compared to those drugs under clinical trials (except the drug, Lopinavir) (Fig. 2D). Lopinavir has shown the highest binding affinity toward M^{pro} with the binding affinity of -9.5kcal/mol, which is comparatively less than the PC786 drug. Similarly, Remdesivir and Zanamivir have shown high binding affinities toward trimeric S protein, i.e., -7.6 and -6.9 kcal/mol. It is known that Remdesivir targets only a highly conserved RNA-dependent RNA polymerase (RdRp) in diverse RNA viruses, providing a basis for designing broad-spectrum antiviral drugs based on

nucleotide analogs (14). So, for further validation and comparison of the FDA approved antiviral drugs with the best screened drugs, we employed Molecular Dynamics (MD) simulation along with molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) calculations (Fig. S1S and Fig. 2, E and F). The findings from the MD and MM-PBSA calculations are further explored in MD section.

Molecular docking analyses of screened drugs against S and M^{pro}

Our *in-silico* strategy enables us to design and screen the small molecules targeting the trimeric S protein that contains key structural domains (Fig. 3A). Specifically, RBD domain plays a pivotal role in viral-host attachment (3). These specific structural domains can be targeted with small molecules to disrupt the viral attachment to the host proteins. From the post-docking interaction analysis, PC786 has shown most favorable binding affinity toward RBD domain of all the chains in trimeric S protein. PC786 also targets the junction of HR1/CH domain. The specificity of PC786 binding to RBD domain of trimeric S protein could be of potential interest for experimentalists or clinicians for further validation. PC786 drug interaction with the trimeric S protein showed conventional carbon-hydrogen bonds with Gly413 of B chain and Asp427, Cys379 of C chain, halogenic bond with Pro986 B chain, alkyl bonds with Leu752 and Pro987 (Fig. 3. B to C and Fig. S2A). The halogen bond with Pro986 indicates that there is a net attractive interaction between an electrophilic region associated with a fluorine atom and a nucleophilic region Pro986. Pi-anion bond with Glu988 also has been observed for the same. Apart from PC786, Lorecivivint and Tegavivint also yielded good binding affinity of -10.7 kcal/mol, -10.2 kcal/mol in case of trimeric S protein (Fig. S1, B and C).

Lorecivivint is a phase II clinical trial anti-inflammatory drug that inhibits or even reverses the progression of osteoarthritis. Tegavivint is an antineoplastic drug that inhibits the Wnt/beta-catenin pathway with potential antineoplastic activity. Among the FDA approved drugs, Zanamivir turned out to be the second-best (next to Remdesivir) having a binding affinity of -6.9 kcal/mol with S protein (Fig. S1. D). Zanamivir showed conventional hydrogen bonding with Asn978, Val976 of A chain, and rest of the key residues involved were from the C chain of trimeric S protein (Fig. 3, D to E and Fig. S2. D). In the case of Lorecivivint, there are many conventional hydrogen bonds with interacting with the A chain of the S protein, i.e., the O atom and N atom of the Lorecivivint forms bond with Arg1014 and Arg1019 respectively (Fig. S2. B). In the case of Tegavivint, many alkyl bonds and an unfavorable donor-donor bond with Arg1014 of A chain seem to be formed during the interaction with the trimeric S protein of SARS-CoV-2 (Fig. S2. C). All the bonding patterns mentioned above that have been interacted with the

spike protein are in the vicinity of the active site pockets, as illustrated in Fig. S1. I and supplementary data file 2. The target receptor hydrogen bonding and solvent accessible surfaces were depicted in Fig. S1, E to H.

The SARS-CoV-2 main protease (M^{pro}) protein tends to show high binding affinities with PC786 with a binding affinity of -9.3 kcal/mol. JNJ-449095397 (JNJ), which is an inhibitor for chronic obstructive pulmonary disease, also has high the binding affinity of -9.2 kcal/mol with the M^{pro} of SARS-CoV-2 (Fig. S1. L). The interaction analysis of the best screened drugs with M^{pro} revealed that, the drugs mostly binds to junction of Domain II and III that have two- β -barrel fold conformations and five helices (Fig. 3F)

PC786 showed conventional hydrogen bonds with Gln110 and many hydrophobic interactions with Phe134, Pro108, Ile249, Val202, Pro203 and Phe294 (Fig. 3, G to H) whereas, JNJ showed conventional hydrogen bonds with Lys137, Thr199, Tyr239, Asn277, Met276 (Fig. S1, K and L, Fig. S2, E and F). Additionally, Lorecivint also showed a better binding affinity of -8.9kcal/mol that has a high binding affinity than FDA approved drug Lopinavir (Fig. S1. M and Fig. S2. G) with conventional hydrogen and halogen bonds. In the case of FDA approved drug Lopinavir, Pro293, Val297, Val104, Ile206 forms hydrophobic interactions with the drug molecule along with two Pi-sigma bonds associated with Ile249 and Phe 294 (Fig. 3 I to J, Fig. S1. N and Fig. S2. H). All the interacted residues in the M^{pro} protein were in the active site region, as shown in Fig. S1. J and supplementary data file 2. The target receptor hydrogen bonding and solvent accessible surfaces were depicted in Fig. S1, O to R. The final screened best antiviral and FDA approved compounds with virtual screening binding affinity scores have been provided in supplementary data file 1. The therapeutic description of best selected antiviral drugs was also described in supplementary data file 1.

MD simulation with free energy (MM-PBSA) calculations

To validate the intrinsic atomic interaction and binding conformation of the best screened antiviral drugs, we have employed all atom-based molecular dynamics simulation using GROMACS v.2019.2 (22) for ten nanoseconds (10 ns) for the protein-drug conjugates. Fig. 4 A to S illustrates the results from MD calculations i.e., conformations of the proteins bound to the antiviral drugs and the molecular interactions antiviral drugs. We have taken PC786 and Zanamivir with S protein (Fig. 4, A to H) and PC786, JNJ and Lopinavir with M^{pro} (Fig. 4 I to S). The clusters of the 10ns simulated structures for S protein and M^{pro} conjugated with the drugs mentioned above were illustrated in Fig. S3, A to B and Fig. S4, A to C respectively. We have taken the FDA approved drugs Zanamivir and Lopinavir to compare with the virtually

screened drugs against S protein and M^{pro} , respectively as the binding affinities of the respective compounds is higher toward respective target receptors. We have compared the 10ns simulated structure with the initial configuration (0ns) using the structure comparison tool (MatchMaker) from UCSF Chimera to align the pair of protein chains that uses BLOSUM-62 matrix and Needleman-Wunsch algorithm. Fig. 4, A and E represents the overlap of 0ns to 10ns simulated structure of S protein bound to PC786 and Zanamivir, respectively. Similarly, we have compared the simulated structure of M^{pro} bound to PC786, JNJ and Lopinavir that are illustrated in Fig. 4. I, M and Q, respectively.

The 10ns molecular dynamics simulation revealed that the trimeric S protein seems to be stable that showed Root Mean Square Deviation (RMSD) in the range of ~0.2nm to ~0.5nm in both the drug conjugates i.e., PC786 - trimeric S protein and Zanamivir - trimeric S protein (Fig. 4D) respectively. The ligand (drug molecules) bound to trimeric S protein seems to fluctuate during the 10ns simulation that varied in the RMSD range of ~0.15nm to ~1.3nm as shown in Fig. 4H for both systems, respectively. PC786 showed less RMSD deviation (Fig. 4B) from which the stability of the drug can be delineated whereas, Zanamivir showed significant deviation (Fig. 4F). Similarly, Root Mean Square Fluctuation (RMSF) analysis also showed significant fluctuations in Zanamivir compared to PC786 (Fig. S3, E and F). The compactness of the S protein receptor upon binding with Zanamivir has been significantly reduced (Fig. S3. G). PC786 seems to retain the native bonding patterns (at 0ns), whereas Zanamivir showed significant deviation in bonding patterns (Fig. 4, C and G), respectively. Moreover, the total interaction energies calculated from the average short-range Coulombic interaction energy and the short-range Lennard-Jones energy is significantly higher in the case of Zanamivir that shows a less binding affinity toward the target receptor (likely to deviate from its original orientation) (Fig. S3, C and D).

Furthermore, to estimate the interaction free energies of the biomolecular interaction, the complex systems have been subjected to molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) calculations (23). The combination of molecular dynamics with MM-PBSA incorporates conformational fluctuations and entropic contributions to the binding energy. The binding energy decomposition analysis divulged into various free energies associated with the ligand upon binding to the protein, e.g., ΔE_{MM} , ΔG_{polar} , and $\Delta G_{nonpolar}$ were calculated separately and later combined to predict the total energies of the individual components. The net contribution energy of the residues involved in the interaction to the drug molecules have been depicted in Fig. S3. H and supplementary data file 3). The energy components such as ΔE_{MM} , ΔG_{polar} , and $\Delta G_{nonpolar}$ were calculated for 10ns extracted at each 10ps interval from the production trajectories. The binding energy

$\Delta G_{\text{binding}}$ calculated from MM-PBSA calculation is higher in magnitude in the case of PC786 (-376.95 kJ/mol) as compared to Zanamivir (-109.74 kJ/mol) (Fig. 2F).

