Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

European Journal of Pharmaceutical Sciences 162 (2021) 105820

Contents lists available at ScienceDirect

European Journal of Pharmaceutical Sciences

journal homepage: www.elsevier.com/locate/ejps

In-silico screening for identification of potential inhibitors against SARS-CoV-2 transmembrane serine protease 2 (TMPRSS2)

Sagar Barge [a,1], Dhananjay Jade [b,1], Gokul Gosavi [c], Narayan Chandra Talukdar [a,d], Jagat Borah [a,*]

[a] Chemical Biology Lab I, Institute of Advanced Study in Science and Technology, Paschim Boragaon, Guwahati-35, Assam, India.
[b] JSS College of Pharmacy, Department of Pharmaceutical Chemistry, Ooty, 643001, Tamilnadu, India.
[c] Institute of Plant Protection Chinese Academy of Agricultural Sciences, Beijing, China
[d] Assam Down Town University, Panikhaiti, Guwahati, Assam 781006, India.

ARTICLE INFO

Keywords:
SARS-CoV-2
Transmembrane serine protease 2
Virtual screening
Molecular Docking
Drug repositioning

ABSTRACT

A new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a respiratory infection out broke in December 2019 in Wuhan, Hubei province, China, resulted in pandemic conditions worldwide. COVID-19 spread swiftly around the world over with an alert of an emergency for an adequate drug. Therefore, in this research, we repurposed the FDA-approved medicines to find the prominent drug used to cure the COVID infected patients. We performed homology modeling of the transmembrane serine protease 2 (TMPRSS2), responsible for the viral entry. The prediction of the transmembrane region and the Conserved Domain in TMPRSS2 protein was made for docking. 4182 FDA-approved compounds from the ZINC database were downloaded and used for the calculation of physicochemical properties. Two thousand eight hundred fifteen screened compounds were used for molecular docking against the modelled protein structure. From which top hit compounds based on binding energy were extracted. At 1st site pose, ZINC3830554 showed the highest binding energy -12.91kcal/mol by forming Salt Bridge at LYS143, Hydrogen bond at ALA8, VAL45, HIS47, SER142, ASN277, ASN359, and TRP363. The hydrophobic Interactions at PHE3, LEU4, ALA7, ALA8, ALA139, PRO197, and PHE266. In the 2nd site pose, ZINC203686879 shows the highest binding energy (-12.56 kcal/mol) and forms a hydrophobic interaction with VAL187, VAL189, HIS205, LYS301, GLN347, TRP370 and hydrogen bond was at GLY300, THR302, GLN347, SER350 residues. These hit compounds were subjected to stability checks between the protein-ligand complex through the dynamics simulation (MD), and binding free energy was calculated through the Molecular Mechanics energies combined with Poisson-Boltzmann (MM/PBSA) method. We hope that hit compounds would be an efficient inhibitor that can block the TMPRSS2 activity and resist the entry of the SARS-CoV-2 virus into targeted human cells by reducing the virus's infectivity and transmissibility.

1. Introduction

Coronaviridae is an enveloped, positive-stranded RNA virus tending to be circulated in the human population and causing severe, mild respiratory diseases (van der Hoek et al., 2006). The Middle East respiratory syndrome coronavirus (MERS-CoV) and Severe acute respiratory syndrome coronavirus (SARS-CoV) are types of coronaviruses that have the capability of transmission (Kannan et al., 2020). The SARS outcome started in the year 2002 from Guangdong with 8096 cases and 774 death. Due to the unavailability of potable drugs or vaccines against SARS, the pandemic was tried to be ceased by travel restriction and patient isolation (Berger and Preiser, 2011).

In December 2019, from Wuhan, China's seafood market, a new respiratory infection originated, named COVID-19 in March by the World Health Organization (Mackenzie and Smith, 2020). It had the potential of human-to-human transmission. It caused 139468 deaths in

Abbreviations: BE., Binding Energy; MD, Molecular Simulation; MMPBSA, Molecular Mechanics energies combined with Poisson-Boltzmann; nm, Nanometer; ns, Nano Seconds; PAINS, Pan Assay interference compounds; Rg, Radius of Gyration; RMSD, Root Mean Square Deviation; RMSF, Root mean square fluctuations; SASA, Solvent-Accessible Surface Area; Tc, Tanimoto; 2D, 2Dimensional.

* Corresponding author at: Institute of Advanced Study in Science and Technology (IASST), Guwahati-35, Assam.
E-mail addresses: borahjc@gmail.com, jagatborahiasst@gmail.com (J. Borah).

[1] Contributed equally

https://doi.org/10.1016/j.ejps.2021.105820
Received 12 December 2020; Received in revised form 2 March 2021; Accepted 21 March 2021
Available online 26 March 2021

the United States, after which it was considered an epicenter for virus infection as one-third of all diseases raised from the US. By the middle of 2020, it became defile for the human population by infecting 216 Countries worldwide by transmitting from one person to another through respiratory droplets. The minor symptoms showed by the infected person were Fever, Cough, Headache, Malaise, Sore throat, Vomiting, Nausea, Diarrhea, and trouble breathing (Zhang et al., 2020; Zhu et al., 2020).

There are three COVID vaccines' for which certain national regulatory authorities have authorized the use. In the United Kingdom, the BNT162b2 vaccine developed by Pfizer & BioNTech (Oliver et al., 2020) and the ChAdOx1 nCoV-19 (AZD1222) developed by the University of Oxford and AstraZeneca are currently in use (Voysey et al., 2021). These two vaccines have been authorized for emergency use by the Medicines and Healthcare products Regulatory Agency (MHRA), another third vaccine developed by Moderna will be available from Spring 2021 (https://vk.ovg.ox.ac.uk/vk/covid-19-vaccines). Before these three vaccines, the US Food and Drug Administration (FDA) has approved the Remdesivir first drug against COVID-19 (Rubin et al., 2020). This remdesivir is an antiviral medicine given an Intravenous therapy (IV) for patients needing hospitalization; this drug was California based Gilead Science Inc, which called Veklury, which reduce the time to recover from 15 days to 10 days shown by the US National Institutes of Health (Liang et al., 2020) (https://www.nih.gov/news-events/news-releases/nih-clinical-trial-shows-remdesivir-accelerates-recovery-advanced-covid-19). Pfizer vaccine shows the efficacy of 95% at preventing symptomatic COVID-19, which is measured starting from seven days after the 2nd dose was administered (Polack et al., 2020; Rubin and Longo, 2020). The Moderna vaccine shows 94.1% effectiveness at preventing symptomatic COVID infection, which is measured starting from 14 days after the 2nd dose (Mahase, 2020).

Recently scientists have found a new variant of coronavirus; due to this new strain of the COVID -19, coronavirus has emerged in the UK, spreading fast than the previous strain. This new strain (VUI-202012/01 variant) common in London, the South East and the East of England. After identifying the new strain Pfizer and Oxford vaccines were started, there was a new strain, existing coronavirus treatments. Studies of the new strain so far suggest it has 17 differences from previous strains. That includes a change to the Spike on the surface of the virus that may make it easier to get inside our body cells. Changes in the virus are very common; in new strains formed due to the virus entering the body, it multiplies very fast. With this quick multiply process, there is a chance that the virus may again change this quick-change process called a mutation. Due to only a few vaccines' availabilities or drug and mutation into COVID-19, scientists are still trying to discover more effective treatments by combining antiviral and convalescent plasma therapy.

Along with this, repurposing existing drugs is also used to provide a valuable medicinal procedure, which is achieved by inhibiting the different mechanisms responsible for COVID-19 infection (Baron et al., 2020; Elfiky, 2020; Fan et al., 2020; Zhou et al., 2020). Zygotic induction is one mechanism in which virus DNA is transferred from the virus and replicates into the host, leading to an increase in the virus's infection. Several enzymes like Furin, trypsin, and other proprotein-convertases, cathepsin, transmembrane proteases (TMPRSS), and elastases are involved in coronavirus cell entry (Matsuyama et al., 2018; Millet and Whittaker, 2015). The human entry is more frequent due to viral spike proteins on the surface of coronavirus, which binds to the carboxypeptidase angiotensin-converting enzyme-2 (ACE2) receptor present on human cells. The spike protein has two subunits S1 and S2, in which the S1 subunit contains the receptor-binding domain, which helps in binding to a cellular receptor, and another subunit, S2, which is a membrane-anchored subunit responsible for infusion of virus and hosts membrane (Bertram et al., 2011a; Glowacka et al., 2011; Shulla et al., 2011).

In the preliminary studies, serine protease anchored on the plasma membrane helps the virus enter and cause infection in the host (Millet and Whittaker, 2015). The S1 subunit initially binds to the cellular receptor, then priming S protein by the host transmembrane serine protease 2 (TMPRSS2) that breaks out the viral S protein at right upstream of the fusion peptide, which helps in membrane fusion via irreversible conformational changes (Bertram et al., 2011b; Hoffmann et al., 2018). Inhibition of the TMPRSS2 can help to avoid the entry of the SARS-CoV-2 virus into the human cells. Some earlier study also shows that a decrease in expression and activity of TMPRSS2 is a safe and effective method for the treatment of viral infection caused by viruses (Hoffmann et al., 2020; Matsuyama et al., 2020). Therefore, many researchers are involved in the discovery of efficient medicament for inhibiting the TMPRSS2 of coronavirus. Camostat, Nafamostat, and Aprotinin are some of these TMPRSS2 preceding inhibitors, which had reduced the rate of infection of the lung cell line (Calu-3) by SARS-CoV2 (Azimi, 2020; Hoffmann et al., 2020).

