Sequence analysis of Indian SARS-CoV-2 isolates shows a stronger interaction of mutant receptor-binding domain with ACE2

Pujarini Dash\textsuperscript{a,1,2}, Jyotirmayee Turuk\textsuperscript{a,1,}, Santosh K. Behera\textsuperscript{a,2}, Subrata Kumar Pal\textsuperscript{a,2}, Sunil K. Raghav\textsuperscript{b,\textsuperscript{a},\textsuperscript{3}}, Arup Ghosh\textsuperscript{b,\textsuperscript{3}}, Jyotsnamayee Sabat\textsuperscript{a,2}, Sonalika Rath\textsuperscript{a,2}, Subhra Subhadra\textsuperscript{a,2}, Khokan Rana\textsuperscript{a,2}, Debudutta Bhattacharya\textsuperscript{a,2}, Srikantha Kanungo\textsuperscript{a,2}, Jaya Singh Kshatrii\textsuperscript{a,2}, Bijaya Kumar Mishra\textsuperscript{a,2}, Saroj Dash\textsuperscript{a,4}, Ajay Parida\textsuperscript{b,3}, Sanghamitra Pati\textsuperscript{a,2}

\textsuperscript{a} Indian Council of Medical Research Regional Medical Research Centre, Bhubaneswar, Odisha, India
\textsuperscript{b} Institute of Life Sciences, Nalco Square, Chandrasekharpur, Bhubaneswar, Odisha, India

A R T I C L E   I N F O

Article history:
Received 4 September 2020
Received in revised form 7 January 2021
Accepted 7 January 2021

Keywords:
SARS-CoV-2
Indian isolates
Spike
Receptor-binding domain
ACE2

A B S T R A C T

Objective: Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has affected the whole world, including Odisha, a state in eastern India. Many people have migrated to the state from different countries as well as other states during this SARS-CoV-2 pandemic. The aim of this study was to analyse the receptor-binding domain (RBD) sequence of the spike protein from isolates collected from throat swab samples of COVID-19-positive patients and further to assess the RBD affinity for angiotensin-converting enzyme 2 (ACE2) of different species, including humans.

Methods: Whole-genome sequencing for 35 clinical SARS-CoV-2 isolates from COVID-19-positive patients was performed by ARTIC amplicon-based sequencing. Sequence analysis and phylogenetic analysis were performed for the spike region and the RBD region of all isolates. The interaction between the RBD and ACE2 of five different species was also analysed.

Results: The spike region of 32 isolates showed one or multiple alterations in nucleotide bases in comparison with the Wuhan reference strain. One of the identical mutations, at position 1204 (Ref A, RMRC 22 C), in the RBD coding region of the spike protein showed stronger binding affinity for human ACE2. Furthermore, RBDs of all the Indian isolates showed binding affinity for ACE2 of different species.

Conclusion: As mutant RBD showed stronger interaction with human ACE2, it could potentially result in higher infectivity. The binding affinity of the RBDs for ACE2 of all five species studied suggests that the virus can infect a wide variety of animals, which could also act as natural reservoir for SARS-CoV-2.

© 2021 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Coronaviruses have been studied for more than 50 years and are known to infect multiple animal species, including humans. Although the pathogenesis of the diseases caused by these viruses and their mechanism of replication have already been well described because of previous outbreaks (severe acute respiratory syndrome coronavirus (SARS-CoV) in China in 2003 and Middle East respiratory syndrome coronavirus in Saudi Arabia in 2012), the current COVID-19 pandemic has propelled the whole world to study and investigate more deeply the pathogenesis of diseases caused by these viruses. Coronaviruses belong to a diverse virus family which consists of four genera: Alphacoronavirus, Betacoronavirus, Gammaronavirus and Deltacoronavirus (Fehr and Perlman, 2015). According to the nucleic acid sequence similarity, SARS-CoV-2, the organism that causes COVID-19, is a betacoronavirus. The spike glycoprotein of coronaviruses facilitates their entry into the host cell and also gives the virus a crown-like structure on its surface. Binding of pathogenic particles of SARS-CoV-2 with host cell receptors remains the crucial step for

https://doi.org/10.1016/j.ijid.2021.01.020
1201-9712/© 2021 The Authors. Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
initiation of infection. Besides, the ability of the virus to bind to the specific receptor of other host species is an essential requirement for transmission across different species (Lu et al., 2015). SARS-CoV interacts with human angiotensin-converting enzyme 2 (ACE2) for entry, and after the COVID-19 outbreak, researchers found that SARS-CoV-2 also interacts with ACE2 for entry into host cells (Li et al., 2003; Zhou et al., 2020). The SARS-CoV-2 spike protein is cleaved by host proteases into an S1 domain and an S2 domain, which mediate receptor recognition and membrane fusion, respectively (Lai et al., 2007). Wang et al. (2020) reported that the S1 domain of the spike protein of SARS-CoV-2 contains a receptor-binding domain (RBD), which forms a complex with human ACE2 and facilitates viral entry. However, emerging mutations in the SARS-CoV-2 genome might alter the processes of infection transmission and replication and the potential for viral attachment to ACE2. According to epidemiological data, SARS-CoV-2 was first identified in bats in Wuhan, China, and then spread to other parts after its zoonotic transmission via Malayan pangolins (Chinazzi et al., 2020; Zhang et al., 2020). In a country such as India with a diversified geographical distribution, it is important to understand the origin of different strains of SARS-CoV-2 isolated from different parts of the country. As interaction with ACE2 is the main pathway of entry of this virus into its host, knowledge of the RBD binding affinity of different Indian SARS-CoV-2 isolates for ACE2 of the natural reservoirs, including humans, is very important. However, this has remained a grey area with little information available. Our study aimed to analyse the spike sequences of Indian SARS-CoV-2 isolates collected from COVID-19 patients in Odisha (an eastern state in India) and further understand the interaction between their RBDs and ACE2 of different probable natural hosts of coronavirus: bats, pangolins, hamsters and humans. The mutant RBD of one isolate shows stronger binding affinity for human ACE2 than the wild-type RBD, providing important information regarding its virulence as well as drug targeting.

**Material and methods**

*Sequencing of different Indian isolates of SARS-CoV-2 from throat swab samples*

The current study was part of a whole-genome sequencing study performed by the Odisha Study Group, which consists of different government organizations: the Regional Medical Research Centre (RMRC), Bhubaneswar, and the Institute of Life Sciences, Bhubaneswar. As Indian Council of Medical Research RMRC, Bhubaneswar, is a government-authorized testing laboratory for COVID-19 testing, we received throat swab samples of suspected cases from different hospitals in Odisha. For COVID-19 diagnosis, viral RNA isolation from all samples was performed with a QIamp Viral RNA Mini Kit followed by quantitative PCR (TaqPath™ 1-step master mix, ThermoFisher Scientific). The libraries were prepared for whole-genome sequencing with use of ACTIC amplicon-based sequencing kits from Qiagen as per the manufacturer’s recommended protocol. Among all positive samples, whole-genome sequencing of 35 isolates from different

<table>
<thead>
<tr>
<th>Serial no.</th>
<th>Sample ID</th>
<th>RMRC spike nucleotide sequence identity with the reference (%)</th>
<th>RMRC RBD nucleotide sequence identity with the reference (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>RMRC 2</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>RMRC 5</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>RMRC 6</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>4</td>
<td>RMRC 7</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>RMRC 22</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>6</td>
<td>RMRC 23</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>RMRC 24</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>RMRC 27</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>RMRC 28</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>10</td>
<td>RMRC 30</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>11</td>
<td>RMRC 38</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>RMRC 46</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>13</td>
<td>RMRC 90</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>14</td>
<td>RMRC 103</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>15</td>
<td>RMRC 104</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>16</td>
<td>RMRC 106</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>17</td>
<td>RMRC 108</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>18</td>
<td>RMRC 110</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>19</td>
<td>RMRC 112</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>20</td>
<td>RMRC 154</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>21</td>
<td>RMRC 155</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>22</td>
<td>RMRC 156</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>23</td>
<td>RMRC 157</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>24</td>
<td>RMRC 158</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>RMRC 159</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>26</td>
<td>RMRC 160</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>27</td>
<td>RMRC 162</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>28</td>
<td>RMRC 163</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>29</td>
<td>RMRC 164</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>RMRC 165</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>31</td>
<td>RMRC 166</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>32</td>
<td>RMRC 167</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>33</td>
<td>RMRC 168</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>34</td>
<td>RMRC 170</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>35</td>
<td>RMRC 171</td>
<td>99</td>
<td>100</td>
</tr>
</tbody>
</table>
COVID-19 patients with different travel histories was performed with the Illumina platform. Detailed information and a description of the method can be found in Raghav et al. (2020).

**Sequence alignment of RMRC spike genes with the reference sequence**

For sequence alignment, the full-length genome sequence of SARS-CoV-2 isolate Wuhan-Hu-1 (accession no. NC_045512) was downloaded from the US National Center for Biotechnology Information (NCBI GenBank Database) database and used as the reference sequence for all further analyses. Alignment of all 35 RMRC spike nucleotide sequences with the reference genome was performed with BLASTN (align two/more sequences).

**Identification and phylogenetic analysis of the RBD of RMRC SARS-CoV-2 isolates**

From the data available for the reference strain, we retrieved information on the coding region of the spike RBD sequence, and all RMRC spike sequences were aligned with the reference sequence to identify the respective RBD coding region mutations by our using BLASTN. Multiple sequence alignment of all 35 RBDs (amino acid sequences) was performed with Clustal X. Mutations specific to RMRC isolates were identified by our comparing the RBD coding regions with the reference strain. A phylogenetic tree was generated with MEGA version 6 with 1000 bootstrap replications as in the instructions for MEGA.

**Figure 1.** (A) Multiple sequence alignment of receptor-binding domain (RBD) protein sequences of Regional Medical Research Centre (RMRC) severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) isolates. All the sequences were aligned with the reference strain, and the arrow indicates the alteration found in the RBD of the RMRC 22 isolate. (B) Phylogenetic analysis of RBD protein sequences of RMRC isolates using the neighbour-joining method. The bootstrap consensus tree inferred from 1000 replicates has been taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% of bootstrap replicates were collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The evolutionary distances were computed by the Poisson correction method and are in the unit of the number of amino acid substitutions per site.
isolated sequences of the were from C7ECV1, sphinx model human CoV-2. was percentage interaction Sequence ACE2 receptor-binding (UniProtKB 2. RBD.

Figure 2. Three-dimensional structure analysis and protein–protein interaction of angiotensin-converting enzyme 2 (ACE2) of *Manis javanica* with wild-type and mutant receptor-binding domain (RBD): (A) predicted 3D structure of ACE2; (B) Ramachandran plot and Z score; (C) ACE2 interaction with wild-type RBD; (D) ACE2 interaction with mutant RBD.

Sequence and 3D structure analysis of ACE2

From the phylogenetic tree analysis, four RMRC RBD sequences were selected from four random clusters for further investigation of their interactions with ACE2 of probable natural hosts of SARS-CoV-2. Hamsters were reported to be the most suitable animal model to perform SARS-CoV-2–related experiments; therefore, it was essential to understand the interaction of hamster ACE2 with isolated Indian SARS-CoV-2 RBDs (Imai et al., 2020). For the interaction study, hamster (*Mesocricetus auratus*) ACE2 (UniProtKB C7ECV1, 785 amino acids), pangolin (*Manis javanica*) ACE2 (NCBI XP_017505752, 805 amino acids), Chinese bat (*Rhinolophus sinicus*) ACE2 (UniProtKB E2DH17, 805 amino acids), Indian bat (*Cynopterus sphinx*) ACE2 (UniProtKB A0A6M3WQ69, 807 amino acids) and human (*Homo sapiens*) ACE2 (UniProtKB Q9BYF1, 805 amino acids) sequences were included. Before inception of structure prediction, the pangolin, hamster, Chinese bat and Indian bat ACE2 sequences were aligned with the human counterpart to identify the percentage of similarity and dissimilarity between these sequences.

The experimental 3D structure of human ACE2 was retrieved from the RCSB Protein Data Bank (PDB) (ID 6M0J) with a resolution of 2.45 Å positioning from 19 to 615 amino acids. The PDB did not provide any experimental structures of pangolin, hamster, Chinese bat or Indian bat ACE2, which prompted us to predict their 3D structure through homology modelling using Modeller version 9.19 followed by structure validation. Suitable templates were identified for 3D model building of ACE2 by a BLASTp search against the PDB (Altschul et al., 1990). The templates with PDB IDs 1R42, 6CS2, 6LZG, 3SCI and 2AJF were found to be good homologues for target–template alignment and modelled 3D structure prediction. On the basis of the optimized target–template alignment, Modeller 9.19 tool was used to predict the three dimensional structures of the proteins with unknown structures through homology modelling, the models with the lowest discrete optimized protein energy score were retained for further structural refinement (Webb and Sali, 2017). Side chain optimization was performed with WHATIF and GalaxyRefine (Hekkelman et al., 2010; Heo et al., 2013). The optimized models of ACE2 were finalized on the basis of the overall quality and stereochemical geometry and energy. The geometry of the predicted model was evaluated with PROCHECK and Ramachandran plot analysis (Laskowski et al., 1993; Pontius et al., 1996). The program ERRAT was used to calculate the accuracy of the non-
Figure 3. Three-dimensional structure analysis and protein–protein interaction of angiotensin-converting enzyme 2 (ACE2) of *Mesocricetus auratus* with wild-type and mutant receptor-binding domain (RBD): (A) predicted 3D structure of ACE2; (B) Ramachandran plot and Zscore; (C) ACE2 interaction with wild-type RBD; (D) ACE2 interaction with mutant RBD.

bonded atoms for the predicted model (Colovos and Yeates, 1993). Verify3D was used to evaluate the compatibility of the 3D model with its own amino acid sequence by our assigning a structural class based on its location and environment and comparing the results with good-quality structures (Bowie et al., 1991). The structure was uploaded to the Qualitative Model Energy Analysis (QMEAN) server to resolve the model quality. The energy potential of the predicted model was calculated with the ProSA-web server (Wiederstein and Sippl, 2007).

In silico translation of RBD sequences and their interaction with ACE2 of different natural reservoirs

Four nucleotide sequences of RMRC RBD coding regions were selected and translated to the protein sequence with use of the EMBOSS Transeq tool of the European Bioinformatics Institute. The 3D structure of the reference human RBD (UniProtKB P0DT2, amino acids 333–526 of the spike protein, PDB ID 6M0J) was considered as the wild type. The experimental structure of the wild-type RBD was altered at the position 402 (I402L) with use of Discovery Studio Visualizer (version 4.1) to obtain a mutant RBD (RMRC 22) as required for further computational analysis. Finally, protein–protein interaction between wild-type/mutant RBDs and ACE2 from different organisms was studied with the online HawkDock server, which is a powerful tool to predict the binding structures and identify the key residues of protein–protein interactions (Weng et al., 2019).

Results

Sequence information and analysis of the spike gene of Indian SARS-CoV-2 isolates

All SARS Co-V-2 isolates included in the study were from patients with a travel history from either outside the country or from other states in India. Among the 35 isolates, one was from a patient with a foreign travel history, 11 were from patients who had returned from Nizamuddin (cluster detected in New Delhi, India, during April 2020) and the rest (23) were from patients who had migrated from Surat, Gujarat, India. The detail of the demographic and clinical status of the patients included in the study has been described in one of our unpublished reports (Turuk
et al., unpublished). We did not record any deaths among the patients whose samples were included in the current study.

In the current study, the spike region was identified at the position 21,563 to position 25,384 of the whole genome and consists of 3822 nucleotides. The BLAST alignment analysis of RMRC spike nucleotide sequences showed that three spike sequences (RMRC 104, 157 and 158) were 100% identical to the Wuhan reference spike sequence, whereas all other RMRC spike sequences shared 99% identity, with one or multiple altered bases at different positions (Table 1).

Sequence identification, alignment and phylogenetic analysis of RBD

The alignment of the spike sequences of RMRC isolates with the reference strain RBD revealed the RBD region of Indian SARS-CoV-2 isolates is located in the 996–1578 bp region of the spike gene, consisting of a total of 582 bases. The protein sequence of the RBD spans from amino acid 333 (threonine) to amino acid 526 (glycine) of the spike protein (Lan et al., 2020). RBD sequence alignment results showed that RMRC 22 has 99% sequence identity with the reference RBD and harbours a mutation at position 1204 (nucleotide: Ref A, RMRC 22 C; protein: Ref I, RMRC 22L) (Figure 1A). Another three isolates shared 100% identity with the reference RBD (UniProtKB P0DTC2, amino acids 333–526 of the spike protein). As only RMRC 22 had a mutation, it was named a mutant isolate, whereas all other isolates were considered to be the wild type.

The phylogenetic analysis was performed with RBD sequences of all 35 RMRC isolates and showed that they formed four different clusters. RMRC 22, having a mutation, belonged to the first cluster, RMRC 171 RBD formed the second cluster, four RBDs (RMRC 2, 5, 6 and 7) formed the third cluster and all other RBDs were in the fourth cluster, which describes their phylogenetic distribution (Figure 1B).

Protein structure analysis and interaction of RBD with ACE2

The sequence alignment analysis showed that ACE2 of pangolin, hamster, Chinese bat and Indian bat shared 85%, 84%, 80% and 78% identity with human ACE2, respectively (Supplementary Figure 1).
The quality of the modelled structures of ACE2 was validated by several computational methods (pangolin in Figure 2A; hamster in Figure 3A; Chinese bat in Figure 4A; Indian bat in Figure 5A; human in Figure 6A). In pangolin, of 805 amino acid residues, the Ramachandran plot analysis illustrated 637 residues (88.0%) in the most favoured regions, 66 (9.1%) in additional allowed regions, 14 (1.7%) in generously allowed regions and 7 (1.0%) in disallowed regions (Figure 2B). In hamster, of 785 amino acid residues, the Ramachandran plot analysis illustrated 613 residues (87.2%) in the most favoured regions, 70 (10.0%) in additional allowed regions, 12 (1.7%) in generously allowed regions and 8 (1.1%) in disallowed regions (Figure 3B). In Chinese bat, of 805 amino acid residues, the Ramachandran plot analysis illustrated 634 residues (88.8%) in the most favoured regions, 66 (9.2%) in additional allowed regions, 12 (1.7%) in generously allowed regions and 2 (0.3%) in disallowed regions (Figure 4B). In Indian bat, of 807 amino acid residues, the Ramachandran plot analysis illustrated 630 residues (86.8%) in most favoured regions, 73 (10.1%) in additional allowed regions, 15 (2.1%) generously allowed regions and 8 (1.1%) in disallowed regions (Figure 5B). We determined the model quality using the QMEAN server. The overall quality of the model was good as indicated by its QMEAN Z score and QMEAN4 global score. Low-quality models are expected to have a negative QMEAN Z score. QMEAN4 ranges from 0 to 1, with a higher value indicating a good-quality model (Benkert et al., 2008). Additionally, the overall quality of the model was evaluated with ProSA-web, which provides a quality score (Z score) as compared with all known protein structures from X-ray crystallography as well structural NMR spectroscopy. The Z scores obtained were −6.33 (pangolin), −4.76 (hamster), −6.24 (Chinese bat) and −6.73 (Indian bat), which indicates the high quality of the models compared with known protein structures (Figures 2B, 3B, 4B and 5B).

As the 3D structure of human RBD was already available in the database (PDB ID 6M0J), we inserted the observed mutation of the RMRC 22 isolate (protein: Ref 1, RMRC 22L) and predicted the structure of the mutant RBD (Figure 6B). Since the specific binding of the viral RBD with host ACE2 determines the establishment of infection, we analysed the interaction of Indian SARS-CoV-2 RBDs (both mutant and wild type) with human ACE2 as well as ACE2 of other species which are reported to be natural reservoirs for this virus. The interaction analysis showed that the mutant RBD of the RMRC 22 isolate has stronger interaction with human ACE2 (Figure 6D) as compared with the wild-type RBD (Figure 6C). The interaction between mutant RBD and human ACE2 has a binding
energy of $-65.95$ kcal/mol, with six hydrogen bonds, whereas wild-type RBD–human ACE2 interaction has a binding energy of $-63.09$ kcal/mol, with four hydrogen bonds. For ACE2 of all other species (pangolin in Figure 2C and D; hamster in Figure 3C and D; Chinese bat in Figure 4C and D; Indian bat in Figure 5C and D), the wild-type RBD seems to have stronger binding affinity, with no difference in the number of hydrogen bonds (except for hamster, for which wild-type RBD–ACE2 has two hydrogen bonds and mutant RBD–ACE2 has three hydrogen bonds). The details of the interaction analysis with hydrogen-bond-forming residues and the average distance of hydrogen bonds are given in Table 2. However, if we compare the interaction of the mutant and wild-type RBD with ACE2 of all species, the interaction between the mutant RBD and human ACE2 is the strongest, with the highest binding energy and highest number of hydrogen bonds.

**Discussion**

The complex trajectory of the recent COVID-19 pandemic in India poses great risk towards control and containment of the infection. It is high time to understand its mobility pattern in the country and the viral genetic properties favouring virulence. Although during the early phase of the pandemic Odisha had comparatively very few positive cases as well as a small number of deaths, the virus has gradually become rapidly infectious, making the clinical scenario worsen. Many people from Odisha were working outside the state, and during the pandemic they returned to their home state for many reasons. The samples included in our study were collected mainly from individuals with suspected COVID-19 who had a travel history from different states or a foreign travel record.

As the spike protein of SARS-CoV-2 mediates viral entry into the host and houses the RBD, which binds to ACE2 of the host cell, understanding the spike RBD distribution in the genome of Indian isolates is crucial for therapeutic design. The SARS-CoV-2 spike protein is reported to have stronger binding affinity for ACE2 than the SARS-CoV spike protein, and higher affinity means a lower viral load is required to infect the cell, which may explain the high transmission of SARS-CoV-2 (Chen et al., 2020). In the current study, the spike region of 32 isolates showed altered nucleotide bases at multiple positions as compared with the Wuhan reference strain, suggesting mutations in these Indian isolates during the
spread. For SARS-CoV-2, specific RBD–ACE2 binding ensures infection as well as serves as a potential target for developing treatment strategies for this infection (Chen et al., 2020). According to Premkumar et al. (2020), the RBD of SARS-CoV-2 is an immunodominant and potential target of antibodies in COVID-19 patients. An earlier study confirmed the presence of 25 alterations in the spike protein of SARS-CoV-2 Indian isolates, and some of those were capable of altering secondary structure and dynamicity of the spike protein (Chand et al., 2020). Our earlier published work provided detailed information on the D614G alteration in the spike protein of these isolates, which falls parallel with the findings of other researchers (Chand et al., 2020; Raghav et al., 2020). However, the current study does not focus on the D614G alteration as the RBD sequence (which spans amino acids 333–526 of the spike protein) analysis is of prime importance rather than analysis of the whole spike protein. The mutation found in the RMRC 22 isolate in our study might play a role in altering the antigenicity or binding affinity of the respective RBD. The beauty of the current study is that it describes the role of this mutation in increasing the binding affinity of the RBD of this isolate for ACE2, which could be a potential cause of aggressive infection. Because of rapid spread and evolution, the SARS-CoV-2 RBD is known to acquire several alterations leading to increased binding affinity for human ACE2 (Ortega et al., 2020). In France, multiple alterations were identified in the RBD of SARS-CoV-2 contributing to higher receptor binding capacity, which might be responsible for increased virus spread and infectivity (Ou et al., 2020). On the other hand, an alteration in the spike protein has been found to be associated with a decrease in receptor binding affinity (Jia et al., 2020; Saha et al., 2020). In the current study, mutation in the RBD region of Indian isolates did not seem to affect interaction of the RBD with ACE2 of other species prominently except that of humans. Surprisingly, the patient from whom the RMRC 22 isolate was obtained had a travel history of returning from Nizamuddin (cluster detected in New Delhi, India, during April 2020) before he tested positive for SARS-CoV-2. However, no mutation was observed in the isolates obtained from his other family members (his father and two brothers), who were also COVID-19 positive. A study including SARS-CoV-2 isolated from Indian patients showed that a few genes, such as ORF6 and ORF10, completely lack any mutation and the E gene contains a single mutation, suggesting therapeutic strategies against these genes could be beneficial (Hassan et al., 2020). It appears that the emergence and role of a mutation in any region of the SARS-CoV-2 genome depends on multiple factors, including the geographical distribution, rate of spreading, alteration in the virulence of the virus and immune response of the host. The interaction analysis of mutant and wild-type RBDs with ACE2 indicated that bats and pangolins could be suitable natural reservoirs for Indian isolates of this virus, and this finding agrees with earlier reports (Lam et al., 2020; Zhou et al., 2020). As hamster has been reported to be a suitable animal model to study SARS-CoV-2 related pathogenesis, the interaction of the RBD of the Indian isolates included in the current study makes the earlier report more relevant (Imami et al., 2020). Although the susceptibility to infection and the death rate could be affected by several factors, mutation in the virus genome and its ability to adapt to a new environment could be crucial.

Being an important determinant in SARS-CoV-2 infection, RBD–ACE2 interaction has already become a potential target for developing treatment against this deadly pathogen. The current study provides important information regarding the structural basis of the spike and RBD regions of a few Indian SARS-CoV-2 isolates and gives an idea about their evolution and spreading. The mutation observed in the RBD region of one of the isolates sheds light on drug targeted therapy for different strains of the virus. Further studies with a larger number of isolates of a wider origin would be helpful to understand this mutation pattern in the RBD region of Indian isolates.

### Financial support and sponsorship

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

### Conflicts of interest

None.

### Ethical approval

Has been approved from state research and ethics committee. This work was done and was part of the Raghav et al. (2020) study.
Acknowledgments

We acknowledge the contribution of all the VRDL staff (Laboratory technicians, Multi tasking workers) whose relentless effort in COVID testing has given the strength to further plan up new experiments and studies. We also acknowledge the support of Dr. L.M. Ho for his contribution. We acknowledge Dr. Ira Praharaj and Dr. Sidhartha Giri for their expert advice and scientific inputs. We acknowledge Dr Namita Mahapatra who contributed in analysis during the study, however is no more with us and has succumbed to COVID-19.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.ijid.2021.01.020.

References

Analysis of Indian SARS-CoV-2 Genomes Reveals Prevalence of D614G Mutation in Spike Protein Predicting an Increase in Interaction With TMPRSS2 and Virus Infectivity

Sunil Raghav*†, Arup Ghosh†, Jyotirmayee Turuk2, Sugandh Kumar1, Atimukta Jha1, Swati Madhulika1, Manasi Priyadarshini1, Viplov K. Biswas1, P. Sushree Shyamli1, Bharati Singh1, Neha Singh1, Deepika Singh1, Ankita Datey1, Kiran Avula1, Shuchi Smita1, Jyotsnamayee Sabat1, Debdutta Bhattacharya2, Jaya Singh Kshatri2, Dileep Vasudevan1, Amol Suryawanshi1, Rupesh Dash1, Shantibhusan Senapati1, Tushar K. Beuria1, Rajeeb Swain1, Soma Chattopadhyay1, Gulam Hussain Syed1, Anshuman Dixit1, Punit Prasad1, Odisha COVID-19 Study Group1, ILS COVID-19 Team1, Sanghamitra Pati2 and Ajay Parida*

†These authors have contributed equally to this work

Coronavirus disease 2019 (COVID-19), caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, has emerged as a global pandemic worldwide. In this study, we used ARTIC primers–based amplicon sequencing to profile 225 SARS-CoV-2 genomes from India. Phylogenetic analysis of 202 high-quality assemblies identified the presence of all the five reported clades 19A, 19B, 20A, 20B, and 20C in the population. The analyses revealed Europe and Southeast Asia as two major routes for introduction of the disease in India followed by local transmission. Interestingly, the 19B clade was found to be more prevalent in our sequenced genomes (17%) compared to other genomes reported so far from India. Haplotype network analysis showed evolution of 19A and 19B clades in parallel from predominantly Gujarat state in India, suggesting it to be one of the major routes of disease transmission in India during the months of March and April, whereas 20B and 20C appeared to evolve from 20A. At the same time, 20A and 20B clades depicted prevalence of four common mutations 241 C > T in 5′ UTR, P4715L, F942F along with D614G in the Spike protein. D614G mutation has been reported to increase virus shedding and infectivity. Our molecular modeling and docking analysis identified that D614G mutation resulted in enhanced affinity of Spike S1–S2 hinge region with TMPRSS2 protease, possibly the reason for increased shedding of S1 domain in G614 as compared to D614. Moreover, we also observed an increased concordance of G614 mutation with the viral load, as evident from decreased Ct value of Spike and the ORF1ab gene.

Keywords: SARS-CoV-2, COVID-19, phylogeny, India, D614G, viral RNA sequencing, protein-protein interaction
INTRODUCTION

The coronavirus disease 2019 (COVID-19) pandemic is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a betacoronavirus belonging to the Coronaviridae family. The first occurrence of this novel coronavirus was observed in Wuhan, China, in late December 2019, which later spread globally via human-to-human contact transmission (Lu et al., 2020). According to the World Health Organization (WHO) weekly epidemiological update until September 14, 2020, the virus has spread to more than 180 countries with 28.6 million total confirmed cases and 0.9 million deaths worldwide (World Health Organization, 2020). The first occurrence of a coronavirus case in India was observed in mid-January, but the number of cases started to increase from the first week of March. According to the WHO, the number of cases in India has reached 4.7 million with 78,500 deceased (World Health Organization, 2020).

Genomic studies performed to understand the origin of SARS-CoV-2, a positive single-stranded RNA virus, unraveled that it has a zoonotic origin and is transmitted to humans from bats via Malayan pangolins (Zhang T. et al., 2020). The nucleotide sequence of SARS-CoV-2 is ~79% similar to SARS-CoV-1 and about 50% with MERS-CoV (Middle East respiratory syndrome coronavirus) (Lu et al., 2020). Approximately 30-kb genome of SARS-CoV-2 features a cap structure in 5′ and 3′ poly (A) like other members of the coronavirus family. A major portion of the genome is covered by two ORFs (ORF1a and ORF1b), which code for 15 non-structural proteins, including crucial proteins required for viral replication, such as viral proteases nsp3, nsp5, and nsp12, also known as RNA-dependent RNA polymerase (RdRP) (Kim et al., 2020). The genomic RNA (gRNA) also codes for structural proteins like Spike protein (S), nucleocapsid protein (N), membrane protein (M), and envelope protein (E), which are required for packaging of the virus and at least accessory proteins, but their ORFs are still not experimentally validated (Kim et al., 2020).

The entry of viral particles in the human body occurs through binding with angiotensin I-converting enzyme 2 (hACE2) receptor present on lung epithelial cells (Hoffmann et al., 2020). The predominant infection site is the respiratory tract due to the route of infection. Other than the lungs, it is known to infect other organs of the body, such as the kidney, liver, and intestine, as ACE2 expression is found to be quite high. After the viral entry into host cells, to initiate viral replication, negative-sense RNA intermediates are first synthesized by RdRP activity; these templates are then utilized for synthesis of gRNA and subgenomic RNAs (Kim et al., 2020). Because of the low fidelity of RdRP, mutations are incorporated with high frequency in gRNA, and such mutations are often known to increase the pathogenicity and fitness of the virus (Graepel et al., 2017; Ferron et al., 2018; Korber et al., 2020).

Recently, a predominant mutation, i.e., D614G in Spike protein, has been identified in virus strains sequenced from European population (Korber et al., 2020). There are several reports depicting that D614G mutation in Spike protein is associated with enhanced infectivity and spread of the virus owing to increased interaction with ACE2 receptor present on host cells (Korber et al., 2020). It has also been indicated that this mutation is present on the S2 domain of the Spike protein that is important for cleavage by TMPRSS2 enzyme for cleavage of S1 to facilitate the fusion of the viral Spike with the host cell membrane (Korber et al., 2020; Zhang L. et al., 2020). It is indeed interesting to understand further at a molecular level the changes in the protein structure induced by this mutation and its functional association with the infectivity and disease severity.

At the same time, it has been reported that D614G mutation co-occurs with three more mutations, i.e., 241 in UTR, 3307, and 14408 (Korber et al., 2020). The functional significance of these co-occurring mutations in evolution and virus selectivity is interesting to understand.

In the present study, we have sequenced 225 COVID-19 isolates from patient samples from the state of Odisha, those migrated from the 13 most affected Indian states as a part of DBT’s PAN-India 1000 SARS-CoV-2 RNA genome sequencing consortium, using an amplicon sequencing–based methodology. The travel history of these patients was collected along with their symptomatic and asymptomatic behavior. We performed phylogenetic analysis from sequenced data to understand the genetic diversity and evolution of SARS-CoV-2 in the Indian subcontinent. From the sequenced isolates, we have identified 247 single-nucleotide variants; most of them are observed in ORF1ab, Spike and nucleocapsid protein coding region. Moreover, we have analyzed the D614G mutations in the samples to obtain information on its evolution in Indian population. Protein-modelling analysis of D614G mutation was carried out to identify the impact on structural changes at protein levels. We further performed protein–protein docking simulation to predict the impact of D614G mutation on the interaction between wild-type and mutated Spike protein with TMPRSS2 enzyme to assess its impact on binding and perhaps viral infectivity.

MATERIALS AND METHODS

Sample Collection

All hospitalized and quarantined patients (March 2020 to June 2020), based on their clinical symptoms (fever or respiratory symptoms) or travel history, were preliminarily involved in this study. We received throat swabs in viral transport media samples of these patients used for SARS−CoV−2 detection. Patients with missing or with negative SARS−CoV−2 test results were excluded from this study based on Ct values obtained by quantitative polymerase chain reaction (qPCR) of isolated RNA. All patients involved in this study were residents of Odisha, India, during the outbreak period of COVID−19. The samples were collected and processed as per the guidelines of the Institutional Ethics and Biosafety Committee. Institutional Biosafety Committee (IBSC) approval (IBSC file no. V-122-MISC/2007-08/01) was taken before processing the samples in BSL3 laboratory.

Viral Load Detection

RNA isolation for all the 248 human patients was performed using QIAamp Viral RNA Mini Kit (Qiagen, cat. no. 52906). The isolated RNA was subjected to qPCR for determining viral load by
Ct values. For qPCR, we performed one-step multiplex real-time PCR using TaqPath™ 1-Step Multiplex Master Mix (Thermo Fisher Scientific, cat. no. A28526), targeting three different gene-specific primer and probe sets—envelope glycoprotein Spike (S), nucleocapsid (N), and open reading frame 1 (ORF1).

**Viral RNA Library Preparation and Sequencing**

We prepared amplicon libraries for viral genome sequencing using QIAseq FX DNA Library Kit and QIAseq SARS-CoV-2 Primer Panel (Qiagen, cat. no. 180475, cat. no. 333896) as instructed by the manufacturer’s manual, and the library was subsequently sequenced using Illumina platform. The adapter sequence used for each sample was compatible with Illumina sequencing instrument with 96-sample configurations (Qiaseq unique dual Y-adapter kit). The average insert length was in the 250–500 bp range. Prepared libraries were then pooled as a batch of 96 samples and sequenced using Illumina NextSeq 550 platform in 150 × 2 layout.

**Raw Data Preprocessing**

Quality of the sequenced files was checked using FastQC tool (0.11.9) (Andrews, 2010), followed by removal of low quality bases (–nextseq-trim, Q < 20), Illumina Universal adapter sequence and reads with less than 30-bp length using Cutadapt (2.10) (Martin, 2011). To access the quantity of host genomic DNA and other contaminants, Kraken (2.0.9-beta) (Wood et al., 2019) was used, and the reports were summarized using Kroma (2.7.1) (Ondov et al., 2011). All the files were then aligned to human genome (assembly version GRCh38) using HISAT2 (2.2.0) (Kim et al., 2015), and unmapped reads were extracted using SAMTOOLS (1.10) (Li et al., 2009) and converted to FASTQ format using BEDTOOLS (2.29.2) (Quinlan and Hall, 2010) bamToFastq option.

**Alignment With Viral Genome**

The unmapped reads were then aligned to SARS-CoV-2 reference assembly (NCBI accession NC_045512) using HISAT2 (2.2.0) (Kim et al., 2015). Amplicon primes from the aligned file were removed using iVar (1.2.2) (Grubaugh et al., 2019) guided by Artic Network V3 primer scheme1. The aligned files were then deduplicated using Picard Tools (2.18.7)2. Alignment quality was checked using SAMTOOLS (1.10) (Li et al., 2009) flagstat option.

**Consensus Sequence Generation and Variant Calling**

A consensus sequence for each isolate was generated using Bcftools (1.10) (Li, 2011) and SEQTK3. After generating a reference-based consensus sequence, we selected 202 genomes with less than 5% Ns and more than 10× coverage for phylogenetic and mutation analysis. Single-nucleotide variants were called and filtered (QUAL > 40 and DP > 20) using Bcftools (1.10) (Li, 2011). Effects of the filtered variants were annotated using SnpEff (4.5) (Cingolani et al., 2012). All of the consensus sequences were deposited in GISAID (Shu and McCauley, 2017); accession IDs are provided in Supplementary Table 2.

**Phylogenetic Analysis**

Phylogenetic tree analyses of all the samples were performed using SARS-CoV-2 analysis protocol standards and tools provided by Nextstrain (Hadfield et al., 2018) pipeline. First, all the sequences are aligned against the WH01 reference genome using Augur wrapper of MAFFT (Katoh and Standley, 2013), and low-quality variant sites are masked from the alignment. The initial maximum likelihood tree was generated by the IQTREE2 tool (Nguyen et al., 2015) with 1,000 bootstraps. We have provided the maximum likelihood tree with bootstrap percentage marked for branches with ≥ 60 support and clade information in Supplementary Figure 3. Further refinement of the tree was done using the Augur refine command, and the tree was rooted using the reference sequence with timeline information incorporation using TimeTree (Kumar et al., 2017). To finalize the tree for Nextstrain auspice visualization the tree was annotated using ancestral traits, clades, nucleotide mutation and amino acid mutation. The resulting tree was visualized using an Auspice instance, and the visualization was refined using the ggtree R package.

We have used Nextstrain year-letter clade nomenclature that started with 19A and 19B branched by C8782T and T28144C nucleotide changes and was initially prevalent in Asia during initial outbreak. Later, 20A emerged in European outbreak from 19A parents having C3037T, C14408T, and A23403G as distinctive features. 20B emerged as a distinct clade in Europe with three consecutive mutations, e.g., G28881A, G28882A, and G28883C. Further the 20C emerged as a North America–specific clade with C1059T and G25563T nucleotide changes.

**Haplotype Network Analysis**

For haplotype network analysis, we took a total of 287 (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, and South Korea) SARS-CoV-2 whole-genome sequences with less than 1% N and with collection date of March, April, and May from GISAID database. The selected samples are then aligned to the WH01 reference genome using MAFFT (Katoh and Standley, 2013). After filtering aligned sequences, we used POPART (Leigh and Bryant, 2015) software to generate haplotype network using a median joining method with 2,000 iterations.

**Modeling of the Protein Structures**

The sequences of SARS-CoV-2 proteins (NPS3, NSP4b, NSP6, nucleocapsid, and Spike) were retrieved from NCBI. As most of the proteins do not have a three-dimensional (3D) structure in protein data bank (PDB), they were modeled using Modeler 9.21 (Webb and Sali, 2016). The suitable templates for modeling of the proteins (Supplementary Table 1) were selected by DELTA-BLAST (Boratyn et al., 2012) against the PDB proteins. One hundred models were generated for each of the proteins. The best model was selected based on the lowest DOPE score (Shen and 1https://github.com/artic-network/artic-ncov2019/tree/master/primer_schemes/nCoV-2019/V3
2https://broadinstitute.github.io/picard/
3https://github.com/lh3/seqtk
Sali, 2006). The D614G mutant of the Spike protein was generated by Modeler 9.21. The loop in Spike protein (670–690) was refined using loop modeling procedure in Modeler by generating 100 loop models. The model with the lowest dope score was finally chosen for both the mutant and wild-type protein. Similarly, the host transmembrane serine protease 2 (TMPRSS2) was also modeled using Modeler. The PROCHECK (Laskowski et al., 1993) server was used to assess the stereochemistry of the generated models.

Protein–Protein Docking
The standalone version of HADDOCK2.2 (van Zundert et al., 2016) was used to perform protein–protein docking with SARS-CoV-2 Spike protein with human TMPRSS2. The docking was conducted by the restraining of the receptor (Spike) and ligand (TMPRSS2) residues known to be at the interface Spike-TMPRSS2 interface. Specifically, residues within 5 Å of the reported cleavage site (Arg685, Ser686) (Hoffmann et al., 2020) of the Spike and catalytic triad of TMPRSS2 (H296, D345, and S441) and binding residue D435 were restrained to be at the docking interface. A total of 100 docking poses were generated and ranked based on the HADDOCK2.2 docking score (van Zundert et al., 2016), which is composed of van der Waals energy, electrostatic energy, restraints energy, etc. The best-ranked docking pose was visualized using Pymol.

Statistical Analysis and Plotting
All the statistical analysis and plots were generated in R (3.6.1) statistical programming language using ggplot2, dplyr, reshape2, lubridate, ggsci, and ggpubr package available from CRAN and Bioconductor repository.

RESULTS

Demographics, Clinical Status, and Travel History
The average age of the 225 patients was 30.98 ± 11.79 years with age range 1–75 years and median age of 30 years. The overall gender ratio of male-to-female was 201:24 with median age range 1–75 years and median age of 30 years. The average age of male patients 31.75 and 24.5 years for female patients overall gender ratio of male-to-female was 201:24 with median age range 1–75 years and median age of 30 years. The Travel History
In patients, we performed phylogenetic analysis of our dataset with 1,042 high-coverage Indian SARS-CoV-2 whole-genome sequences (N < 1%) obtained from GISAID on July 12, 2020 (Supplementary Figure 1F). From the transmission map generated using Nextstrain, we observed that most of the sequences in 20A clades share sequence similarity with samples from Gujarat, which has very high prevalence of this clade (Supplementary Figure 1E). The 19A clade is very prevalent in Delhi and Telangana region and Tamil Nadu, but the 20B clade is only prevalent in the southern parts of India (Supplementary Figure 1E). Both 20A and 20B clades predominantly contain a mutation in protein coding sequence 23403A > G, and the mutation is traced back to the West European region in late January (Korber et al., 2020). The missense mutation in Spike protein coding gene causes a change in 614 D > G position of Spike protein and reported to increase the shedding of S1 subunit of the protein, which leads to the increased infectivity (Korber et al., 2020).

Mutation Analysis
After filtering out low-quality genomes with <5% N’s in the assembly and at least 10 × average coverage, we observed a total of 247 single-nucleotide variants from 202 SARS-CoV-2 isolates. Of these variants, 156 variants observed only in single isolates, 25 variants were classified as common variants with occurrence of more than 5% and 19 variants as rare with 2–5% occurrence (Supplementary Table 1). Among the common variants, the most frequent mutations are 23403 A > G (D614G, S gene), 241 C > T (5’ UTR), 14408 C > T (P4715L, RdRP gene), 3037 C > T (F942F, NSP3), 28881 G > A (R203K, N gene), 28882 G > A (R203R, N gene), 28883 G > C (G204R, N gene), and 28144 T > C (L84S, ORF8) with presence in more than 15% of all of our sequenced genomes (Supplementary Table 1). Plotting mutation diversity (>2%) in a clade-wise manner, we
observed distinct mutation signatures in different clades. The isolates that were grouped in 19A clade depicted prevalence of mostly ORF1ab mutations with one distinct N gene $C > T$ mutation at 28311 position (Figure 2C). In clade 19B samples, we observed a very distinct ORF8 $T > C$ mutation at 28144 position, two N gene mutations at positions 28326 and 28878 with some ORF1ab mutation in lower frequency (Figure 2C). Clades 20A and 20B have almost similar mutation profile with
FIGURE 2 | SARS-CoV-2 clade distribution and their prevalent mutation profiles. (A) Dot plot representing the number of single-nucleotide mutation (occurred in more than 2% of the samples) present in different genomic segments of SARS-CoV-2 genome. (B) The ORF1ab region codes for a polypeptide which is later cleaved to several mature peptides. The dot plot represents the amino acid changes (location of amino acid acids as per location in polypeptide sequence) in the mature peptides of ORF1ab. (C) Clade-wise occurrence of nucleotide mutations with presence in more than 2% of sequenced samples (n = 202). Color of the dots represents the clade and size of the dots represents number of the samples showing presence of the single-nucleotide variant. (D–I) The mutation sites on the modeled structures of the SARS-CoV-2 proteins. The mutation site(s) of the NSP3, NSP4b, NSP6, RdRP, and nucleocapsid proteins are marked as sphere, while the rest of the structure is shown in cartoon representation.
major mutated positions 241 C > T mutation in leader sequence, ORF1ab 3037 C > G, ORF1ab 14408 C > T (RdRP), and S gene 23403 A > G mutations. The very distinct characteristic that we observed for 20B clade is three consecutive N gene mutations at positions 28,881, 28,882, and 28,883 (Figure 2C). Of these three mutations, two are missense mutations resulting in change of protein sequence. In clade 20A, we observed two mutations at ORF3a and M protein coding gene at position 25,563 G > T; 26,735 C > T but in less frequencies in comparison with other mutated sites (Figure 2C). Overall, to conclude, we summarized all the mutated sites present in all the samples and observed ORF1ab is the most mutated region, followed by N gene, S gene, and ORF8 in SARS-CoV-2 samples from India (Figure 2A). Among the ORF1ab mutations, 4,715 P > L change in nsp12, also known as RdRP (the viral RNA dependent RNA polymerase), was the most common one followed by a synonymous change (F924F) in nsp3 protein (Figure 2B).

To understand the impact of identified prominent missense mutations (present on > 10% of the samples) on the annotated viral proteins in our sequenced population, we used protein crystal structures and protein models as crystal structures were not available. The PROCHECK (Laskowski et al., 1993) results showed that the generated models have acceptable stereochemistry. First, we looked into the location of the mutated amino acid with respect to its functional domains as any change in the functional domain has high probability to perturb the protein function. The sites of mutation on the modeled protein structures (nucleocapsid, NPS3, NSP4b, NSP6, ORF8, and RdRP) are marked to indicate their location (Figures 2D–I).

Haplotype Network Analysis

First, we created a median-joining haplotype network to look for transmission within India, and we observed two major branches (Supplementary Figure 2A). When we colored the sequences with clade information, the branches represented 19A and 20A, which later furcate into 19B, 20B, and 20C (Supplementary Figure 2A). When we overlaid migration information in the network, we observed the majority of 19A and 19B isolates migrated from Gujarat (Figure 3A). Migration of isolates identified in newly prevalent clades occurred from the southern path of India (Figure 3A). To understand the transmission source of SARS-CoV-2 infection in India, we constructed haplotype network using our sequencing data combined with genome sequences obtained from GISAID (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, and South Korea). From the haplotype network, we observed distinct clusters of genome sequences that were grouped in four major nodes. A large group of sequences clustered in two major haplotype clusters, one with genome sequences from China, Singapore, and South Korea and the other one with Italy, Saudi Arabia, and Germany with 2- to 4-nucleotide substitutions (Figure 3B). From the collection date, we observed that 20A clade, which is prevalent in Europe, became abundant in Odisha from April, representing the top cluster, whereas the bottom cluster represents samples belonging to clades 19A and 19B having a common origin in Southeast Asia (Figure 3B).

Effect of D614G on Viral Load

To understand the prevalence and effect of G614 in our sequencing dataset, we plotted week-wise count (based on collection date) of D614 and G614 and observed that the occurrence of the mutation has been observed in early March (12th week), and the cumulative frequency increased over time (Figure 4A). We also assessed the frequency of D614G mutation from Covid19 Beacon database (CSIRO and CSIR-IGIB, access date: February 8, 2020) in global and all the SARS-CoV-2 sequence published from India, and the occurrence of G614 in Indian genomes is 76.31% where the global frequency is 43.8% (Supplementary Figure 2C). Every sample we sequenced as a part of the study was also checked for the levels of ORF1 and S gene using qPCR. The Ct obtained is also a direct indicator of viral load (lesser the Ct, higher the viral load) in the individual. When we plotted the Ct values of all sequenced samples, we observed that except week 21, the Ct values of the isolates having G614 mutation are less in comparison to isolates having D614 (Figures 4B,C). We also had Ct values (ORF1, S gene) of 637 positive isolates available as a partner institute of the COVID19 surveillance program in Odisha, India. Plotting the data against the date of sample collection for testing, we observed a sharp and significant (p < 0.05) decline (median ~5 Ct change) in the Ct values (Figure 4E) from April 2020 to May 2020, and there was median ~1 Ct change of S gene, as well as ORF1ab (Figures 4D–F) from the month of May to June 2020. In the case of ORF1 expression, we also observed a sharp change between April and May, but there is almost no change between May and June 2020 (Supplementary Figure 2A).

Molecular Modeling Depicted Enhanced Interaction of D614G-Mutated Spike Protein With TMPRSS2 Protease

As reported in earlier publications, we also noticed D614G as a highly prevalent mutation in highly transmitted and evolved strains belonging to clades 20A and 20B in the Indian scenario; therefore, we carried out molecular modeling analysis. It was observed that the wild-type (D614) and mutated form (G614) of Spike protein showed a slightly different arrangement of structural elements near the mutation site, which is present in hinge region linking S1 and S2 domain, i.e., S1 furin cleavage site (Figure 4G). The S2 domain of the Spike protein is reported to interact with TMPRSS2 protease necessary for shedding of S1 domain for viral entry inside the host cells by facilitating the merging of virus with the host cell membrane. The proteolytic cleavage of Spike protein by TMPRSS2 results in the shedding of S1 domain, which is one of the key steps of the virus infection in host cells. Cryo–electron microscopy structures indicated that side chains of D614 protomer and T859 of the neighboring protomer form a hydrogen bond in between bringing together S1 domain with S2 (Lu et al., 2020; Walls et al., 2020). This substitution could modulate the glycosylation at N616 site as well, perturbing the interaction between the neighboring protomer. Our Spike protein model depicted this hydrogen bonding between D614-T859 of S1 and S2 domain.
FIGURE 3 | Haplotype network analysis of SARS-CoV-2 sequences. (A) Haplotype network of 202 SARS-CoV-2 whole-genome sequences from our dataset colored by their respective place of migration. (B) Haplotype network of 100 high-coverage SARS-CoV-2 genomes obtained from GISAID (China 15, Germany 23, Italy 25, Saudi Arabia 23, Singapore 14, South Korea 17) combined with 170 samples sequenced from Odisha with less than <5% N’s present in consensus sequence.
Raghav et al. Prevalence of D614G in Indian SARS-CoV-2 Genomes

**FIGURE 4 |** D614G in Spike gene increases infectivity portrayed by Ct values as a surrogate for viral load. (A) Cumulative count of the occurrence of D and G in 614 position of Spike protein in sequenced genomes (\( n = 202 \)). (B, C) Ct value distribution of S gene and ORF1ab for the sequenced genomes (\( n = 202 \)). (D–F) Ct value distribution of S gene and ORF1ab in all the positive samples tested at Institute of Life Sciences until June 17, 2020. (G–I) The superimposed 3D structures G614 mutant and wild-type Spike protein. (G) The mutant site is highlighted with a circle at 614 position. (H) The hydrogen bond (D614-T859) shown as dotted line between Spike S1 and S2 domain in wild type. (I) The hydrogen bond is lost as a result of D614G mutation.

(Figure 4H). The mutation of D614 to G614 eliminates this side-chain hydrogen bonding between S1 and S2 domain (Figure 4I), leading to increased main-chain flexibility enabling a more favorable orientation of Q613, possibly facilitating cleavage by TMPRSS2 by perturbing its affinity with the S1-furin cleavage site. It has also been proposed that D614 forms an intrasalt bridge...
with R646, which makes the conformation unfavorable for S1 association with S2 domain (Zhang L. et al., 2020). Interestingly, the protein docking analysis depicted better hydrogen bonding interactions between the Spike protein cleavage sites (Arg685, Ser686) with the catalytic triad of TMPRSS2 in mutant condition as compared to wild-type. In the case of mutant Spike protein, the Arg682 and residues at primary cleavage site of Spike protein (Arg685 and Ser686) formed six hydrogen bonding interactions with Gln299, Lys300, Asp338, and Gln438 residues of TMPRSS2 (Figures 5A,B), whereas in the D614 wild-type form there were five hydrogen bonds observed between the cleavage site of Spike protein S2 domain and TMPRSS2 (Figures 5C,D). The binding energy was observed to be better for the G614 mutant (−143.03 kcal/mol) as compared to that of the wild type (−113.67 kcal/mol), indicating better binding of TMPRSS2 with the mutated Spike protein (Figure 5E).

**DISCUSSION**

The pattern of COVID-19 pandemic spread in India is quite different from all over the world in terms of slow transmission...
and lower mortality rates in the beginning. Therefore, it was extremely important to understand the transmission dynamics of SARS-CoV-2 and its evolution during different phases of disease in India. The mutations that accumulate in the virus genome with time act as a molecular clock that can provide insight into emergence and evolution of the virus. These analyses could be helpful to prevent or control the transmission of the virus. To track the COVID-19 outbreak and understand the genomic clades of SARS-CoV-2 prevalent in India as compared to rest of the world, we performed the genome sequencing of oropharyngeal and nasopharyngeal swabs samples collected from individuals migrated from different regions of India to the state of Odisha after the reported incidences of COVID-19 pandemic in India.

For SARS-CoV-2 genome sequencing, we did amplicon-based sequencing for 225 SARS-CoV-2 genomes based on their assembly coverage (<5% gap) for further detailed analysis. We selected migrant groups from different parts of the country so that diversity of COVID-19 prevalent in the country could be captured in the phylogenetic analysis to understand disease transmission and virus evolution with time. The sequenced samples represent migratory populations from North, East, West, and Southern parts of India. Our analysis depicted that all the genomes analyzed in our study were grouped into four major clades, 19A, 19B, 20A, and 20B, according to the new Nextstrain clade nomenclature. Clade 19A is the Wuhan clade from China. Interestingly, we were able to capture occurrence of a rare clade with very less occurrence in India, i.e., 19B with 17% (n = 36) in our samples. This clade was found to be prevalent in East and Southeast Asia during the early outbreak of the pathogen. This confirmed that the foundation of the source of early outbreak infection in India also came from Southeast Asian countries. The migration information for these early collected samples were not well defined; therefore, it was difficult for us to pinpoint exactly which Southeast Asian country the transmission of 19B started. The mutation analysis depicted that both 19A and 19B clades almost evolved in parallel as prevalent mutations in both the clades are highly variable. On the other side, the phylogenetic analysis showed that clades 20A and 20B evolved quite rapidly in the Indian population and are a major source of disease transmission in the country, whereas the 20C strain is rarely detected and appeared to be less adapted or somehow contracted by contact tracing at early stages of infection. The haplotype network construction also pointed to the later strains belonging to 20A; 20B clades originated from Western Europe and transmitted directly or via Saudi Arabia are mostly prevalent in the southern and western part of India. In the clade, we also observed a very less frequent clade 20C, prevalent only in a handful of the Middle Eastern countries making up a marginal portion (n = 4) of our sequenced samples. This suggests the requirement of constant monitoring of SARS-CoV-2 with sequencing technology to understand the source of infection and design prevention mechanisms such as strategic lockdowns and region specific travel restrictions.

The fitness of the virus strain and its transmission depend on the adaptive mutations that it acquires with time. From initial observation, we have seen a 10.34% occurrence of C6312A, which has been associated with an India-specific clade called I/A3i (Banu et al., 2020). The mutation is dominant only India as the global frequency of the mutation is around 1.2% according to the Covid19 Beacon database. We found that four common variants, i.e., 241 C > T in the UTR region, 3,037 C > T in NSP3 gene, 14,408 C > T in the NSP12, and 23,403 A > G in S gene coevolved mostly in the 20A and 20B clades. As 20A and 20B clade frequency increased with time in the population, which indicates that these strains have some selective advantage with time for increased transmission. It has been reported that the leader sequence present in the UTR region of positive strand RNA viruses like SARS-CoV is important for the replication and strand switching to generate negative strands (Kim et al., 2020). This mutation in leader sequence is 20 nucleotides upstream of translation start site of ORF1ab gene. Therefore, it might be providing an advantage to virus in preventing stem-loop generation required during strand switching by RdRP, which needs further experimental evaluation. At the same time, several reports documented that 23,403 A > G (D614G) missense mutation in the Spike protein enhanced the infectivity rate of the virus. One of the reports showed that the shedding of S1 domain of Spike protein changes due to change in the hydrogen bonding between S1 and S2 domains. We observed in our protein modeling analysis that TMPRSS2 binding to Spike protein is enhanced by this mutation of aspartic acid to glycine (D614G) as it resulted in increased hydrogen bonding interactions. This change enhances the interaction of TMPRSS2 with the S2 domain, which is important for the cleavage of the S1 domain and virus entry into cells by facilitating its entry into the host cells. The overall primary in silico docking study showed that mutant Spike protein has a greater number of hydrogen bonds with TMPRSS2 at the cleavage site as compared to the wild type, resulting in better docking energy. Overall analysis indicates that the breakage of a hydrogen bond as a result of the mutation may facilitate greater cleavage of the mutant Spike protein as compared to the wild type. All the sequenced genomes were submitted in the GISAID database.

ODISHA COVID-19 STUDY GROUP
(AUTHOR NAMES ARE ARRANGED IN ALPHABETICAL MANNER)

**ILS COVID-19 GROUP (AUTHOR NAMES ARE ARRANGED IN ALPHABETICAL MANNER)**


**DATA AVAILABILITY STATEMENT**

In the data availability statement the name of the Online repository is GISAID (https://www.gisaid.org/) accession ids are available in Supplementary table 2.

**ETHICS STATEMENT**

The studies involving human participants were reviewed and approved by the Institutional Human Ethics Committee, Institute of Life Sciences. Written informed consent to participate in this study was provided by the participants’ legal guardian/next of kin.

**AUTHOR CONTRIBUTIONS**

AP, SP, SR, and JT planned and designed the study. AJ, SM, MP, VB, PS, BS, NS, DS, AD, SK, KA, SSm, and JS did the experiments. SC, GS, RD, SSe, RS, TB, and PP coordinated sampling and COVID-19 testing analysis. AG and SR did the genomic data analysis. SK and AD performed protein modeling and docking analysis. All authors have read and approved the manuscript.

**REFERENCES**


**FUNDING**

Authors would like to acknowledge Department of Biotechnology, India for providing the funding and approvals under the umbrella of DBT’s PAN-India 1000 SARS-CoV-2 RNA genome sequencing consortium, ILS intramural grant for supporting for sequencing resources and Odisha state authorities to provide the clinical samples for testing where the virus RNA were isolated for sequencing analysis. PS and SM were supported by the DST inspire fellowship.

**ACKNOWLEDGMENTS**

We acknowledge GISAID, the genomic data sharing platform for influenza viruses and the contributors sequencing data used in this study. We gratefully acknowledge the efforts of volunteers who helped in coordinating the testing and related activities that supported implementation of this work.

**SUPPLEMENTARY MATERIAL**

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.594928/full#supplementary-material

Supplementary Figure 1 | Patient demographics, transmission map and diversity of COVID-19 in India. (A–D) Represents the distribution of gender, age, clinical status and geographical location of sampled patients (n = 225). (E–F) Transmission map and Timetree of the 1,042 high coverage Indian SARS-CoV-2 whole genome sequences (N < 1%) obtained from GISAID on 12th July, 2020.

Supplementary Figure 2 | ORF1ab Ct values in tested positive samples (n = 637) and haplotype network of sequenced data colored by clade. (A) ORF1ab Ct values in tested positive samples (n = 637) binned in their respective collection months. (B) Haplotype network of 202 SARS-CoV-2 whole genome sequences from our dataset colored by their respective clade. (C) Comparison of D614G mutation in global and Indian SARS-CoV-2 sequences.

Supplementary Figure 3 | Maximum likelihood tree of SARS-CoV-2 genomes (n = 202). (A) Maximum likelihood tree of SARS-CoV-2 genomes with 1,000 bootstraps and branch support values with >60 threshold.

Supplementary Table 1 | Information about patients, List of single nucleotide mutations and the annotated genes, corresponding amino acid changes and its predicted impact on protein function.

Supplementary Table 2 | Accession ids of GISAID submissions.


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Raghav, Ghosh, Turak, Kumar, Jha, Madhusudana, Priyadarshini, Biswas, Shyamli, Singh, Singh, Singh, Datey, Avula, Smita, Sabat, Bhattacharya, Khatri, Vasudev, Suryawanshi, Dash, Senapati, Bearia, Swain, Chattopadhyay, Syed, Dixit, Prasad, Odisha COVID-19 Study Group, ILS COVID-19 Team, Pati, and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
Identification of multipotent drugs for COVID-19 therapeutics with the evaluation of their SARS-CoV2 inhibitory activity

Sugandh Kumar a,b,1, Bharati Singh a,b,1, Pratima Kumari a,d, Preethy V. Kumar a,b, Geetanjali Agnihotri c, Shaheerah Khan a,d, Tushar Kant Beuria a, Gulam Hussain Syeda c, Anshuman Dixita a,a

a Institute of Life Science, Nalco Square, Bhubaneswar, Odisha 751023, India
b School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT) University, Bhubaneswar, Odisha 751024, India
c School of Chemical Technology, Kalinga Institute of Industrial Technology (KIIT) University, Bhubaneswar, Odisha 751024, India
d Regional Centre for Biotechnology (RCB), 3rd Milestone, Faridabad-Gurgaon, Haryana 121001, India

A R T I C L E   I N F O
Article history:
Received 13 January 2021
Received in revised form 4 April 2021
Accepted 4 April 2021
Available online 07 April 2021

Keywords:
Coronavirus
COVID-19
Drug repurposing
Network analysis
Docking
Polypharmacology
Molecular dynamics

A B S T R A C T
The SARS-CoV2 is a highly contagious pathogen that causes COVID-19 disease. It has affected millions of people globally with an average lethality of ~3%. There is an urgent need of drugs for the treatment of COVID-19.

In the current studies, we have used bioinformatics techniques to screen the FDA approved drugs against nine SARS-CoV2 proteins to identify drugs for repurposing. Additionally, we analyzed if the identified molecules can also affect the human proteins whose expression in lung changed during SARS-CoV2 infection. Targeting such genes may also be a beneficial strategy to curb disease manifestation. We have identified 74 molecules that can bind to various SARS-CoV2 and human host proteins.

In the end of 2019 a highly infectious novel coronavirus, which was later renamed as severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), emerged, which causes the disease termed “coronavirus disease-19” (COVID-19) [2]. COVID-19 has affected millions of people globally and the number is increasing alarmingly with the continuous surge in the number of cases in many countries world wide. COVID-19 is associated with an average lethality of ~3%. The elderly and those with comorbidities are at high risk of developing severe clinical manifestations. Infection with SARS-CoV2 results in acute respiratory distress syndrome (ARDS) leading to lung injury, respiratory distress, and lethality. The extremely infectious nature of the disease, emergence of new hyperinfective strains, limited supply of vaccines and unavailability of effective and specific drugs is a serious cause of concern.

The SARS-CoV2, SARS-CoV1 and MERS-CoV belongs to the family of Coronaviridae and β-coronavirus genus [3]. While bats are considered to be the origin of SARS-CoV1 and SARS-CoV2, the intermediate host that led to human transmission of SARS-CoV2 is still unknown. Sequence analysis reveals that SARS-CoV2 is similar to coronavirus identified in Malayan pangolins (Manis javanica) [4]. The SARS-CoV2 genome is 29.8–29.9 kb positive-sense single stranded RNA with 5'-cap and 3'-poly-A tail. Its genome is organised into two segments that encode non-structural (Nsp) and structural proteins. The first segment is directly translated by ribosomal frameshifting into polyprotein 1a (486 kDa) or 1ab (790 kDa) (ORF1a, ORF1ab), which results in synthesis of non-structural proteins and formation of replication-transcription complex (RTC) [5,6]. The ORF1a/1ab covers two-thirds of the whole genome.
genomic length and encodes for the 16 non-structural proteins (Nsp1-16), which play critical roles in various viral processes. The discontinuous transcription of the second segment of viral genome results in formation of subgenomic RNAs (sgRNAs) containing common 5’-and 3’-leader and terminal sequences which serve as the template for subgenomic mRNA production [6]. The subgenomic mRNA’s encode for the 4 structural proteins (spike (S), membrane (M), envelope (E), and nucleocapsid (N) proteins) and 6 accessory proteins. [6]. The life cycle of SARS-CoV2 starts with its entry into the host cell through endocytosis initiated by its spike protein binding to the ACE2 receptor [7]. Subsequently, uncoating of the virus particle releases the genome, which is translocated to generate replication-transcription complex proteins. The viral RTC complex then generates full-length negative sense RNA that is subsequently transcribed into full-length genome. The viral genome and structural proteins are assembled into virions near the ER and Golgi interface and are transported out of the cell through vesicles by exocytosis [8].

The detailed understanding of the clinical manifestations and the underlying molecular mechanisms that drive disease pathogenesis are still unclear. There is no standard cure for the disease and the current COVID-19 therapeutic guidelines approved by FDA recommends the use of remdesivir and symptomatic treatment in hospitalized patients (www.covid19treatmentguidelines.nih.gov). Worldwide efforts to develop vaccines and drug against SARS-CoV2 are ongoing. While some of the vaccine candidates are being approved for use, however it will take many months to validate their efficacy and safety in a large population. The ever-evolving mutant strains also pose a risk of making the vaccine ineffective. Moreover, it has been estimated to take 2–3 years to vaccinate majority of the people on the globe. The ever-growing COVID-19 trajectories mandates to identify new therapeutics with potent SARS-CoV2 inhibitory activity. The repurposing of approved drugs is among the best and rapid strategies to identify potential therapeutics [9]. In this context, the computational techniques can help quickly identify novel molecules that target viral proteins to suggest candidates for repurposing. Hence, during the COVID-19 pandemic a lot of studies have been reported using a variety of such strategies [10–12].

The in-silico studies have helped in identification of many drugs that can target viral proteins viz. RNA-dependent RNA polymerase (RdRp), Spike, proteases (3CLpro and PLpro) and human proteins such as angiotensin converting enzyme 2 (ACE2), transmembrane serine protease (TMPRSS2), and PIKfyve etc. [13]. Among them zanamivir, indinavir, saquinavir, and lopinavir are notable [14,15]. There are many drugs such as baricitinib [16], favipiravir [17], duvelisib, ivermectin [18] and arbidol [19] etc., that are currently under clinical trials (https://clinicaltrials.gov/) to treat SARS-CoV-2 infection. In the early stage of the COVID-19 pandemic, remdesivir and hydroxychloroquine were widely used for treatments of SARS-CoV2. Later on, it became evident that despite the efficacy observed in-vitro, hydroxychloroquine did not show any clinically relevant benefits [20]. Recent reports suggest that the SARS-CoV2 not only causes infection in the lungs but may also cause infection in brain tissues [21]. The experiments on mice have shown that the SARS-CoV2 infections can cause neuronal distraction and death [22]. In COVID-19 also similar effects have been seen [23]. In light of the above, the therapeutic agents with good CNS penetration ability could have additional advantage [24].