A similar approach has been undertaken for M^{pro} as well, where we compared PC785 and JNJ with FDA approved drug Lopinavir as it showed high binding affinity. The RMSD of the M^{pro} protein tends to deviate more in the case of Lopinavir as compared to PC786 and JNJ (Fig. 4L). The ligand RMSD of PC786, JNJ and Lopinavir showed comparatively similar behavior in terms of RMSD deviation (Fig. 4P) and can be visualized in Fig. 4, J, N and R respectively. However, the RMSF of Lopinavir tends to have more fluctuation as compared to PC786 and JNJ (Fig. S4, F and G) whereas, the compactness of the proteins remains the same in all three cases (Fig. S4H). Fig. 4, K, O and S shows the 2D interaction plots of PC786, JNJ and Lopinavir interacted with M^{pro} after 10ns simulation.

The total interaction energy of Lopinavir is significantly high as compared to PC786 and JNJ (Fig. S4, D and E). The binding energies $\Delta G_{\text{binding}}$ of PC786 and JNJ (-179.79 and -177.56 kJ/mol) are higher in magnitude as compared to Lopinavir (-131.49 kJ/mol) (Fig. 2E). The net contribution energy of the residues involved in the interaction to the drug molecules have been depicted in Fig. S4. I and supplementary data file 3). The contributions from van-der Waals interactions is -247.146 kJ/mol and -233.90 kJ/mol to the total binding free energies of PC786 with S protein and M^{pro} respectively, suggesting that the complexation process is driven by hydrophobic interactions. If we compare the known antivirals, the contributions from van-der Waals interactions -92.82kJ/mol in the case of Zanamivir with S protein whereas, Lopinavir has van-der Waals contribution of -188.88 kJ/mol with M^{pro} . In both the cases i.e., S and M^{pro} , PC786 binding free energy and van-der Waals contributions are higher than all the other known antiviral compounds which determines its specificity. Thus, the binding free energy calculations confirmed the favorable binding of antiviral drugs with the SARS-CoV-2 S and M^{pro} proteins and demonstrate the use of computational screening and free energy calculations on compounds from open-source chemical space toward successful research in drug design.

Conformational changes of trimeric S protein RBD-ACE2 complex upon interaction with antiviral drugs

The trimeric S protein RBD domain mediates receptor recognition and membrane fusion upon interaction with the ACE2 enzyme (Fig. 5A). The trimeric S protein S1 subunit that contains the receptor-binding domain (RBD) directly binds to the host receptor ACE2, which may exploit host infection. The S protein S2 subunit is mainly responsible for membrane fusion that is exposed and is cleaved by the ACE2 protease domain that is critical for viral infection. Initially, we have performed an ensemble docking of best-screened drugs

obtained from virtual high-throughput screening (Fig. 5B) with the complex structure. The ensemble docking provides us similar results obtained using single target receptors, i.e., S protein and M^{pro} . Again, PC786 was proven to be having a higher binding affinity toward the complex as well (-12.1kcal/mol) (Fig. 5C). Furthermore, we have checked the binding affinity of PC786 with another configuration of the complex of RBD-ACE2 involving only the RBD domain (Chain A) of trimeric S protein interacted with ACE2 (Fig. S5, A and B). The binding affinity of -12.1 kcal/mol was obtained majorly involving residues of ACE2 protein. His374 and His505 of ACE2 forms conventional hydrogen bonding with the PC786 drug (Fig. S5C). The full trimeric S protein RBD-ACE2 complex conjugated with PC786 drug results in forming conventional hydrogen bonding with Cys379 with the binding affinity of -11.3kcal/mol (Fig. S5, D to F).

Next, we performed a protein-protein interaction analysis between the trimeric S protein and ACE2 using PatchDock server (24). We have also used FireDock server for further refinements of the complexes. We observed a significant conformational change between the native trimeric S protein RBD-ACE2 complex and the PC786 drug conjugate complex. When the full trimeric S protein RBD binds to ACE2, closed conformation with the highest binding affinity of -40.6 kcal/mol with key residues involved has been observed as depicted by Yan *et al.* (25) (Fig. 5D). The closed conformation involves Asp615, Thr376, Lys417, Tyr453, Cys488, Gly404, Tyr508, Tyr421, Tyr453 of the RBD domain of S protein and Tyr613, Glu23, Ala386, Ala384, Tyr613, Ala614, Ser19, Tyr20, Lys26, Lys353, Gln388, Lys475 and Arg559 of the ACE2 subunit (Fig. 5E). The closed conformation significantly involves hydrophobic residues that play a pivotal role in signal transduction processes/ signaling cascade.

When PC786 binds to the RBD domain of the S protein, upon interaction with ACE2 enzyme, the conformation changed from close to open with the least binding affinity of -6.7 kcal/mol according to the protein-protein interaction analysis that may affect viral attachment and infection (Fig. 5F). The significant residual changes observed in this complex interaction generally involve, i.e., Asn343, Ala344, Arg355, Arg357, Ser399, Lys424, Tyr453, Val512, Lys129, Ser155 of the RBD and N-terminal domain as well. ACE2 enzyme involves Thr608, Asp609, Trp606, Ser602, Asn250, Glu238, Asp597, Glu589, Ser602, Asp157, Asp136 in which most of the residues are from the C-Terminal domain of ACE2 upon interaction with the drug conjugate complex (Fig. 5G). The analyses mentioned above can be attributed to a predictable conclusion that, when the small molecule inhibitor binds to the S protein, it significantly affects the binding to the ACE2 domain, which may be helpful to reduce the chances of the signal cascading process in host viral infection. Again, to confirm that we have also considered the FDA approved drug,

Zanamivir, for protein-protein interaction analysis. The analysis revealed that, when Zanamivir binds to the RBD-ACE2, the interaction remains in closed conformation with a score of -25.1 kcal/mol, which is more or less similar to the native complex of RBD-ACE2. (Fig. S5, G to I). The higher binding affinity reveals that ACE2 strongly binds to the RBD domain of trimeric S protein that can mediate the cascade of viral replication. Upon binding of antiviral drugs, the binding affinities were less in comparison to the native complex for which PC786 proved to be better from our analyses.

Immunoinformatics approach for designing T cell and B cell epitopes

In addition to the drug screening approach, integrating immunogenomics using a bioinformatics approach to design various treatments and finding potential candidates in the form of a drug or peptide for SARS-CoV-2 could help during this pandemic outbreak. This immunoinformatics approach could help in designing new vaccines and can overcome the problem of experimentation and time-consuming development process. The discovery of epitope is a key first step in antigen-targeted immunotherapy against viral diseases (26). Several studies have proposed strategies to achieve this in the last decade, particularly in the light of antiviral immunotherapy (27). The current strategy mainly focuses on identifying B cell epitopes using key structural proteins i.e., S and M^{pro} of SARS-CoV-2 (28). The epitopes identified using structure-based approach are expected to be most valuable leads. The implied *in-silico* method identifies human leukocyte antigens (HLA) and T cell epitopes (29). That specifies the selection of potent vaccine candidates associated with the transporter of antigen presentation (TAP) molecules. The designed peptides determined through the immunoinformatics approach served as a foreign substance for the human host cells, thus producing inflammation, demonstrating an allergic reaction. The present study was undertaken to design T cell and B cell epitope-based peptide vaccines against SARS-CoV-2 using the immunoinformatics approach. The approach can serve as a fast-track approach for experimentalists to validate the prediction. The query for SARS-CoV-2 structural and non-structural proteins resulted in a few pieces of evidence, and thus not much information was obtained. Since the S and M^{pro} proteins were available for SARS-CoV-2, structure-based epitope design was taken into consideration. Both the structural protein, i.e., S protein and the non-structural protein, i.e., main protease (M^{pro}) were antigenic in nature according to VaxiJen 2.0 (30). The server predicted both the proteins to be antigenic based on an overall protective antigen prediction score of 0.4512 and 0.4159, respectively, which is beyond the threshold 0.4. The NetCTL 1.2 webserver (31) was used for both the structures to predict CD8⁺ T-cell epitopes based on MHC binding affinity, C-Terminal Cleavage Affinity, TAP transport

efficiency, and NetCTL Prediction Scores (Fig. 6A). The study unraveled 6 T cell epitopes for the spike protein and 8 T cell epitopes for the main protease protein based on a combined score of NetCTL.

The selection is thus chosen to be the best T cell epitopes keeping in mind the MHC binding affinity and can interact with MHC alleles creating an effective immune response. The conservancy of the T cell epitopes is thus necessary to design effective vaccines and can provide immunization effectively. MHC class I immunogenicity score also revealed some significant insights to select the epitopes as high score designates the probability of eliciting an immune response. The selected best T cell epitopes with residue numbering, e.g., WTAGAAAYY, CVADYSVLY, TSNQVAVLY, KTSVDCTMY, STECSNLLL, LTDEMLAQY and SEDMLNPNY, TANPKTPKY, QTFSVLACY, GSVGFNIDY, DYDCVSFCY, GTDLEGNFY, TVNVLAWLY, LLEDEFPTPF of the spike (Fig. 6D) and M^{pro} (Fig. 6E) proteins respectively were further subjected for MHC class-I binding and processing analysis according to IEDB resources. The epitopes were also evaluated for MHC binding affinity, C-Terminal Cleavage Affinity, TAP transport efficiency, Proteasome score, and MHC(IC50) < 200nm for ensuring high immune response and higher affinity. For the processing of MHC-I, the IEDB analytics method produces an average score for the intrinsic ability of each epitope to be a T-cell epitope dependent on the proteasomal synthesis, TAP transport, and MHC-I binding capacity (supplementary Table S1 and S2).