In this study, we focused on the inhibitor of TMPRSS2, a host LDLRA domain region (1st docking site) and serine protease (2nd docking site), which is less prone to mutation over time compared to viral protein using the 3D homology model of TMPRSS2 was predicted using Modeller software. The FDA-approved drugs were downloaded from the ZINC database and screened against the modeled structure of TMPRSS2. The binding site prediction was also made performing Molecular Docking. Further, the protein-ligand stability was studied using Molecular dynamics (MD) simulation and Free Energy Calculation (**Fig. 1**).

2. Materials and methods

2.1. Retrieve and analysis of sequence

In this study, the inhibitor was proposed through the *in-silico* approach for COVID19. Hence, the protein sequence was retrieved from the UniProt database (Consortium, 2018) by searching "Transmembrane protease serine 2". From hits, we selected the TMPRSS2 gene of *Homo sapiens* (Human).

2.2. Prediction of the transmembrane region and physicochemical parameters

The transmembrane part of the selected TMPRSS2 sequence was predicted through TMHMM Server v. 2.0 (Krogh et al., 2001). After that, prediction of Physicochemical parameters of TMPRSS2 such as Extinction coefficient, molecular weight, theoretical isoelectric point (pI), half-life, aliphatic index, instability index, the total number of positive and negative residues, and grand average hydropathy (GRAVY) was computed using Expasy's ProtParam (Gasteiger et al., 2005).

2.3. Finding conserved domain and motif

The Domain site in the TMPRSS2 gene was found by submitting the amino acid sequence to the PROSITE tool of ExPasY (de Castro et al., 2006). This tool provided resources for identifying and annotating conserved sites on the protein, covering protein families, domains, and motifs. Along with this, PROSITE, which identified the protein domain, families, and functional sites, was also used. A motif site on the TMPRSS2 sequence was also predicted through the online motif tool (https://www.genome.jp/tools/motif/).

2.4. TMPRSS2 model generation through modeller

For homology modeling, the TMPRSS2 protein sequence was used to construct a 3D structure by the modeller. The BLAST search was done for the TMPRSS2 protein sequence, which was retrieved from the UniProt database. Two templates were selected based on similarity score and maximum identity. The model was generated using Modeller9.12 (Eswar et al., 2006). TMPRSS2 sequence was put into the PIR format that is readable by the modeller. Subsequently, a search for potential was

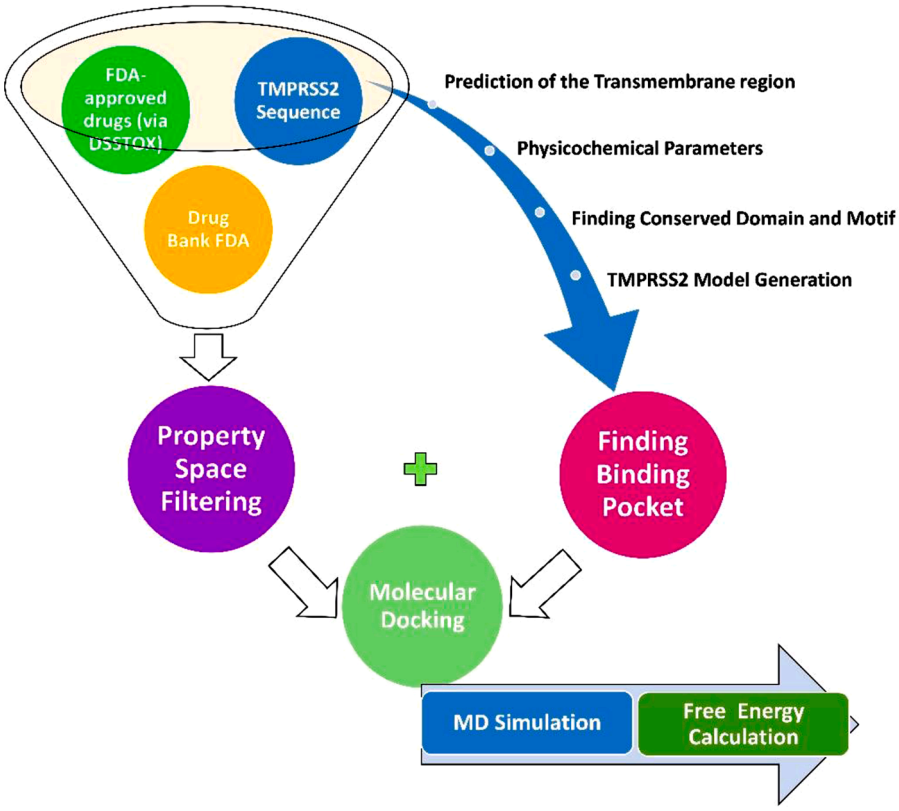

Fig. 1. Description of the *In-silico* virtual screening workflow of the iterative process to find a potent inhibitor for TMPRSS2.

aligned with the template and constructed the model.

2.5. Analysis of constructed model

After the model's construction, the DOPE score was checked, and the best model was selected, which were analyzed by ERRAT and Ramachandran plot. ERRAT and Ramachandran plot was calculated by SAVES v5.0 online program (Colovos and Yeates, 1993; Laskowski et al., 1996). This Ramachandran plot is used to check the protein conformation of φ and ψ angles, are possible for amino acid residue in protein and show the empirical of data points observed in a single structure in usage for structure validation or else in a database of many structures and usually shown against for the theoretically favored regions. ERRAT score is a program for verifying protein structure determined by crystallography.

2.6. Finding binding pocket

The Binding pockets finding was achieved with the MetaPocket 2.0 open Meta serve (Huang, 2009; Zhang et al., 2011). Binding sites on the protein surfaces play an essential role in protein function. It is based on a consensus method. The predicted binding sites from eight modes: LIGSITEcs, PASS, Q-SiteFinder, SURFNET, Fpocket, GHECOM, ConCavity POCASA are combined to improve the prediction. To find the binding pocket, we used the model's selected model and used it as an input query. Along with these pockets, we use the binding site residues extracted from literature, and these amino acids are Q185, H205, E208, K209, P210, K249, K251, E298, K299, L328, Q347, S350, W370 present in an active or catalytic pocket of TMPRSS2. A Triad of catalytically active site was included in this study, which contains H205, I255, D254, and S350 amino acid (David et al., 2020; Hussain et al., 2020; Vishnubhotla et al., 2020). Also, we verified the binding site identified through Metapocket by using the PDBsum tool and phyre2 tool.

2.7. Preparing FDA approval compounds

ZINC Database (Sterling and Irwin, 2015) has a wide range of compounds from which we used FDA approval compounds DrugBank FDA only (https://zinc.docking.org/catalogs/dbfda/), and FDA-approved drugs (via DSSTOX) (https://zinc.docking.org/catalogs/fda/) were downloaded from the ZINC database and used for high throughput virtual screening against TMPRSS2 protein of COVID-19.

2.8. Property space filtering

To assess drug-likeness, through Lipinski rule and verber rules were performed for downloaded compounds. This property space filtering is based on physicochemical parameters. This Lipinski rule (Benet et al., 2016; Lipinski et al., 1997) and verber rules (Veber et al., 2002) consist of Molecular weight (MW), LogP, Hydrogen Bond Accepter (HBA), Hydrogen Bond Donor (HBD), Topological polar surface area (TPSA) and rotatable bonds. MW, logP, HBA, and HBD belong to Lipinski rules and rotatable bonds, and TPSA includes verber rules. Along with these parameters, Heavy Atom and Aromatic Ring were included in this study. ChemAxon InstantJChem software (https://chemaxon.com/products/instant-jchem) was used for ligand preparation, such as adding a hydrogen atom, converting it into the 3D format, and calculating the Physicochemical properties; the physicochemical cut-off range was decided based on Lipinski rule and verber rules.

2.9. Molecular docking of FDA approval compounds

FDA approval Compounds used for docking through AutoDock Tools (Morris et al., 2009). A three-dimensional (3D) structure of the compounds was generated, and hydrogen was added by employing (command line) ChemAxon software-molconvert (Jade et al., 2020). All input file was prepared by using AutoDock Tools (ADT) 1.5.4. Addition of Polar Hydrogen in protein and assigned charges by Kollman charges

method. All docking calculation parameters were kept as a default value. Ligands were docked using the Lamarckian Genetic Algorithm, with an initial population of 150 randomly placed individuals, a maximum number of 2500000 energy evaluation, a mutation rate of 0.02, and a crossover rate of 0.8 and a total of ten conformations were generated for each screened compound. We performed docking on two sites, the first site was the binding pocket found by the Metapocket tool, and 2nd site was the substrate binding site & triad of the active catalytic site. For 1st Metapocket pocket site, we use the number of a grid point in XYZ 80 × 74 × 70 (x, y, and z) and grid box center is -15.997 × 24.256 × -24.858 was assigned and for 2nd substrate binding site and triad of the active catalytic site, number of a grid point in XYZ 110 × 88 × 98 (x, y, and z) and grid box center is -18.418 × (-0.107) × (-9.329) was assigned on the protein with the spacing 0.375Å. Autogrid4 was used to calculate grid maps, and Autodock4 was used to perform docking procedures. Generate the lowest binding energy conformation and find the molecular interaction between receptor and ligands using PLIP. After that, we check the binding poses of selected active hit compounds with TMPRSS2 protein on the 1st and 2nd site based on Binding energy and most populous cluster.