Only a few studies have reported targeting more than one viral protein with a single molecule or using combination therapy [25–29]. In this study, we attempted to identify molecules that can simultaneously bind to multiple proteins of the SARS-CoV2. The strategy to target multiple proteins originates from the fact that individual viral proteins play specific role in multiple aspects of vir al lifecycle such as attachment, entry, replication, morphogenesis and egress. Single molecules that can potentially target many viral proteins can perturb viral lifecycle at multiple points and thereby can be highly efficient in curbing SARS-CoV2 infection. Such strategy will also have a higher barrier towards emergence of resistant mutants.

In this work, we have used the 3D-structures of the SARS-CoV2 proteins to identify FDA approved drugs that can bind to these proteins using computational methods. The FDA approved drugs were chosen so that they can be quickly repurposed for treating COVID-19. Additionally, we also analyzed if the identified molecules can affect the host proteins that get differentially expressed as a result of SARS-CoV2 infection. We have also tested these molecules using an in-vitro SARS-CoV2 infection model in Vero E6 cells. These molecules can be used as modulators of both the SARS-CoV2 and human proteins.

2. Methods

2.1. Protein structure modelling

The SARS-CoV2 proteins for which there is no crystal structure reported were modelled using Modeller v9.22 [30] (homology modeling) (Table 1). The modelling template for each protein was identified by performing Delta-BLAST against the PDB database. Proteins were modelled using either single or multiple templates based on the query coverage. Further, the model stereochemistry and other structural parameters were assessed using standalone PROCHECK [31] tool.

2.2. Molecular docking of FDA approved drugs in SARS-CoV2 proteins

The ensemble docking approach increases the efficiency by allowing virtual screening against multiple conformations [32]. Therefore, the selected protein structures were subjected to 20 ns MD run (total 180 ns) using NAMD 2.6 [33] to explore the flexibility of the binding site (Supplementary methods Sec 1.1). The health of the molecular dynamics simulation was evaluated for stability using root mean square deviation (RMSD) (Supplementary Fig. 2), and radius of gyration (ROG) (Supplementary Fig. 3A-I). Thereafter, five snapshots were generated at equal time (4 ns) points during MD simulation for each protein. The structures of the FDA approved drugs were obtained from the e-drugs (https://chemoinfo.ipmc.cnrs.fr/TMP/tmp.13454/e-Drug3D_1993.sdf) repository containing 1993 molecules in the current library (updated till July 2020). The molecules were prepared by Schrodinger LigPrep wizard ligands using the default parameters. The protein structures were prepared by addition of missing atoms, hydrogens, assignment of bond orders and proper protonation states. The structure of each of the protein was minimized by keeping heavy atoms fixed and then the whole structure was minimized until a RMS gradient of 0.3 kcal/molÅ as implied in Schrodinger.

The active site of the modelled proteins were identified using either of the following methods 1) the ligand binding pocket, if the co-crystal structure is available or 2) the ligand bound co-crystal structure of a close homolog or 3) the active site was predicted using SiteMap algorithm in Schrodinger v9.3 molecular modelling software [34]. The proteins with active site pocket volume of <150 Å³ were removed as smaller pockets may not be amenable to docking. The pockets were further selected by sequence comparisons and available literature. Finally, 9 proteins were selected for docking. The molecular docking was performed using the Glide module of Schrodinger molecular modelling software (www.schrodinger.com/glide). The docking was performed...
using default settings except that the formation of intramolecular hydrogen bonds was rewarded and the enhancement of planarity of conjugated π groups was penalized and the pose may get removed from final results. A maximum of 10 poses were generated for each of the molecules. The final ranking of the molecules was obtained by calculating the average glide score in the five snapshots of a viral protein generated by molecular dynamics simulation to include the effect of binding site dynamics. The molecules showing a docking score of −8.5 [11] or better were selected for further analysis.

2.3. Binding free energy calculation (MM-GBSA):

The obtained hits were subjected to MM-GBSA analysis as implied in Glide module of Schrodinger modeling software for further selection of better hits. The receptor residues within 5 Å of the ligands were considered flexible for the MM-GBSA procedure with other default settings. Since the MM-GBSA binding energies reflect approximate free energies of binding, a more negative value indicates stronger binding. Similar to average glide score, average MMGBSA score was also calculated for each of the ligand for each viral protein.

2.4. The differential gene expression (DEGs) and protein–protein interaction network analysis

The differentially expressed genes were obtained from the data reported by Blanco-Melo et al. [35]. The identified DEGs were mapped for their interactions with other human proteins using HIPPIE v2.2 which contains 14,855 proteins and 411,430 interactions. The reported protein–protein interactions with a minimum score of 0.63 (medium confidence, 2nd quartile) [36] were used for creation of the network using Cytoscape v3.7.2. The largest interconnected component was extracted and degree for individual nodes was calculated to assess their importance in the network.

2.5. Interaction with human proteins

The drug-gene interaction database (DGIdb) that contains information about the drugs and their target genes was employed to identify the drugs that can modulate the differentially expressed genes in COVID-19. The drug gene interaction was obtained from the data reported by Blanco-Melo et al. [35]. The identified DEGs were mapped for their interactions with other human proteins using HIPPIE v2.2 which contains 14,855 proteins and 411,430 interactions. The reported protein–protein interactions with a minimum score of 0.63 (medium confidence, 2nd quartile) [36] were used for creation of the network using Cytoscape v3.7.2. The largest interconnected component was extracted and degree for individual nodes was calculated to assess their importance in the network.

2.6. Virus and cell culture

Vero E6 cells were cultivated from the American Type Culture Collection (ATCC CRL-1586) and maintained at 37 °C with 5% CO2 in Dulbecco’s modified Eagle’s medium (DMEM; Gibco), supplemented with 10% heat-inactivated fetal bovine serum (Gibco) and 1 × Pen-Strep solution (Gibco). SARS-CoV-2 virus (IND/ILS-01/2020, Genebank Accession MW559533) of clade 19A was isolated from the oropharyngeal swab sample of laboratory-confirmed COVID-19 individual and was propagated in Vero E6 cells. Viral titers were determined by TCID50 assays in Vero cells. All experiments using SARS-CoV2 were performed in the Biosafety level 3 containment facility at the Institute of Life Sciences, Bhubaneswar. The approval for performing the drug repurposing studies against SARS-CoV2 was obtained from the institutional biosafety committee (No: V-122-MISC/2007-08/01).

Drug Assay: The FDA-approved drug library obtained from Enzo Life Sciences (BML-2843-0100, V.1.0) was used to source the compounds. 10 mM stocks were prepared in DMSO for further testing. To evaluate the effect of compounds in vitro, Vero E6 cells were seeded in 96 well culture plates. 16 h after seeding the cells were infected with SARS-CoV2 virus at 0.1 MOI for 2 h at 37 °C. The virus inoculum was replaced with fresh 2% FBS media containing the compounds at various dilutions. After 24 h incubation the cells and supernatants were collected to quantify cell-associated (intracellular) and cell-free (extracellular) viral loads by quantitative real-time RT-PCR using Takara PrimeScriptTM One Step RT-PCR Kit (RR055A) with forward (5'-GTAAGATGCGATGTCGCCG-3') and reverse (5'-CAAGATTCATGAGAATTA-3') primers and probe (5'-FAM-CAGGCTGACCTCCTAGCAGAGATGC-BHQ-3') targeting the SARS-CoV2 RdRp gene. Standard curve for absolute quantification of viral genome copies was generated using log-fold dilutions of plasmid pLVX-EF1alpha-nCoV2019-nsp12-2xStrep p-IRES-Puro plasmid harbouring the SARS-CoV2 RdRp gene [37]. RT-PCR assay was performed on ABI 7500, Applied Biosystems PCR machine. Dose-response curve analysis (GraphPad Prism) was used to determine the half maximal inhibitory concentrations (IC50) of the compounds. Cytotoxicity of the compounds was evaluated in Vero E6 cells using the cell cell Vibryant MTT assay kit (Thermo Scientific).

Immunofluorescence Assay: Vero E6 cells were seeded in 96 well culture plates. 16 h after seeding the cells were infected with SARS-CoV2 virus at 0.1 MOI for 2 h at 37 °C. The virus inoculum was replaced with fresh 2% FBS media containing the compounds at their IC50 concentration. After 48 h incubation cells were washed with 1 × PBS and fixed with 4% paraformaldehyde. Subsequently the cells were permeabilized and blocked for 1 hr with PBS containing 0.1% TritonX-100 and 3% BSA and probed with primary antibody specific for SARS-CoV2 nucleocapsid (Abgenex, cat. No. 11-2003) overnight at 4 °C. After 3 washes with PBS, the cells were probed with anti-rabbit secondary antibody tagged with Alexa Fluor 568 (Invitrogen, Carlsbad, CA), for 1 h at RT. The cells were counter stained with DAPI for 10 min to stain the nucleus. Images
were captured under a 20x objective using an Olympus IX83 inverted fluorescence microscope. Images were quantified using ImageJ software (National Institutes of Health, Bethesda, Maryland, USA).

3. Results

As stated earlier nine viral proteins (Table 1) were selected for molecular docking. The computational analysis of ligands binding to various proteins is a powerful method to quickly identify potential molecules for further studies. These methods have been successfully used in various studies [38–40]. In the first stage, the molecules were docked into the SARS-CoV2 protein snapshots obtained by molecular dynamics using Glide module of Schrodinger in standard precision (SP) mode. The molecules were then ranked using average Glide score. The MM-GBSA score was also calculated for each of the ligand in individual snapshots. (B) Transcriptomics data from SARS-CoV2 infected and normal human samples identified significantly differentially expressed genes (\(|\log_{2}\text{FC}>1\), p-value < 0.01) as a result of infection. A protein–protein-interaction network was created using these genes. They were also analyzed for their involvement in biological pathways using Ingenuity Pathways Analysis (IPA). The drug gene interactions were also analysed and molecules were tested using SARS-CoV2 infection model in Vero E6 cells for antiviral activity.

3.1. Molecules docking to SARS-CoV2 structural proteins

The hallmark feature of coronaviruses is their transmembrane spike (S) glycoprotein as this protein is the reason for its name “Corona” in Latin meaning, “Crown”. SARS-CoV-2 uses its spike (S) protein to attach to host cells. The spike protein exists as homo-trimers. Each monomer is about 180 kDa and has two distinct subunits S1 and S2. While the receptor binding is mediated by S1 subunit with the help of receptor binding domain (RBD), the fusion between the viral envelope and the host cellular membranes is facilitated by the S2 subunits upon the cleavage of S1–S2 junction by host proteases [41]. The S1 subunit of spike protein in SARS-CoV2 has four distinct domains: NTD, CTD1, CTD2 and CTD3,
of these the “up” conformation of CTD1 is responsible for binding with ACE2 receptor [42]. The S protein, due to its important role in the first stage of infection, is an important target for development of therapeutics and vaccines. The co-crystal structure of the S-protein with small molecule ligand is not available, therefore we used the sitemap algorithm in Schrodinger to identify the active site on S-protein. The sitemap revealed a site that is very close to the receptor binding domain and trimerization interface lined by the residues Ser 46, Leu 48, Leu 303, Lys 304, Ser 305, Glu 309, Thr 732, Asn 758, Thr 827, Phe 833, Tyr 837, Arg 847, Lys 854, Asn 856, Val 860, Glu 949, Val 952, Asn 955, Glu 957, Asn 960, Val 963, and His 1058. Many of these residues are highly conserved among coronaviruses. The site is overlapping to the site suggested by Kalathiya et al. [43]. Recently, some of the SARS-CoV2 strains containing mutants of the S-protein (D614G) with high infectivity have been reported. This mutant does not change the structure of S-protein but increases its binding with human TMPRSS2 protein [44]. This mutation is away from the identified binding site.

Our molecular docking analysis suggest that capreomycin, posaconazole, melfloquine, nebivolol, angiotensin II, celecoxib and trimethoprim bind to spike protein with appreciable affinity (Supplementary Table 1) (Fig. 2). Other groups have also predicted the binding of posaconazole to spike protein which further substantiates our analysis [45]. Posaconazole is an antifungal drug used in the prevention of invasive fungal infections and is also shown to inhibit the entry of Chikungunya virus [46] and replication of Zika and Dengue viruses by binding to oxysterol-binding protein (sterol transporter) [47]. Mefloquine is an antimalarial drug used in chloroquine resistant malaria. Nebivolol is an antihypertensive molecule with a very good safety profile in subjects with obstructive respiratory comorbidities [48] and can be an important drug to consider in SARS like diseases. Capreomycin is a polypeptide (iso-ates our analysis [45]. Posaconazole is an antifungal agent used in the treatment of mul-

molecules binding to SARS-CoV2 enzymes:

2′-O-Methyl Transferase (Nsp16) of SARS-CoV2 belongs to the S-adenosylmethionine-dependent methyl transferase family and is activated upon binding to Nsp10. Nsp10 binds with a conserved four amino acid sequence ‘KDKE’ of Nsp16 in its catalytic pocket and activates its methyltransferase activity. Capping of viral mRNA at 5′-end is one of the viral strategy for protecting viral transcripts from host 5′ exoribonucleases and escaping the host innate immune response by mimicking as host mRNAs, thus Nsp16 is the potential target for antiviral therapeutics [54].

The crystal structure of SARS-CoV2 Nsp16 (PDB ID: 6W4H, co-crystallized with S-adenosyl methionine) was used in the current studies. The binding site was found to be lined by the residues Phe 70, Gly 71, Ala 72, Gly 73, Asp 99, Leu 100, Leu 111, Gly 113, Met 131, Tyr 132, Asp 133, Phe 149, Asp 114, Ala 116, Cys 115, and Val 118. Our study shows that methotrexate, viomycin, saralin, saquinavir, venetoclax, vidarabine, histrelin, triptolide and ribavirin binds to Nsp16 with high affinity (Supplementary Table 1). Methotrexate forms hydrogen bonds with Asn6841, Asp6928, Lys6968, Asp6897, Asn6899 and Asp6876 of Nsp16 (Fig. 4).

Methotrexate acts as an antimetabolite and thus used as an antineoplastic drug. It is also used in treatment of inflammatory diseases like rheumatoid arthritis. It decreases the de novo synthesis of purines and pyrimidines and forms dimers with thymidylate synthase (TS), hence also has anti-parasitic effect [55]. Methotrexate is also shown to effectively inhibit Zika and Dengue virus replication [56]. Zidovudine is used in HIV1 treatment [21], histrelin and triptolide are gonadotropin-releasing hormone analogs used in the treatment of central precocious puberty and endometriosis [57]. Lanreotide is a long-acting analog of somatostatin and is used for the management of acromegaly, a condition caused by excess secretion of growth hormone. Octreotide is also a somatostatin analog currently used for the treatment of watery diarrhea and flushes caused by certain carcinoid tumors. Vidarabine (ara-A) is a purine analog and an antiviral drug used for infections caused by herpes simplex and varicella zoster viruses.

Among all the proteins encoded by SARS-CoV2 genome, PLpro (papain-like protease) and 3CLpro (3C chymotrypsin-like protease) are two important viral proteases that cleave the two polyproteins (pp1a and pp1ab) into individual viral proteins (Nsp2-Nsp16). The two proteases are important for replication and controlling the host cell response and hence they are among the key targets for the development of therapeutics against SARS-CoV2. These proteases have cysteine in the active site that has also been targeted for the development of covalent inhibitors. There are many small molecules, peptides and peptidomimetics that have been developed against these proteases [58–60].

The 3CLpro is a cysteine protease having three domains: β-barrel Domain I (residues 8–101) and II (residues 102–184) and α-helix domain III (residues 201–306) similar in structure to chymotrypsin [61]. The functional protease is a dimer that cleaves polyprotein 1ab in 11 regions at its specific cleavage site (P1) of Leu-Gln (Ser, Ala, Gly). The sequences of SARS-CoV2 and SARS-CoV main protease are highly similar (96% identity) and so their 3D structures, barring some surface residues. However, enantiomically the inhibitors of SARS-CoV 3CLpro lopinavir and ritonavir that were also recommended for use against SARS-CoV2 have not shown expected results in the clinical trials for COVID-19 [62]. The binding site for 3CLpro was defined as residues falling within 5 Å of the co-crystallized ligand (PDB: 6W63). The residues Thr 25, His 41, Cys 44, Thr 45, Ser 46, Met 49, Asn 142, Gly 143, Ser 144, Cys 145, His 164, Met 165, Glu 166, Leu 167, Pro 168, Asp 178, Arg 188, Gln 189, and Gln 192 were used for defining the active site. Rupintrivir, alatrofloxacin, cangrelor, capreomycin, naldemedine, lopinavir and indinavir are among the drugs predicted to bind to 3CLpro (Supplementary Table 1) (Fig. 5). It is important to note that most of the molecules are making HB interactions with the oxygenation hole residues (Asn 142, Gly 143, Ser 144) of the 3CLpro.

Previous studies report α-ketoamides, lopinavir and ritonavir as inhibitor of 3CLpro [63,64]. Rupintrivir inhibits human rhinovirus (HRV) 3C protease and has shown broad-spectrum anti-HRV activity [65]. Others have also indicated it to be useful against SARS-CoV2 main protease [66]. Indinavir is shown to inhibit HIV protease by blocking its active site and leads to immature virus
particle formation, however high doses have been linked to lipodystrophy syndrome [67]. Naldemedine, is a μ-opioid receptor antagonist used for the treatment of opioid-induced constipation [68].

PLpro is a domain within nsp3 of pp1a/pp1ab with proteolytic activity. It cleaves three sites at 181–182, 818–819, and 2763–2764 at the N-terminus of PP1ab [69]. It is the least explored among coronavirus proteins and only a few inhibitors are known for this protein [70]. Our study predicts that galidesivir, pralatrexate, methotrexate, daunorubicin, ganciclovir, folic acid, montelukast and itraconazole are among the molecules binding to the protease PLpro (Supplementary Table 1). Galidesivir has broad-spectrum antiviral activity (in vitro) against many RNA viruses in nine different families, including the coronaviruses [71]. The binding of galidesivir with PLpro is shown in Fig. 6. This drug has been under clinical trials for COVID-19 (NCT03891420). Daunorubicin (DNR) is the anthracycline compound used in the Kaposi’s sarcoma

---

**Fig. 2.** Virtual screening of FDA drugs against Spike protein. (A) The surface view of spike protein where domains are represented in different colors. The ligand binding site is shown inside white dotted circle at RBD (receptor binding domain) of the spike protein. (B) Cartoon depiction of the RBD showing the secondary structure elements and binding of capreomycin. (C) Closeup of RBD-capreomycin (average docking score −9.10) interaction showing residues making hydrogen bonds interactions (red dotted lines) with capreomycin. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

---

**Fig. 3.** Nelarabine binding to nucleocapsid (N) protein of SARS-CoV2. (A) The surface is coloured by the charge on the amino acids. The red, white and blue surface area depict negative, neutral and positive surface respectively. Nelarabine binds in a predominantly positive area at the nucleocapsid homodimer interface. (B) Nelarabine docked in five MD snapshots of N-protein (average docking score −8.73). (C) Ligand-protein interaction showing nelarabine makes multiple hydrogen bond interactions with N-protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
and lymphoma treatment of HIV-1 infected patients [72]. Moreover its derivative N,N-dimethyl daunomycin (NDMD) is used as the inhibitor of Herpes simplex virus (HSV) [73]. Montelukast has been predicted by other groups as well to bind to main protease of SARS-CoV2 [74]. An interesting observation is the identification of folic acid as a high affinity ligand of PLpro.

Helicase enzyme (Nsp13) of SARS-CoV2 is motor protein essential for unwinding of both dsDNA and dsRNA and has metal binding (Zn2+) N-terminal and helicase domain (Hel). It is involved in formation of RTC of SARS-CoV2 along with RdRp, which is known to enhance its activity [75]. The SARS-CoV2 helicase has 99.8% sequence similarity with that of SARS-CoV. Since
it is one of the most conserved proteins in Nidoviruses and is essential for viral RNA synthesis, it is an attractive target for antiviral drug development. A recent review summarizes its importance as a drug target in COVID-19 [76]. In the current studies, the cryo electron microscope structure of helicase-RdRp (PDB: 6XEZ) was used. The residues within 5 Å of the ADP bound to helicase enzyme were defined as the active site. Our analysis shows that eratapenem, methotrexate, clofarabine, trimethoprim, ascorbic acid, cefxime, and pibrentasvir bind to the helicase with high affinity (Fig. 7). Clofarabine is a potent HIV-1 inhibitor [77], Pibrentasvir, is a HCV NS5A inhibitor effective against all HCV genotypes [78].

The most vital enzyme responsible for the replication/transcription of the viral genome is the RNA-dependent RNA polymerase (RdRp) also known as Nsp12. The primer for RdRp RNA synthesis is synthesized by Nsp8 [79]. Nsp12 has two main functional domains namely nidovirus RdRp associated nucleotidyl transferase (NIRAN) domain and RNA dependent RNA polymerase (RdRp) domain. The NIRAN domain helps in nucleotide transfer while RdRp domain is involved in the polymerisation. The RdRp is conserved in structure and function among RNA viruses [80]. This enzyme, due to its importance in viral replication and also to the fact that humans are devoid of it, is a very attractive target [81,82]. Moreover due to the availability of its structure with cofactors Nsp7 and Nsp8 (PDB: 6M71) and remdesivir (PDB: 7BV2) the structure based design is feasible. A number of studies have been done on development of RdRp inhibitors and some molecules e.g. remdesivir, favipiravir etc. have been approved for emergency use in COVID-19.

In the current studies, we have used the structure of RdRp complexed with remdesivir (PDB: 7BV2). The residues falling withing 5 Å of the remdesivir were defined as active site. Our analysis shows that fludarabine, cobicistat, capreomycin, regadenoson, doxazocin, pibrentasvir, elbasvir, indinavir and remdesivir among others that can bind with RdRp (Fig. 8).

Fludarabine is used for the treatment of hematological malignancies. It inhibits various critical enzymes and results in the inhibition of DNA synthesis. It has been predicted to be active against SARS-CoV2 RdRp by other groups as well [74,83]. Ribavirin is broad spectrum antiviral used for treatment of RSV infection, hepatitis C and viral hemorrhagic fevers [84]. It is a well known RdRp inhibitor. Cobicistat is known to inhibit the cytochrome-mediated metabolism of HIV protease and was approved in 2012 by FDA as pharmacoenhancer for HIV treatment [85]. Other groups have also predicted that cobicistat and capreomycin can inhibit SARS-CoV2 protease [86,87]. Pibrentasvir and elbasvir are HCV NS5A inhibitors and indinavir is potent HIV protease inhibitor [88]. Another molecule monteleukast, a leukotrine inhibitor used as antihistamine was also showing good affinity towards RdRp (docking score ~9.42). The molecules we identified to bind to RdRp can serve as potential alternatives to remdesivir.

The Nsp15 is EndoRNase with endoribonuclease activity. It cleaves the 5' and 3' of uridylate residues in RNA by forming 2'-3'cyclic phosphodiester. Its mechanism is similar to that of RNase A, RNase T1 and XendoU [89]. Its XendoU activity can interfere with the host’s innate immune response and masks the exposure of viral dsRNA to host dsRNA sensors [90]. The crystal structure of SARS-CoV2 Nsp15 cocrystallized with USP5 (PDB: 6WLC) was used in the current studies. The active site was defined by the residues falling with 5 Å of the co-crystallized ligand. The active site is situated near the N-terminal and is surrounded by beta sheets and a helix. In our analysis, drugs such as quinapril, octreotide, folic acid, and macimorelin were found to bind to Nsp15 with acceptable affinity (Fig. 9). Quinapril is an angiotensin converting enzyme (ACE) inhibitor and the ACE inhibitors have been suggested to be beneficial for COVID-19 patients [91]. Folic acid is essential for DNA and protein synthesis and in the adaptive immune response [92]. The dose dependent effect of folic acid on rotavirus infected mice has been reported indicating its antiviral activity [93]. Additionally, the role of folic acid in the prevention of cellular entry...
Fig. 7. (A) Binding of eratapenem (average docking score – 9.86) at the helicase protein of SARS-CoV2. (B) The five receptor frames are shown with docked eratapenem. (C) The ligand plot is showing interactions between eratapenem and helicase binding site.

Fig. 8. Binding of fludarabine with RNA-dependent RNA polymerase (RdRp) of SARS-CoV2. (A) The surface view of RdRp protein with co-crystallized ligand (remdesivir). The red white and blue surface area depict negative, neutral and positive charged surface respectively. (B) Fludarabine docked into MD snapshots of RdRp (average docking score – 11.04). (C) Fludarabine (yellow sticks) and remdesivir (blue sticks) in RdRp active site. The hydrogen bonding interactions are shown in dotted lines (blue-remdesivir, red-fludarabine). It is clear from the binding pose and interactions that fludarabine binds to RdRp quite similar as remdesivir. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
of SARS-CoV2 has been reported [94]. Macimorelin is used for the diagnosis of adult growth hormone deficiency [95]. Interestingly other groups have also predicted it to be active against SARS-CoV2 [96,97].

Nsp14 is the 3′-5′ exonuclease that plays a role in proofreading mechanism [98]. Nsp14 contains four conserved DE-D-D acidic and a zinc-finger (ZnF) domain [99]. The homology model of Nsp14 based on the crystal structure of closely related Nsp14 of SARS CoV (PDB: 5C8T_chainB, 95.07% identity) was used for the current studies. The binding site was defined by comparison with the cocrystallized ligand (PDB: 5C8T, chainB). The SARS-CoV2 Nsp14 active site was found to be lined by the residues Arg289, Trp292, Asn306, Arg310, Asp314, Ala363, Leu366, Asn386, Asn388, Phe401, Tyr419, Asn422, Phe426, His455, Arg476, Tyr491, and Phe506.

Our molecular docking predicted that cangrelor, venetoclax, pimozone, nilotinib, droperidol, nebivolol, indacaterol, ezetimibe, simprevir, siponimod, lapatinib, elagolix bind to Nsp14 (Supplementary Fig. 1).

Pimozone, a calmodulin inhibitor is shown to inhibit Chikungunya virus secretion [100]. Moreover, it binds to the envelope protein of HCV and inhibits infection with many HCV genotypes [101]. Droperidol is also predicted by other groups to be effective against SARS-CoV2 infection [102]. Ezetimibe is shown to inhibit formation of capsid-associated relaxed circular DNA of Hepatitis B Virus (HBV) [103] and is also shown to inhibit Dengue infection by interfering in formation of replication complex [104]. Indacaterol is the β2-adrenoceptor agonist and used in the treatment of chronic obstructive pulmonary disease (COPD) since it induces bronchodilation [105]. It is a promising candidate for therapeutics against SARS-CoV2 due to its ability to regulate genes involved in suppressing proinflammatory cytokine production and attenuation of airway hyper-responsiveness [106]. However, dose and treatment schedule needs to be evaluated due to its counter effect on the expression of RNase L which is vital for antiviral response.

Since one of our major objectives was modeling of the intrinsic flexibility of the SARS-CoV2 proteins by molecular dynamics simulation and finding drugs that can adjust with the site flexibility. We provide a summary of the top drugs for individual proteins and their docking scores in the frames generated by molecular dynamics along with the average MMGBSA score Table 2. The drugs with consistently good docking scores will have a better average. This approach is novel and is not reported anywhere before for screening of drugs against SARS-CoV2 as per the best of our knowledge.

3.3. Drugs targeting multiple SARS-CoV2 proteins

A heatmap (Fig. 10) was generated using the docking scores to summarize the binding of important drugs to multiple proteins. Individually or in combinations these drugs can serve as potential therapeutics with the capacity to modulate both the viral as well as human proteins. The identification of molecules targeting multiple viral proteins simultaneously will effect the virus life cycle at multiple stages and will also have a higher barrier towards evolution/ emergence of drug resistant mutants, a common problem with many direct acting antivirals (DAA) against RNA viruses. Using a combination of drugs that target different viral proteins we can achieve synergistic antiviral effect. The detailed list of drugs and their docking scores is given in Supplementary Table 1.

3.4. Analysis of the transcriptome data and drug interactions of few differentially expressed genes

As stated earlier, the differentially expressed genes (DEGs) were obtained from the data reported by Blanco-Melo et al [35]. The DEGs were selected based on the following criteria: $|\log_{2}\text{FC}| \geq 1$, P-value $\leq 0.01$ (Supplementary Table 2). This criteria was chosen to select genes showing the most significant variation. The gene ontology (GO) enrichment analysis on these DEGs indicates immune system processes, such as type II interferon signaling (IFNG), innate defence response, cytokine and chemokine signaling, RAGE receptor binding, secretory granule are among the most enriched ontologies (Supplementary Table 3). It is important to note that these are among the processes usually activated in infection-associated inflammation. Protein-protein-interaction network analysis was done using Cytoscape to identify the
Table 2
Docking and MMGBSA scores of top drugs targeting different SARS-CoV2 proteins.

<table>
<thead>
<tr>
<th>Drugs/Frames</th>
<th>F1(^1)</th>
<th>F2(^3)</th>
<th>F3(^3)</th>
<th>F4(^3)</th>
<th>F5(^3)</th>
<th>Avg_score(^*)</th>
<th>MMGBSA_Avg(^*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RdRp</td>
<td>−11.32</td>
<td>−10.85</td>
<td>−9.83</td>
<td>−11.34</td>
<td>−11.86</td>
<td>−11.04</td>
<td>−126.47</td>
</tr>
<tr>
<td>Ribavirin</td>
<td>−10.83</td>
<td>−9.63</td>
<td>−10.24</td>
<td>−10.06</td>
<td>−11.38</td>
<td>−10.43</td>
<td>−110.14</td>
</tr>
<tr>
<td>Ivermectin</td>
<td>−10.94</td>
<td>−10.28</td>
<td>−9.08</td>
<td>−9.28</td>
<td>−10.04</td>
<td>−10.30</td>
<td>−99.25</td>
</tr>
<tr>
<td>Remdesivir</td>
<td>−10.53</td>
<td>−9.69</td>
<td>−8.49</td>
<td>−10.02</td>
<td>−10.21</td>
<td>−9.79</td>
<td>−107.23</td>
</tr>
<tr>
<td>Cangrelor</td>
<td>−9.39</td>
<td>−8.36</td>
<td>−9.03</td>
<td>−9.42</td>
<td>−10.36</td>
<td>−9.31</td>
<td>−118.86</td>
</tr>
<tr>
<td>Nevirapine</td>
<td>−9.56</td>
<td>−9.89</td>
<td>−8.69</td>
<td>−7.98</td>
<td>−10.13</td>
<td>−9.25</td>
<td>−88.87</td>
</tr>
<tr>
<td>Spike</td>
<td>−9.61</td>
<td>−8.82</td>
<td>−9.16</td>
<td>−8.24</td>
<td>−9.68</td>
<td>−9.10</td>
<td>−168.25</td>
</tr>
<tr>
<td>Hydrocortisone</td>
<td>−8.56</td>
<td>−8.69</td>
<td>−7.89</td>
<td>−9.02</td>
<td>−9.10</td>
<td>−8.65</td>
<td>−105.84</td>
</tr>
<tr>
<td>Hydroxychloroquine</td>
<td>−8.56</td>
<td>−8.59</td>
<td>−8.47</td>
<td>−7.89</td>
<td>−9.26</td>
<td>−8.55</td>
<td>−135.63</td>
</tr>
<tr>
<td>Ivermectin</td>
<td>−10.77</td>
<td>−8.68</td>
<td>−7.28</td>
<td>−8.02</td>
<td>−7.89</td>
<td>−8.53</td>
<td>−96.82</td>
</tr>
<tr>
<td>Angiotensin II</td>
<td>−9.23</td>
<td>−7.58</td>
<td>−8.08</td>
<td>−7.89</td>
<td>−8.95</td>
<td>−8.35</td>
<td>−83.21</td>
</tr>
<tr>
<td>Lopinavir</td>
<td>−11.79</td>
<td>−10.58</td>
<td>−11.06</td>
<td>−11.03</td>
<td>−11.28</td>
<td>−11.15</td>
<td>−147.54</td>
</tr>
<tr>
<td>Prapavenoxal</td>
<td>−9.90</td>
<td>−10.03</td>
<td>−10.07</td>
<td>−9.58</td>
<td>−10.01</td>
<td>−9.92</td>
<td>−165.58</td>
</tr>
<tr>
<td>Pimozide</td>
<td>−9.68</td>
<td>−9.50</td>
<td>−9.89</td>
<td>−10.05</td>
<td>−10.21</td>
<td>−9.87</td>
<td>−90.28</td>
</tr>
<tr>
<td>Nevirapine</td>
<td>−9.23</td>
<td>−9.45</td>
<td>−10.25</td>
<td>−10.84</td>
<td>−9.58</td>
<td>−9.87</td>
<td>−115.42</td>
</tr>
<tr>
<td>Droperidol</td>
<td>−9.34</td>
<td>−9.26</td>
<td>−10.21</td>
<td>−9.87</td>
<td>−8.51</td>
<td>−9.44</td>
<td>−100.52</td>
</tr>
<tr>
<td>Octreotide</td>
<td>−8.52</td>
<td>−7.85</td>
<td>−9.28</td>
<td>−8.97</td>
<td>−10.03</td>
<td>−8.93</td>
<td>−108.20</td>
</tr>
<tr>
<td>Quinapril</td>
<td>−9.40</td>
<td>−8.12</td>
<td>−8.58</td>
<td>−10.87</td>
<td>−11.03</td>
<td>−9.60</td>
<td>−86.25</td>
</tr>
<tr>
<td>Celecoxib</td>
<td>−9.63</td>
<td>−8.69</td>
<td>−9.02</td>
<td>−9.36</td>
<td>−10.02</td>
<td>−9.34</td>
<td>−108.02</td>
</tr>
<tr>
<td>Ribavirin</td>
<td>−9.47</td>
<td>−9.57</td>
<td>−8.24</td>
<td>−10.24</td>
<td>−9.02</td>
<td>−9.31</td>
<td>−98.57</td>
</tr>
<tr>
<td>Folic Acid</td>
<td>−8.06</td>
<td>−9.28</td>
<td>−8.26</td>
<td>−10.03</td>
<td>−10.58</td>
<td>−9.24</td>
<td>111.58</td>
</tr>
<tr>
<td>Ascorbic Acid</td>
<td>−8.56</td>
<td>−8.77</td>
<td>−8.69</td>
<td>−8.96</td>
<td>−9.09</td>
<td>−8.81</td>
<td>−88.26</td>
</tr>
<tr>
<td>Helicase</td>
<td>−10.34</td>
<td>−9.51</td>
<td>−9.67</td>
<td>−10.21</td>
<td>−9.58</td>
<td>−9.86</td>
<td>−120.58</td>
</tr>
<tr>
<td>Eratapenem</td>
<td>−9.99</td>
<td>−8.28</td>
<td>−10.69</td>
<td>−9.14</td>
<td>−10.78</td>
<td>−9.78</td>
<td>−111.21</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>−9.07</td>
<td>−8.82</td>
<td>−9.26</td>
<td>−8.87</td>
<td>−9.05</td>
<td>−8.99</td>
<td>−98.82</td>
</tr>
<tr>
<td>Ivermectin</td>
<td>−8.33</td>
<td>−8.55</td>
<td>−8.69</td>
<td>−8.78</td>
<td>−9.87</td>
<td>−8.84</td>
<td>−99.58</td>
</tr>
<tr>
<td>Clofarabine</td>
<td>−8.56</td>
<td>−8.77</td>
<td>−8.69</td>
<td>−8.96</td>
<td>−9.09</td>
<td>−8.81</td>
<td>−88.26</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>−12.36</td>
<td>−11.39</td>
<td>−12.06</td>
<td>−11.03</td>
<td>−12.68</td>
<td>−11.90</td>
<td>−126.95</td>
</tr>
<tr>
<td>Prapavenoxal</td>
<td>−11.10</td>
<td>−11.25</td>
<td>−11.02</td>
<td>−12.02</td>
<td>−12.24</td>
<td>−11.76</td>
<td>−105.13</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>−10.84</td>
<td>−10.76</td>
<td>−10.76</td>
<td>−10.76</td>
<td>−10.76</td>
<td>−10.78</td>
<td>−83.26</td>
</tr>
<tr>
<td>Remdesivir</td>
<td>−9.42</td>
<td>−9.04</td>
<td>−8.98</td>
<td>−10.06</td>
<td>−11.06</td>
<td>−9.71</td>
<td>−82.06</td>
</tr>
<tr>
<td>Daunonubicin</td>
<td>−8.95</td>
<td>−8.78</td>
<td>−9.06</td>
<td>−10.25</td>
<td>−10.17</td>
<td>−9.44</td>
<td>−69.20</td>
</tr>
<tr>
<td>Itraconazole</td>
<td>−8.57</td>
<td>−8.98</td>
<td>−9.26</td>
<td>−8.98</td>
<td>−9.58</td>
<td>−9.07</td>
<td>−90.58</td>
</tr>
<tr>
<td>Methotrexate</td>
<td>−8.56</td>
<td>−8.56</td>
<td>−8.56</td>
<td>−8.26</td>
<td>−8.56</td>
<td>−7.74</td>
<td>−65.89</td>
</tr>
<tr>
<td>Vinudarabine</td>
<td>−8.36</td>
<td>−8.55</td>
<td>−8.45</td>
<td>−8.60</td>
<td>−9.02</td>
<td>−8.60</td>
<td>−92.05</td>
</tr>
<tr>
<td>Saquinavir</td>
<td>−9.78</td>
<td>−9.68</td>
<td>−10.06</td>
<td>−8.89</td>
<td>−9.08</td>
<td>−9.49</td>
<td>−94.28</td>
</tr>
<tr>
<td>Capreomycin</td>
<td>−7.85</td>
<td>−7.89</td>
<td>−8.26</td>
<td>−8.65</td>
<td>−8.79</td>
<td>−8.29</td>
<td>−100.60</td>
</tr>
<tr>
<td>Viomycin</td>
<td>−8.29</td>
<td>−8.88</td>
<td>−8.76</td>
<td>−8.23</td>
<td>−7.26</td>
<td>−8.28</td>
<td>−95.28</td>
</tr>
<tr>
<td>Nucleocapsid</td>
<td>−8.41</td>
<td>−8.96</td>
<td>−8.10</td>
<td>−7.58</td>
<td>−8.74</td>
<td>−8.36</td>
<td>−84.85</td>
</tr>
</tbody>
</table>

\(^1\) Docking score in individual snapshots generated from molecular dynamics (F1-F5)  
\(^3\) Average docking score.  
\(^*\) Average MMGBSA score.
network hubs based on their interactions with other proteins. The giant component was extracted from the network with 1823 nodes containing 2546 interactions. Top 5% of the proteins (total 91) were selected based on the connections they make to other nodes in the network for further analysis (Supplementary Table 4).

It is worth to mention that among the human proteins many of them (e.g. ARRB2, JUN, CDC73, SUMO, TNF, IL2RG, MCM7, IFIT1, FOS and ISG15 etc.), with prominent role in inflammation and immune response, are hubs i.e. very important proteins in the generated protein–protein-interaction network. A search for the selected DEGs (292) at drug-gene interaction database (DGIdb) resulted in the identification of 658 unique drugs for 97 proteins (Supplementary Table 5). An intersection of these with drugs binding with viral proteins (docking score \( \leq -8.5 \)) resulted in identification of 74 drugs that can target at least one viral protein whereas there are 31 drugs that can target atleast two viral proteins and one or more human proteins differentially expressed as a result of SARS-CoV2 infection. Recently, Li et al. [107] have reported a set of drugs for repurposing in COVID-19 using analysis of transcriptomic data of human tissue samples. The study identified a list of monoclonal antibodies along with FDA approved drugs. There are many drugs such as methotrexate, indinavir, chloroquine, sequinavir, and ritonavir, that are common between our study and the said study further corroborating our findings. It is interesting to find many drugs with multi-targeting ability against hub proteins as well as SARS-CoV2 proteins. Such drugs can have a significant therapeutic utility for COVID-19 (Fig. 11).

3.5. In-vitro validation of identified high affinity binders

We assessed the activity of few of the compounds that bind to more than one SARS-CoV2 proteins in an in-vitro cell culture model of SARS-CoV2 in VeroE6 cells. We used the SARS-CoV2 virus (ILS-01) isolated from oropharyngeal swab sample of confirmed COVID19 positive patient.

Seven different concentrations of the drugs ranging from 0.01 \( \mu \)M to 10 \( \mu \)M were used to determine their IC\(_{50}\) values. Almost all of the drugs tested showed anti-SARS-CoV2 activity further validating our in-silico analysis. All of the compounds showed more than 50% reduction in viral loads at minimal or non-toxic concentrations. Ivermectin, and hydroxychloroquine, two of the compounds known to inhibit SARS-CoV2 replication in-vitro, were used as positive controls in this study that further validated the activity and potency of the molecules tested in this study. Among the 5 drugs tested mefloquine has the lowest IC\(_{50}\) value at 0.37 \( \mu \)M and montelukast, which was also predicted by others to possess anti-SARS-CoV2 potential has a higher IC\(_{50}\) value of 18.82 \( \mu \)M (Fig. 12, Table 3).

To further validate the anti-SARS-CoV2 potency of the tested drugs we performed an immunofluorescence assay by staining for SARS-CoV2 nucleocapsid protein in SARS-CoV2 infected VeroE6 cells subjected to treatment with the vehicle control or drugs at their IC50 concentrations.

We observed marked reduction (around 50% or higher) in the percentage of infected cells subjected to treatment with the drugs in comparison to the vehicle control (Fig. 13A and B). We also observed that the drug treatment at the indicated IC50 concentrations had very minimal effect on the total cell number in comparison to the vehicle control (Fig. 13C). Overall these observations suggest that the treatment with drugs resulted in marked inhibition of viral gene expression with minimal effect on cellular viability.

Overall these observations strongly validate our in-silico finding, however further screening is required in physiologically relevant cell lines and in-vivo animal models to fully establish the anti-SARS-CoV2 potential of the identified leads.

4. Discussion

The computational drug repurposing studies came into forefront because of the speed and memory of the modern computers supplemented by the availability of the algorithms and data from studies in the past. Moreover, the crystal structures of many SARS-CoV2 proteins were reported during 2020 that made the structure based screening studies feasible. Various libraries (e.g. FDA approved) were screened against targets like RdRp, main protease (3Clpro), spike protein, membrane protein, and non-structural proteins (Nsps) using various strategies [74,108–112]. For example, the main protease (Mpro) was screened by three docking algorithms. The authors selected the molecules that are commonly predicted by all algorithms [108]. Similarly, structure based docking followed by molecular dynamics studies have been used by Al-Khafaji et al. [113], Mittal et al. [114] and Wang et al. [111]. Wu et al. [74] screened ZINC database molecules against a number of SARS-CoV2 and host proteins. A set of antiviral drugs
was screened for their possible activity against 5 SARS-CoV2 proteins using deep learning [115]. Zhou et al [116] identified set of molecules against SARS-CoV2, using a network proximity analysis combining information from HCoV-host interactions, and human protein interactome.

The strategy of combining multiple approaches (gene expression, graph based algorithms, docking and molecular dynamics etc.) and targeting both the SARS-CoV2 and human proteins is the novelty of this study. The overall goal was to identify molecules that can bind with multiple SARS-CoV2 proteins that play vital role(s) in various stages of the viral lifecycle as well as target the host factors that drive viral persistence and disease pathogenesis. Our analysis predicted drugs that can bind to viral proteins (both the structural and non-structural proteins) with high affinity and can effectively inhibit viral entry as well as the post entry events like viral genome replication and transcription.

Capreomycin is a promising candidate with potential to inhibit SARS-CoV2 at mutltiple stages of viral lifecycle, as it can bind with high affinity to spike protein and the viral proteases and methyl transferase, which play a crucial role in viral entry, replication and transcription. In our in-vitro asays it has shown good inhibitory activity.

Mefloquine, an antimalarial drug, has shown good affinity towards spike protein in our in-silico studies. It is pertinent to note that recently mefloquine has been shown to prevent the entry of SARS-CoV-2 into host cells. It has shown potent inhibitory activity against SARS-CoV2 in multiple cell lines [117]. Many other studies reported similar observations corroborating our findings [109,112,118].

Our analysis further indicated that some drugs that bind viral proteins also target some of host proteins that are differentially expressed in lung tissue during SARS-CoV2 infection. The predicted drug candidates that interact with the viral protein(s), in parallel can also specifically target the host signalling pathways vital to control viral infection or disease manifestation. For instance, nebivolol a β-adrenoreceptor blocker, which stimulates nitric oxide production by endothelial nitric oxide synthase [48] is found to bind to PLP and exonuclease of SARS-CoV2. Nitric oxide is used to reverse pulmonary hypertension and shown to improve severe hypoxia in SARS-CoV1 [119] and SARS-CoV2 patients. Hence, nebivolol can be a promising therapeutic strategy with dual benefit; i) to curb SARS-CoV2 infection and ii) reversal of severe hypoxia manifestation in critical Covid-19 patients via its direct effect on nitric oxide synthase. Corroborating our findings, another recent study also reported that nebivolol can efficiently inhibit SARS-CoV2 in submicromolar range [120].

A major contributor of COVID-19 pathology is hyper-inflammation and cytokine hyperactivity. Strategies to reduce the inflammation and cytokine hyperactivity have shown promising results. Celecoxib, a selective cyclooxygenase-2 (COX-2) inhibitor, which lowers the effect of proinflammatory cytokines IL-6 and IL-1β [121], can also target Spike and RdRp protein of SARS-CoV2. Thus administrating celecoxib to COVID-19 patients can have dual benefit of reducing systemic inflammation as well as inhibition of viral replication. It has been shown that COX inhibitors can inhibit viral replication and production of virus particles in other coronaviruses [122]. A recent review [123] suggested selectively targeting COX2 and closely related cascades could be significant in the treatment of COVID-19. The authors were of the opinion that celecoxib has potential and should be evaluated clinically.

The leukotriene inhibitor montelukast is shown to reduce proinflammatory cytokines e.g. TNF-α, IL-6 and IL-1β levels [124,125]. A previous study suggests that it inhibits Zika virus by disrupting the integrity of the virions [126]. Durdagi et al. [127] using a multiscale modeling approach and in-vitro evaluation identified it to interfere with viral entry through Spike-ACE2 interface and by inhibiting the main protease. Apart from the anti-asthmatic effect it is also reported to cross BBB and reduce

![Fig. 11. Human host and SARS-CoV2 protein drug interactions. Drugs capable of binding to multiple viral as well as human proteins related with SARS-CoV2 infection. The pink circles show the name of the SARS-CoV2 proteins, the green squares depict the drugs and the triangles show the name of the human proteins. The lines show the interaction between drugs and target protein. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
Fig. 12. *In-vitro* validation of anti-SARS-CoV2 activity: Vero E6 cells were infected with SARS-CoV2 at 0.1 MOI for 2 h. Subsequently, the virus inoculum is replaced with fresh media containing the 0.1% of vehicle (DMSO) or indicated concentration (0.01, 0.1, 0.5, 1, 2, 5, 10 μM) of various drugs. 24 h post treatment the viral load in the cells (cell associated) or supernatant (secreted virus) was determined using real-time PCR as described in methods. IC50 values were determined from the dose response curve analysis (GraphPad prism). Vybrant MTT cell viability assay was used to determine cytotoxicity at low–high concentrations (0.01, 0.1, 0.5, 1, 5, 10 & 20 μM). The percentage cell viability was calculated with respect to vehicle treated control and 50% cytotoxic concentration (CC50) was determined by dose–response curve (GraphPad prism). Experiments were done in duplicates.
neuroinflammation [128]. A recent paper reviewed its antiviral, anti-inflammatory, anti-allergic and anti-fibrotic activities. It has also been suggested that montelukast should be tried as therapeutic option in COVID-19 [129].

Interestingly, the drugs montelukast, celecoxib and nebivolol can cross blood brain barrier [128,130,131], which gives additional advantage to counter neurological manifestations of COVID-19. It is again pertinent to note that these drugs have shown appreciable inhibitory activity (IC50) against SARS-CoV2 in our study.

Lapatinib binds to Nsp14, a viral protein crucial for viral RNA synthesis [132]. Lapatinib is a HER2 inhibitor, which can also trigger TBK1 activation that plays a crucial role in anti-viral signalling. Computational studies have predicted lapatinib to be able to bind many SARS-CoV2 proteins including the main protease [110,133]. Thereby, lapatinib has the dual advantage of inhibiting SARS-CoV2 replication as well as upregulating anti-viral signaling [134–136].

Saralasin belongs to a class of drugs called angiotensin receptor blockers (ARBs). It is worth mentioning some other ARBs (e.g. losartan) are also in clinical trials as therapeutics for COVID-19

Table 3
The activity (IC50) and cytotoxicity (CC50) of various drugs.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Drug</th>
<th>IC50 (µM)</th>
<th>CC50 (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Capreomycin</td>
<td>1</td>
<td>&gt;20</td>
</tr>
<tr>
<td>2</td>
<td>Celecoxib</td>
<td>3.25</td>
<td>&gt;20</td>
</tr>
<tr>
<td>3</td>
<td>Mefloquine</td>
<td>0.37</td>
<td>&gt;20</td>
</tr>
<tr>
<td>4</td>
<td>Montelukast</td>
<td>18.82</td>
<td>&gt;20</td>
</tr>
<tr>
<td>5</td>
<td>Nebivolol</td>
<td>5.21</td>
<td>&gt;20</td>
</tr>
<tr>
<td>6</td>
<td>Hydroxychloroquine</td>
<td>5</td>
<td>&gt;20</td>
</tr>
<tr>
<td>7</td>
<td>Ivermectin</td>
<td>5.54</td>
<td>14.96</td>
</tr>
</tbody>
</table>

Fig. 13. Immunofluorescence assay for validating the anti-SARS-CoV2 potency of the identified drugs. (A) Immunofluorescence images of SARS-CoV-2-infected Vero E6 cells stained with antibody against SARS-CoV2 nucleocapsid. (B) Graphical representation of the percent infected cells with respect to vehicle control. (C) Graphical representation of percent total number of cells with respect to vehicle control. Cells were counter stained with Dapi nuclear stain. The experiment was performed in triplicates and the data presented here is the mean ± SEM. Hydroxychloroquine is represented as HCQ.
The transcription complex activator protein 1 (AP1) is composed of homo/hetero dimers of Fos, Jun, CREB and other activated transcription factors (ATFs). The studies on the SARS-CoV1 infection in the Vero and Huh7 cell shows that nucleocapsid protein is the potent activator of (AP-1) [143]. Interestingly, asthmatic patients show higher expression of c-Fos in their epithelial cells. It is also observed that TNF-α induced ROS and intracellular glutathione depletion in the airway epithelial cells induces the production of AP-1 and leads to the pulmonary fibrosis [144,145]. Our analysis suggests that paclitaxel and bromocriptine, which dock with nucleocapsid and Nsp4 proteins can also effectively bind to c-Fos and thereby would be beneficial in inhibiting SARS-CoV2, as well as in alleviating lung injury observed in COVID-19 patients. Interestingly, other groups have also reported that bromocriptine is able to bind to main protease [146] or Nsp14 [147].

The transcriptome analysis revealed that S100/calgranulin is upregulated during SARS-CoV2 infection. This protein is also found in higher quantity in the Bronchoalveloar Lavage Fluid (BALF) and sputum of patients with inflamed lungs, COPD, and ARDS [148]. Calgranulin is polypeptide released by the activated inflammatory cells such as leukocytes, PBMC phagocytes and lymphocytes and is accumulated at the sites of chronic inflammation. It is the ligand for RAGE receptors and is the major initiator of cascading events that amplify inflammation [149]. Our analysis suggests that the anti-inflammatory agent methotrexate which has high affinity to the Nsp16 protein of SARS-CoV2 also shows appreciable binding to calgranulin and can thereby be useful to curtail systemic inflammation observed in lungs during COVID-19 in addition to its inhibitory effect on SARS-CoV2. Interestingly, methotrexate is recently reported to inhibit the replication of SARS-CoV2 [150]. Another paper suggested to use methotrexate with leucovorin rescue for the treatment of severe COVID-19 [151]. Methotrexate is also in clinical trial currently for the treatment of mild COVID-19 (https://clinicaltrials.gov/ct2/show/NCT04610567).

The expression of endogenous prolactin is also upregulated during SARS-CoV2, which leads to prolactin induced STAT5 activation and its pathways. Prolactin has a dual role in human physiology functioning as a hormone (secreted from anterior pituitary gland) and cytokine (secreted by immune cells). It causes anti-apoptotic effect and induces proliferation in immune cells in response to antigens leading to increased production of immunoglobulin, cytokines, and autoantibodies [152]. We envisage that prolactin may be one of the significant player in trigger of cytokine storm implicated in COVID-19. Interestingly, our study suggests that bromocriptine, which targets O'-methyl transferase (Nsp16) can also bind to prolactin and can be of high significance in management of COVID-19 due to dual ability to affect Nsp16 and prolactin.

The COVID-19 creates an inflammatory state involving proinflammatory cytokines e.g. IL-6, TNF-α etc. IL-6 stimulate ferritin and hepcidin synthesis [153]. The hyper-ferritinemia is associated with generation of ROS and RNS leading to enhanced systemic inflammation. As a result a devastating cycle is propagated where increased ferritin leads to higher inflammation (increased IL6) resulting in further increase in ferritin levels [154]. Hyperferritinemia has been linked with poor prognosis in COVID-19 patients, evidenced by high levels of ferritin in non-survivors as compared to survivors [155,156]. In this milieu, iron chelators can be extremely helpful by reducing the hyper-ferritinemia and systemic inflammation. Deferoxamine is an iron chelator that also increases degradation of ferritin by lysosomes leading to reduction of free radicals and subsequently inflammation. It also limits the chances of ARDS and tissue fibrosis. Our analysis indicates that deferoxamine binds to RdRp and PLpro protein of SARS-CoV2 with good affinity. Therefore it can be a good candidate for the therapeutics of COVID-19. Currently, deferaox is under clinical trials for the treatment of COVID-19.

We could test a few molecules that showed potent anti-SARS-CoV2 activity in in-vitro models. The identified molecules are commonly used drugs and hence can be quickly repurposed. Their combinations can also have synergistic effects against SARS-CoV2. We hope these molecules will prove to be useful in our fight against COVID-19. Apparently, the cytotoxicity of some molecules is high. As the in-vitro assays were performed in Vero-E6 cells, a monkey kidney epithelial cell line, we expect that the cytotoxicity may be lower in human cell lines. However, they are FDA approved drugs already being used for treating various clinical conditions at the recommended dosage and may be beneficial for SARS-CoV2 therapeutics. Also the treatment for SARS-CoV2 will likely last only for a short duration, therefore weighing the potential benefits vs toxicity they could be very useful in curbing SARS-CoV2 infection.

5. Conclusions

Currently, there are no approved coronavirus treatments and therefore there is a pressing need for drugs that can be effective therapeutics for COVID-19. Our study predicted promising drug candidates with high binding affinity towards many of SARS-CoV2 proteins. These drugs are expected to be more effective than drugs that target single viral proteins due to their ability to affect multiple aspects of viral lifecycle and enhance the barrier towards the evolution of drug-resistant mutants, a usual phenomenon observed in RNA viruses.

Overall our study predicts promising agents with potential to inhibit crucial viral processes, upregulate anti-viral host response and alleviate severe lung disease condition thereby providing attractive avenues for design of potential and multipronged thera peutic strategies against COVID 19.
Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.014.

References


S. Kumar, B. Singh, P. Kumari et al.


[154] Kell DB, Pretorius E. Serum ferritin is an important inflammatory disease marker, as it is mainly a leakage product from damaged cells. Metallomics 2014;6(4):748–73. PubMed PMID: 24549403.


Quantitative proteomics of hamster lung tissues infected with SARS-CoV-2 reveal host factors having implication in the disease pathogenesis and severity

Voddu Suresh1,2 | Varshasnata Mohanty1 | Kiran Avula1,2 | Arup Ghosh1,3 | Bharati Singh1,3 | Rajendra Kumar Reddy1 | Deepti Parida1,2 | Amol Ratnakar Suryawanshi1 | Sunil Kumar Raghav1 | Soma Chattopadhyay1 | Punit Prasad1 | Rajeeb Kumar Swain1 | Rupesh Dash1 | Ajay Parida1 | Gulam Hussain Syed1,5 | Shantibhusan Senapati1,5

1Institute of Life Sciences, Bhubaneswar, India
2Regional Centre for Biotechnology, Faridabad, India
3Kalinga Institute of Industrial Technology, Bhubaneswar, India

Abstract

Syrian golden hamsters (Mesocricetus auratus) infected by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) manifests lung pathology. In this study, efforts were made to check the infectivity of a local SARS-CoV-2 isolate in a self-limiting and non-lethal hamster model and evaluate the differential expression of lung proteins during acute infection and convalescence. The findings of this study confirm the infectivity of this isolate in vivo. Analysis of clinical parameters and tissue samples show the pathophysiological manifestation of SARS-CoV-2 infection similar to that reported earlier in COVID-19 patients and hamsters infected with other isolates. However, diffuse alveolar damage (DAD), a common histopathological feature of human COVID-19 was only occasionally noticed. The lung-associated pathological changes were very prominent on the 4th day post-infection (dpi), mostly resolved by 14 dpi. Here, we carried out the quantitative proteomic analysis of the lung tissues from SARS-CoV-2-infected hamsters on day 4 and day 14 post-infection. This resulted in the identification of 1585 proteins of which 68 proteins were significantly altered between both the infected groups. Pathway analysis revealed complement and coagulation cascade, platelet activation, ferroptosis, and focal adhesion as the top enriched pathways. In addition, we also identified altered expression of two pulmonary surfactant-associated proteins (Sftpd and Sftpb), known for their protective role in lung function. Together, these findings will aid in understanding the mechanism(s) involved in SARS-CoV-2 pathogenesis and progression of the disease.

Abbreviations: ABSL3, animal biosafety level 3; AERD, aspirin exacerbated respiratory disease; ARDS, acute respiratory distress syndrome; BAL, bronchoalveolar lavage; BCA, bicinchoninic acid assay; CF, cystic fibrosis; DAD, diffuse alveolar damage; DAPI, 4′, 6-diamidino-2-phenylindole; DTT, dithiothreitol; HPLC, high performance liquid chromatography; IAA, iodoacetamide; PANTHER, protein analysis through evolutionary relationships; PBST, phosphate-buffered saline with tween detergent; PCA, principal component analysis; TEABC, triethylammonium bicarbonate.

Voddu Suresh, Varshasnata Mohanty and Kiran Avula are contributed equally.
1 | INTRODUCTION

The recent outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has emerged as a global public health crisis affecting millions of people worldwide. SARS-CoV-2, the causative agent of the coronavirus disease 2019 (COVID-19), primarily infects the respiratory tract resulting in respiratory failure consistent with the acute respiratory distress syndrome (ARDS). The aggressive inflammatory response associated with infection leads to tissue damage and fatal lung injury. The death of the virus affected epithelial cells and endothelial cells, and activation of resident dendritic cells, monocytes, and macrophages result in the dysregulated inflammatory response. This triggers further activation and recruitment of immune cells leading to tissue damage and exacerbation of respiratory distress and disease severity. However, the levels of circulating cytokines in COVID-19 patients are found to be lower in comparison to patients experiencing ARDS due to other reasons and post-mortem investigation of COVID-19 patients revealed severe vascular injury including alveolar microthrombi. Virus-induced coagulopathy has also been implicated to trigger COVID-19 associated pneumonia and ARDS. Currently, our understanding of COVID-19 associated lung injury is limited and requires further studies to elucidate the intricate and specific mechanisms.

Severe clinical manifestations involving multi-organ failure have resulted in high morbidity and mortality, thus demanding immediate therapeutic measures and disease management. In this regard, a better understanding of viral and host factors involved in this disease’s pathogenesis has paramount significance. After identifying the SARS-CoV-2 virus, multiple genomic and proteomics-based approaches have been adapted to understand the host response to this viral infection. In the recent past, different studies have been conducted using proteomics technology to understand the virus-host protein interactome, changes in host protein expression upon virus infection, and for diagnosis of SARS-CoV-2 infection. Most of these studies have been carried out with patients’ liquid specimens like serum, plasma, sputum, and bronchoalveolar lavage. Clinical manifestations of COVID-19 are primarily related to the lung; hence clinical proteomics of infected lung tissues are essential to understand and combat this disease. The information obtained from such studies will provide the rationale for designing novel diagnostic and therapeutic interventions. Using lung tissues obtained from a small number of deceased COVID-19 patients, an earlier study has identified differentially expressed proteins in biological processes like blood coagulation, metabolism, immune response, angiogenesis and cell microenvironment regulation. Another group has extracted proteins from Formalin-Fixed Paraffin-Embedded (FFPE) lung tissues of COVID-19 deceased patients and identified high expression of proteins associated with SARS-CoV-2 entry (cathepsins B and L) and inflammatory response modulator (S100A8/A9). In a recent study, proteomic analysis of autopsy samples of nineteen COVID-19 patients showed the elevation of cathepsin L1, rather than ACE2 in infected lung tissues and highlighted the dysregulation of biological processes like angiogenesis, coagulation, and fibrosis in different organs including lungs. Most of these studies report the upregulation of proteins predominantly implicated in the hyperinflammatory state, repair state, and lung fibrosis. However, so far no experimental controlled lung proteomics study is available to characterize the molecular mechanisms underlying COVID-19 pathogenesis at different stages of the disease progression. Preclinical animal models that recapitulate multiple sequential events associated with human diseases are precious tools to understand the mechanistic aspects of human disease progression. In human patients, the COVID-19 associated lung pathology is a major clinical concern.

At present, no animal model recapitulates all aspects of COVID-19 in humans. Among different available animal models, so far hamsters have been widely utilized in SARS-CoV-2 infection studies. Hamster model of COVID-19 mimics a mild pattern of human disease with full recovery. Although SARS-CoV-2 infection of hamsters induces lung pathologies like pulmonary edema, consolidation, and interstitial pneumonia, but failed to develop diffuse alveolar disease, a prominent clinical feature noticed in COVID-19 patients experiencing severe disease. Further characterization of this model is essential for the development of effective therapeutics and vaccines against this virus. An in-depth understanding of host response to SARS-CoV-2 infection in hamsters will elucidate this model’s similarity or dissimilarity with human patients. Although efforts to identify differentially expressed proteins in diverse body fluids of COVID-19 patients were made, there is a dearth of evidence related to differentially expressed proteins in human lung tissues at acute and convalescent stages of SARS-CoV-2 infection. Fresh lung tissues of COVID-19 patients during infection or recovery are ethically impossible to obtain. Hence, so far, data obtained...
from tissues of deceased COVID-19 patients are only available. In this regard, tissues obtained from animals infected with SARS-CoV-2 at different days post-infection will prove beneficial. In this study, efforts have been made to quantitatively compare the lung proteome in SARS-CoV-2-infected hamsters at various days post-infection. The differentially expressed proteins identified in this study will provide information on various host proteins that might have a significant role in the pathogenesis of SARS-CoV-2 infection or disease manifestation. Some of the differentially expressed proteins identified in this study can also be validated in easily accessible patient samples such as body fluids as potential biomarkers for predicting the disease course.

2 | MATERIALS AND METHODS

2.1 | Animal ethics

In this study, attempts were made to evaluate the infectivity of one of the local isolates of SARS-CoV-2, IND-ILS01/2020 (GenBank accession ID-MW559533.2) in the Syrian Golden Hamster model. All the experiments were performed with prior approval of the Institutional Biosafety Committee (IBSC) and Institutional Animal Ethical Committee (IAEC). The study was carried out adhering to the guidelines of the Committee for the Purpose of Control and Supervision of Experiments on Animals (CPCSEA), Govt. of India.

2.2 | Animal studies

For this study, eleven hamsters of the age group 6-7 months were acclimatized at the ILS ABSL3 facility for 4-6 days prior to the experiments. As shown in the schematic (Figure 1A), on day zero, six animals were infected intranasally with SARS-CoV-2 (10^5 TCID50) and five animals were mock-infected (only PBS) under intraperitoneal ketamine (200 mg/kg) and xylazine (10 mg/kg) anesthesia. The SARS-CoV-2 virus used in this study was isolated from a clinically confirmed local COVID-19 patient (IND-ILS01/2020; GenBank accession ID-MW559533.2). Virus stock from the 10th passage was titrated by plaque and TCID50 (Median Tissue Culture Infectious Dose) assays and used for this animal challenge study. Out of the six infected animals, three animals were sacrificed on 4 dpi (day post-infection), and the other three were sacrificed on 14 dpi. Out of the five mock-infected animals, three were sacrificed on 4 dpi, and the other two were sacrificed on 14 dpi. Throughout the experiment, all the animals were monitored daily, and body weights were recorded on alternate days. On the day of sacrifice, tissues from all the vital organs and other organs like the pancreas, spleen, and gastrointestinal tracts were harvested, preserved, and further processed for histopathological analysis and viral load estimation. All the groups had two male and one female animal, except the 14 dpi mock-infected group, which had one male and one female animal in it.

2.3 | Sample collection, storage, and processing

Tissues harvested from different organs, including lungs, were divided into three parts. One part was stored in buffered formalin and further processed for histopathological analysis. The second part was immediately put into TRIzol RNA isolation buffer, and the third part was snap-frozen and stored at −80°C until further use. Formalin-fixed tissues were further processed and sectioned for Hematoxylin and Eosin (H&E), Immunohistochemistry (IHC), and Immunofluorescence analysis (IF) as reported earlier.

2.4 | Hematoxylin and Eosin (H&E) staining and lung pathology scoring

Tissues from multiple lobes of all the animals were processed and stained with H&E staining. Stained lung sections were assigned with different numbers and evaluated in a blinded fashion. Lung pathology scoring was performed similarly as reported by Li et al. The scoring was performed in a blinded fashion and cross-validated by two experienced evaluators. Briefly, the sections were scored for three major pulmonary pathological features associated with SARS-CoV-2 infection (bronchiolitis, alveolitis, and vasculitis/endotheliitis) (Supplementary Figure S1). Each feature was scored on a scale of 0-4 and the total score was obtained by adding values for each feature. The total score ranged between 0 and 12.

2.5 | Immunohistochemistry

Collected tissues were processed and sectioned as reported previously. Sections were deparaffinized, rehydrated, and subjected to antigen retrieval (Vector Laboratories) treatment for 20 minutes followed by blocking the endogenous peroxidase with 3% hydrogen peroxide in methanol for 20 minutes. Horse serum (Vector Laboratories) was used for blocking the sections for 30 minutes at room temperature and incubated with Ki-67 antibody (#VP-RM04; Vector Laboratories, 1:100) or rabbit polyclonal anti-human Sftpd antibody (1:500) antibody overnight at 4°C. Sections were washed twice
with 1× PBS and treated with biotinylated anti-rabbit/mouse IgG secondary antibody (Vector Laboratories) for 45 minutes, followed by ABC reagent for 30 minutes. Diaminobenzidine (Vector Laboratories) was used as a substrate to develop the stain. Hematoxylin was used as a counterstain followed by dehydration with alcohol, clearing with xylene, and mounting with permanent mounting media (Vector Laboratories). Stained sections were observed under the microscope (Leica DM500), and images were taken at different magnifications.