Before being introduced to the T-cells on the cell's plasma membrane, the protein is reduced by the cytosolic proteases to tiny peptides in the proteasome and MHC-I also forms a complex with the peptides. Then, the MHC-I peptide complex is transferred by heat shock proteins and transport-associated proteins (TAP) to the endoplasmic reticulum (ER). However, the higher the overall score of the epitopes with the HLA alleles assures appearance to the T-cell, so it significantly relies on a positive immune response. Eliciting effective immune response relies not only on the positive identification by HLA molecules with substantial affinity to epitopes but also on the score to antigenicity and immunogenicity. Therefore, the epitopes which were identified by the vast number of HLA alleles and provided the highest immunogenicity, antigenicity and non-toxic to humans were known as the possible epitope for a robust immune response (supplementary Table S1 and S2). Distribution of MHC class I HLA alleles differs across various geographic areas and ethnic groups across the world. Therefore, when planning an efficient vaccine, account must be taken of population coverage. A substantial community density was noticed for the best selected epitopes selected in various geographic regions of the world for both the S and M^{pro} T cell epitopes (Fig. 6C). The S protein T cell epitopes were majorly found to be evident in the European

region (Fig. S6, A). In contrast, M^{pro} T cell epitopes are distributed majorly in the East Asian region (Fig. S6, B).

The B cell epitope is an appropriate portion of an antigen that is detected in a humoral reaction by either a particular B cell receptor or the elicited antibody (32). There are two major groups of B cell epitopes (i) the B cell epitope continuous or linear (ii) B cell epitope discontinuous or conformational (33–35). Most B cell epitopes have been shown to be conformational epitopes, and the amount of this epitope reaches 90 percent. The detection of antigenic protein B cell epitopes is the main phase in the design of epitope dependent vaccines. Based on B Cell epitope prediction methods e.g., BepiPred and Ellipro which are linear B cell epitope prediction methods combinedly predicted, 94 linear epitopes using the S protein sequence (Fig. 6I and Fig. S6 I and J) and 17 linear epitopes for M^{pro} protein (Fig. 6J and Fig. S6, L and M).

Discotope analysis gave propensity scores of the individual residues to be evaluated as a discontinuous B cell epitope as depicted in supplementary data file 4 and Fig. 6, K and L. Likewise, other B cell epitope prediction methods such as Chou & Fasman Beta Turn prediction, Emini Surface Accessibility prediction and Kolaskar & Tongaonkar Antigenicity also unraveled some B cell epitope regions which could be of a potential interest for researchers in designing new vaccines for SARS-CoV-2. Currently, the vaccinations are mainly based on immunity from B Cells. But recently, vaccine based on T cell epitope has been promoted as the host will produce a powerful immune response by CD8 + T cell against the infected cell (Fig. 6F). With time, due to antigenic drift, every foreign particle will escape the response of the antibody memory; however, the immune response of T cells also provides long-lasting immunity.

The predictive antigenicity system of Kolaskar and Tongaonkar assesses the antigenicity of a particular epitope. The antigenicity depends on the physiochemical properties of amino acids present in the epitopes. The average antigenic propensity of the protein was 1.037 with maximum of 1.261 and minimum of 0.866 in case of spike protein and average antigenicity of 1.042 with minimum 0.844 and maximum 1.220 was predicted for the protease protein (Fig. S6, E and H) respectively. The Chou & Fasman Beta turn prediction is a method useful for selecting protein regions to be synthesized in order to produce anti-peptide antibodies cross-reacting with the parent protein. As illustrated in Fig. S6, C and F, the predicted B cell epitopes can be experimentally synthesized with a beta-turn propensity in order to produce anti-peptide antibodies for both S protein and M^{pro}, respectively. The Emini Surface accessibility prediction of the B cell epitopes provides an evidence that the epitopes found on the surface can be easily accessed by the antibodies and thus elicits an effective immune response. Fig. S6, D and G represents the epitopes predicted above the threshold value with a

maximum propensity of 5.960 and 8.294 in both spike and protease proteins respectively.

MHC class I molecules and T cell epitope interaction analysis

In addition to the designing of T cell epitopes, we have predicted the 3-dimensional model of the selected epitopes both for S and M^{pro} proteins using PEP-FOLD (36) web-based server to analyze the interactions with particular HLAs respectively (Fig. 6F). To ensure the binding between HLA molecules and the predicted epitopes, the protein-peptide docking was performed using Cluspro2.0 (37) (Fig. 6B). The protein-peptide interactions revealed that T604-Y612 binds to the peptide-binding groove of MHC class I HLA-A*01:01 molecules with interaction energy of -289 kcal/mol (Fig. 6G). Similarly, the T cell epitope derived from M^{pro} i.e., G146-Y154 also predicted to show high binding affinity (-341.8 kcal/mol) toward MHC class I HLA-A*01:01 molecule (Fig. 6H). Based on IEDB MHC-I binding analysis and NetMHCpan, various HLA class I antigens were selected for the receptor molecules based on the availability of the PDB (Protein Data Bank structures). For the T cell epitopes derived from S protein, all the selected T cell epitopes bound to the peptide-binding groove (cleft) of the MHC class I (HLA-A) antigens (Fig. S7A). Once a peptide is bound to MHC-class I HLA complexes and presented on the cell surface, CD8 + T cell epitopes can be detected. Epitope conformation in the HLA-A and HLA-B groove are of paramount importance not only for epitope affinity and stability, but also for epitope interactions. Similar HLA based epitope interactions were studied for the M^{pro} proteins and found evident interactions in the peptide binding groove of MHC class I antigens (Fig. S7B). The corresponding epitope-HLA class I antigens docking energies have been illustrated in Fig. 6B. The corresponding interaction energies obtained for the S and M^{pro} based epitope interactions with HLA-A and HLA-B suggested that most of the epitopes have significant binding affinities toward HLA-A class I molecules. Although, detailed study based on the interactions have not been explored in this context.

Discussion

The quick and efficient development of active antiviral agents for therapeutic use is exceptionally challenging because traditional drug development methods usually take years of research and cost billions of dollars. Repurposing approved pharmaceutical drugs, and drug applicants provide an alternative approach to quickly identify potential medication leads and rapidly manage evolving viral infections. In the current study, the combined drug designing and immunoinformatics approach provide a detailed understanding of the vital structural domains involved in either acting as a substrate-binding site or epitope recognition site. Research teams in

companies and universities are currently developing more than 90 vaccines against SARS-CoV-2 (38). At least eight types of coronavirus are being tested, relying on different viruses or viral parts (38). The structure-based immunoinformatics approach may help identify vital structural domains and active sites that can provide a basis to develop protein-based vaccines targeting mainly spike glycoprotein of SARS-CoV-2. We do acknowledge several limitations to validate our in-silico proposed work due to the lack of experimental support. However, the strategies mentioned above can reduce the resources and expenses for researchers involved directly with the experimental and clinical studies with a higher probability of obtaining the desired responses and fewer trials and recurrences of mistake. Our strategy will reduce the translational distance between pre-clinical test results and clinical outcomes, which is a significant challenge in the rapid development of practical treatment approaches for the ongoing SARS-CoV-2 pandemic.

Materials and Methods

Protein structure retrieval

We have retrieved the crystal structures of prefusion SARS-CoV-2 spike glycoprotein with a receptor-binding domain (S) (PDB ID: 6VSB) and the main protease (M^{pro}) in complex with an inhibitor N3 (PDB-ID: 6LU7) from Protein Data Bank. For the complex interaction analyses with RBD domain of S protein, we have taken the structure of Native Human Angiotensin Converting Enzyme-Related Carboxypeptidase (ACE2) (PDB ID 1R42). We have used UCSF Chimera and Discovery Studio Visualizer to visualize and analyze the interactions.

Virtual Screening and Molecular Docking

The antiviral drug compounds were retrieved from the ChEMBL database with a search query term “antiviral drugs” and “coronavirus” that resulted in 640 chemical compounds with the corresponding filters for data availability, e.g., SMILES and Structure Data File formats. We have screened the antiviral drug molecules for RO5 violations and refined them to get 3D coordinates using the Open Babel command-line tool. All the structures that have passed the RO5 rule have been subjected for further refinement using Marvin Sketch. The drug compounds and the corresponding target receptors, e.g., 6VSB and 6LU7, were submitted for Virtual screening using Autodock Vina. The virtually screened best compounds were then docked with the target receptors again to ensure the conformations poses and binding affinities. We have performed blind docking as the location of binding site is unknown for both target receptors. The grid for the target receptors was set to 126Å x126Å x 126Å with a spacing of 1.000 Å. The interactions were visualized using Discovery Studio Visualizer. Furthermore, we have again docked the best-screened compounds to the complex of RBD-ACE2 in

two conformations (i) partial chain A- RBD domain of S protein with ACE2 complex (ii) full trimeric S protein with ACE2 complex using Autodock Vina.