2.10. ADME prediction and PAINS filter analysis

For ADME Prediction, we used SwissADME online software (Daina et al., 2017) for the prediction of pharmacokinetic parameters of a drug-like compound through a molecular structure called ADME Absorption, Distribution, Metabolism, and Excretion. Compounds with low toxicity and high bioactivity by a drug-like compound are not the right criteria to qualify the compound as a promising drug. Finding a good drug is very important to calculate the ADME properties for new drug-like compounds. After that, we perform PAINS (pan assay interference compounds) filter analysis. PAINS is a prominent source to remove false positive compounds with poor pharmacokinetics and toxicity. We apply PAINS using PAINS-Remover (Baell and Holloway, 2010) (http://cbligand.org/PAINS) to remove scrutinize the false positives from selected compounds.

2.11. Clustering for a selected compound

After docking and ADMET calculation, the virtual hit compounds were further subjected to Clustering. Clustering was performed using Chem BioServer online tool (Karatzas et al., 2020) based on the Hierarchical Clustering method. The hit compounds were used for Clustering through Soergel (Tanimoto Coefficient) Distance method and Ward linkage Clustering method with clustering Threshold was 0.9. After this, we check the structural difference between known Camostat and other hit compounds using MolAlign Server based on property-based small molecule alignment; for this alignment, we used full-flexible alignment (Brown et al., 2019).

2.12. Molecular dynamics simulation and free energy calculation

MD simulation was performed by GROMACS-4.6.5 (Abraham et al., 2015), and protein topology was generated by GROMOS 54A7 force field. The initial orientation of the ligand towards the protein was obtained from the docking studies. The topologies for the ligands were created using PRODRUG (Schüttelkopf and van Aalten, 2004). The protein-ligand complex was put in a triclinic box, and the complex structure was solved with simple point charge (spc216) water. The neutralization of the system was done by adding ions and relaxation through the energy minimization process. The electrostatic interaction was estimated by using the PME algorithm. MD simulations with reasonable initial velocity follow the steepest descent path on the potential energy surface to a local minimum. The temperature and pressure equilibrium step of 1ns was performed before the 20ns production simulation. The root means square deviation (RMSD), root mean square fluctuation (RMSF), Radius of gyration (Rg) were calculated using g_rms, g_rmsf, and g_gyrate.

Along with this, the calculation of the hydrogen bond formation between protein-ligand, protein solvent was calculated. After that, we calculate the Principle compound analysis. We check the change in the protein's secondary structure and complex with time through using the do_dssp program (Kabsch and Sander, 1983; Touw et al., 2015). After the simulation, we estimated the ligand-protein complex's binding free energy using Molecular Mechanics energies combined with Poisson-Boltzmann (MM/PBSA) (Genheden and Ryde, 2015; Kumari et al., 2014). In this study, we calculated the protein-ligand complex binding fee energy using the MM-PBSA method. Calculation of Free binding energy in drug discovery plays a vital role, which gives the quantitative estimation of binding free energy. We use the 20ns MD trajectory to calculate MM-PBSA using the g_mmpbsa Tool (Kumari et al., 2014).

The protein-ligand complex interaction analysis and visualization were done through AutoDock ADT and PyMOL. Docking conformations and MD simulation results were generated using PyMOL. 2D graph of RMSD, RMSF, the radius of gyration, hydrogen bonds, and protein secondary structure plot was generated using xmgrace.

3. Result and discussion

3.1. Calculation of physicochemical and secondary structure parameters

The amino acid sequence was retrieved from UniProt ID: C9JKZ3, which belongs to the Transmembrane protease serine 2 (TMPRSS2) gene of Homo sapiens (Human), as shown in **Supplementary Figure 1**. This TMPRSS2 consists of 489 amino acids. The selected sequence of TMPRSS2 protein sequence was analyzed for prediction of the transmembrane region through TMHMM Server, which shows one transmembrane helix region consists of 21.8427 amino acids, 1-83 amino acids present inside, 84-106 amino acids consist of transmembrane helix region and 107-489 amino acids present outside the cell wall shown in **Supplementary Figure 2**.

For this study, the outer region and some helix regions with 416 amino acids were used for further research. Using this selected sequence, we calculated the physicochemical properties. The results revealed that the total length of protein as 416 amino acids, with the protein's average molecular weight, was 45537.92 Daltons. The sequence has about 31 negatively charged residues (Aspartic acid + Glutamic acid) and 36 positively charged residues (Arginine + Lysine). pI of Protein was 8.29, a pH at which the surface of protein was covered with charge, but the protein's net charge would be zero. Based on the instability index, Expasy's ProtParam classified the optimized protein as stable by showing the Instability index value as 36.97. The estimated half-life of the selected protein was calculated; in the N-terminal of the sequence considered is G (GLY). The apparent half-life is 30 hours (mammalian reticulocytes, in vitro), >20 hours (yeast, in vivo), >10 hours (Escherichia coli, in vivo). The aliphatic index of protein was 76.10. The grand average of hydropathicity (GRAVY) is -0.172. The extinction coefficient of optimized protein in the water at 280nm was 106100 M^{-1} cm^{-1}.

The FoldIndex program predicted the folding state of the TMPRSS2 protein sequence, which estimates the probability of a protein sequence folding under specified conditions. The TMPRSS2 sequence showed a small unfolding of the sequence (0.176%). **Supplementary Figure 3** shows the positive and negative numbers representing ordered or folded (Green) region and disordered or unfolded (Red) protein. Amino acids are suggested as ordered, shown in green and disordered in red, respectively (**Supplementary Figure 3**). Fifty-five amino acids belonged to the disordered region at 76-130, responsible for unfolding the protein.

3.2. Analysis of conserved domain

The PROSITE tool of ExPasY was used to determine the domain in the TMPRSS2 Amino acid sequence. In a selected sequence of TMPRSS2, we found the six hits, among which the three hits belong to distinct profiles that consist of LDL-receptor class A (LDLRA) domain profile-2, SRCR domain profile (SRCR_2) and Serine proteases, trypsin domain profile (TRYPSIN_DOM), shown in **Supplementary Figure 4**. LDLRA domain consists of a region 112-149, which shows a 10.662 score. These domains contain disulfide bonds, shown in **Supplementary Figure 4**. SRCR domain profile consists of a region 158-189, which shows an 8.560 score. Serine proteases TRYPSIN_DOM show 36.611 scores in region 256-489. The other three hits have distinct patterns LDL-receptor class A (LDLRA) domain signature at region 126-148, (TRYPSIN_HIS) Serine proteases, trypsin family, histidine active site at 292-297 region, and (TRYPSIN_SER) Serine proteases, trypsin family, serine active site at 435-446 shown in **Supplementary Figure 4**.

3.3. Construction of model throughmodeller

MODELLER is a computer program used for comparative protein structure modeling. TMPRSS2 amino acid sequence was used for searching the template through default modeller settings. The input sequence was aligned to be modeled with the template structures, which automatically generated the model. The top two templates 5TJX and 1Z8G, were selected based on maximum Query cover and identity similarity, based on a sequence of identity value and query cover. They showed 48% and 69% query cover and identity similarity score 42.74 and 33.52, respectively. The alignment of the query sequence with two selected target sequences for alignment is shown in **Supplementary Figure 5**. After alignment, we generate 20 models of TMPRSS2. From the 20 models, the 14th best model showed a good DOPE score -38215.922Kcal/mol. This final model was further used for further studies.

3.4. Analysis of constructed model

The modeled structure was analyzed by the ERRAT score and Ramachandran plot generated by SAVES v5.0. Regions of the structure that can be rejected at the 95% confidence level are Grey color; 5% of a good protein structure is expected to have an error value above this level. Regions that can be rejected at the 99% level are shown in black. The model ERRAT overall quality factor score is 44.324, which revealed that the model's nonbonded interactions lie within a reasonable normal range shown in **Supplementary Figure 6**.

After ERRAT, we calculated the Ramachandran plot, suggesting that 86.2% of residues are in the most favored region. The additional allowed region shows 11.1%, the generously allowed region show 2.3% and 0.3% of the modeled amino acids have disallowed geometry, as shown in **Supplementary Figure 7**. The obtained results suggest that the predicted protein model is reliable and suitable for further studies. The predicted final model was shown in **Supplementary Figure 8**.