**FIGURE 1** Syrian Golden Hamster model of SARS-CoV-2 infection by a local SARS-CoV-2 isolate. A, Study design to evaluate the infectivity of local-isolate in Syrian Golden Hamster infection model. B, Graph showing percent body weight change in hamsters after mock-infection or SARS-CoV-2 infection. C, Digital images of lungs harvested from mock-infected or infected animals. At 4 dpi, infected lungs have massive congestions visible from the surface (highlighted with white border). No gross changes were noticed in mock-infected (4 dpi) and infected (14 dpi) lungs. Images showing H&E stained lung tissues harvested from mock-infected or infected (4 or 14 dpi) hamsters (scale bar = 200 µm). D, Immunofluorescence images of mock-infected or infected lung tissue sections showing presence of Nucleocapsid protein (N) in bronchial and alveolar epithelial cells. E, Immunoblot analysis showing the Nucleocapsid protein expression in the lungs previously archived lysate (normal), mock-infected (4 dpi), infected 4 dpi or 14 dpi lung tissues. F, Graph showing the viral RNA quantification by RT-qPCR in hamster lung tissues. G, Immunohistochemistry staining with Ki67 showing cell proliferation of bronchial and alveolar cells (scale bar = 50 µm). H, Graph showing pathology score of the lung tissues after the infection (4 and 14 dpi).
2.6 | Immunofluorescence

Sections were deparaffinized, rehydrated, and subjected to antigen retrieval treatment and serum blocking as reported earlier.22 Sections were incubated with SARS-CoV-2 N protein (Nucleocapsid) (#11-2003; Abgenex, 1:200) or SOD2/Mn-SOD (#NB100-1992; Novus biologicals, 1:100) in a humidified chamber overnight at 4°C. Sections were washed twice with PBST for 5 minutes each and incubated with anti-Rabbit Alexa Fluor 594 (#A-11037; Life technologies, 1:500) or anti-Mouse Alexa Fluor 594 (#A-11005; Life technologies, 1:500) for 45 minutes under dark conditions at room temperature. Sections were washed with PBST twice and mounted with ProLong Gold Antifade reagent with DAPI (#P36935; Invitrogen) and visualized using Leica TCS SP8STED confocal microscope.

2.7 | Western blot analysis

For Western blot analysis, cells were re-suspended in Radioimmunoprecipitation assay (RIPA) buffer (20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 50 mM NaF, 1 mM Sodium orthovanadate (Na3VO4), 0.1% SDS, and 0.5% TritonX-100) containing the protease inhibitor cocktail (Thermo Scientific). The whole-cell lysates (WCL) were subjected to SDS-PAGE and transferred to nitrocellulose membrane (Thermo Scientific), followed by blocking and immunoblotting with antibodies specific for SARS-CoV-2 Nucleocapsid protein (#11-2003, Abgenex) or β-actin (#4970, CST).26

2.8 | qRT-PCR

RNA isolation was carried out from hamster tissue samples using TRizol reagent (#10296010, Invitrogen). The isolated RNA was subjected to qRT-PCR for determining the viral load. We performed one-step multiplex real-time PCR using TaqPath 1-Step Multiplex Master Mix (#A28526, Thermo Fisher Scientific), targeting SARS-CoV-2 gene with primer and probe set specific for nucleocapsid (N). The standard curve for absolute quantification of viral genome copies was generated using log-fold dilutions of plasmid harboring the SARS-CoV-2 nucleocapsid gene.27

2.9 | Sample preparation for proteomics analysis

The lung tissue samples of SARS-CoV-2-infected on 4 and 14 dpi along with mock-infected (4 dpi) were processed and homogenized using liquid nitrogen and lysed in lysis buffer containing 4% sodium dodecyl sulfate (SDS) and 50 mM triethylammonium bicarbonate (TEABC). The samples were subjected to sonication three times for 10 seconds storing on ice to prevent overheating between the sonication and were followed by heating at 90°C for 5 minutes. The lysates were then incubated at room temperature for cooling and centrifuged at 12 000 rpm for 10 minutes. The protein concentration present in the supernatant was determined using a bicinchoninic acid assay (BCA) kit (Thermo Scientific Pierce) and an equal amount of protein from each group was pooled for further analysis. LC-MS/MS approach was used using isotopomer labels, “tandem mass tags” (TMTs), to determine the relative quantification of proteins.28

2.10 | In-solution digestion and TMT labeling

Protein lysate of 300 µg from pooled samples of each group was reduced by incubating in 10mM Dithiothreitol (DTT) at 60°C for 20 minutes. Alkylation was carried out in the dark with 20 mM iodoacetamide (IAA) at room temperature for 10 minutes. The lysate was further subjected to acetone precipitation, and the pellet was dissolved in 50 mM TEABC. The reaction was acidified using 0.1% formic acid, and the peptides were lyophilized and stored at −80°C until further use.

2.11 | LC-MS/MS acquisition

The digested peptides were fractionated into six fractions using StageTip fractionation. Fractionated peptides (6 fractions in triplicate, total of 18 runs) were analyzed on Orbitrap Fusion Trubrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) interfaced with EASY-nLC 1000 nanoflow liquid chromatography system (Thermo Scientific, Odense, Southern Denmark). Each fraction was reconstituted in solvent A (0.1% formic acid) and loaded onto trap column (75 µm × 2 cm) Thermo Scientific Acclaim PepMap 100 C18 (#164535; Thermo Scientific) (3 µm particle size, pore size 100 Å) at a flow rate of 5 µl/min with solvent A (0.1% formic acid in water).

The peptides were resolved on an analytical column (EASY-Spray C18 Reversed Phase HPLC Column, 2 µm, 75 µm × 500 mm; Thermo Scientific) using a linear gradient of 7%-30% solvent B (0.1% formic acid in 95% acetonitrile) over 100 minutes at a flow rate of 300 nl/min. Data-dependent Mass Spectrometry acquisition was carried out at top speed mode with full scans (350-1500 m/z) acquired using an Orbitrap mass analyzer at a mass resolution of 120 000 at 200 m/z. For MS/MS, top intense precursor ions from a 3-second duty cycle were selected and subjected to higher-energy collision dissociation (HCD) with 35% normalized collision energy. The
2.12 Bioinformatics and statistical analysis

The Proteome Discoverer 2.3 (Thermo Scientific, Bremen, Germany) was used to carry out protein identification and quantitation. All raw files were searched against a Mesocricetus auratus (Syrian Golden hamster) protein database in Universal Proteins Resource Knowledgebase (UniProt) (32,336 entries) supplemented with common contaminants (116 entries) using SequestHT as a search algorithm. The search parameters included trypsin as the proteolytic enzyme with a maximum allowed missed cleavages to two. Oxidation of methionine and acetylation of protein N-terminus were set as dynamic modifications. In contrast, static modifications included cysteine carbamidomethylation and TMT modification at the N-terminus of the peptide and lysine residue. Precursor mass tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.05 Da. Peptide Spectrum Matches were identified at 1% False Discovery rate. The differential expression ratios between the groups were calculated. Proteins with differential expression ratios ≥1.5 (upregulated) or ≤0.67 (downregulated) were considered as differentially expressed. The significance of differences between groups was calculated using Student’s t test (two-tailed), and a P-value ≤ .05 was considered statistically significant. All the differentially expressed proteins across the two groups (Infected 4 dpi and infected 14 dpi) were compared using a Venn diagram (Venny 2.1, https://www.bioinfogp.cnb.csic.es/tools/venny/) and further analyzed with PANTHER classification system, version 16.0 (http://www.pantherdb.org). Gene Ontology (GO) and pathway analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed using the Enrichr online tool (http://www.amp.pharm.mssm.edu/Enrichr/).

Principal component analysis was performed using sample-wise scaled unfiltered normalized protein abundance data using PCAtools. Heatmaps were generated using protein-wise scaled and filtered protein abundance (1.5 and 1.3 up- or down-regulation; P-value ≤ .05 in any comparison) and k-means clustering (k = 3).

3 RESULTS

3.1 Clinical features, viral load, and histopathological changes in lungs

To investigate the pathogenicity of SARS-CoV-2, Syrian hamsters were infected with the virus, and harvested tissue samples were collected at two different time points (4 & 14 dpi) for viral load and pathological analysis (Figure 1A). Corroborating earlier reports, a significant weight loss was noticed in all the infected animals at 4 dpi. On 4 dpi, all the animals lost around 15% of their initial body weight (Figure 1B). After 6 dpi, the infected animals started regaining their body weight. During the course of the experiments, no mortality was found in both infected and mock-infected animals. At the time of organ isolation, congestion of lungs was grossly visible only in infected animals at 4 dpi (Figure 1C). Further histopathological analysis of the 4 dpi SARS-CoV-2-infected tissue samples showed the presence of severe pathological lesions in the lungs (Figure 1C and Supplementary Figure S1A-G). Multifocal necrosis and the desquamation of bronchial epithelial cells and infiltration of inflammatory cells were present in the SARS-CoV-2-infected lung tissues (Figure 1C). Around 50% lung area was affected in all the infected animals, and the lesions were patchy throughout the lungs. Necrosuppurative bronchitis and interstitial pneumonia were evident in all the infected animals at 4 dpi (Supplementary Figure S1D,E). Different areas of the infected lung tissues showed consolidation of lungs and hemorrhage (Supplementary Figure S1C). These hamsters exhibited severe interstitial pneumonia, as evidenced by the thickening of the alveolar wall, altered alveolar structure, and immune cells’ infiltration (Supplementary Figure S1D). At 4 dpi, infected lungs have occasional features of diffuse alveolar damage (DAD) including necrosis of alveolar epithelial cells, presence of intra-alveolar immune cells, cellular debris, and protein exudates (Supplementary Figure S1F). However, hyaline membrane formation was very rarely noticed, which indicates a less severe form of DAD. Endothelium near the damaged areas was reactive, as evidenced by mononuclear cells’ adhesion to the endothelium (Supplementary Figure S1A). In certain instances, the immune cells have invaded the vessel wall and caused endotheliitis (Supplementary Figure S1B). In one of the infected animals (4 dpi), visceral pleural invasion of immune cells was also noticed (Supplementary Figure S1G). However, no noticeable histopathological changes were observed in mock-infected hamster lung at any time point. After 14 dpi, hamsters infected with SARS-CoV-2 exhibited only mild inflammatory infiltration and tissue damage suggestive of the resolution of disease manifestation (Figure 1C). Histopathological evaluation of tracheal tissues from all the animals also showed severe tracheal epithelial and endothelial damage in 4 dpi-infected tissues compared to 14 dpi-infected or mock-infected tissues (Supplementary Figure S2). The significantly higher lung pathology score at 4 dpi-infected tissues than 14 dpi corroborates the earlier reports and indicates the self-limiting nature of this disease in the hamsters (Figure 1H).
Immunofluorescence (IF) staining of lung tissue sections with SARS-CoV-2 nucleocapsid (N) protein showed the viral antigen-positive bronchial and alveolar epithelial cells at 4 dpi, which were not detected at 14 dpi tissues (Figure 1D). Immunoblot analysis of the lung tissue protein lysates also corroborated the IF findings (Figure 1E). The viral genome copy number estimation showed the presence of viral genome in both 4 dpi and 14 dpi-infected lung tissues; however, the copy number was significantly low in 14 dpi tissues compared to 4 dpi tissues (Figure 1F). Immunohistochemical staining for Ki67, a marker of cell proliferation, showed marked cellular proliferation (hyperplasia) of bronchial and alveolar cells at 4 dpi (Figure 1G). This finding corroborates with the earlier report.20

3.2 | Quantitative proteomics analysis

3.2.1 | Molecular alterations in pulmonary pathology induced by SARS-CoV-2 infection

Analysis of all the animals’ body weights, lung pathology, and status of viral protein in lung tissues clearly demonstrated normal animal health in both the 4 dpi and 14 dpi mock-infected animals. At the same time, lung tissues from the 4 dpi and 14 dpi-infected hamsters manifested features of acute and convalescent stages of the infection, respectively. One of the major objectives of the current study was to compare the lung proteome of SARS-CoV-2-infected hamster lung tissues at acute and convalescent stages of the infection. Hence, we employed a quantitative proteomics approach to identify the proteomic alterations in hamster lung tissue induced by SARS-CoV-2 at different time points (4 dpi and 14 dpi) compared with only 4 dpi mock-infected control tissues (as a representative of mock-infected groups). The quantitative proteomics data were analyzed using high-resolution LC-MS/MS in triplicates where the raw files were searched using Proteome Discoverer 2.3. The search resulted in the identification of 1585 proteins expressed across all the samples. Of these, 50 and 18 proteins were differentially expressed (cut off 1.5-fold, \( P \leq .05 \)) at 4 dpi and 14 dpi, respectively. The 50 differentially expressed proteins in 4 dpi included 33 upregulated and 17 downregulated proteins whereas 18 differentially expressed proteins in 14 dpi included nine upregulated and nine downregulated proteins. A complete list of the proteins identified is provided in Supplementary Table S1. All the differentially expressed proteins were further represented as a heat map using hierarchical clustering analysis comparing mock-infected with the infected samples (Figure 2A). Interestingly, the principal component analysis (PCA) demonstrated distinct protein expression patterns among the mock-infected 4 dpi and 14 dpi-infected samples, depicting the number of common and differentially expressed proteins. This comparison was carried out using a Venn diagram (Venny 2.1, https://www.bioinfogp.cnb.csic.es/tools/venny/). Upon comparison, we identified 41 proteins exclusive to 4 dpi tissue samples and nine exclusive to 14 dpi tissue samples when compared with mock-infected (4 dpi) tissue samples, while nine proteins were shared between the groups (Supplementary Figure S3). Of the total identified proteins, 50 proteins were differentially expressed (33 upregulated and 17 downregulated; \( P \leq .05 \)) in 4 dpi-infected groups. Among these proteins, majority were involved in the blood coagulation, integrin signaling pathway, alternative complement activation signaling pathway, and plasminogen activating cascade (PANTHER classification system, version 16.0).32 Similarly, 18 proteins were differentially expressed (nine upregulated and nine downregulated; \( P \leq .05 \)) in the infected 14 dpi group. This included proteins belonging to the plasminogen activating cascade and PI3K-Akt signaling pathway. The differentially expressed proteins are graphically represented as a heat map and volcano plots representing distinct proteomic patterns between the two groups compared to mock-infected (Supplementary Figure S4A-C). We also evaluated the altered protein expression during the acute and convalescent stages of infection (14 dpi vs 4 dpi) (Supplementary Figure S4D).

We further compared the differentially expressed proteins of statistical significance across the SAR-CoV2-infected (4 dpi and 14 dpi) and mock-infected (4 dpi) groups. A comparison of the downregulated proteins and upregulated proteins from both groups did not result in any common protein. However, among the 33 proteins upregulated in the 4 dpi tissue samples, three proteins such as superoxide dismutase (Sod2), myosin-2 (Myh2), and calgranulin-B (S100a9) were downregulated in the 14 dpi tissue samples. These proteins are known to regulate various cellular processes such as antioxidant defense, cellular localization, cell adhesion, and cellular migration. Similarly, among the 33 proteins upregulated in the 4 dpi lung tissue, four proteins, including Fibrinogen beta chain (Fgb), Ferritin (Fth1), Calpactin I (S100a10), Thymosin (Tmsb4x) were found to be upregulated in a 14 dpi sample. These proteins were involved in the regulation of biological processes and exhibit catalytic activity. On further comparison of downregulated proteins identified across
FIGURE 2  Lung tissue proteome of hamsters infected with SARS-CoV-2 at different time points. A, Heatmap of the differential expressed proteins ($P \leq .05$; fold change cut off of 1.5) were plotted sample wise suggesting the proteomic alterations across the infected samples (4 and 14 dpi) compared to mock-infected (4 dpi). B, Principal Component Analysis (PCA) was performed and plotted using PCAtools with technical replicates marked by R1, R2, and R3. The PCA plot represents the variance explained by the principal components (denoted by PC) indicating a clear separation of the samples among the three groups, ie, mock-infected (4 dpi), infected (4 dpi), and infected (14 dpi). C, Pathway analysis was performed using Enrichr online tool using the KEGG database. The enrichment analysis was performed for upregulated and downregulated differentially expressed proteins for all the three groups infected (4 dpi) versus mock-infected (4 dpi); Infected (14 dpi) versus mock-infected (4 dpi); and infected (14 dpi) versus infected (4 dpi). The horizontal axis represents the enrichment score $-\log_{10}$ ($P$-value) of the pathway and the vertical axis represents the pathway category. The red color represents upregulated proteins and the blue bar represents downregulated proteins.
both 4 dpi- and 14 dpi-infected tissues, two proteins were common in both groups, the serine/arginine-rich splicing factor 1 (Srsf1) and Guanine nucleotide-binding protein G (q) subunit alpha (Gnaq).

### 3.2.2 Functional characteristic of significantly altered proteins

To systematically investigate the molecular differences in the hamster lung owing to SARS-CoV-2 infection, we carried out a Gene Ontology (GO) analysis of the differentially expressed proteins. The biological process such as “platelet degranulation,” “regulated exocytosis,” and “fibrinolysis” were enriched mainly in the proteins upregulated in infected (4 dpi) lung tissues, whereas “wound healing,” “collagen fibril organization,” and “actin cytoskeleton reorganization” were enriched in downregulated proteins (Supplementary Table S2). The proteins associated with these pathways include fibrinogens (Fga, Fgb, and Fgg); complement factors (C4a, C3); alpha-2-antiplasmin (Serpinf2); Alpha-1-B Glycoprotein (A1bg), Apolipoprotein A1 (Apoa1), and Alpha-1-acid glycoprotein 1 (Orm1). Notably, the terms such as “regulation of substrate adhesion-dependent cell spreading” and “endothelial cell migration” were mostly observed in upregulated and downregulated proteins in 14 dpi (Supplementary Table S3).

### 3.2.3 Pathway analysis

KEGG pathway analyses of the differentially expressed proteins in lung tissue at 4 dpi compared to mock-infected (4 dpi) tissue revealed that SARS-CoV-2-triggered the activation of complement and coagulation cascade and ferroptosis. The analysis also resulted in the identification of other aberrant pathways such as platelet activation, focal adhesion, and tight junction. Interestingly, we also observed an aberrant expression of proteins associated with necroptosis, cholesterol metabolism, ferroptosis, and interleukin-17 signaling pathway in the 14 dpi tissue samples suggestive of the underlying mechanisms that lead to tissue damage and lung injury during SARS-CoV-2 infection (Figure 2C).

### 3.3 Secretory proteins

To investigate the amount of secretory proteins among the pool of differentially expressed proteins in the infected tissue samples, we compared our data with the proteins annotated as “secretory proteins (2640 proteins),” “Secreted in the blood (729 proteins),” “Lung enriched (13 proteins),” “Lung proteome (19 649 proteins),” and “Group enriched (61 proteins expressed in the lung)” (Supplementary Figure S4E). We found four proteins, Advanced Glycosylation End-Product-Specific Receptor (Ager), secretoglobin family 1A member 1 (Scgb1a1), Surfactant associated protein B (Sftpba), and Surfactant associated protein D (Sftpd) identified in the current study were specific to lung proteome.

### 3.4 Validation of selected proteins’ expression status

To validate the hamster lung proteome data identified in this study, we randomly picked two proteins, SOD2 and Sftpba, and checked their expression through IF and IHC, respectively. We found a significantly higher level of SOD2 expression in the infected 4 dpi tissues than 4 dpi mock-infected or 14 dpi-infected tissues (Figure 3A). Similarly, analysis of Sftpba expression showed lower expression of this protein in bronchial epithelial cells of 4 dpi and 14-dpi infected lung tissues than 4 dpi mock-infected samples (Figure 3B). Together, these data successfully validated the results of the mass spectrometry-based proteomics analysis.

### 4 DISCUSSION

Multiple studies across the globe have demonstrated that the hamster model of SARS-CoV-2 infection mimics the milder form of COVID-19 in humans. Earlier study by using USA-WA1/2020 (NR-52281, BEI Resources) isolate showed that 5 x 10^4 TCID50 dose of viral infection in hamsters lead to weight loss and full recovery by 14 days. However, an infection dose of 5 x 10^5 TCID50 dose induced severe weight loss and partial mortality. Similarly, in a different study with an infection dose of 8 x 10^5 TCID50 of BetaCoV/Hong Kong/VM20001061/2020 (GISAID# EPI_ISL_412028) virus, the investigators have reported a self-limiting model of SARS-CoV-2 with the manifestation of earlier lung tissue damage followed by recovery. All these studies also corroborate the self-limiting nature of this model as reported by others. In the recent study, Mohandas et al have used an Indian SARS-CoV-2 isolate (NIV-2020-770) at different infection doses and noticed that 10^{1.5} TCID50 and 10^{4.5} TCID50 infection dose of the virus manifested a self-limiting disease with lung pathologies at the earlier days of infection. Based on the aforementioned studies we picked a dose of 10^5 TCID50 for the inoculation of hamsters with our isolate IND-ILS01/2020 (GenBank accession ID- MW559533.2). Similar to the earlier report, our analyses also showed a non-lethal and self-limiting model of SARS-CoV-2 infection in hamster. The histopathological and viral analysis of lung samples clearly showed that 4 dpi- and 14 dpi-infected animals represent acute and convalescent
The absence of the severe form of DAD features also corroborates earlier reports and indicates the limitation of this model.\textsuperscript{18}

Our quantitative proteomic analysis data showed the dysregulation of biological processes such as “collagen fibril organization” and “actin cytoskeleton reorganization” in infected 4 dpi and “regulation of substrate adhesion-dependent cell spreading” and “endothelial cell migration” in the infected 14 dpi group. Collagen network organization occurs in response to tissue damage and is a part of the wound healing process and can trigger the onset of fibrosis. The actin cytoskeleton is an important cellular component, essential for maintaining the shape and structure of the cells. Many viruses promote the rearrangement of the host cell cytoskeleton to facilitate their dissemination.\textsuperscript{36} We observed reduced expression of the proteins that contribute to dysregulation in actin polymerization at both 4 dpi and 14 dpi (Supplementary Tables S2 and S3). Reports also indicate that any alteration in the cell-matrix adhesion and extracellular matrix during injury repair can disrupt lung structure, further leading to lung damage. Our results also indicate the dysregulation of collagens in the 14 dpi lung tissues. The proteins associated with “endothelial cell migration” include Thymosin beta-4 (Tmsb4x) and High Mobility Group Box 1 (HMGB1), both of which have been reported to be potential therapeutic targets in drug discovery. Tmsb4x is a small and water-soluble peptide known for its role in angiogenesis, wound healing, and increased metastatic potential of tumor cells.\textsuperscript{37,38} Its ability to induce fibrinolysis makes it an interesting molecular drug target. Recently, HMGB1 has also emerged as a potential target for therapeutic interventions for COVID-19. It is known to play a critical role in various infections, and its elevated expression has been reported in many viral infections, including COVID-19.\textsuperscript{39,40} This association further results in receptor-dependent responses suggesting its possible role in SARS-CoV-2 infection. Further studies are warranted to uncover the functional role of HMGB1 and evaluate its inhibitors in the COVID-19 treatment.

We also observed multiple pathways such as the dysregulation of complement and coagulation cascades, platelet degranulation, and ferroptosis upon viral infection across both the groups (4 dpi and 14 dpi). Existing literature also indicates the critical role of the complement pathway in pathogenesis and disease severity of SARS-CoV-2.\textsuperscript{41,42} The complement system serves as the host systems’ first response to foreign pathogens and regulates processes such as opsonization, chemotaxis, leukocyte recruitment, activation, and phagocytosis. However, unrestricted activation of this pathway contributes to acute and chronic inflammation, coagulation, cell injury, and facilitates multiple organ failure leading to death.\textsuperscript{43}
Significant dysregulation of complement and coagulation cascades along with elevated levels of D-dimer, fibrinogen, and von Willebrand factor have been observed in SARS-CoV-2-infected patients. In the current study, we identified higher expression levels of C3, Fgb, Fga, Fgg, Serpinf2, and Cfb in 4 dpi hamster lung tissues as compared to mock-infected (4 dpi). The higher levels of these proteins were well studied as a marker of the activation of early complement and coagulation cascade. The significant enrichment of coagulation proteins in SARS-CoV-2-infected tissues indicates the disruption of coagulation mechanisms during SARS-CoV-2 infection. In addition, the higher expression of proteins such as Fgb, Fga, and Fgg identified in the study are also known to be involved in platelet activation. Platelets are non-nucleated cell fragments derived from megakaryocytes and essential for physiological hemostasis. Additionally, they are also known for their diverse role in inflammatory and immune response by acting as inflammatory effector cells. They also serve as an indispensable element in coagulation and inflammation and their activated state is associated with cancer progression. Recently, the lung has also been proposed as a platelet biogenesis site, accounting for almost 50% of total platelets. There are reports demonstrating the role of platelets in inflammatory lung syndromes/disorders such as acute respiratory distress syndrome (ARDS), chronic obstructive pulmonary disease (COPD), cystic fibrosis (CF), aspirin-exacerbated respiratory disease (AERD), and asthma.

Our pathway analysis with significantly downregulated proteins list also identified certain proteins involved in platelet activation (Col1a2, Gnaq, Rhoa), focal adhesion (Col1a2, Parva, Rhoa), and leukocyte transendothelial migration (Cldn18, Rhoa) in day 4 infected vs mock-infected animals. Even though the pathways are commonly shared among the up and downregulated proteins, the biological properties of each protein need to be considered while concluding. The contrasting results from the pathway analysis of upregulated and downregulated proteins highlight the complex nature of these biological processes. Together, we believe that the current pathway analysis only gives a preliminary idea about the differentially expressed proteins in each group and their known biological function. Thus, more studies are required to elucidate the role of each pathway component in the pathophysiology of COVID-19.

Ferroptosis is a form of programmed cell death associated with unchecked lipid peroxidation due to the accumulation of lipid reactive oxygen species (ROS) in cells. Its role is well documented in pathophysiological processes of various diseases, such as tumors, nervous system diseases, ischemia-reperfusion injury, kidney injury, and blood diseases. Iron is a pivotal component of the ferroptosis pathway and disruption of iron metabolism has been reported in COVID-19 patients. In our study, we observed increased ferritin (Fth1) levels in 4 dpi tissues sample. Its increased expression has also been reported in COVID-19 patients. Studies on COVID-19 patients have demonstrated the usefulness of inflammatory markers such as procalcitonin, C-reactive protein, erythrocyte sedimentation rate, and serum amyloid A as an indicator of disease progression, however, little is known about the increased ferritin levels (hyperferritinemia) in these patients. In our study, we observed increased ferritin (Fth1) levels in both 4 dpi and 14 dpi lung tissues sample which corroborates with the recent findings where its increased expression has been reported in COVID-19 patients. Further studies are warranted to evaluate its role as a pathogenic mediator in COVID-19. In addition, we also report here increased levels of Fgb, and S100a10 in both groups when compared to mock-infected. Fgb is a blood-borne glycoprotein, with an active role in biological functions such as angiogenesis, wound healing, and inflammation. It serves as a marker for vascular injuries along with other pathological conditions such as colitis, lung and kidney fibrosis. Consistent with lung fibrosis observed in hamster post-infection with SARS-CoV-2 in the current study, Fgb was higher among both groups suggesting its role in blood clot formation where its higher expression is also reported in COVID-19 patients. Calpain I, belongs to the calcium-binding S100 family, and is ubiquitously expressed in the majority of cells. It is known for its role in wound healing, fibrinolysis, and angiogenesis. Binding of S100a10 with plasminogens facilitates its conversion to plasmin which in turn enhances the virulence and pathogenicity of SARS-CoV-2 by cleaving the spike protein. Moreover, we identified lower levels of Srsf1 and Gnaq among both the infected group as compared to the mock-infected. Srsf1 belongs to the family of splicing regulators and its interaction with different proteins enables it to regulate a plethora of cellular pathways. It plays a significant role in providing genomic stability and thus, viral infection results in the depletion of this protein along with other RNA binding proteins (RBPs) in the nucleus. Such association of the viral genome with host RBPs initiates apoptosis, further resulting in the release of virally directed infection by compromising the host machinery. Gnaq, constitute the family of largest cell surface membrane receptors, expressed ubiquitously in mammalian cells, and known to be involved in multiple ways during viral infection. Studies have also demonstrated its role in regulating both innate and adaptive immunity. Here we report the reduced expression of Gnaq in hamster lung tissues infected with SARS-CoV-2. Similar data were also observed in virus-infected macrophages implicating its role as a negative regulator of antiviral immune responses.

Pulmonary surfactant proteins constitute a type of lipoprotein complex comprising 90% lipids and 10% surfactant proteins (Sftpa, Sftpb, Sftpc, and Sftpd). These surfactant proteins contribute to providing defense against pathogens and play a critical role in efficient gaseous exchange at the air-liquid interface in the alveoli and provide lung stability.
We observed aberrant expression of the surfactant proteins (Sftpβ, and Sftpδ) across the 4 and 14 dpi-infected samples compared to the mock-infected samples suggesting the decline in the normal functionality of lung respiratory gaseous exchange owing to virus infection. The previous report has suggested the role of viral proteins in modulating the surfactant metabolism and thus, resulting in host immune compromise.

As reported earlier, we also noticed distinct damage to the epithelial-endothelial barrier in 4 dpi-infected lung tissues. This barrier system’s damage is believed to be the major mediator of ARDS in different respiratory viral infections, including SARS-CoV-2.67 The loss of the epithelial-endothelial barrier allows leakage of blood components into the alveolar lumen and lung interstitium. At the same time, it also allows the leakage of lung proteins into circulation.68 In the recent past, it is shown that COVID-19 patients who developed ARDS have significantly higher IL-6 and Sftpδ circulatory levels compared to patients who did not have ARDS.69 These findings further indicate that the identification of pneumo-proteins in circulation is an indicator of the severity of lung pathology. In our lung tissue proteome analysis, we noticed a significant downregulation of Sftpδ protein in tissues with high lung pathology score (4 dpi), which suggests that certain proteins might have a differential pattern of expression level in circulation and at the primary site of infection (lungs). Based on several features of pulmonary surfactants, they are believed to have importance in COVID-19 pathogenesis, diagnosis or therapy. It has been proposed that a lower concentration of pulmonary surfactant is a critical risk factor for COVID-19.70 Studies have also reported that concentrations of SP-A and SP-B were low in bronchoalveolar lavage (BAL) of patients before and after the onset of ARDS.71 The low level of pulmonary surfactant proteins detected in 4 dpi lung tissues underscores their possible role in the COVID-19 pathology. Altogether, these pulmonary lung surfactants might be considered as potential therapeutics to aid in COVID-19 treatment. Secretoglobin Family 1A Member 1 (Sgb1a1) encodes a member of the secretoglobin family of small secreted proteins, a component of pulmonary surfactant, which is expressed in mucus-secreting cells. This protein is known for its anti-inflammatory/ immunomodulatory and anti-fibrotic functions.72 We identified severe downregulation of Sgb1a1 in the lung tissue post-infection (4 dpi), suggesting respiratory distress owing to virus-mediated lung injury. Some literature have also reported its altered expression following lung injury and where its absence is marked with the greater inflammatory response.73-75 It is also known to regulate alveolar macrophage-mediated inflammatory response upon virus invasion.75 However, further studies are required to study its effectiveness and the underlying mechanism(s) associated with virus infection. Further, it suggests the usefulness of the hamster model of SARS-CoV-2 infection in evaluating the therapeutic efficacies of these proteins for COVID-19.

Through an ultra-high-throughput clinical proteomics approach, Messner et al, identified protein expression signatures in serum/plasma samples including complement factors, components of coagulation systems, immunomodulators, and proinflammatory factors that can classify COVID-19 patients based on WHO grading.4 The authors have proposed 27 protein groups (23 upregulated and 4 downregulated) as potential biomarkers of disease severity. Out of these upregulated proteins, we noticed six proteins Complement factor B (Cfb), Fibrinogens (Fga, Fgb, Fgg), Haptoglobin (Hp), and Galectin-3-binding protein (Lgals3bp) are also present in the list of upregulated proteins at 4 dpi vs mock-infected groups of our study. However, albumin (Alb), and transferrin (Tf) whose downregulation correlated with COVID-19 severity4 are found to be upregulated at the 4 dpi infection group of our study. There are multiple possibilities for this discrepancy like (a) blood proteome and lung proteome might be different in SARS-CoV-2-infected humans or animals, (b) difference in techniques or methods used to analyze samples, and (c) species-dependent differential (hamsters or human) host response to SARS-CoV-2 infection. Similarly, proteomic and metabolomic profiling of serum samples obtained from 46 COVID-19 and 53 control individuals showed deregulation of three major pathways namely complement system, macrophage function, and platelet degranulation in severe COVID-19 patients.41 A study by Park et al, employed an in-depth proteome profiling of undepleted plasma revealed signatures of proteins involved in neutrophil activation, platelet function, and T cell suppression.8 Based on their findings, the authors have proposed specific plasma proteins as predictive biomarkers of COVID-19.6 In a different study, proteomic analysis of serum samples from early COVID-19 patients also identified differentially expressed proteins known to have a function in SARS-CoV-2 infection-associated inflammation and immune signaling.7 The findings of our study also corroborate the aforementioned findings.

Efforts to investigate the alteration in bronchoalveolar lavage fluid (BALF) proteome in COVID-19 patients compared to the non-COVID-19 controls demonstrated that SARS-CoV-2 infection induces alteration in BALF proteome with enrichment of proteins involved in proinflammatory cytokine-mediated signaling and oxidative stress response.5 Superoxide dismutase (SOD) is vital for human health and upregulation of SOD2 expression upon challenge with human pathogens suggests its role in innate response. Antioxidant enzymes such as SOD2 are pivotal to protect from free superoxide anion, which can damage epithelial cells and impair their function. Thus, enhanced expression of Sod2 in SARS-CoV-2-infected hamster lungs, suggests the upregulation of antioxidant response
to prevent oxidative stress-induced tissue damage, lung injury, and respiratory distress.

Taken together, the current study highlights the proteomic alterations caused owing to SARS-CoV-2 infection during the course of infection and provides an insight on the molecular pathogenesis and possible therapeutic targets of COVID-19. Importantly, the current study provides strong molecular evidence that shows the similarities between SARS-CoV-2 infection in humans and hamsters and supports the clinical relevance of this model in COVID-19 research. However, the present study has some limitations of the small sample size considered for the analysis and use of limited experimental conditions (single viral dose and strain). This demands similar extensive research in a large number of animals and different experimental conditions for further validation. Moreover, due to the small number of animals, the effects of age, sex, and comorbidities await further investigation. The unavailability of complete genome information of Syrian golden hamsters has also substantially restricted our findings. We hope in the future with the availability of complete genome sequence and hamster-specific reagents like antibodies will help in obtaining more relevant information from our current findings.

ACKNOWLEDGMENTS

The study is financially supported by the DBT-Institute of Life Science (ILS) and DBT-BIRAC (BT/CS0004/CS/02/20). We sincerely acknowledge the technical supports given by Madan Mohan Mallick, Sushanta Kumar Swain, and Bhabani S. Sahoo, ILS. We are thankful to Dr. Kunal Sahu, Veterinary pathologist, ILS for his valuable support in the histopathological analysis. We are grateful to Dr. Taruna Madan Gupta, NIRRH, Mumbai, for providing us the SftpD (SP-D) antibody. Voddu Suresh is a recipient of the Council of Scientific and Industrial Research (CSIR) students’ research fellowship, Government of India. Varshasana Mohanty is a recipient of the Women Scientist-A award from the Department of Science and Technology (DST), Government of India. G. H. Syed acknowledges the support from DBT-Wellcome Trust India Alliance (IA/I/15/1/501826).

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

S. Senapati and G. H. Syed, conceived the idea, and supervised the overall experiments. V. Suresh, and K. Avula contributed to animal inoculation and tissue analysis. B. Singh, K. Avula, and G. H. Syed contributed to viral copy number estimation and viral culture. V. Mohanty and R. K. Reddy carried out sample preparation for proteomics analysis. V. Mohanty, A. Ghosh, and V. Suresh were involved in proteomics data analysis, data interpretation, prepared figures and tables. A. R. Suryawanshi and S. K. Raghav crosschecked the bioinformatics data analysis. V. Suresh and D. Parida did tissue processing and staining. S. Senapati carried out the pathologic evaluation of tissue sections as one of the evaluators. S. Chattopadhyay, P. Prasad, R. K. Swain, R. Dash, and A. Parida facilitated in establishing Standard Operating Procedures (SOPs) for ABSL3-related activities and intellectually contributed in manuscript writing and interpretation of the data. All authors were involved in manuscript writing and editing.

DATA AVAILABILITY STATEMENT

The raw data files and the MSF files were submitted to the PRIDE partner repository with data set identifier PXD024547.

ORCID

Gulam Hussain Syed https://orcid.org/0000-0001-8540-6162
Shantibhusan Senapati https://orcid.org/0000-0001-7108-8255

REFERENCES


SUPPORTING INFORMATION
Additional Supporting Information may be found online in the Supporting Information section.

Supplementary figure 1

- Reactive endothelium
- Necrotizing bronchiolitis
- Pleural invagination
- Alveolitis
- Endotheliitis
- Consolidation of lungs and haemorrhage
- Diffuse alveolar damage (DAD)
- Pleural invagination
Mock-infected (4dpi) vs. Infected (4dpi)

- Significantly altered
- 50

Mock-infected (4dpi) vs. Infected (14dpi)

- p-Value ≤ 0.05
- Fold Change ±1.5
- Significantly altered
- 18

Upregulated: 33
Downregulated: 17

Infected (4dpi)
(Downregulated, p≤0.05)

Infected (14dpi)
(Upregulated, p≤0.05)

Supplementary figure 3
Secretory proteins (2,641)
Tissue Distribution of ACE2 Protein in Syrian Golden Hamster (Mesocricetus auratus) and Its Possible Implications in SARS-CoV-2 Related Studies

Voddu Suresh1,2, Deepti Parida1,2, Aliva P. Minz1,2, Manisha Sethi1,2, Bhabani S. Sahoo3 and Shantibhusan Senapati1*

1Tumor Microenvironment and Animal Models Lab, Institute of Life Sciences, Bhubaneswar, India, 2Regional Centre for Biotechnology, Faridabad, India, 3Institute of Life Sciences, Bhubaneswar, India

The Syrian golden hamster (Mesocricetus auratus) has recently been demonstrated as a clinically relevant animal model for SARS-CoV-2 infection. However, lack of knowledge about the tissue-specific expression pattern of various proteins in these animals and the unavailability of reagents like antibodies against this species hampers these models’ optimal use. The major objective of our current study was to analyze the tissue-specific expression pattern of angiotensin-converting enzyme 2, a proven functional receptor for SARS-CoV-2 in different organs of the hamster. Using two different antibodies (MA5-32307 and AF933), we have conducted immunoblotting, immunohistochemistry, and immunofluorescence analysis to evaluate the ACE2 expression in different tissues of the hamster. Further, at the mRNA level, the expression of Ace2 in tissues was evaluated through RT-qPCR analysis. Both the antibodies detected expression of ACE2 in kidney, small intestine, tongue, and liver. Epithelium of proximal tubules of kidney and surface epithelium of ileum expresses a very high amount of this protein. Surprisingly, analysis of stained tissue sections showed no detectable expression of ACE2 in the lung or tracheal epithelial cells. Similarly, all parts of the large intestine were negative for ACE2 expression. Analysis of tissues from different age groups and sex didn’t show any obvious difference in ACE2 expression pattern or level. Together, our findings corroborate some of the earlier reports related to ACE2 expression patterns in human tissues and contradict others. We believe that this study’s findings have provided evidence that demands further investigation to understand the predominant respiratory pathology of SARS-CoV-2 infection and disease.

Keywords: lung, COVID 19, angiotensin converting enzyme-2, SARS CoV-2, hamster (Mesocricetus auratus)

INTRODUCTION
The current outbreak of COVID-19 (Corona Virus Disease 2019) caused by the SARS-CoV-2 virus was declared a pandemic on March 11, 2020. It has now succumbed to around 44 million people worldwide, and almost 1.17 million people have lost their lives due to this pandemic (till October 28, 2020). Therefore, there is an urgent need to study this viral disease transmission, pathogenesis, prevention, and treatment. In this regard, the role of clinically relevant experimental animal models...
is crucial. No single species of animal might be able to recapitulate all the SARS-CoV-2 infection-related events in humans exactly. However, using different animal models will help to address questions in a more reliable and clinically relevant manner. Simultaneously, exploring multiple species susceptible to this virus might also help to identify the natural reservoir and potential carriers of this pathogen.

Syrian golden hamsters *(Mesocricetus auratus)* being a permissive host to multiple other viruses and a recognized model for respiratory syndrome coronavirus (SARS-CoV) infection has drawn immediate attention for COVID-19 related studies (Roberts et al., 2005). Studies have shown that the Syrian hamster infected by SARS-CoV-2 manifest various clinical signs of COVID-19 in human patients (Chan et al., 2020a; Chan et al., 2020b; Sia et al., 2020). Moreover, the pathology of this disease in hamsters resembles humans. The model has also demonstrated high transmissibility of SARS-CoV-2 among close contact animals. In a recent study, male Syrian golden hamsters at 4–5 weeks of age were used to check the potential of hamster as a model of SARS-CoV-2 infection (Sia et al., 2020). Analysis of tissue samples from the infected animals showed pathological changes in the respiratory tract and SARS-CoV-2 N protein at bronchial epithelial cells, pneumocytes, and nasal epithelial cells. Duodenum epithelial cells also stained positive for viral N protein. No apparent histopathological changes were observed in the brain, heart, liver, and kidney on five dpi (days post-infection) (Sia et al., 2020). Although no infectious virus was detected in the kidney, low copies of the viral genome were detected on two and five dpi (Sia et al., 2020). In a similar type of study, by using male and female Syrian hamsters of 6–10 weeks old, identified tissue damage and viral N protein presence at different parts of the respiratory tract (nasal turbinates, trachea, and lungs) (Chan et al., 2020a). The viral N protein was abundantly present in bronchial epithelial cells, macrophages, type I, and II pneumocytes. At four dpi, N protein expression was found all over the alveolar wall (Chan et al., 2020a). The histopathological analysis also showed tissue damage and/or inflammatory lesions at multiple extra-pulmonary organs (intestine, heart, spleen, bronchial, and mesenteric lymph nodes); however, N protein expression was only detected in the intestinal epithelial cells (Chan et al., 2020a). In the very recent past, the Syrian hamster model of SARS-CoV-2 infection has been instrumental in establishing that passive transfer of a neutralizing antibody (nAb) protects SARS-CoV-2 infection (Rogers et al., 2020).

Studies have clearly shown SARS-CoV-2 binds to human angiotensin-converting enzyme 2 (ACE2) expressed by its target cells and use it as a functional receptor to enter into cells (Hoffmann et al., 2020; Walls et al., 2020). Hence, drugs that could inhibit the binding of viral proteins (S-protein) to the ACE2 expressed on the target cells are assumed to be potential therapeutics against COVID-19. A recent study has shown that human recombinant soluble ACE2 (hrsACE2) blocks the early stage of SARS-CoV-2 infection (Monteil et al., 2020). We have also proposed the bioengineered probiotics expressing human ACE2 as a potential therapeutics against SARS-CoV-2 infection (Senapati et al., 2020a). The ACE2 protein’s sequence alignment of different species has suggested that the S protein may interact more efficiently with Cricetidae ACE2 than murine ACE2 (Luan et al., 2020). *In silico* analysis also shows possible interaction between SARS-CoV-2 spike proteins with Syrian hamster ACE2 (Chan et al., 2020a).

At the time of the ongoing COVID-19 pandemic, in addition to the vaccine and antiviral development, attempts have been made to target host proteins for therapeutic purposes. As discussed above, the pharmaceutical modulation of ACE2 expression or inhibition of its interaction with SARS-CoV-2 spike protein for COVID-19 therapy is a matter of current investigation at different parts of the world (Kai and Kai, 2020). In these efforts, animal models will be instrumental in checking potential drug candidates’ efficacies and safety against COVID-19. Although the Syrian hamster is a clinically relevant model for multiple infectious diseases, the unavailability of reagents like antibodies against hamster proteins and lack of publicly available gene or protein expression data for this species are the major constraints to using these models up to their full capacity (Suresh et al., 2019). Before utilizing hamster as a model to understand the role of ACE2 in the pathogenesis of SARS-CoV-2 infection and/or to evaluate the efficacy of ACE2-targeted drugs, knowledge about the basal level of ACE2 expression in different tissues of hamster is essential. In the current study, we have checked the expression pattern of ACE2 in different tissues of normal Syrian hamsters through immunoblotting, immunohistochemistry, and immunofluorescence staining techniques.

**MATERIAL AND METHODS:**

**Isolation of Hamster Tissue Samples**

The tissue samples used for initial antibody standardization are from archived samples collected during our previous studies (Suklabaidya et al., 2016; Suresh et al., 2019). To analyze ACE2 expression in hamsters of different age groups and sexes, a separate Institutional Animal Ethical Committee (Institute of Life Sciences, Bhubaneswar, India) approval was obtained before conducting the study (Project no.: ILS/IAEC-195-AH/Jul-20). All the methods associated with animal studies were performed according to the Committee for the Purpose of Control and Supervision of Experiments on Animal (CPCSEA), India guidelines. Three age groups of animals comprising of young (~2–4 months old), adult (~6–8 months old), and old (~15–17 months old) were included in this study. For each age group, organs from six different animals (three males and three females) were harvested and preserved for further processing and analysis.

**Western Blot Analysis**

Using an electric homogenizer, tissues were lyzed in ice-cold RIPA buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM Na2 EDTA, 1 mM EGTA, 1% NP-40, 1% sodium de-oxo-cholate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate, 1 mM Na3VO4) supplemented with a protease inhibitor.
cocktail (MP Biomedicals) and soluble proteins were collected. Protein concentrations were measured by Bradford assay (Sigma). 20 μg of protein was loaded for each sample and electrophoresed through 8% SDS-polyacrylamide gels. Proteins were transferred to polyvinylidene difluoride membrane (Millipore) and blocked with 5% bovine serum albumin. Membranes were probed with ACE2 (#MA5-32307; Invitrogen; 1:3000 or #AF933; R&D Systems: 1 μg/ml) or β-actin (#A2066; Sigma-Aldrich; 1:1,000) primary antibody and horseradish peroxidase-conjugated secondary antibody. Antibody binding was detected with electrochemiluminescence substrate (#12757P; CST) and chemiluminescence visualized with ChemiDoc™ MP Gel Imaging System (BioRad).

Immunohistochemistry
All the tissue samples were processed and sectioned as reported earlier (Suklabaidya et al., 2016; Suresh et al., 2019). Paraffin-embedded sections were de-paraffinized using xylene, rehydrated in graded ethanol, and deionized water. Sections were subjected to antigen retrieval treatment by boiling in acidic pH citrate buffer (Vector Laboratories) for 20 min in a steam cooker. 3% hydrogen peroxide in methanol was used to block the endogenous peroxidase for 20 min and washed with 1X PBS two times, followed by blocking with horse serum (Vector Lab) for 30 min at room temperature. Sections were treated with ACE2 antibody (#MA5-32307, 1:200 or #AF933; 2 μg/ml) overnight in a humidified chamber at 4°C. Sections were washed twice with 1X PBS for 5 min each. In MA5-32307 antibody case, slides were treated with horse anti-rabbit/mouse IgG biotinylated universal antibody (Vector Laboratories) for 45 min at room temperature and with ABC reagent for 30 min. For AF933 antibody, the slides were incubated with Goat IgG VisUCyte HRP Polymer (#VC004, R&D Systems) and incubated for 45 min at room temperature (without ABC incubation). To develop the stain, 3, 3′-diaminobenzidine (DAB; Vector Laboratories) was used as a substrate according to the manufacturer’s instructions, and hematoxylin was used as a counter-stain. Sections were dehydrated with ethanol, cleared with xylene, and mounted with Vecta mount permanent mounting medium. Sections were observed under the microscope (Leica ICC500), and images were captured at ×40 magnification.

Immunofluorescence
Paraffin sections were subjected to de-paraffinization, rehydration, and antigen retrieval treatment. Sections were blocked with horse serum (Vector Lab) for 30 min at room temperature and probed with ACE2 primary antibody (#MA5-32307, 1:200 or #AF933; 2 μg/ml) overnight in a humidified chamber at 4°C. Sections were washed twice with PBST and treated with anti-Rabbit Alexa Fluor 594 (#A-11037; Life technologies, 1:500) or anti-Goat Alexa Fluor 594 (#A-11080; Life technologies, 1:500) or anti-Goat Alexa Fluor 594 (#A-11080; Life technologies, 1:500) or anti-Rabbit Alexa Fluor 594 (#A-11037; Life technologies, 1:500) at room temperature for 45 min under dark conditions. After washing, slides were mounted with SlowFade Gold Antifade mountant with DAPI (#S36938; Life Technologies) and visualized using Leica TCS SP8STED confocal microscope.

Antigen Preadsorption Test
Antibodies (#MA5-32307 or #AF933) were pre-incubated with human recombinant ACE2 protein (#933-ZN; R&D systems) at 1:5 weight to weight ratio or antibodies alone in microcentrifuge tubes containing 1% BSA. Incubation was performed by keeping the tubes with slow rotations overnight at 4°C. After the incubation, antibodies were used for IHC and IF staining of selected hamster tissues.

Quantitative Real-Time PCR
Tissues from animals were harvested for RNA isolation following the Trizol method using TRizol® LS Reagent (#10296028; Thermo). Synthesis of cDNA was performed using High Capacity cDNA Reverse Transcription Kit (#4368814; Applied Biosystems), and qRT-PCR was performed using the BRYT green PCR master mix (#A6002; Promega) according to the manufacturer’s protocol. Gapdh or β-actin was used for normalization. The expression status of Ace2 was analyzed using the following specific primers: forward: 5′- GAGAG GCTGTCAGGTTGTC -3′ and reverse 5′- TGCCAACCCTA CAATTCCC -3′.

RESULTS AND DISCUSSIONS
This study’s major objective was to check the status of ACE2 protein expression in hamster tissues of different age groups. Traditionally, immunohistochemistry or immunocytochemistry are widely used techniques to localize specific proteins or epitopes in cells and tissues. The authenticity of interpretations based on these techniques requires multiple controls in an application- and context-dependent manner. The International Working Group for Antibody Validation (IWGAV) has suggested five different pillars for antibody validation. A minimum of one pillar should be used to validate an antibody’s suitability for a particular application (Uhlen et al., 2016). To get a convincing and reliable interpretation of our study, we have adopted multiple controls, which includes correlation between Ace2 mRNA and protein expression in tissues (orthogonal strategy), use of two different antibodies and comparison of staining pattern (independent antibody strategy), reactivity toward purified human recombinant ACE2 protein (rHu-ACE2), and antigen preadsorption test. We have also conducted both IHC and IF staining for most of the organs.

In this study, the ACE2 protein expression was evaluated by two different antibodies. The ACE2 recombinant rabbit monoclonal antibody (Invitrogen; clone SN0754; Cat No MA5-32307) used in this study was generated by using synthetic peptide within human ACE2 aa 200–230 as an immunogen. As per the information available by different companies, this clone (SN0754) has reactivity against human, mouse, rat, and hamster (Invitrogen: MA5-32307 and Novus Biologicals: NBP2-67692). The polyclonal goat IgG (R&D Systems; Cat No AF933) was generated by using mouse myeloma cell line N50-derived human recombinant ACE-2 (Gln18-Ser740; Accession# Q9BYF1). The information
available with the product indicates its reactivity against human, mouse, rat, and hamster proteins. There are 26 species of hamsters in the world. In the recent past Syrian golden hamster (*Mesocricetus auratus*) has been demonstrated as a clinically relevant model for SARS-CoV-2 infection (Chan et al., 2020a; Sia et al., 2020; Chan et al., 2020b). Hence, to

**FIGURE 1** Validation of MA5-32307 and AF933 antibodies’ reactivity and specificity toward hamster ACE2. (A) Immunoblot analysis showing the expression status of ACE2 with two different antibodies (MA5-32307 and AF933) in multiple organ tissues of the Syrian golden hamster. β-actin was used as an internal control. (B) Immunoblots showing ACE2 expression in hamster kidney, human colorectal adenocarcinoma cells (Caco2), and recombinant human ACE-2 protein (rHu-ACE2) with two different antibodies. (C) Graph showing Ace2 relative mRNA expression in hamster kidney and spleen tissues. (D) Micrographs showing immunofluorescence and immunohistochemical staining of hamster kidney with antibody alone or antibody preincubated with rHu-ACE2 protein (in IHC, scale bar = 25 µm).
check whether these antibodies have reactivity against the Syrian golden hamster ACE2, we initially did an immunoblot analysis of proteins isolated from these animals’ different tissues. The antibodies showed clear reactivity with a protein of molecular weight of ∼120 kDa, which matches with ACE2 (Figure 1A). The protein lysates used in this experiment were harvested from an adult male hamster. Although similar conditions were adapted for both the antibodies, AF933 showed more nonspecific bands than MA5-32307, which could be due to the polyclonal nature of AF933. To further confirm these antibodies’ reactivity against hamster and human ACE2, we carried out immunoblot analysis of samples, including hamster kidney lysates, Caco2 human colon cancer cells lysate (known to express ACE2) (Wang et al., 2020), and rHu-ACE2. The data shown in Figure 1B confirms the reactivity of both the antibodies against human and hamster ACE2.

The immunoblot analysis results showed a high amount of ACE2 expression in hamster kidney tissues and almost no expression in spleen tissues (Figure 1A). We adapted an orthogonal strategy to further validate the antibodies by measuring mRNA expression in these two tissues. The qPCR analysis results showed a high level of Ace2 expression in kidney tissues and minimal expression in the spleen (Figure 1C), matching with the immunoblot results. Together, these findings validate the reactivity of both the antibodies against hamster ACE2.

In experimental conditions, proteins are generally in the native form in the absence of denaturing agents; however, they are partly or fully denatured during western blotting or immunohistochemistry. Detection of human recombinant ACE2 protein by both the antibodies in the western blotting indicates the reactivity of these antibodies against the corresponding epitopes present in ACE2 proteins. On the other hand, successful neutralization/blocking of AF933 by preadsorption with rHu-ACE2 (Cat No 933-ZN; actual immunogen used to generate this antibody), but failure with MA5-32307 in IHC and IF staining (Figure 1D and Supplementary Figure S1A) indicates inaccessibility of MA5-32307-specific epitope in native rHu-ACE2 protein. During the antibody and immunogen incubation/preadsorption step, no denaturing agents were present; hence, it is highly expected that rHu-ACE2 was in its native form at this step. Although the peptide used for the generation of MA5-32307 Ab (ACE2 aa 200–230) is present in the rHu-ACE2 protein, its possible non-surface localization might have led to unsuccessful blocking of antigen-binding sites of the antibody. At the same time, a similar type of cell and region-specific staining by both the antibodies in kidney tissues (Figure 1D) indicates that tissue fixation with formalin and further downstream processing might have exposed the MA5-32307-reactive epitopes present on the target protein (ACE2). At present, ACE2 aa 200–230 peptides’ unavailability has restricted us from doing antigen preadsorption with the actual immunogen for the MA5-32307 antibody. However, a comparable pattern of reactivity of both the antibodies in detecting ACE2 protein by western blotting, IHC and IF staining strongly validate the reactivity of both the antibodies against hamster ACE2 (Figures 1A,B,D). Moreover, positive staining with MA5-32307 Ab in organs known to express ACE2 (e.g., kidney or ileum) and absence in organs known to have no or minimal expression (e.g., spleen or caecum) of other species has further validated its specificity toward hamster ACE2 (Figure 1D and Supplementary Figure S1B) (Hashimoto et al., 2012).

In our study, irrespective of antibodies, age, and sex of animals, kidney tissues showed a very high ACE2 expression (Figures 1A,B,D). It’s expression was mostly at the apical surface of proximal tubules, whereas glomeruli were negative (Figure 1D). In addition to the expected membrane staining, AF933 Ab showed some level of nuclear and cytoplasmic staining in IHC, which did not get abolished after antibody preadsorption with the immunogen (Figure 1D). Hence, the nuclear staining seen with AF933 in certain tissues, including kidney, might be due to its nonspecific reactivity toward some unknown nuclear protein(s). So far, most of the literature and publicly available protein expression database have clearly shown high expression of ACE2 protein in human kidney tissues (Monteil et al., 2020; Zhang and Zhang, 2020). High expression of ACE2 in the kidney is believed to contribute to SARS-CoV-2 virus pathogenesis and disease severity (Zhang and Zhang, 2020). Detection of kidney injuries in tissues of COVID-19 patients’ post-mortem tissues further supports the importance of considering kidney function-related issues for COVID-19 treatment and management (Diao et al., 2020). Detection of the SARS-CoV-2 viral genome in certain patients’ urine samples has also been reported (Ling et al., 2020). Using a kidney organoid model, Monteil V et al. have demonstrated that the proximal tubules express ACE2 and SARS-CoV-2 replicated in these organoids (Monteil et al., 2020). Despite this clinical and experimental evidence from human patients or tissues, so far, none of the SARS-CoV-2–related studies in hamsters have reported any kidney-related histopathological changes (Chan et al., 2020a; Chan et al., 2020b; Sia et al., 2020). In the future, further investigation and analysis are required to confirm whether hamster kidney epithelial cells are permissible for this viral infection. The tropism of the virus to different organs depends on multiple factors, including the organ-specific microenvironment. Simultaneously, the infectivity of isolates from different parts of the world is yet to be experimentally compared. In the future, studies with different isolates and/or hamsters with different comorbidity conditions might help to decipher the role of kidney cells in SARS-CoV-2 pathogenesis.

In human patients, the lung-associated pathology is the predominant feature of SARS-CoV-2 infection (Jain, 2020). Certain earlier studies have shown the expression of ACE2 transcripts or protein by lung epithelial cells (Lukassen et al., 2020; Hamming et al., 2004; Sungnak et al., 2020; Zhang et al., 2020). Hence, just after the reports that ACE2 binds with SARS-CoV-2 spike (S) protein, the research and clinical communities assumed a high level of ACE2 expression in the lung or other respiratory tract parts might be a major driving factor in the pathogenesis of this respiratory virus. Our initial immunoblot analysis data showed a very trace amount of ACE2 expression in lung tissue lysate (Figure 1A). To get an idea about the spatial and cell-type distribution of ACE2 expressing cells in the lungs and trachea, further IHC analysis was conducted. Interestingly, with
both the antibodies (MA5-32307 and AF933), we did not find any visible positive staining in the epithelial cells of the lung bronchioles, and alveoli (Figure 2A). The pattern was consistent irrespective of age and sex of animals (data not shown). Endothelial cells and smooth muscle cells associated with the wall of blood vessels were also negative for ACE2 staining (data not shown). Corroborating our IHC data, IF staining of lung tissue samples did not show any positive staining compared with corresponding without antibody stained tissue sections or antibody and immunogen preadsorption controls (Figure 2A). In certain instances, we noticed some mild staining of bronchial epithelial cells with AF933 Ab; however, a similar pattern was also noticed after antibody preadsorption with the immunogen (Figure 2A). We believe that the trace amount of lung-associated
ACE2 detected in immunoblot analysis (Figure 1A) might have come from some non-epithelial cells with low abundance and scattered distribution in the lung parenchyma, whose presence might not be obvious in stained tissue sections. Our findings corroborate multiple earlier reports, including information available at the human protein atlas (https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue/lung), which have shown no or minimal expression of ACE2 in human lung tissues (Hikmet et al., 2020; Hikmet et al., 2020; Fu et al., 2020; Aguiar et al., 2020). Tracheal epithelial cells showed no positive staining with MA5-32307 antibody. However, AF933 showed strong staining of tracheal epithelial cells, which did not get neutralized after preadsorption of the antibody with the corresponding immunogen. Based on these data, we concluded that tracheal epithelial cells do not have ACE2 expression (Figure 2B). We conducted a qPCR analysis to get an idea about Ace2 mRNA expression level in hamster lung tissues of different age groups. The results shown in Figure 3B clearly show a very minimal expression...
of Ace2 in lung tissues of all the animals. We did not find any obvious difference in the expression level of Ace2 in males of adult and young hamsters (Figure 3B). However, the average expression level of Ace2 in old age males’ lung tissues was comparatively higher than adult (3.51 fold) or young (3.71 fold) males. Simultaneously, the Ace2 expression level in old age males’ was higher (3.17 fold) than corresponding females (Figure 3B). Old age and male sex have been identified as independent risk factors for poor prognosis of COVID-19 patients (Grasselli et al., 2020; Williamson et al., 2020; Zhou et al., 2020). Although through our IHC or IF analysis, we failed to detect any expression of ACE2 protein in lung tissues, but very low level of ACE2 expression in lung tissues cannot be totally ruled out. Together, based on our preliminary findings and available reports, we believe that SARS-CoV-2 related lung pathology might be independent or minimally dependent on ACE2 expression status in the lungs, which warrants further investigation. Recent studies have reported the presence of SARS-CoV-2 viral protein in respiratory tract epithelial cells and lungs of infected hamsters.

FIGURE 4 | ACE2 expression pattern in different parts of the hamster gastrointestinal tract and tongue tissues. Immunohistochemical analysis showing representative images of ACE2 expression at different parts of the hamster gastrointestinal tract and tongue tissues. The arrow symbol indicates positively stained cells in different tissues (scale bar = 50 µm).
(Sia et al., 2020), hence our findings suggest the possible involvement of some other proteins than ACE2 in the entry of SARS-CoV-2 virus into respiratory and lungs cells (Jeffers et al., 2004; To and Lo, 2004).

Out of other vital organs in the brain, mostly the cerebral cortex neurons were positively stained for ACE2 expression by MA5-32307 (Figure 3A). This finding corroborates information available at the human protein atlas (https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue). Expression of ACE2 in hamster brain neuronal cells might help investigate the possible neurological tissue damage due to SARS-CoV-2 infection (Baig et al., 2020). In liver tissues, mostly the sinusoidal endothelial cells stained positive for ACE2, but hepatocytes were negatively stained (Figure 3A and Supplementary Figure S1A). Our immunoblot analysis data also showed ACE2 expression in hamster liver tissues (Figure 1A). The sinusoidal endothelial expression of ACE2 in the hamster does not match the expression of ACE2 pattern reported for human liver (Hamming et al., 2004), and warrants further investigation to understand these contradictory findings. Our analysis, did not notice any positive staining in the bile duct and gall bladder epithelial cells (data not shown).

Like ACE2 expression in kidney tissues, we consistently observed a high level of ACE2 expression at different parts of the hamster gut (Figure 4). Prior studies have reported high expression of ACE2 in human gut tissues (Hamming et al., 2004; Hikmet et al., 2020). Single-cell transcriptomic analysis of gut tissues has identified the expression of ACE2 in upper epithelial and gland cells of the esophagus and absorptive enterocytes of ileum and colon (Zhang et al., 2020). Our IHC data shows the expression of ACE2 in the surface epithelial cells of hamster esophagus, duodenum, ileum but no expression in the large intestine (caecum, colon, and rectum) (Figure 4). Although AF933 Ab showed low intense staining of esophagus and duodenum tissues compared to MA5-32307, both the antibodies showed very high surface expression of ACE2 in surface epithelial cells of the ileum. ACE2 expression in enterocytes indicates the possibilities for these cells being infected by SARS-CoV-2 through intestinal routes. Our study and other parts of the world have clearly shown shedding of viral genome in COVID-19 patients stool samples (Senapati et al., 2020b). Importantly, studies have also reported the isolation of infectious SARS-CoV-2 virus from the stool samples (Amirian, 2020). In the recent past, some elegant studies have clearly shown infection of gut epithelial cells by SARS-CoV-2 (Lamers et al., 2020; Xiao et al., 2020). Diarrhea is a common finding in multiple COVID-19 patients across the world (Liang et al., 2020), and in certain COVID-19 patients, gut epithelial cells damage has also been reported. Hamster model with SARS-CoV-2 infection also showed intestinal epithelial damage and expression of viral protein in enterocytes (Chan et al., 2020a). These data suggest that ACE2 expressed by gut epithelial cells might have a role in the pathogenesis of SARS-CoV-2 infection. Considering these points, the hamster might be a suitable model to investigate the intestinal pathogenesis of SARS-CoV-2 infection and evaluate different therapeutics that target ACE2.

The human oral mucosal cavity expresses ACE2, and specifically, this is highly enriched in tongue epithelial cells (Xu et al., 2020). Our IHC study’s data also shows the expression of ACE2 in both the dorsal and ventral stratified squamous epithelium of the hamster tongue. Interestingly, the ventral side epithelial cells have a very high ACE2 expression level than the dorsal side (Figure 4). The absence of ACE2 expression in our immunoblot analysis (Figure 1A) could be due to less proportion of cellular proteins contribution into the total tissue lysate (also a possible reason for a low level of β-actin detection). With both the antibodies, we did not observe any positive staining in hamster spleen tissues (data not shown).

Together, our study has provided a comprehensive idea about ACE2 expression patterns in the hamster’s different tissues. We believe that this information will be instrumental in the optimal use of the Syrian golden hamster as a model for SARS-CoV-2 infection. At the same time, whether the expression of ACE2 in hamsters depends on other pathophysiological or diseased conditions warrants further investigation. In this context, a recent study has demonstrated that ACE2 is an interferon-stimulating gene (ISG), hence at the time of infection or tissue injury, interferon might upregulate ACE2 expression in different organs of hamster (Ziegler et al., 2020). In the future, attempts should also be taken to understand the possible effect of tissue-specific post-translational modification of ACE2 on antibody reactivity in hamster tissues.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Institutional Animal Ethical Committee (IAEC), Institute of Life Sciences, Bhubaneswar, India.

AUTHOR CONTRIBUTIONS

SS conceived the idea, helped in manuscript writing and supervised the study. VS, DP, AM, MS and BSS executed the experiments and helped in manuscript writing.