Molecular Dynamics Simulation

Molecular dynamics simulation using GROMACS v.2019.2 has been performed for the complex molecules (drug bound proteins). We obtained the topologies for all the small antiviral molecules from the PRODRG database. We have optimized the parameters of the target receptor and the drug molecules using the GROMOS96 54a7 force field. The complex systems were placed in a periodic cubic box solvated with SPC solvent molecules. Periodic boundary conditions with a 15 Å cutoff for nonbonded interactions were applied, with the particle mesh Ewald (PME) method applied to account for the long-range electrostatic interactions. The system was neutralized with Na⁺ counterions to attain equilibration. Energy minimization and equilibration were carried out in three steps, as follows: (i) We minimize the whole system containing ions, solvent, protein, and ligand for up to 50 000 steps using a steepest-descent algorithm (ii) Constraints were added to protein and the ligand dimer for 100 ps during heating using NVT (No. of atoms, Volume, Temperature) ensemble with leapfrog integrator, and LINCS holonomic constrains. (iii) NPT (No. of atoms, Pressure, Temperature) ensemble was used at constant pressure (1 bar) and temperature (300 K) for 100 ps using a time step of 2 fs for equilibration phase 2. The SHAKE algorithm was used to constraint hydrogen to heavy atom bonds. The MD production phase for all the systems have been simulated for 10ns with a time step of 2 fs. Furthermore, after 10 ns simulation, the protein–ligand interaction energy was evaluated to compute the nonbonded interaction energy and short-range nonbonded energies, which were quantitatively reproduced with energy profiles generated by Gromacs tools. Furthermore, we employed MM/PBSA to calculate the polar, non-polar solvation energies with corresponding binding energy decomposition of the complexes. MM/PBSA calculates the free energy of the docked complex (the binding free energy of the protein with ligand in a solvent medium) where the general expression of the term can be depicted as:

$$\Delta G_{binding} = G_{complex} - (G_{protein} + G_{ligand}) \quad (1)$$

where, $G_{complex}$ is the total free energy of the protein–ligand complex and $G_{protein}$ and G_{ligand} are total free energies of the isolated protein and ligand in solvent, respectively.

$$G_x = (E_{MM}) - TS + (G_{solvation}) \quad (2)$$

where x is the protein or ligand or protein–ligand complex. (E_{MM}) is the average molecular mechanics potential energy in a vacuum. TS refers to the entropic contribution to the free energy in a vacuum where T and S denote the temperature

and entropy, respectively. The $\Delta G_{\text{solvation}}$ term is the free energy of solvation.

$$E = E_{\text{bonded}} + E_{\text{nonbonded}} = E_{\text{bonded}} + (E_{\text{vdW}} + E_{\text{elec}}) \quad (3)$$

where E_{bonded} is bonded interactions consisting of bond, angle, dihedral and improper interactions. The nonbonded interactions ($E_{\text{nonbonded}}$) include both electrostatic (E_{elec}) and van der Waals (E_{vdW}) interactions depicted using a Coulomb and Lennard-Jones (LJ) potential function, respectively.

Moreover, the free energy of solvation which is the energy required to transfer a solute from a vacuum into the solvent has been calculated including polar and non-polar solvation energies that can be depicted as

$$G_{\text{solvation}} = G_{\text{polar}} + G_{\text{nonpolar}} \quad (4)$$

where G_{polar} and G_{nonpolar} are the electrostatic and non-electrostatic contributions to the solvation free energy, respectively.

Protein-Protein interaction

To predict the conformational changes upon binding of ACE2 to the trimeric S protein RBD domain, we have used PatchDock and FireDock for protein-protein interaction analysis. UCSF chimera was used for the post protein-protein interaction analyses.

Antigenicity and T cell epitope identification

We have retrieved the protein FASTA sequence of Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome sequence bearing ID NC_045512.2 for the epitope screening. The prediction of protective antigens and subunit vaccines were evaluated using VaxiJen 2.0 with default parameters. NETCTL 1.2 server was used for the T cell epitope identification. The method integrates MHC-I binding, proteasomal C terminal cleavage, TAP transport and combinatorial scores for the prediction of epitopes. For MHC-I binding, Immune Epitope Database tools have been used to get best selected epitopes based on stabilized matrix base method (SMM) and inhibitory concentrations (IC50) values for peptide binding to MHC-class I molecules. Further, the selected epitopes were further processed to obtain the specificity to TAP transport, proteasomal cleavage, TAP transport and MHC-I. The web-based tool IEDB has been used for population coverage analysis.

Immunogenicity prediction

The IEDB MHC class I immunogenicity tool and EMBOSS has been used for immunogenicity prediction. The algorithm prediction was based on immunogenicity and antigenic scores.

T cell epitopes structure prediction

The selected T cell epitopes were subjected to PEP-FOLD server to predict the 3-Dimensional structure to be able to

perform the protein-peptide interaction with HLA-A and HLA-B class I molecules.

B cell epitope prediction

IEDB resources were used to classify B cell antigenicity such as Kolaskar and Tongaonkar antigenicity scale, Emini surface usability prediction, Karplus and Schulz versatility prediction, and Bepipred linear epitope prediction analysis. The Chou and Fasman beta turn prediction tool is used as the antigenic sections of a protein belong to the beta turn areas. We have also used Ellipro and Discotope to predict linear and discontinuous peptides, respectively.

Molecular interaction of epitopes to HLA class I molecules

The T cell epitopes were further processed for interaction analysis using HLA-Class I molecules using ClusPro 2.0. ClusPro 2.0 was based on ranking models by cluster size where the ligands were rotated 70,000 conformations. The server also predicts the cluster size, interacting members and give best models with lowest energies.

Statistical analysis

All statistical data analyses were performed in Origin 2018. Linear curve fitting has been performed using independent and dependent variables, with the goal of defining a "best fit" model of the relationship. Use of weighted least-square method to fit a linear model function to specified data has been performed. Box plots, scatter plots and bar graphs have been depicted to represent the data.

REFERENCES AND NOTES

1. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020). doi:10.1016/S0140-6736(20)30183-5 [Medline](#)
2. J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Y. Yip, R. W. S. Poon, H. W. Tsoi, S. K. F. Lo, K. H. Chan, V. K. M. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. C. Cheng, H. Chen, C. K. M. Hui, K. Y. Yuen. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *Lancet* **395**, 514–523 (2020). doi:10.1016/S0140-6736(20)30154-9 [Medline](#)
3. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. **367**, 1260–1263 (2020). doi:10.1126/science.abb2507
4. F. Li, W. Li, M. Farzan, S. C. Harrison. Structural biology: Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*. (2005). doi:10.1126/science.1116480.
5. T. M. Gallagher, M. J. Buchmeier. Coronavirus spike proteins in viral entry and pathogenesis. *Virology* **279**, 371–374 (2001). doi:10.1006/viro.2000.0757 [Medline](#)