3.5. Finding an active binding site

MetaPocket 2.0 server predicted a total of Six binding pockets, as shown in **Supplementary Figure 9 (A)**. In the MetaPocket server, our protein successfully ran on the CON, FPK, GHE, LCS, PAS, SFN method, but the QSF and PCS method failed among the total methods. **Supplementary Figure 9 (B)** shows the top 3 binding sites and **Supplementary Figure 9 (C)** shows the top 3 ligand binding sites surface in MetaPocket. MetaPocket shows six pockets, which shows Z-Score 19.88. To validate the above-predicted result through docking analysis. In docking studies, we perform docking on two different sites 1st site Fig. 2 **(A)** contain pocket 1, pocket two found in Metapocket tool (LDLRA domain region) and 2nd site Fig. 2 **(A)** is an amino acid is taken from literature which is present in a substrate binding site and Triad of the active catalytic site; these amino acids are H205, I255, D254, S350 (Triad of the active catalytic site) (magenta color) and Q185, H205, E208, K209, P210, K249, K251, E298, K299, L328, Q347, S350, W370 (active or catalytic site) (yellow color) show in Fig. 2 **(B)**. We also run the phyre2 and PDBsum online tool to confirm the predicted pocket. This tool shows the same binding cavity site shown by Metapocket. We performed more analysis to check the binding site and compared identified binding site. In this study, we select the binding site based on the Metapocket tool; for confirmation of this binding site, we use other tools to find binding pocket/cavity; **Supplementary Figure 10** shows the binding, catalytic sits and pocket (red color top, binding pocket) find by PDBsum tool (**Supplementary Figure 10 (A)**), phyre2 (**Supplementary Figure 10 (B)**) and Metapocket (**Supplementary Figure 10 (C)**) (Circled in

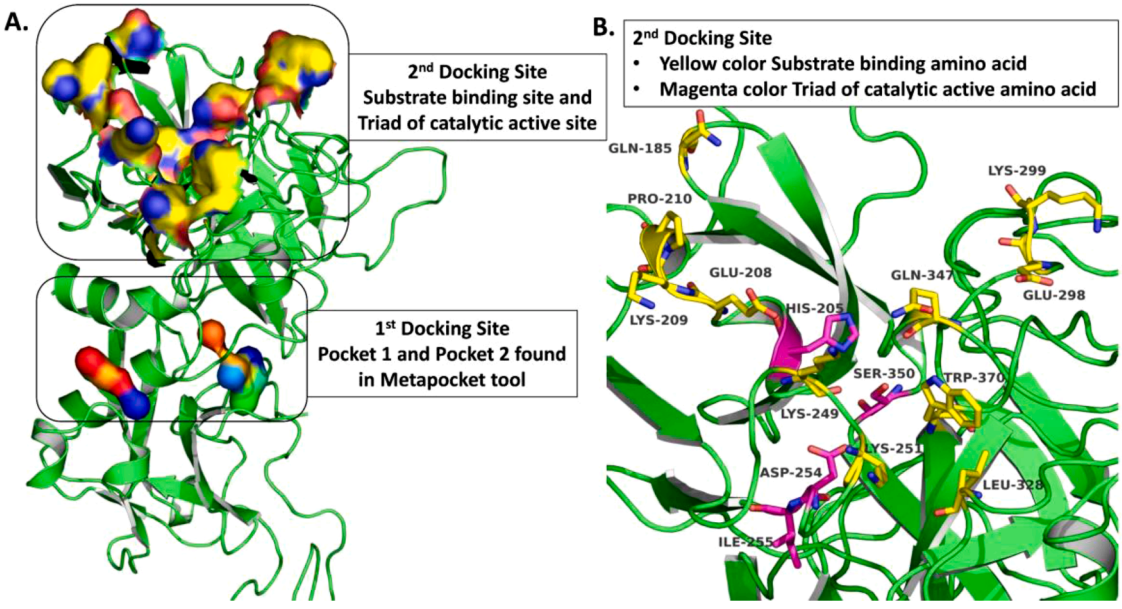

Fig. 2. Binding pocket site (A.) showing 1st site on pocket 1, pocket two found in Metapocket tool, and 2nd site is substrate binding site & Triad of the active catalytic site (B.) The amino acid present in the 2nd site is the substrate binding site (Yellow color) and the Triad of the active catalytic site (Magenta color).

Magenta color). Selected active sites 1st and 2nd shown red color (circled in magenta and brick color) cleft or pocket was present at the same place we predicted through Metapocket.

3.6. Calculation of physicochemical properties for FDA approval compounds

The FDA approval compounds (DrugBank FDA 1657 Compounds and DSSTOX 2525 Compounds) from the ZINC database in total 4182 compounds were converted to 3D coordinates. After that, selected compounds were further refined by using physicochemical parameters. The ChemAxon (command line) software calculated the Lipinski rule and Verber rules properties. In the given threshold, 2815 compounds were fit, which is shown in **Supplementary Figure 11**.

3.7. Molecular docking analysis

The docking for FDA approval 2815 compounds was done using AutoDock on-site predicted by Metapocket tool and substrate binding site and Triad of the active catalytic site shown in Fig. 3. Once docking finished, the compounds were selected based on binding energy (BE). **Supplementary Table 1** shows that the selected FDA approval compounds name/ID with their 2D structural and Supplementary Table 2 shows that the FDA approval compounds name/ID and binding energy in kcal/mol.

Docking on 1st site in pocket 1, pocket two found in Metapocket tool

Top virtual hit compounds selected based on binding energy docked in a selected site shown in Fig. 3. Inhibitor Camostat forms a Hydrophobic Interactions, Hydrogen Bonds, and Salt Bridges at LEU4 (LEU4 form 2 hydrophobic interaction with Camostat, which shows the 3.46Å and 3.81Å distance), ALA8 (3.59Å), VAL45 (3.84Å), ASN359 (3.31Å), and ILE361 (3.43Å). The residues SER46, CYS140, ASN277, ASN359, and TRP363, can form hydrogen bonds. In contrast, HIS47 AND LYS143 are involved in Salt Bridges formation (**Supplementary Figure 12**). Inhibitor Differin shows the -11.07 kcal/mol binding energy. Differin forms a hydrophobic interaction with THR2 (3.63Å), LEU4 (2.97Å), VAL5 (3.81Å AND 3.32Å), VAL45 (3.74Å), LEU121 (3.03Å), PRO278 (3.17Å). The hydrogen bond was formed at GLY279 residue. The pi-Cation Interaction was observed at LYS358 residues shown in **Supplementary Figure 12**.

E155 inhibitor shows the Hydrogen bonds at ALA8, ALA10, LYS16, VAL45, HIS47, CYS140, LYS143, and ASN277; along with this; there is a Hydrophobic interaction at PHE3 (3.07Å), LEU4 (3.46Å), ALA10 (3.16Å) and LEU13 (3.89Å), and whose distance was not more than 4Å. Dasabuvir is another FDA approval compound called Exviera and used for antiviral medication for hepatitis-C; this drug is used to combine ombitasvir/paritaprevir/ritonavir for treatment of hepatitis-C virus type 1. This drug forms a hydrophobic interaction and Hydrogen bonds. Hydrogen bonds are at TRP41 and GLY279 and hydrophobic interaction at amino acid LEU4 (3.41Å), VAL5 (3.45Å), PHE17 (3.38Å), TRP41 (3.92Å), ASN122 (3.21Å), and PRO278 (3.23Å).

We get the hit compound Sorafenib, which is a kinase inhibitor-approved drug to treat primary kidney cancer through our screening. It formed hydrophobic interaction at VAL5, ASP43, LYS120, ASN122, and PRO278 and Hydrogen bonds at ASN40, TRP41, GLY44, LEU121, and ASN122. Another drug Betrixaban hit compound, is an oral anti-coagulant drug, which formed a Halogen bond with TMPRSS2 protein at MET280 at a distance of 3.89Å. The hydrogen bonds were formed at ASN40, TRP41, LYS120, ASN122, ASP391, and Hydrophobic interaction at ASP43, LYS120, LEU121, PRO278, THR390, and TYR394. Other hit compounds ZINC3830554 form salt Bridges, Hydrogen Bonds and Hydrophobic Interaction. LYS143 forms a Salt Bridge at a distance of 3.28Å with TMPRSS2 protein. The hydrogen bond at ALA8, VAL45, HIS47, SER142, ASN277, ASN359, and TRP363. The hydrophobic Interactions are at PHE3, LEU4, ALA7, ALA8, ALA139, PRO197, and PHE266, as shown in **Supplementary Figure 12** and all hit compounds hydrogen bond and hydrophobic interaction were shown in **Supplementary Table 3**.