ACKNOWLEDGMENTS

The authors are thankful to the Director, ILS, Bhubaneswar, for his support. VS, DP, and AM are recipients of Council of Scientific and Industrial Research (CSIR) students’ research fellowship, Government of India. MS is recipient of University Grants Commission (UGC) fellowship, Government of India. We sincerely acknowledge the technical supports given by Madan Mohan Mallick and Sushanta Kumar Swain, ILS.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2020.579330/full#supplementary-material.
REFERENCES


Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Suresh, Parida, Minz, Sethi, Sahoo and Senapati. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
PAN-INDIA 1000 SARS-CoV-2 RNA Genome Sequencing Reveals Important Insights into the Outbreak


1 National Institute of Biomedical Genomics (NIBMG), Kalyani
2 Institute of Life Science (ILS), Bhubaneswar
3 Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad
4 National Centre for Biological Sciences (NCBS), Bengaluru
5 Institute for Stem Cell Science and Regenerative Medicine (InStem), Bengaluru
6 National Centre for Cell Science (NCCS), Pune
7 ICMR-Regional Centre for Medical Research, Bhubaneswar

* Joint First Authors

*Corresponding Authors
Saumitra Das: sdas@nibmg.ac.in
Ajay Parida: drajayparida@gmail.com
Murali D. Bashyam: bashyam@cdfd.org.in
Dasaradhi Palakodeti: dasaradhip@instem.res.in
Aswin Sai Narain Seshasayee: aswinsainarain@gmail.com
Yogesh Sauche: yogesh@nccs.res.in
Arindam Maitra: am1@nibmg.ac.in

Key words: SARS-CoV-2 genome, Mutation analysis PAN India, Phylodynamics, Haplotype diversities
Abstract

The PAN-INDIA 1000 SARS-CoV-2 RNA Genome Sequencing Consortium has achieved its initial goal of completing the sequencing of 1000 SARS-CoV-2 genomes from nasopharyngeal and oropharyngeal swabs collected from individuals testing positive for COVID-19 by Real Time PCR. The samples were collected across 10 states covering different zones within India. Given the importance of this information for public health response initiatives investigating transmission of COVID-19, the sequence data is being released in GISAID database. This information will improve our understanding on how the virus is spreading, ultimately helping to interrupt the transmission chains, prevent new cases of infection, and provide impetus to research on intervention measures. This will also provide us with information on evolution of the virus, genetic predisposition (if any) and adaptation to human hosts.

One thousand and fifty two sequences were used for phylodynamic, temporal and geographic mutation patterns and haplotype network analyses. Initial results indicate that multiple lineages of SARS-CoV-2 are circulating in India, probably introduced by travel from Europe, USA and East Asia. A2a (20A/B/C) was found to be predominant, along with few parental haplotypes 19A/B. In particular, there is a predominance of the D614G mutation, which is found to be emerging in almost all regions of the country. Additionally, mutations in important regions of the viral genome with significant geographical clustering have also been observed. The temporal haplotype diversities landscape in each region appears to be similar pan India, with haplotype diversities peaking between March-May, while by June A2a (20A/B/C) emerged as the predominant one. Within haplotypes, different states appear to have different proportions. Temporal and geographic patterns in the sequences obtained reveal interesting clustering of mutations. Some mutations are present at particularly high frequencies in one state as compared to others. The negative estimate Tajimas D (D = -2.26817) is consistent with the rapid expansion of SARS-CoV-2 population in India. Detailed mutational analysis across India to understand the gradual emergence of mutants at different regions of the country and its possible implication will help in better disease management.

Background

The ongoing pandemic of Severe Acute Respiratory Syndrome (SARS-CoV-2) has emerged as a global health problem reaching dimensions unparalleled in the history of humankind. Since its emergence in Wuhan, China in December 2019, it has spread rapidly to the rest of the world, with 17,354,751 people infected (as on August 1, 2020) and 674,291 deaths (1). Unlike SARS-CoV and MERS-CoV, the SARS-CoV-2 virus is characterised by an efficient person to person transmission, mostly via droplets, resulting in a basic reproductive number ($R_0$) of 2.79 (1.5 and 6.68) (2) compared to $R_0$ of 2.0–3.0, 0.9 and 1.5 for the SARS-CoV and the 1918 influenza pandemic, MERS-CoV, and the 2009 influenza pandemic respectively (3). Beginning from the first reported case in India in January 2020, the number of infections has risen to 1,695,988 including 36,511 deaths as on date (1, 4). Until date, 1,98,21,831 individuals in India have been tested for the infection (5).

Whole-genome sequencing of pathogens, especially viruses, is a powerful tool to generate rapid information on outbreaks, resulting in effective understanding of the introduction of the infection,
dynamics of transmission, contact tracing networks and impact of informed outbreak control decisions (6-8). This will also provide us with information on evolution of the virus, genetic predisposition (if any) and adaptation to human hosts. In earlier outbreaks of the West African Ebola virus infection, rapid whole genome sequencing and analysis coupled with epidemiological information has been used effectively for public health decision making. In the ongoing pandemic, large scale whole genome sequencing initiatives in other countries are being used to establish the origins of the spreading infections as well as to determine how the virus has mutated during its transmission (9). Viral genome sequencing and epidemiological data have also been successfully used to estimate the impact of informed public health decisions on the spread of the virus and contact tracing of cluster cases (10). Hence, large scale generation of whole-genome sequence data on SARS-CoV-2 from clinical samples collected from multiple geographic locations and at different time points of the outbreak and rapid dissemination of the same in public databases like the Global Initiative on Sharing All Influenza Data (GISAID) (11), is expected to provide valuable information.

Since the initial sequence data from the first reported cases, there have been attempts by various medical and scientific research organisations to generate whole-genome sequence data from infected cases in India (12-15). Understanding the necessity of generation of large scale viral whole-genome sequences representing geographic locations all over the country and at different time points of the outbreak, a PAN-INDIA 1000 SARS-CoV-2 RNA Genome Sequencing Consortium of scientific research organisations and their collaborating clinical partners was formed under the aegis of Department of Biotechnology, Ministry of Science and Technology, Government of India (Figure 1). The Consortium members collected 1,062 clinical samples in the form of nasopharyngeal and oropharyngeal swabs from individuals testing positive for COVID-19 as per the guidelines of the Indian Council of Medical Research (ICMR), isolated and sequenced viral whole genomes. Collection and sequencing of additional samples is ongoing. The data is being jointly analysed along with available epidemiological information on these samples, and all efforts are being undertaken by the Consortium members to disseminate the sequence and related information in GISAID in real-time. It is expected that this large scale initiative will facilitate informed public health decisions to control the outbreak and provide major thrust in the development of intervention measures.

**Results**

**Phylodynamic Analysis**

Phylodynamic analysis using 1052 SARS-CoV-2 RNA sequences of DBT-Pan-India-Consortium was performed through Nextstrain/ncov pipeline, and time-tree was constructed after default quality control (Figure 2). 962 sequences out of 1052 sequences passed the stringent QC criteria in the Nextstrain/ncov pipeline. Sequences that are distributed over different regions of India (Northern India = 104, Western India = 201, Eastern India = 385, Southern India = 272) were classified into 5 different haplotypes, namely 19A (8%), 19B (5%), 20A (38%), 20B (48%) and 20C (0.5%) as described in Figure 2. In previously reported studies, 20A, 20B and 20C were reported as A2a haplotype. Further, 770 additional Indian SARS-CoV-2 sequences from GISAID (11) outside of DBT-Pan-India-Consortium were added to our dataset, and a phyldynamic time-tree was generated (Figure 3) which showed a similar pattern in viral haplotype distribution.
SARS-CoV-2 RNA sequences generated by the Consortium were distributed in different states: 1) Northern India (Delhi, Uttar Pradesh, Haryana, Uttarakhand), 2) Eastern India (Odisha, West Bengal), 3) Western India (Maharashtra) and 4) Southern India (Telangana, Karnataka) as described in Figure 4.

We have observed differential distributions of viral haplotypes in different parts of India as depicted by Figure 5 (state-wise) and Figure 6 (region wise). Viral clade 20A and 20B (previously known as A2a) dominated in all geographic regions of India. In Northern and Eastern India, the frequency of 20A (51.9% and 55.6%) was highest whereas 20B was found to be in higher frequencies in Western (70.6%) and Southern India (77.6%) compared to the rest. The frequency of ancestral 19A and 19B was higher in Northern (12.5%) and Eastern India (10.6%) as compared to Western (5%) and Southern India (5.1%). We have observed the highest frequency of 19B (13.2%) in Eastern India. We studied the differences in the haplotype distributions at pan India and regional levels over time. We find that although multiple haplotypes were introduced at the beginning of the outbreak (March-May 2020) in all regions, the predominant A2a haplotypes (20A, B and C) have overtaken others in June 2020.

Further, we have hypothesised that SARS-CoV-2 in India might be introduced from multiple foreign countries. To test our hypothesis, SARS-CoV-2 RNA sequences all over India were aligned to a subset of globally available sequences (~8000 random sequences representative of all months (Dec - July), all global region and all country locations) (Figure 7). Indian sequences that were classified into 19A and 19B haplotypes are majorly clustered with sequences from other Asian countries (Figure 7). In contrast, 20A, 20B and 20C sequences are clustered with European and American countries (Figure 7).

In order to estimate possible SARS-CoV-2 introduction from foreign countries, we have utilised the country confidence information from phylodynamic timetree. Globally all sequences belonging to a particular clade were further divided into subgroups based on amino acid mutations (to estimate subgroup, the amino acid mutation is considered only if it is present in at least 10 samples for a particular clade). The country confidence information was extracted from the most recent common ancestors of each subgroup.

We have estimated SARS-CoV-2 introduction from multiple regions in India described in Figure 8. Further, we have visually validated the country confidence information from Nextstrain/auspice results for each geographic region of India (Figure 9, Figure 10, Figure 11 and Figure 12). Introduction of SARS-CoV-2 belonging to haplotype 19A and 19B mostly came from China and 20A, 20B and 20C from the United Kingdom, Italy and Saudi Arabia (Figure 8). We have noticed multiple introductions of the same haplotype in some geographic regions. Although we have found the same clade in different geographical regions of India, not all of them were introduced by the same foreign country. 20A haplotype of SARS-CoV-2 viruses came from Italy and Saudi Arabia in all Indian regions, but in the case of Eastern Indian, 20A came additionally from the United Kingdom and Switzerland. Similarly, 20B was introduced mostly from the United Kingdom in all regions, additionally in Western India from Brazil and in Southern India introduced from Italy and Greece. 19A was introduced from China in all regions. In contrast, 19B was introduced from Oman in North India and from China and Saudi Arabia in Eastern India.

**Mutation Analysis**

To understand temporal and geographic patterns in the development of various genetic variants across India, we used a set of ~1,700 genomes combining those sequences by this consortium and others.
available in GISAID. We first defined sequence types based on the nucleotide sequence of 10 sites as defined by Guan et al. (16), and numbered according to the date of their first emergence globally, as described earlier for the CoVa pipeline (17). The most predominant sequence type among Indian samples was ST4 (Figure 13) which is characterized by two non-synonymous and 1 synonymous mutations (S D614G + RdRp P323L + nsp3 F106F). It represented 65 % of all samples. The second most predominant type was ST1 same as the reference type, and which accounted for 22 % of the dataset, followed by ST2 which is characterized by two mutations (ORF8 L84S + nsp4 S76S). The relative abundance of these types across states has changed over time (Figure 14). In most states, ST4 has increased at the expense of ST1, except in Delhi, where the distribution has shifted in the opposite direction.

The predominant type can also be observed by the frequency of its mutations. S D614G + RdRp P323L has attained frequencies above 75% in most states over time (Figure 15). In Delhi, however, the current majority is represented by 3 mutations (RdRp A97V + nsp3 T1198K + N P13L) which, considering the relative abundance of sequence types, have appeared not on the background of ST4 but ST1. Accordingly, this triad has decreased in frequency over time across other states. A sub-type seems to have attained high frequencies in multiple states; this one is characterized by two consecutive point mutations (N RG203KR) and appears to have originated on the background of type ST4.

Some mutations, other than the ones mentioned above, are present at particularly high frequencies in one state as compared to others. For example, ORF3a Q57H in Gujarat. One particular mutation that has gradually increased in frequency over time starting as a minority in a specific state is nsp3 A994D in Maharashtra, where it has reached a frequency above 50% while in other states, it is present at best at a frequency of 20%. The sharp peak in its curve towards the end corresponds to 86 sequences collected from Aurangabad in a span of 4 days. We note that such temporal and spatial sampling bias could affect other frequencies as well. However, the plot (Figure 15) is informative of temporal bias wherever the rise or decline of a mutation appears abrupt and is supported by a few data points, such bias is a likely explanation.

As seen above, frequency trajectories of many mutations appear correlated and indicate the presence of a few prevailing sub-types of the virus in the Indian population. Specifically, based on the average time-series correlation across states, two major clusters of mutations could be identified (Figure 16A). The two characteristic mutations of ST4 are strongly anti-correlated with a group of 4 mutations (nsp3 T1198K, RdRp A97V, N P13L, nsp6 L37F). On the other hand, they are positively correlated to a sub-cluster of 3 mutations (nsp3 A994D, N RG203KR). Some mutations might appear linked because of the artifacts of time-series correlation. If two mutations are not strongly monotonic in either direction, then they would appear correlated. Therefore, to directly assess the degree of co-occurrence of mutations, we performed clustering on the patterns of presence/absence of pairs of mutations across 1630 genomes. The results were largely similar (Figure 16B). However, the two pair of mutations, (ORF3a Q57H + NS194L) and (ORF8 L84S + N S202N) were separated into their own clusters.

Haplotype network analysis

From a total of 1034 SARS-CoV-2 genome sequences we selected 717 high quality sequences representing the pathogen diversity in nine Indian states i.e. Delhi (n=8), Haryana (n=31), Karnataka (n=118), Maharashtra (n=115), Odisha (n=173), Telangana (n=24), Uttar Pradesh (n = 4), Uttarakhand (n=63) and West Bengal (n =181) were used for haplotype network re construction (Figure 17). From the
haplotype network, we observed Maharashtra, Karnataka created three distinct haplotype nodes and sequences from Odisha, West Bengal and Uttarakhand sparse in different haplotype nodes. We also observed a haplotype node with the majority of the genomes from West Bengal, Odisha and a small percentage of the samples belonging to Uttarakhand. Geographically Odisha and West Bengal share borders, and the shared SARS-CoV-2 haplotypes might be because of the high interstate travelling. On the bottom right-side of the network we see a group of samples from Maharashtra, Delhi, Haryana and Uttarakhand grouped together with 2-4 single nucleotide variants (SNVs), suggesting the infection might have spread in a short duration of time. On the left of the network, there is a portion of the samples from Haryana and Karnataka sharing the same parent haplotype, representing possible transmission by migration. The negative estimate Tajima’s D (D = -2.26817, AMOVA p(D<=-2.26817) = 0.00281) (18) is consistent with the rapid expansion of SARS-CoV-2 population in India.

Discussion

Our analysis shows the presence of multiple lineages of SARS-CoV-2 in India in different geographic regions. 20A and 20B together (belonging to the former haplotype of A2a) are the predominant haplotypes at pan India level and in each region. Interestingly, the haplotypes were differentially distributed in different regions. While the 20A were most abundant in Northern and Eastern India, 20B was found to be the most abundant haplotype in Western and Southern India. The ancestral haplotypes of 19A and 19B were mostly found in Northern and Eastern India, with 19B being the most abundant in the latter region. Our data shows interesting temporal changes in haplotype diversity of SARS-CoV-2 in India. While multiple haplotypes were introduced during the early part of the outbreak in March-May 2020, the 20A, B and C haplotypes (A2a) became the predominant haplotypes in all regions by June 2020. This is consistent with the analysis of global data (12) and the reported enhanced efficiency of transmission ability of the A2a haplotype and its association with high viral load (13, 14).

Analysis of probable country of origin of these SARS-CoV-2 sequences in India revealed that they had been probably introduced by travel from multiple countries across the globe. 20A, B and C haplotypes were introduced from multiple countries in Europe and also American continents. Interestingly, 20A alone is predicted to have been introduced by travel from Italy, Saudi Arabia, United Kingdom and Switzerland. Similarly, 20B was introduced from the United Kingdom, Brazil, Italy and Greece. In contrast, 19A was introduced from China alone while 19B was introduced by travel from China, Oman and Saudi Arabia.

The most predominant SARS-CoV-2 lineage across India is the one with D614G variant of the Spike protein. Across most states, it's relative abundance has increased over time to 70% and above, except in Delhi, where an earlier lineage, ST1 (known as type L in GISAID) is the predominant type present at a similar proportion. We found three characteristic mutations for this lineage of type ST1 (nsp3 T1198K, RdRp A97V, N P13L). The frequency time-series of these two lineages across states was strongly anticorrelated. Based on the co-occurrence patterns and frequency dynamics, we identified a sublineage of D614G, characterized by two adjacent mutations in the nucleocapsid protein (N RG203KR), in many states. In Maharashtra, it was the predominant lineage and was strongly correlated to another mutation nsp3 A994D. This particular mutation had attained frequencies above 50% in Maharashtra while its end point frequency elsewhere was no greater than 20%. Further work would be required to understand the effect, if any, of these mutations on the infectivity or virulence of the virus.
An interesting set of mutations which has enriched during the course of the outbreak in India has been the emergence of the RG203KR in the N gene. Since its detection in few sequences obtained from individuals in April 2020 (15), this mutation has been found to gain abundance, especially in the West, South and North region but not in the East in our sequence data. Further analysis and studies are required to be undertaken to understand the implication of this observation, since this mutation is expected to alter potential miRNA binding sites and cause changes in the structure of the nucleocapsid.

Our analysis indicates that the haplotype diversities pan India and in each region continued to increase until May 2020, subsequent to which it reduced drastically with the emergence of the A2a haplotypes which has overtaken other lineages by June 2020. Further data is required to understand whether this observation might also reflect no new introductions of the virus in the country subsequent to May 2020 since India implemented a national lockdown between April to May. Such interpretations might enable improved understanding of such informed public health decisions. In recent times the number of COVID-19 occurrences in India has increased drastically. Although most of the states have their own strategic lockdown devised to control the outbreak, it will be more efficient if we can incorporate the geographical transmission pattern information in the planning of such strategies. In the current haplotype network, we have tried to explore the transmission of the infection among different states of India. It is necessary to incorporate more genomic datasets to draw a clearer picture.

**Methods**

**Viral Genome Sequencing and analysis:**

RNA isolated from nasopharyngeal and oropharyngeal swabs was used to prepare genome sequencing library using the following kits as per manufacturer’s instructions; (i) TruSeq Stranded Total RNA Library Preparation Kit (Illumina Inc, USA) for shotgun metagenomics RNA sequencing, or (ii) QIAseq SARS-CoV-2 Primer Panel (Qiagen GmbH, Germany) for amplified of viral genome sequencing or (iii) Maxima H Minus Double-Stranded cDNA Synthesis Kit (ThermoFisher Scientific, USA), Nextera Flex Enrichment Kit with Respiratory Virus Oligo Panel (Illumina Inc, USA) for viral RNA capture and sequencing. All the sequencing libraries were checked using high sensitivity D1000 ScreenTape in 2200 TapeStation system (Agilent Technologies, USA) and quantified by Real-Time PCR using Library Quantitation Kit (Kapa Biosystems, USA). Next-Generation Sequencing were carried out using MiSeq Reagent Kit v3 or MiSeq Reagent Kit v2 Micro or Misex reagent Kit v2 Nano in Miseq system (Illumina Inc, USA) or using NovaSeq 6000 SP Reagent Kit (Illumina Inc, USA). All of these libraries were prepared for 2x100 bp sequencing reads or 2x150 bp sequencing reads. For shotgun RNA sequencing data and captured viral RNA sequencing data, sequencing reads were mapped to reference viral genome sequence and consensus sequence for each sample was built using Dragen RNA pathogen detection software (version 9) in BaseSpace (Illumina Inc, USA). For amplified whole-genome sequencing, the viral sequences were assembled, and a consensus sequence for each sample was generated using CLC Genomics Workbench v20.0.3 (Qiagen GmbH, Germany). In both cases, the SARS-CoV-2 isolate Wuhan-Hu-1 (Accession NC_045512.2) was used as the reference genome.

**Phylodynamic Analysis:**
SARS-CoV-2 RNA sequences (N=1052) generated by DBT-all-India-Consortium representing major geographic regions in India were combined. To analyze the data, community standard Nextstrain/ncov pipeline was used. Nextstrain/ncov pipeline (19) includes quality control, alignment, phylogenetic inference, temporal dating of ancestral nodes and discrete trait geographic reconstruction which utilises MAFFT (20), augur and auspice tools.

Prior to analysis, preliminary quality control as sequence duplicate removal from multifasta and fixing of metadata information was performed. Sequences that did not meet minimum length criteria, maximum ‘N’ content were removed. Over a period of time, Nextstrain curates extremely divergent samples in their GitHub repository, which were also removed by the pipeline from the dataset before alignment. 1010 sequences that passed QC criteria were finally aligned using MAFFT. Post alignment 46 sequences were removed, which includes clustered mutations and divergent samples. The timescale and branch lengths of phylogenetic tree were estimated using the IQ-TREE method (21) considering Wuhan/WHO1/2019 as ancestral. Viral clades could be assigned to 962 sequences. Further, frequency-trajectories of mutations, genotypes and haplotypes (clades) were estimated by Augur. Finally, the phylogenetic timetree was constructed by auspice tool.

Transmission profiles of SARS-CoV-2 sequences in India

In order to get an estimation of viral introduction from a foreign country to a geographical region in India, we prepared a global subset of 9201 sequences (Including all DBT-all-India-Consortium sequences). Around eight thousand foreign sequences were randomly subsampled from GISAID using Nextstrain subsample module, which represented all months (Dec - July), all global region and all country locations. A maximum of 75 sequences per month per country were taken during subsampling. Phylodynamic analysis with Indian SARS-CoV-2 sequences with global subset was performed as was described in the previous section.

Sequences assigned to a particular clade were further divided into subgroups based on amino acid mutations. Only those amino acid mutations that present in at least 10 samples in a particular clade were considered while performing subgrouping. The most recent common ancestors (MRCA) for each subgroup were determined using the ‘Phylo’ module of Biopython (22) from the phylogenetic tree. The country confidence information of MRCAs for each subgroup was extracted and curated from auspice timetree using python scripts. Further, we have visually validated the introduction analysis results from the auspice generated web-portal.

Mutation Analysis

1. Data description

1586 SARS-CoV-2 genome assemblies were available from India on GISAID as of July 18, 2020. Additionally, 631 assemblies, which were not yet available publicly, were directly provided by NIBMG and NCCS. A total of 1058 genomes were sequenced by the consortium till that date. Out of these 2217 genomes, 1928 had information on the collection date as well as on the state location. 8 Indian states (Karnataka, Maharashtra, Telangana, Odisha, West Bengal, Uttarakhand, Gujarat and Delhi) had more than 50 genomes each, and a total of 1782 genomes were selected from these states. The consortium sequenced sufficiently large samples from each of these states except Delhi and Gujarat. Sequences with more than 10% ambiguous characters were discarded leaving 1696 genomes. CoVa was employed...
to perform variant analysis on this set of genomes. After generating a Multiple Sequence Alignment (MSA) limited to the sites in the reference genome (NC_045512.2), duplicate genomes were excluded, and further analysis was performed on a final set of 1630 genomes. In this final set, 698 genomes were from the DBT consortium, and the rest were from the other institutes in India.

2. Types distribution

The SARS-CoV-2 genomes selected above were classified into distinct sequence types using CoVa. This sequence typing was based on the nucleotide sequence extracted from 10 positions, proposed to be used for barcoding in Guan et al. (16). Sequence types were labeled in the order of their first appearance based on a global dataset of ~17000 sequences collected from GISAID, as of May 21, 2020. Based on this dataset, 17 sequence types were identified and incorporated in the sequence typing program of CoVa. The reference genome belonged to type ST1.

3. Allele frequency dynamics

Allele (mutation) frequency time-series were built for 8 states using collection dates metadata for the selected genomes. The fraction of genomes with a certain mutation of all the genomes that were collected till a certain day were plotted against the number of days passed from the first collection. Only the mutations which were present in at least 1-3rd of the genomes at least once in any state were included in the plot. All the data points from across states were plotted on the same temporal scale i.e. the day 0 marked March 01, 2020, the first day of collection in the entire dataset. Further, the earlier data points with less than 20 genomes were excluded from the plot as their frequency estimates would be highly unreliable.

The error in the raw frequency resulting from a small sample size was estimated under the assumption that the raw counts were binomially distributed. For a binomial distribution, the maximum likelihood estimate (MLE) is the same as the raw frequency. Therefore, considering raw frequencies as MLE, the likelihood ratio test can be employed to generate confidence intervals. Briefly, For a significance threshold of $\alpha = 0.05$, a $100(1 - \alpha)$ confidence interval can be generated by moving the parameter, binomial probability $p$ in this case, away from the MLE to the extent that the null hypothesis of MLE being the true $p$ still holds.

4. Binary Clustering

The mutations selected above were clustered based on their presence/absence across 1630 genomes. The binary distance was used to measure the extent of dissimilarity in the occurrence patterns of two mutations. This is the proportion of genomes in which the only one was present out of all the genomes in which either or both of the mutations were present. The mutations were clustered using these pairwise distances in a hierarchical clustering algorithm of complete linkage.

5. Time-series correlation

Mutations were also clustered based on their frequency dynamics. The correlation between any two time-series was estimated using the Spearman’s method, which doesn’t require the relationship between two variables to be linear but only monotonic. The correlation estimates were averaged over the 8 states, and the pairwise correlation matrix was clustered using the same approach as above.

Haplotype network analysis
For haplotype network, 815 sequences were selected out of the 1034 genome sequences that were generated as a part of the DBT’s PAN-India 1000 SARS-CoV2 RNA genome sequencing consortium. The genome sequences were aligned to WH01 reference genome using MAFFT (20) with Nextstrain (19) Augur wrapper script. Then the aligned fasta file was converted to Phylip format using a custom Biopython (22) script for haplotype network reconstruction. Using POPART (v 1.7) (23) 717 SARS-CoV-2 genomes were selected with less than 5% undefined states and a TCS Network was created. The network was coloured using the respective residence place (State) of the patient.

†PAN-INDIA 1000 SARS-CoV-2 RNA Genome Sequencing Consortium Team


Centre for DNA Fingerprinting and Diagnostics, Hyderabad: Ashwin Dalal, Murali Bashyam, Pratyusha Bala, Vinay Donipadi, Divya Vashisht, Debashis Mitra


National Centre for Cell Science, Pune: Dhiraj Paul, Kunal Jani, Janesh Kumar, Radha Chauhan, Vasudevan Seshadri, Girdhari Lal, Arvind Sahu, Yogesh S Shouche, Manoj Kumar Bhat

ICMR-National Institute of Cholera and Enteric Diseases, Kolkata: Shanta Dutta, Mamta Chawla Sarkar, Ananya Chatterjee, Hasina Banu, Agniva Majumdar

Institute of Post Graduate Medical Education and Research, Kolkata: Monimoy Banerjee, Raja Ray, Jayeeta Halder, Aritra Biswas

Translational Health Science and Technology Institute, Faridabad: Guruprasad Medigesi, Gagandeep Kang, Sharanabasava Patil, Anbalagan Ananthraj, Madhu Pareek, Imran Khan, ESIC Hospital and Medical College, Faridabad, Gurugram Civil Hospital, Gurugram and Palwal Civil Hospital, Palwal


Government Medical College, Aurangabad: Jyoti Iravan, Dhaaval Khatri, Maitrik Dave

All India Institute of Medical Sciences, Rishikesh: Ravi Kant, Deepjyoti Kalita, Amit Mangla

Mahatma Gandhi Institute of Medical Sciences, Wardha: Vijayshri Deotale, Rahul Narang, Deepashri Maraskolhe
Nizam’s Institute of Medical Sciences, Hyderabad: K Manohar, Madhumohan Rao, Vijay Dharma Teja

Maulana Azad Medical College, Delhi: Sonal Saxena, Vikas Manchanda, Oves Siddiqui

Regional Medical Research Center, Bhubaneswar: Sanghamitra Pari, Jyotirmayee Turuk

Armed Forces Medical College, Pune: Sourav Sen, Santosh Karade, Kavita Bala Anand, Shelinder Pal Singh Shergill, Rajiv Mohan Gupta

Byramjee Jeejeebhoy Government Medical College, Pune: Rajesh Karyakarte, Suvarna Joshi, Murlidhar Tambe

Acknowledgments

Department of Biotechnology, Ministry of Science and Technology, Govt. of India, DBT Wellcome Trust India Alliance Intermediate Fellowship IA/I/16/2/502711 awarded to A.S.N.S., J.C. Fellowship of Department of Science and Technology to SD, Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.04-0800, National Genomics Core at NIBMG and CDFD. NCCS thanks Dr. T. P. Lahane, Director, Department of Medical Education and Research for the permission to use the Maharashtra clinical specimens. The Consortium team would like to thank Dr. Renu Swarup, the Secretary, Department of Biotechnology, Govt. of India, for her enthusiastic support and critical suggestions for this initiative of DBT-AIs.

References


Figure 01. Regions of India covered by the PAN-INDIA 1000 SARS-CoV-2 RNA Genome Sequencing Consortium.
Figure 02. Radial phylogenetic timetree based on 962 SARS-CoV-2 RNA sequences. The concentric circles represent the sample collection date; the earlier date of sample collection is located close to the centre.
Figure 03. Radial phylogenetic timetree based on 1732 India-specific SARS-CoV-2 RNA sequences collected from this study and other studies.
Figure 04. Distributions SARS-CoV-2 RNA sequences belonging to different states within India.
Figure 05. State-wise viral haplotype distributions.
Figure 06. Region-wise viral haplotype distributions

20A = A2a, 20B = A2a, 20C = A2a
19A = O, A2, A3
19B = B1, B4
Figure 07. Phylodynamic timetree of Indian SARS-CoV-2 sequences with global subset consists of Asian, European, African, Oceania and American countries.
Figure 08. Introduction of SARS-CoV-2 haplotypes by foreign countries in different Indian regions. Foreign countries from which introductions were estimated are colour coded.
**Figure 09.** Northern Indian sequences in Timetree and estimated introduction by foreign countries.

**Figure 10.** Eastern Indian sequences in Timetree and estimated introduction by foreign countries.
Figure 11. Western Indian sequences in Timetree and estimated introduction by foreign countries.

Figure 12. Southern Indian sequences in Timetree and estimated introduction by foreign countries.
Figure 13. Phylogeny of 1630 SARS-CoV-2 genomes from India, depicting clustering and relative abundance of CoVa sequence types.
Figure 14. Changes in the relative abundance of sequence types across 8 Indian states.
Figure 15. Frequency dynamics of major variants across 8 Indian states. Major variants were defined as those that were present in 1-3rd of the genomes in any state at least at one point of time.
Figure 16. Clustering of mutations, A) Based on the frequency time-series correlation. B) Based on the pattern of co-occurrence across genomes.
Figure 17: Haplotype network of 717 SARS-CoV-2 genomes (Delhi (n=8), Haryana (n=31), Karnataka (n=118), Maharashtra (n=115), Odisha (n=173), Telangana (n=24), Uttar Pradesh (n = 4), Uttarakhand (n=63), West Bengal (n =181)) from DBT’s PAN-India 1000 SARS-CoV2 RNA genome sequencing consortium. The circle size is proportional to the number of samples belonging to the haplogroup.
Large-scale Testing for SARS-CoV-2 using Tapestry Pooling

Anirudh Chakravarthy†1,2, Srikar Krishna†1,2, Sumana Ghosh1, Ajay Tomar1,3, Sriram Varahan1, Ajit Rajwade5, Sabyasachi Ghosh5, Nimay Gupta5, Rishi Agarwal5, Himanshu Payal5, Prantik Chakraborty5, Krishna Vishal Vemula5, Akanksha Vyas5, Ritesh Goru5, Sandeep Krishna*,4, Dasaradhi Palakodeti*,1, Manoj Gopalkrishnan*,5

1 Institute for Stem Cell Science and Regenerative Medicine (inStem), Bangalore, India; 2 SASTRA University, Thanjavur, India; 3 The University of Trans-Disciplinary Health Sciences and Technology; 4 Simons Centre for the Study of Living Machines, National Center for Biological Sciences TIFR, Bangalore, India; 5 Indian Institute of Technology Bombay, Mumbai, India.

†Co-first authors; *Corresponding authors

Email: manoj.gopalkrishnan@gmail.com, dasaradhip@instem.res.in, sandeep@ncbs.res.in

Abstract

We have previously described Tapestry Pooling, a scheme to enhance the capacity of RT-qPCR testing, and provided experimental evidence with spiked synthetic RNA to show that it can help to scale testing and restart the economy. Here we report on validation studies with Covid19 patient samples for the Tapestry Pooling scheme with prevalence in the range of 1% to 2%. We pooled RNA extracted from patient samples that were previously tested for Covid19, sending each sample to three pools. Following three different pooling schemes, we pipetted 320 samples into 48 pools with pool size of 20 at prevalence rate of 1.6%, 500 samples into 60 pools with pool size of 25 at prevalence rate of 2%, and 961 samples into 93 pools with pool size of 31 at prevalence rate of 1%. Of the 191 RT-qPCR experiments that we performed, only one pool was incorrect (false negative). Our recovery algorithm correctly called results for the individual samples, with a 100% sensitivity and a 99.9% specificity, with only one false positive across all the 1,781 blinded results required to be called. We show up to 10X savings in number of tests required at a range of prevalence rates and pool sizes. These experiments establish that Tapestry Pooling is robust enough to handle the diversity of sample constitutions and viral loads seen in real-world samples.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Introduction

The SARS-CoV-2 virus is now active in all populated parts of the world, with the cumulative number of infections globally being estimated over 33 million as of September 2020. While the virus was initially met with large-scale lockdowns, these have largely been lifted and replaced by smaller-scale lockdowns around known clusters of infection. Several establishments such as schools, offices, airports, hospitals, shopping malls remain restricted or working at less than full capacity. Communities are forced to trade off economic distress against morbidity and mortality. Such decisions would be eased with the availability of an accurate and affordable method for large-scale and relatively rapid testing of individuals for SARS-CoV-2.

Pooling samples enhances the capacity of qRT-PCR assays to meet the growing demand for testing. If a pool tests negative, all samples in that pool are declared negative. When the prevalence of infection is low, many pools turn out to be negative. Each test gives results for multiple samples, allowing available resources to be deployed to perform more tests. One such two-round pooling strategy that has been used in India as well as in Israel, Germany, China, and the USA has been to take pools of 5 samples each. If a pool tests negative by qRT-PCR, all samples participating in that pool are declared negative. For each positive pool, every sample within that pool is individually re-tested to determine if it is positive or negative for the virus. We term this “simple pooling”.

We have previously proposed an algorithmic method of pooling, known as Tapestry Pooling, to achieve substantial reductions of the number of tests required per sample, compared to simple pooling (1, 2). Tapestry Pooling identifies positives in a single round of testing. It can give savings at prevalence as high as 20%. At very low prevalence, it can deliver more than 10 results per test. These advantages make Tapestry Pooling an attractive alternative to simple pooling. In Tapestry Pooling, each sample goes to three pools while ensuring that any two samples share at most one pool. Our novel reconstruction algorithm recovers the presence or absence as well as amount of the virus in each individual sample from a single round of qRT-PCR results of the pools. We have combined ideas from the fields of Combinatorial Group Testing, Compressed Sensing and Sparse Regression to devise the algorithm (1).

We previously validated Tapestry Pooling in experiments where fake positive samples were created by spiking with synthetic template RNA (2). We have also argued in Ghosh et al that Tapestry Pooling has various advantages over other algorithmic pooling schemes (3). In this paper, we describe experiments on Covid19 patient samples and show that Tapestry Pooling works well for population sizes ranging from 300 to 1000 when prevalence rates are in the range of 1-2% and pool sizes range from 20 to 31. Elsewhere we will report the validation of Tapestry Pooling for scenarios where the prevalence rate is as high as 15-20% (manuscript under preparation).
Methods

Study Design

We performed three case studies. Each case study was implemented as follows:

1. We chose the number of individuals and the prevalence rate to be tested.
2. Given these numbers and without knowing precisely which samples will be positive or negative, we chose a pooling scheme that determines which samples participate in which pools. The pooling schemes can be found in Supplementary Table S1.
3. The experimental team chose which samples were negative and which were positive, in accordance with the prevalence rate chosen above. Positive samples were chosen randomly from among previously-tested positive patient samples (Supplementary Table S1), so that the variation in their viral loads was indicative of natural variation in viral loads for patient samples arriving at a testing lab.
4. The experimental team then pooled pre-extracted RNA from the samples according to the pooling matrix provided, and conducted qRT-PCR tests on the pools. In each pooling scheme, every sample participated in exactly 3 pools, while the pool sizes ranged from 20 to 31.
5. The Ct values obtained from the qRT-PCR runs were reported back to an analysis team. The Ct values for the three pooling schemes tested are shown in supplementary table S2a, S2b and S2c. Which individual samples were positive or negative was blinded from the analysis team.
6. The analysis team applied the reconstruction algorithm to the qRT-PCR results. The algorithm classified each individual sample into one of three categories: positive, negative and undetermined.
7. This prediction was tallied with the ground truth, namely which samples were actually positive or negative, to compute the sensitivity and specificity of the method.

The following cases were examined:

1. 320 samples, 48 pools, 5 positives (1.6% prevalence), pool size 20.
2. 500 samples, 60 pools, 10 positives (2% prevalence), pool size 25.
3. 961 samples, 93 pools, 10 positives (1% prevalence), pool size 31.

The pooling schemes were chosen in order to explore population sizes ranging from several hundred to around a thousand keeping prevalence rates in the 1-2% range while allowing all pools and controls to fit within a 96-well plate for a single round of qRT-PCR.

Sample handling and RNA isolation:
All pooling experiments done for this study were performed using RNA isolated from patient nasopharyngeal swab samples. The handling of these samples was done as prescribed by the Indian Council of Medical Research. RNA isolation was done using the Qiagen QIAamp Viral RNA Mini kit. The isolated RNA was tested independently by qRT-PCR in addition to being used for this experimental pooling study.

**Pooling strategy:**

RNA samples that had been isolated previously from clinical swab samples and stored at -80°C were used for the pooling experiments.

48 pools for 320 samples: 315 negative samples and 5 positive samples were chosen for pooling. Each pool contained 20 = 320*3/48 RNA samples. Since this was the first study we were doing with a large pool size, we were concerned about possible over-dilution of positive RNA samples. So we made provision to reduce dilution by pooling 1µL of each negative sample and 2.5µL of each positive sample. Each pool contained between 20 to 23 µL of pooled RNA that was used for subsequent qRT-PCR testing.

60 pools for 500 samples: 490 negative samples and 10 positive samples were selected. Each pool contained 25 = 500*3/60 individual samples. From our previous experiment, we discovered that the dilution of RNA samples due to pooling did not drastically increase the Ct values of positive pools in our qRT-PCR runs. Therefore, we added 1µL of each RNA sample (positive and negative samples) into each of its designated pools. Each pool contained 25µL of pooled RNA that was used for subsequent qRT-PCR testing.

93 pools for 961 samples: Each pool contained a total of 31 = 961*3/93 samples. We chose 10 positive samples and 547 negatives. We identified the pools into which the 10 positive samples were added. These amounted to a total of 23 pools. We prepared these pools by pooling the participating positive samples along with several other negative samples per pool. Pooling was done by adding 1uL of each RNA sample (positive and negative samples) resulting in a total of 31µL of RNA per pool that was used for subsequent qRT-PCR testing. The remaining 70 reactions which contained only negative samples and no positives were simulated using 70 unique negative samples as the template.

**Testing using qRT-PCR:**

Once the pooling of RNA samples was completed, aliquots from the pools were subjected to qRT-PCR testing for Covid-19. The pools were tested using the MyLab qRT-PCR kit (PathoDetect Covid-19). This kit is a multiplex kit capable of detecting the SARS-Cov2 RdRp and E genes. In order to reduce the dilution factor of the positive RNA samples, we used double the volume of template RNA and half the volume of nuclease free water prescribed by the MyLab kit protocol. All other components of the reaction mixture (primers, probes and RT-PCR master mix) were added exactly as prescribed by the MyLab kit instructions with appropriate control reactions. The reaction compositions for each pooling scheme are shown in the tables below.
The thermocycler was set up as recommended in the kit instructions. Results from the qRT-PCR machine were recorded and submitted to the Tapestry recovery algorithm for identification of Covid19 positive samples.

### Reagents

<table>
<thead>
<tr>
<th>Reagents</th>
<th>48 pools for 320 samples</th>
<th>60 pools for 500 samples</th>
<th>93 pools for 961 samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>162</td>
<td>198</td>
<td>303</td>
</tr>
<tr>
<td>Detection mix (RDRP gene)</td>
<td>54</td>
<td>66</td>
<td>101</td>
</tr>
<tr>
<td>Detection mix (E gene)</td>
<td>54</td>
<td>66</td>
<td>101</td>
</tr>
<tr>
<td>PCR master mix</td>
<td>270</td>
<td>330</td>
<td>505</td>
</tr>
<tr>
<td>Total volume</td>
<td>540</td>
<td>660</td>
<td>1010</td>
</tr>
<tr>
<td>Template from pooled RNA</td>
<td>10 µL per reaction</td>
<td>10 µL per reaction</td>
<td>10 µL per reaction</td>
</tr>
</tbody>
</table>

**Table 1: Reaction composition for the pooling schemes used.** Total volume per reaction: 20 µL

**Performance Statistics**
The Tapestry Pooling reconstruction algorithm classifies samples into three categories: positive, negative, and undetermined. The provision for undetermined samples ensures that even when the number of positives in a batch exceeds the capacity of the pooling scheme, we do not compromise on sensitivity or specificity. Therefore the algorithm’s performance must be judged using sensitivity and specificity, as well as the number of undetermined samples it returns. We calculated sensitivity = $\frac{TP}{TP+FN}$ and specificity = $\frac{TN}{TN+FP}$, where TP = number of predicted positives that were true positives, FP = number of predicted positives that were actually negative, TN = number of predicted negatives that were true negatives, and FN = number of predicted negatives that were actually positive.

**Results**

Tapestry Pooling was found to work in all three studies with sensitivity of 100% and specificity of 99.9% and only 13 undetermined from 1781 samples tested as shown in Table 2. All pools with positive samples were positive except for one pool in the third study. Note that the results of the algorithm were robust to this false negative. The raw data for the pooling schemes used, the measured Ct values (Supplementary table S2a, S2b and S2c), and the ground truth and predicted positive samples is included in the supplementary information (Supplementary table S3a, S3b and S3c).

<table>
<thead>
<tr>
<th>#Samples</th>
<th>#Positives</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>#Tests for Tapestry Pooling</th>
<th>#Tests in Simple Pooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>320</td>
<td>5</td>
<td>100%</td>
<td>100%</td>
<td>48 + 0</td>
<td>64 + 25</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
<td>100%</td>
<td>99.8% (1 false positive)</td>
<td>60 + 10</td>
<td>100 + 50</td>
</tr>
<tr>
<td>961</td>
<td>10</td>
<td>100%</td>
<td>100%</td>
<td>93 + 3</td>
<td>192 + 50</td>
</tr>
</tbody>
</table>

Table 2: High Specificity and Sensitivity of Tapestry Pooling. Tapestry pooling requires substantially fewer tests than simple pooling while preserving high sensitivity and specificity. #Tests represents the number of first round tests plus the number of second round tests. For Tapestry Pooling, this is the number of undetermined samples.

**Discussion**

The Tapestry Pooling method is made available as a web app at www.tapestry-pooling.com, making it easy to deploy. The user is assisted to pick a particular matrix size based on their requirements of number of samples to be tested, testing capacity available, prevalence rate expected, and any constraints of pool size for their assay. After pooling and completing the assay,
the user enters the Ct value of each pool obtained from the qRT-PCR into the web app. The recovery algorithm solves and returns a result for each sample.

We have demonstrated that Tapestry Pooling works with high sensitivity and specificity as a method for testing large numbers of individuals at low prevalence for SARS-CoV-2. It allows a wide range of pooling schemes to be deployed so that it can be customized over a large range of prevalence rates, number of samples, pool sizes, and number of tests. We have shown that pools of size upto 31 can be managed without false negatives by adjusting the reaction composition, specifically by adding a larger volume of template and less water. This means for example that a positive sample in a pool of size 31 is diluted only 15.5-fold so that the Ct shift is kept within detectable limits. In summary, Tapestry Pooling is a flexible pooling scheme that returns large savings while still maintaining high accuracy.

Acknowledgements

Sandeep Krishna thanks the Simons Foundation for funding as well as the Department of Atomic Energy, Government of India under project no. 12-R\&D-TFR-5.04-0800. AR, MG thank SERB MATRICS grants MTR/2019/0006, EMR/2017/004089, MTR/2018/000817, IITB WRCB grant 10013976, and DST-Rakshak grant 10013980. DP acknowledges support by CSR funding from Azim Premji philanthropy, Standard Chartered Bank, and Punjab National Bank,

References

**Supplementary Data**

**Supplementary Table S1: Pooling schemes used**
The 3 pooling schemes tested are provided as Supplementary Excel sheets. In each sheet, the columns represent the pools to be created and list the sample IDs that are grouped together. The positive samples selected for each pooling scheme are marked in red in their respective pools.

**Supplementary Table S2: qRT-PCR results**
The results of the qRT-PCR runs are shown in Table S2 for each pooled test for each of the 3 pooling schemes we studied. The tests for which a cycle threshold (Ct) value was detectable are as shown in the table. The tests not included in the table were undetected in the qRT-PCR reaction. The table shows results for both the RdRp and E genes. The former Ct values were what were fed into the reconstruction algorithm to obtain the results shown in the main text.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>RdRp gene</th>
<th>E gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ct values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>B1 B3 C3 C11 D7 E5 E7 F5 G3 G5 G11 H1 H3 H7 NTC</td>
<td></td>
</tr>
<tr>
<td>Ct values</td>
<td>32.84 23.69 30.35 26.72 23.55 26.25 23.43 27.08 33.21 32.88 30.35 27.15 30.23 30.18 NA</td>
<td></td>
</tr>
</tbody>
</table>

| Test        | B1 B3 C3 C11 D7 E5 E7 F5 G3 G5 G11 H1 H3 H7 NTC |
| Ct values   | 32.45 22.73 29.54 26.50 22.86 26.55 22.66 26.04 32.31 31.92 29.25 29.15 29.27 29.53 NA |

**Table S2a:** Cycle threshold values obtained for detectable tests amongst the 48 pooled tests performed with 320 samples.

<table>
<thead>
<tr>
<th>Target gene</th>
<th>RdRp gene</th>
<th>E gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ct values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>A1 A2 A3 A6 A7 A8 A12 B1 B7 B9 B10 C1 C2 C4</td>
<td></td>
</tr>
<tr>
<td>Ct values</td>
<td>31.52 29.31 31.25 28.95 25.94 23.26 26.69 41.12 25.59 28.75 36.22 37.14 34.02 23.04</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>C5 C8 C11 D1 D3 D4 D5 D6 D8 D10 E3 E9 E10 NTC</td>
<td></td>
</tr>
<tr>
<td>Ct values</td>
<td>33.91 33.10 34.35 22.96 25.28 33.28 35.33 33.53 25.55 30.59 34.32 34.98 NA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target gene</th>
<th>RdRp gene</th>
<th>E gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ct values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>A1 A2 A3 A6 A7 A8 A12 B1 B7 B9 B10 C1 C2 C4</td>
<td></td>
</tr>
<tr>
<td>Ct values</td>
<td>30.83 28.21 30.34 27.98 24.50 22.09 24.79 32.78 24.26 27.56 31.66 31.07 31.40 21.69</td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td>C5 C8 C11 D1 D3 D4 D5 D6 D8 D10 E3 E9 E10 NTC</td>
<td></td>
</tr>
</tbody>
</table>

**Table S2b:** Cycle threshold values obtained for detectable tests amongst the 60 pooled tests performed with 500 samples.
<table>
<thead>
<tr>
<th>Target gene</th>
<th>RdRp gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>A12</td>
</tr>
<tr>
<td>Ct values</td>
<td>22.34</td>
</tr>
<tr>
<td>Test</td>
<td>E10</td>
</tr>
<tr>
<td>Ct values</td>
<td>28.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target gene</th>
<th>E gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test</td>
<td>A12</td>
</tr>
<tr>
<td>Ct values</td>
<td>23.90</td>
</tr>
<tr>
<td>Test</td>
<td>E10</td>
</tr>
<tr>
<td>Ct values</td>
<td>30.29</td>
</tr>
</tbody>
</table>

**Table S2c:** Cycle threshold values obtained for detectable tests amongst the 93 tests performed with 961 samples.

**Supplementary Table S3: Prediction results for the pooling scheme**

<table>
<thead>
<tr>
<th>Ground truth (sample IDs that were positive - 5 total)</th>
<th>46,116,167,238,314</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positives predicted from Tapestry algorithm</td>
<td>46,116,167,238,314</td>
</tr>
<tr>
<td>Number of undetermined from Tapestry algorithm</td>
<td>0</td>
</tr>
<tr>
<td>Number of false positives</td>
<td>0</td>
</tr>
<tr>
<td>Number of false negatives</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table S3a:** Prediction results for the pooling scheme consisting of 48 pools for 320 samples.

<table>
<thead>
<tr>
<th>Ground truth (sample IDs that were positive - 10 total)</th>
<th>29, 79, 141, 142, 214, 238, 333, 384, 404, 498</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positives predicted from Tapestry algorithm</td>
<td>29, 79, 141, 142, 214, 232, 238, 333, 384, 404, 498,</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Number of undetermined from</td>
<td>10</td>
</tr>
<tr>
<td>Tapestry algorithm</td>
<td></td>
</tr>
<tr>
<td>Number of false positives</td>
<td>1 (sample ID 232)</td>
</tr>
<tr>
<td>Number of false negatives</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table S3b:** Prediction results for the pooling scheme consisting of 60 pools for 500 samples.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground truth (sample IDs that</td>
<td>58, 103, 274,</td>
</tr>
<tr>
<td>were positive - 10 total)</td>
<td>397, 411, 544,</td>
</tr>
<tr>
<td></td>
<td>632, 700, 788,</td>
</tr>
<tr>
<td></td>
<td>933</td>
</tr>
<tr>
<td>Positives predicted from</td>
<td>58, 103, 274,</td>
</tr>
<tr>
<td>Tapestry algorithm</td>
<td>397, 411, 544,</td>
</tr>
<tr>
<td></td>
<td>632, 700, 788,</td>
</tr>
<tr>
<td></td>
<td>933</td>
</tr>
<tr>
<td>Number of undetermined from</td>
<td>3</td>
</tr>
<tr>
<td>Tapestry algorithm</td>
<td></td>
</tr>
<tr>
<td>Number of false positives</td>
<td>0</td>
</tr>
<tr>
<td>Number of false negatives</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table S3c:** Prediction results for the pooling scheme consisting of 93 pools for 961 samples.
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors


Abstract

Many viruses utilize the host endo-lysosomal network for infection. Tracing the endocytic itinerary of SARS-CoV-2 can provide insights into viral trafficking and aid in designing new therapeutic strategies. Here, we demonstrate that the receptor binding domain (RBD) of SARS-CoV-2 spike protein is internalized via the pH-dependent CLIC/GEEC (CG) endocytic pathway in human gastric adenocarcinoma (AGS) cells expressing undetectable levels of ACE2. Endocytic expression of ACE2 (AGS-ACE2) results in RBD traffic via both CG and clathrin-mediated endocytosis. Endosomal acidification inhibitors like BafilomycinA1 and NH4Cl, which inhibit the CG pathway, reduce the uptake of RBD and impede Spike-pseudoviral infection in both AGS and AGS-ACE2 cells. The inhibition by BafilomycinA1 was found to be distinct from Chloroquine which neither affects RBD uptake nor alters endosomal pH, yet attenuates Spike-pseudovirus entry. By screening a subset of FDA-approved inhibitors for functionality similar to BafilomycinA1, we identified Niclosamide as a SARS-CoV-2 entry inhibitor. Further validation using a clinical isolate of SARS-CoV-2 in AGS-ACE2 and Vero cells confirmed its antiviral effect. We propose that Niclosamide, and other drugs which neutralize endosomal pH as well as inhibit the endocytic uptake, could provide broader applicability in subverting infection of viruses entering host cells via a pH-dependent endocytic pathway.

Author summary

This study investigates the cellular mechanisms by which SARS-CoV-2 can gain entry into human cells. We find that the virus employs diverse endocytic processes to enter cells and the acidic environment within these endocytic compartments is essential for infection. Using these observations from first principles, we screened a small set of FDA-approved drugs which could potentially inhibit endosomal acidification and therefore prevent viral entry and infection. The routinely prescribed anti-helminthic drug, Niclosamide, was observed to have this capability. Our study proposes that drugs altering both endocytic entry as well as endosomal acidification can assist in the clinical management of viral infections.


Editor: Benhur Lee, Icahn School of Medicine at Mount Sinai, UNITED STATES

Received: May 26, 2021; Accepted: June 10, 2021; Published: July 12, 2021

Copyright: © 2021 Prabhakara et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: All relevant data are within the manuscript and its Supporting Information files.

Funding: We acknowledge NCBS-TIFR graduate fellowship (for CP and PS), UGC graduate fellowship (for RG), INSPIRE fellowship (for SG) and NCBS postdoctoral fellowship (for SJ). TsvZ acknowledges EMBO postdoctoral fellowship (ALTF 1519-2013) and NCBS Campus fellowship; AC thanks India Alliance DBT – Wellcome Trust Early career fellowship (IA/151/150235); VS and SM acknowledge the support of Department of Atomic Energy, Government of India, under Project Identification No. RTI 4006 and NCBS-TIFR-core funds; SM acknowledges J.C. Bose Fellowship from DST, Government of India, and India Alliance DBT – Wellcome Trust Margdarshi fellowship (IA/M/151/502018). CP, PS, TsvZ and SJ received support from Margdarshi fellowship grant (IA/M/151/502018). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Coronaviruses (CoVs) are a group of related enveloped RNA viruses of which some are known to cause respiratory tract infections in humans. The recent emergence of SARS-CoV-2 and its rapid spread across the world in multiple waves has posed a global health emergency [1]. Several therapeutic strategies are currently being used to alleviate the respiratory symptoms of patients infected with SARS-CoV-2 [2, 3]. However, strategies aimed to directly counter the virus have met with only limited success. A search for antivirals affecting the endocytic entry of viruses is particularly useful as infections from multiple emerging CoV-2 strains and other related viruses can be controlled through the inhibition of a common step [4].
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

Virus entry into host cells is a multistep process. A key step in successful invasion is the release of viral genomic content into the host cell cytoplasm. To achieve this, viruses bind to specific cell surface receptors and subsequently undergo endocytic fusion either directly at the plasma membrane or following endocytic uptake. While fusion directly at the plasma membrane is well established for HIV and Influenza virus infections [5,6], both alternatives of entry are feasible for CoV infections depending on the availability of receptors and proteases at the host cell surface. Different CoVs interact with a range of specific receptors for entry [7–11]. Although angiotensin converting enzyme 2 (ACE2) is a well-studied receptor for SARS-CoV-2 [12], other receptors and co-receptors are being discovered [13–17]. Additionally, CoVs require proteolytic processing of the viral spike protein by host cell proteases to gain entry [16,19]. Therefore, these viruses can directly fuse at the cell surface if the Spike protein is cleaved by a cell surface serine protease like Tmprss2 [12,20], or utilize an endo-lysosomal route for fusion, where the Spike protein is primed by cysteine protease cathepsins [12,21–23]. Viral entry and infection in different host cells is dependent on the expression of these key host factors [24]. Cell tropism studies revealed that SARS-CoV-2 infection is not only restricted to airway epithelium but the cells along the Gastrointestinal tract are also infected [25]. An understanding of the entry pathways across various host cell types is important as it allows better interpretation of cell-based drug screens and translatable of cellular models of infection.

The role of the endo-lysosomal network appears to be crucial in delivering viruses to acidic compartments. For instance, cathepsins function optimally in a low pH environment [18,26]. Inhibitors of acidification which increase the pH of endosomal compartments significantly reduce the infection of endosomal compartments. Drugs like Apilimod and YM201636 reduce CoV infections [27,31,32] by inhibiting the maturation of late endosome to lysosome without altering the endosomal pH directly. These studies emphasize the importance of an optimal endocytic network in viral entry and infection. The applicability of therapeutics that act by dissolving pH gradients across intracellular compartments remains to be explored in the clinical management of SARS-CoV-2 infection.

Multiple endocytic pathways operate at the cell surface [33,34] which can be exploited by viruses. However, the endocytic routes preferred by SARS-CoV-2 in different host cell types is largely unknown. The clathrin and dynamin independent CLIC/GEEC (CG) endocytic pathway [35] is of particular interest here as uptake through this pathway is known to be pH-dependent. Vascular ATPases (V-ATPases), which actively pump protons into the endocytic compartments [36], play a crucial role in the formation of CG endosomes as established using genetic and pharmacological perturbations [37,38]. By contrast, uptake through clathrin-mediated endocytosis (CME) remains unaltered upon V-ATPase perturbation [37]. Homotypic fusion of nascent CG endosomes (called CLICs–clathrin-independent carriers) forms highly acidic early endosomal compartments of the CG pathway (called GEECs–GPI anchored protein enriched early endosomal compartments) with an estimated luminal pH of 6.0 [39]. Thus, GEECs could provide a conducive environment for viral uncoating and membrane fusion. Interestingly, Adeno-associated virus (AAV) hijacks the CG pathway for infection [40] and SARS-CoV has also been reported to enter cells through a clathrin and dynamin independent endocytic pathway [20]. These observations prompted us to study the role of CG endocytosis in the context of SARS-CoV-2 entry and infection.

In this report, we study the endocytosis of receptor binding domain (RBD) of SARS-CoV-2 Spike protein in gastric epithelial cells (AGS) in the presence and absence of ACE2. We show that RBD is endocytosed via the CG pathway and its uptake is sensitive to pharmacological perturbations of this pathway in AGS cells which express undetectable levels of ACE2. Overexpression of ACE2 in AGS cells (AGS-ACE2) results in RBD employing both AGS and CME pathways for entry. Endosomal acidification inhibitors such as BafilomycinA1 and NH4Cl only alter the endosomal pH but also blocked RBD uptake, similar to CAG cargo uptake. Spike-proteolytic processing inhibitors impede early steps of viral entry. We find that Chloroquine which also blocks early steps of Spike-pseudovirus entry acts differently from BafilomycinA1 and NH4Cl. We also observed a positive correlation between the amount of RBD endocytosed and surface levels of ACE2 [41,42]. Interestingly, Adeno-associated virus (AAV2) hijacks the CG pathway for infection [40] and SARS-CoV has also been reported to enter cells through a clathrin and dynamin independent endocytic pathway [20]. These observations prompted us to study the role of CG endocytosis in the context of SARS-CoV-2 entry and infection.

In this report, we study the endocytosis of receptor binding domain (RBD) of SARS-CoV-2 Spike protein in gastric epithelial cells (AGS) in the presence and absence of ACE2. We show that RBD is endocytosed via the CG pathway and its uptake is sensitive to pharmacological perturbations of this pathway in AGS cells which express undetectable levels of ACE2. Overexpression of ACE2 in AGS cells (AGS-ACE2) results in RBD employing both AGS and CME pathways for entry. Endosomal acidification inhibitors such as BafilomycinA1 and NH4Cl not only alter the endosomal pH but also blocked RBD uptake, similar to CAG cargo uptake. Spike-proteolytic processing inhibitors impede early steps of viral entry. We find that Chloroquine which also blocks early steps of Spike-pseudovirus entry acts differently from BafilomycinA1 and NH4Cl. We also observed a positive correlation between the amount of RBD endocytosed and surface levels of ACE2 [41,42]. Interestingly, Adeno-associated virus (AAV2) hijacks the CG pathway for infection [40] and SARS-CoV has also been reported to enter cells through a clathrin and dynamin independent endocytic pathway [20]. These observations prompted us to study the role of CG endocytosis in the context of SARS-CoV-2 entry and infection.

In this report, we study the endocytosis of receptor binding domain (RBD) of SARS-CoV-2 Spike protein in gastric epithelial cells (AGS) in the presence and absence of ACE2. We show that RBD is endocytosed via the CG pathway and its uptake is sensitive to pharmacological perturbations of this pathway in AGS cells which express undetectable levels of ACE2. Overexpression of ACE2 in AGS cells (AGS-ACE2) results in RBD employing both AGS and CME pathways for entry. Endosomal acidification inhibitors such as BafilomycinA1 and NH4Cl not only alter the endosomal pH but also blocked RBD uptake, similar to CAG cargo uptake. Spike-proteolytic processing inhibitors impede early steps of viral entry. We find that Chloroquine which also blocks early steps of Spike-pseudovirus entry acts differently from BafilomycinA1 and NH4Cl. We also observed a positive correlation between the amount of RBD endocytosed and surface levels of ACE2 [41,42]. Interestingly, Adeno-associated virus (AAV2) hijacks the CG pathway for infection [40] and SARS-CoV has also been reported to enter cells through a clathrin and dynamin independent endocytic pathway [20]. These observations prompted us to study the role of CG endocytosis in the context of SARS-CoV-2 entry and infection.

Results

Generation of SARS-CoV-2 probe to study its endocytosis itinerary

Virus can enter cells via multiple endocytic routes (Fig 1A) [33]. A cellular system that exhibits clathrin-dependent and independent routes is required for identifying the possible pathways that the virus may utilize to enter cells. Therefore, we chose a human adenocarcinoma gastric cell line (AGS cells) [43] as a surrogate model system to study SARS-CoV-2 endocytosis and infection. To explore the trafficking itinerary, we purified the receptor binding domain (RBD) of SARS-CoV-2 Spike protein [44] and fluorescently labeled the same (S1A and S1B Figs and Materials and Methods). We tested the specificity of the labelled RBD probe in AGS cells transiently overexpressing myc-tagged ACE2 and found that more RBD was bound to cells overexpressing ACE2 (S1C Fig). We also observed a positive correlation between the amount of RBD endocytosed and surface levels of ACE2 (S1D and S1E Figs), supporting the notion that ACE2 is one of the cell surface receptors of RBD [12].

https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

Several viruses utilize the macropinocytosis pathway as an entry route into cells [50]. Since 10kDa dextran marks both CG cargo as well as larger endocytic compartments like those derived from macropinocytosis [51], we tested if macropinocytosis plays any role in RBD uptake. Macropinocytosis is dependent on amiloride-sensitive Na+/H+ exchangers [52]. Upon treatment with Amloride, we found no alteration in the uptake of RBD, dextran and transferrin (S2G and S2H Fig). Whereas macropinocytic dextran uptake stimulated by PMA (phorbol 12-myristate 13-acetate) was completely rescued upon co-treatment with Amloride (S2I Fig). This

Fig 1. RBD uptake is sensitive to CG Pathway inhibitors in AGS cells.
A: Schematic describing endocytic pathways at the plasma membrane with specific cargoes for each endocytic pathway: transferrin (CME Cargo) and 10kDa dextran (CG Cargo). AN96, ML141 and BatfA1 specifically affect the uptake of CG cargoes. B: AGS cells were pulsed with RBD, dextran and transferrin for 30 minutes and imaged at high resolution after fixation. Images are shown in B and quantification of Manders' co-occurrence coefficient is shown in C. This compares the fraction of RBD endosomal intensity with transferrin or dextran (p-value < e-06), transferrin or dextran intensity with transferrin or RBD (p-value < e-05) and dextran endosomal intensity with transferrin or RBD (p-value = 0.18). RBD is more co-localized to dextran endosomes. Number of cells = 10. White arrow represents endosomes containing RBD, dextran and transferrin. Yellow arrow represents endosomes with RBD and dextran without transferrin. Dashed white line in B represents the approximate cell boundary. D, E: AGS cells were pretreated with Control (0.6% DMSO) or AN96 25μM for 30 minutes and pulsed with RBD, dextran and transferrin for 30 minutes with or without the inhibitor. Treatment with AN96 reduces RBD (p-value < e-19) and dextran (p < e-44) uptake while minimally alters transferrin uptake (p = 0.02). Images are shown in D and quantification in E. Numbers of cells > 100 for each treatment. F, G: AGS cells were treated with Control (0.2% DMSO) or BatfA1 200nM for 30 minutes and pulsed with RBD and dextran for 30 minutes with or without the inhibitor. Treatment with BatfA1 reduces RBD (p-value < e-33) and dextran (p-value < e-18) uptake. Images are shown in F and quantification in G. Numbers of cells > 100 for each treatment. H, I: AGS cells were treated with Control (0.2% DMSO) or BatfA1 200nM for 30 minutes and pulsed with RBD and transferrin for 30 minutes with or without the inhibitor. The surface transferrin receptor (TIR) was labelled after fixation. Treatment with BatfA1 reduces RBD uptake (p-value < e-27) and increased normalized transferrin uptake (p-value < e-03). Images are shown in H and quantification in I. Numbers of cells > 80 for each treatment. Data (E, G, I) is represented as a scatter with box plot. Black dots represent per-cell data points. Box plot represents the distribution (25% to 75% percentile) with the red line indicating the median and red dot indicating the mean of the distribution. Whiskers represent distribution up to 1.5 times interquartile range and + indicates outliers beyond the whiskers. In the entire manuscript, ***, ** , * and ns indicate p-value of Wilcoxon rank-sum test < 0.001, <0.01, <0.05 and not significant, respectively. Scale bar: 20μm (B) and 40μm (D, F, H).
https://doi.org/10.1371/journal.ppat.1009706.g001

RBD is internalized via CG endocytosis and RBD uptake is sensitive to CG Pathway inhibitors

We employed the methodology of tracking RBD uptake along with cargoes specific to CME (transferrin) and CG (10kDa dextran) endocytic pathways to determine the endocytic route taken up by RBD (Fig 1A). At 10 minutes post internalization, transferrin endosomes of the CME pathway are distinct from dextran endosomes of the CG pathway [43]. At these times, internalized RBD is colocalized with endosomes containing the CG cargo but not the CME cargo (S2A and S2B Fig). At 30 minutes post internalization, as well, this segregation remains. While a small fraction of RBD endosomes were colocalized with endosomes containing both transferrin and dextran, a large fraction of RBD endosomes were localized to compartments uniquely marked by dextran (Fig 1B and 1C; compare % RBD with transferrin and dextran). This suggests that the itinerary of uptake of RBD is similar to CG cargo and different from CME cargo.

CG pathway is regulated by small GTPases—CDC42 [45], Arf1 [46] and GEF of Arf1, GBF1 [47]. Inhibitors that block the function of these regulators affect the formation of CG endosomes without altering uptake through the CME pathway. The inhibitor AN96, which is a stable analog of LG-186 [48,49], targets GBF1 and specifically affects the CG pathway (Godbole et al., Manuscript in preparation). Towards determining the trafficking route of RBD, we examined if macropinocytosis plays any role in RBD uptake. We observed that AN96 treatment reduced both RBD and dextran uptake but had minimal effects on the amount of transferrin internalized (Fig 1D and 1E). We also observed that the peri-nuclear transferrin recycling endosomal pool was redistributed throughout the cytoplasm upon treatment with AN96 without affecting the net amount of transferrin internalized. Another CG pathway inhibitor, ML141 (CDC42 inhibitor) [49], also significantly decreased both dextran as well as RBD uptake (S2E and S2F Figs).
confirms that macropinocytosis does not play a role in RBD trafficking in AGS cells. Together, the co-localization studies and pharmacological inhibition experiments strongly suggest that RBD uptake occurs via the CG pathway and is inhibited by specific blockers of the CG pathway.