6. J. Kirchmair, S. Distinto, K. R. Liedl, P. Markt, J. M. Rollinger, D. Schuster, G. M. Spitzer, G. Wolber, Development of anti-viral agents using molecular modeling and virtual screening techniques. *Infect. Disord. Drug Targets* **11**, 64–93 (2011). doi:10.2174/187152611794407782 [Medline](#)
7. Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L. W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature* **582**, 289–293 (2020). doi:10.1038/s41586-020-2223-y [Medline](#)
8. Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, F. Cheng, Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov.* **6**, 14 (2020). doi:10.1038/s41421-020-0153-3 [Medline](#)
9. A. R. Fehr, S. Perlman, Coronaviruses: An Overview of Their Replication and Pathogenesis. *Coronaviruses: Methods and Protocols* **1282**, 1–23 (2015). https://link.springer.com/10.1007/978-1-4939-2438-7_1
10. Z. Song, Y. Xu, L. Bao, L. Zhang, P. Yu, Y. Qu, H. Zhu, W. Zhao, Y. Han, C. Qin, From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses* **11**, 59 (2019). doi:10.3390/v11010059 [Medline](#)
11. X. Xue, H. Yu, H. Yang, F. Xue, Z. Wu, W. Shen, J. Li, Z. Zhou, Y. Ding, Q. Zhao, X. C. Zhang, M. Liao, M. Bartlam, Z. Rao, Structures of two coronavirus main proteases: Implications for substrate binding and antiviral drug design. *J. Virol.* **82**, 2515–2527 (2008). doi:10.1128/JVI.02114-07 [Medline](#)
12. J. Cohen, Can an anti-HIV combination or other existing drugs outwit the new coronavirus? *Science*. (2020), doi:10.1126/science.abb0659.
13. B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, X. Li, J. Xia, N. Chen, J. Xiang, T. Yu, T. Bai, X. Xie, L. Zhang, C. Li, Y. Yuan, H. Chen, H. Li, H. Huang, S. Tu, F. Gong, Y. Liu, Y. Wei, C. Dong, F. Zhou, X. Gu, J. Xu, Z. Liu, Y. Zhang, H. Li, L. Shang, K. Wang, K. Li, X. Zhou, X. Dong, Z. Qu, S. Lu, X. Hu, S. Ruan, S. Luo, J. Wu, L. Peng, F. Cheng, L. Pan, J. Zou, C. Jia, J. Wang, X. Liu, S. Wang, X. Wu, Q. Ge, J. He, H. Zhan, F. Qiu, L. Guo, C. Huang, T. Jaki, F. G. Hayden, P. W. Horby, D. Zhang, C. Wang, A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N. Engl. J. Med.* **382**, 1787–1799 (2020). doi:10.1056/NEJMoa2001282 [Medline](#)
14. W. Yin, C. Mao, X. Luan, D.-D. Shen, Q. Shen, H. Su, X. Wang, F. Zhou, W. Zhao, M. Gao, S. Chang, Y.-C. Xie, G. Tian, H.-W. Jiang, S.-C. Tao, J. Shen, Y. Jiang, H. Jiang, Y. Xu, S. Zhang, Y. Zhang, H. E. Xu, Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, eabc1560 (2020).
15. A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**(D1), D1100–D1107 (2012). doi:10.1093/nar/gkr777 [Medline](#)
16. O. Trott, A. J. Olson, AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010). [Medline](#)
17. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi:10.1002/jcc.20084 [Medline](#)
18. D. S. Biovia, H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, T. Darden, D. York, L. G. Pedersen, G. Bussi, D. Donadio, M. Parrinello, U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, M. Parrinello, A. Rahman, V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, H. J. C. Berendsen, J. P. M. Postma, W. F. Van Gunsteren, A. Dinola, J. R. Haak, R. W. Hockney, S. P. Goel, J. W. Eastwood, C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, Dassault Systèmes BIOVIA, Discovery Studio Visualizer, v.17.2, San Diego: Dassault Systèmes, 2016. *J. Chem. Phys.* (2000). 10.1016/0021-9991(74)90010-2
19. A. Grifoni, J. Sidney, Y. Zhang, R. H. Scheuermann, B. Peters, A. Sette, A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671–680.e2 (2020). doi:10.1016/j.chom.2020.03.002 [Medline](#)
20. R. Vita, J. A. Overton, J. A. Greenbaum, J. Ponomarenko, J. D. Clark, J. R. Cantrell, D. K. Wheeler, J. L. Gabbard, D. Hix, A. Sette, B. Peters, The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**(D1), D405–D412 (2015). doi:10.1093/nar/gku938 [Medline](#)
21. D. W. Brookes, M. Coates, H. Allen, L. Daly, S. Constant, S. Huang, M. Hows, A. Davis, L. Cass, J. Ayrton, I. Knowles, P. Strong, G. Rapeport, K. Ito, Late therapeutic intervention with a respiratory syncytial virus L-protein polymerase inhibitor, PC786, on respiratory syncytial virus infection in human airway epithelium. *Br. J. Pharmacol.* **175**, 2520–2534 (2018). doi:10.1111/bph.14221 [Medline](#)
22. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005). doi:10.1002/jcc.20291 [Medline](#)
23. R. Kumari, R. Kumar, A. Lynn: Open Source Drug Discovery Consortium, g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. *J. Chem. Inf. Model.* **54**, 1951–1962 (2014). doi:10.1021/ci500020m [Medline](#)
24. D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H. J. Wolfson, PatchDock and SymmDock: Servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**(Web Server), W363-7 (2005). doi:10.1093/nar/gki481 [Medline](#)
25. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo, Q. Zhou, Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. **367**, 1444–1448 (2020).
26. M. Tahir ul Qamar, S. Saleem, U. A. Ashfaq, A. Bari, F. Anwar, S. Alqahtani, Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: An immune-informatics study. *J. Transl. Med.* **17**, 362 (2019). doi:10.1186/s12967-019-2116-8 [Medline](#)
27. L. M. Walker, D. R. Burton, Passive immunotherapy of viral infections: 'super-antibodies' enter the fray. *Nat. Rev. Immunol.* **18**, 297–308 (2018). doi:10.1038/nri.2017.148 [Medline](#)
28. X. Tian, C. Li, A. Huang, S. Xia, S. Lu, Z. Shi, L. Lu, S. Jiang, Z. Yang, Y. Wu, T. Ying, Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect.* **9**, 382–385 (2020). doi:10.1080/22221751.2020.1729069 [Medline](#)
29. A. Patronov, I. Doytchinova, T-cell epitope vaccine design by immunoinformatics. *Open Biol.* **3**, 120139 (2013). doi:10.1098/rsob.120139 [Medline](#)
30. I. A. Doytchinova, D. R. Flower, VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics* **8**, 4 (2007). doi:10.1186/1471-2105-8-4 [Medline](#)
31. T. Stranzl, M. V. Larsen, C. Lundegaard, M. Nielsen, NetCTLpan: Pan-specific MHC class I pathway epitope predictions. *Immunogenetics* **62**, 357–368 (2010). doi:10.1007/s00251-010-0441-4 [Medline](#)
32. D. C. Parker, T cell-dependent B cell activation. *Annu. Rev. Immunol.* **11**, 331–360 (1993). doi:10.1146/annurev.iy.11.040193.001555 [Medline](#)
33. Y. El-Manzalawy, V. Honavar, Recent advances in B-cell epitope prediction methods. *Immunome Res.* **6**(Suppl 2), S2 (2010). doi:10.1186/1745-7580-6-S2-S2 [Medline](#)
34. J. L. Sanchez-Trincado, M. Gomez-Perosanz, P. A. Reche, Fundamentals and Methods for T- and B-Cell Epitope Prediction. *J. Immunol. Res.* **2017**, 2680160 (2017). doi:10.1155/2017/2680160 [Medline](#)

35. L. Potocnakova, M. Bhide, L. B. Pulzova, An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. *J. Immunol. Res.* **2016**, 6760830 (2016). [doi:10.1155/2016/6760830](https://doi.org/10.1155/2016/6760830) [Medline](#)
36. J. Maupetit, P. Derreumaux, P. Tuffery, PEP-FOLD: An online resource for de novo peptide structure prediction. *Nucleic Acids Res.* **37**, W498-503 (2009). [doi:10.1093/nar/gkp323](https://doi.org/10.1093/nar/gkp323) [Medline](#)
37. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**, 255–278 (2017). [doi:10.1038/nprot.2016.169](https://doi.org/10.1038/nprot.2016.169) [Medline](#)
38. E. Callaway, The race for coronavirus vaccines: A graphical guide. *Nature* **580**, 576–577 (2020). [doi:10.1038/d41586-020-01221-y](https://doi.org/10.1038/d41586-020-01221-y) [Medline](#)

ACKNOWLEDGMENTS

Funding: The authors acknowledge the financial support from the Swedish Research Council, (VR grant no. 2016–06014). SNIC and HPC2N have been acknowledged for providing computational resources. We also thank Suman Mishra, former Drug Safety Associate from IQVIA, India, for helping to assess the therapeutic usage of antiviral drugs. **Author contributions:** Conceptualization: P.K.P., R.A., M.N.A.; computational analyses: P.K.P.; writing (original draft preparation): P.K.P., R.A.; writing (review and editing): P.K.P., R.A., M.N.A., P.P., S.K.V., Y.K.M., H.G.R., W.L., M.S.; supervision: R.A., Y.K.M., H.G.R., M.S.; funding acquisition: R.A. **Competing interests:** The authors declare that they have no competing interests **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

advances.sciencemag.org/cgi/content/full/sciadv.abb8097/DC1

Submitted 19 March 2020

Accepted 27 May 2020

Published First Release 24 June 2020

10.1126/sciadv.abb8097

A Epidemiological Comparison of Respiratory Viral Infections

Disease	Flu	COVID-19	SARS	MERS
Disease Causing Pathogen	Influenza virus	SARS-CoV-2	SARS-CoV	MERS-CoV
Basic Reproductive Number R_0	1.3	2.0 - 2.5 *	3	0.3 - 0.8
Case Fatality Rate CFR	0.05 - 0.1%	~3.4% *	9.6 - 11%	34.4%
Incubation Time	1 - 4 days	4 - 14 days *	2 - 7 days	6 days
Hospitalization Rate	2%	~19% *	Most cases	Most cases
Community Attack Rate	10 - 20%	30 - 40% *	10 - 60%	4 - 13%
Annual Infected (global)	~1 billion	N/A (ongoing)	8098 (in 2003)	420
Annual Infected (US)	10 - 45 million	N/A (ongoing)	8 (in 2003)	2 (in 2014)
Annual Deaths (US)	10,000 - 61,000	N/A (ongoing)	None (since 2003)	None (since 2014)

* COVID-19 data as of March 2020.

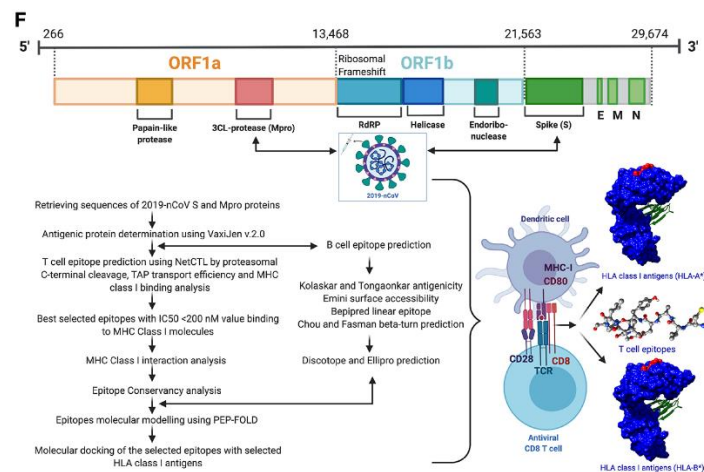
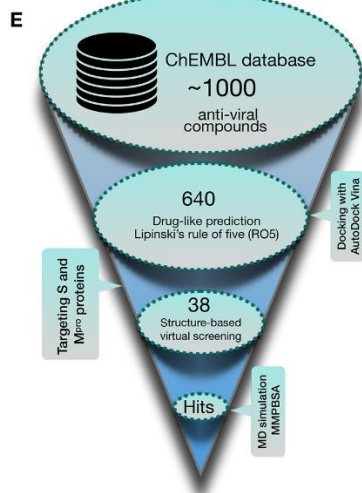
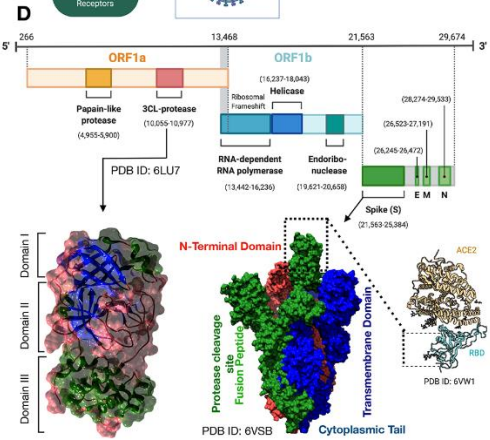
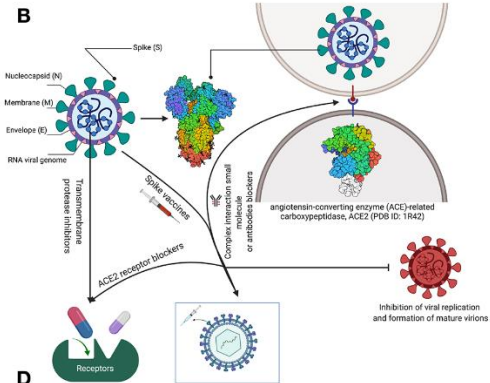
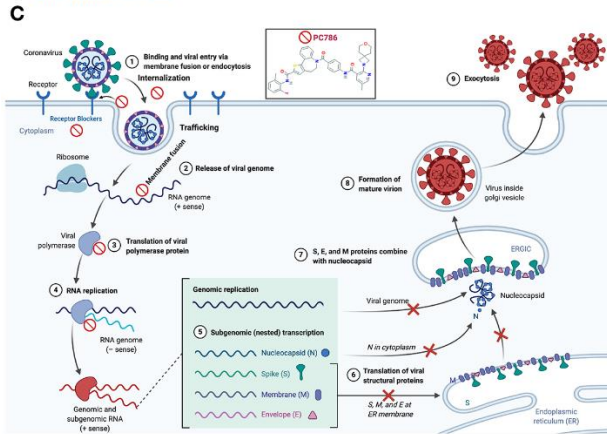