Docking on 2nd site in substrate binding site and Triad of the active catalytic site

Top virtual hit compounds selected based on binding energy docked

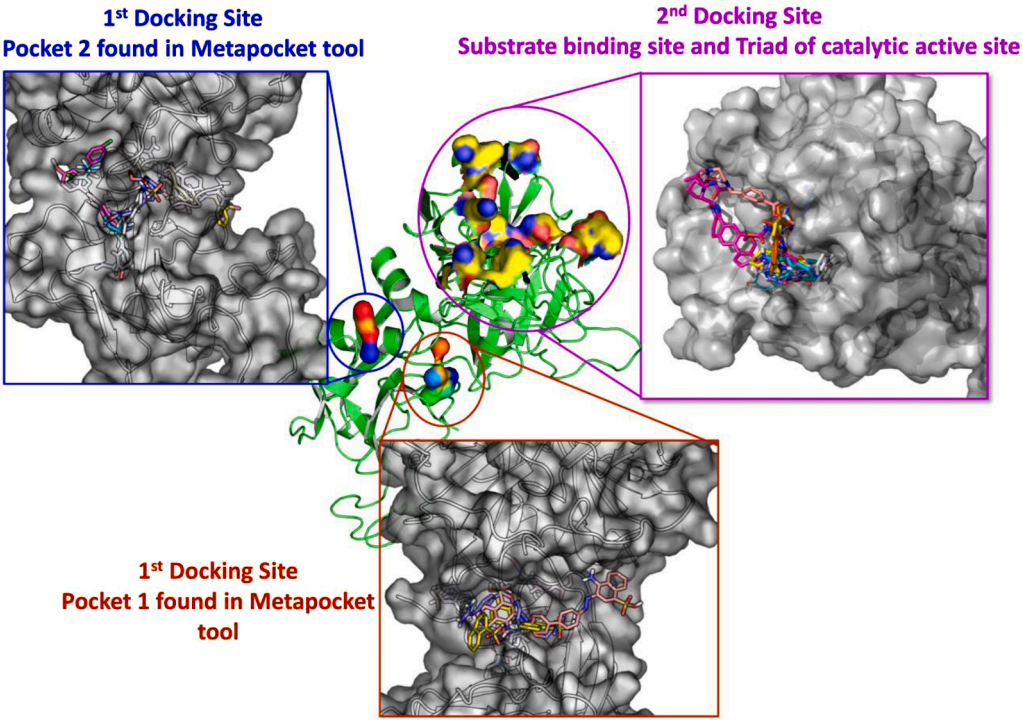

Fig. 3. Screened virtual hit compounds binding in selected 1st site on pocket 1, pocket two and 2nd site are substrate binding sites and Triad of the active catalytic site.

in a selected site shown in Fig. 3. In the substrate-binding site and Triad catalytic active site, compound Camostat binds and forms Hydrophobic Interactions at TRP370, Hydrogen Bonds at SER345, TRP370 and GLY373. Along with this bond, there is a Salt Bridges at HIS205 and ASP344 interaction with Camostat. All these amino acids were involved in Camostat binding in the serine protease domain of TMPRSS2, shows in **Supplementary Figure 13**. Inhibitor ZINC203686879 shows the -12.56 kcal/mol highest binding energy among all selected seven hit compounds. ZINC203686879 forms a hydrophobic interaction with VAL187, VAL189, HIS205, LYS301, GLN347, and TRP370. The hydrogen bond was at GLY300, THR302, GLN347, SER350 residues. This study finds the Imatinib (ZINC19632618), which is used as a cancer growth blocker called tyrosine kinase inhibitor. When this compound binds to the TMPRSS2 protein, it shows the -12.37 kcal/mol binding energy and forms a pi-stacking interaction at HIS205, Hydrophobic interaction at GLU298, GLN347, THR368, TRP370 and Hydrogen bonds at GLN347, SER350, GLY373. Another drug from seven hit compounds Darifenacin (ZINC1996117), used to treat urinary incontinence and discovered by scientists at the Pfizer research site, forms hydrogen bonds TYR246 pi-cation interaction at HIS205. We see the Hydrophobic interaction created by Darifenacin (ZINC1996117) at HIS205, TYR246, LYS251, ASP254, THR368, and TRP370 (**Supplementary Figure 13**).

Compound Fazadinium (ZINC3830826) and Raloxifene (ZINC538275) show the Hydrophobic interaction at HIS205, LYS251, ASP254, THR368, and TRP370. Hydrogen bond for Raloxifene (ZINC538275) at GLU208, LYS251, ASP254, SER369, GLY371 Fazadinium (ZINC3830826) at GLN347, SER350. Fazadinium (ZINC3830826) shows Pi-cation interaction and Raloxifene (ZINC538275) shows pi-Stacking interaction at HIS205. Compound ZINC1530886 and Brexpiprazole (ZINC84758479) show the same binding energy -11.36 kcal/mol. Brexpiprazole (ZINC84758479)shows the hydrophobic (HIS205, LYS251, TRP370) and hydrogen bond interaction (SER345, GLN347, SER350, SER369), along with this, there is salt Bridges formation by GLU208. Compound ZINC1530886 shows TYR246, LYS251, ASP254, LEU328, TRP370 hydrophobic interaction and LYS251, GLY371 hydrogen bonds. TRP370 forms a pi-sacking interaction. All these selected seven hit compounds most common occurrence of a hydrogen bond at GLN347and SER350. These amino acids are to the catalytic site or pocket of TMPRSS2 protein, highly conserved with other TTPs (type II transmembrane serine proteases). All hit compounds hydrogen bond and hydrophobic interaction were shown in **Supplementary Table 3.**

After that, we check the binding poses of selected active hit compounds bind in a 1st and 2nd site based on binding energy and most populous cluster. Docked poses of selected compounds were shown in a different color binding pose based on binding energy were shown by magenta color and most populous cluster were shown by Green color (**Supplementary Figure 14 and 15**), but some compounds show the same pose where highest binding energy and most populous cluster were found in same poses, in 1st binding site differin and ZINC3830554 found (**Supplementary Figure 14)** and 2nd binding site Camostat, Darifenacin and Imatinib (**Supplementary Figure 15**).

3.8. ADME prediction and PAINS analysis

In Bioactive radar, the Pink area exhibits an optimal range of particular properties for selected hit compounds. In bioactive radar, LIPO shows the lipophilicity as XLOGP3; SIZE is size as molecular weight; POLAR is polarity as topological polar surface area (TPSA); INSOLU is insolubility in water by logS scale; INSATU is in saturation as per fraction of carbons in the sp3 hybridization and FLEX is flexibility as per rotatable bonds. **Supplementary Table 4** shows the bioavailability radar for all virtual hit compounds. Selected compounds (1st binding sites) other parameter values were shown in **Supplementary Table 5**; Selected compounds (2nd binding site) other parameter values are shown in **Supplementary Table 6**. Selected compounds were submitted to 'PAINS remover' to identify the false positive compounds; through this analysis, selected compounds passed the PAINS filtration; all compounds are PAINS negative.

3.9. Clustering analysis

We performed Clustering analysis using ChemBioServer online tool based on the Hierarchical Clustering, using Soergel (Tanimoto Coefficient) Distance method, Ward Linkage clustering method and Clustering Threshold is 0.9. compounds from the 1st docking site, we found three clusters shown in Fig. 4 **(A)**. We selected all three clusters, which contain seven compounds. Among these compounds, the known Camostat compound is used to inhibit TMPRSS2, and these compounds have formed a cluster with ZINC3784182 (Differin) and ZINC95616937 (Dasabuvir), which fall in cluster first. On the other binding site (2nd Site), we found two clusters shown in Fig. 4 **(B)**, First cluster contains five compounds, including Camostat, and 2nd cluster found three compounds. We check the structural difference between known Camostat and other hit compounds using these virtual hit compounds (**supplementary figure 16)** using MolAlign Server. Their alignment score was shown in **Supplementary Table 7**. Based on the Maximum atom distance (Dmax) score, Dasabuvir shows the 0.971583 top scores and Differin shows the 2nd top Dmax score of 0.961961 on the first docking site. In the 2nd Docking site, ZINC1530886 shows the top Dmax score of 1.06116 and Raloxifene shows the 2nd top score of 1.02178. We use all these compounds for 20ns MD simulation and free energy calculation.

3.10. Analysis of MD simulation

The virtual hit compounds with the highest binding energy were subjected to 20ns MD simulation to check the modeled structure and conformational stability of protein and protein-ligand complex structure by calculating RMSD, RMSF, and radius of gyration.

RMSD Analysis

MD analysis for hit compounds present in 1st site

RMSD for 20ns was run for checking the change in the Cα atom of a protein (without ligand) and protein-ligand complex, which show the same fluctuation at the start of simulation up to 1.5ns between 0.38 to 0.5nm. After 1.5ns, Betrixaban shows the sudden increase in the fluctuation and ZINC3830554 shows a slight decrease in the fluctuation and other compounds are show the slow increase in the fluctuation up to 4ns simulation as shown in Fig. 5 **(A)**. After 10ns Camostst, and Sorafenib showed more movement (between 0.6-0.8nm) than the other hit compounds. The other five compounds Betrixaban, Differin, Dasabuvir, E155, and ZINC3830554, show the fluctuation between 0.38-0.7nm from 1.5ns to 20ns. In comparison, all compounds showed the same fluctuation. RMSD values are shown in Fig. 5 **(A)**.

MD analysis for hit compounds present in 2nd site

Fig. 5 **(B)** shows the protein's RMSD plot and selected seven hit compounds, one known compound-complex. All the eight complexes show the same pattern of fluctuation as shown by protein without the ligand. Still, the RMSD value of hit compounds is higher in the RMSD value due to hit compound binding to a protein complex. This fluctuation is between 0.2 to 1nm throughout the 20ns simulation. At the initial of simulation, the increased fluctuation was observed up to 3ns. Raloxifene, Brexpiprazole and Fazadinium show more RMSD values, but other compounds showed the same RMSD value. After 7ns, Brexpiprazole slowly decreases in the RMSD value and Raloxifene reveals a gradual decrease in the RMSD value after 11.4ns. The Fazadinium compound shows an increase in RMSD up to 9ns; after that, it shows a steady fluctuation with a slight rise and a decrease in RMSD value. Compound ZINC1530886 shows the low RMSD value up to 12ns simulation after that it was gradually increased up to 20ns and at the end of 20ns simulation, it shows the 1.3nm fluctuation. The RMSD result reveals that the hit compounds-protein complexes are stable throughout the 20ns simulation period, excluding the ZINC1530886 compound shown in Fig. 5 **(B)**.