**Bafilomycin A1 and NH₄Cl block RBD uptake**

Given the relevance of acidification in both formation of CG endosomes [37, 39] and viral infection [53], we focused on studying the role of acidification inhibitors on the uptake of RBD. We assessed the effect of Bafilomycin A1 (BafA1), a specific inhibitor of V-ATPase [54], on RBD, dextran and transferrin uptake in AGS cells. A strong reduction in RBD and dextran uptake and an increase in normalized transferrin uptake was observed when cells were treated with BafA1 (Fig S3A and S3B Fig). The increase in transferrin uptake could be because BafA1 also retards the transferrin recycling from the recycling endosomes [55] and thereby increasing the net amount of transferrin internalized within cells as observed. We also examined the effect of NH₄Cl, a weak base known to alter endosomal acidification [56], on the uptake of these 3 cargoes. Similar results as with BafA1 were observed (S3A and S3B Fig), confirming our earlier [37] finding that uptake via the CG pathway is pH sensitive and blocking acidification results in reduced CG uptake.

Towards understanding the mechanism of action for acidification inhibitors in bringing about these changes in trafficking, we assessed their effect on two parameters—numbers of endosomes (S3Biii Fig) and per-endosome intensity in the presence/absence of inhibitor (S3Biii Fig). We observed that both BafA1 and NH₄Cl reduced the total number of RBD and dextran endosomes without affecting the per-endosome intensity. However, while the total number of transferrin endosomes remained unchanged, the per-endosome intensity of transferrin increased with BafA1 and NH₄Cl treatment. This indicates that the reduction in RBD and dextran is likely due to a block in the entry while an increase in per-endosome transferrin intensity could be because of a block in the formation of recycling endosome carriers, as proposed earlier.

These results are not specific to AGS cells alone. HEK-293T cells, which are also permissive to Spike-pseudovirus transduction, showed similar inhibition of RBD and dextran uptake, and an increase in transferrin uptake with BafA1 (S10A–S10D Fig).

**RBD is localized to acidic compartments**

Internalized cargoes can be recycled along with the bulk membrane [57] or directed towards degradation with the fluid phase [58]. Typically, transferrin bound to its receptor marks the early sorting/recycling endosomes and lysotracker labels the acidic degradative compartments within a cell [34]. At 30 minutes of pulse with the three cargoes, while a small fraction of RBD (~36%) associated with transferrin, the majority of RBD (~84%) co-localized with dextran suggesting that RBD is directed predominantly towards the degradation route rather than the recycling route (Fig 1B and 1C). The lysotracker labelling showed highly acidic tubular compartments with significant co-localization with RBD. At 30 minutes of pulse with RBD, around 55% of RBD co-localized with lysotracker marked compartments. At longer time points (3 hours) of pulse with RBD, an even increased proportion of RBD (85%) associated with compartments marked by lysotracker, confirming that RBD is trafficked to acidic compartments (Fig 2A and 2B).

**Fig 2.** RBD trafficking with CG cargo is localized to acidic compartments and endosomal acidification inhibitors neutralize these endosomes. A, B: AGS cells were pulsed with RBD for 30 mins or 3 hours, labelled with lysotracker in the last 15 minutes of pulse and imaged live at high resolution. Images are shown in A and quantification of Manders’ co-occurrence coefficient is shown in B. RBD is colocalized with Lysotracker positive compartments. With increase in time more RBD is associated with Lysotracker (p-value < e-06) and more Lysotracker positive compartments have RBD (p-value < e-06). Each condition has >12 cells. Dashed white line in A represents approximate cell boundary. C: Schematic describing the experimental protocol for estimating the pH of endosomes by ratiometric measurements using pH-sensitive (FITC) and pH-insensitive (TMR) dextran. D–F: AGS cells were pulsed with FITC and TMR dextran for 2 hours, chased for 1 hour with BafA1 200nM/400nM, NH₄Cl 30mM or control and imaged live. Endosomal pH is increased upon addition of acidification inhibitors (p-values < e-118 for BafA1 200nM, < e-122 for BafA1 400nM, < e-223 for NH₄Cl). Images along with pH maps are shown in D (and in S4C) and quantification in E (and in S4D). Enlarged regions of pH maps indicated by white boxes are shown in F. Box plot in E represents the distribution of medians of each repeat which is denoted by red dots. Violin plot indicates all the data points from repeats. Colour bar in F corresponds to the estimated endosomal pH. Control₁ is 0.2% DMSO, Control₂ is 0.4% DMSO and Control₃ is 0% DMSO. Number of repeats ≥ 3 for each treatment and each repeat has >80 cells. Scale bar: 20μm (A) and 40μm (D). https://doi.org/10.1371/journal.ppat.1009706.g002

**Bafilomycin A1 and NH₄Cl alter the pH of acidic endosomal compartments**
We next focused on determining the change in endosomal pH brought about by various inhibitors within the acidic compartments populated by RBB. Cells were labelled with RBD-sensitive (FITC) and pH-insensitive (TMR) dextran for 2 hours and chased for 1 hour with or without inhibitors (Fig 2C and Materials and Methods). The above pulse and chase durations were chosen to allow accumulation of labelled dextran in late endosomes and lysosomal compartments. Additionally, since the acidification inhibitors also have inhibitory roles in the early steps of CG endocytosis as discussed in the previous section, to evaluate their effect on endosomal pH, cells were incubated with inhibitors only during the chase. While the ratio of the fluorescence of these probes is used to estimate endosomal pH by comparing the ratio with the calibration curve [50] (S4A and S4B Fig and Materials and Methods), quantifications of the endosomal intensities and the endosomal number of TMR dextran aids in understanding the effect of various drugs on late endosomal trafficking.

Treatment of cells with acidification inhibitors showed an increase in endosomal pH. The average pH of the late endosomes in control cells was 5.8. The pH of these compartments increased to 6.2 and 7.1 in the presence of BafA1 200nM and 400nM respectively. Incubation with NH4Cl also resulted in increasing the pH of these endosomes to 6.6 (Figs 2D and 2E and S4C). While BafA1 marginally changed the TMR intensity per endosome, NH4Cl greatly increased the TMR intensity indicating that NH4Cl also brings about the fusion of endosomes (S4D Fig). All the acidification inhibitors also reduced the numbers of endosomes (S4D Fig) and this effect was most prominent with NH4Cl wherein the endosomes were organized close to the perinuclear region (S4C Fig). The spatial pH maps show the distribution of the pH of endosomes within a cell. Cells treated with BafA1 400nM and NH4Cl showed a homogenous distribution of endosomes with increased pH similar to the respective cell averages. On the other hand, cells treated with BafA1 200nM, showed heterogeneity in endosomal pH with some endosomes depicting high pH while others were closer to the average (Fig 2F).

To assess the effect of BafA1 on the pH of early time point endosomes, AGS cells were labelled with FITC and TMR dextran for 20 minutes and chased for 10 minutes with or without BafA1 for the entire duration of pulse and chase (S4E Fig). While the total amount of dextran uptake is not affected significantly, the endosomal FITC intensity and the endosomal ratio of FITC/TMR, which can be considered as a proxy for endosomal pH, show a robust increase with BafA1 treatment (S4F Fig). This indicates that BafA1 also affects the endosomal pH of early time point endosomes.

Endosomal acidification inhibitors affect the entry of Spike-pseudovirus

The observations made using RBD internalization were extended to SARS-CoV-2 Spike-pseudotyped lentiviral particles (Spike-pseudovirus), generated using a previously established methodology [80]. Pseudoviral infection was assessed using reporter (mCherry) expression (S5A Fig and Materials and Methods). The specificity of the Spike-pseudovirus was validated by comparing infection with an alternatively pseudotyped virus (VSVG; S5DI Fig) and bald pseudoparticles (SSC Fig). Dilutions of the supernatant containing bald pseudoparticles showed no transduction, while similar dilutions of supernatant with Spike-pseudovirus showed high levels of transduction in AGS (SSC Fig). Independently, a competition experiment was conducted to assess the effect of excess soluble RBB and trimeric RBB on transduction of Spike-pseudovirus in AGS cells (Materials and Methods). The transduction efficiency was reduced in the presence of soluble RBB as well as trimeric RBB, indicating that the Spike-pseudovirus competes for the same binding sites as RBB (S5F Fig).

Since our experiments were aimed at understanding the entry mechanism of Spike-pseudovirus, we designed the transduction assays with shorter incubation times and followed the infection efficiency by tracing reporter expression at a later time point (Fig 3A and Materials and Methods). We characterized the transduction efficiency of the pseudovirus and found 0.5 MOI (4/8 hours) or 1 MOI (2 hours) of pseudovirus incubation to be optimal (S5F Fig). This design was chosen to reduce the long-term toxicity of the inhibitors to the cells and minimize any secondary effects on the translational processes of the reporter gene post entry.

Fig 3. BafA1 and NH4Cl affect Spike-pseudovirus infection.

A: Schematic describing the experimental protocol for SARS-CoV-2 Spike-pseudovirus transduction assay. AGS cells were pre-incubated with the inhibitors (BafA1 50nM, NH4Cl 20nM, CQ 50μM) for an hour and transduced with Spike-pseudovirus at MOI 0.5 or MOI 1.0 for indicated incubation times. Following this, virus and inhibitor-containing medium was removed, and cells were further incubated with inhibitor-free media. B and D show images of AGS cells expressing the reporter mCherry protein and C and E show the quantification of normalized % transduction of cells treated with inhibitors compared to respective controls. B-E: Transduction efficiency is reduced with BafA1 (p-values < e-86 for 8 hours, < e-60 for 4 hours) and NH4Cl (p-values < e-83 for 8 hours, < e-72 for 4 hours, < e-65 for 2 hours) at all time points of incubation. Control in B is 0.05%DMSO and in D is 0%DMSO. Number of repeats ≥ 2 for each treatment. F: AGS cells were treated with control or BafA1, pre/during/post incubation with virus as well as throughout all three stages. BafA1 significantly reduces transduction when used throughout (p-value < e-63) as well as during virus presentation (p-value < e-59). Only pre-treatment or post-treatment of BafA1 does not significantly change the percentage of transduction. Number of repeats = 7 for control and 3 for each condition. Data (C, E, F) is represented as percentage transduction normalized to control on left Y-axis along with percentage viability, calculated from the total number of nuclei, normalized to control on the right Y-axis. Black dots represent the mean of each repeat. Box plot represents the distribution of the means (25% to 75% percentile) with red line indicating the median of the distribution. Asterix with error bars in red represents the mean +/- SD of % viability. Black dotted line marks 100%. Scale bar: 100μm (B, D).

https://doi.org/10.1371/journal.ppat.1009706.g003

Spike-pseudovirus infection has been previously reported to be reduced upon treatment with NH4Cl, BafA1 and Chloroquine in different cell types [27,61]. Using our methodology, we tested the effect of NH4Cl and BafA1 in AGS cells and observed a significant reduction in Spike-pseudovirus infection with no difference in cell viability at all time points of viral incubation (Fig 3B–3E). HEK-293T (S10E–S10I Fig) and A549-ACE2 (S10H Fig) cells also exhibited a similar inhibition of transduction upon treatment with acidification inhibitors.

RBD uptake is reduced upon treatment with AN96 and ML141, albeit to a lesser extent compared to the effect of BafA1 and NH4Cl. Therefore, we assessed the effect of AN96 and ML141 on Spike-pseudovirus transduction in AGS cells. Different concentrations of AN-96 were tested, and we observed no reduction in transduction even at the highest concentration of 25μM (S5G Fig), without any compromise in cell viability. Similarly, treatment with ML141 did not alter the normalized percentage transduction compared to the control (SS5G Fig). Blocking of the CG pathway often results in the redistribution of CG cargo towards the CME pathway [43].

https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

Chloroquine does not affect RBD uptake, minimally alters endosomal pH but affects Spike pseudovirus infection

Chloroquine, a diprotic weak base, is expected to accumulate in acidic compartments and neutralize lysosomal pH [62]. While, mounting evidence shows that Chloroquine and its analogs can inhibit the infection by several viruses such as Ebola, Dengue, Chikungunya, HIV, etc [63], many studies point towards differences between the mode of action of Chloroquine and acidification inhibitors—BafA1 and NH4Cl [64,65]. We, therefore, tested the effect of Chloroquine on RBD, dextran and transferrin uptake to verify if it behaves like BafA1. We found that upon treatment with Chloroquine, the uptake of neither RBD nor transferrin was altered significantly (Fig 4A and 4B). Dextran uptake was marginally higher upon treatment with Chloroquine (Fig 4B).

The effect of Chloroquine in changing the endosomal pH of late endosomes was assessed. At different concentrations of Chloroquine tested, the endosomal pH measured using FITC/TMR ratio was only minimally increased (Fig 4C and 4D). We also observed that both FITC and TMR endosomal intensities increased with the concentration of Chloroquine. To confirm our results, we used another method to estimate endosomal pH. FITC has a pH-sensitive (488nm) and a pH-insensitive excitation (450nm) [56]. We used the 488/458 excitation ratio of FITC dextran as a readout of pH and found that this ratio also showed only a small albeit significant increase when compared to control cells, unlike the increase brought about by NH4Cl (p-value < e-27). Image data shown in E and quantification in F. Number of cells > 75 for each treatment. G, H: Assay as described in 3A; Transduction efficiency is reduced with CQ with 8 hours of incubation (p-value < e-90). Number of repeats = 3 each for 0% DMSO Control and CQ. Data representation in B, D is as described in Fig 2; F as described in Fig 1; H as described in Fig 3. Scale bar: 40μm (A, C, E), 100μm (G).

Effects of BafA1 on RBD uptake and Spike-pseudovirus transduction in AGS-ACE2 cells.

To ascertain the mode of RBD and Spike-pseudovirus entry into AGS cells, we next measured the levels of known modulators–BafA1 and NH4Cl [64,65]. We, therefore, tested the effect of Chloroquine on RBD, dextran and transferrin uptake to verify if it behaves like BafA1. We found that upon treatment with Chloroquine, the uptake of neither RBD nor transferrin was altered significantly (Fig 4A and 4B). Dextran uptake was marginally higher upon treatment with Chloroquine (Fig 4B).

Fig 4. Chloroquine does not affect RBD uptake, minimally alters endosomal pH but affects Spike pseudovirus infection.
A, B: AGS cells were treated with Control (0% DMSO) or CQ 10μM for 30 minutes with and pulsed with RBD, transferrin and dextran for 30 minutes with or without the inhibitor. Images shown in A and quantification in B show no change in the uptake of transferrin and RBD and marginal change in dextran uptake with CQ (p-values = 0.2 for RBD, 0.24 for Tf, 0.013 for Dex). Box plot represents the distribution of medians of each repeat which is denoted by black dots. Number of repeats = 18 and 12 for Control and CQ, respectively and each repeat has >100 cells. C, D: AGS cells were pulsed with FITC and TMR dextran for 2 hours and chased for 1 hour with Control (0.2% DMSO), 10, 50 or 100μM of CQ and imaged live. CQ minimally increased the FITC/TMR ratio (p-value < e-23, 0.38 and < e-03 for 10, 50 and 100 μM of CQ, respectively). Numbers of cells in each condition is >150 cells. E, F: AGS cells pulsed with pH-sensitive (FITC) dextran for 2 hours and chased for 1 hour with Control or 100μM CQ and imaged live. CQ increases the 488/458 ratio of FITC dextran only slightly (p < e-10) compared to the increase brought about by NH4Cl (p-value < e-27). Image data shown in E and quantification in F. Number of cells > 75 for each treatment. G, H: Assay as described in 3A; Transduction efficiency is reduced with CQ with 8 hours of incubation (p-value < e-90). Number of repeats = 3 each for 0% DMSO Control and CQ. Data representation in B, D is as described in Fig 2; F as described in Fig 1; H as described in Fig 3. Scale bar: 40μm (A, C, E), 100μm (G).

To ascertain the mode of RBD and Spike-pseudovirus entry into AGS cells, we next measured the levels of known modulators–BafA1 and NH4Cl [64,65]. We, therefore, tested the effect of Chloroquine on RBD, dextran and transferrin uptake to verify if it behaves like BafA1. We found that upon treatment with Chloroquine, the uptake of neither RBD nor transferrin was altered significantly (Fig 4A and 4B). Dextran uptake was marginally higher upon treatment with Chloroquine

ACE2 biases RBD uptake via the clathrin-mediated endocytic pathway

ACE2 biases RBD uptake via the clathrin-mediated endocytic pathway

ACE2 biases RBD uptake via the clathrin-mediated endocytic pathway

Fig 5. Effects of BafA1 on RBD uptake and Spike-pseudovirus transduction in AGS-ACE2 cells.
A: HEKs, A549 and AGS cell lysates were evaluated for expression of ACE2 protein. Lysates from A549 and AGS cells by western blot analysis (Fig 5A). However, low expression of ACE2 transcripts was observed by qPCR in all three cell types (Fig 5B). Thus, AGS can be considered as a cell line with undetectable levels of endogenous ACE2. TMPRSS2 levels measured using qPCR also showed lower levels in all the cell lines tested. Spike-pseudovirus infection was competed out by excess RBD (S5E Fig), indicating an RBD-binding modality of entry in AGS cells. However, specific RBD-receptor interactions are unknown currently. Further studies interrogating specific RBD-receptor interactions in AGS cells will be required to determine the exact binding mechanisms.
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

To determine the effect of ACE2 on uptake of RBD in AGS cells, we generated a stable AGS cell line ectopically expressing ACE2 (AGS-ACE2); the expression of ACE2 was confirmed using qPCR and western blot analysis (Fig 5A). RBD uptake in AGS-ACE2 was about 3-fold higher than AGS cells (Fig 5E and 5F). On characterizing the RBD endocytic itinerary in AGS-ACE2 cells, we observed an increase in the co-occurrence of RBD with transferrin and slightly reduced co-occurrence of RBD with dextran compared to AGS cells (Fig 5C and 5D). This indicates that in addition to trafficking via the CG pathway, RBD is now trafficked via the CME in AGS-ACE2 cells. The AGS-ACE2 cell line has a heterogenous population of cells with a range of RBD uptake. We observed a positive trend between the co-occurrence coefficient of RBD with transferrin, and cellular RBD uptake (Fig 5D). This is consistent with the possibility that ACE2 biases RBD uptake towards CME.

Effects of BafA1 on RBD uptake and Spike-pseudovirus infection in AGS-ACE2 cells

We next evaluated the effect of BafA1 on RBD and dextran endocytosis and observed a significant reduction in the uptake of both in AGS and AGS-ACE2 cells (Fig 5E and 5F). However, the absolute reduction of RBD uptake in AGS-ACE2 is not to the same extent as in AGS cells. This is likely because BafA1 only affects the CG fraction of uptake and has little effect on RBD entering via the CME. Additionally, BafA1 shows a similar reduction of normalized RBD uptake (normalized to the surface levels of ACE2) in AGS cells transiently overexpressing myc-ACE2 (Materials and Methods, S3C and S3D Fig).

AGS-ACE2 cell line provides an important tool to distinguish the effects brought about by acidification inhibitors on endocytosis and neutralization of acidic endosomes. This is because ACE2 biases RBD uptake towards CME and therefore, the effects of BafA1 in this cell line will more predominantly be due to its neutralization role in the endo-lysosomal network. In pseudovirus transduction assays, we observed a significant reduction in the percentage of infected AGS-ACE2 cells with BafA1 (Fig 5G). Thus, BafA1 robustly affects Spike-pseudovirus infection in experimental regimes where RBD enters cells via both endocytic pathways, taking it one step closer to being more universal in its action of preventing infection.

We next designed time-of-addition experiments to disentangle the possibilities of the involvement of BafA1 at different stages of the viral entry process in both AGS (Fig 5F) and AGS-ACE2 (Fig 5H) cells. These assays revealed that the BafA1 sensitive step is during the virus presentation (~45 minutes) in both cell lines. However, pre-treatment with BafA1 or post-treatment with BafA1, even as early as 45 minutes after pseudovirus presentation, does not inhibit viral entry in both cell lines. This confirms that the effect of BafA1 is restricted to the early time points of entry and the endosomal neutralization role of BafA1 is a necessary and sufficient step in controlling the infection. Additionally, in cell lines with low ACE2 (like AGS and HEK293T), BafA1 also restricts infection by restricting entry via the CG pathway.

Identifying FDA-approved drugs functioning similar to BafA1 and NH₄Cl

Armed with the knowledge on the modes of action of acidification inhibitors in reducing the uptake of RBD, increasing the pH of endosomes and abrogating the infection of Spike-pseudovirus, we screened a small subset of FDA-approved drugs with the potential to alter the pH of endosomes (Fig 6A). We selected a panel of 6 drugs which includes those acting on Na+/K+ ATPase (Omeprazole, Esomeprazole, Pantoprazole, SCH-28080, Lansoprazole) and a protonophore that disrupts proton gradient (Niclosamide). We developed a quantitative high throughput screening pipeline for testing these drugs in both endocytic assay as well as pH estimation assay in AGS cells. The screen was carried out at a concentration of 10μM for all drugs.

Fig 6. Identifying FDA-approved drugs functioning similar to BafA1 and NH₄Cl.

A: Table describing various methods of altering endosomal pH along with the chosen subset of drugs to screen for entry and acidification inhibition B: C: High-throughput assay in which AGS cells were treated with an array of drugs at 10μM concentration for 30 minutes and pulsed with RBD, dextran and transferrin for 30 minutes. Niclosamide shows reduction in RBD (p-value < e-195) and dextran (p-value < e-133) uptake and increase in transferrin uptake (p-value < e-155). Omeprazole (p-value < e-34) and Pantoprazole (p-value < e-38) show an increase in transferrin uptake and minimally affects RBD or dextran uptake. Images are shown in B (and in S7A), quantification is shown in C and p-value table for all the inhibitor treatments is indicated in S1 Table. Control1 is 0% DMSO and Control2 is 0.3% DMSO. Number of repeats ≥ 3 for each treatment and each repeat has >80 cells. D: E: High-throughput assay in which AGS cells were pulsed with FITC and TMR dextran for 2 hours, chased for 1 hour with an array of drugs or control and imaged live. Endosomal pH increases upon addition of Niclosamide (p-value < e-110). Images along with pH maps are depicted in D (and in S7B), quantification is shown in E (and in S7C) and p-value table for all the inhibitors is indicated in S1 Table. Control1 is 0% DMSO and Control2 is 0.2% DMSO. Number of repeats ≥ 3 for each treatment and each repeat has >80 cells. Data representation in C and E is as described in Fig 2. Scale bar: 40μm (B, D). https://doi.org/10.1371/journal.ppat.1009706.g006

Of the 6 drugs tested in the endocytosis assay, Niclosamide showed the strongest effect on the uptake of the 3 probes (RBD, dextran and transferrin) similar to what we observed for the acidification inhibitors. Niclosamide-treated cells showed reduced RBD and dextran uptake (Fig 6B and 6C). It is interesting to note that Omeprazole had minimal effects on RBD or dextran uptake at the concentration tested, Omeprazole and Pantoprazole showed a
significant increase in transferrin uptake (Figs 6C and S7A). This suggests that these two drugs could specifically act on the transferrin-containing endosomes and not in the compartments of relevance for RBD and dextran uptake, while Niclosamide inhibits the RBD and dextran uptake.

Of the 6 drugs tested in the late endosomal pH estimation assay, Niclosamide also showed the strongest neutralization effect on the pH of acidic endosomes (Fig 5E) by increasing the endosomal ratio of FITC/TMR (S7C Fig). The other drugs had minimal effects on the pH of late endosomes at the concentration tested (Figs 6E and S7B). The spatial pH maps of Niclosamide-treated cells show an increase in pH in the majority of endosomes within the cell (Fig 6D). Niclosamide increased the FITC endosomal intensity and reduced the numbers of endosomes (S7C Fig) similar to the effect of BafA1 on these endosomal trafficking parameters.

Omeprazole and other proton pump inhibitors are prodrugs that are used for treating Gastro-esophageal reflux disease (GERD) [66]. They are activated by low pH, bind covalently to H+/K+ ATPase and inhibit the enzymatic function [67]. We tested the hypothesis if these drugs could also similarly block the proton pumps in the late endosomes and thus increase the endosomal pH [68-69]. Earlier studies have indicated that Omeprazole [70], Lansoprazole [71], and Pantoprazole [72], neutralize the endosomal pH only when used at very high concentrations (> 1mM) in EMT-6 and MCF-7 cells. However, the plasma concentration of these proton pump inhibitors varies between 1–25μM [66]. Thus, at least in the concentration range of relevance, we find no effect of these drugs on the acidification of endosomes and the uptake of RBD.

Niclosamide functions similar to BafA1 and NH4Cl as an acidification and entry inhibitor

Niclosamide is an anti-helmintic FDA-approved drug and has been in use since the 1960s (Ditzel, 1967). Many recent studies show that Niclosamide has broader clinical applications and has also been identified as an antiviral against SARS-CoV, human Rhinovirus, Influenza virus, Dengue virus [73-74]. As Niclosamide emerged as a potential drug candidate in both the RBD entry and endocytosis as well as endosomal pH neutralization screens, we investigated the dose-dependent role of Niclosamide in reducing RBD uptake by neutralizing endosomal pH and inhibiting Spike-pseudovirus infection. We found that Niclosamide reduced both RBD and dextran uptake, as well as increase transferrin uptake in a dose-dependent manner (1–25μM) (Figs 7A and S8A). We observed Niclosamide’s effect on RBD endocytosis even at concentrations as low as 1μM. On analyzing the effect of Niclosamide on endosomal numbers and intensity, we found that Niclosamide increased the endosomal intensity of transferrin endosomes and reduced the number of RBD and dextran endosomes (S8B Fig). These effects are remarkably similar to the effects observed with acidification inhibitors—BafA1 and NH4Cl. We also confirmed the inhibitory effect of Niclosamide on RBD and dextran uptake in another cell line—HEK-293T (S10A–S10D Fig), and on normalized RBD uptake in AGS cells overexpressing ACE2 (S3C and S3D Fig). In AGS-ACE2 cells, similar to the effect of BafA1, Niclosamide also showed a reduction in RBD uptake (Fig 7B).

Fig 7. Niclosamide functions as an acidification and entry inhibitor. A: High-throughput endocytic assay in which AGS cells were treated with different concentrations of Niclosamide for 30 minutes followed by pulse of RBD, dextran and transferrin for 30 minutes. RBD and dextran uptake decreases, and transferrin uptake increases in a dose-dependent manner. Images are shown in S8A, quantification in 7A and S8B and p-value table for all the concentrations is indicated in S1 Table. Number of repeats = 4 for Control (0.6% DMSO) and 2 each for each concentration of Niclosamide. Each repeat has >80 cells. B: AGS and AGS-ACE2 cells were treated with Control (0.2% DMSO) or 10μM Niclosamide for 30 minutes and pulse with RBD and dextran for 30 minutes with or without the inhibitor. Treatment with Niclosamide reduces RBD (p-value < e-24 in AGS, < e-3 in AGS-ACE2) and dextran (p-value <e-17 in AGS, < e-16 in AGS-ACE2) in both cell types. Numbers of cells > 70 for each treatment. C: D: High-throughput pH estimation assay in which AGS cells were treated with FITC and TMR dextran for 2 hours, chased for 1 hour with different concentrations of Niclosamide and imaged live. A dose-dependent increase in endosomal pH is seen with increasing Niclosamide concentrations. Images along with pH maps are shown in S7A and S9C. p-value table is indicated in S1 Table. Number of repeats = 6 for Control, 2 each for each concentration of Niclosamide and 1 for 10μM Niclosamide. Each repeat has >80 cells. E, F, G: Spike-pseudovirus transduction assay in which AGS cells (7E, 7F, S8A) or AGS-ACE2 (7G) were preincubated for an hour with different concentrations of Niclosamide or DMSO and incubated along with virus (MOI = 0.5) for 8 hours (7E, 7F) or 4 hours (S8A, 7G) followed by the continued presence of 100μM Niclosamide or 0.005% DMSO in the media after the removal of virus. Images of cells expressing the reporter mCherry protein in E and F, and normalized percentage transduction in F (AGS) and G (AGS-ACE2) show a dose-dependent reduction in transduction efficiency upon treatment with Niclosamide compared to pooled control from different DMSO treatments. Black dots with red error bars represent mean +/- SD. Dotted line represents the sigmoidal fit of the means across different Niclosamide concentrations. Refer to S9A(i) for % viability quantification, S9A(ii) for transduction efficiency after 4 hours of incubation with the virus in AGS cells and S1 Table for p-value table. H: 2-dimensional dose-response heat map in H depicts the combinatory Spike pseudovirus transduction assay in AGS cells with indicated concentrations of Niclosamide (0–5μM) and Hydroxychloroquine (0–10μM). Normalized percentage transduction of cells is represented as a heat map. Refer to S9F for % viability quantification. Data representation in A and D is as described in Fig 2, in B is as described in Fig 1. Scale bar: 40μm (C) and 100μm (E).

Further, we observed a dose-dependent effect of Niclosamide on neutralizing the pH of late endosomes and neutralization effects were seen even at 2.5μM (Fig 7C and 7D). The dose-response effect is seen on the ratio of endosomal FITC/TMR as well as other endosomal trafficking parameters—FITC and TMR endosomal intensities and numbers of endosomes (S8C Fig). The spatial pH maps of cells also show a gradual shift of endosomal pH from acidic to neutral pH with different doses of Niclosamide (Fig 7C), especially at 2.5μM wherein some endosomes within the cell are still acidic while some others are neutralized. Towards evaluating the effect of Niclosamide on the pH of early time point endosomes, AGS cells were labelled with FITC and TMR dextran for 20 minutes and chased for 10 minutes with or without Niclosamide for the entire duration of pulse and chase (S4F Fig). Unlike BafA1, Niclosamide reduced the net uptake of dextran. However, similar to BafA1, Niclosamide increased the endosomal FITC intensity and endosomal FITC/TMR ratio of early time point (30 minutes) endosomes (S4F Fig), indicating that Niclosamide neutralizes the pH of these endosomes as well.

We assessed the dose-dependent effect of Niclosamide on Spike-pseudovirus entry in AGS and AGS-ACE2 cells, using the experimental strategy designed to assess virus entry as described before. We observed a strong reduction of transduction efficiency as a function of increasing Niclosamide concentration at different viral incubation durations with negligible toxicity (Figs
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

A combinatorial approach of drugs with varying mechanisms of inhibition works as an effective therapy to combat infection [75]. Given that Niclosamide exhibits a short half-life [76], has poor bio-availability (~10%) [77] and our observations indicate moderate IC\textsubscript{50} for inhibition of Spike-pseudovirus transduction, we tested if the action of Niclosamide can be enhanced in the presence of another FDA-approved drug known to be effective against SARS-CoV-2 infection. Since published reports and commonly practised treatments against SARS-CoV-2 infection employ Hydroxychloroquine (HCQ), a less toxic variant of Chloroquine [78], we tested the effect of HCQ on altering late endosomal pH and Spike-pseudovirus transduction assay. Like Chloroquine, cells treated with 50μM HCQ also minimally altered the late endosomal pH (S9B and S9C Fig). However, we observed the pseudovirus transduction to be markedly reduced at HCQ concentrations of 50μM and 25μM (S9D and S9E Fig) and only modestly reduced at the concentrations of 10μM or lower in a dose-dependent manner (See the first box plot in S9F-Ill Fig). To assess the synergistic effect of the two drugs, we chose a concentration range with the maximum concentrations of 10μM HCQ and 5μM Niclosamide. The 2-dimensional dose-response map shown in Fig 7H summarises the effect of the two drugs on transduction. We observed an augmented reduction in infection when HCQ was used at a concentration of 10μM along with varying concentrations of Niclosamide compared to where HCQ was used at 0, 2 and 5μM (S9F-iii Fig). These results indicate an additive effect on inhibition of pseudovirus transduction when effective concentrations of HCQ is added along with effective concentrations of Niclosamide (Fig 7H). Thus, Niclosamide could potentially enhance the efficacy of other treatments currently being used to combat SARS-CoV-2 infection.

Bafilomycin and Niclosamide inhibit SARS-CoV-2 infection in AGS-ACE2 and Vero cells

Our studies on RBD trafficking and Spike-pseudovirus infections were extended to the infectious SARS-CoV-2. Evaluation of viral gene transcripts following infection using qRT-PCR and estimation of cytopathic effects (CPE—host cell death due to viral infection) are predominant to assess transduction (S7F, S10 and S9A Fig). Appropriate experimental conditions (MOI and time of incubation) for each cell line were standardized by evaluating the time at which CPE appears. Infection was confirmed in AGS and AGS-ACE2 cells with qRT-PCR against viral genes; infected cell lysates of AGS and AGS-ACE2 cells show about 4 and 9 fold increase, respectively at 18 hpi (hours post-infection) compared to corresponding uninfected controls (Fig 8B). Further, viral gene expression in infected AGS cell lysates and their supernatants indicated a persistent presence of these transcripts at 24, 48 and 72 hpi (S11C Fig). SARS-CoV-2 infection in AGS and AGS-ACE2 was additionally validated by collecting supernatants of infected cells and using them to re-infect Vero cells. Cell viability was reduced in Vero cells infected with AGS supernatant after 96 hpi. However, much lower volumes of supernatants from AGS-ACE2 cells affected Vero cell viability after 72 hpi, in line with the higher viral gene transcripts seen in these cells (Fig 8C).

Fig 8. Validating the effect of endosomal acidification inhibitors on infection with SARS-CoV-2 virus.

A: Schematic of viral infection assays in three cell types: AGS, AGS-ACE2 and Vero. Cells were pre-treated with vehicle controls or inhibitors for 1 hour and then presented with viruses at specified MOI for indicated duration. Viruses were then removed, and cells were further incubated in presence or absence of inhibitors. Infection efficiency was estimated using viral gene expression, cell viability or immunostaining for Spike antigen. B: Viral gene expression from cell lysates of infected AGS and AGS-ACE2 cells depicted as log fold change with respect to uninfected controls. Expression is normalized to 18s rRNA levels. Cells were infected with viruses at MOI 0.5 for AGS and MOI 0.1 for AGS-ACE2 and Vero, washed and further incubated for 18 hours. Number of repeats = 3 for all conditions. C: Evaluating infectability of supernatants from infected AGS and AGS-ACE2 cells. AGS and AGS-ACE2 were infected with viruses at MOI 0.5 and MOI 0.1 respectively for 8 hours, and cells were then treated with media containing supernatants for 24 hours. Indicated volumes of culture supernatants were used for infecting Vero cells for a total of 96 hours (for supernatant from AGS) and 72 hours (for supernatant from AGS-ACE2). Following this, Vero cells were stained with crystal violet to assess cell viability post infection. Image of the stained plate is shown in (i) and quantification in (ii). Number of repeats = 2 (AGS/Vero) and 3 (AGS-ACE2 /Vero) for both uninfected and infected conditions. D: Assessing effect of endosomal acidification inhibitors on viral infection. AGS-ACE2 (i) and Vero cells (ii) were pre-treated with control or inhibitors for 1 hour followed by presentation of viruses at MOI 0.1 for 30 minutes (AGS-ACE2) or 8 hours (Vero) in the presence or absence of inhibitors. Viruses were removed and cells further incubated for 16 hours (AGS-ACE2) or 48 hours (Vero) and cell viability assessed by ATP quantification assay. Number of repeats = 3 for each condition. Bars represent mean +/- SD of % viability upon infection. Asterix with error bars in red represents the mean +/- SD of % viability indicating drug toxicity in the absence of viruses. E: In cells expressing low endogenous ACE2, Spike RBD endocytosis is aided by ACE2/other host cell factors via the CLIC/GEEC pathway. Overexpression of ACE2 results in trafficking of RBD through both CG and CME pathways. Acidification inhibitors BafA1 and FDA approved drug Niclosamide, neutralize the pH of endosomes as well as block entry via the CG pathway. Infection assays using both Spike pseudovirus and infectious SARS-CoV-2 virus confirm that BafA1 and Niclosamide prevent viral infection.

https://doi.org/10.1371/journal.ppat.1009706.g008

CPE was observed as early as 8 hours of incubation with the virus in AGS-ACE2 cells, while Vero cells showed similar effects at a much later time point of 72 hours (S11A-Fig). Interestingly, no such effects were observed in AGS cells even at a higher MOI of the virus at 8 hours (S11A-Fig) and lower MOI of 96 hours (S11A-Fig). The extent of CPE was also quantified using cell viability assays (S11B Fig). Interestingly, similar to AGS cells, Calu-3 and Caco-2 cell lines which are permissive hosts to SARS-CoV-2 also do not exhibit CPE despite robust viral replication [79].

The effect of endosomal acidification inhibitors on SARS-CoV-2 infection was examined using CPE as a readout in AGS-ACE2 and Vero cells. Both BafA1 and Niclosamide showed a significant rescue in cell survival in AGS-ACE2 (Fig 8Di) and Vero cells (Fig 8Di). Similar rescue effects were observed in AGS-ACE2 cells in a time-dependent manner (S11D Fig). Hydroxychloroquine was also effective in inhibiting viral infection in Vero cells with no associated toxicity (Fig 8Di). Further, immunostaining for Spike antigen in infected AGS-ACE2 cells suggested a decrease in viral load, as shown by a decrease in Spike intensity, in the presence of BafA1 and Niclosamide (S11E Fig). These observations emphasize the relevance of endosomal acidification inhibitors and validate their use as potential therapeutic candidates in restricting SARS-CoV-2 infection.

Discussion
Understanding the molecular mechanisms of viral entry into relevant target cells is critical to design effective treatment and prevention strategies against infection. Employing various methodologies, we report for the first time that fluorescently labelled RBD of SARS-CoV-2 enters cells through a pH-dependent CG pathway in gastric origin AGS cells that have low or undetectable levels of ACE2 and utilizes CME in addition to CG upon ACE2 overexpression. High-resolution quantitative imaging approaches enabled us to detect the localization of RBD to acidic compartments. Endosomal acidification inhibitors that affect the uptake of CG cargo also inhibit RBD uptake. Complementing our observations with RBD, we show that infection by Spike-pseudovirus and clinical isolate of SARS-CoV-2 is also dependent on endosomal acidification. Further, by employing a targeted drug screen, we have identified Niclosamide as a potential inhibitor against SARS-CoV-2 entry and infection (Fig 8E).

The choice of viral entry into host cells is influenced by cell surface interacting partners and co-factors [12–28]. Although ACE2 has been identified as the receptor for SARS-CoV-2, other receptors are being uncovered. These include Neutrophil [13,14], CD147 [15], Heparan sulphate proteoglycans [16], and other highly glycosylated proteins [17]. In addition, note that the inhibition of Spike protein could also confer the ability to interact with yet unidentified receptors. It is well recognized that virus-receptor interactions, influenced by the levels of host receptor(s) and additional co-factors, dictate the endocytic route employed by the virus [81]. This is further exemplified by our observation that although RBD uptake is reduced upon blocking the CG pathway in AGS cells, residual RBD re-routes towards the CME and enables pseudovirus infection. Re-routing could be due to binding to different receptors that follow alternative internalization mechanisms or to activation of internalization machinery in AGS cells. Additionally, we therefore, affect ACE2 in AGS cells results in trafficking RBD via CME in addition to the CG pathway. Whether the Spike-pseudovirus and SARS-CoV-2 follow routes of entry like RBD, can be addressed with tractable pseudoviruses or synthetic virus-like particles. SARS-CoV-2 infection has been reported in multiple cell types expressing low levels of ACE2 – pulmonary and olfactory epithelial cells [13]. CG pathways could be a preferred route of entry in such cell types while CME can be employed in cells with ACE2 expression (as reported here and [82]). Future studies can explore the contribution of host factor interactions and endocytic routes in understanding cell-type-specific pathological outcomes of viral infection.

Even though respiratory symptoms dominate the clinical presentation of COVID-19, a subset of patients also face gastrointestinal symptoms [83–84]. There is growing evidence of the involvement of the gastrointestinal tract in SARS-CoV-2 infection [25,85]. Transcripomic and histochemical studies show that the gastrointestinal tract comprises of cell-types differentially expressing ACE2, cесophagus and proximal stomach have undetectable ACE2, distal stomach, duodenum, colon and rectum express high levels of ACE2. While our data demonstrate that both AGS (derived from stomach carcinoma) and AGS-ACE2 cells are permissive to SARS-CoV-2 infection, AGS-ACE2 cells show more severe cytopathic effects and extensive virus replication compared to AGS cells. These two cell lines could therefore offer suitable model systems to study SARS-CoV-2 infection in the gastrointestinal tract.

Known inhibitors of endosomal acidification, BafliomycinA1 and NH4Cl, play an important role in neutralizing acidic lysosomes and thus subverting viral membrane fusion and entry of several viruses [12,27–30]. Here, we report that these inhibitors also play a more upstream role by inhibiting the endocytosis of RBD itself. Both these treatments inhibited the uptake of CG cargo and RBD, most likely by interfering with the endosomal pH. It is interesting to note that the inhibition of acidification in addition to dramatically reducing CG uptake did not cause re-trafficking of RBD through another endocytic pathway, as was observed for other CG inhibitors. This suggests that the acidification inhibitors could negatively influence the RBD-receptor interactions at the cell surface along with further ramifications of blocking the CG pathway. In both AGS and AGS-ACE2 cells, BafliomycinA1 was effective in inhibiting Spike-pseudovirus transduction when added at the time of virus infection but was ineffective when added post-infection (as late as 45 minutes post-infection). BafliomycinA1 was, therefore, affects an early step in viral entry and not any downstream events such as viral protein translation or replication. The finding that BafliomycinA1 affects Spike-pseudovirus infection in AGS-ACE2 cells which employs both CG and CME routes of RBD entry, further validated that the inhibitor affects infection in the common acidic endo-lysosomal compartments where multiple endocytic pathways converge. We and others show BafliomycinA1 also affects other viral entry by while pseudotyped virus (S27 and S5Dii Fig) inhibits entry known to be endocytosed primarily via the CME [86,87]. Thus, targeting endosomal acidification can inhibit infection in a variety of cell types by affecting the very first step of endocytosis or the subsequent step of viral fusion in acidic endocytic compartments or both.

Dual mechanisms of inhibition offered by BafliomycinA1 encouraged us to screen a subset of FDA-approved compounds for similar inhibitory characteristics: proton pump inhibitors (Omeprazole, Lansoprazole, Pantoprazole, Esomeprazole, SCH-28080), and protonophore (Niclosamide). Of all the 6 compounds tested only Niclosamide inhibited CG cargo and RBD uptake, elevated endosomal pH and concomitantly inhibited Spike-pseudovirus infection, all in a dose-dependent manner with an IC50 of 1.27 μM in AGS cells. Niclosamide also rescued cytopathic effects upon infection with a clinical isolate of SARS-CoV-2 in AGS-ACE2 and Vero cells. Among several mechanisms of action [78], Niclosamide disrupts proton gradient across mitochondrial [89] and endosomal [73] membranes. The elevated endosomal pH brought about by Niclosamide was shown to inhibit human rhinovirus infection [73]. Additionally, Niclosamide has been identified as an anti-viral agent against SARS-CoV [90], Dengue [74], MERS-CoV [91] and more recently proposed for SARS-CoV-2 (with IC50 of 0.28 μM in Vero cells) [92]. Niclosamide also affects calcium-activated chloride channels and scramblases and interferes with sncytia formation induced by SARS-CoV-2 spike protein [42]. Our finding that Niclosamide inhibits SARS-CoV-2 entry provides an additional mechanism for its anti-viral property. In contrast, the proton pump inhibitors used in our study failed to interfere with RBD uptake. This could be because they remained inactive [88] or the concentrations tested predominantly affect H+K+ ATPases, while mM concentrations are required to inhibit V-ATPases [88]. Along these lines, studies show that proton pump inhibitors inhibit Ebola-virus [93]. SARS-CoV and SARS-CoV-2 [94] infection only when used beyond achievable plasma concentrations [96].

Surprisingly, Chloroquine did not affect RBD uptake and only marginally raised the endosomal pH. However, it still affected Spike-pseudovirus infection in the initial steps of entry. The mechanism of action of Chloroquine appears to be distinct from BafliomycinA1 and Niclosamide. Chloroquine is likely to function in many pH-independent ways to inhibit SARS-CoV-2 infections. For example, by altering terminal glycosylation of ACE2 [95]; via its activity as a zinc ionophore affecting ACE2 activation [96,97]; by interacting with ER-resident Sigma receptors that initiate cell stress response [98]; by its ability to strongly bind a viral protease essential for Spike activation [99].

In conclusion, our study reports that the high-capacity CG pathway serves as a potential endocytic route for SARS-CoV-2 in cells with low/undetectable levels of ACE2, and via both CG and CME upon ectopic expression of ACE2. We further show that endosomal acidification is critical for SARS-CoV-2 entry and infection and can be a promising therapeutic target across multiple host cell types as observed by the results seen with Niclosamide, BafliomycinA1 and NH4Cl observed by this study also paves way for large-scale screens to test different chemical libraries including FDA-approved drugs as acidification inhibitors and scrutinize for more Niclosamide-like drugs that might have better bioavailability or can be used in combination with other antiviral drugs. Moreover, the methods described in our study can be effectively extended to primary cells and emerging organoid systems that represent the more natural hosts for infection.

Materials and methods

https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706

10/20
Cell lines, constructs, and antibodies: See S1 Text for more details.

Chemicals and reagents

Niclosamide and AN96 were chemically synthesized and proton pump inhibitors, Esomeprazole and Pantoprazole, were extracted from commercially available 3000 library, and Omeprazole was procured from Sigma (O104).

Endocytosis assays

AGS or HEK-293T cells were plated in 35mm coverslip bottom dishes and processed after 48 hours at 60–70% confluency. Cells were washed twice with HEPES buffer (wash and imaging buffer composition: 150mM NaCl, 20mM HEPES, 5mM KCl, 1mM CaCl2, 1mM MgCl2, 2mg/ml Glucose, pH 7.5) at 37°C. Endocytosis was monitored using fluorescently labelled RBD (Alexa/Atto 488, 10μg/ml), 10kDa TMR-dextran (1mg/ml) and/or iron-loaded transferrin (10μg/ml, Alexa 647) in serum-free medium for indicated time points at 37°C. Endocytosis was stopped using ice-cold wash buffer and cells were subsequently fixed with 2.5% paraformaldehyde (PFA) for 20 minutes at room temperature (RT). Cells were then washed and imaged. For inhibitor experiments, cells were pre-treated with various inhibitors (AN96 25μM, ML141 50μM, Amiloride 1mM, BafA1 200nM or 400nM, NH4Cl 30μM) and respective controls in serum-free medium for 30 minutes at 37°C and inhibitors were maintained during endocytic assays.

To measure normalized transferrin or normalized RBD uptake (Figs 1H–1I and S3C and S3D), cell surface-bound probes after the endocytic pulse with transferrin or RBD were stripped using two washes with ice-cold ascorbate buffer (160mM sodium ascorbate, 40mM ascorbic acid, 1mM MgCl2, 1mM CaCl2, pH 4.5), followed by three washes with ice-cold wash buffer at 4°C. Cells were then fixed with ice-cold 2.5% PFA for 5 minutes at 4°C and 15 minutes at RT. Transferrin receptor (TR) was labelled by incubating cells with fluorescently labelled anti-HTR (OKT-9) for 2 hours at RT. To label surface ACE2, fixed cells were blocked with 10mg/ml bovine serum albumin (30 minutes) followed by incubation with anti-myc primary antibody (1 hour) and secondary antibody (45 minutes) in blocking buffer at RT. Cells were then washed and imaged.

pH estimation assays

For estimating the pH of late endosomes, cells were pulsed with pH-sensitive 10kDa FITC-dextran (1mg/ml) and pH-insensitive 10kDa TMR-dextran (1mg/ml) for 2 hours in serum-free media, chased for 1 hour in the presence of inhibitors or control and imaged live. The above pulse and chase times were chosen to allow the accumulation of labelled dextran in acidic late endosomal and lysosomal compartments. To estimate the endosomal pH, the ratio of FITC to TMR fluorescence was computed and compared to a pH calibration curve (S4A and S4B Fig) which was generated by equalizing the endosomal pH to that of an external buffer. After the pulse with FITC and TMR-dextran and chase, cells were incubated with 5μg/ml nigericin containing buffers of different pH for 10 minutes and imaged to evaluate FITC/TMR ratios for each pH.

For estimating the pH of late endosomes using the 488/458 excitation ratio of FITC-dextran (Fig 4E and 4F), cells were pulsed with FITC-dextran at 1mg/ml for 2 hours, followed by chase in the presence or absence of inhibitors and imaged live.

For estimating the FITC/TMR ratio of early endosomes (S4E and S4F Fig), cells were incubated with pH-sensitive 10kDa FITC-dextran (1mg/ml) and pH-insensitive 10kDa TMR-dextran (1mg/ml) for 20 minutes, chased for 10 minutes and imaged live. Throughout the pulse and chase duration, the cells were incubated in serum-free media with control (0.2%DMSO) or BafA1 400nM or Niclosamide 10μM.

Spike-pseudovirus transduction assays

AGS/HEK-293T cells were plated in optical bottom 96-well plates. 36 hours post-plating, when cell numbers were ~4000, transduction was carried out at indicated MOIs. For inhibitor treatment, cells were pre-incubated with indicated concentrations of NH4Cl BafA1/ CQ/ Niclosamide/ HCQ/ AN96/ ML141, for 1 hour. This was followed by addition of the Spike-pseudoviruses in presence or absence of the inhibitors. At the end of 2/4/8 hours, media containing pseudoviruses and inhibitors was removed, and cells were washed once with drug-free media. This was followed by addition of media with or without inhibitor: NH4Cl, BafA1 and CQ were removed from the media; Niclosamide, AN96 and HCQ were maintained at a low concentration of 100nM, 1μM and 500nM respectively. This was done to assess the effects of the inhibitors at the initial stages of infection, minimize long-term toxicity to the cells as well as to avoid effects on the translational processes of the reporter gene post entry. After 60 hours, cells were fixed, nuclei were labelled with Hoechst and assessed for transduction efficiency based on mCherry reporter expression. In the case of HEK-293T cells (S10G Fig), MTT cell viability assay was performed to check toxicity (assay described in S1 Text).

SARS-CoV-2 infection assays

SARS-CoV-2 (NR-52284 obtained from BEI-Resources) infections were conducted in AGS, AGS-AE2 and Vero cells at the indicated MOIs and incubation durations as mentioned in the legends of Figs 8 and S11. Post infection, different assays were used to evaluate the extent of infection: Cell viability assays, Spike immunostaining and qPCR. Each assay is detailed in S1 Text. The effect of endosomal acidification inhibitors on SARS-CoV-2 infection were tested in AGS-AE2 and Vero cells. Cells were pre-treated with inhibitors/vehicle controls at different concentrations for 1 hour, followed by infection in the presence of inhibitors/vehicle control. Virus and inhibitors were removed after indicated time of infection, cells were washed 3–4 times and incubated in virus free media with or without inhibitors. Post infection, Niclosamide treated cells were maintained in media with lower concentration of Niclosamide, while Bafilomycin treated cells were maintained in virus and drug free growth media until termination of the assay. To determine cytotoxicity of inhibitors, cells were treated with indicated concentrations of inhibitors in the absence of any virus presentation.

Imaging and analysis

a. Endocytic and pH estimation assays.

For 35mm dish-based endocytic experiments, fixed samples were imaged using confocal microscopy (Olympus FV3000, 20X/0.85NA objective) to image RBD, dextran and transferrin endosomes with Z sections of 1μm. Maximum intensity projected images were used for further analysis. Cell ROIs were drawn and features such as cell mean intensity in each channel was extracted.

https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706

11/20
For high-throughput endocytic and pH estimation experiments, automated imaging (Spinifex, Phenix, Perkin Elmer, 40X/W1.1NA objective) was used to image nucleuses along with RBD, dextran and transferrin (for endocytosis) or FITC and TMR dextran (for pH) with Z sections of 1μm each. For both assays, cell profiler based pipeline was used to segment cells, nucleus and endosomes and extract features as described in S1 Text. For pH calibration, the mean of the endosomal ratio distributions at different extracellular pH was fit to a sigmoidal equation. For both assays, custom MATLAB routines were used to estimate the endosomal intensities, the number of endosomes and cell mean intensities. In addition, for pH assays, endosomal ratio (FITC/TMR) and endosomal pH (using the calibration curve) for each endosome was computed. As the endosomal intensity distribution within cells is a heavy right-tailed distribution, median endosomal intensity for each probe for each cell was computed. The distributions of cell mean intensity/endosomal intensities/numbers of endosomes per cell per treatment (for endocytosis) and endosomal intensities/ratio/pH per cell per treatment (for pH) is represented in each quantification.

For 488/458 endosomal ratio estimation experiments, live imaging was done using confocal microscopy (Zeiss LSM 780, 40X/1.4NA objective). Excitation lasers 488nm, 458nm were used and emission was detected using a spectral detector (490nm-560nm). Images were processed as described above to estimate endosomal intensities and endosomal ratios per cell.

b. Colocalization analysis

Confocal microscopy (Olympus FV3000, 60X/1.42NA objective) with Z sections of 0.4μm each was employed to image cells across all channels. A MATLAB routine was written to extract colocalization indices. For each cell, endosomes in each channel were segmented based on threshold values. The segmentation in each channel was made finer using morphological operations (dilation followed by erosion). Segmented endosomes were considered for colocalization analysis. Manders’ coefficients and Pearson’s correlation coefficients were computed as described before (100).

c. Pseudovirus transduction assays

Automated imaging (Widefield, Phenix, 10X/0.3NA objective) of 96 well assay plates was used to image nucleus as well as mCherry positive cells. A cell profiler based pipeline was used to segment nucleus and extract features, as described in S1 Text. Approximately 50,000 nucleus (cells) were scored for each treatment. A MATLAB routine was written to estimate the % transduction. Mean intensities of the segmented nucleus in the nuclei channel and the mCherry channel for each nucleus across all fields were extracted. Each assay plate included “No-Virus” negative control. This control was used to estimate the background intensities of the mCherry channel within each segmented nucleus. The median of this distribution was considered the background. All nuclei with mCherry intensities of at least 1.8–2.2 times (empirically determined) the intensity of the background were considered positive. For each field, the fraction of positive nuclei to the total number of nuclei was determined. The mean of % transduction across all fields for each treatment was calculated. The % transduction was normalized to that of the control and is represented in all the quantifications. The total number of nuclei for each treatment is also represented to understand the effect of the toxicity of drugs.

Supporting information

S1 Fig. Generation of SARS-CoV-2 probe to study its endocytosis itinerary.
A: Schematic describing the protocol for purification and fluorescent labelling of RBD. B: i) Image of a 10% SDS PAGE Gel showing the output from Ni-NTA purification of his-tagged RBD. Input is the culture supernatant containing secreted RBD (marked by a black box on the gel). FL is the flowthrough after binding the supernatant to the Ni-NTA column. RBD is eluted in fractions containing increasing concentrations of imidazole (50, 100, 150, 200, 250 mM). ii) Image of a 10% SDS PAGE Gel showing purified RBD after Gel filtration step of purification. L represents the ladder lane. C: AGS cells were transfected with mycin-ACE2 and pulsed with RBD and transferrin for 30 minutes. Surface ACE2 was marked using anti-mycin antibody. Mycin-ACE2 transfected cells show increased RBD. D. E: AGS cells were transfected with mycin-ACE2 and pulsed with RBD for 30 minutes. The cell surface-bound RBD was stripped using ascorbate buffer and cell surface ACE2 was labelled using anti-mycin antibody. Images in D and scatter plot in E shows a positive correlation between the amount of RBD endocytosed and levels of surface ACE2. Number of cells >50. Scale bar: 40μm (C, D).
https://doi.org/10.1371/journal.ppat.1009706.s001 (TIF)

S2 Fig. RBD uptake is sensitive to CG pathway inhibitors.
A: AGS cells were pulsed with RBD, dextran and transferrin for 10 minutes and imaged at high resolution after fixation. Images in A and quantification in B shows that dextran and RBD are more correlated compared to dextran and transferrin (p-value < e-04) or transferrin and RBD (p-value < e-05) as measured using Pearson’s correlation coefficient (PCC). Number of cells = 10. C, D: AGS cells were treated with Control (0.6% DMSO) or AN96 25μM for 30 minutes, pulsed with RBD, dextran and transferrin for 30 minutes with Control or AN96 and imaged at high resolution upon fixation. Images are shown in C and quantification of Manders’ co-occurrence coefficient is shown in D. This depicts the fraction of RBD endocytosed in transferrin dextran and the fraction of dextran endosomal intensity with transferrin or RBD. As seen in D(i), in control cells, the fraction of RBD endosomal intensity is more associated with dextran than transferrin (p-value < e-07). With AN96, internalized RBD and dextran is associated more with transferrin compared to control cells. Numbers of cells in each condition >10. p-value table is indicated in S1 Table. E, F: AGS cells were treated with Control or ML141 50μM for 30 minutes and pulsed with RBD and Dextran for 30 minutes with or without the inhibitor. RBD (p-value < e-9) and Dextran (p-value < e-20) uptake is significantly reduced upon treatment with ML141. Images are shown in E and quantification in F. Numbers of cells > 100 for each treatment. G, H: AGS cells were treated with Control (0.2% DMSO) or Amloride 1mM for 30 minutes and pulsed with RBD, transferrin and dextran for 30 minutes with or without the inhibitor. RBD (p-value = 0.05), Dextran (p-value = 0.04) and transferrin (p-value = 0.013) uptake is not altered with Amloride. Images are shown in G and quantification in H. Numbers of cells > 80 for each treatment. I: AGS cells were serum starved and treated with Control (0.2%DMSO), PMA alone (100nM), Amloride alone (1mM) or in combination and pulsed with dextran for 30 minutes. Dextran uptake is enhanced with PMA; co-treatment with Amloride abolishes this increase. Data representation is as described in Fig 1. Scale bar: 10 μm (A, C) and 40μm (E, G).
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

https://doi.org/10.1371/journal.ppat.1009706.s002
(TIF)

S3 Fig. RBD uptake is sensitive to acidification inhibitors.

A: AGS cells were treated with Control (0.3% DMSO, 0.6% DMSO, 0% DMSO) or inhibitors (BafA1 200nM, BafA1 400nM, NH4Cl 30mM) for 30 minutes and then pulsed with RBD, transferrin and dextran for 30 minutes with or without inhibitors. Images are shown in A and quantification of A are total mean intensity shown in (i), the number of endosomes shown in (ii) and intensity per endosome shown in (iii) for each probe in each condition. Control0 is 0.3% DMSO, Control1 0.6% DMSO and Control3 0% DMSO. Number of repeats ≥ 4 for each treatment and each repeat has >80 cells. C, D: AGS cells transfected with myc-ACE2 were treated with Control (0.2%-DMSO) or BafA1 400nM or Niclosamide 10μM for 30 minutes and then pulsed with RBD for 30 minutes. The cell surface-bound RBD was stripped using ascorbate buffer and cell surface ACE2 was labelled using anti-myc antibody. Normalized RBD uptake is quantified as the ratio of the amount of internalized RBD to the amount of surface ACE2. Images depicted in C and quantification in D show that there is a reduction of RBD uptake upon treatment with BafA1 (p-value < e-08) or Niclosamide (p-value < e-07) in transfected as well as untransfected cells. Number of cells > 50 for each condition. Data representation in B and D are as described in Figs 2 and 1, respectively. Scale bar: 40μm (A, C).

https://doi.org/10.1371/journal.ppat.1009706.s003
(TIF)

S4 Fig. Acidification inhibitors neutralize endosomal pH.

A: pH calibration in AGS cells. AGS cells pulsed with pH-sensitive (FITC) and pH-insensitive (TMR) dextran were incubated in buffers of different pH with 5μg/ml of Nigericin and imaged live. A steady increase in the endosomal ratio of FITC/TMR with increasing pH is observed. The observed ratio vs. clamped pH is fit to a sigmoidal curve (red curve) which is used as a calibration curve to estimate the pH of endosomes. Numbers of cells in each condition is >100 cells. The data in B is represented as mean +/- SD. C, D: For the experiment described in Fig 2F-2H, images including pH maps are shown in Fig 2F, Fig 2H. S4C and quantification in Fig 2G. S4D. FITC and TMR endosomal intensities, numbers of endosomes and FITC/TMR endosomal ratio are quantified in S4D. BafA1 200nM/400nM and NH4Cl increases FITC intensity and reduces numbers of endosomes. NH4Cl also affects trafficking as seen with an increase of TMR intensity. Endosomal ratio (as a proxy for endosomal pH) also shows an increase with all the acidification inhibitors. Control0 is 0.2% DMSO, Control1 0.4% DMSO and Control3 0% DMSO. Number of repeats ≥ 3 for each treatment and each repeat has >80 cells. p-value table is indicated in S1 Table. E, F: Estimation of FITC/TMR ratio of early endosomes. AGS cells were pulsed with FITC and TMR dextran for 20 minutes, chased for 10 minutes and imaged live. Throughout the pulse and chase duration, the cells were incubated with Control (0.2%-DMSO) or BafA1 400nM or Niclosamide 10μM. Dextran uptake and TMR endosomal intensity are marginally reduced with Niclosamide while unaffected with BafA1. An increase in FITC endosomal intensity as well FITC/TMR endosomal ratio is observed with both inhibitors. Number of cells > 35 for each condition. p-value table is indicated in S1 Table. Data representation in D and F is as described in Figs 2 and 1, respectively. Scale bar: 40μm (A, C, E).

https://doi.org/10.1371/journal.ppat.1009706.s004
(TIF)

S5 Fig. Characterization of SARS-CoV-2 spike pseudotyped virus.

A: Schematic showing the strategy for generating SARS-CoV-2 Spike-pseudovirus. 2nd generation lentiviral helper plasmid psPAX was co-transfected with the reporter plasmid pHRmCherry and SARS-CoV-2 Spike protein-encoding plasmid psTwist Spike in HEK-293T cells to generate Spike pseudotyped virus particles. pHRmCherry reporter plasmid was used to score for infected cells by mCherry expression. B: Western blot showing bands of different molecular weights as detected by the anti-Strep-tag antibody which recognizes the C-term 2X Strep-tag on the Spike proteins incorporated into the pseudovirus particles. C: Comparison of infection by Spike-pseudotyped or bald pseudotyped (lacking spike) virus particles in AGS cells. Quantification shows that Spike-pseudotyped viruses infect AGS cells at various dilutions while bald-pseudoviruses do not infect at same dilutions. Number of repeats for each condition = 2 for each dilution. D: Transduction of AGS cells by Spike-pseudotyped viruses compared to an alternatively pseudotyped virus. (i) Quantification of infection at indicated MOI shows that VSV-G pseudotyped viruses are capable of transducing AGS cells at a higher efficiency. Number of repeats is 3 for Spike-pseudotyped viruses and 2 for VSV-G pseudotype. Data is plotted as mean +/- SD. (ii) Infection by VSV-G pseudotyped viruses is also susceptible to BafA1. Number of repeats = 2 for each treatment. E: Specificity of Spike protein-ACE-2 dependent pseudovirus entry was tested in a competition assay in the presence of excess purified RBD of Spike. Quantification of percentage transduction shows a reduction in transduction efficiency with both monomeric and trimeric Spike RBD in AGS cells. Number of repeats is 2 for each condition. F: Characterization of transduction efficiency in AGS cells of Spike-pseudovirus at varying MOIs and varying incubation times. Quantification of percentage transduced mCherry positive cells depicts a steady increase in transduction as a function of MOI and incubation times. Number of repeats = 2 for each condition. The data is plotted as mean +/- SD. G: (i) Treatment of AGS cells across different concentrations of AN96 shows no significant reduction in transduction efficiency of Spike-pseudotyped viruses compared to the pooled controls from 0.6%, 0.24%, 0.12%, 0.048% and 0.024% DMSO treatments. Number of repeats = 15 for pooled controls and 3 for each concentration of AN96. (ii) Treatment of AGS cells with 5μM of ML141 shows no significant reduction in transduction efficiency compared to control (p-value = 0.55). Number of repeats = 3 for control (0% DMSO) and ML141 respectively. Data representation in C, D, E, G is as described in Fig 3.

https://doi.org/10.1371/journal.ppat.1009706.s005
(TIF)

S6 Fig. Chloroquine does not affect RBD uptake and minimally affects endosomal acidification in AGS cells.

A, B: Using AGS cells treated with Control or CQ for 12 hours, RBD, dextran and transferrin uptake experiment (A) and FITC/TMR endosomal ratio estimation experiment (B) was conducted. Quantification in G and H show that RBD uptake, dextran uptake and FITC/TMR endosomal ratio are unaffected by long term treatment with CQ. Number of cells > 80 for each condition. Data representation is as described in Fig 1.

https://doi.org/10.1371/journal.ppat.1009706.s006
(TIF)

S7 Fig. Identifying FDA-approved drugs functioning similar to BafA1 and NH4Cl.

A: For the experiment described in Figs 6B and 6C, images are shown in 6B and 7A and quantification in Fig 6C. p-value table is indicated in S1 Table. B, C: For the experiment described in Fig 6D and 6E, images including pH maps are shown in Figs 6D and 7B and quantification in Figs 6E and 7C. FITC and TMR endosomal intensities and numbers of endosomes are quantified in S1 Table. Number of cells > 35 for each condition. The data is plotted as mean +/- SD. (i) Comparison of infection in AGS cells of Spike-pseudovirus at varying MOIs and varying incubation times. Quantification of percentage transduction shows a reduction in transduction efficiency compared to control (p-value = 0.55). Number of repeats = 3 for control (0% DMSO) and ML141 respectively. Data representation in C, D, E, G is as described in Fig 3.

https://doi.org/10.1371/journal.ppat.1009706.s007
(TIF)
**S8 Fig**, Niclosamide functions like BafA1 in inhibiting RBD uptake and neutralizing the endosomal pH.