Fig. 1. Repurposed therapeutics for SARS-CoV-2 with broad-spectrum antiviral activities. (A) Epidemiological comparison of the recent outbreak SARS-CoV-2 with previous respiratory viral infections (B) In-silico proposed analyses targeting trimeric S protein, M^{pro} and trimeric S protein RBD-ACE2 complex (C) Potential repurposed drug candidates for SARS-CoV-2 targeting viral entry mechanism (D) Genomic organization of SARS-CoV-2 with structural domains representing M^{pro} and trimeric S protein (surface representation with the colors indicating secondary structure i.e., Blue: Beta strands, Green: Helix, Brown: Coils for M^{pro}) (Green: Chain A, Blue: Chain B, Red: Chain C for trimeric S protein). The RBD domain of trimeric S protein involved in the interaction with the human host ACE2 enzyme (E) Virtual library screening workflow discerning repurposed antiviral drugs targeting trimeric S and M^{pro} of SARS-CoV-2. (F) Approaches to predict potential vaccine candidates (T cell and B cell epitopes) for SARS-CoV-2. Schematic representation of the major histocompatibility I (MHC) class I displaying antigenic peptides to CD8 T cells. The surface representation of the human leukocyte antigen (HLA) MHC class I molecules bound to T cell epitopes (red color).

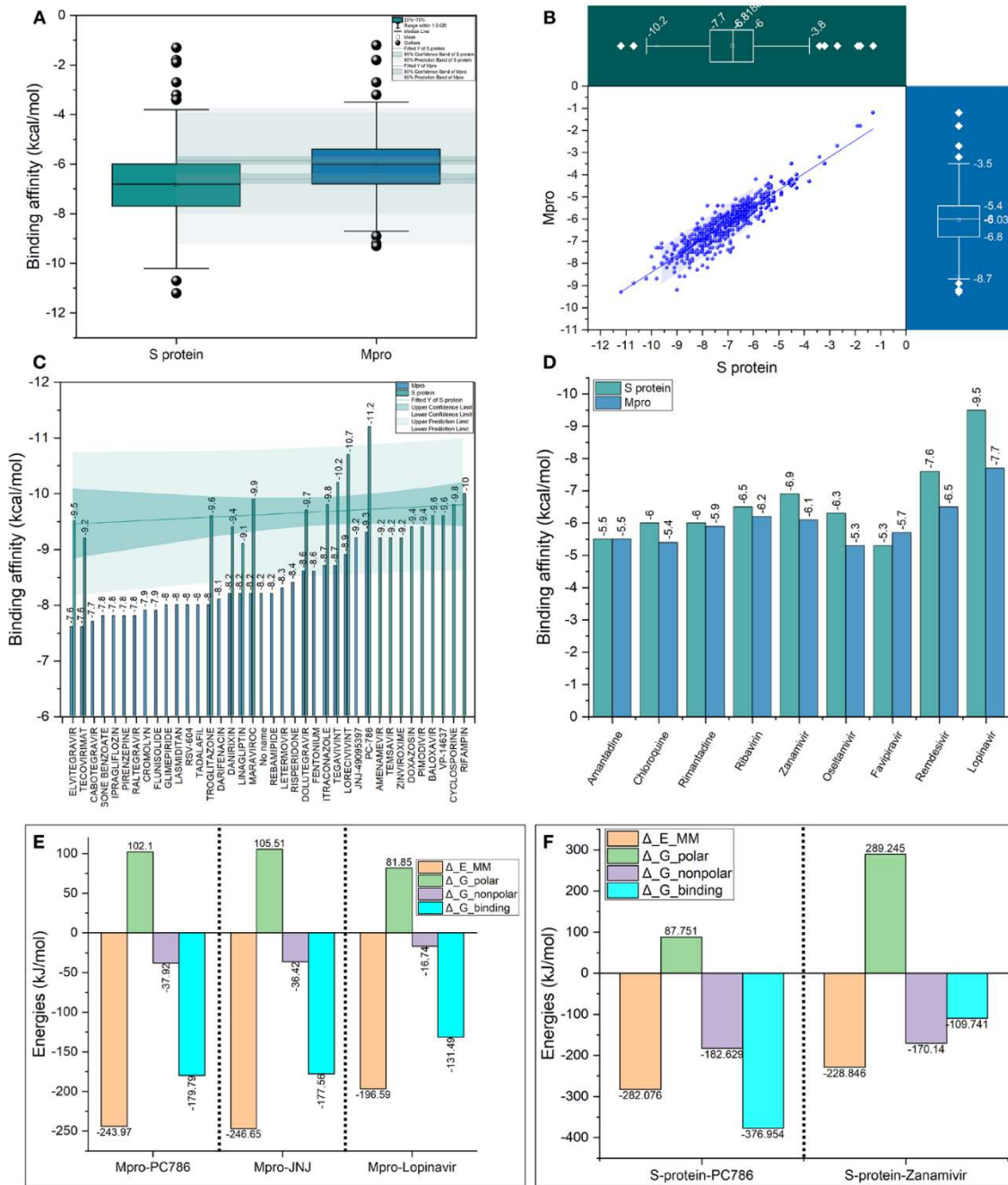


Fig. 2. Virtual screening of potential antiviral drugs against S and M^{pro} proteins of SARS-CoV-2. (A) Box plot with normal distribution with two different groups of the same size (n=640) showing the antiviral compounds binding affinities in kcal/mol for trimeric S protein (Green) and M^{pro} (Blue) with optimal means and 95% confidence intervals (CI). The outliers were shown in black spheres. (B) 2D scatter plot (blue color) showing the distribution of binding affinity scores in kcal/mol for all the virtually screened n=640 antiviral drugs against both trimeric S protein and M^{pro}. The two box and whisker plots (green color for S protein and blue color for M^{pro}) shows the distribution of binding affinity scores (kcal/mol) with median (m= -6.81 & -6.03 kcal/mol) for S protein and M^{pro} respectively. A horizontal box chart represents the S protein scatter data. A vertical box chart represents the M^{pro} scatter data. (C) Bar chart showing the binding affinity scores (kcal/mol) for the selected antiviral compounds ranging from -7.5 kcal/mol to -12.0 kcal/mol with 95% CI (green shade). The maximum binding affinity (high negative scores indicates maximum binding affinity). (D) Bar chart showing binding affinities scores in kcal/mol for selected approved drugs (E-F) Free energy terms obtained from MM-PBSA calculations relative to two selected drugs from virtual screening for M^{pro} and one for S protein in comparison to approved drugs.