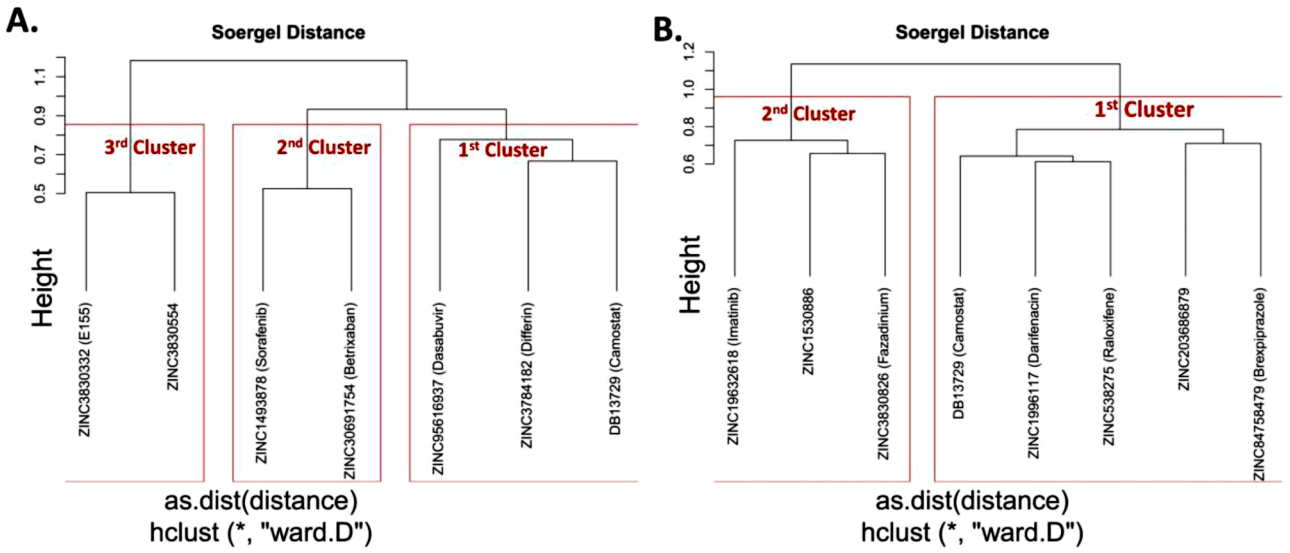

Fig. 4. (A) Three clusters were formed by hit compounds found in the 1st site and (B) Two clusters were formed by hit compounds found in the 2nd site based on the Hierarchical Clustering method.

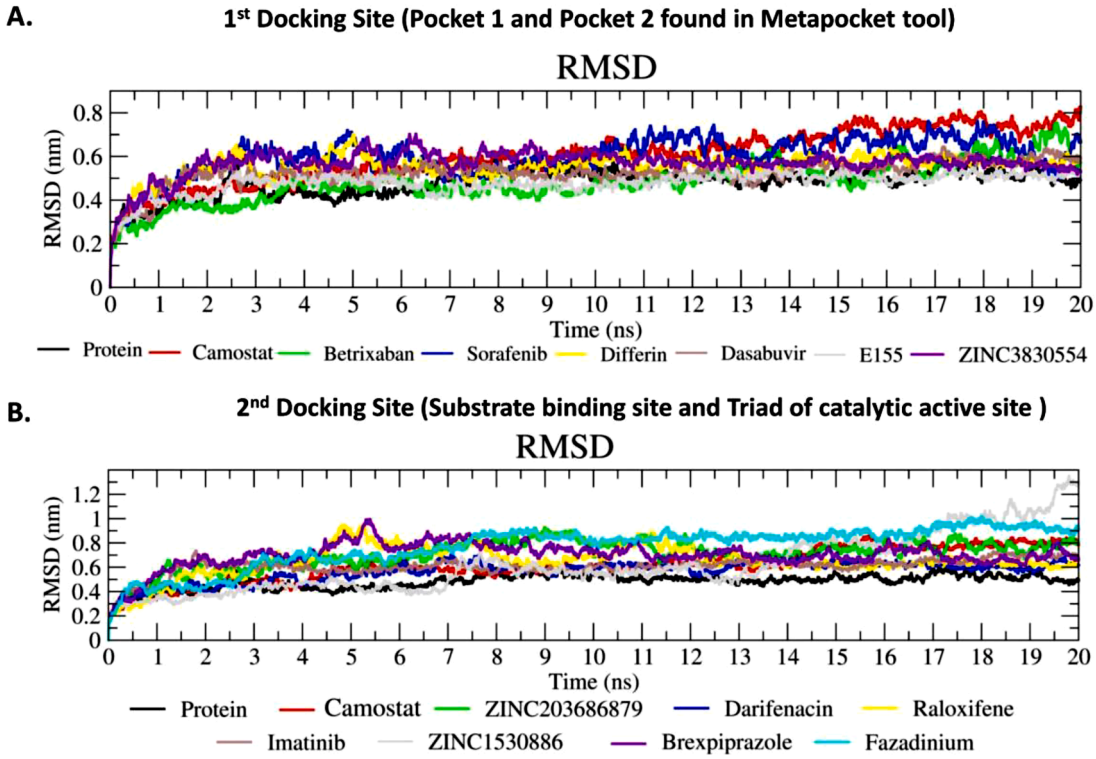

Fig. 5. RMSD for protein and selected hit compounds for 20ns MD simulation. **(A)** RMSD for hit compounds found in 1st site and **(B)** RMSD for hit compounds found in 2nd site.

Root Mean square Fluctuation Analysis

MD analysis for hit compounds present in 1st site

The RMSF (Root Mean square Fluctuation) value was calculated by plotting nm vs. residues. **Supplementary Figure 17 (A)**, all screened seven compound protein complexes, and protein (free ligand) showed the same fluctuation, but Differin (Yellow color) showed more change than other compounds, as shown in **Supplementary Figure 17 (A)**. The RMSF plot's backbone showed more fluctuation at 60-105 amino acids because of the β-sheet, loop, and small helix region. There is a little movement at region 110-135 amino acid due to the helix and β-sheet region. After that, the connected fluctuation was observed in the 145-165 region. A motion was shown as screened compounds bound from 1 amino acid to 170 amino acids, resulting in the changes in protein structure and RMSF moment. When a comparison was made between protein (without ligand) and protein-ligand complex with selected seven hits compound, the same fluctuation was observed. Simultaneously, some regions exhibited differences in fluctuation, as shown in **Supplementary Figure 17 (A)**.

MD analysis for hit compounds present in 2nd site

2nd docking site present between 205-370 amino acid. At initial

residues of the protein are not very stable. It shows more fluctuation; specifically, Fazadinium and ZINC1530886 show more RMSF values, shown in **Supplementary Figure 17 (B)**. The remaining all other complexes have shown similar RMSF patterns. However, in 2nd docking sites, an amino acid which is present in a substrate binding site and Triad of the active catalytic site; these amino acids are H205, I255, D254, S350 (Triad of the active catalytic site) and Q185, H205, E208, K209, P210, K249, K251, E298, K299, L328, Q347, S350, W370 (active or catalytic site). In this region, we observed fluctuation and when we compare these with Figures A and B, we see slightly more fluctuation in Figure B, which is the 2nd docking site. Some residues show higher fluctuations in a few regions, depending on the inhibitor compound's nature and overall complex stability.

The radius of Gyration Analysis

MD analysis for hit compounds present in 1st site

The radius of Gyration (Rg) analyzes the protein and ligand's compactness level, which was done for the 20ns time scale, are shown in **Supplementary Figure 18 (A)**. At the MD simulation start, the Rg value for seven FDA approval hit compounds between 2.3-2.5nm was 3700ps. The Sorafenib and Differin protein complex's compactness showed an increase in the Rg value as simulation time increases. Still, after 16000ps, it shows a decrease in Rg value and ends at 2.45nm Rg. The protein (Black color) showed more Rg value (up to 2.5nm), but after 3200ps, it shows an Rg value decrease between 2.35-2.45nm. At the end of the simulation, all compounds showed the Rg value between 2.3-2.5nm. Compound Camostat, Betrixaban, E155, and ZINC 3830554 show the same fluctuation between 2.33-2.5nm Rg. Rg values are shown in **Supplementary Figure 18 (A)**.

MD analysis for hit compounds present in 2st

At the start of the simulation, all compounds, including protein were shows the Rg value between 2.35-2.55nm up to 1000ps, that compounds show the different Rg value, compound Imatinib shows the lowest Rg value throughout the simulation and Raloxifene and Brexpiprazole shows the highest Rg value at initial of simulation after 8000ps it shows the decrease in the Rg value. After the 15000ps simulation, the ZINC1530886 compound shows a sudden increase in Rg value and slowly decreases in it up to 2.55nm at the end of the simulation. Throughout the 20ns simulation Camostat (Red color) compound the Rg value decrease and ends with 2.2nm Rg value **Supplementary Figure 18 (B)**.