A, B: For the endocytosis assay experiment described in Fig 7A, images are shown in S8A Fig and quantification in Figs 7A and S8B, with total cell mean intensity shown in Fig 7A, number of endosomes shown in S8B(i) Fig and intensity per-endosome shown in S8B(ii) Fig. The number of RBD endosomes and dextran endosomes decrease, while transferrin endosomal intensity increases with increasing concentrations of Niclosamide. p-value table is indicated in S1 Table. C: For the pH estimation assay described in Fig 7C and 7D, quantification of endosomal FITC intensities, TMR intensities and FITC/TMR endosomal ratio is shown in S8C. A dose-dependent increase in FITC endosomal intensity, as well as ratio, is seen with increasing Niclosamide concentrations. p-value table is indicated in S1 Table. Data representation in B, C is as described in Fig 2. Scale bar shown in A is 40μm. https://doi.org/10.1371/journal.ppat.1009706.s008 (TIF)

**S9 Fig**, Niclosamide and Hydroxychloroquine affect Spike-pseudovirus transduction in a dose-dependent manner.

A: Quantification in A shows the normalized percentage transduction of Spike-pseudo virus across different concentrations of Niclosamide for incubation times of 0hours(i) and 4hours(ii) in AGS cells. Both the inhibitor and the virus were removed beyond the indicated times and cells were incubated with the continued presence of 100nM Niclosamide or 0.005% DMSO until termination. Images shown in Fig 7E and dose-response curve depicted in Fig 7E are related to the experiment in S9A(i). Number of repeats = 3 for each condition in S9A(ii) and S9B, respectively for the 4 hours set. For the experiment in Fig 7G in AGS-ACE2 cells, number of repeats = 3 for each concentration of Niclosamide. p-value table is indicated in S1 Table. B, C: AGS cells pulsed with pH-sensitive (FITC) dextran for 2 hours and chased for 1 hour with Control or 50μM HCQ and imaged live. HCQ increases pH only slightly. pH maps are shown in B and quantification in C. Numbers of repeats: Control = 22, HCQ = 2. D, E: Images of AGS cells expressing the reporter mCherry protein transduction with Spike-pseudovirus in D and quantification in E show a dose-dependent reduction in transduction efficiency upon treatment with HCQ at the two concentrations tested compared to control (p-value < e-66 for HCQ 25μM, p-value < e-96 for HCQ 50μM). Number of repeats = 4 for control (0% DMSO) and 3 each for each concentration of HCQ. F: For the experiment described in Fig 7H, Quantification in S9F shows the normalized percentage transduction across indicated concentrations of Niclosamide in combination with indicated Hydroxychloroquine concentration of 2μM(i), 5μM(ii) and 10μM(iii). The percentage of cell viability for each condition is also indicated. Number of repeats = 2 for HCQ 5μM + Niclosamide 1μM combination and 3 each for all other combinations. p-value table is indicated in S1 Table. Data representation in A, E and F are as described in Fig 3 and C as described in Fig 2. Scale bar: 100μm (B). https://doi.org/10.1371/journal.ppat.1009706.s009 (TIF)

**S10 Fig**, BafA1 and Niclosamide affect RBD uptake in HEK-293T cells and reduce Spike pseudovirus infection in HEK-293T and A549-ACE2 cells.

A-D: HEK-293T cells were treated with Control (0.4%DMSO), BafA1 400nM or Niclosamide 10μM for 30 minutes and pulsed with RBD and transferrin (A and B) or RBD and dextran (C and D) for 30 minutes with or without inhibitors. Images are shown in A and C. Quantification is shown in B and D. RBD and dextran uptake is robustly reduced, while transferrin uptake increases upon treatment with BafA1 and Niclosamide. Number of cells ≥ 75 for each treatment, p-value table is indicated in S1 Table. E: Images show Spike-pseudovirus transduced mCherry positive HEK-293T cells in the presence of NH4Cl 20mM, CQ 10μM, BafA1 50nM and Niclosamide 5μM. F: Quantification of the normalized area of mCherry positive cells as a proxy for transduction, shows a reduction in transduction efficiency upon treatment with NH4Cl 20mM and CQ 10μM compared to 0% DMSO (Control), 25nM and 50nM of BafA1 compared with 0.05% DMSO (Control) and 0.1μM and 0.2μM, respectively for the 8 hours set and (1μM and 2μM) for the 8 hours set. Number of repeats = 3 for each condition except 4 for 0% DMSO. The data is represented as mean +/- SD. G: Toxicity, as assessed by MTT colorimetric assay, is reduced compared to 0.1%DMSO. H: Quantification of the mCherry positive cells in A549-ACE2 cells, as a proxy for transduction, shows a reduction in transduction efficiency upon treatment with 50nM of BafA1 in H and range of concentrations of Niclosamide in I (compared to respective DMSO controls). Data representation in B, D as are described in Fig 1 and H, I as in as Fig 3. Scale bar: 40μm (A, C), 100μm (E). https://doi.org/10.1371/journal.ppat.1009706.s100 (TIF)

**S11 Fig**, Cytopathic effect of SARS-CoV-2 in AGS-ACE2 and Vero cells is rescued by BafA1 and Niclosamide.

A, B: AGS, AGS-ACE2 and Vero cells were infected with viruses at indicated MOI for 8, 8 and 72 hours respectively. AGS cells were also infected with indicated MOI for 96 hours. Post infection, bright field images show cytopathic morphology in both AGS-ACE2 and Vero but not in AGS. Images shown are representative of multiple independent experiments. B: Evaluation of cell viability at early time points post infection in AGS and AGS-ACE2 cells. Cells were infected with viruses at indicated MOI for 4 hours or 8 hours. Cell viability, assessed using an ATP quantification assay, indicates cytopathic effects in AGS-ACE2 cells. Percentage viability relative to uninected control is depicted. Number of repeats = 3 for each condition. C: Viral gene expression in AGS cell lysates (i) and supernatants (ii) as a function of duration of infection. AGS and AGS-ACE2 cells were infected with viruses at indicated MOIs for the specified time periods post infection. Expression is depicted as log fold change compared to uninfected cells in (i), raw Ct values of viral gene transcripts from culture supernatants in (ii). NTC: no transcript control; PTC: positive transcript control. Number of repeats = 3 (uninfected and infected AGS), 1 (uninfected AGS-ACE2) and 2 (infected AGS-ACE2). D: Effect of endosomal acidification inhibitors on SARS-CoV2 infection in AGS-ACE2 cells. Cells were pre-treated with control/inhibitor for 1 hour at indicated concentrations and infected with virus for 6 hours in the presence/absence of inhibitors. Viruses were then removed, and cells were further incubated for 0, 6 or 12 hours. Cells treated with Niclosamide were maintained at 1μM post infection. Upon termination, cell viability was assessed by ATP quantification assay. Number of repeats = 3 for each condition. E: Detection of SARS-CoV-2 Spike antigen in infected AGS-ACE2 cells. Cells were pre-treated with control/inhibitor for 1 hour followed by infection with viruses at indicated MOI for 30 minutes in the presence/absence of inhibitors. Viruses were removed and cells further maintained for 16 hours. Cells treated with Niclosamide were maintained at 1μM post infection. Cells were fixed and stained for Spike antigen. Representative images are shown in (i) and quantification as violin plot in (ii). Number of repeats = 2 for each condition. Intensity of Spike staining is compared to uninfected and DMSO-treated cells. https://doi.org/10.1371/journal.ppat.1009706.s101 (TIF)

**S1 Table**, p-value table for various experiments (in bold) is detailed in S1 Table. Statistical tests between control and treatment were performed using Wilcoxon rank-sum test. https://doi.org/10.1371/journal.ppat.1009706.s012 (TIF)
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

https://doi.org/10.1371/journal.ppat.1009706.s013

(PDF)

Acknowledgments

We thank Nevan Krogan (UCSF, USA) for Spike expression construct, Florian Krammer (Mt. Sinai, USA) for secreted RBD expression construct, Biocon, Ltd (India), Raghuvar Varadarajan (IISc, India) for providing purified RBD for initial experiments and trimeric RBD for competition experiments, Minhaj Sirajuddin (inSTEM, India) for mCherry expression plasmid, Vinoth Kumar (NCBS, India) for discussions on Spike-pseudovirus characterization and Mylan Laboratories (India) for providing HCQ. The following reagents were obtained through BEI Resources, NIAID, NIH: SARS-Related Coronavirus 2, Isolate Italy-INMI1, NR-52284 (deposited by Dr. Maria R. Capobianchi); Human Lung Carcinoma Cells (A549) Expressing Human Angiotensin-Converting Enzyme 2 (HA-FLAG), NR-53522; Polyclonal Anti-SARS-Related Coronavirus 2 Spike Glycoprotein (IgG, Rabbit), NR-52947. We thank the Central Imaging and Flow Cytometry Facility (CIFF) and the Screening Facility at NCBS for imaging and high content screening, and biosafety facility for virus infection assays. We sincerely thank Abrar Bhat (for help with RBD purification), Chandrima Patra (for help with imaging), Greeshma Pradeep (for help with microscopes), Sarayu Beri (for help with cell culture), Jagadish Sampath (for assistance with BSL2 and BSL3 facilities), Amrutha K. Malayath and Umer Farooq (for qPCR primers) and all members of SM lab for exciting discussions on this project. We acknowledge the technical, administrative and hospitality staff at NCBS for tremendous help especially during the national lockdown.

References


https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706

15/20
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors


Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors


https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1009706
Strategies to target SARS-CoV-2 entry and infection using dual mechanisms of inhibition by acidification inhibitors

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   https://doi.org/10.1016/j.cellsig.2017.04.001 pmid:28386414

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar


   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar

   View Article  PubMed/NCBI  Google Scholar
View Article  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar

View Article  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar


View Article  •  PubMed/NCBI  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar

View Article  •  PubMed/NCBI  •  Google Scholar
COVID-19 co-infection mathematical model as guided through signaling structural framework

Bhavnita Soni, Shailza Singh*
National Centre for Cell Science, NCCS Complex, Ganeshkhind, SPPU Campus, Pune 411007, India

1. Introduction

In 2019, coronavirus disease (COVID-19) originated in Wuhan, China and has been declared as global public health emergency [1,2]. It is caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) and the clinical symptoms includes dry cough, fever, shortness of breath, pharyngitis and tiredness [3]. When virus enters the lower respiratory tract, it shoots up the cytokine storm including IL6, IL12, IL1β, TNFx leading to viral sepsis, Acute Respiratory Distress Syndrome (ARDS) and multiple organ failure that ultimately leads to the death of the patient. The clinical symptoms and disease transmission mechanism of SARS-CoV-2 infection are similar to influenza virus infection [4]. During the early stage of the pandemic, one of the studies on 5700 patients hospitalized with COVID-19 infection in the New York City revealed that nearly 2.1% patients were found to have co-infection with other respiratory viruses including influenza A virus [5]. Furthermore, the percentage was increased in Wuhan, China where 4.35% confirmed cases of SARS-CoV-2 and influenza virus co-infection were reported [6]. Many parts of Europe and USA, considered it as ‘Double threat’ as its been anticipated in October 2020 that SARS-CoV-2 and Influenza virus might surge simultaneously [7].

One of the characteristic aspects of COVID-19 infection is the generation of cytokine storm. The molecular structure of SARS-CoV-2 has four major proteins namely SARS-CoV-2 spike (S), envelope (E), membrane (M) and nucleocapsid (N) [8]. Through its spike (S) glycoprotein, it interacts with alveolar epithelial type II cells which forms a protective layer on the inner respiratory tract. The high affinity interaction of Spike glycoprotein occurs with angiotensin-converting enzyme 2 (ACE2) and additionally processed by transmembrane serine protease 2 (TMPRSS2) both being present on the host cell [9]. The virus gets entry inside the cell, leading to the excessive production of pro-inflammatory cytokines such as IL6, IL1β, TNFx, IL12 and IL18. Most of these cytokines functions in acute inflammatory responses and are liable for causing Cytokine release syndrome (CRS) or Cytokine Storm syndrome [10–13]. CRS is defined as cytokine-mediated systemic inflammatory response caused by excessive production of pro-inflammatory cytokines which ultimately results in clinical symptoms such as unrelenting high fever, lymphadenopathy, cytopaenia, central nervous system (CNS) abnormalities and if not treated leads to multiple organ failure (MOF) and death [14]. The hallmark of CRS is the unchecked feed forward activation and amplification of host cellular immune signaling which in turn activates the nearby cells to produce similar pattern of cytokines.
leading to cytokine storm [15]. Similar to COVID-19 infection, the cytokine storm plays a crucial role in severe influenza infected patients where aggressive pro-inflammatory response and in sufficient anti-inflammatory response leads to MOF and death of the patient [16–18]. Literature survey suggested that risk of death with SARS-CoV-2 was double in people which were prior infected with flu [19]. Furthermore, the current data suggested that the co-infection brings out synergistic effect of these two viruses which drastically increase the mortality rate as well as the demands of the health services [20]. This necessitated the design of new therapeutics targeting co-infection dynamics at basic cellular level and apparently help reduce the mortality rate.

Macrophage plays crucial role in disease manifestation during both the infection. The activated alveolar epithelial cells release excessive amount of MCP-1 and GM-CSF which in turn recruits macrophages at the site of infection. The macrophages engulf the infected epithelial cells for phagocytosis. The single-cell transcriptomic profiling study of broncho alveolar lavage fluid samples from 88 patients with SARS-CoV-2-induced respiratory failure suggested that SARS-CoV-2 infects alveolar macrophages, which in turn respond by producing T cell chemo attractants [21]. Moreover, SARS-CoV-2 transcriptome has been found in alveolar macrophages. This shows that the capacity of SARS-CoV-2 to infect cells, is not only limited to alveolar epithelial cells but it can also infect alveolar macrophage. Further, macrophages play a central role in antiviral responses, tissue repair and fibrosis. Macrophages can be reprogrammed by environmental cues and thus can change their phenotype during an antiviral immune response as the viral infection progresses [22]. Thus, macrophages are considered as prominent candidate for therapeutics development in COVID-19 infections. The nature of both the viruses is similar. SARS-CoV-2 is positive strand, non-segmented RNA virus having their genetic material protected in viral envelope. The outermost layer has a spike glycoprotein which is recognised by the host cell whereas influenza is negative strand, segmented RNA virus, protected by viral envelope. The surface protein includes Haemagglutinin and Neuraminidase [23,24]. Literature survey suggested that the RNA genome of SARS-CoV-1 and H1N1/H1N5 is recognized by Toll-like receptor 3 which is one of the major pattern recognition receptor (PRR) functioning in innate immune system [25–27]. Like TLR7, TLR8, TLR9 and TLR3 is present intracellular in an endosomal compartment and is known to recognize viral genome [28]. Consequently, TLR3 is likely to be involved in pathogenesis of SARS-CoV-2 virus infection. Further in SARS-CoV-1 infection, the stimulation of TLR3 activates NFKB transcription factor through TRIF signaling resulting in activation of pro-inflammatory genes such as IL6, TNFα, IL12 and IFNγ [27]. Influenza A utilizes PI3K/Akt pathway to activate transcription factor NFKB [29]. These cytokines are likely to activate nearby macrophages and epithelial cells resulting in massive production of cytokines and activation of macrophages, the phenomena called as Macrophage activated syndrome (MAS). One of the major genes activated by pro-inflammatory cytokines is Interleukin Adhesion Molecule (ICAM) and Vascular Cell Adhesion Molecule (VCAM). ICAM and VCAM function in cell–cell interaction. It is expressed by endothelial cell, epithelial cell and immune cells including macrophages. ICAM-1 interacts with LFA-1 present on leukocytes and facilitates its entry in tissues. ICAM-1 is found to enhance influenza viruses infection and their survival during an early stage of infection [30]. Similarly, VCAM also functions in leukocyte-endothelial cell interaction and both the cell adhesion molecules are likely to play a role in SARS-CoV-2 and influenza virus infection [31–33]. Together with IL6, TNFα, IL12 and IFNγ, Macrophage Colony Stimulating Factor (MCSF) levels were also found to be elevated in COVID-19 infected patients [34]. MCSF is responsible for proliferation and differentiation of monocytes. MCSF plays a major role in establishing MAS during SARS-CoV-2 infection.

Literature survey suggested the presence of curated population model defining spread of COVID-19 infection (e.g. BIOMD000000957 or BIOMD000000958) but is unable to define mathematically the co-infection dynamics. In the present work, using numerical simulation [35,36], combination of bits and pieces of literature evidences has acquired us to build new mathematical model which define the co-infection dynamics at cellular level and reveal crucial signaling proteins that can be targeted for generating new therapeutic regime for COVID-19 infection (Fig. 1).
2. Materials and methods

2.1. Databases for signaling network

For identifying each component in the interlinked signaling cascade, we have employed various immune signaling based databases. For IL6 signaling pathways we have used String database. TLR signaling pathway (map04620), PI3K-Akt signaling pathway (map04611), TNFα signaling pathways (map04668), influenza A pathway (map05164) and SARS-CoV-2 infection pathway (map05171) has been obtained from KEGG pathway database [37]. For identifying targets of transcription factor NFκB, STAT1, STAT4 and STAT3 we have used TRRUST (version 2) database [38].

2.2. Construction of mathematical model

Here, we have constructed two mathematical models namely COVID-19 model and co-infection model (SARS-CoV-2 and influenza virus) using SimBiology toolbox from Matlab v7.11.1.866. SimBiology is a programmatic tool to model, simulate and analyse dynamic systems focusing on pharmacokinetic/pharmacodynamics (PK/PD) and systems biology applications. It provides a block diagram editor for building models or one can create models programmatically using the MATLAB® language. It supports systems biology mark-up language (SBML) format. Following are the steps used to create and simulate mathematical model of COVID-19 and co-infection model:

2.2.1. Building model

The models were constructed in diagram editor using compartment, species, reactions and plot building blocks. In “Model Building: Reaction Properties” we defined reactions and assigned specific rate laws (e.g. for association/dissociation, we used Law of Mass action, for enzyme kinetics, we used Michaelis-Menten equation and for gene expression, we used Hill equation). Following this, concentration and parameter values are provided with respective units. As we are dealing with immune signaling, the concentration needs to be set between 10^3–10^6 signaling molecules [39]. Further, the concentration of IL6 signaling pathway has been obtained from FACS experimental data [40] and has been incorporated to mimic the mathematical model close to physiological condition.

2.2.2. Simulation

The reactions were defined along with concentration and parameter followed by model simulation using ode15s (stiff/NDF) solver. The solver computes the model’s state at the next time step.
using variable-order numerical differentiation formulas (NDFs). The values of absolute and relative tolerance are kept as default. The model is then simulated for 100 min close to 60 min of infection time taken by SARS-CoV-1 to infect the macrophages [41]. The behaviour of each component in the system have been obtained in the form of state v/s time graph.

2.2.3. Parameter estimation
Parameter estimation is required to improve the accuracy of the mathematical model. The parameter of each reaction in reconstructed signaling network has been assigned within a range of parameter space. The simulation is performed to fine tune these parameters till a reproducible graph depicting the biochemical behaviour has been obtained. The process is iterative in nature as fine tuning of parameter space is required for proper estimation together with algorithmic data fitted to reach physiological accuracy.

2.3. Mathematical model analysis
The mathematical model is analysed using various systems driven techniques such as Sensitivity analysis, Flux analysis, determining network properties and principal component analysis. The model is analysed through various aspects in order to churn out the crucial reactions that are governing the dynamics of the system (Fig. 2).

2.3.1. Elucidating network properties
Mathematically each biological network can be defined in terms of an interacting network of nodes and edges where biological components are the nodes and their interactions such as association, dissociation, phosphorylation, de-phosphorylation, ubiquitination, translocation etc. are the edges. Since the present work is based on immune signaling, the network is called as Protein-Protein Interaction (PPI) network. The network is analysed using Cytoscape (v 3.4.0) in terms of degree, clustering coefficient and centrality.

2.3.2. Sensitivity analysis
The mathematical models were subjected to sensitivity analysis in order to check the robustness. Sensitivity analysis is a technique, wherein, the robustness of the mathematical model is achieved by perturbing the parameters of each reaction in the system. It further aids in parameter estimation close to physiological condition. Sensitivity analysis is obtained in terms of sensitivity coefficient and for the present work, we have performed local sensitivity analysis using SimBiology tool box from Matlab v7.11.1.866. We have calculated the time dependent sensitivities of the entire component in the system using SimBiology tool box. The SimBiology sensitivity analysis uses the “complex-step approximation” to calculate derivatives of reaction rates. It calculates the time-dependent sensitivities of all the species states with respect to species initial conditions and parameter values in the model. Thus, if a model has species x and two parameters y, and z, then the time-dependent sensitivities of x with respect to each parameter value can be represented at the time-dependent derivatives as Eq. dx/dy, dx/dz where the numerators are the sensitivity output and the denominators are the sensitivity inputs to sensitivity analysis.

2.3.3. Principal component analysis
Principal component analysis was performed to reduce the dimensionality of the large amount of sensitivity coefficients obtained from sensitivity analysis. The principal component (PC) score depicts the variation in sensitivities of each component in the system. The high PCA score depicts the high sensitivity of the component in the system or in other words the system might collapse if that component would be targeted. In network biology, highly connected nodes are those which has tendency to pass maximum biological information for one end to the other. The term “collapse” here means, the transfer of biological information which is interrupted and which in turn greatly impacts the output of the system. PCA is meant for reducing dimensionality and removing the background noise of the biological data. For calculating PCA, we have used Matlab’s function as score_coefficient = princomp A, where A = m*n matrix, sensitivity scores for each component with the other component in the system.

2.3.4. Flux analysis
The flux determines the productivity of each reaction, thereby determining the contribution of each reaction in the system. Comparative flux analysis is one of the optimum strategies to determine higher productive reactions in the system that may contribute to the disease pathogenesis. In the present work, we
have obtained flux of each reaction by using COPASI pathway simulator (v4.11).

2.4. Model reduction

The process is employed to reduce the complexity of the system by eliminating reaction with no or low impact on overall dynamics of the system. It simplifies the mathematical model through the elimination of redundant parameters which are not contributing to the system, thus easing in accurate prediction of the biological system. Here, we are employing sensitivity, flux and principal component analysis to churn out key components governing the dynamics of the system.

3. Results

3.1. Mathematical modeling and simulation

3.1.1. COVID-19 model

The mathematical model has four compartments namely membrane, cytosol, nucleus and endosome. There are total 61 signaling components, 106 parameters, 58 reactions and kinetic laws (Fig. 3a). The simulation was performed for 100 min, using Stiff Deterministic ODE15s solver (SimBiologiing toolbox) which generates the first order non-linear ODEs for each reaction in the system. The network shows production of four components at the end of simulation i.e. ICAM, VCAM, iNOS and MCSF (Fig. 3b). The ODEs of the two major reactions are given below and for the entire system (S6). The ODEs represents the kinetic equations involving two major components of the system i.e. NFκB and STAT1, obtained after simulating the entire system for 100 min. The model was submitted to Biomodel database with identifier MODEL21011130002.

3.1.2. Co-infection model

The mathematical model has four compartments namely membrane, cytosol, nucleus and endosome. There are total 67 signaling components, 118 parameters, 64 reactions and kinetic laws (Fig. 4a). The simulation was performed for 100 min, using Stiff Deterministic ODE15s solver (SimBiologiing toolbox) which generates the first order non-linear ODEs for each reaction in the system. The network shows production of four components at the end of simulation i.e. ICAM, VCAM, iNOS and MCSF (Fig. 4b). The ODEs of two major reactions are given below (S6). The ODEs represents the kinetic equations involving two major components of the system i.e. NFκB and STAT1, obtained after simulating the co-infection model system for 100 min. The model was submitted to Biomodel database with identifier MODEL21011130002.

\[
\frac{d([\text{NFκB}] \cdot \text{Vcytosol})}{dt} = +\text{VCytosol} \left( \frac{100 \cdot \text{IKKβComplex}}{50 \cdot \text{IKKβComplex}} \right) \frac{\text{Vcytosol}}{+\text{VCytosol} \left( \frac{100 \cdot \text{TAK1}}{50 \cdot \text{TAK1}} \right) \frac{\text{Vcytosol}}{+\text{VCytosol} \left( \frac{355 \cdot \text{IKKβComplex}}{800 \cdot \text{IKKβComplex}} \right) \frac{\text{Vcytosol}}{}}{}}
\]

\[
\frac{d([\text{STAT1}] \cdot \text{P}) \cdot \text{Vcytosol})}{dt} = +\text{VCytosol} \left( \frac{200 \cdot \text{JAK1.JAK2}}{50 \cdot \text{JAK1.JAK2}} \right) \frac{\text{Vcytosol}}{-\text{VCytosol}(1 \cdot - 0.5 \cdot [\text{STAT1}] \cdot P)\frac{+\text{VCytosol} \left( \frac{500 \cdot \text{JAK1.JAK2}}{100 \cdot \text{JAK1.JAK2}} \right) \frac{\text{Vcytosol}}{+\text{VCytosol} \left( \frac{400 \cdot \text{TRAF6}}{100 \cdot \text{TRAF6}} \right) \frac{\text{Vcytosol}}{}}}{}}
\]

By analysing both the models, we have deciphered the concentration of ICAM, VCAM, iNOS and MCSF, which is increased in co-infection model system as compared to COVID-19 model indicating the severity of the disease during co-infection (Fig. 5).

3.2. Analysis of mathematical models

The COVID-19 model consists of 66 nodes and 112 edges where as co-infection model consist of 72 nodes and 112 edges (Fig. 6). The clustering coefficient of the COVID-19 network is 0.302 whereas co-infection network is 0.290, depicting that both the network are highly connected, therefore the flow of biological information is faster. Digging into the biology of the network, we have done network based ranking and identified five components which are found to have high degree as well as centrality value in both the network (Fig. 7). Degree, Closeness, Betweenness centrality is the classic centrality measures. The betweenness centrality symbolizes the communication or flow of biological information in the system. In both the model systems, cytoplasmic and nuclear NFκB, phosphorylated STAT1, JAK, JAK1-JAK2 complex are the components through which maximum amount of biological information is transferred (Fig. 7a).

Closeness centrality determines the higher connectivity of the respective components in the system (Fig. 7b) and degree centrality determines the local connectivity of the components (Fig. 7c). Cytoplasmic NFκB is found to be the cross talk identified between TNFα, IL1β, and TLR3 signaling pathway whereas phosphorylated STAT1 is the cross talk between IL6 and IFNγ signaling pathway (Fig. 7d).

3.3. Flux analysis

The flux of each reaction in both the models have been calculated in terms of molecules/ second and are compared to get broader picture of functioning of the two different models (Fig. 8a). The comparative flux analysis shows that the reaction associated with major pro-inflammatory cytokines such as IL12, IL1β, IL6 and TNFα has high amount of flux (Fig. 8b & c). This signifies that these reactions are responsible for major output of the system i.e., in the gene expression of ICAM, VCAM, MCP-1, iNOS etc. Targeting any one of the reactions may alter the expression levels of the above-mentioned gene resulting in decrease severity of the disease. Moreover, the flux of IL1β and IL6 reactions are increased which indicates their role in aggravated cytokine storm causing MAS (Fig. 8b & c).
Fig. 3. Mathematical model showing Cytokine storm during Covid 19 infection in alveolar macrophage (a) Diagrammatic view of interconnected signaling network (b) Simulated graph of the model depicting the increased levels of ICAM, VCAM, iNOS and MCSF as compared to the basal levels.
Fig. 4. Mathematical Model showing Cytokine storm during Covid 19 and influenza co-infection in alveolar macrophage (a) Diagrammatic view of interconnected signaling network (b) Simulated graph of the model depicting the increased levels of ICAM, VCAM, iNOS and MCSF as compared to the basal levels.
3.4. Principal component analysis

The comparative analysis of PC score shows that major components common in both the systems are IL6, gp130, IL1β, VCAM, ICAM, NFκB and phosphorylated STAT1. The high score of IL6 and IL1β indicate their role in establishing pro-inflammatory immune response during infection stage. The data is supported by experimental data, where IL6 and IL1β levels has been found elevated in COVID-19 patients and IL6 is also considered as a parameter for measuring disease severity [42,43]. Apart from this, PCA shows novel components such as cytoplasmic NFκB and phosphorylated STAT1. These components can be considered as targets as they have high centrality values. Targeting these components may affect the immune signaling established during SARS-CoV-2 and Co-infection. VCAM and ICAM is the end product of the system responsible for Macrophage Activated Syndrome (MAS).

Principal component analysis shows the most sensitive or the most crucial components in both the model systems (Fig. 9). The sensitivity coefficient of each component with respect to other components in the system have been documented in supplementary data (S7).

3.5. Model reduction

The process of model reduction involves overlapping components and reactions identified in centrality, flux, sensitivity and principal component analysis. The complexities of both the models were reduced that resulted in 89.65% reduction in COVID-19 model whereas 90.62% reduction in co-infection model system. The reactions that are playing major role in regulating the dynamics of the system are as follows:

Fig. 5. Comparative analysis of end product of COVID-19 and Co-infection model.

(a) (b) (c) (d)
<table>
<thead>
<tr>
<th>SN.</th>
<th>Reactions</th>
<th>COVID-19</th>
<th>Co-infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IL12(Membrane) + IL12R1 + IL12 R2 -&gt; “IL12 signaling complex”</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>IL6R + gp130 + IL6(Membrane) -&gt; “IL6 signaling complex”</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>IL1B(Membrane) + IL1R1 + IL1R2 -&gt; “IL1B signaling complex”</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>TNFR1 + TNFA(Membrane) + TNF R2 -&gt; “TNFA signaling complex”</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>IKKb complex -&gt;NFkB</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>JAK1-JAK2 -&gt; STAT1.P</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

NFkB and phosphorylated STAT1 are further selected for therapeutic intervention in COVID-19 infection due to their high centrality as PCA scores values (Figs. 7 and 9).

4. Discussion

The formation of cytokine storm is the hallmark for the severity of COVID-19 infection as well as influenza infection. The in-depth analysis of cellular signaling reveals several components that play crucial role in disease progression. The high flux of IL6, IL1β, TNFα, IL12, IFNγ signaling indicate the establishment of cytokine storm in COVID-19 and Co-infection model systems. After 100 min simulation, the end products are inflammatory proteins which include ICAM, VCAM, MCP-1, MCSF and iNOS. The co-infection model depicting the simultaneous infection of COVID-19 as well as influenza virus shows higher production of these inflammatory proteins indicating that the co-infection state is even worse than the COVID-19 infection as alone. The condition can be controlled once the productions of these inflammatory proteins are regulated at cellular level. Further, the model reduction reveals cytoplasmic NFkB and phosphorylated STAT1 as crucial therapeutic targets to regulate the intensity of cytokine storm established during infection. At present, there are various drugs available to inhibit the activity of NFkB such as Non-Steroidal Anti-inflammatory Drugs (NSAIDs). Aspirin and sodium salicylate are examples of anti-inflammatory agents that target NFkB activity during chronic inflammation [44]. Cyclosporin A (CsA) and tacrolimus (FK-506) are immunosuppressive agents used during organ transplant and are found to inhibit calcineurin dependent NFkB activation [45]. The present line of therapeutics has their own potential side effects whereas use of peptide in therapeutical application provides more immunogenicity and specificity as well as has less off target effects. Its envisaged that the therapeutical design of the peptides is ideal to target the activity of NFkB for shorter duration of time.

With the beginning of 2021, new strain of COVID-19 has emerged in United Kingdom [46]. The information on host-parasite interaction, proteins involved, mechanism of action is still unknown. It is foreseen that since we are targeting common cellu-
Fig. 8. Comparative flux analysis of COVID-19 and Co-infection model. (a) Flux of all the reactions. (b and c) Reactions with high flux values.

Fig. 9. Principal Component analysis depicting the crucial components in the model system.
lar proteins responsible for aggravated immune response the adopted strategy might be helpful or equally efficient against rising mutant strains by providing long-term immunity.

5. Conclusion

The host immune signaling plays a crucial role in establishing or combating any viral infection. The present piece of work provides an in-depth system analysis of COVID-19 and Co-infection mathematical model which reveals major cellular protein (including transcription factors, cytokines, chemokines, receptors etc.) that can be targeted for future therapeutic intervention. The major components produced after 100 min simulations are IL6, IL12, IL1β, TNFa, IFNγ, iNOS, VCAM, ICAM and MCSF. The production of iNOS, VCAM, ICAM and MCSF are high during Co-infection model indicating the disease severity during co-infection of influenza and SARS-CoV-2 virus. Further the network analysis highlighted the cellular protein JAK2, phosphorylated STAT1, cytoplasmic and nuclear NFkB showing high centrality value and indicating their role in establishing pro-inflammatory immune response during infection. These high centrality nodes are perfect targets to rewire the impaired immune response during SARS-CoV-2 infection. The models were further reduced to 90% through flux, sensitivity and principal component analysis and we have identified cytoplasmic proteins NFkB and phosphorylated STAT1as targets for future therapeutic intervention in case of both SARS-CoV-2 or Co-infection with the influenza virus. The underlying purpose here is to provide list of common potential targets out of numerous cellular proteins activated during SARS-CoV-2 infection and co-infection stage, which further might pave a way for generation of single therapeutics that is efficient in combating both the infection stages. The data generated here has opened new avenues for experimental investigations with accessible realms.

6. Author statement

Bhaviniita Soni performed the mathematical modeling and was involved in acquisition, analysis and interpretation of the data. Shaliza Singh was involved in the conception and design of study. Bhaviniita Soni and Shaliza Singh edited the manuscript and all authors approved the final version of the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Director, NCCS for supporting the Bioinformatics and High performance Computing Facility at National Centre for Cell Science (NCCS), PUNE, INDIA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.03.028.

References


Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV-2), which initiated as an endemic from China, converted into a pandemic disease worldwide within a couple of months’ time. This has led researchers from all over the world to come together to find and develop possible curative or preventive strategies, including vaccine development, drug repurposing, plasma therapy, drug discovery, and cytokine-based therapies. Herein, we are providing, a summarized overview of immunopathology of the SARS-CoV-2 along with various therapeutic strategies undertaken to COVID-19 with a vision for their possible outcome. High levels of proinflammatory cytokines such as interleukin (IL)-7, G-CSF, IP-10, TNF-α, monocyte chemoattractant protein-1 (MCP-1), and IL-2 in severe cases of COVID-19 have been observed. Immune responses play significant roles in the determination of SARS-CoV-2 pathogenesis. Thus, exploring the underlying mechanism of the immune system response to SARS-CoV-2 infection would help in the prediction of disease course and selection of intensive care and therapeutic strategy. As an effort toward developing possible therapeutics for COVID-19, we highlighted different types of vaccines, which are under clinical trials, and also discussed the impact of genome variability on efficacy of vaccine under development.

Keywords: COVID-19, immunopathology, SARS-CoV-2, cytokine storm, complement system, vaccine development

Introduction

Severe Acute Respiratory Syndrome-Coronavirus (SARS-CoV-2), a causative agent of SARS-CoV-2 infection, is a positive-sense, single-stranded RNA virus of family Coronaviridae. The SARS-CoV-2 infection has originated from Wuhan, China, and now spreading in different parts of the world causing pandemic state (Huang and others 2020). The World Health Organization (WHO) has declared it as a global health emergency, adversely affecting various countries across the globe. No effective treatment regimen has yet been available; only guidelines issued by WHO suggest for isolation of suspected and confirmed patients, followed by supportive care for severely affected individuals through oxygen, fluid therapy, and antibiotics treatment for secondary infections.

The present line of treatment includes antiviral drugs used against human immunodeficiency virus, antimalarial drugs, convalescent plasma therapy, and few compounds inhibiting the viral replication. One such example is Lopinavir/ritonavir (Kaletra), which are protease inhibitors interfering with viral replication, which are found to show inhibitory activity against Middle East respiratory syndrome coronavirus (MERS-CoV) (de Wilde and others 2014), does not show any benefit to severely ill COVID-19 patients (Cao and others 2020). Apart from this, broad spectrum antiviral drug such as remdesivir, which is known to inhibit viral replication by binding to RNA-dependent-RNA polymerase, does not seem to show much clinical benefits in initial trials, but was able to reduce time required for clinical improvement in patients with SARS-CoV-2 infection (Wang and others 2020). In this queue, cytokine-based therapies also uphold a promising approach. Few reports suggest that interferon gamma (IFN-γ) could also be considered as a treatment strategy because it interferes with viral replication and promotes T cell response (Thijsen and others 2020).

Another important cytokine is interleukin-6 (IL-6), which is a pleotropic cytokine, functions in establishing initial inflammatory response. IL-6 levels are elevated in severely ill COVID-19 patients, therefore considered as a potential biomarker of COVID-19 progression (Ulhaq and Soraya 2020).

Herein, the present review focuses on immunological response in SARS-CoV-2 infection, special focus on IL-6 along with various systemic rationale-based treatment approaches for COVID-19 and their status.
Immunopathology of SARS-CoV-2

The pathophysiology of SARS-CoV-2 infection involves robust inflammatory responses, which cause damage to airways, suggesting that not only infection but also the host response to the infection contributes to disease severity (Naqvi and others 2020; Zhang and others 2020c). High levels of proinflammatory cytokines such as IL-7, G-CSF, IP-10, TNF-α, monocyte chemoattractant protein-1 (MCP-1), and IL-2 in severe cases of COVID-19 have been observed (Jose and Manuel 2020; Zhou and others 2020b). This is called the “cytokine storm.” These results in heightened immune response shown by patients’ body, characterized by acute respiratory distress syndrome (ARDS) observed in severe cases of COVID-19, which includes difficulty in breathing and low levels of oxygen in the blood that may cause respiratory failure leading to death in most of the severe cases. Lymphopenia and pneumonia are also observed in most of the patients with a characteristic CT of chest showing pulmonary ground-glass opacity changes. The cytokine storm is released as an immune response to viral infection, other secondary infection, and symptoms of sepsis resulting in the death of the individual. This increased and uncontrolled inflammation can also cause multiorgan damage especially cardiac, hepatic, and renal (Pedersen and Ho 2020; Zhang and others 2020a). The worldwide mortality rate of COVID-19 is around 2.4%. One of the major cytokine comes into picture is IL-6. It mediates inflammatory responses and is responsible for acute respiratory failure associated with the need for mechanical ventilation and mortality in COVID-19 patients (Herold and others 2020). In the meta-analysis study of COVID-19 patients, including 9 independent studies, a total number of 1,426 patients show that IL-6 levels were >3 times higher in patients with complicated COVID-19 compared with those with uncomplicated disease, and IL-6 levels were associated with mortality risk (Aziz and others 2020). Studies even have considered IL-6 as a prognosticator in patients with COVID-19 on the basis of calculating risk factors for the combined endpoint progression to severe COVID-19 and/or in-hospital mortality (Grifoni and others 2020b).

Innate Response to SARS-CoV-2

Infection caused by the virus followed by lung cell destruction initiates immune response, which recruits macrophages and monocytes, cytokine release, and T and B cell immune response. Not much is known about innate response against COVID-19, except few reports which suggest increased neutrophils, decreased lymphocytes along with elevated levels of IL-6, and C-reactive protein (Prompetchara and others 2020).

As discussed above, the secretion of cytokines and chemokines is increased during infection at the local site and these might cause increased infiltration of monocytes and macrophages, but not of neutrophils from blood to local site of infection (Blanco-Melo and others 2020). Recruitment of immune cells from blood to the pulmonary site and lymphocyte infiltration in the airways might explain the lymphopenia and increased ratio of neutrophil-lymphocyte in

FIG. 1. Immune dysregulation caused by cytokine storm and macrophage-activated syndrome in SARS-CoV-2-infected patients. IFN, interferon; IL, interleukin; SARS-CoV-2, severe acute respiratory syndrome-coronavirus; TNF-α, tumor necrosis factor alpha.
most of the COVID-19 patients (Fu and others 2020). IL-6 is one of the major cytokines released during acute SARS-CoV-2 infection. The exact mechanism of IL-6 functioning in SARS-CoV-2 infection is still unknown, but it has been found to induce hyperinfiltrate inflammatory response during SARS-CoV-1 infection in the respiratory tract (Okabayashi and others 2006). High levels of IL-6 and IL-8 are found to be associated with lung lesions in SARS-CoV-1 patients (Okabayashi and others 2006). The high concentration of IL-6 causes immune dysregulation, which is one of the major characteristics of respiratory failure in SARS-CoV-2 patients (Giamarellos-Bourboulis and others 2020) (Fig. 1).

SARS-CoV-1 has the ability to induce oxidized phospholipid (OxPL) in acute lung injury (ALI). This OxPL shows production of proinflammatory cytokine such as IL-6 through TLR4 activation (Imai and others 2008). Thus, in ALI, IL-6 plays a crucial role during SARS-CoV-1 infection. Further, among the structural protein of SARS-CoV-1, nucleocapsid protein is found to induce IL-6 secretion in epithelial cell, which further contributes to inflammatory response (Zhang and others 2007). From the above discussion, it can be concluded that immunopathological condition of SARS-CoV-2 is similar to SARS-CoV-1 (Fig. 1).

SARS-CoV-2 is a cytopathic virus that causes changes/injury in the virus-infected host cells and tissues that is a part of the replication cycle of the virus. Pyroptosis, which is a highly inflammatory type of cell death, is caused in airway epithelial cells leading to vascular leakage as observed in COVID-19 patients, and IL-1β is also found to be elevated in the SARS-CoV-2 patients (Tay and others 2020). Innate response against viral infection involves IFN type I responses, which are followed by downstream events to control the replication of the virus and further activation of adaptive cell response. SARS-CoV-2 uses ACE2 as entry receptors, which are usually present on alveolar cells of the lungs (Friedman and others 2008).

The generation of innate immune response always requires pathogen-associated molecular patterns (PAMPs). It has been explored for RNA viruses such as coronavirus that viral genomic RNA or intermediate during viral replication such as dsRNA work as PAMPs, and are recognized by TLR3 and TLR7, which are present in the endosomal compartment and RIG-I/MDA5 which are cytosolic RNA sensor (Prompetchara and others 2020). During the normal course of infection, further downstream signaling cascade, including NF-κB and IRF3, is activated, followed by their nuclear translocation and transcription of proinflammatory cytokines in the nuclei, which serve as the first line of defense. Type I IFN is generated further, which activates the JAK/STAT pathway that initiates transcription of IFN-stimulated genes. Initially, this type I IFN response is sufficient to control viral replication, but in the case of SARS-CoV, this response of type I IFN is suppressed (Tay and others 2020).

Since much is not known about this new SARS-CoV-2 virus, the mechanism of how it subverts host innate antiviral mechanism is still not clear. However, its 79% and 50% genomic similarity with SARS-CoV and MERS-CoV, respectively, gives some partial speculations that SARS-CoV-2 utilizes a similar subversion mechanism to dampen the host innate immune response, especially type I IFN (Zhang and Holmes 2020). Along with these, the additional mechanism needs to be explored. Both SARS-CoV and MERS-CoV use various strategies to interfere with the signaling pathway, which directs type I IFN production and further downstream pathway. It has been shown by various researchers that various viral structural and nonstructural proteins subvert IFN response. Sometimes they help in preventing the recognition of PRR of viral RNA and interfere with the downstream signaling of RNA sensors either by a direct or indirect mechanism. These include ubiquitination and degradation of RNA sensor adaptor molecules MAVS and TRAF3/6 along with inhibition of nuclear translocation of IRF3 (Beug and others 2012). Repressive histone modification is one more strategy followed by MERS-CoV along with the above said. Once the type I IFN is secreted, both these viruses have the mechanism to inhibit the IFN signaling, for example, STAT1 phosphorylation is decreased. Both structural and nonstructural proteins are involved in modulating type I IFN response.

Excessive infiltration of inflammatory cells causes damage to the lung by the production of reactive oxygen species and secreting proteases. This along with direct damage caused by viruses results in diffused damage to the alveolar tissue, the formation of the hyaline membrane, leading to pulmonary edema (Xu and others 2020). Histological observation of patients with travel history from Wuhan showed bilateral diffused damage of alveolar tissue along with cellular fibromyxoid exudates. Desquamation of pneumocytes and formation of hyaline membrane indicated for ARDS. Intra-alveolar spaces showed the presence of multinucleated sycntial cells and enlarged pneumocytes with large nuclei and amphophilic granular cytoplasm (Tian and others 2020a). Pathology of 2 patients infected with COVID-19 showed edema, proteineaceous exudates along with hyperplasia of pneumocytes. Infiltration of inflammatory cells and multinucleated giant cells were seen and the hyaline membrane was not prominently observed. These were considered as early pathology of lung in COVID-19 (Tian and others 2020b). Such changes reduce the capacity of gas exchange of lungs leading to a lower level of oxygen in the blood. Septic shock and multiorgan failure are other major consequences of elevated levels of cytokines. An analysis of data of 150 patients in Wuhan, China, showed that these may result in fulminant myocarditis in many patients (Ruan and others 2020). One of the reasons for dysfunctional immune response and disease pathology in older people and people with various comorbidities could be aging in lung, which can cause altered maturation of dendritic cells and thereby defective activation of T cells. At even high viral titer, children do not develop severe disease; it might be because innate immune response is highly effective.

The question which has very controversial observations is if the high viral titer is needed for disease progression. As discussed above, many studies have shown high viral titer even before the onset of symptoms of pneumonia. Another study with 191 patients suggested that in the case of survivors, the median duration for viral shedding was found to be 20 days, however, in nonsurvivor’s, viral load was detectable until death. The longest duration of viral shedding was recorded to be 37 days (Zhou and others 2020a). This suggests a strong correlation between viral persistence and the poor outcome of the disease. Many studies with SARS-CoV have shown that virus has other target cells also except lung cells, such as T lymphocytes, macrophages, and dendritic cells derived from
macrophages. This provides a reason for lymphopenia observed in patients as the virus might directly kill the lymphocytes. The infection of macrophages and other immune cells could cause robust cytokine production. However, the infection mechanism of SARS-CoV-2 in these cells is not defined. This mechanism of immune dysfunction, if deciphered, could help in designing immunomodulatory treatments against the virus.

However, based on these facts, various immunesuppressive therapies are underway against COVID-19 such as the clinical trial of IL-6 antagonist, tocilizumab, targeting GM-CSF (granulocyte macrophage-colony stimulating factor) (Wicks and Roberts 2016), use of gimsilumabs, and so on. (Sanders and others 2020). Another novel therapy under-study is cytoSorbs, which can absorb a broad spectrum of cytokines (Cao 2020). We have discussed all these therapies in detail in upcoming sections.

Role of the Complement System in SARS-CoV-2 Infection

The role of the complement system in innate response against viral infection and their role as initial stimulators of proinflammatory responses are well established (Risitano and others 2020). It has been seen in the case of SARS-CoV, which is closely related to SARS-CoV-2, that the C3 component exacerbates the disease. In a mouse model of SARS-CoV, complement cascade was studied from day 1 of infection in the lungs. They used mice deficient in C3 (C3−/−) for the study. It was observed that C3 knockout mice showed less weight loss and respiratory dysfunction compared to normal despite the same viral load in both the groups. C3 knockout group showed less infiltration of neutrophils and inflammatory monocytes in the lungs compared to the control group along with lower levels of cytokines and chemokines in both serum and lungs of knockout mice. These data suggested the role of complement system activation in SARS-CoV-induced disease and its role in regulating proinflammatory responses. These observations indicate that inhibiting C3 could reduce the inflammatory complications of lungs in SARS-CoV-2 infection (Magro and others 2020).

The cytokines and chemokines such as IL-5, IL-6, CXCl1, and G-CSF, which are known to promote neutrophil recruitment, were found in abundance in control mice compared to C3 knockout mice. This decrease in infiltrating neutrophils and IL-6 suggested that inhibiting IL-6 along with anti-IL-6 therapy could be a potential strategy against SARS-CoV-2 (Wu and Yang 2020; Zhang and others 2020b). As discussed above the complement, C3 activation has been implicated for initial exacerbation of lung injury in SARS-CoV infection, it was speculated that C3-targeted intervention might prove as a strategy to control inflammatory damage mediated by complement activation in COVID-19 patients. A patient of COVID-19 showing severe ARDS was treated with a compstatin-based inhibitor of C3, AMY-101. Drastic improvements were observed in patients after 48 h of initiation of AMY-101 treatment with speedy resolution of inflammatory responses. C-reactive protein and lactate dehydrogenase were normalized progressively with leukocytosis and lymphocytopenia improving slowly (Misra and others 2020; Sarzi-Puttini and others 2020). Oxygen requirements decrease gradually. It is the possibility that inhibition of C3 could block C3a and C5a simultaneously as well as activation of intrapulmonary C3 and release of IL-6 from resident macrophages or other cells, thereby reducing the injury to lungs (Risitano and others 2020).

Another study suggested that the N protein of SARS-CoV-2 binds with MASp-2, which is key serine protease in the lectin pathway of activation of complement. This results in aberrant activation of complement and aggravates injury in the lung due to inflammation. Blocking the interaction of N protein with MASp-2 or suppressing the activation of complement could assuage N protein-induced hyperactivation of complement and thereby reduce lung injury (Zheng and others 2020). When deteriorating patients were treated with an anti-C5a monoclonal antibody, promising suppressive effects were observed, again ascertaining that suppression of complement could be a therapeutic approach against SARS-CoV-2 (Wilk 2020).

T Cell Response

After about 1 week of onset of infection, both T and B cell responses are detected in the blood of COVID-19 patients (Grifoni and others 2020a). It is known that for direct killing of virus CD8+ cells are important and CD4+ T cells are essential for priming of both CD8+ T cells and B cells as well for the production of cytokines for recruiting immune cells. Immunological characteristics of 16 patients admitted to Yunnan Provincial Hospital of Infectious Diseases, Kunming, China, showed T cell exhaustion. It was observed that multiple molecules related to activation and regulation of T cells increased significantly, but functional molecules decreased remarkably. These molecules included decreased levels of IFN-γ and TNF-α in CD4+ T cells in patients with severe disease compared to mildly infected individuals. Moreover, levels of granzyme B and perforin were found to be higher in CD8+ cells of severely infected compared to mild infection (Xiong and others 2020).

No difference in activation molecules in CD4+ T cells was seen, however, in CD8+ T cells, levels of Human Leukocyte Antigen-DR isotype and T cell immunoreceptor with Ig and ITIM domains were found to be higher in individuals of the highly infected group compared with mildly infected group. Data suggested that similar to few other chronic infections, COVID-19 causes T cell exhaustion and excessive activation followed by probable exhaustion of CD8+ cells. The perturbation of T cell response might eventually lead to diminished immunity of host against the virus (Xiong and others 2020). First autopsy of a COVID-19 patient revealed that the count of hyperactive CD4+ and CD8+ T cells decreased substantially in peripheral blood. The increased concentration of CCR6+ Th17 in CD4 T cells and CD8 T cells showed a higher concentration of cytotoxic granules. This suggested that increased Th17 cells and high cytotoxicity of CD8+ T cells could be partially responsible for the immune injury of the patient. It was observed in SARS-CoV infection that virus-specific memory T cells against the spike (S) protein were found for a long-time postinfection. Most of the CD4+ T cells were central memory CD45RO+(+) CCR7+(+) CD62L(−) cells, and CD8+ T cells were CD45RO+(+) CCR7+(+) CD62L(−) effector memory cells (Yang and others 2006).

Compared to non-ICU patients, ICU patients have even higher levels of IL-2, IL-7, IL-10, G-CSF, IP-10, MCP-1, MIP1A, and TNF-α.1 (Huang and others 2020). Various cytokines from among these are involved in TH17 type
responses. For example, IL-1β and TNF-α both promote TH17 responses and vascular permeability and leakage. TH17 cells majorly produce IL-17, which has broad proinflammatory effects responsible for granulopoiesis and recruitment of neutrophils, IL-6, IL-1β, and TNF-α responsible for systemic inflammatory symptoms, such as fever (Wu and Yang 2020). It is observed that MERS-CoV and SARS-CoV patients show enhanced IL-17-related pathways (Faure and others 2014). Apart from this, Pandemic H1N1 influenza virus also induces strong TH17 (and TH1) responses (Bermejo-Martin and others 2009). Thus, we can conclude that TH17 type response contributes to the cytokine storm in pulmonary viral infection, including SARS-CoV-2.

T cells specific to SARS-CoV expressed IFN-γ, TNF-α suggesting induction of Th1 type protective cellular response in the patients (Prompetchara and others 2020). Various animal studies have shown that immunization of mice resulted in an increased number of CD4+ and CD8+ cells specific against viruses in the lungs with increased survival (Qin and others 2020). It was also found that the adoptive transfer of splenocytes from mice immune to SARS-CoV or T cells generated under the in vitro condition in immunodeficient mice resulted in increased survival and reduced titer of virus in the lungs. These studies indicated the protective role of T cells in SARS-CoV clearance and a rationale for vaccine design. However, vaccination studies also showed some controversial data. A vaccine formulation when tested in the animal model showed a remarkable increase in both Th1 and Th2 response along with downregulation of IL-10 and TGF-β. This resulted in the infiltration of neutrophils, eosinophils, and lymphocytes along with the thickening of alveolar epithelium increasing the severity of pneumonia. This suggests that a fine balance between protective and detrimental T cell responses must be studied more intensely for vaccination strategies. Moreover, without results in humans, nothing concrete could be concluded.

B Cell Response

B cell response in COVID-19 patients initiates nearly a week after the onset of symptoms. In the case of SARS-CoV infection, antibodies were detected as early as 4 days after the onset of disease and many patients showed the presence of antibodies by day 14 (Kellam and Barclay 2020). Neutralizing antibodies were reported even after 2 years of infection. In the case of MERS-CoV infection, delayed seroconversion is seen, usually in the second or third week of infection. Very less serological data are available for SARS-CoV-2 infection. In a study for doing serology, nucleocapsid (N) protein from bat SARS-CoV Rp3 was taken as antigen for IgG and IgM, since this protein was found to share 92% homology in amino acid sequence to N protein of novel COVID-19 and was specific in cross-reactivity to SARS-CoV (Thevarajan and others 2020). The presence of antibodies was detected in <40% among patients within 1 week from onset and rapidly increased to 100% (total antibodies), 94.3% (IgM) and 79.8% (IgG) from day-15 after onset (Zhao and others 2020).

IgG and IgM titer were found to increase with time except that in one last sample, IgM was decreased. In another analysis, a serum sample of 5 patients was evaluated for the presence of viral antibody day 20 postinfection, and it was recorded that all the samples showed strong seropositivity for viral IgG and 3 for IgM. To ascertain the neutralizing effect of viral IgG+ serum samples, neutralization assay was conducted in Vero-E6 cells using a serum sample of 5 patients, and it showed that all samples neutralized 100 TCID₅₀ of COVID-19. This suggested that a humoral response could have a positive outcome in the case of SARS-CoV-2 infection. (Guo and others 2020).

Role of Neutrophils and Dendritic Cell

Neutrophils are the abundant polymorphonuclear phagocytic cells that function in innate immune system. The in vitro studies of COVID-19 patient sera reveal the induction of neutrophil extracellular traps (NETs). The NETs are found to be linked with obstructive ALI and ARDS (Laridan and others 2019; Zuo and others 2021). This clearly indicates the role of neutrophils in pathogenesis of COVID-19. Further, the ScRNA analysis of bronchoalveolar lavage fluid of severely and mild symptomatic COVID-19 patients reveals the higher percentage of neutrophils in severe condition than in mild condition (Liao and others 2020). Another study of 82 patients revealed similar results. Approximately, 74.3% of the patients showed neutrophilia, whereas 94.5% of the patients showed higher neutrophil to lymphocyte ratio (Zhang and others 2020). Neutrophil activation is identified as an indicator of COVID-19 severity and the activated neutrophils express hepatocyte growth factor, resistin, lipocalin-2, IL-8, and G-CSF (Mathew and others 2020). These cytokines, chemokines, and receptor are responsible for aggravated inflammatory response in severe COVID-19 patients.

Dendritic cell also plays a pivotal role in causing severity of COVID-19. They are professional antigen presenting cell function in eliminating viral infection. They migrate to the site of infection and get activated. Once they are activated, they move toward lymphoid organ to activate T cells for specific immune response. One of the reports on SARS-CoV shows the induction of plasmacytoid dendritic cells (pDCs) that activates the production of type I IFNs. This causes the accumulation of large number of monocytes, macrophages, and neutrophils at the site of infection, which eventually causes immune exhaustion. In case of infection, the depletion of pDCs and conventional DCs has been observed (Wilk and others 2020). Apart from that, delayed activation and migration of DCs from lungs to lymph nodes has also been observed. This may correlate with the slow clearance of viral infection in severe COVID-19 patients (Brufsky and others 2020).

The Correlation of Host Genetic Composition with Severity of COVID-19 Disease

One of the researchers has identified major genetic risk factor associated with the severity of the disease. They have identified 2 genomic regions: first is a cluster of 6 gene which is present on chromosome 3, and second region is present on chromosome 9 that determines ABO blood groups (Zeberg and Pääbo 2020). The evolutionary studies showed that these segments are conserved and has been inherited from Neanderthals. They are carried by 50% population of South Asia and 16% population of Europe (Zeberg and Pääbo 2020). Similarly, transcriptomic studies of lung tissue revealed that high expression of the monocyte–macrophage chemotactic receptor CCR2 is associated with severe COVID-19 disease (Paiaro-Castineira and others 2021). These studies indicate that
the genetic composition of host plays a huge role in shaping the disease outcome and can be studied further to reveal the mechanism of disease severity.

Potential Anti-inflammatory Therapies

Targeting major cytokine causing cytokine storm can be a potential base for designing therapeutics against this novel SARS-CoV-2 virus. The level of IL-6, a potential proinflammatory cytokine, is found to be elevated in serum samples of infected patients and is correlated with respiratory failure (Herold and others 2020). IL-6, through its trans form of signaling, is found to target many somatic cell types such as endothelial cells, B cell, T cells, and so on. Its effect on endothelial cells would result in expression of vascular endothelial growth factor (VEGF), MCP-1, and IL-8; further secretion of VEGF and reduction of E-cadherin expression contribute to vascular permeability and leakage, which participate in the pathophysiology of hypotension and pulmonary dysfunction in ARDS (Magro 2020). The potential target strategy is to block IL-6 signaling through Tocilizumab, which is currently been given to ARDS patients. Another strategy is to target specific proinflammatory pathway through SGP130Fc. Other than this, potential inhibitors of IL-6 pathway are ruxolitinib and baricitinib (Virtanen and others 2019). Despite of the higher hopes with the inhibition strategy of IL-6 signaling in SARS-CoV-2-infected patients, trial results of tocilizumab are damping enthusiasm. This antibody against IL-6R fails in phase III trial conducted on 450 patients with severe COVID-associated pneumonia (Mullard 2020). Apart from this fedratinib (SAR302503, TG101348), a JAK2-specific inhibitor approved by FDA for myeloproliferative neoplasms, suppress the production of several TH17 cytokines (and likely to affect IL-6 cytokine functioning as well), therefore very promising approach to prevent the deteriorating outcomes of TH17-associated cytokine storm in COVID-19 (Wu and Yang 2020). Another drug in this list could be etoposide, which is used to deplete monocytes and suppress cytokine release in hemophagocytic lymphohistiocytosis. Etoposide can be potential drug to combat the inflammatory monocyte–macrophage immune sensing responses involved in causing lethal pneumonia in COVID-19 patients (Channappanavar and others 2016).

WHO is working in collaboration with scientists, industries, and global health organizations through the ACT Accelerator to develop therapeutic strategies to treat SARS-CoV-2 infection. There are currently over 169 COVID-19 vaccine candidates under development, with 26 of these in the human trial phase (Table 1). A new organization led by WHO, GAVI, and CEPI called as COVAX will facilitate the equitable access and distribution of these vaccines to treat COVID-19 patients all over the world (WHO 2020).

Potential Target for Vaccination and Their Challenges

The present line of therapies includes anti-inflammatory and antiviral therapies, which are only supportive in nature with limited efficacy, moreover, they are unable to prevent viral replication and its transmission. Effective vaccination here plays a crucial role in establishing host acquired immune response, which neutralizes viral particles and inhibits its transmission. In April 2020, vaccine tracker was launched by London School of Hygiene & Tropical Medicine to follow vaccine candidates as they progress through the developmental pipeline (Parker and others 2020). As of March 2021, the total number of vaccine candidate in phase trials are 24 in phase I trial, 7 in phase II trial, 16 in Phase III trial, and 4 vaccine candidate in phase IV trial (London School of Hygiene & Tropical Medicine 2020). Till date, total 11 vaccines are authorized by National Regulatory Authority for public use (London School of Hygiene & Tropical Medicine 2020). There are various forms of vaccines under preclinical stage or in phase trial stage, which has been discussed below:

(a) Inactivated vaccines: In this case, the virus is modified in such a way that it can raise immune response, but unable to multiply or transmit. The virus particles are chemically modified to inactivate by the use of formaldehyde or beta-propiolactone (Delrue and others 2012). To generate sufficient amount of antibody titer, they are either required to be given in multiple doses or given with adjuvants (Strizova and others 2021). But due to regressive mutation in present COVID strains, the efficacy of inactivated vaccines is suspected to be low. Till date, there are 2 inactivated vaccines which have reached to phase III trial and are under evaluation, namely NCT04508075 and NCT04456595 (Strizova and others 2021).

(b) DNA vaccines: The basic principle of DNA vaccines is the transcription and translation of viral protein in host cell to generate acquired immunity. For the said purpose, the DNA vaccine often requires a vector/carrier (eg, Plasmid) to transfer the target gene into host nucleus. DNA vaccine either targets the monocytes or the keratinocytes depending upon their way of administration, that is, intradermal, intramuscular, or subcutaneous. The target protein is either expressed through MHC I/MHC II or secreted through exosomes where it is recognized by antigen presenting cells to generate immunity. One of the key features that needs to be considered during the construction of DNA vaccine is that the vector needs to have low tendency of chromosomal integration to avoid mutation and dysregulation of native gene function. N and S proteins were used for DNA vaccination against SARS-CoV-2 infection (Silveira and others 2021). Till date, there are NCT04673149, NCT04591184, NCT04445389, NCT04715997, and NCT04447781 in phase I/II trial.

(c) RNA vaccines: The mRNA vaccines include plasmid DNA having mRNA that encodes for SARS-CoV-2 viral protein that eventually activates hosts humoral as well as cellular immunity. The limitation of the mRNA vaccine is the instability of mRNA in host environment. The conventional mRNA vaccine involves the expression of target protein in the cytoplasm of host cell, whereas the advance mRNA vaccine includes RNA-dependent RNA polymerase gene for self-amplification of the vaccine. The advantages of mRNA vaccine include inability to incorporate into host genome (Zhang and others 2019). The mRNA vaccine under phase I/II trial includes NCT04728347, NCT04480957, NCT04380701, and NCT04839315; phase III trial includes NCT04816643, NCT04800133, and NCT04368728; and phase IV trial includes NCT04760132.
# Table 1. List of Few Vaccines Against COVID-19 in Progress with the Current Status

<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Company</th>
<th>Phase</th>
<th>Trial No.</th>
<th>Status</th>
</tr>
</thead>
</table>
| BCG live-attenuated vaccine                       | University of Melbourne and Murdoch Children’s Research Institute (Australia); Radboud University Medical Center (The Netherlands); Faustman Lab at Massachusetts General Hospital (United States) | III    | NCT04327206, NCT04328441       | (1) The randomized, controlled, phase III BRACE trial in Australia: Recruitment ongoing  
(2) Netherlands phase III trial: Recruitment planned  
(3) The Faustman Lab: Efforts to get funding for BCG trial in health care workers to prevent COVID-19 |
| Absorbed COVID-19 (inactivated)                   | Butantan Institute                           | III    | NCT04456595                     | Randomized double blinded, Placebo controlled trials are going on for the age group of (18–59 years). |
| SARS-CoV-2 (inactivated)                          | PT Bio Farma                                  | III    | NCT04508075                     | Randomized, observer blinded, Placebo controlled trials are going on for the age group of (18–59 years). |
| mRNA-1273                                         | Moderna                                      | I      | NCT04283461                     | Recruitment finished. Company received grant from the BARDA. Application submitted for phase II to FDA |
| Ad5-nCoV [recombinant vaccine with the adenovirus type 5 vector (Ad5)] | CanSino Biologics; Tongji Hospital; Wuhan, China | Phase I safety data allowed to start phase II | NCT04313127                      | Initial safety data from phase I, company announced phase II    |
| ChAdOx1 (a new vaccine candidate for COVID-19, a chimpanzee adenovirus vaccine vector called ChAdOx1.) | The University of Oxford                     | A phase I/II single-blinded, multicenter study | NCT04324606                      | On 24 April, human testing investigation on 1,100 people initiation reported |
| INO-4800                                           | Inovio Pharmaceuticals                        | A nonrandomized, open label phase I trial Phase I/II parallel trial to be initiated in South Korea alongside with the U.S. trial. | NCT04336410                      | April 16, Inovio and The International Vaccine Institute+Korea National Institute of Health announced phase I/II clinical trial in South Korea. 40 people recruited on April 28. |
| Formalin-inactivated and alum-adjuvant candidate vaccine | Sinovac                                      | A randomized controlled phase I trial of 144 healthy participant | NCT04352608                      | Phase I randomized control trial initiated |

(continued)
<table>
<thead>
<tr>
<th>Vaccine</th>
<th>Company</th>
<th>Phase</th>
<th>Trial No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Novavax</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVX-CoV2373 (full length recombinant SARS-CoV-2 glycoprotein nanoparticle vaccine)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Janssen Pharmaceutical Companies</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Imperial College London</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Medicago</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>ZydusCadila</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Maken Biotech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Takis Biotech</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Altimmune</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Johns Hopkins University</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Osaka University/RIKEN/National Institutes of Biomedical Innovation, Japan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>National Institute of Infectious Disease, Japan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>EpiVax/Univ. of Georgia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Biological E Ltd.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>Clover Biopharmaceuticals Inc.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>WRAIR/USAMRIID</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA vaccine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preclinical vaccine candidates</td>
<td>AdaptVac (PREVENT-nCoV consortium)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RBD, receptor binding domain; SARS-CoV-2, severe acute respiratory syndrome-coronavirus.

Table 1. (Continued)
As per the Centre for Disease Control and Prevention, the sequencing of SARS-CoV-2 virus has enabled the researchers to track the spread of various lineages all across the world. The global initiative of widespread whole-genome sequencing of SARS-CoV-2 virus has enabled the researchers on Vaccine Development.

SARS-CoV-2 Mutant Strains and Its Effect on Vaccine Development

The global initiative of widespread whole-genome sequencing of SARS-CoV-2 virus has enabled the researchers to track the spread of various lineages all across the world. As per the Centre for Disease Control and Prevention, there are 5 variants of concern (VOCs) for COVID-19 being identified in United States (Centers for Disease Control and Prevention 2021):

- B.1.1.7: This variant was first identified in the United States in December 2020, which was initially detected in the United Kingdom.
- B.1.351: This variant was first identified in the United States at the end of January 2021, which was initially detected in South Africa in December 2020.
- P.1: This variant was first detected in the United States in January 2021. P.1 was initially identified in travelers from Brazil, who were tested during routine screening at an airport in Japan, in early January.
- B.1.427 and B.1.429: These 2 variants were first identified in California in February 2021 and were classified as VOCs in March 2021.

The present line of viral vaccines are either live attenuated (eg, measles) or the inactivated form (eg, hepatitis A). Both these categories can give rise to a polyclonal response characterized by prolong B cell and T cell immunity against not just single viral antigen, but whole virus itself. The longevity of the immune response generated determines the efficacy of the developed vaccine. Exception to this is the vaccines against influenza virus where genome variability (in terms of antigenic shift and antigenic drift) plays a crucial role and on the basis of that reformulation of vaccine occurs periodically.