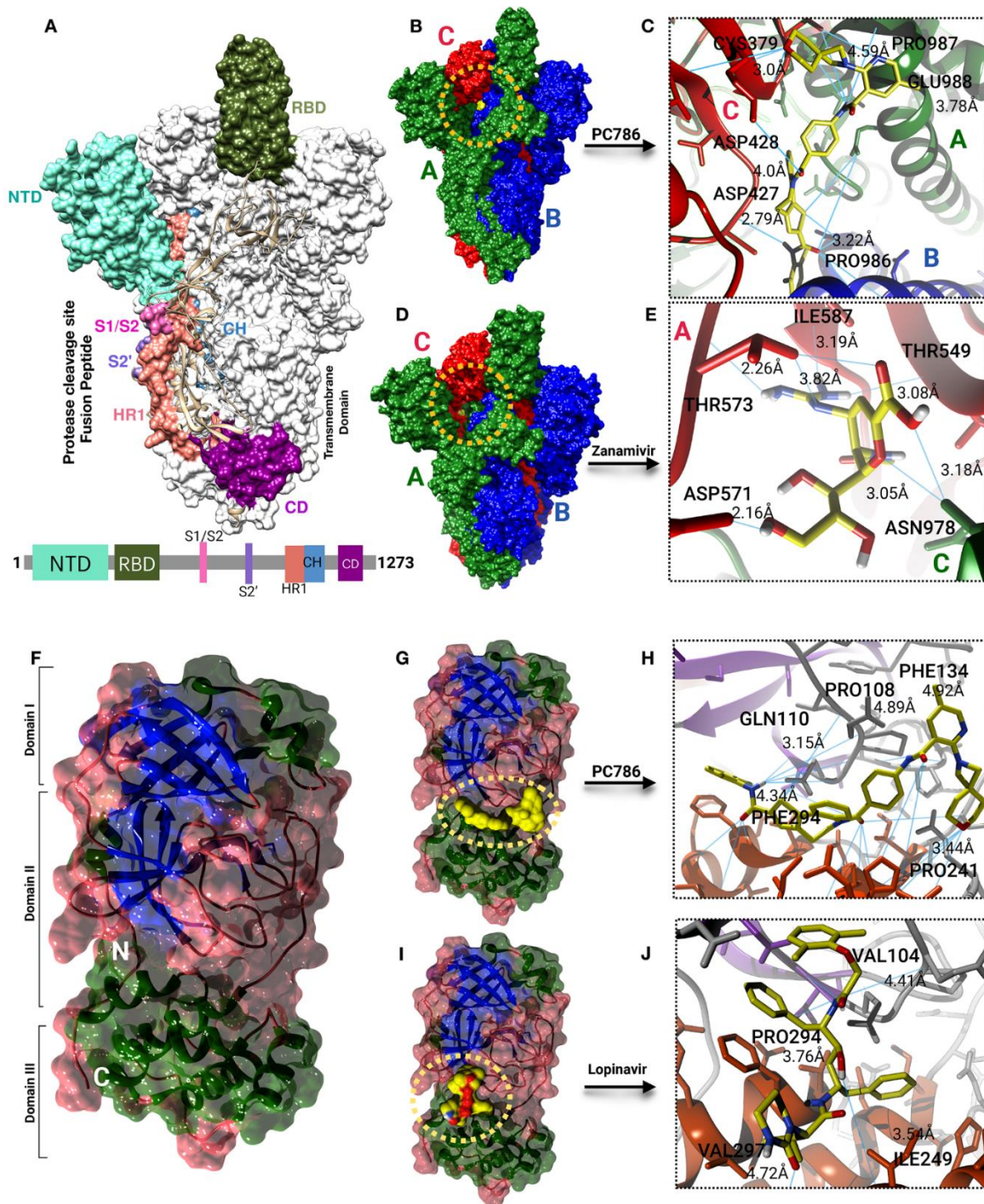


Fig. 3. Antiviral drugs binding modes for S and M^{pro}. (A) Surface representation of the trimeric S protein of SARS-CoV-2 with structural domains. Schematic of S protein primary structure colored by the domain (below). (B, D) PC786 & Zanamivir binding mode (yellow) to S protein (colors indicating chains). (C, E) Close-up view of PC786 (yellow surface) & Zanamivir (yellow sticks) binding to S protein chains (ribbons). Sky blue lines indicate hydrogen bonds. (F) Surface representation of the SARS-CoV-2 M^{pro} structural domains. The structure is represented by its secondary structure components (Blue: Beta strands, Green: Helix, Brown: Coils). (G, I) PC786 & Lopinavir binding mode (yellow) to M^{pro}. (H, J) Close-up view of PC786 (yellow surface) & Lopinavir (yellow surface) binding to M^{pro} (ribbons). Sky blue lines indicate hydrogen bonds.

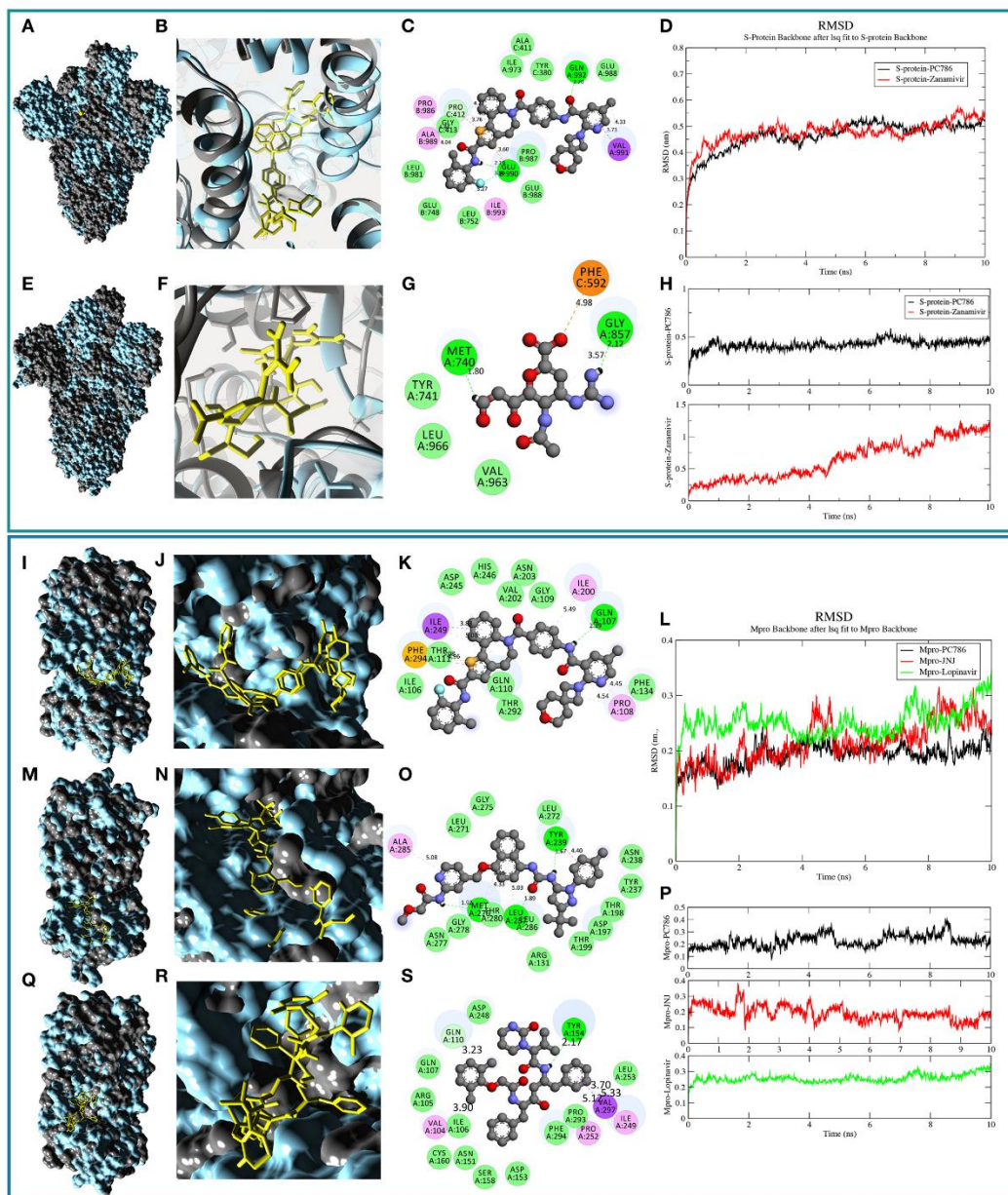


Fig. 4. Molecular dynamics (MD) simulation of the proposed antiviral drugs bound to trimeric S protein and M^{pro}. The green panel shows the MD simulation results for S protein drug conjugates and the blue panel represents the M^{pro} protein drug conjugates. (A, E) Surface representation of the trimeric S protein from 10ns simulation (Grey: initial configuration of trimeric S protein at 0ns, Cyan: 10ns simulated structure) (B, F) Conformations of the antiviral drugs with S protein i.e., PC786 and Zanamivir (yellow sticks) respectively, after 10ns simulation. (C, G) 2D representation of the drug interaction (key residues involved in the interaction mechanism. Green circles representing conventional hydrogen bonds, Pink: Alkyl bonds Violet: Pi-Sigma bonds Orange: Pi-Anion bonds, Yellow: Pi-Sulfur bonds). (D) Root Mean Square Deviation (RMSD) plot for the trimeric S protein during the 10ns simulation. (H) RMSD plot of the antiviral drugs (I, M, Q) Surface representation of the M^{pro} from 10ns simulation (Grey: initial configuration of M^{pro} protein at 0ns, Cyan: 10ns simulated structure) (J, N, R) Conformations of the antiviral drugs with M^{pro} i.e., PC786, JNJ & Lopinavir (yellow sticks) respectively after 10ns simulation. (K, O, S) 2D representation of the drug interaction (critical residues involved in the interaction mechanism. (refer (C, F) for legends. (L) RMSD plot for M^{pro} proteins conjugated with drugs following 10ns simulation. (P) RMSD plot of the antiviral drugs.