Further, the change in the protein's secondary structure with time through the do_dssp program was done. Through this, we read a trajectory file and compute the secondary structure for each time frame. Secondary structure changes observed when the ligand binds to the protein, as shown in **Supplementary Figure 19** and **Supplementary Figure 20**. When selected FDA-approved compounds bind to the protein, a change was observed in protein structure compared between protein and protein-ligand complex. When the ligand binds, some changes occurred in turn (yellow color) to Bends (Green color). In the 1st docking site (**Supplementary Figure 19**), there is a Coil (White color) region between 60-100 amino acids in protein when the ligand binds, and there are changes in this region to form Bends (Green color). In some places, β-sheets (Red color) are changes to turn (yellow color) and Bends (Green color). The case of the 2nd docking site region between 200-350 shows the changes in the structure in which A-helix to turns, coil to β-sheets/ Bends. Compounds Darifenacin observe the A-helix region between 225-250 amino acid shown in **Supplementary Figure 20**.

3.11. Hydrogen bond analysis

Hydrogen bonds play an essential role in stabilizing the protein-ligand complex. This hydrogen bond is responsible for drug specificity, metabolization and adsorption in the body. MD trajectories were analyzed to check the number of hydrogen bonds between protein-hit compounds complex and protein solvent interaction was observed during the 20ns simulation. The number of hydrogen bonds helps us to understand the stability of the protein-ligand complex. 1st Docking site, the formation of a hydrogen bond between protein and ligands was 0 to 7 bonds for 20 ns simulation. Still, compound E155 forms more hydrogen bonds than the other compound hydrogen bond throughout the simulation shown in **Supplementary Figure 21 (A)**.

On the other hand, we check the hydrogen interaction between protein and solvent throughout the 20 ns simulation. This shows the constant hydrogen bond throughout the simulation, between 690-870 number of hydrogen bonds. All hit compounds and known reference compound shows the same number of hydrogen bonds shown in **Supplementary Figure 22 (A)**, but Differin (yellow color) shows the more hydrogen bond after 10.5ns up to 19ns. 2nd docking site, protein-ligand interaction shows the hydrogen bond less than four, but compound Camostst shows more hydrogen bond than the other compound-complex. Raloxifene also shows more hydrogen formation than the other compounds but less than the Camostst complex shown in **Supplementary Figure 21 (B)**. Once we check protein-ligand interaction, we check the protein solvent hydrogen bond interaction, which shows the hydrogen bond formation between 690-870, shown in **Supplementary Figure 22 (B)**. We conclude that all hit compounds and known compounds can bind to protein effectively and tightly through this hydrogen bond interaction.

3.12. Principal component analysis (PCA)

To perform PCA, the eigenvectors, eigenvalues and their projection were calculated using the essential dynamics methods. Through this PCA analysis, we analyzed the motion during hit compounds binding in different complexes. These eigenvectors determine the overall motion of the particular protein. The protein-ligand complexes can be explained by 2D projection plot generation in PCA shown in **Supplementary Figure 23**. For that, we select the first two principal components, which are PC1 and PC2, to predict the significant motions. From **Supplementary Figure 23 (A),** we observe that all hit complexes from the 1st docking site show stable clusters as similarly shown by known camostat compound. Still, Sorafenib, Dasabuvir show a slight difference in the moments. 2nd docking site Fazadinium and ZINC1530886 show the difference in the moment of the protein complex. Other remaining compounds show the same moment shown in **Supplementary Figure 23 (B),** which offers a stable cluster.

3.13. Free energy calculation

After MD simulation, free binding energy was calculated by MMPBSA for selected seven FDA approved compounds as shown in **Supplementary Table 8**; the post-simulation free binding energy for 1st docking site calculations showed maximum binding free energy for ZINC3830554 compound, which is -288.702kJ/mol and van der Waal energy, Electrostatic energy, Polar solvation energy, SASA energy was -296.495kJ/mol, -237.287kJ/mol, 271.750kJ/mol, and -26.671kJ/mol respectively. The lowest binding free energy was shown by Betrixaban (ZINC30691754), which is -90.847kJ/mol, van der Waal energy, Electrostatic energy, Polar solvation energy, and SASA energy was -253.444kJ/mol, -203.530kJ/mol, 390.250kJ/mol, -24.123kJ/mol respectively. Free binding energy for 2nd docking site, highest binding free energy was shown by ZINC1530886 compound -164.738kJ/mol and lowest binding free energy was shown by Camostat (DB13729) which is -55.609kJ/mol. The rest of the compounds binding free energy and Van der Waal energy, Electrostatic energy, Polar solvation energy, and SASA energy were shown in **Supplementary Table 7**.

We further validated the ligand-protein complex interaction before simulation and after simulation (20ns MD simulation run). Residues involved in the binding interaction of pre and post-simulation were compared. Almost the same residues of TMPRSS2 within 4Å of the ligand were involved, as shown in **Supplementary Table 9**, suggesting the same pocket preference and stable ligand binding interactions.

4. Conclusion

COVID-19 is a global infection increasing the high mortality rate. In this study, in total, 4182 FDA-approved compounds from the ZINC database were used to interact with TMPRSS2. The target TMPRSS2, which consists of 489 amino acids transmembrane region, was predicted to show one transmembrane helix region at 84-106 amino acid and remaining part of the present inside (amino acid 1-83) and outside (107-489 amino acid). The outer region and some parts of the helix region, which consist of 415 amino acids, were used to calculate the physicochemical properties.

The modelled structure of the TMPRSS2 protein was generated and validated. Simultaneously, physicochemical properties space filtering for FDA-approved compounds was done to reduce them to 2815 compounds. These 2815 compounds' binding energy was revealed by molecular docking. The top virtual hit compounds were selected based on Docking Binding energy. In the 1st docking site, we selected seven compounds on them. ZINC3830554 showed promising binding energy -12.91kcal/mol and 2nd docking site. We selected eight compounds, including the known Camostat ZINC203686879 compound shows the highest binding energy, which is -12.56 kcal/mol. Through docking study, we observed that selected virtual hit compounds with docking scores comparable to or more than that of the known Camostat inhibitor, screened FDA approval compounds on 1st and 2nd docking site shows the higher docking score, which is considered to be the most potent inhibitor. These virtual hit compounds found in the 1st and 2nd docking site protein-ligand complexes have maintained stability at 0.3-1.2nm during the simulation. RMSF of the protein and complex were similar, which shows the strength when binds to protein. Among the virtual hit compounds, the Differin (ZINC3784182) has been mostly used for external use to treat skin disease. Our hit compounds show the more binding energy they known Camostat compounds and offer more stability when formed 20ns MD simulation. Therefore, other screened FDA compounds might be effective against COVID-19, and these compounds can be explored further for drug repurposing treatment against coronavirus to inhibit the COVID-19 successfully.

CRediT authorship contribution statement

Sagar Barge: Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Dhananjay Jade:** Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Gokul Gosavi:** Formal analysis, Writing - original draft. **Narayan Chandra Talukdar:** Writing - review & editing, Supervision. **Jagat Borah:** Conceptualization, Data curation, Writing - review & editing, Supervision, Project administration.

Acknowledgments

The authors thank the Director of IASST for providing resources for the completion of our research.

Funding

No funding to declare.

Conflicts of Interest

The authors have declared that no conflict of interest exists.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejps.2021.105820.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindahl, E., 2015. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2, 19–25.

Azimi, A., 2020. TMPRSS2 inhibitors, Bromhexine, Aprotinin, Camostat and Nafamostat as potential treatments for COVID-19. Center for Open Science.

Baell, J.B., Holloway, G.A., 2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J. Med. Chem. 53, 2719–2740.

Baron, S.A., Devaux, C., Colson, P., Raoult, D., Rolain, J.-M., 2020. Teicoplanin: an alternative drug for the treatment of COVID-19? Int. J. Antimicrob. Agents 55, 105944-105944.

Benet, L.Z., Hosey, C.M., Ursu, O., Oprea, T.I., 2016. BDDCS, the Rule of 5 and drugability. Adv. Drug Deliv. Rev. 101, 89–98.

Berger, A., Preiser, W., 2011. SARS. Encyclopedia Environ. Health 596–604.

Bertram, S., Glowacka, I., Müller, M.A., Lavender, H., Gnirss, K., Nehlmeier, I., Niemeyer, D., He, Y., Simmons, G., Drosten, C., Soilleux, E.J., Jahn, O., Steffen, I., Pöhlmann, S., 2011a. Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. J. Virol. 85, 13363–13372.

Bertram, S., Glowacka, I., Müller, M.A., Lavender, H., Gnirss, K., Nehlmeier, I., Niemeyer, D., He, Y., Simmons, G., Drosten, C., Soilleux, E.J., Jahn, O., Steffen, I., Pöhlmann, S., 2011b. Cleavage and activation of the severe acute respiratory syndrome coronavirus spike protein by human airway trypsin-like protease. J. Virol. 85, 13363–13372.

Brown, B.P., Mendenhall, J., Meiler, J., 2019. BCL::MolAlign: Three-Dimensional Small Molecule Alignment for Pharmacophore Mapping. J. Chem. Inf. Model 59, 689–701.

Colovos, C., Yeates, T.O., 1993. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 2, 1511–1519.

Consortium, T.U., 2018. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515.

Daina, A., Michielin, O., Zoete, V., 2017. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Sci. Rep. 7, 42717.

David, A., Khanna, T., Beykou, M., Hanna, G., Sternberg, M.J.E., 2020. Structure, function and variants analysis of the androgen-regulated TMPRSS2, a drug target candidate for COVID-19 infection. bioRxiv, 2020.2005.2026.116608.

de Castro, E., Sigrist, C.J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P.S., Gasteiger, E., Bairoch, A., Hulo, N., 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res. 34, W362–W365.