SARS-CoV-2 shows low mutation rate than that of the other RNA viruses (dos Santos 2021), and till now there is no such evidence of antigenic drift that has completely changed the antigen genotype. Overall, it is possible to come up with high-efficacy vaccine until and unless SARS-CoV-2 does not show the evolutionary pattern similar to Influenza virus. Advancement in genome variability studies could possibly provide a better insight into its relationship with host immune response in near future.

Conclusions

The pathogenesis of SAR-CoV-2 has been reviewed extensively with context to its immunological responses possibly involved in establishing the infection. The role of cytokine storm in establishing hyperinflammatory responses has been explained in detailed to correlate it with the disease severity. The present line of antiviral treatment involve killing of virus by interfering with its replication machinery. This treatment does not show clinically significant results as alone, but may be in combination with cytokine therapy, it could reduce the infectivity of the virus along with reducing the hyperinflammatory responses. Thus, combinational therapy would be the best approach suited in the present scenario to treat SARS-CoV-2-infected patients. Apart from this, various SAR-CoV-2 viral proteins have been used to develop vaccines, which are currently under clinical trials. Spike membrane protein is found to induce effective B cell as well as T cell responses, which makes it a potential target for vaccine development. One such example is the ongoing phase III trial of an adenovirus-vectorized spike-based vaccine (NCT04505722; Johnson & Johnson). Since there is not much evidence regarding the rate of mutation in spike protein of SARS-CoV-2 virus, it is early to comment on the efficacy of the vaccines under development. With the
adancement in SARS-CoV-2 evolution studies, we would be able to formulate a vaccine with better efficacy or periodically reformulate the vaccine as per the emerging mutation in targeted antigen.

Authors’ Contributions

All authors have contributed to the article. B.S. contributed in writing and editing the original draft and designing of figures and tables. R.K. contributed in writing and editing the original draft. S.S. contributed in initialization, supervision, writing, and editing of original draft. All authors agree with the submission of the final version of the article.

Acknowledgments

We are thankful to the Director, National Centre for Cell Science (NCCS) for supporting the Bioinformatics and High-Performance Computing Facility at NCCS, India.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

Intramural Funding (Dr. Shailza Singh), National Centre for Cell Science, India.

References


Address correspondence to:
Dr. Shaiiza Singh
National Centre for Cell Science
NCCS Complex
Ganeshkhind
SP Pune University Campus
Pune 411007
India

E-mail: singhs@nccs.res.in

Received 23 October 2020/Accepted 28 June 2021
Evolutionary artificial intelligence based peptide discoveries for effective Covid-19 therapeutics

Ritika Kabra, Shailza Singh

National Centre for Cell Science, NCCS Complex, Ganeshkhind, SP Pune University Campus, Pune 411007, India

ARTICLE INFO

Keywords:
Covid-19
Artificial intelligence
Evolutionary peptides
Computational biology
SARS-CoV-2

ABSTRACT

An epidemic caused by COVID-19 in China turned into pandemic within a short duration affecting countries worldwide. Researchers and companies around the world are working on all the possible strategies to develop a curative or preventive strategy for the same, which includes vaccine development, drug repurposing, plasma therapy, and drug discovery based on Artificial intelligence. Therapeutic approaches based on Computational biology and Machine-learning algorithms are specially considered, with a view that these could provide a fast and accurate outcome in the present scenario. As an effort towards developing possible therapeutics for COVID-19, we have used machine-learning algorithms for the generation of alignment kernels from diverse viral sequences of Covid-19 reported from India, China, Italy and USA. Using these diverse sequences we have identified the conserved motifs and subsequently a peptide library was designed against them. Of these, 4 peptides have shown strong binding affinity against the main protease of SARS-CoV-2 (M\text{pro}) and also maintained their stability and specificity under physiological conditions as observed through MD Simulations. Our data suggest that these evolutionary peptides against COVID-19 if found effective may provide cross-protection against diverse Covid-19 variants.

1. Introduction

A novel positive-sense, single-stranded RNA virus of family Coronaviridae named SARS-CoV-2 has become a serious threat to life emerging from China and spreading worldwide [1]. It has been declared a global health emergency by the World health organization (WHO). No effective treatment regimen is available against this virus. Affected nations are trying all the possible strategies for treatment and prevention of the spread of this deadly infection. Guidelines of WHO, along with a suggestion for isolation of suspected and confirmed patients, recommended supportive care for individuals infected with SARS-CoV-2 is by oxygen, fluid therapy and for secondary infections, antibiotics have been suggested.

The major ongoing treatments include antiviral drugs used against human immunodeficiency virus (HIV), anti-malarial drugs, and few compounds preventing the viral replication and convalescent plasma [2–4]. Anti-viral therapy included Lopinavir/ritonavir (Kaletra) which are protease inhibitors interfering with the replication and synthesis of HIV. Few reports suggest that IFN-\text{γ} could also be considered as a treatment strategy because it interferes with viral replication and promotes both arms of immune response i.e. Innate and adaptive. Cumulatively, we can say that the treatment strategies adopted against Covid-19 are going on three fronts viz., drug repurposing, designing new drugs, and ongoing studies for developing vaccines against Covid-19 [5].

Herein, our efforts are to report epidemiology of the disease, various systemic treatment methodologies under pursue and our Artificial Intelligence rationale-based treatment approach for Covid-19 and its therapeutics.

1.1. Epidemiology of COVID-19

Family of coronavirus including SARS-2003, MERS-2012, and COVID-19 has affected a larger population in the last 2 decades. COVID-19 caused by SARS-CoV-2 which started from Wuhan in China has spread rapidly in many countries around the globe, turning into a
deadly pandemic and more than three million confirmed cases have been reported [6,7]. After the initiation of disease at the end of 2019, nearly 84,900 cases have been reported from China, Wuhan being the epicenter. It was jointly reported by the World Health Organization (WHO)-China fact-finding mission that the peak of the epidemic was observed in January and February 2020 and the emergence of new cases decreased by March. Although by the end of March; 938,113 total cases and 44,779 total deaths have been reported all over the world. Presently, data confirms cases from all the continents except Antarctica. After China, Italy was the second country to be affected badly by Covid-19 with a very high fatality ratio to the number of cases closely followed by USA, India and Brazil. The case fatality rate of USA was found to be highest among all the countries with more than 195,239 deaths till September 9, 2020. All over the world, 908,000 deaths have been reported as of September 9, 2020. All 50 states of the United States have been affected by the pandemic. India has also reported nearly 75,091 deaths with 87,700 average new cases every day (September 9, 2020). The total number of cases has been more than 4.46 million. Johns Hopkins data reports 6.54 million cases in the USA as of September 9, 2020.

Initially, when the disease originated, the transmission was found to be associated with the seafood market which sells live animals in Wuhan [8]. However, after the progression of the disease, the main transmission mode is found to be individual-to-individual transmission, which is the reason social-distancing is being followed by affected countries very strictly. This individual to the individual transmission of SARS-CoV-2 usually occurs by respiratory droplets, similar to influenza when the infected individual coughs, sneezes, or talks in close contact with other person directly reaching the mucous membrane [9]. A study was conducted using experimentally generated aerosol to evaluate the stability of SARS-CoV-2 in aerosols using a Bayesian regression model for the decay rate. They found that SARS-CoV-2 could exist in aerosols for 3 h. They also suggested that the stability of the virus was more on plastic and stainless steel than on copper and cardboard, and virus although low titer was found nearly up to 72 h on these surfaces [2].

A study based on data from two hospitals in Wuhan found that the isolation wards and ventilated patient room had very less viral RNA and was found to be highly increased in the toilet area. Moreover, the area of medical staff was also found to be initially having a high concentration of viral RNA which was later on reduced by rigorous sanitization [10]. It has been suggested by few studies that these respiratory droplets could be carried in a gas cloud and their horizontal trajectories were estimated up to 6 ft. Overall studies suggest that recommendations on airborne settings could vary but precautions for avoiding airborne transmission must be universally followed, mainly where the generation of the aerosol is possible [11].

Along with respiratory samples, the presence of SARS-CoV-2 has also been studied in various other samples such as stool, blood, ocular secretions. Real-time PCR was conducted to analyze the presence of the virus in the pharyngeal swab, blood, and anal swab. Data indicated the presence of the virus in both blood and anal swab. Data confirmed that viral RNA could be present in extra-pulmonary sites [12,13].

Another study with 205 patients found live viruses in feces which suggested that the fecal route could also be a mode of viral transmission [14]. A 65 years old female having a travel history from Wuhan to Italy represented with nonproductive cough, sore throat, coryza, and bilateral conjunctivitis. Along with nasal swab, her ocular swab was also collected due to persistent conjunctivitis and viral RNA was detected in the ocular swab also. Surprisingly, SARS-CoV-2 RNA was detected in ocular swabs even after it was not detected in nasal swabs. This further strengthens the need for control measures like avoiding touching face, eyes, and nose. However, more studies need to confirm fecal-oral transmission [15].

The other major issue is for long the individual once infected with Covid-19 is infectious. It is speculated that the transmission of the virus could occur even before the symptoms appear as well as during disease.

In 17 patients after onset of symptoms, a high viral load was detected in the nose as compared to the throat. It was suggested that nucleic acid shedding of the virus was similar to the influenza virus. Moreover, the viral load was similar in asymptomatic and symptomatic patients suggesting that asymptomatic patients have the equal potential of transmission and transmission could occur in the early stage of infection [16]. A cohort study conducted in two hospitals in Hong Kong suggested that viral load was initially high during the first week after the symptoms appear but declined subsequently with time, even in one patient viral load was detected after 25 days of onset of symptoms [17]. These suggest that viral RNA levels are higher in the upper respiratory tract after the onset of symptoms as compared to later illness [18].

Clinical data of viral shedding was compared with incubation period, epidemiological data and interval between cases in the transmission chain for a conclusive inference on the profile of infectivity. The group reported a temporal pattern of viral shedding in 94 patients with confirmed Covid-19 infection and the infectiousness profile of Covid-19 was modeled from a separate 77 transmission pairs. The highest viral load was found in throat swabs at the onset of symptoms and it was inferred that infectivity reached very high before the onset of symptoms. It was calculated that nearly 44% of secondary cases got an infection during the pre-symptomatic stage of index cases [19]. Most of these studies suggest that the infected individual has a higher possibility of causing transmission at the initial stage, which needs to be further validated.

1.2. Life cycle of SARS-CoV-2

The exact life cycle and disease proliferation of SARS-CoV-2 is yet to be deciphered, but due to its similarity with other coronaviruses, like SARS-CoV and MERS-CoV, its believed to share similar life cycle. The replication of coronavirus occurs inside host cell and different stages of its life cycle are shown in Fig. 1a. As previously mentioned, coronavirus is a positive-sense RNA virus with a genome size of ~30 kb which acts as mRNA for translation of replicase polyproteins. Of this genome, only one-third encodes for structural and accessory protein, while the rest genome encodes for non-structural proteins. The virus comprises of mainly four structural proteins; S (Spike) protein, E (Envelope) protein, M (Membrane) protein and N (Nucleocapsid) protein [20]. The S protein initiates the attachment of virion with the receptor angiotensin-converting enzyme 2 (ACE2) of host cell. Upon entering the host cell, S protein is cleaved by TMPRSS2 protease exposing fusion peptide, which in turn is responsible for binding of cellular membranes and subsequently release of viral RNA into cytoplasm [21,22].

The viral RNA released in cytosol now translates the replicase gene encoding two ORFs, rep1a and rep1b, which further expresses two overlapping polyproteins, pp1a and pp1ab. Coronavirus family usually has 2–3 proteases for cleaving these polyproteins. They may be papain like protease (PLpro) or 3-chymotrypsin-like protease (3CLpro). 3CLpro is the main protease, also known as Mpro cleaving at 11 distinct sites (Fig. 1b). Post processing, these polyproteins assemble at replicase – transcriptase complex (RTC), serving as negative strand RNA for the actual replication and transcription of positive strand viral RNA [23–25]. The newly transcribed positive strand mRNAs expresses the structural and non-structural accessory proteins including viral proteins. These viral proteins translocate to a zone between ER and Golgi apparatus that acts as an assembly site for new virions. Post maturation, new virions bud off from golgi apparatus as vesicles and are released from the host cell [26,27].

1.3. Therapies against Covid-19

The major issue for urgently designing therapies against Covid-19 is no background data available. The reason being, SARS-CoV-2 the causative organism of the disease was never heard and studied for any type of therapeutic strategy [28]. However, the genetic sequence of SARS-
CoV-2 was deciphered in January 2020 in a very fast manner, which triggered scientists around the world to identify various methodologies to fight it, which is the need of time visualizing the scenario in the world, due to the pandemic. All researchers are targeting one of the four major pathways to block the spread of disease, i.e., inhibiting the interaction of virus with human cell receptors, or counteracting the synthesis and replication of viral RNA, or re-establishing and boosting host innate immunity, or inhibiting host specific receptors or enzyme [29].

1.3.1. Vaccine technologies deployed against Covid-19

A successful vaccine is essential for the prevention of future infections and mortalities [30]. This will not only reduce pressure on the healthcare system but also on the economy, which is badly affected by
this pandemic. Various vaccination strategies including whole virus vaccine, subunit vaccine, nucleic acid vaccines are in pipeline initiated by multinational companies like Johnson and Johnson along with research organizations in many nations. Johnson and Johnson in collaboration with Beth Israel Deaconess Medical Center, part of Harvard Medical School initiated their efforts in January 2020 once the sequence of the virus was deciphered. Their vaccines are based on Janssen’s proven AdVac® and PER.C6® technology which the company has used to design Zika, RSV, and HIV vaccines which are mostly in Phase2 or Phase 3 trials. Its estimated availability is in early 2021 [31]. A live-attenuated vaccine has been developed by Serum Institute of India in partnership with US-based biotech drug research company Codagenix against SARS-CoV-2 and is under the pre-clinical trial phase [32].

Various other strategies, based on inducing neutralizing antibodies against surface-exposed spike (S) glycoprotein or S protein are underway to develop the Covid-19 vaccine [33]. Most of these spike protein vaccines are based on the rationale of inducing a robust immunological response against the spike protein to avoid its interaction with the ACE2 receptor. Full length S protein or S1-receptor binding domain is being expressed in viral-like particles, DNA, or any other suitable expression vector. Clover bio-pharmaceuticals, a clinical-stage, Biotechnology Company based on research, are in process to develop a recombinant subunit vaccine for SARS-CoV-2. The company is using its patented Trimer-Tag® technology for designing a subunit vaccine, which is the S protein subunit-trimer vaccine (S-Trimer) and the expression system used is rapid mammalian cell-culture based expression system. Once a successful vaccine candidate is obtained, Clover can speed up the production of vaccines as they have in-house cGMP biomanufacturing capabilities [3]. Texas Children’s Hospital Center for Vaccine Development at Baylor College of Medicine has been working on the development and testing of a subunit vaccine which comprises only the receptor-binding domain (RBD) of the SARS-CoV S-protein. This antigen has undergone cGMP manufacture and therefore can undergo clinical trials soon. It was found that this vaccine candidate did not induce significant eosinophilia or eosinophilic lung pathology in the mouse challenge model supporting its safety [34].

Texas Children’s Hospital has developed one vaccine based on yeast-derived (Pichia pastoris) recombinant protein comprised of the receptor-binding domain (RBD) of the SARS-CoV which has been formulated on alum is termed as SARS-CoV 219-N1. This vaccine may be a potential heterologous vaccine. The Australian group has developed a vaccine that includes synthetic viral proteins and an additional stabilization domain called the molecular clamp. This is based on the idea that synthetic proteins to be held in the conformation and viral fusion proteins be maintained before they are merged with the host cells. This molecular clamp was used so that vaccines could be recognized by the host immune system more efficiently [35].

Novavax has also identified a SARS-CoV-2 Recombinant Spike Protein Nanoparticle vaccine, NVX-CoV2373 which has shown high immunogenicity and neutralizing antibodies were produced in high levels, as the pre-clinical studies data indicates. The nano-particle technology used is the propriety of Novavax. In the animal studies wherein, they were immunized with one dose, a high level of antibodies specific to spike protein with Angiotensin-converting enzyme 2 (ACE-2) human receptor binding domain blocking activity and neutralizing antibodies for SARS-CoV-2 wild-type virus were observed. Eight folds increase in antibody titer was observed after the second immunization dose [36].

This is the first vaccine against COVID-19 which would be tested outside of the USA, and as for clinical trials are concerned, it is third in the world, conducted by Australia’s largest Phase 1 clinical trials specialist Nucleus Network. Nucleus will commence its Phase 1 trial in Melbourne and Brisbane clinics in the coming weeks. Inovio is also using nucleic acid-based technology but instead of RNA, they are using DNA as a base for the vaccine. Their concept also involves the ability of the body to produce proteins that mimic the SARS-CoV-2 spike protein. Their vaccine relies on electroporation of plasmid DNA into human cells. Phase 1 trial is in pipeline for this candidate vaccine [37].

Various companies based on biotech research are using advanced nucleic acid-based vaccine platform for Covid-19. Various companies like Inovio Pharmaceuticals, Moderna therapeutics are involved using this strategy. Moderna has come up with an mRNA-1273 based platform for the vaccine in close association with the National Institute of Health. These RNAs when administered will direct the cells of the recipient to produce protein mimicking “spike protein” and elicit an immune against the Covid-19 virus. They are about to initiate Phase 1 clinical trials.

The above data suggests that there are many vaccines in the pipeline with different strategies involved. However, there is a lengthy process of phase trials including manufacturing and testing toxicity. The coming few months are very critical for determining the fate of this pandemic. We urgently require vaccines and even FDA or other regulatory bodies approve it, after evaluating the data of clinical trials for safety and efficacy, another major hurdle to prevail would be to generate a bulk supply of the candidate vaccine to meet the need of the large population.

1.3.2. Passive antibody therapy

The general strategy in this therapy is to draw blood from an individual who was sick with COVID-19 and has recovered for the screening of virus-neutralizing antibodies [38,39]. The samples in which neutralizing antibodies are found in high levels; these serum samples with virus-neutralizing antibodies could be administered prophylactically in individuals at high risk such as health care professionals, close contacts of confirmed cases of COVID-19 patients and to individuals who are having vulnerable health conditions. Trials to use the convalescent serum for patients also have been initiated, to reduce the symptoms and mortality. In a small clinical study conducted in China recruiting 10 severely infected patients and it was found that one dose of convalescent plasma was well tolerable by patients and neutralizing antibodies levels were significantly increased or levels were maintained. Viremia was found to disappear after 7 days with improvement in clinical and paraclinical symptoms suggesting its potential efficacy to treat patients [40,41].

FDA has initiated guidelines for using convalescent plasma therapy and Methodist Hospital in Houston recruited plasma donors a few days back and initiated the first plasma transfusions to COVID-19 patient the following day [42–44]. The Drug Controller General of India (DCGI) has also approved clinical trials using convalescent plasma therapy. According to The Print, the US has started a clinical trial for evaluating the efficacy of gimsilumab which is an artificially synthesized monoclonal antibody and targets granulocyte macrophage-colony stimulating factor, a key driver in lung hypertension [45,46].

1.3.3. Drug based therapeutic approaches

This arm of treatment against COVID-19 is bifurcated into two strategies, one is drug repurposing and other is search for novel drug molecules based on structure-based drug designing using computational biology and Artificial intelligence. FDA and many other big organizations are insisting researchers for long, for the repurposing of drugs. Broad-spectrum antiviral agents (BSAs), which are small molecules inhibiting spectrum of human viruses have been considered as safe in humans by early phase clinical trials. A database has been created for these nearly 120 drugs wherein they suggested that few can be effective against COVID-19 (https://drugvirus.info/). Various therapeutic strategies and their proposed outcomes are presented in Fig. 2.

Remdesivir (GS-5734), which is a viral RNA-dependent RNA polymerase inhibitor, has been under study for early-stage SARS-CoV-2 and is in Phase III clinical trial (ClinicalTrials.gov Identifier: NCT04252664; NCT04280705). It has shown potential against SARS-CoV and the Middle East respiratory syndrome (MERS-CoV) [47,48]. Intravenous
remdesivir was used to treat the first case of SARS-CoV-2 patient in the USA when his condition deteriorated. However, to evaluate safety and efficacy, randomized controlled trials are needed [49].

Chloroquine is in use against malaria and some autoimmune diseases [50]. Potential of hydroxychloroquine has been studied in combination therapy for viral pneumonia (ClinicalTrials.gov Identifier: NCT04261517). In vitro studies for SARS-CoV-2 showed that chloroquine is potentially active in suppressing it. Hydroxychloroquine clinical trial for COVID-19 indicated that in patients treated with hydroxychloroquine, the total time of recovery was shortened and it also helped in resolving pneumonia [4,51]. A study coordinated by the Méditerranée Infection University Hospital Institute in Marseille suggested that viral load was reduced or it disappeared after treatment with COVID-19 and this effect was reinforced by azithromycin [13]. U.S. National Institutes of Health has also initiated a clinical trial for hydroxychloroquine for COVID-19.

Umifenovir/Arbidol is the drug used against influenza infection in Russia and China and is membrane fusion inhibitor which targets the entry of virus [52]. Few pieces of evidence along with a retrospective cohort study suggest that Umifenovir alone or along with antiviral drugs is beneficial against COVID-19 pneumonia and various randomized clinical controlled trials are in progress to study the efficacy of

**Fig. 2.** Various therapeutic strategies underway against COVID-19 (a) and their proposed outcomes (b).
Arbidol on COVID-19. Umifenovir with another antiviral drug lopinavir/ritonavir in different combinations is currently considered for Phase IV clinical trial for pneumonia associated with COVID-19 (ClinicalTrials.gov ID: NCT04255017) [53].

Phase II clinical trials are ongoing for Favipiravir which is an inhibitor for viral RNA polymerase in combination for pneumonia associated with COVID-19 (Chinese Clinical Trial Registry Identifier: ChiCTR2000029544). Japan-based company Fujifilm is working on Phase III clinical trials of Favipiravir in Japan and initiated Phase II trials in the US [54]. Other therapeutic approaches are discussed in other communications [5,55–60].

Another protein acting as potential therapeutic target is the main protease of SARS-CoV-2, Mpro. Mpro is a 33.8 kDa protein which cleaves polypeptides formed from translation of viral RNA. These polypeptides are usually two overlapping products pp1a and pp1ab and their proteolytic cleavage results in functional polypeptides (Fig. 1b), which are then recruited in RTC region wherein translation of other viral structural and non-structural proteins occur. The main protease Mpro, also known as 3CLpro (3-chymotrypsin-like protease) cleaves polypeptides at 11 distinct sites. Mpro protein, being functionally important in the survival and replication of SARS-CoV-2 virus and with the absence of homologues in human, makes them a potential drug target in fighting against SARS-CoV-2 [61]. Zhenming et al. have demonstrated that N3 is a potential inhibitor of Mpro not only for SARS-CoV-2 but also for SARS-CoV and MERS-CoV using structure-based ab initio drug design, virtual screening and high-throughput screening methodologies [62]. Kanhed et al. have screened a FDA approved drug library and the Asinex Bio-Design Library to determine potential molecules which can be useful as therapeutic agents for COVID-19 [61]. Even various bioactives from medicinal plants have been screened for potential candidacy against Mpro [63]. Attempts have been made to design protein specific peptides but due to their small size (8 aa residues) they were unable to maintain secondary structure thus were less effective [64,65]. Here, in this study, we have used machine-learning approach to design protein specific peptides inhibiting SARS-CoV-2-Mpro, which would be able to provide effective cross-protection against various COVID-19 variants.

2. Materials and methodology

2.1. Sequence retrieval

Nucleotide Sequence database downloaded from GISAID server [66] for Covid-19. Nucleotide Sequence data set consists of 2765 sequences of different Covid-19 patients from different countries and/or region. We analyzed sequence data of mainly four regions including Wuhan, Italy, USA and India.

2.2. Sequence alignment

The sequence alignment of all sequences belonging to 4 regions (India, Wuhan, Italy, USA) was performed using European Bioinformatics Institute (EBI) ClustalW [67]. The parameters were defined as DEALIGN INPUT = NO, MBED-LIKE CLUSTERING = Yes, NUMBER OF COMBINED = 0, MAX GUIDE TREE = DEFAULT, MAX HMM = 1 and order = aligned. In addition, the alignment results were analyzed using phylogenetic tree, aligned sequences and scoring matrix.

2.3. Generation of alignment kernel and classification

Once alignment is done, we proceeded further with defining sequence kernels on neighborhoods by taking into account the most conserved central residue of the sequence fragment. Thereafter, we created matrices by assuming a fixed serial ordering of the residues. Let P and Q be m x k and m x l matrices respectively. Let R and S be m x m and l x k matrices, such that R and extension of S to n x n matrices, n = max (m, k, l), which is a positive semi definite matrix. Then, (P, Q) = tr (P R Q) S is a pair sum and direct alignment kernel.

Generalizing this result, hereafter, we defined our generated sets of sequence-based kernels on neighborhoods. We took minimalistic set of neighborhoods that cover all residues in an alignment.

Training dataset: Here, we divided sequence of the training data set, as mentioned in equation below:

\[ D = \{(A,T)\}_{i=1}^n = 1 \]

into groups according to their class labels. We obtained frequent occurring subsequences as seed sequences. Thereafter, identified frequent subsequences from each clustered aligned group and then took the merged subsequences sets together as the seed set S’. For each seed Ss and for each A in the training dataset, we used \( f \) to define the feature vector. For two sets of sequences, A, A’, we constructed the complete weighted bipartite sequences and computed the kernel \( K_s(A,A’) \) using equation:

\[ K_s(A,A’) = \max_{\pi} \sum_{V \in \pi} f(V) \]

where, \( \pi : V(A) \rightarrow V(A’) \) denotes an alignment of sequence A and A’ and \( f(V) \) is a set of features associated with the sequence.

Thereafter, we trained the predictive model obtained through the above equation using a kernel classifier.

Testing dataset: We used the model which was trained to forecast about sequence in the test data set. Here, for each seed Ss and for each A in the testing data set, \( f \) is used to label the sequence in A and thereby create feature vectors as performed above for training data set. The above equation illustrated is a representation, wherein, the kernel function \( K_s(A,A’) \) is computed for each sequence A in the test data set and for each sequence S’ in the training data set. Hereafter, we have used kernel classifier and thereby models were trained to obtain prediction accuracy of the test data set. The obtained sets of kernel functions were used as an input for the generation of motifs.

2.4. Motif generation

Motifs one of a kind for the main protease of SARS-CoV-2 (Mpro) was recognized by the use of MEME suite v4.11.2 [68]. A motif recognizable proof measure incorporates any number of repetition sites of motifs within the given set of input arrangements. The number of motifs that ought to be found was set as 6 with a least length of 6 builds to a most extreme length of 20 builds were distinguished. The motif shortlisting was done based on their secondary structure and from that point 6 motifs were shortlisted.

2.5. Peptide library design

A synthetic library of peptides was designed where the 6 shortlisted motifs, as mentioned above, served as template. For the peptide designing and optimization purpose, a deterministic search approach was analyzed with physiochemical properties was implemented. This algorithm is known as Dead End Elimination (DEE) Algorithm. Its main motive is to design novel sequences with an exhaustive searching to identify and eliminate those who don’t fit in global minimum energy conformation. The total energy of a given pair is represented as:

\[ E_{Tot} = E_i + \sum_{j \neq i} E(i,j) + \sum_{j \neq i} E(i,j) \]

where, \( E_{Tot} \) is the total energy of the system; \( E_i \) is the initial energy of
the template; i, is the rotamer state at position i; E(iri) is the self-energy of rotamer at position i; and E(iri, js) is the pairwise energy between rotamers at position i and j [69]. A brief overview of the algorithm is shown in Fig. 3. Contemplating with overlapping combinations and permutations, a library of 97 peptides was created with an average residue length of 12 to 14 amino acids. Applying various filters narrowed the peptides library search space down. The different filtering criteria's considered were structural stability of protein-peptide complex; geometric and chemical complementarity to ensure minimum stearic clashes between residues; degree and extent of interactions between receptor-ligand, high binding affinity. The concept of reverse docking was also implemented to minimize non-specific binding and toxic effects on humans. Based on these criteria; 4 peptides were shortlisted.

2.6. Structure prediction and docking

PEPstr [70] was used to predict the 3D structure of the designed peptides. From the library of 97 peptides, 4 peptides were chosen as potential candidates based on their structural geometry and chemical complementarity, stearic hindrance, structural stability, hydrophathy index, net charges, binding affinity, specific interactions, etc.

AutoDock Vina [71] and MGL Tools 1.5.6. [72] respectively, were used for protein-peptide docking and analysis. While preparing the files for protein and peptides, Gasteiger charges and polar hydrogen atoms were added and subsequently saved as ‘pdbqt’ file. AutoDock follows empirical type of scoring method where total energy is calculated as:

\[ \Delta G_{\text{total}} = \Delta G_{\text{bond}} + \Delta G_{\text{angle}} + \Delta G_{\text{dihedral}} + \Delta G_{\text{vdw}} + \Delta G_{\text{elec}} + \Delta G_{\text{conf}} + \Delta G_{\text{cdot}} + \Delta G_{\text{sur}} \]

The docking was achieved using Genetic Algorithm with rigid type of docking. PyMOL v1.7.4.5 (Delano Scientific) was used for the visualization of docked complexes and Ligplot + [73] for the mapping of protein-peptide interactions.

2.7. Molecular dynamics simulation

The Molecular Dynamics (MD) Simulations were used to check the stability and extent of interactions maintained between protein-peptide complexes in physiological conditions using DESMOND 3.2 with maestro-v11.6 (D.E. Shaw Research) [74]. Schrodinger 2018-2 was used as GUI interface for 50 ns long MD Simulations. MD Simulation was done with Normal Pressure and Temperature (NPT) ensemble class at 300 K temperature and 1.01325 bar pressure with Nose-Hoover chain thermostat and Martyna-Tobias-Klein barostat methods. The results were analyzed using inbuilt analysis Simulation Interaction Diag. It provides information about protein-ligand interactions in a graphical form, RMSD (Root Mean Square Deviation), protein and ligands RMSF (Root Mean Square Fluctuation) and protein-ligand interactions (hydrophobic, ionic and water-bridge contacts).

RMSD is calculated as:

\[ \text{RMSD} = \frac{1}{N} \sum_{i=1}^{N} \left( r'(t_i) - r(t_{ref}) \right)^2 \]

where, \(N\): total number of atoms for which RMSD has to be calculated; \(t_{ref}\): the reference time, (usually the first frame); and \(r'\) is the position of the selected atoms in frame x after overlaying on the reference frame, where frame x is recorded at time \(t_x\). The procedure is iterated for every frame in the simulation trajectory.

RMSF is calculated as:

\[ \text{RMSF} = \sqrt{\frac{1}{T} \sum_{i=1}^{T} \left< (r'_i(t)) - r_i(t_{ref})^2 \right>} \]

where, \(T\): RMSF is calculated based on the trajectory time, \(t_{ref}\): the reference time, \(r_i\): the position of residue i; \(r'\): the location of atoms in residue i after superposition on the reference frame, and the angular brackets signifies the average of the square distance is taken over the selection of atoms in the residue.

The total energy of the system is calculated as:

\[ E_{\text{total}} = E_{\text{bond}} + E_{\text{vngl}} + E_{\text{ torsion}} + E_{\text{coup}} + E_{\text{cross}} + E_{\text{nonbond}} \]

where, \(E_{\text{bond}}\) is the bond length energy based on either Morse or LJ Potential, and is defined as:

\[ E_{\text{bond}} = \sum_{\text{bonds}} \frac{1}{2} k_b (b - b_0)^2 \]
E\text{angle} is the bond angle energy and is defined as:

\[ E_{\text{angle}} = \sum_{\theta} H_\theta (\theta - \theta_0)^2 \]

E\text{torsion} is the torsion energy associated with the rotation of two atoms of a molecule relative to each other. It is defined as:

\[ E_{\text{torsion}} = \sum k_\phi [1 - \cos(n_\phi + \phi)] \]

E\text{oop} is the out of plane interactions and is defined as:

\[ E_{\text{oop}} = \sum H_\phi X_\phi^2 \]

E\text{cross} is the cross term between different energies like angle energy and torsion energy, etc.

E\text{nonbond} includes the energies of non-covalent bonds like van-der-Waals, Coulombs and hydrogen bonds [75].

3. Results

We know that algorithm based machine learning approach is not only fast but chances of failure of drug discovery in lab are reduced largely because so many filters and conditions are implicated while performing these algorithms. Therefore, this systematic approach aided by merit of reduced time required could show the way for a better outcome.

In pursuit of this aim, we defined our therapeutic strategy using machine learning algorithms based on two perspectives, one is selection of diverse viral sequences of Covid-19 reported from India, China, Italy and USA and using these diverse sequences to identify peptides with activity against diverse strain of COVID-19 virus worldwide. Once identified, these peptides were tested against Proteases of COVID-19 for their efficacy or based on diverse peptides identified, peptides of different amino-acid composition could be designed based on identified peptides and tested for cross protection against various viral strains worldwide. Viral sequences were taken from GISAID wherein more than 95,000 viral sequences have been submitted since the emergence of pandemic till date. We aligned these sequences using European Bioinformatics Institute (EBI) tool Clustal Omega and the alignment results were analyzed using phylogenetic tree, aligned sequences and scoring matrix. As we know that identification and designing of these peptides is resource consuming approach, thus, we relied on computational method of high accuracy for prediction of a library of these peptides (Supplementary Table 1) and hence used Dead End Elimination algorithm which is one of the most popular algorithm for model development and further identification of peptides from molecular diverse sequences. We selected proteases as a target because the main protease of SARS-CoV-2 (M\text{pro}) is an important or can say key enzyme, which is involved in the process of replication and transcription of virus.

Through the virtual screening against the M\text{pro} identified 4 peptides as the potential candidates in inhibiting protease function with strong binding activity (Table 1). Docking of these selected peptides along with their interactions against the M\text{pro} and their docked poses are shown in Figs. 4 and 5a–d and Table 2. We further carried out short scale MD Simulation of 50 ns time scale to check the stability, strength and amount of interactions being maintained between protein and peptide during simulation time. For model1 and model4, interactions increased by 3 fold and 4 fold respectively but for model2 and model3 interactions increased by more than 5folds during MD simulations (Table 3 and Fig. 6). In the Fig. 5e–h, only those residues that share more than 1 fraction of interactions, i.e. more than one atom of that residue remain in contact with peptide throughout simulation time, are highlighted. All pre-MD interactions were observed to remain during MD. It was also observed that during MD simulation, the negatively charged amino acids aspartate and glutamate (position 153, 240, 245, 248) forms strong-ionic interaction in all the four protein-peptide models indicating that these residues might be crucial while inhibiting the activity of M\text{pro}. The MD Simulation trajectory of all four models in form movie is available as supplementary files (Movies M1–M4). These peptides need further validation in laboratory conditions. In short, we can say we have used the approach of sequence alignment and machine learning algorithm for identifying peptides based on molecular diversity of COVID-19.
4. Discussion

We have extensively reviewed the pathogenesis mechanism of SARS-CoV-2 in the context of immunological features to find possible drug candidates. It is based on the strategy that such drugs will prevent the entry of viruses in cells via inhibition of endocytosis and combination therapy with antiviral drugs could reduce the infectivity of the virus and its replication along with preventing inflammatory responses.

As the above data suggests, lot of efforts are being taken by both industry and research organizations for developing a therapeutic strategy that might be a positive click for treatment of COVID-19. However, most of the studies are underway, and scientific, medical, and regulatory authorities will not neglect the validation of these aspects, so rigorous validation of the hypothesis is equally important. Strategy along with having an advantage of a short period with accuracy is majorly focused on molecular diversity of pathogen, which is the main

![Fig. 5. Mpro residues interacting with peptides. (a) to (d) are protein-peptide complexes before MD Simulations and (e) to (h) are protein-peptide complexes post MD Simulations. Peptide is represented in orange stick form while residues of Mpro protein interacting with peptides are shown in cyan color.](image-url)
lacunae and dilemma for the efficacy of various strategies in process, to be effective against various strains of the virus. To add to the crux of the story, ML has provided us a novel and a promising technique for peptide design, calculating energies for training set of sequences and helps assign an energy criterion that is only applicable to one or few of the amino acid residues. Most importantly, it outscore the generation of possible rotamers and thereby rendering it to be combinatorial easier with fewer amino acid residues. Through this work, we augmented simple kernel representations recognizing important topological similarities and incorporating structural information into kernel functions, which actually helped in approximate matching of structure features of the peptide library created. We demonstrated machine learning algorithms to address biological applications in high-dimensionality datasets. Leveraging the novel framework implemented in this paper opens a newer dimension of generation of peptide library through kernel function. Finally, optimal generation of diverse sequences with favorable conformation has yielded the strategy adopted to be statistically equivalent to discretedesign as we are able to design peptides with high selectivity towards their desired target that warrants experimental validation. Considering the evolutionary aspect of SARS-CoV-2, we have retrieved 2765 sequences containing wide variety of mutants from COVID-19 patients. This large dataset was used to generate alignment kernels, which further provided conserved motifs across all mutants. Against these motifs, a library of 97 peptides was designed. This tactic enabled us to cover a wide spectrum of existing as well as probable mutants of SARS-CoV-2 ensuring the efficient targeting of main protease M^pro_. Based on binding energies, peptide conformation and strength and type of interactions with protein, 4 peptides were shortlisted. These 4 protein-peptide complexes were subsequently subjected to MD Simulation studies to ensure their stability in physiological conditions.

### Table 1

<table>
<thead>
<tr>
<th>S. No</th>
<th>Motif sequences</th>
<th>Motif E-value</th>
<th>Translated sequence</th>
<th>Total energy (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>CGGGTTTGCGGTGCTAAGTCGAGCAGCCCGTCCTACACCCCGCCGCAACGACG</td>
<td>1.0e-038</td>
<td>RVGCVSAARLTGCGTG</td>
<td>−23.060</td>
</tr>
<tr>
<td>Model 2</td>
<td>ACAGGGGAGTCGAGCAGCGCTCTACCTAGTCAGGAGGACGGAGGAGGGCAGCAGCTTCTCTGTGGTG</td>
<td>5.5e-038</td>
<td>TGSSRATWVLWRLRGG</td>
<td>−38.704</td>
</tr>
<tr>
<td>Model 3</td>
<td>GGCCTCTAGACGAGGAGGCGAGGCCAGGCGACGCCAGCGCTCTTCTCTG</td>
<td>3.0e-037</td>
<td>GYAESRRGSGGQASSR</td>
<td>−37.035</td>
</tr>
<tr>
<td>Model 4</td>
<td>CCAGCCGCACGGAGGGAACTCTCTGCTAGAAATGGGCTGGGACATGGG</td>
<td>3.0e-037</td>
<td>PGSRRGTSRPMAGNG</td>
<td>−21.124</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Interacting residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>D153, Y154, D245, D248, I249, F294, V303, T304, F305</td>
</tr>
<tr>
<td>3</td>
<td>Q107, D153, L227, T243, Q244, D245, H246, I249, F294</td>
</tr>
<tr>
<td>4</td>
<td>F8, N151, D153, D245, D248, I249, F294, D295, V297, R298</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Model</th>
<th>Interactions pre-MD</th>
<th>Interactions during MD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>45</td>
</tr>
</tbody>
</table>

Fig. 6. Short scale Molecular Dynamics Simulation studies to check the stability, strength and amount of interactions being maintained between protein and peptide. The first row represents the protein – ligand (here peptide) RMSDs with protein in blue and ligand in red color. The middle row represents the type and fraction of interactions maintained by various protein residues during simulation time. H-bond is represented in green, hydrophobic in purple, ionic in pink and water bridges in blue color. The bottom row also represents the extent of protein-peptide interactions for a given residue at a particular simulation time. In the timeline graph, darker the color more is the fraction of contacts between protein and peptide.
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We are thankful to the Director, National Centre for Cell Science (NCCS) for supporting the Bioinformatics and High Performance Computing Facility at NCCS, India.

References


Phylogenomic analysis of SARS-CoV-2 genomes from western India reveals unique linked mutations

Dhiraj Paul¹, Kunal Jani¹, Janesh Kumar¹, Radha Chauhan¹, Vasudevan Seshadri¹, Girdhari Lal¹, Rajesh Karyakarte², Suvarna Joshi², Murlidhar Tambe², Sourav Sen³, Santosh Karade³, Kavita Bala Anand³, Shelinder Pal Singh Shergill³, Rajiv Mohan Gupta³ Manoj Kumar Bhat¹, Arvind Sahu¹,Yogesh S Shouche¹*

¹National Center for Cell Science Pune, India
²B. J. Government Medical College, Pune, India
³Armed Forces Medical College Pune, India

*Crossponding author

Yogesh S Shouche, Email id: yogesh@nccs.res.in

Abstract

India has become the third worst-hit nation by the COVID-19 pandemic caused by the SARS-CoV-2 virus. Here, we investigated the molecular, phylogenomic, and evolutionary dynamics of SARS-CoV-2 in western India, the most affected region of the country. A total of 90 genomes were sequenced. Four nucleotide variants, namely C241T, C3037T, C14408T (Pro4715Leu), and A23403G (Asp614Gly), located at 5'UTR, Orf1a, Orf1b, and Spike protein regions of the genome, respectively, were predominant and ubiquitous (90%). Phylogenetic analysis of the genomes revealed four distinct clusters, formed owing to different variants. The major cluster (cluster 4) is distinguished by mutations C313T, C5700A, G28881A are unique patterns and observed in 45% of samples. We thus report a newly emerging pattern of linked mutations. The predominance of these linked mutations suggests that they are likely a part of the viral fitness landscape. A novel and distinct pattern of mutations in the viral strains of each of the districts was observed. The Satara district viral strains showed mutations primarily at the 3´ end of the genome, while Nashik district viral strains displayed mutations at the 5´ end of the genome. Characterization of Pune strains showed that a novel variant has overtaken the other strains. Examination of the frequency
of three mutations i.e., C313T, C5700A, G28881A in symptomatic versus asymptomatic patients indicated an increased occurrence in symptomatic cases, which is more prominent in females. The age-wise specific pattern of mutation is observed. Mutations C18877T, G20326A, G24794T, G25563T, G26152T, and C26735T are found in more than 30% study samples in the age group of 10-25. Intriguingly, these mutations are not detected in the higher age range 61-80. These findings portray the prevalence of unique linked mutations in SARS-CoV-2 in western India and their prevalence in symptomatic patients.

**Keywords:** SARS-CoV-2; Single Nucleotide Polymorphism; Phylogenetics; Age-wise; unique mutation pattern; genome sequencing of SARS-CoV-2 from clinical specimens, Maharashtra, India.

**Importance**
Elucidation of the SARS-CoV-2 mutational landscape within a specific geographical location, and its relationship with age and symptoms, is essential to understand its local transmission dynamics and control. Here we present the first comprehensive study on genome and mutation pattern analysis of SARS-CoV-2 from the western part of India, the worst affected region by the pandemic. Our analysis revealed three unique linked mutations, which are prevalent in most of the sequences studied. These may serve as a molecular marker to track the spread of this viral variant to different places.

**Introduction**
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) which is the causal agent for COVID-19, belongs to the category of betacoronaviruses. The respiratory illness caused by the virus varies from mild disease to severe disease and death (Guo et al., 2020; Tian et al., 2020). This virus has spread rapidly from Wuhan, China since late 2019, and on 11th March 2020, WHO declared the disease caused by SARS-CoV-2 as a pandemic (Li et al., 2020). This crisis has rapidly escalated due to globalization and highly contagious nature of the virus. Till-date, more than 16
million people have been infected globally, and ~0.6 million have succumbed to it (https://covid19.who.int/). Currently, after the USA and Brazil, India is in the third position on the basis of adversity of infection, and more than a million Indians have been infected until now. Among the various states of India, the Maharashtra state is a major hotspot for this disease, having around 1/5th of total reported infections in India and thus needs more attention (https://www.covid19india.org/).

The genome sequences from across the world for SARS-CoV-2 are available through the Global Initiative on Sharing All Influenza Data (GISAID) platform since 12th January 2020 (Shu and McCauley, 2017). The sequences of novel coronavirus (CoV) show a close similarity with severe acute respiratory syndrome-related coronaviruses (SARSr-CoV), and like SARS-CoV, it also utilizes ACE2 as their entry receptor (Zhang et al., 2020). The virus contains ~30 kb positive-sense, single-stranded RNA genome, which encodes four structural and multiple non-structural proteins (Astuti 2020). Structural viral proteins form the capsid and the virion envelope. Non-structural proteins help in various stages of the virus life cycle like replication, followed by translation, packaging, and release (Lai and Cavanagh, 1997; Li 2016; Lu et al., 2020). Many mutations have emerged in the SARS-CoV-2 genome that can modulate viral replication, transmission and virulence efficiency (Jia et al., 2020; Pachetti et al., 2020). Genome sequencing of SARS-CoV-2, followed by the identification of genetic variants, have been reported from different parts of the world (Dorp et al., 2020). Recent studies have speculated that, under selective pressure, genetic variability accumulate and persist in the SARS-CoV-2 genome, for its better survival and transmission (Dorp et al., 2020). Most of the reported variants belong to Orf1a, Orf1b, S, N, and 5’ UTR region of its genome. Three sites in Orf1ab (region Nsp6, Nsp11, Nsp13), and one in the spike protein are characterized by an unusually abundant number of recurrent mutations that may indicate convergent type of evolution. Specific interest is imparted in the context of adaptation of SARS-CoV-2 in the human host (Dorp et al., 2020).

In this study, we explored the phylogenomic diversity among the SARS-CoV-2 from three different regions of Maharashtra, the western state of India. In particular, we generated whole-genome sequences from 90 samples and examined if there are any predominant and unique mutations responsible for different cluster formation among the present study samples. Further, we have also analyzed whether select variants are associated with the gender, age and symptoms.
Materials and Methods

Sample collection

Nasopharyngeal/throat swabs of suspected Covid-19 patients were collected in April and May 2020. Samples considered in the study confirmed positive by real-time PCR for SARS-CoV-2 at the National Center for Cell Science, B. J. Government Medical College Pune, or Armed Forces Medical College, Pune. The RT-PCR assay utilized WHO suggested primers and probes to target E, ORF1b and RdRp genes. Ethical clearance was taken from the Institutional ethical committee for the present study. Samples were anonymized except the information about gender, age, collection date, travel histories, and symptoms. The study sample details are given in Supplementary Data 1.

RNA extraction and genome sequence

The COVID-19 patient samples that showed Ct value for E gene ranging from 18 to 24 employing RT-PCR were selected for genome sequencing. Viral RNA was extracted from 300 µl of viral transport media (VTM, Himedia, India) using QIAamp 96 Virus QIAcube HT Kit (Qiagen, Germany), following the manufacturer’s instructions. Extracted RNA present in 50 µl of elution buffer was immediately used for cDNA synthesis using random hexamer primers, following the manufacturer’s protocol (Qiagen, Germany). The extracted RNA and cDNA were stored at -80 °C until analysis. PCR based COVID-19 viral genome enrichment was performed using QIAseq SARS-CoV-2 Primer Panel Kit (Qiagen, Germany). A total of 98 primer pairs specific to COVID-19 were used for enrichment where primers pairs were distributed in two pools (Pool-1 and Pool-2). For multiplex PCR amplification, 5 µl cDNA of each sample was used as a template. The amplification was done with initial denaturation step at 98°C for 2 minutes, followed by 30 cycles of denaturation at 98°C for 20 sec, annealing/elongation at 72°C, and final extension step at 72°C for 20 minutes. Amplified PCR product of both the pooled primers set of the samples was run on 2% agarose electrophoresis gel. After EtBr staining, samples that showed amplification product with both the pooled primers were selected for further use.

The ~400 bp amplicon product produced using both Pool-1 and Pool-2 from the same clinical sample was combined and purified using a 1X concentration of Agencourt AMPure XP beads (Beckman Coulter); the final product was eluted in 30 µl of elution buffer (Qiagen, Germany). The purified product was quantified using the HS DS DNA assay kit using Qubit
For library preparation around 100-300 ng of purified amplicon were used. DNA library was made using QIAseq FX DNA Library Kit (Qiagen, Germany). To retrieve larger fragment library size, during digestion, 3 minutes incubation time was set followed by enzyme inactivation. Adapter ligation was performed using the supplied ligase enzyme by incubation for 15 minutes at 20°C followed by enzyme inactivation by heating at 65°C for 20 minutes. After library preparation, library size selection and washing were carried out following manufactures’ protocol using 0.8X and 1 X concentration of AmpureXP. The average library size was ~500 bp. The purified library was sequenced by 2x250 bp chemistry using the Illumina MiSeq platform.

Data analysis

Fastqc tool was used to check the quality of the raw paired-end sequences after sequencing (Andrews 2010). Adapter sequences and poor quality sequences were removed, and good quality sequences (Q>30) were selected using Trimmomatic (Bolger et al., 2014) for further analysis. Reference-based genome assembly was done using BWA (Burrows-Wheeler Aligner; Li 2013) to generate the consensus sequence. SARS-CoV-2 genome (NCBI GenBank accession MN908947.3) was used as a reference for mapping. Unmapped reads were discarded. Consensus sequences were used for checking completeness and coverage calculation. Cross verification of the assembly process using CLC genomics was achieved using the reference genome (Accession ID MN908947.3). Genome sequences were deposited in the GISAID database. The complete list of accession IDs of the study samples is listed in Supplementary Data 1.

Phylogenomics analysis and functional evaluation of variants

Following the neighbor-joining method, a phylogenetic analysis was carried out to understand the relationship among the study sequences, with 1000 bootstrap. SARS-CoV-2 isolate Wuhan-Hu-1 (NCBI GenBank accession MN908947.3) was included as a reference genome in the tree. To identify the mutations and their position, a variant analysis was performed for all the samples utilizing SARS-CoV-2 sequence Wuhan-Hu-1 (NCBI GenBank accession MN908947.3) as a reference. The protein variants identified in the clade were assessed to know their functional effects using PROVEAN (Protein Variation Effect Analyzer) program, considering the protein sequences of the Wuhan-Hu-1 genome as reference and a default threshold value of -2.5 (Choi and Chan, 2015). It provided a universal method to calculate the functional effects of protein sequence
variations that might be amino acid substitutions, deletion, and insertions at single or multiple levels. The pair-wise alignment-based score was calculated to estimate the sequence similarity change of the query sequence to a protein sequence homolog before and after the insertion/substitution of an amino acid variation to the query sequence.

**Structural and bioinformatics analysis of SARS-CoV-2 variants**

Multiple sequence alignment: ClustalOmega (Sievers et al., 2011) and MUSCLE (Edgar 2004) as multiple sequence alignment tool were used to align protein specific regions. 3D structures of protein were retrieved from PDB database. Specifically, following PDB were used: Spike protein: 6MOJ, 6MIV, 6LXT, Orf3a: 6XDC; RdrP: 6X2G, 7C2K; and Nucleocapsid: 6WJI. Structural mapping and analysis of mutations was carried out in PYMOL (DeLano et al., 2002).

**A comparative study among the Indian samples**

The phylodynamic analysis was done by Nextstrain pipeline following the standard protocol (Hadfield et al., 2018). The dataset of Indian samples till 24th June 2020 was downloaded from the GISAID database and used for the analysis. A total of 943 sequences from different regions of India were reported till that day. The details of samples used from the GISAID database, including sample id, sampling time, and submitting Institutes are listed in Supplementary Data 2. In addition, 90 samples from the present study were also included in the analysis. In this pipeline, all the sequences, including our study samples, were aligned using MAFFT (Multiple alignments using fast Fourier transform) (Katoh and Toh, 2008). Using IQTREE, the phylogenetic tree was made by the Augur tree implementation. Furthermore, the raw tree was processed with Augur for generating TimeTree, annotating ancestral traits, inferring mutations, and finally identifying clades (Nguyen et al., 2015). The resulting tree was viewed using Nextstrain. After phylodynamics and clade reclassification, only samples that were collected during April and May 2020 from other parts of India with a suitable representation were selected for comparative study.

**Statistical analysis**

Relationship/association among the mutations with respect to their presence and absence in study samples was determined by Unweighted Pair Group Method with Arithmetic Mean analysis.
(UPGMA). The binomial and one way ANOVA was used to check the significance level of effect of gender, age, symptoms of the study samples.

Results

Demographics and data analysis
A total of 90 COVID 19 positive samples were sequenced, and were collected from Maharashtra, India, primarily from Pune, Satara, Nashik and Kolhapur districts. The age of the patients selected in the present study ranged from 2-78, with 80% patients were in the age range of 30-60 years. More than 1000x coverage and almost 99.99% genome completeness was achieved for all the study samples with respect to the reference genome, i.e., Wuhan COVID 19 genome. The metadata for the specimens have been deposited after sequencing in the public domain (Supplementary Data 1).

Identification of phylogenomics cluster and responsible mutation pattern
A phylogenetic analysis of the viral genome sequences in the present study performed using Wuhan SARS-CoV-2 isolate’s genome (NCBI GenBank accession MN908947) as reference. The sequences analyzed in the study formed four clusters. Except for cluster 1, all other clusters were rooted from a single clade. Within this clade, clusters 2, 3, and 4 formed major groups (Figure 1). Cluster 1 formed by six sequences in the present study showed a close association with Wuhan reference sequence, indicating that these strains were genetically closer to the original sequences (Wuhan, China). Interestingly, ~45% of the sequences in the study fell in cluster 4 which is most distant from cluster 1 (Figure 1). Cluster 2 and 3 formed the other two significant groups, each consisting of 18 sequences.

To identify the basis of the formation of different phylogenetic clusters by the sequences of the present study, all the SNPs at different locations of the 90 genomes were mapped, followed by a heat map analysis based on the presence and absence of SNPs (Figure 2). Interestingly, a distinct pattern of mutations in different clusters was noticed in the present study (Figure 2). Of note, cluster 1 in the phylogenetic tree showed a distinct pattern of mutations compared to other clusters. Mutations C6312A, G11083T, C13730T, C23929T, and C28311T, were present in all the sequences of cluster 1, making them characteristic for this cluster 1 (Figure 2). The major cluster,
i.e., cluster 4, was rooted from the same clade where clusters 2, 3, and 4 were also formed and variants C241T, C14408T, and A23403G were found to be key for the formation of this particular clade (Figure 2). These variants were detected in all the sequences, except six sequences of cluster 1. In addition to these variants, SNPs C313T, C5700A, G28881A were explicitly present in more than 95% sequences belonging to cluster 4, i.e., supercluster, and were variants for the formation of cluster 4 (Figure 2). This unique pattern of mutation, i.e., the simultaneous presence of the three mutations (C313T, C5700A, G28881A), has not been reported so far. Only mutation C313T and G28881A have been reported separately. Variants C3634T, C15324T were unique for all the sequences belonging to cluster 3, which were found in the region of Orf1a and Orf1b, respectively. Another major cluster, i.e., Cluster 2, where mutation C18877T, G25563T, C26735T, were unique and were found in ~90% of the sequences of this cluster (Figure 2).

**Mutations and functional consequences**

A total of 125 SNPs/variants were found in the present study. Most of the SNPs were found in Orf1a (46 SNPs) followed by Orf1b (31 SNPs), S (17 SNPs), N (10 SNPs), ORF3a (8 SNPs), 5′ UTR (4 SNPs), M (3 SNPs), E (1 SNP) regions of the genome (Supplementary Data 3). Among these variants, four were found in more than 90% study samples, i.e., C241T (5′ UTR), C3037T (ORF1a), C14408T (Pro4715Leu), and A23403G (Spike protein, Asp614Gly). Three variants C313T (ORF1a), C5700A (ORF1a, Arg to Lys), G28881A (N protein, Gly to Arg), were found in close to 50% of the sequenced samples. Mutations C3634T (Orf1a), C15324T (Orf1b), G18877T (NSP11 region of Orf1b), G20326A (Orf1b, Val6688Ile), G25563T (APA3_viroporin region ORF3a, Asn57His), G26152T (APA3_viroporin region ORF3a), were found in more than 20% of the study sequences (Supplementary Data 3). In addition, there were 100 other mutations that were found at a lower frequency (1-10%).

Protein Variant Effect Analyser (PROVEAN) was used to understand the functional consequences of non-synonymous variants formed due to amino acid substitution. A database for SARS-CoV-2 was constructed using the protein sequences of the WH1 reference genome (NC_045512) and the non-synonymous variants were analyzed. Two highly abundant (90%) variants of ORF1a/RdRP (Pro4715Leu) and Spike protein (Asp614Gly) were suggestive of neutral functional consequences as predicted by PROVEAN (Supplementary Data 3). Some mutations
were found deleterious, based on PROVEAN analysis (Supplementary Data 3). However, the exact effect of this mutation should be validated experimentally.

Region-wise mutation pattern
Region-wise mutation patterns among the viral sequences from Pune, Satara, and Nashik districts are depicted in Figure 3. Intriguingly, a specific pattern of mutation was found to be prevalent in all districts. Though mutations C241T, C3037T, C14408T (RdRp: Pro4715Leu), and A23403G (S protein: Asp614Gly) were dominant (80% of sequences) in all districts, Pune district sequences had additional predominant mutations C313T, C5700A, and GGG28881..28883AAC. Satara sequences, on the other hand, had unique mutations like G20326A (Nsp15: Val6688Ile), G24794T (S: Ala1078Ser), G26152T (Orf3a: Gly254Arg), and variants like C18877T and G25563T (Orf3a: Gln57His) in higher frequency, located towards the 3´ end of the viral genome, i.e., region NSP15 of Orf1b, S, and Orf3a. Sequences from Nashik district had unique mutations C190T, C1404T (Nsp1: Pro380Leu), T1872G (Nsp1: Phe536Cys), C3634T (Nsp3) and C5512T (Nsp3). It is interesting to note that unlike Satara sequences, the novel mutations observed in Nasik sequences were located at the 5´ end of the genome, i.e., in 5´ UTR, and Orf1a region of the genome (Figure 3).

Mutation pattern with gender, age, and symptoms
The observed mutation pattern was further analyzed for any relationship with gender, age, and symptoms. With respect to gender, no predominant or unique mutation or mutation pattern was observed (Figure 4). A distinct pattern was observed in age-wise distribution (Figure 5). Its noteworthy that mutations C6843T, C18877T, G20326A, G24794T, G25563T, G26152T, and C26735T were not detected in the age range of 61-80 (Figure 5). Out of these specific seven mutations, C6843T, G20326A, G24794T, G25563T, and G26152T are non-synonymous mutations (Supplementary Data 3). These mutations are less prevalent (5-10%) in the age range of 26-60. However, in the age range of 10-25, the mutations were found significantly (p<0.05, Supplementary Data 4) at a higher proportion (>30%) (Figure 5). This indicated that though these mutations (i.e., C6843T, C18877T, G20326A, G24794T, G25563T, G26152T and C26735T) are observed in the age range 26-60, they are more prevalent in the age range of 10-25.
The relative changes in the frequency of mutations at positions C313T, C5700A, G28881A was observed when symptomatic and asymptomatic individuals were compared (Figure 6). The proportion of mutations C313T, C5700A, and G28881A were found relatively higher (~80%) in symptomatic subjects as compared to asymptomatic (40-50%) (Figure 6). Interestingly, these differences in mutations between symptomatic and asymptomatic individuals were more pronounced females compared to males (Supplementary Figure 1 and 2). In the case of symptomatic females, the relative frequency of these mutations was close to 60% whereas for asymptomatic females, it was about less than half, i.e., 15-20%. In contrast, in case of males relative changes of these mutations, i.e., C313T, C5700A, and G28881A was observed less between symptomatic and asymptomatic individuals (Supplementary Figure 2).

The pattern of mutation: Maharashtra compared to Indian scenario

Phylodynamic and clade reclassification of the previously sequenced genomes from India (943 genomes) and our study (90 genomes) was performed using Nextstrain. Notably 50% strains belonged to the 20A clade followed by 20B, 19A, and 19B (Supplementary Figure 3). The temporal changes in the appearance of different clades were further investigated. Data obtained from previous studies on six Indian states which were generated from the sample collected during April–May 2020 were only considered. The temporal change was studied by calculating proportions of sequences belonging to the five types in each of the two months under consideration. It was found that in capital Delhi (Northern part of India) during both the months, there were no changes. Only 19A type was dominant (100%), which was an ancestral China type (Figure 7). In Maharashtra (Western part of India), in April, 20B and 19A were prevalent where as in May, 19A proportion was reduced, and 20A and 20B were found in almost equal proportion. This indicated that the recent clade has become dominant with time in Maharashtra. Drastic changes were observed in Telangana (Southern part of India) (Figure 7). In April, only 19A clade was dominant (100%), but in month May, 19A type was found <5% and 20B and 20A were found to be more prevalent. Whereas the eastern part of India especially in Odisha, an opposite trend was observed in April, 20A type (90%) was dominating followed by 19A (10%) but in month May, 19A (50%) was prevalent followed by 20A, 20B (Figure 7). Thus, the overall pattern suggests that based on geographical location, the four sites of India showed unique SARS-CoV-2 prevalence. This may be one of the reasons for variable infection rates in different parts of this country.
Discussion
In the present study, the goal was to explore the phylogenetic relationship among the Indian strains and to characterize the prevalent and unique genetic signature/pattern of the local strains sequenced from patients belonging to different regions of Maharashtra, the western part of India. A total of 90 viral whole-genomes were sequenced using nasopharyngeal and throat swabs. Recently, the pandemic caused by SARS-CoV-2 from Wuhan (China), clade O (Forster et al., 2020), has been reclassified as a 19A clade, which thereafter has evolved into multiple clades. Our phylogenetic analysis revealed 6 sequences with close proximity to the original Wuhan sequence, and others belong to three distinct clusters. This indicates that the sequences underwent several mutations in their genome, resulting in radiating phylogenetic clusters. In general, RNA viruses containing error-prone replication causes the mutations in their genetic material, which can be analyzed to track the viral evolution. The study until now proposes that SARS-CoV-2 arised not very long before the first reported case of pneumonia in Wuhan (Wu et al., 2020). Having emerged recently, the observed level of diversity in SARS-CoV-2 is much lower compared to other known RNA viruses (e.g., dengue) where many subclones or lineages occur with multiple SNPs and associated functions (Wu et al., 2020). In the present study, 42 sequences belonged to supercluster 4 and mutations C313T, C5700A, G28881A were responsible for separating this supercluster from other clusters. Interestingly, all three mutations were found in the same genomes and UPGMA based cluster analysis (Supplementary Figure 4) indicating that they may be linked-mutations. Mutations C313T and G28881A have been reported previously in a separate study, but they were not present together with C5700A forming a supercluster. To the best of our knowledge, this finding in the present study is being reported for the first time.

About 50% of mutations in this study were synonymous, indicating null amino acid changes due to nucleotide substitution. At the same time, 50% of non-synonymous sites were detected, which might be due to convergent evolution. Of these non-synonymous mutations, the four most robust SNPs were present in above fifty percent frequency in the population, i.e., C5700A (\textgreater{}50\%), C14408T (\textgreater{}90\%), A23403G (\textgreater{}90\%), G28881A (\textgreater{}50\%) and are located in Orf1a, Orf1b, spike glycoprotein and Corona_nucleoca of N protein, respectively. Non-synonymous mutation at the A23403G position located in the Spike glycoprotein (Asp614Gly), which has a vital role in the binding of the virus to the ACE2 receptor in the host was observed.
The Asp614Gly mutation is very close to the Furin recognition site for cleavage of the Spike protein, which plays a decisive role in the virus entry (Korber et al., 2020; Hu et al., 2020). Interestingly, it has been shown that Gly614 mutant protein is more stable than Asp614 enabling increased transmission. Due to this, pseudovirions with Gly614 infected ACE2-expressing cells more efficiently than those with Asp614 (Zhang et al., 2020; Korber et al., 2020). We analyzed the Spike protein variants of Maharashtra stains using the known 3D structure of the spike protein (Wrapp et al., 2020; Wall et al., 2020). Asp614 site is located at the hinge region of the S1 and S2 domain and is shown to be involved in hydrogen bonding with T859. Any perturbation in this is proposed to destabilize the conformation of spike protein. Interestingly, this mutation was reported mostly in recent clades of SARS-CoV-2, which has a high frequency (more than 90%) in our study. A similar observation was made from other countries where the SARS-CoV-2 variant carrying the amino acid change in Spike protein at Asp614Gly position has become the most prevalent (Korber et al., 2020; Hu et al., 2020). This is in agreement with our results, which indicate similarity in the same pattern in three districts of Maharashtra. In addition to this prevalent mutation, the other 13 mutations in the Spike protein were also detected in the present study but at lower frequencies. Out of this G24794T, corresponding to Ala1078Ser lies in the S2 domain and located at the interface of trimeric oligomer and may affect the stability of the trimeric assembly of the spike protein. S2 domain is shown to be important for the fusion of viral membrane to the host-cell membrane. It was observed in 84 out of 90 tested samples (93.3%).This mutation is also reported in the European strains of SARS-CoV-2 (source: GISAID). We observed that this variant (6/84 samples) is coexisting together with Asp614Gly mutation in our data. Another SNP named C23929T, which is a synonymous, and neutral mutation was found in six sequences. Other SNPs were observed in one or two individuals/sequences only. Most of these mutations are unique for the Spike proteins and have not been reported before. Thus, it could be speculated that these new mutations might be emerging in our chosen geographical area. However, many factors may contribute, and their exact role in virus evolution or infection warrants to be explored further. Interestingly, no mutations in the furin cleavage site (PRRAR; 681-685) located at 23603-23617 (as per Wuhan-Hu-1 genome) which is responsible for the proteolytic cleavage of S protein and increased pathogenesis was observed in ourr samples.

In our sequences, a large number of SNPs (45 positions) were found in the Orf1a region of their genome, such as C313T, C3037T, C3634T, and C5700A. The C313T, C3037T, and C3634T
are silent mutations, whereas a C5700A result in the Ala1812Asp substitution, and it is found in more than 50% of the sequences. The mutation C5700A in ORF1a results in missense mutation Ala1812Asp corresponding to Ala994Asp in the nsp3 protein was observed in 47.8% sequences. Interestingly, this mutation is reported in GISAID from other geographic locations of India. Orf1a containing of nsp 1 to nsp 10 plays an essential role in coping with cellular stress, retaining the functional integrity of the cellular components, and essential roles in the viral replication. Orf1a products play a vital role in viral structure and pathogenicity. Although the major mutations observed here were mostly silent, mutations in the Orf1 a region may affect viral structure and pathogenicity (Harcourt et al., 2004; Stobart et al., 2013). A number (35) of mutations were also found in the Orf1b region, among which C14408T, C15324T, C18877T, and G20326A were observed frequently. Mutation C14408T was found in more than 90% of the sequences, and it is a non-synonymous natural mutation resulting in Pro4715Leu substitution. This mutation has been dominant in the sequences across European countries. It is found in the RdRp region, which is involved in proofreading activities in the presence of other viral cofactors, like ExoN, nsp7, and nsp8 (Pachetti et al., 2020; O'Meara et al., 2020). It is anticipated that this mutation might affect proofreading capability. A minor alteration in the RdRp structure, without changing its catalytic activity, can also alter its binding efficiency with other cofactors such as ExoN, nsp7 or nsp8, and may contribute to change in the rate of mutation (Pachetti et al., 2020). Extensive studies are needed to understand the effect of mutations in RdRp, especially on viral replication. Two synonymous mutations at position C15324T and C18877T in NSP11 with moderate frequency (about 20%) were also observed in our study. We also observed another mutation with >6% frequency at position C13730T. Interestingly, this mutation is also reported in other states of India as well as in Singapore and Malaysia. This leads to a non-synonymous mutation Ala4489Val (Orf1ab) that corresponds to Ala97Val substitution in the RdRp protein. It has been suggested to substitute the alpha helix at positions 94-96 with beta-sheets, altering the tertiary conformation thereby might affect the fidelity of polymerase (Banerjee et al., 2020). It is also likely that these mutation lead to minor changes without affecting the functional activity of these proteins and thus adaptation in the host cell.