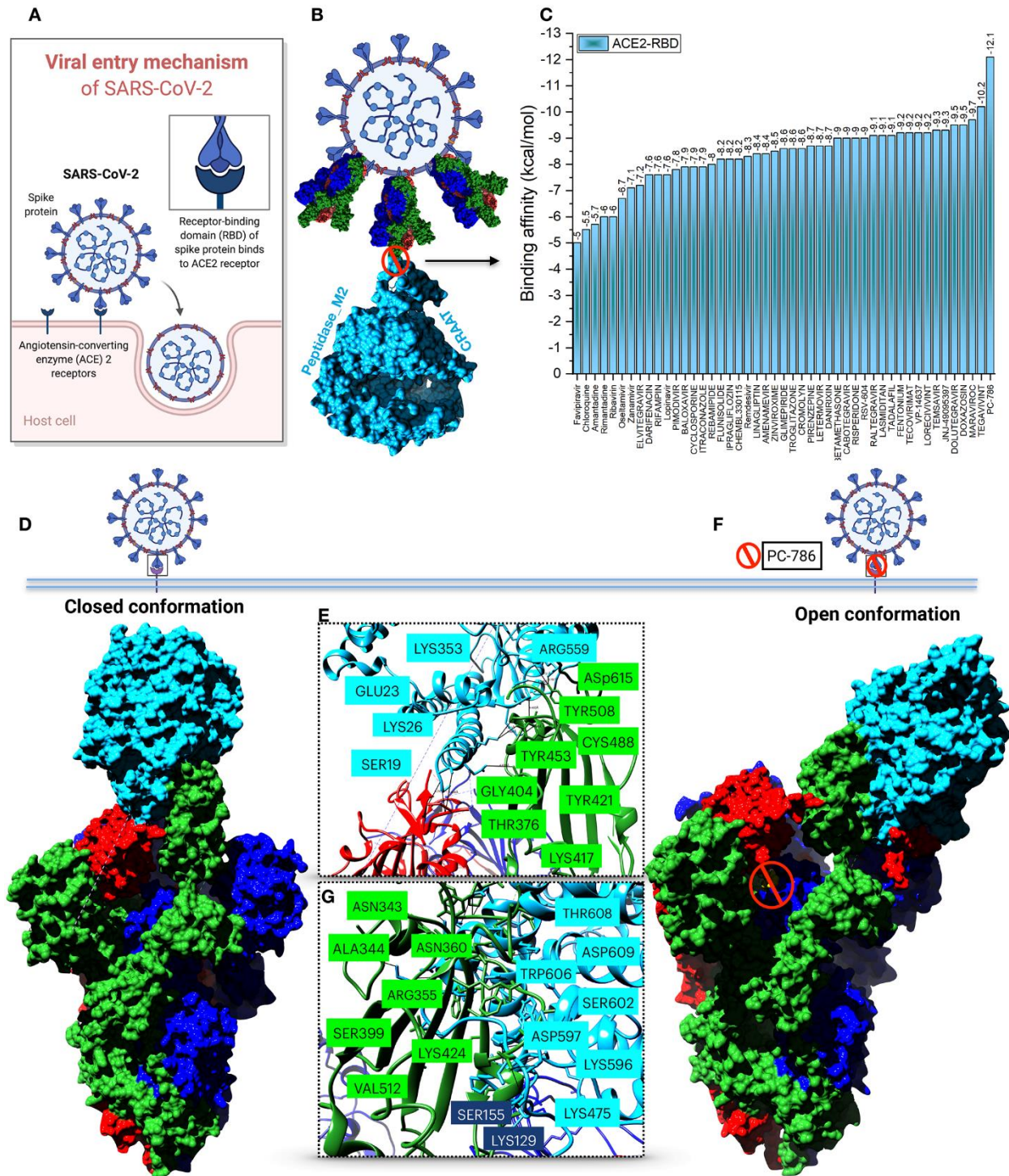


Fig. 5. Structural basis of the RBD-ACE2 complex protein-protein interaction (A) Schematic illustration of the viral entry mechanism of SARS-CoV-2. (B) Structural representation of the trimeric S protein RBD domain interaction inhibition with ACE2 by repurposed antiviral drugs. (C) Bar plot depicting binding affinities (kcal/mol) of selected antiviral drugs from virtual screening to RBD-ACE2 complex. (D) ACE2 binding to trimeric S protein RBD domain in a closed conformation. (E, G) Key residues involved in the interaction mechanism. Blue colored residues are from ACE2 enzyme and green shaded residues from trimeric S protein (F) Open conformation of antiviral drug (PC786) conjugate RBD-ACE2 complex.

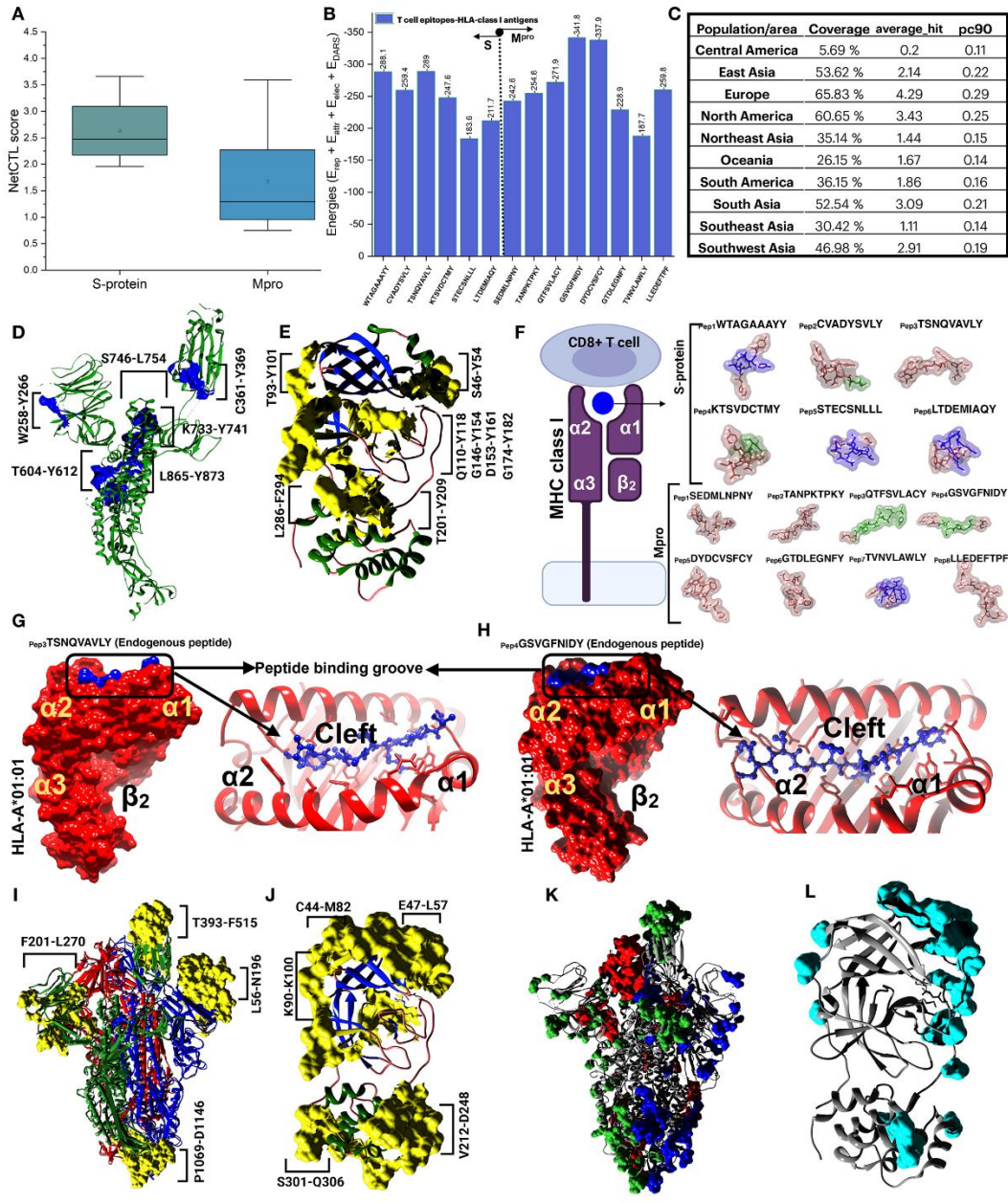


Fig. 6. Immunoinformatics approach for finding potential T cell and B cell epitopes (A) Box plot depicting NetCTL scores for predictions of Cytotoxic T lymphocyte (CTL) epitopes for rational vaccine design against SARS-CoV-2 S protein (Green) and M^{pro} (Blue). **(B)** Bar chart plot showing T cell epitope- HLA class I antigens interaction energy scores predicted using ClusPro 2.0 server. **(C)** Combined population coverage analysis of the T cell epitopes predicted from S protein and M^{pro} of SARS-CoV-2. **(D, E)** Cytotoxic T lymphocyte (CTL) epitopes (blue and yellow color) identified from S protein (green color: chain A of trimeric S protein) and M^{pro}, respectively. **(F)** Schematic of Cytotoxic T lymphocyte (CTL) epitopes binding to MHC class I molecules (representing peptide-binding groove present in between two α domains and one β_2 - microglobulin domain). 3D representation of T cell epitopes identified from the PEP-Fold server for S protein and M^{pro} of SARS-CoV-2, respectively. **(G, H)** The T cell epitopes e.g., TSNQVAVLY and GSVGFNIDY of S protein and M^{pro} interaction with HLA-A*01:01 molecules, respectively. The T cell epitopes bind in the cleft (peptide-binding groove) between the two α domains. **(I, J)** Identified linear B cell epitopes using BepiPred and Ellipro analyses. The B cell epitopes were represented in yellow color on both trimeric S protein and Mpro. **(K, L)** The predicted discontinuous epitopes (surface presentation) from Discotope analysis using trimeric S protein and M^{pro}.