Elfiky, A.A., 2020. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. Life Sci. 248, 117477-117477.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-Y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. Curr. Protoc. Bioinformatics. Chapter 5, Unit-5.6.

Fan, H.-H., Wang, L.-Q., Liu, W.-L., An, X.-P., Liu, Z.-D., He, X.-Q., Song, L.-H., Tong, Y.-G., 2020. Repurposing of clinically approved drugs for treatment of coronavirus disease 2019 in a 2019-novel coronavirus-related coronavirus model. Chin. Med. J. (Engl.) 133, 1051–1056.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, NJ, pp. 571–607.

Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discov. 10, 449–461.

Glowacka, I., Bertram, S., Müller, M.A., Allen, P., Soilleux, E., Pfefferle, S., Steffen, I., Tsegaye, T.S., He, Y., Gnirss, K., Niemeyer, D., Schneider, H., Drosten, C., Pöhlmann, S., 2011. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. J. Virol. 85, 4122–4134.

Hoffmann, M., Hofmann-Winkler, H., Pöhlmann, S., 2018. Priming Time: How Cellular Proteases Arm Coronavirus Spike Proteins. Activation of Viruses by Host Proteases 71–98.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M.A., Drosten, C., Pöhlmann, S., 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280.e278.

Huang, B., 2009. MetaPocket: a meta approach to improve protein ligand binding site prediction. Omics 13, 325–330.

Hussain, M., Jabeen, N., Amanullah, A., Baig, A.A., Aziz, B., Shabbir, S., Raza, F., 2020. Structural Basis of SARS-CoV-2 Spike Protein Priming by TMPRSS2. bioRxiv, 2020.2004.2021.052639.

Jade, D.D., Pandey, R., Kumar, R., Gupta, D., 2020. Ligand-based pharmacophore modeling of TNF-α to design novel inhibitors using virtual screening and molecular dynamics. J. Biomol. Struct. Dyn. 1–17.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22, 2577–2637.

Kannan, S., Shaik Syed Ali, P., Sheeza, A., Hemalatha, K., 2020. COVID-19 (Novel Coronavirus 2019) - recent trends. Eur. Rev. Med. Pharmacol. Sci. 24, 2006–2011.

Karatzas, E., Zamora, J.E., Athanasiadis, E., Dellis, D., Cournia, Z., Spyrou, G.M., 2020. ChemBioServer 2.0: an advanced web server for filtering, clustering and networking of chemical compounds facilitating both drug discovery and repurposing. Bioinformatics 36, 2602–2604.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580.
Kumari, R., Kumar, R., Lynn, A., 2014. g_mmpbsa—A GROMACS Tool for High-Throughput MM-PBSA Calculations. J. Chem. Inf. Model. 54, 1951–1962.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486.
Liang, C., Tian, L., Liu, Y., Hui, N., Qiao, G., Li, H., Shi, Z., Tang, Y., Zhang, D., Xie, X., Zhao, X., 2020. A promising antiviral candidate drug for the COVID-19 pandemic: A mini-review of remdesivir. Eur. J. Med. Chem. 201, 112527.
Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J.J.A.d.d.r., 1997. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. 23, 3-25.
Mackenzie, J.S., Smith, D.W., 2020. COVID-19: a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't. Microbiol. Aust. MA20013-MA20013.
Mahase, E., 2020. Covid-19: Moderna applies for US and EU approval as vaccine trial reports 94.1% efficacy. BMJ 371, m4709.
Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., Nagata, N., Sekizuka, T., Katoh, H., Kato, F., Sakata, M., Tahara, M., Kutsuna, S., Ohmagari, N., Kuroda, M., Suzuki, T., Kageyama, T., Takeda, M., 2020. Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. Proc. Natl. Acad. Sci. U.S.A. 117, 7001–7003.
Matsuyama, S., Shirato, K., Kawase, M., Terada, Y., Kawachi, K., Fukushi, S., Kamitani, W., 2018. Middle East Respiratory Syndrome Coronavirus Spike Protein Is Not Activated Directly by Cellular Furin during Viral Entry into Target Cells. J. Virol. 92 e00683-00618.
Millet, J.K., Whittaker, G.R., 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus Res. 202, 120–134.
Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791.
Oliver, S.E., Gargano, J.W., Marin, M., Wallace, M., Curran, K.G., Chamberland, M., McClung, N., Campos-Outcalt, D., Morgan, R.L., Mbaeyi, S., Romero, J.R., Talbot, H. K., Lee, G.M., Bell, B.P., Dooling, K., 2020. The Advisory Committee on Immunization Practices' Interim Recommendation for Use of Pfizer-BioNTech COVID-19 Vaccine - United States, December 2020. MMWR Morb. Mortal Wkly. Rep. 69, 1922–1924.
Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., Bailey, R., Swanson, K.A., Roychoudhury, S., Koury, K., Li, P., Kalina, W.V., Cooper, D., Frenck Jr., R.W., Hammitt, L.L., Türeci, Ö., Nell, H., Schaefer, A., Ünal, S., Tresnan, D.B., Mather, S., Dormitzer, P.R., Şahin, U., Jansen, K.U., Gruber, W.C., 2020. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N. Engl. J. Med. 383, 2603–2615.
Rubin, D., Chan-Tack, K., Farley, J., Sherwat, A., 2020. FDA Approval of Remdesivir - A Step in the Right Direction. N. Engl. J. Med. 383, 2598–2600.
Rubin, E.J., Longo, D.L., 2020. SARS-CoV-2 Vaccination - An Ounce (Actually, Much Less) of Prevention. N. Engl. J. Med. 383, 2677–2678.
Schüttelkopf, A.W., van Aalten, D.M., 2004. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallogr. D. Biol. Crystallogr. 60, 1355–1363.
Shulla, A., Heald-Sargent, T., Subramanya, G., Zhao, J., Perlman, S., Gallagher, T., 2011. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. J. Virol. 85, 873–882.
Sterling, T., Irwin, J.J., 2015. ZINC 15–Ligand Discovery for Everyone. J. Chem. Inf. Model. 55, 2324–2337.
Touw, W.G., Baakman, C., Black, J., te Beek, T.A.H., Krieger, E., Joosten, R.P., Vriend, G., 2015. A series of PDB-related databanks for everyday needs. Nucleic Acids Res. 43, D364–D368.
van der Hoek, L., Pyrc, K., Berkhout, B., 2006. Human coronavirus NL63, a new respiratory virus. FEMS Microbiol. Rev. 30, 760–773.
Veber, D.F., Johnson, S.R., Cheng, H.-Y., Smith, B.R., Ward, K.W., Kopple, K.D.J.J.o.m.c., 2002. Molecular properties that influence the oral bioavailability of drug candidates. 45, 2615-2623.
Vishnubhotla, R., Vankadari, N., Ketavarapu, V., Amanchy, R., Avanthi, S., Bale, G., Reddy, D.N., Sasikala, M., 2020. Genetic variants in TMPRSS2 and Structure of SARS-CoV-2 spike glycoprotein and TMPRSS2 complex. bioRxiv, 2020.2006.2030.179663.
Voysey, M., Clemens, S.A.C., Madhi, S.A., Weckx, L.Y., Folegatti, P.M., Aley, P.K., Angus, B., Baillie, V.L., Barnabas, S.L., Bhorat, Q.E., Bibi, S., Briner, C., Cicconi, P., Collins, A.M., Colin-Jones, R., Cutland, C.L., Darton, T.C., Dheda, K., Duncan, C.J.A., Emary, K.R.W., Ewer, K.J., Fairlie, L., Faust, S.N., Feng, S., Ferreira, D.M., Finn, A., Goodman, A.L., Green, C.M., Green, C.A., Heath, P.T., Hill, C., Hill, H., Hirsch, I., Hodgson, S.H.C., Izu, A., Jackson, S., Jenkin, D., Joe, C.C.D., Kerridge, S., Koen, A., Kwatra, G., Lazarus, R., Lawrie, A.M., Lelliott, A., Libri, V., Lillie, P.J., Mallory, R., Mendes, A.V.A., Milan, E.P., Minassian, A.M., McGregor, A., Morrison, H., Mujadidi, Y.F., Nana, A., O'Reilly, P.J., Padayachee, S.D., Pittella, A., Plested, E., Pollock, K.M., Ramasamy, M.N., Rhead, S., Schwarzbold, A.V., Singh, N., Smith, A., Song, R., Snape, M.D., Sprinz, E., Sutherland, R.K., Tarrant, R., Thomson, E.C., Török, M.E., Toshner, M., Turner, D.P.J., Vekemans, J., Villafana, T.L., Watson, M.E. E., Williams, C.J., Douglas, A.D., Hill, A.V.S., Lambe, T., Gilbert, S.C., Pollard, A.J., Oxford, C.V.T.G., 2021. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet 397, 99–111.
Zhang, J., Xie, B., Hashimoto, K., 2020. Current status of potential therapeutic candidates for the COVID-19 crisis. Brain Behav. Immun. 87, 59–73.
Zhang, Z., Li, Y., Lin, B., Schroeder, M., Huang, B., 2011. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. Bioinformatics 27, 2083–2088.
Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., Cheng, F., 2020. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 6, 14-14.
Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R.J.N.E.J.o.M., 2020. A novel coronavirus from patients with pneumonia in China, 2019.