Putative cation channel encoded by Orf3a (about 274 amino acid long) is important for viral release, cell death, and inflammation. Our data indicated a significant number of mutations (G25552T, G25563T, C25886T, G26056T, G26065T, C26110T, and G26152T). All these
mutations have been reported earlier in globally available sequences (source GISAID). It is known that the deletion of Orf3a in related SARS-CoV-1 leads to reduced viral titers and morbidity. Hence, it is important to understand how these Orf3a mutations contributing to COVID19 disease severity. Orf3a region translates a unique membrane protein with three-transmembrane, and G25563T variation in this region seems important (Cortey et al., 2020). We detected G255563T mutation at a frequency of 16.7% (n=15/90). This leads to mutation of Gln57 that lines the ion channel pore and forms the inner hydrophilic constriction to histidine. However, a recent study showed that the mutation Gln57His does not affect the assembly or functional properties of this putative cation channel (Kern et al., 2020).

SARS-CoV-2 N protein, a promising drug target for COVID-19, is a multifunctional RNA-binding protein required for RNA transcription and replication in virus. Thus play an important role in viral RNA genome packaging. Three distinct yet conserved domains of the N proteins are an N-terminal RNA-binding domain (NTD), a C-terminal dimerization domain (CTD), and intrinsically disordered central Ser/Arg (SR)-rich linker. Our study revealed a list of N protein mutations that is similarly reported in globally available sequences at GISAID. We noted that consecutive mutations at positions GGG28881--28883AAC on SARS-CoV-2 was observed in 47.8% sequences and is widely reported across the countries. These mutations lead to a non-frameshift substitution on Orf9. Notably, this has been reported in 23% of worldwide cases. G28881A and G28883C exist in the nucleocapsid gene and lead to variations in amino acid Arg203Lys and Gly204Arg, respectively. The Gly204Arg results in amino acid change with significantly different isoelectric point, and thus could potentially affect nucleocapsid protein structure and function (Pachetti et al., 2020). Non-synonymous mutation C28311T (Pro13Leu) in the nucleocapsid protein which is mandatory for the viral entry into the cells will likely have a significant effect on virulence was found in 6.7% of the sequenced samples.

The mutation C241T was found in 90% of all the sequences and is located in the 5’ UTR region and was found predominantly severely affected patients. It has also been shown to coincide with other mutations like C3037T (nsp3), C14408T (RdRP) and A23403G (S) in mildly affected, whereas in severe infections it coincided mostly with A23403G (S) (Biswas and Mudi, 2020). However, the role of this mutation has not been unraveled yet.

In the present study, a distinct novel pattern of mutations in closely located different districts of the same state was noted, perhaps due to strict, and effective lockdown. In Pune, a novel
variant has overtaken the other strains. Satara and Nasik also have unique mutations but at a lower frequency. It would be interesting to study how these mutations affect viral fitness and virulence. Nonetheless, the lockdown resulted in an independent evolution of SARS-CoV-2 viral variants.

The host may also play a significant role in influencing the mutation pattern. One such host factor could be age, and age-wise specific pattern of mutations was detected in the present study. Mutations C6843T, C18877T, G20326A, G24794T, G25563T, G26152T, and C26735T were prevalent in more than 30% of the sequences were in the age group of 10-25 years. Surprisingly, these mutations were not found (0%) in a higher range of age, i.e., 61-80 years. The prevalence of viral variants might be dependent upon host-immunity and host genetic state, which is very different across elderly and young humans. Our results indicate that certain viral variants having specific set of mutations might be able to infect the young population more than the other less evolutionally advanced virus sub-clones. This is an exciting finding and demands further study in a larger cohort.

The mutations, i.e., C313T, C5700A, G28881A, were found more in samples from symptomatic subjects, specifically in symptomatic females. The same mutation was also found as a pattern in the present study supercluster 4. Hence, the linked-mutation, which is also evolving as a cluster in the present study, may have a role in exhibiting symptoms in a host-specific manner. Based on host immunity and physiology, it may be associated with the symptoms in that population where these mutations are prevalent. However, this observation needs to be validated with a much larger population-based study.

Additionally, in the Indian context, it was found that distinct sub-clones of virus were prevalent in different parts of India at the same time period. During April-May, 2020, the type 19A clade virus was predominant in the northern part of India (Delhi), which is an older clade whereas the western part (Maharashtra) was dominated by a more evolved clade, i.e., 20A, 20B. At the same time southern part (Telangana), where 19A clade was dominant in April, shifted completely to 20A and 20B in May 2020. Due to lockdown, most confounding factors (mobility of humans, inter-district transport, etc.) in the transmission of SARS-CoV-2 was restricted. Thus, we could assert that the prevalence of a specific viral variant in a region could be attributed to human host susceptibility for specific viral variants. In turn, this susceptibility seems to be based on mutations prevalent in the viral variants in that region. These factors also support the observation that
indifferent parts of India, infection rate, viral load as well as mortality rate are variable during the same period.

**Acknowledgement**

All the authors acknowledge the GISAID database and the concerned contributors of genomic data. The authors wish to thank the Department of Biotechnology, Ministry of Science and Technology, Government of India for the financial support and Dr. T. P. Lahane, Director, Department of Medical Education and Research for the permission to use the clinical specimens for this study. The authors would also like to appreciate the efforts of all those involved in the diagnosis, without their contribution it would not have been possible to generate the samples for sequencing.

**Contribution**

Conceptualized and designed the study: YSS, AS, MKB, VS; Sample processing and sequencing: DP, KJ; Bioinformatics and statistical analysis: DP; Protein modeling: JK, RC; Data interpretation: DP, AS, YSS, JK, RC, VS, GL, MKB, RK, SS; Manuscript writing: DP with input YSS, AS; Manuscript improvement: AS, YSS, JK, MKB, VS, SK; Sample supply and coordination: RK, SJ, MT, SS, SK, KBA, SPSS, RM. All authors have read and approved the manuscript.

**Conflict of Interest:** Authors have no conflict of interest

**References**


Cortey M, Li Y, Díaz I, Clilverd H, Darwich L, Mateu E. SARS-CoV-2 amino acid substitutions widely spread in the human population are mainly located in highly conserved segments of the structural proteins. bioRxiv (2020).


Legends to Figures

**Figure 1:** Phylogenetic analysis of the sequences used in the present study. Whole-genome sequences of the study sequences along with Wuhan SARS-CoV-2 genome (NCBI GenBank accession MN908947.3) as reference genome are used for the analysis.
**Figure 2:** Heat map representing all the mutations found in the present study. Red color indicates the presence of the mutation, and blue color indicates the absence of mutation. The first column represents the mutation position in the nucleotide level, and the first row represents samples Id of the present study.

**Figure 3:** Region-wise mutation pattern of the Maharashtra state during the lockdown condition. The bottom-most bar represents nucleotide positions in various ORFs where mutations occurred on the genome (gene locations are based on Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, NCBI GenBank accession MN908947.3 (https://www.ncbi.nlm.nih.gov/nuccore/MN908947)). Specific mutation observed in more than 5 samples are considered in this plot. Sample size: Pune (n=33), Satara (n=32), and Nashik (n=17).

**Figure 4:** Gender-wise mutation pattern detection. The bottom part of the bar plot represents a nucleotide position where mutations occur on the genome. Specific mutation observed in more than 5 samples are considered in this plot. Sample size for female (n=31) and male (n=51).

**Figure 5:** Age-wise mutation pattern identification. The bottom part of the bar plot represents a nucleotide position where mutations occur on the genome. Specific mutation observed in more than 5 samples are considered in this plot. Sample size for age range 10-25 (n=18), 26-60 (n=60) and 61-80 (n=10).

**Figure 6:** Symptom-wise mutations’ pattern identification. Specific mutation observed in more than 5 samples are considered in this plot. Sample size for patient of symptomatic (n=37) and asymptomatic (n=49).
**Figure 7:** Temporal change in occurrences of five major types of SARS-CoV-2 (based on recent reclassification method) in different states of India. Sequences used for this analysis, retrieved from the GISEAD database.

**Supplementary information**

**Supplementary Figure 1:** Symptom-wise mutation pattern identification females. Specific mutation observed in more than 5 samples are considered in this plot. Sample size for patient of symptomatic (n=13) and asymptomatic (n=15)

**Supplementary Figure 2:** Symptom-wise mutation pattern identification in males. Specific mutation observed in more than 5 samples are considered in this plot. Sample size for patient of symptomatic (n=20) and asymptomatic (n=34)

**Supplementary Figure 3:** Phylodynamics clusters and the clades assigned based on the new reclassification method for the Indian SARS-CoV-2 Genomes (943 samples), including present study samples (90).

**Supplementary Figure 4:** UPGMA-based cluster analysis for identification of closely associated mutation pattern

**Supplementary Data 1:** Metadata information of the present study samples

**Supplementary Data 2:** List of sequences deposited from India retrieved from GISAID database, which are included in the present study for comparative study and phylodynamics analysis

**Supplementary Data 3:** List of all mutations, abundance, and their mutational consequences

**Supplementary Data 4:** Statistical analysis
Figure 4

The figure shows the mutation frequency at various nucleotide positions for both female and male samples. The x-axis represents the nucleotide positions, and the y-axis represents the mutation frequency (%). The bar charts illustrate the distribution of mutations at different positions, with each bar indicating the percentage of mutations at that specific position for each gender.
Figure 5

The figure shows bar charts for mutation frequency (%) in different age groups:

- **Age 61-80**: Mutation frequency is high across various samples, with a concentration of bars above 50%.
- **Age 26-60**: Mutation frequency is lower compared to the older age group, with most bars clustered around 10-20%.
- **Age 10-25**: Mutation frequency is even lower, with most bars below 10%.

The x-axis represents different samples, and the y-axis represents mutation frequency (%).
Figure 7

April, 2020

- Delhi: n=26
  - Clade 19A: 100
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 0

- Gujarat: n=72
  - Clade 19A: 50
  - Clade 19B: 50
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 0

- Maharashtra: n=69
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 50
  - Clade 20B: 0
  - Clade 20C: 0

- Odisha: n=15
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 100
  - Clade 20B: 0
  - Clade 20C: 0

- Telangana: n=68
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 100
  - Clade 20C: 0

- West Bengal: n=18
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 100

May, 2020

- Delhi: n=29
  - Clade 19A: 100
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 0

- Gujarat: n=129
  - Clade 19A: 50
  - Clade 19B: 50
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 0

- Maharashtra: n=66
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 50
  - Clade 20B: 0
  - Clade 20C: 0

- Odisha: n=110
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 100
  - Clade 20B: 0
  - Clade 20C: 0

- Telangana: n=95
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 100
  - Clade 20C: 0

- West Bengal: n=20
  - Clade 19A: 0
  - Clade 19B: 0
  - Clade 20A: 0
  - Clade 20B: 0
  - Clade 20C: 100
Supplementary Figure 2

The relative frequency of mutations (%) for Male and Symptomatic individuals is shown. The x-axis represents different mutation identifiers, and the y-axis shows the percentage frequency.
Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past

Vibhuti Kumar Shah1,2†, Priyanka Firmal1,2†, Aftab Alam2,3, Dipyaman Ganguly3 and Samit Chattopadhyay1,2,3*

1 Department of Biological Sciences, BITS Pilani, K. K. Birla Goa Campus, Goa, India, 2 National Centre for Cell Science, S. P. Pune University Campus, Pune, India, 3 Indian Institute of Chemical Biology, Kolkata, India

After the 1918 flu pandemic, the world is again facing a similar situation. However, the advancement in medical science has made it possible to identify that the novel infectious agent is from the coronavirus family. Rapid genome sequencing by various groups helped in identifying the structure and function of the virus, its immunogenicity in diverse populations, and potential preventive measures. Coronavirus attacks the respiratory system, causing pneumonia and lymphopenia in infected individuals. Viral components like spike and nucleocapsid proteins trigger an immune response in the host to eliminate the virus. These viral antigens can be either recognized by the B cells or presented by MHC complexes to the T cells, resulting in antibody production, increased cytokine secretion, and cytolytic activity in the acute phase of infection. Genetic polymorphism in MHC enables it to present some of the T cell epitopes very well over the other MHC alleles. The association of MHC alleles and its downregulated expression has been correlated with disease severity against influenza and coronaviruses. Studies have reported that infected individuals can, after recovery, induce strong protective responses by generating a memory T-cell pool against SARS-CoV and MERS-CoV. These memory T cells were not persistent in the long term and, upon reactivation, caused local damage due to cross-reactivity. So far, the reports suggest that SARS-CoV-2, which is highly contagious, shows related symptoms in three different stages and develops an exhaustive T-cell pool at higher loads of viral infection. As there are no specific treatments available for this novel coronavirus, numerous small molecular drugs that are being used for the treatment of diseases like SARS, MERS, HIV, ebola, malaria, and tuberculosis are being given to COVID-19 patients, and clinical trials for many such drugs have already begun. A classical immunotherapy of convalescent plasma transfusion from recovered patients has also been initiated for the neutralization of viremia in terminally ill COVID-19 patients. Due to the limitations of plasma transfusion, researchers are now focusing on developing neutralizing antibodies against virus particles along with immuno-modulation of cytokines like IL-6, Type I interferons (IFNs), and TNF-α that could help in combating the infection. This review highlights the similarities of the coronaviruses that caused SARS and MERS to the novel SARS-CoV-2 in relation to their pathogenicity and immunogenicity and also focuses on various treatment strategies that could be employed for curing COVID-19.

Keywords: coronavirus, immune response, COVID-19, T cells, MHC presentation, HLA, memory T cell
INTRODUCTION

The whole world is currently confronting a crisis situation that first appeared in late December 2019 as merely a few cases of pneumonia in Wuhan, China. The patients were exhibiting common symptoms like fever, dry cough, sore throat, breathlessness, and fatigue. Sample swabs from the oral cavity and anal region were collected along with the blood and Bronchoalveolar Lavage Fluid (BALF) from all seven of the patients, irrespective of their age and gender, which were then sent to the Wuhan Institute of Virology for further examination. As the outbreak initiated at the seafood market with the onset of winter, similar to that of the previous Severe Acute Respiratory Syndrome (SARS) infection, the scientists first screened the samples using pan-CoV qPCR primers. Surprisingly, five samples were reported positive for coronavirus. Thorough investigation employing next-generation sequencing and phylogenetic analysis led to the identification of the causative agent of this respiratory disease, a novel coronavirus (2019-nCoV) (1). As more cases started to appear around the world, on February 11, 2020, the World Health Organization assigned a name, COVoronavirus Disease 2019 or COVID-19, to the disease and declared it a pandemic on March 11, 2020. The virus was renamed from 2019-nCoV to SARS-CoV-2 by the International Committee on Taxonomy of Viruses on the basis of its genetic similarity to a previously known coronavirus, Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) (2). Transmission of SARS-CoV-2 occurs when a healthy individual inhales or comes into contact with respiratory droplets from an infected person. The average incubation period before patients exhibit disease symptoms ranges from 2 to 14 days (3). Before the spread of COVID-19, SARS emerged as an epidemic in 2003, followed by Middle East respiratory syndrome (MERS) in 2012, both caused by a novel coronavirus of zoonotic origin and assigned to the genus Beta-coronavirus (4). The worldwide outbreak of SARS-CoV-2 has put life on hold, having a major impact on the world's economy, and has claimed ~436,167 lives globally as of June 15, 2020 (5, 6). Unlike previous episodes of coronavirus spread, where it took months to identify the cause of infection and perform genome sequencing (7), advancement in science and technology made it possible to identify the causative organism swiftly. Within a few weeks of the outbreak, different laboratories across the world had sequenced the whole viral genome and had also provided structural and functional insights into the essential proteins required by the virus for its survival. These immediate scientific inputs helped with developing diagnostic kits and defining treatment strategies for effective prognosis and prevention (8–10). In this review, we are emphasizing the immunological aspect of SARS-CoV-2 pathogenesis by taking into consideration the previous experimental and clinical knowledge obtained from the coronaviruses that were responsible for causing SARS and MERS. This approach will assist in utilizing immunotherapies, repurposing the previously approved antiviral drugs, and developing therapeutic vaccines specific to novel coronavirus more effectively.

CLASSIFICATION AND COMPARISON OF SARS-CoV-2

Initial genome sequencing and phylogenetic analysis of novel coronavirus SARS-CoV-2 has shown that it is genetically similar to previously known coronavirus SARS-CoV and hence is placed under the family Coronaviridae. Coronavirus contains positive-sense single-stranded RNA (+ve ssRNA) as its genetic material, which can be about 30 kb in length and is mostly protected by an outer fatty layer of an envelope that also helps the virus to evade host immune response and assists its entry inside the host cell (11, 12). The subfamily Coronavirinae is further subdivided into four genera, namely alpha-, beta-, gamma-, and delta- coronavirus (α-CoV, β-CoV, γ-CoV, and δ-CoV). Viruses having the potential to infect humans are placed under the genus α-CoV and β-CoV (SARS-CoV & MERS-CoV), whereas viruses of γ-CoV and δ-CoV genera are mostly known to infect avians and pigs (13). The novel coronavirus, SARS-CoV-2 falls under the genus β-CoV, as it shares 88% sequence identity with SARS-CoV-like coronaviruses (derived from bat) but is only 79% identical to SARS-CoV and 50% identical to MERS-CoV (3). Thus, it can be deduced by its genome identity that the immediate host of this virus could be a bat, which then transmitted it to some unknown intermediate host that acted as a source for the transmission of the virus to humans.

Like those of SARS-CoV and MERS-CoV, the SARS-CoV-2 genome comprises of 12 open reading frames (ORFs) in number. At the 5’ end of the viral genome, overlapping ORFs 1a and 1b are present that encode the RNA polymerase and other non-structural proteins of the virus and occupy approximately two-thirds of the genome. Genes encoding structural proteins such as spike (S), membrane (M), envelope (E), and nucleocapsid (N), are present in the remaining one-third of its genome spanning from the 5’ to the 3’ terminal, along with several genes encoding non-structural proteins (NSPs) and accessory proteins scattered in between, as shown in Figure 1. Despite being in the same serogroup, there is a slight difference in the nucleotide number, sequence, gene order, and expression method among previously known coronaviruses and the novel SARS-CoV-2 (1, 14, 15). Recent reports highlight that a few amino acid substitutions have occurred in the novel coronavirus genes encoding the S protein, NSP2, NSP3, and receptor-binding domain (RBD). These mutations in the NSP2 & NSP3 are also believed to impart the enhanced infection abilities of the novel coronavirus (16, 17). RNA viruses are prone to acquiring genetic mutations that eventually help them to escape the host immune system and develop drug resistance. Researchers have also found minor mutations in SARS-CoV-2 genotype in different COVID-19 patients (18). One such hotspot of mutation in the SARS-CoV-2 genome is the RNA-dependent RNA polymerase gene. On analyzing 220 sequences across the globe, eight repetitive novel point mutations were observed. Viral genetic sequences accessed from Europe exhibited five mutation hotspots, whereas the remaining three point mutations were solely present in the sequences from North America. These unique mutations suggest that the viral strains are continuously evolving across the globe.
and that the strains from Europe, North America, and Asia might have co-existed the whole time (19). Another similar report analyzed 7,666 global viral genomic sequences and found 198 unique mutation sites on SARS-CoV-2 genome that encodes NSPs and S protein, suggesting that the virus is trying to adapt to its new host (20). As numerous drugs are currently being designed to target the proteins that are essential for the survival of the virus, rapid genetic mutation occurring in these proteins...
Therefore, the invariable region of the virus could be a better target to avoid drug failures.

Interestingly, SARS-CoV-2, similar to SARS-CoV, exploits the angiotensin-converting enzyme 2 (ACE2) receptor to gain access inside human cells, whereas MERS-CoV binds specifically to Dipeptidyl Peptidase 4 (DPP4) receptor (21, 22). Binding of the virus particle to the specific receptor on the host cell plays a key role in governing its pathogenicity. Functional evaluation was carried out to reveal the potential receptors for different Betacoronaviruses (β-CoV) including SARS-CoV-2, and it was found out that the entry of the virus particle was enhanced in human cells expressing ACE2 receptor instead of DPP4 or Aminopeptidase N (APN) in the case of the novel coronavirus (23). Recent structural insights provided by Cryo-EM studies of S protein in prefusion conformation highlighted that the binding efficiency of ACE2 and S protein of SARS-CoV-2 is 10–20 times greater than for the previously known SARS-CoV (24, 25). The latest reports suggest that the trimeric S protein of SARS-CoV-2 is sliced by transmembrane protease serine 2 (TMPRSS2), similar to SARS-CoV (26, 27). Hence, profound knowledge of the potential receptors to which the virus particle can bind and its associated proteases will help us in designing specific antiviral drugs and neutralizing antibodies and will lead us to foresee whether particular coronaviruses of zoonotic origin could be able to adapt and infect humans.

CORONAVIRUS REPLICATION

All coronaviruses initiate entry inside the target cell by engaging the host receptor with the S glycoprotein present on their surface so as to gain entry inside the target cell. The region of S protein containing the RBD is present on the S1 subunit. In a few coronaviruses, RBD is present at the N-terminus region of S1, whereas in SARS-CoV, it is situated at the C-terminus region (28, 29). The fusogenic activity of virus-cell membrane is governed by two tandem domains, heptad repeats (HR1,2) that are present on the S2 region of S protein (30, 31). Initially, it was believed that SARS-CoV enters the target cell merely by virtue of cell membrane integration of virus particle and host cell membrane (32). Later, it was discovered that an essential proteolytic cleavage event takes place in the S protein at the S2 position of SARS-CoV that results in membrane fusion and facilitates virus entry inside the cell (33).

Once the coronavirus is inside the host cell via membrane fusion, it releases its +ve ssRNA genome into the cytoplasmic compartment, where the translation of ORF-1a and ORF-1b begins resulting in the formation of two large polyproteins (pp1a and pp1ab). Three functional proteases then cleave the polyproteins into 16 non-structural proteins (NSP1-16), which eventually create the viral RNA polymerase and other accessory proteins for virus assembly (34–36). An uninterrupted replication-transcription event results in the formation of various nested sets of subgenomic (sg) mRNAs that eventually translate into numerous structural and accessory proteins (37). The E glycoproteins after synthesis are incorporated into the rough endoplasmic reticulum or Golgi membrane. The +ve ssRNA combines with capsid protein to form the nucleocapsid, followed by budding of assembled virus particles in the ER-Golgi Intermediate Compartment (ERGIC) (38). Lastly, the virus particle-loaded vesicles are fused with the cell membrane for effective shedding of the virus (4). These new virions are now accessible to infect the neighboring healthy cells and are also released into the surrounding environment via respiratory droplets that are highly contagious and hence potentially spread the disease to healthy individuals.

PATHOGENESIS OF COVID-19

The path followed by SARS-CoV-2 to reach the lungs is via the naso-oral cavity. Once the virus is inhaled, it enters the epithelial cells of the nasal cavity by engagement of ACE2 receptor with the viral RBD and initiates its replication (27, 39, 40). This initial asymptomatic phase lasts for about 1–2 days, during which the virus multiplies in the upper respiratory tract, where no major hindrance is caused by the innate immune cells. Within 2–14 days of initial encounter, the common symptoms of COVID-19 start to appear, which are similar to those of SARS and MERS, i.e., fever, dry cough, pharyngitis, shortness of breath, joint pain, and tiredness. Numerous problems arise during this phase of the disease, including nosocomial and fomite transmission of infection, which enhances the chances of community spread (41). Soon, the virus begins to move toward the lower respiratory tract via airways, and this triggers a strong innate immune response. Patients at this stage start exhibiting enhanced pro-inflammatory response that leads to viral sepsis accompanied by other complications, including pulmonary edema, Acute Respiratory Distress Syndrome (ARDS), different organ failures, and death in the worst scenarios (42). The infected individuals rarely show the intestinal symptoms like diarrhea that were evident in other coronavirus infections. Patients are recommended to be quarantined to prevent community spread of this pandemic virus (43). The severity of COVID-19 has been found to be greater in aged individuals and in people with a health history, such as those immune-compromised by HIV infection or by chemotherapy for cancer. Diabetic and asthma patients, along with individuals with hypertension, obesity, or heart, kidney, or liver disorders, are also at higher risk if they acquire the disease (44). Autopsy reports of individuals who died due to SARS show multi-organ dysfunction, with the highest viral titers in the lungs and immune cells in circulation, thus damaging the pulmonary and immune system (45, 46). As opposed to adults, only a very small population of children has been infected with SARS-CoV-2. In one study, the symptoms displayed by children above 15 years were found to be milder as compared to those of younger children, who showed severe symptoms but with rare deaths and better prognosis (47). The study speculated two major possibilities related to COVID-19 severity in children among different age groups. One of these rests on the finding that ACE2 activity is higher in children aged 4–13 years; after this age, it starts to decline until adolescence. This could be one of the reasons why lung fibrosis is observed mainly in younger children.
Secondly, differential CD4⁺ and CD8⁺ T cell populations have been seen in children as compared to adults (48, 49). A large number of clinical and epidemiological criteria were defined to assess probable pediatric cases of COVID-19 (50). A preliminary report from a cross-sectional study of children admitted to US and Canadian Pediatric Intensive Care Units (PICUs) during March 14-April 3, 2020, revealed that the 48 children were admitted in the USA whereas no COVID-19 cases were reported in Canadian PICUs. The study revealed that there are fewer COVID-19 cases in children as compared to adults and that there is a median PICU time of 5 days (51). A recent preprint from Paris reports that 11 children (age 3.7–16.6) were admitted experiencing symptoms similar to Kawasaki disease (KD) along with gastrointestinal issues and elevated inflammatory markers. Further investigation suggested that they were also SARS-CoV-2-positive, speculating that this could be the reason for KD shock syndrome (52). Similar cases have been observed in New York, where four otherwise healthy SARS-CoV-2-positive children started displaying symptoms similar to KD and toxic shock syndrome, thereby needing intensive care (53). Therefore, medical practitioners should be prepared to tackle such sudden post-infection complications to avoid the associated risks.

**IMMUNE RESPONSE TO SARS-CoV-2**

Once the virus gains access inside the target cell, the host immune system recognizes the whole virus or its surface epitopes, eliciting the innate or adaptive immune response (Figure 2). Pathogen recognition receptors (PRRs) present on immune cells, mainly Toll-like receptors 3, 7, and 8, are the first to identify the virus, which leads to enhanced interferon (IFN) production. The function of host innate immune cells is impaired during SARS-CoV and MERS-CoV infection by their non-structural proteins, which affects the overall cytokine production (54–56). Humoral response against SARS-CoV-2 has been found to be similar to that against other coronavirus infections, involving the characteristic IgG and IgM production. At the onset of SARS-CoV infection, B cells elicit an early response against the N protein, while antibodies against S protein could be detected after 4–8 days from the appearance of initial symptoms (57, 58). Although N protein is smaller than S protein, it is highly immunogenic, and the absence of glycosylation sites on it results in N-specific neutralizing antibody production at an early stage of acute infection (59). SARS-CoV-specific IgA, IgG, and IgM antibodies were detected after the onset of symptoms at different time points in infected patients. A persistent level of IgG was detected for a longer period, whereas IgM levels started to decline after 3 months (60, 61). In an observational case study of 16 SARS-CoV-2 patients, anti-S-RBD IgG was detected in all of the subjects, whereas anti-N IgG and anti-S-RBD IgM were detected in 15 patients and anti-N IgM in 14 patients (62). An ELISA-based time kinetics study to detect the COVID-19 specific humoral immune response showed that the patients produced IgM and IgG antibodies that did not cross-react with other human coronaviruses except SARS-CoV. IgM and IgA antibodies were detected 5 days after the onset of initial symptoms, whereas IgG was detected after 14 days (63). Another kinetic study of viral shedding and antibody detection was published in a preprint and reported the presence of higher IgG and IgM antibody titers in severe patients. They also observed that weak responders for IgG antibody had higher viral clearance than strong responders. This observation suggests that robust antibody response leads to disease severity while feeble response is associated with the elimination of virus (64). A case study on pediatric patients reports that 5 out of 6 children showed a protective humoral response, with neutralizing IgG and IgM antibodies targeting the N and S-RBD proteins of SARS-CoV-2 (65). These studies propose that IgM-based ELISA can be used for early diagnosis of patients along with qPCR techniques to improve the sensitivity and specificity of the technique.

In addition to neutralizing antibodies, which are defensive and useful, there are numerous non-neutralizing antibodies in the system that aid the infection of immune cells and APCs. Previously existing SARS-CoV antibodies may promote the viral infection in FcR-expressing cells (66). This ACE2-independent pathway of viral entry does not result in viral replication; rather, viral shedding by macrophages enhances inflammation and tissue injury by myeloid cell activation. This mechanism of viral entry through non-neutralizing antibody that results in aberrant activation of immune cells is called ADE (Antibody-Dependent Enhancement) (66, 67). ADE has been observed in a number of viral infections, including SARS and MERS.

In the case of SARS, anti-S antibodies were observed to be involved in ADE to gain entry into FcR-expressing cells (68), while in MERS, a neutralizing Mab (Mersmab1) targeting RBD aided in MERS pseudo-virus entry via the DPP4 pathway (69). Although there is no clear evidence regarding ADE in SARS-CoV-2 infection, it is still necessary to consider all of the odds in the pursuit of developing vaccines and treatment regimens involving antibodies (70).

**Antigen Presentation**

During viral infection, T cells also recognize the viral antigens presented by MHC class I [MHC; Human Leukocyte Antigen (HLA) in humans], which in turn promotes the cytokine release and cytokotoxic activity of CD8⁺ T cells (71). But in some other cases, MHC class II is also found to present SARS-CoV peptides to CD4⁺ T cells. Due to the genetic polymorphism of HLA, some haplotypes, like HLA-B*07, HLA-B*46, HLA-DRB1*12 (72), and HLA-Cw*08 (73), are found to be more susceptible to coronavirus infection, whereas the HLA-DRB1*03, HLA-A*02, and HLA-Cw*15 haplotypes are protected from SARS-CoV infection (74). Similarly, HLA-DRB1*11 and HLA-DQB1*02 were found to be vulnerable to MERS-CoV infection (75). Additionally, MHC expression is also found to be reduced during the infection due to epigenetic modifications of downstream molecules (76, 77). So far, HLA association is not very well-identified for SARS-CoV-2 infection, and this could be crucial for the prevention and treatment of COVID-19. However, in a recent report, blood plasma from COVID-19 patients was able to block the expression of HLA-DR on CD14⁺ monocytes, which was restored effectively on inhibiting IL-6, suggesting that decreased HLA-DR expression in SARS-CoV-2 patients
plausible host immune responses during COVID-19 infection. The SARS-CoV-2 virus infects through the naso-oral route, followed by infection in cells expressing ACE2 receptor in the lung, such as type 2 alveolar cells. These viruses dampen anti-viral IFN responses by evading the innate immune cells as a consequence of unrestrained virus replication. The infiltration of monocytes/macrophages, neutrophils, and several other adaptive immune cells leads to increased pro-inflammatory cytokines. In the helper T cell subset, stimulation of Th1/Th17 cells with viral epitopes may lead to aggravated inflammatory responses. This inflammatory response results in “cytokine storms” that lead to immunopathologies like pulmonary edema and pneumonia. Cytotoxic T cells recruited to the site of infection try to kill virus-infected cells in the lungs. B cells/plasma cells also recognize viral proteins and are activated to produce antibodies specific to SARS-CoV-2, which may help in deactivating viruses and provide systemic immunity in different organs.

is due to the buildup of hyper-inflammatory conditions (78). Decrease in MHC expression is also evident in cancer cells, which is a mechanism by which they evade the immune response by epigenetically modifying calnexin promoter. But infection with influenza virus in these cancer cells results in enhanced MHC-I presentation due to the increased expression of chromatin remodeling proteins, which stabilizes p53 expression and hence augments the immune surveillance of cancer cells (79). Therefore, molecules that can upregulate chromatin regulators and increase the MHC-I expression could potentially be used for COVID-19. Most of the T-cell epitopes presented by MHC complex are derived from structural proteins such as the S and N proteins of the coronavirus in both humans and animal models, while the NSPs have regulatory effects on the signaling cascade (80, 81). T cells can be stimulated by 14 epitopes, most of which are observed to be located on ORF3 and the S protein in SARS patients (61). In a large cohort study during SARS-CoV infection, S protein was the only immuno-dominant epitope for CD8+ T-cell activation (61), whereas, in MERS, CD8+ response was against the S and N proteins along with some of the M/E epitopes (82). These T-cell epitopes have been tested in animal models by assessing the lung pathology and T-cell response upon infection in BALB/c and C57BL/6 mice (80, 83). The sequence of SARS-CoV-2 being more similar to SARS-CoV than to MERS-CoV, with no mutation in 19 epitopes, provides a prospective subunit vaccine for stimulating a strong T-cell response in COVID 19 patients (84). In a recent study, samples from 20 convalescing COVID-19 patients were analyzed to check the development of adaptive immune response during infection. The results highlighted that helper T cells were eliciting a robust immune response against S, M, and N protein. The effect of adaptive immune response on humoral immunity was also compared, where a strong CD4+ T-cell response against SARS-CoV-2 eventually resulted in an increase in anti-S-RBD-specific IgG and IgA antibody titer. Along with CD4+ T cells, immunogenic epitopes on S, M, and N proteins were also able to activate CD8+ T cells. However, such T-cell response was not specific to recovered patients only but was also present in 40–60%
of the individuals who were not exposed to SARS-CoV-2. Further analysis showed that they had pre-existing cross-reactive CD4+ T cells, which might have been generated in response to some previous coronavirus infection. Hence, these T-cells could impart protective immunity in such individuals against SARS-CoV-2 to some extent (85). These epitopes could be a promising factor in developing immunotherapy by small molecules that can increase the presentation of viral epitopes.

**Cytokine Production**

A rapid and coordinated immune response during viral infection leads to enhanced secretion of various cytokines, which acts as a defense mechanism against the virus. Numerous reports suggest that individuals affected with SARS-CoV or MERS-CoV have dysregulated cytokine production from both innate and adaptive immune cells. In the case of SARS, infected hematopoietic cell, monocyte-macrophages, and other immune cells trigger enhanced secretion of pro-inflammatory cytokines like TNF-α, IL-6, and IFN-α/γ, with reduced anti-inflammatory cytokines (86–88). Similarly, MERS-CoV infection leads to delayed but increased production of IFN-α and pro-inflammatory cytokines like IL-6, IL-8, and IL-1β (89–91). Such elevated levels of cytokines were associated with Multi-Organ Dysfunctional Syndrome (MODS) and ARDS due to the accumulation of numerous immune cells like macrophages, neutrophils, and dendritic cells in the lungs causing alveolar damage and edema (56, 92, 93). Similarly, in COVID-19 patients, secretion of cytokines and chemokines, which attract the immune cells to the lungs, was increased, hence causing ARDS, which is fatal to critically ill individuals (94, 95). Signature cytokines in severely ill COVID-19 patients were consistent with those in SARS and MERS, i.e., enhanced expression of IL-6, TNF-α, macrophage inflammatory protein 1-α (MIP-1α), MCP3, GM-CSF, IL-2, and IP-10 along with elevated chemokines (IP-10, CCL2/MCP1, CXCL1, CXCL5) were also detected in SARS-CoV-2 infection (96–99). In children, the increased inflammatory markers include IL-6, IL-1, and C-reactive protein along with procalcitonin in serum (52). In a case study, a 14-year-old child with cytokine storm was treated with anakinra (IL-1 receptor antagonist) in order to stabilize the respiratory illness and other clinical symptoms (100). Transcriptomic analysis of PBMC and BALF showed that a number of immune regulators were upregulated, particularly CXCL10, with respect to BALF. This study also reported that several apoptotic genes and P53 signaling molecules were upregulated, suggesting a possible reason for lymphopenia in these patients (101). Therapeutic measures to control such cytokines involve neutralizing antibodies or small molecular drugs that can stop the signaling cascade for cytokine production.

**Immune Evasion**

The most potent antiviral machinery acquired by immune cells is the secretion of interferons that act as secondary messengers stimulating the neighboring cells. Most innate immune cells are efficient in producing IFNs that are involved in obstructing cell proliferation, apoptosis, and immunomodulation (54, 102). As an escape mechanism, SARS-CoV or MERS-CoV uses several ways to overcome the host immune response, one of which is by severe leukopenia and lymphopenia (103–105). After gaining entry to the cell, these viruses encode different proteins that interact with downstream signaling molecules of TLRs and the JAK-STAT pathway. MERS-CoV encoded matrix protein, accessory proteins from ORF 4a, 4b, and 5, which directly inhibits the IFN promoter and nuclear localization of IRF3 (106). PLpro, encoded by SARS-CoV and MERS-CoV, prevents the dissociation of NF-κB from IkBα, whereas nonstructural proteins of SARS-CoV, i.e., PLpro and ORF3b, inhibit IRF3 phosphorylation and hence its translocation to the nucleus (4, 107, 108). These viral accessory proteins also inhibit the JAK-STAT pathway, resulting in inhibition of genes by ISRE promoters (109–111) (Figure 3). A new investigation revealed that SARS-CoV-2 infection leads to an overall decrease in the transcription of antiviral genes because of the lower production of Type I and III interferons with sufficient ISG expression, along with elevated chemokine secretion. Results obtained from in-vivo and ex-vivo COVID-19 experiments were in tune with the in-vitro findings. Therefore, a decrease in the innate antiviral response, along with hyper-inflammation, could be one of the causes of COVID-19 severity (112). In addition to reduction in T cells, SARS-CoV-2 infection also enhances the exhaustion of effector T cells, decreasing the immune response against the virus (94, 113). Exhaustion and loss in function of effector T cells is the result of increased expression of inhibitory receptors like PD-1, TIM-3, and TIGIT on its surface as a result of cytokines like IL-6, IL-10, and TNF-α or by decreasing the regulatory T-cell population (114, 115).

**Memory T Cell**

Following viral/antigen clearance, most of the effector T cell undergoes apoptosis in the contraction phase. Subsequently, a pool of memory T cells are generated that are programmed to fight against re-infection. CD4+ memory T cells, upon re-stimulation, trigger B cells and other immune cells by cytokine production, while cytotoxic memory T cells help in destroying the infected cells during subsequent infection (116, 117). Case studies in recovered SARS patients showed that both CD4+ and CD8+ memory T cells were efficient in eliciting immune response from 3 months to 6 years without the presence of any antigens (118). In a case study of 23 recovered SARS-CoV patients, the patients showed very low frequencies of memory B cells, while memory T cells elicited a response against the S protein in 60% of recovered individuals (119). Considering the memory T-cell subset, N-specific helper T cells had more of central memory markers (CD45RA−, CCR7+, CD62L−) while the CD8+ T cell population had the effector memory (CD45RA+, CCR7−, CD62L−) phenotype in a steady-state manner (120). The study suggests that an effective vaccine or T cell epitopes could be used to target a particular population for rapid viral clearance. In recent reports, COVID-19 subjects have shown reduced regulatory T cell populations and memory T cells, which may aggravate the inflammatory response leading to cytokine storm and hence enhance the tissue damage and organ failure (114). In a mouse model, the use of CD4+ memory T cells as a vaccine by the intranasal, but not the subcutaneous, route imparted a protective response against the human coronavirus. The infused CD4+ memory T cell, upon re-stimulation, produces IFN-γ and
recruits CD8$^+$ T cells for rapid clearance in response to SARS-CoV-2 through the intra-nasal route. This tool will be beneficial for evaluating the efficacy of vaccines for COVID-19 and also to study its transmission and pathogenesis (122).

**TREATMENT STRATEGIES FOR COVID-19**

Just like SARS and MERS, there are no specific clinically approved drugs available for COVID-19 as of June 15, 2020 (123). Currently, the treatment regime focuses mainly on providing intensive care in order to alleviate the symptoms and discomfort associated with COVID-19. Conservative fluid therapy accompanied by broad-spectrum antibiotics are also given to the patients as a protective measure to avoid opportunistic bacterial infections. However, ventilator support for respiration is provided to the patient under extreme conditions (124). Numerous FDA-approved antiviral drugs, vaccines, and immunotherapies that are already being used to treat other diseases have also been considered as a possible approach for treating COVID-19 (Table 1). But this approach may reduce the availability of these drugs and vaccines for the intended diseases and for the patients with the greatest need. The molecular, structural, and functional relationships of
SARS-CoV-2 with SARS-CoV might define the use of existing anti-viral drugs against COVID-19 (147, 148), considering the total time it takes to perform clinical trials and get FDA approval for the use of novel drugs and vaccines. The increasing knowledge of the genetic, immunological, and molecular mechanisms behind its enhanced pathogenicity might help in developing specific treatment approaches for COVID-19 in the future.

### Antiviral Agents
Considering the studies on the molecular mechanism of coronavirus infection (147), several antiviral drugs could be repurposed for the treatment of COVID-19. Remdesivir is a nucleotide analog that acts as an antiviral agent for a wide variety of viruses and has been tested widely against previous epidemics of coronavirus infections in both in-vitro and in-vivo models (138, 149–151). This adenosine analog...
gets incorporated into the newly synthesized viral RNA, which inhibits the addition of further nucleotides by viral RNA-dependent RNA polymerase and hence terminates the ongoing transcription. Administration of intravenous remdesivir was found to be effective in treating the first known patient of COVID-19 in the USA (152). A randomized double-blinded clinical trial on 1,059 adult hospitalized COVID-19 patients was sponsored by the National Institute of Allergy and Infectious Diseases, USA, to further test the potency of intravenously administered remdesivir. The preliminary outcomes of the trial reported that remdesivir treatment decreased the median recovery time in the treatment group (11 days) as compared to the placebo group (15 days). The mortality rate was also less in the treatment group (7.1%) in contrast to the placebo group (11.9%) (153). Numerous clinical studies, similar to this, are required so as to validate the proposed drugs for COVID-19. Favipiravir, ribavirin, and galidesivir are also potential nucleoside analogs that might be useful against novel coronavirus infection (154). The combinatorial therapy approach of using remdesivir along with chloroquine, a well-known anti-malarial drug, has also been tested in vitro so as to study its effectiveness against SARS-CoV-2 (141, 155). It has been reported that chloroquine immuno-modulates the host microenvironment and also interferes with the replication of the virus and its interaction with the receptor (156, 157). In a randomized clinical trial (NCT04308668) involving 821 asymptomatic individuals across the US and Canada who had come into close contact with potential COVID-19 patients, the individuals were given either hydroxychloroquine or placebo as a prophylactic measure. The results revealed that hydroxychloroquine treatment had the same effect as did the placebo group. The usage of hydroxychloroquine resulted in minor side effects (40.1%) as compared to the placebo treatment (6.8%). However, no cardiovascular disorder or treatment-related major complications were observed (158). Based on the putative function of hydroxychloroquine on the endosomal acidification, whereby it is presumed to hinder viral uncapping, it can be observed that it has a great potential for prophylaxis, not to prevent infection but to reduce effective viral load in patients and thus lead to milder disease. Numerous clinical trials to further explore the usage of hydroxychloroquine in different combinations are in the pipeline and will finally provide a better understanding of the efficacy of this drug for COVID-19. A few anti-HIV drugs, such as lopinavir/ritonavir in combination with interferon beta (IFN-β), have been tested in vivo for treating coronavirus infections (SARS-CoV, MERS-CoV) and have also been used in the case of COVID-19 (138, 139, 159). Various complementary therapies could also be employed as a preventive measure against viral infections. Many essential proteases, such as chymotrypsin (3C-like protease) and PI-Pro, which are required by coronavirus for completing the replication process, can also be targeted using drugs. Cinanserin, flavonoids, and some small molecules are known to inhibit 3CLpro, whereas diarylheptanoids are used to inhibit PI-Pro (160–162). In a recent study, 16 potential anti-HCoV drugs were identified through a systems biology-based approach, such as melatonin, mercaptourine, sirolimus, dactiomyacin, and toremifene, which are to be tested further for their potency (163).

Plasma Therapy
In the absence of any dependable vaccine or drugs with tested efficacy and when the pandemic onslaught is ongoing, a worthy therapeutic approach is passive immunization using purified antibodies. The source of such antibodies could be the sera of convalescing individuals, mAbs, or genetically modified antibodies from an animal host, which can efficiently neutralize the virus. This is an age-old practice, with pioneering work having been done by the Nobel Laureate, Emil Behring, who applied this approach for diphtheria, and has been used whenever there are sudden outbreaks of viral diseases like SARS, MERS, H1N1, H5N1, Ebola, and many others (61, 164, 165). As opposed to active vaccination, plasma therapy is the only means to provide immediate immunity for viral clearance, as in the case of SARS-CoV-2. As in other epidemic diseases, convalescent sera are currently being employed for COVID-19 in a number of countries (166, 167). Although a randomized controlled trial is yet to be reported, limited studies in 10 patients have been documented with no remission of severe respiratory afflictions on receiving neutralizing antibodies from 39 convalesced donors with antibody titers of 1:160, along with drugs and oxygen support (168). A report from Hong Kong suggested that this therapy had poor outcome in SARS patients, with a number of limitations in their study (169). As with transfusion of any blood products, precautionary screening of infectious agent is warranted in plasma transfusion. Recently, the FDA in the USA has approved trials of convalescent plasma therapy in COVID-19 under specific guidelines; plasma donation is advised 3 weeks after a patient becomes virus-negative on PCR. The major challenge in this therapy is obtaining donors with similar blood antigens with a high antibody titer of SARS-CoV-2 (170). Another potential adverse effect of this approach is ADE of infection, which is common in so many other viruses. But, to date, the incidence of ADE has not been reported in the case of SARS-CoV-2. Another major point of contention is the selection of patients for this therapeutic approach. In most clinical trials, patients with severe diseases are being recruited, while the presumed mechanism of action of convalescent plasma, based on its content of virus-neutralizing antibodies, rather points to plausible favorable outcomes in earlier phases of the disease because in the later, more severe phases, the hyper-immune response, rather than the viral load, becomes the more critical pathology. Finally, there are no available data on the heterogeneity of response to convalescent plasma transfusion, which may further illustrate the importance of careful evidence-based patient selection, as heterogeneity of response may result from both virus and host-intrinsic factors which are, to date, not revealed.

Vaccine Design Strategies
Researchers around the world are working hard to develop a potential vaccine candidate so as to stop the deadly pandemic caused by SARS-CoV-2. However, vaccine development is not an easy task, as a number of successful clinical trials are required before approval for patients. Different approaches are being utilized for designing a specific vaccine targeting either the structural proteins or viral replication process, which
eventually results in the inhibition of viral growth and its further transmission. The common strategies involve the use of live attenuated vaccine (LAV), inactivated virus, subunit vaccines, monoclonal antibody vaccine, virus vectors, protein vaccines, and DNA/RNA-based vaccines (171–174). There are numerous subunit vaccines targeting all or a part of S protein that have already been tested for SARS and MERS in animal models (175) and could be potential candidates for testing against SARS-CoV-2. A recent pilot study with a purified inactivated SARS-CoV-2 virus vaccine displayed very promising outcomes in different animal models. The neutralizing antibodies generated after vaccination were able to effectively target 10 different strains of SARS-CoV-2 without developing any ADE of infection (176). Various randomized controlled trials (NCT04327206, NCT04328441) are also underway to evaluate the effectiveness of the BCG vaccine against SARS-CoV-2 for healthcare professionals. An adenovirus vector-based vaccine candidate, ChAdOx1 (presently AZD1222), developed by Oxford University (licensed to AstraZeneca) for use against SARS-CoV-2 has been reported to activate both the humoral and cell-mediated immune response when tested in rhesus monkey (177). The phase I clinical trial to confirm its potency is also in progress (NCT04324606). Another group has followed a similar approach by using a recombinant adenovirus type 5 (Ad5-nCoV) vector-based vaccine for COVID-19. The full report from the phase I clinical trial (NCT04313127) of Ad5-nCoV shows that it is very effective in generating both humoral and rapid T-cell response post immunization. The group is now ready for the next clinical trial phase to further strengthen the effectiveness of the Ad5-nCoV vaccine (178). It should be noted that there are potential risks associated with the usage of live attenuated viruses, for example, complications resulting in lung damage by infiltrating eosinophils, as seen in in vivo models (179, 180). However, eosinophil immunopathology due to SARS-CoV vaccine could be reduced by using TLR4 agonist as an adjuvant (181). Viral neutralizing antibodies specifically targeting various regions of S, i.e., S1-RBD, S1-NTD, or the S2 region, and blocking the interaction of virus with the receptor are well-known for SARS and MERS (182). These neutralizing antibodies could prove to be the best and potential candidate for cross-neutralization of SARS-CoV-2. Despite being structurally related, some of the SARS-CoV neutralizing monoclonal antibodies failed to interact with the S-protein of SARS-CoV-2, which could be attributable to the substantial differences in their RBD (183). A recent study reported the presence of high titres of neutralizing anti-S-RBD IgG antibodies, but no antibodies were detected against the N protein in recovered COVID-19 patients, suggesting that anti-S IgG persists longer than does anti-N IgG. Along with the humoral immune response, they also observed an S protein-specific T cell-population producing IFN-γ, which further contributes to conferring protective immunity against SARS-CoV-2 infection (184). Recently, a monoclonal antibody (47D11) has been identified from 51 SARS-Spike hybridomas that targets the conserved S-RBD region (residue 338–506) and therefore can very effectively neutralize SARS-CoV-2 along with SARS-CoV (185). On similar lines, a group has isolated a single-domain antibody from a phage display library targeting the S-RBD region of SARS-CoV-2. The fully humanized single-domain antibody was able to neutralize the virus by interacting with a cryptic epitope in S protein (186). These mAb and single-domain antibodies could be used to treat as well as to design quick diagnostic kits for COVID-19.

The new technology of the microneedle array (MNA) has been employed for delivering SARS-CoV-2 S1 subunit vaccine, which could be really helpful in the treatment of the emerging COVID-19 outbreak (187). The transfer of S1 subunit by MNA elicited a strong virus specific-antibody response in SARS-CoV-2 (187). A novel encapsulated mRNA vaccine candidate developed by ModernaTX, Inc. that encodes full length S protein of SARS-CoV-2, is also under clinical trial (NCT04283461). There is an urgent need to develop more such specific vaccines that could neutralize the novel coronavirus effectively (188).

**Immunomodulatory Therapies**

The host innate immune system encounters upcoming infections, and this results in elevated production of various cytokines and type I interferons (IFNs). In the case of prolonged infection, hyperactivation of the immune system may also result in the development of a pro-inflammatory microenvironment, leading to adverse outcomes and even death. The induction of numerous lymphokines, such as IL-6, IL-1β, TNF-α, and CCL2, that are pro-inflammatory in nature has also been observed in the case of COVID-19 (189–191). A previous study in a MERS animal model showed that treatment with recombinant type-1 IFN (rIFN) decreased the viral RNA level in lungs with a decrease in IFN-stimulating gene expression. Early treatment with rIFN resulted in a dampening of cytokine and chemokine release that lowered the migration of neutrophils and other cells in lung (91). An allogenic mesenchymal stem cell-based (Remestemcel-L) therapy developed by Mesoblast, which has been previously used for inflammatory conditions and graft vs. host disease in children and adults, is now being assessed for COVID-19 (192–194). In this therapy, bone marrow-derived MSCs from the donor are grown in vitro and are then transfused to the recipient patients. Upon infusion, these cells exhibit anti-inflammatory activity by reducing pro-inflammatory cytokine production via the recruitment of anti-inflammatory cells in the affected tissue (195). Currently, a randomized placebo-controlled trial (NCT04371393) with 300 patients is ongoing for treating ARDS caused by COVID-19. Treatment with rIFN, inhibitors of the pro-inflammatory pathway, cytokine inhibitors such as tocilizumab, lenzilumab, and many others are still to be used in combination with other drugs for treating COVID-19. So far, there is not much evidence from clinical trials of such inhibitors with which to predict the outcome of these anti-cytokine therapies.

**CONCLUSION**

Considering the current situation of more than 8 million people being infected, with ~436,167 deaths as of June 15, 2020, there is an urgent need to control the SARS-CoV-2 pandemic. The fatality rate of SARS-CoV-2 in lower than those of other coronaviruses that caused catastrophes in the
past, but the higher infectivity rate makes it worse. Raising awareness of this contagious virus is one of the many ways by which its spread can be prevented. The governing authorities concerned in every country have approved guidelines and taken necessary action to quarantine infected people and break the chain of community spread. Antibodies, vaccines, and drugs developed for previously emerged coronaviruses could potentially be used for treating SARS-CoV-2. The combination of various neutralizing antibodies against S protein could enhance the effectiveness of viral clearance. Among various antivirals and other small molecules that are FDA approved, chloroquine/hydroxychloroquine has shown better positive outcome in COVID-19 patients. In clinical trials, some of the combinational antiviral drugs like lopinavir + ritonavir and blockers like angiotensin receptor blocker that were thought to be effective, have failed in curing the disease (139, 196). Cytokine storm being one of the symptoms of infected individuals, anti-cytokine therapy for TNF and IL-6 should be attempted to determine the efficacy of these antibodies in the treatment of SARS-CoV-2 infection. Clinical trial ChiCTR2000029765 with tocilizumab, a monoclonal humanized antibody against IL-6 receptor, has shown some efficacy, but this still needs to be tested in a larger cohort. With the increasing number of deaths, there is an immense need to accelerate the development of rapid and sensitive diagnostic kits and to commence clinical trials of the readily available and safe drugs to reduce the rising infections and COVID-19-related deaths so as to bring life back on track.

**AUTHOR CONTRIBUTIONS**

VS and PF contributed equally in writing the review. Conception of idea was done by SC, VS, and PF. Manuscript writing and editing was done by all the authors.

**FUNDING**

We are thankful to CSIR-Indian Institute of Chemical Biology, Jadavpur, Kolkata and National Centre for Cell Science (NCCS), Pune for providing infrastructure facilities. We would also like to acknowledge Department of Biotechnology-Systems Medicine Cluster (DBT-SyMeC) grant and J C Bose fellowship-SERB (Science and Engineering Research Board) to SC, for the financial support.

---

**REFERENCES**


Shah et al.

63.

64.

65.

66.

67.

68.

69.

70.
71.

72.

73.

74.

75.

76.

77.

78.

79.

80.

Immune Response to SARS-CoV-2

81. Huang J, Cao Y, Du J, Bu X, Ma R, Wu C. Priming with SARS CoV
S DNA and boosting with SARS CoV S epitopes specific for CD4+ and
91. doi: 10.1016/j.vaccine.2007.06.047
82. Zhao J, Alshukairi AN, Baharoon SA, Ahmed WA, Bokhari AA,
Nehdi AM, et al. Recovery from the Middle East respiratory
syndrome is associated with antibody and T cell responses.
Sci
Immunol.
(2017)
2:1–11.
doi:
10.1126/sciimmunol.aan5
393
SARS coronavirus nucleocapsid immunodominant T-cell epitope cluster is
common to both exogenous recombinant and endogenous DNA-encoded
11.042
84. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of
potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV2) Based on SARS-CoV immunological studies. Viruses. (2020) 12, 1–
15. doi: 10.1101/2020.02.03.933226
et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with
doi: 10.1016/j.cell.2020.05.015
86. Cameron MJ, Bermejo-Martin JF, Danesh A, Muller MP, Kelvin DJ. Human
immunopathogenesis of severe acute respiratory syndrome (SARS). Virus
87. Wong CK, Lam CWK, Wu AKL, Ip WK, Lee NLS, Chan
IHS, et al. Plasma inflammatory cytokines and chemokines in
changes in cytokine/chemokine profiles and pulmonary involvement
in
severe
acute
respiratory
syndrome.
Respirology.
(2006)
89. Lau SKP, Lau CCY, Chan KH, Li CPY, Chen H, Jin DY, et al. Delayed
induction of proinflammatory cytokines and suppression of innate antiviral
response by the novel Middle East respiratory syndrome coronavirus:
90. doi: 10.1099/vir.0.055533-0
Active replication of middle east respiratory syndrome coronavirus and
aberrant induction of inflammatory cytokines and chemokines in human
42. doi: 10.1093/infdis/jit504
JE, Mack M, et al. IFN-I response timing relative to virus replication
129:3625–39. doi: 10.1172/JCI126363
92. Ng DL, Al Hosani F, Keating MK, Gerber SI, Jones TL, Metcalfe MG,
et al. Clinicopathologic, immunohistochemical, and ultrastructural findings
of a fatal case of middle east respiratory syndrome coronavirus infection
8. doi: 10.1016/j.ajpath.2015.10.024
93. Kong SL, Chui P, Lim B, Salto-Tellez M. Elucidating the
molecular physiopathology of acute respiratory distress syndrome
95. Zhe Xu, Lei Shi, Yijin Wang, Jiyuan Zhang, Lei Huang, Chao Zhang,
et al. Pathological findings of COVID-19 associated with acute
doi: 10.1016/S2213-2600(20)30076-X
96. Mehta P, Mcauley DF, Brown M, Sanchez E, Tattersall RS, Manson JJ.
COVID-19 : consider cytokine storm syndromes and immunosuppression.
T cells and inflammatory monocytes incite inflammatory storm in

responses during infection by SARS-CoV-2 : an observational cohort study.
2020.03.24.20042382. doi: 10.1101/2020.03.24.20042382
4. doi: 10.1038/s41423-020-0438-3
Iwasaki A, Yang Y. The potential danger of suboptimal
20:1–3. doi: 10.1038/s41577-020-0321-6
Ricke DO, Malone RW. Medical countermeasures analysis of 2019-nCoV
and vaccine risks for antibody-dependent enhancement (ADE). Preprint.
(2020) 1–24. doi: 10.20944/preprints202003.0138.v1
enhancement of SARS coronavirus infection and its role in the pathogenesis
mechanism for antibody-dependent enhancement of coronavirus entry. J
Peeples L. Avoiding pitfalls in the pursuit of a COVID-19 vaccine. Proc Natl
immunodominant peptide presentation strategy: a featured HLA-A∗ 2402restricted cytotoxic t-lymphocyte epitope stabilized by intrachain hydrogen
bonds from severe acute respiratory syndrome coronavirus nucleocapsid
Association of human leukocyte antigen class II alleles with severe acute
Chen YMA, Liang SY, Shih YP, Chen CY, Lee YM, Chang L, et al.
Epidemiological and genetic correlates of severe acute respiratory syndrome
coronavirus infection in the hospital with the highest nosocomial
65. doi: 10.1128/JCM.44.2.359-365.2006
Wang SF, Chen KH, Chen M, Li WY, Chen YJ, Tsao CH, et al. Humanleukocyte antigen class i Cw 1502 and Class II DR 0301 genotypes are
associated with resistance to severe acute respiratory syndrome (SARS)
Hajeer AH, Balkhy H, Johani S, Yousef MZ, Arabi Y. Association
of human leukocyte antigen class II alleles with severe Middle East
11:211–3. doi: 10.4103/1817-1737.185756
Josset L, Menachery VD, Gralinski LE, Agnihothram S, Sova P, Carter VS,
et al. Cell host response to infection with novel human coronavirus
EMC predicts potential antivirals and important differences with
00165-13
Menachery VD, Schäfer A, Burnum-Johnson KE, Mitchell HD, Eisfeld AJ,
Walters KB, et al. MERS-CoV and H5N1 influenza virus antagonize antigen
presentation by altering the epigenetic landscape. Proc Natl Acad Sci USA.
Giamarellos-bourboulis EJ, Netea MG, Rovina N, Koulouris N, Gogos
patients with severe respiratory failure. Cell Host Microbe. (2020) 27:1–
SMAR1 favors immunosurveillance of cancer cells by modulating calnexin
Zhao J, Zhao J, Perlman S. T cell responses are required for
protection from clinical disease and for virus clearance in severe

Frontiers in Immunology | www.frontiersin.org

14

August 2020 | Volume 11 | Article 1949




Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Shah, Firmal, Alam, Ganguly and Chattopadhyay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
COVID-19: Immunology, Immunopathogenesis and Potential Therapies


To link to this article: https://doi.org/10.1080/08830185.2021.1883600

Published online: 27 Feb 2021.

Submit your article to this journal

Article views: 2386

View related articles

View Crossmark data

Citing articles: 4 View citing articles
COVID-19: Immunology, Immunopathogenesis and Potential Therapies

Asha Bhardwaj a#, Leena Sapra a#, Chaman Saini a, Zaffar Azam a, Pradyumna K. Mishra b, Bhupendra Verma a, Gyan C. Mishra c and Rupesh K. Srivastava a

aDepartment of Biotechnology, All India Institute of Medical Sciences, New Delhi, India; bDepartment of Molecular Biology, ICMR-NIIRH, Nehru Hospital Building, Gandhi Medical College Campus, Bhopal, India; cLab # 1, National Centre for Cell Science (NCCS), Savitribai Phule Pune University Campus, Pune, India

ABSTRACT

The Coronavirus Disease-2019 (COVID-19) imposed public health emergency and affected millions of people around the globe. As of January 2021, 100 million confirmed cases of COVID-19 along with more than 2 million deaths were reported worldwide. SARS-CoV-2 infection causes excessive production of pro-inflammatory cytokines thereby leading to the development of “Cytokine Storm Syndrome.” This condition results in uncontrolled inflammation that further imposes multiple-organ-failure eventually leading to death. SARS-CoV-2 induces unrestrained innate immune response and impairs adaptive immune responses thereby causing tissue damage. Thus, understanding the foremost features and evolution of innate and adaptive immunity to SARS-CoV-2 is crucial in anticipating COVID-19 outcomes and in developing effective strategies to control the viral spread. In the present review, we exhaustively discuss the sequential key immunological events that occur during SARS-CoV-2 infection and are involved in the immunopathogenesis of COVID-19. In addition to this, we also highlight various therapeutic options already in use such as immunosuppressive drugs, plasma therapy and intravenous immunoglobulins along with various novel potent therapeutic options that should be considered in managing COVID-19 infection such as traditional medicines and probiotics.

1. Introduction

In December 2019, an outbreak of pneumonia occurred in Wuhan city of China that rapidly spread around the globe and posed serious public health emergency [1]. On 9 January 2020, it was officially announced that novel coronavirus 2019-nCoV is the reason behind the outbreak in Wuhan, China [2]. Later International Committee on Taxonomy named the novel coronavirus as Severe Respiratory Disease Syndrome-Coronavirus-2 (SARS-CoV-2) and World Health Organization (WHO) named the disease as Coronavirus Disease-2019 or COVID-19 [3]. Coronaviruses (CoVs) cause infection in humans and animals and are found to be responsible for various respiratory, renal, gastrointestinal and neurological disorders. CoVs are classified into four genera viz. alpha, beta, gamma and delta CoVs. Alpha and beta CoVs infect humans and usually cause upper respiratory tract infections but in some patients also cause lower respiratory tract infections [4]. SARS-CoV-2 belongs to the family Coronaviridae, order Nidovirales, genus Betacoronavirus and subgenus Sarbecovirus [5]. SARS-CoV-2 is the 7th CoV in the list that is reported to cause infections in humans. The other six CoVs are SARS-CoV, MERS-CoV, HKU1, NL63, OC43 and 229E. SARS-CoV and MERS-CoV cause various fatal and respiratory diseases like SERS-CoV-2 whereas HKU1, NL63, OC43 and 229E cause only minor symptoms [6]. SARS-CoV was found to be responsible for an epidemic in 2002–2003 which started from China and Asia Pacific regions and affected around 8000 people across 37 countries with fatality rate of 10% [7, 8]. Common symptoms observed in SARS-CoV infected patients were fever, dyspnea, dry cough and hypoxemia [9]. Middle East Respiratory Syndrome-Coronavirus (MERS-CoV) is a 2C beta CoV and was first reported in 2012 from Saudi Arabia [10]. MERS-CoV caused severe pneumonia and renal failure in infected patients [11]. The SARS-CoV-2 virus shares 79.6% sequence similarity with the SARS-CoV virus but SARS-CoV-2 is found to be more pathogenic [12]. Due to its pathogenicity and easy

CONTACT Dr. Rupesh K. Srivastava, Assistant Professor rupesh_srivastava13@yahoo.co.in or rupeshk.aiims@gmail.com Department of Biotechnology, All India Institute of Medical Sciences (AIIMS), New Delhi 110029, India. #These authors have contributed equally to the work.

© 2021 Taylor & Francis Group, LLC
transmission from human to human. WHO declared COVID-19 as a pandemic disease on 11 March 2020. As of January 2021, there are 100 million confirmed cases of COVID-19 worldwide with over 2 million reported deaths. SARS-CoV-2 cause mild respiratory disorders to acute pneumonia and multiple organ failure and in severe cases can eventually lead to death [13]. Whole genome sequencing revealed that the SARS-CoV-2 is more closely related to bat CoV RaTG13 which was isolated from *Rhinolophus affinis* with 96.2% sequence similarity [14].

Immune system plays a pivotal role in the pathogenesis of COVID-19. SARS-CoV-2 induces unrestrained innate immune response and impairs adaptive immune responses leading to widespread tissue damage. Till now, there is no effective treatment available for COVID-19. Knowledge of immunopathogenesis of COVID-19 will help in designing suitable immune therapy for the treatment of SARS-CoV-2 infection. In this review, we have discussed the pathogenesis and immunopathogenesis of COVID-19 along with the potential immunotherapeutic interventions that can be targeted toward the dysregulated immune system. We also discuss the plausible relevance of gut microbiota and probiotics in COVID-19.

2. Structure of SARS-CoV-2

The novel CoV is an enveloped, positive sense, single stranded RNA virus with a genome size of 26,000 to 32,000 nucleotides encoding 14 open reading frames (ORFs). The first two large ORFs (orf1ab and orf1a) which are present at the 5' end cover almost two third of the genome (20 kb; Figure 1). They constitute the replicase gene which contains 16 nonstructural proteins (nsps). Replicase gene is required for replication and transcription. Replicase gene codes for two polyproteins: pp1a (contains the 1–11 nsps) and pp1ab gene (contains the 12–16 nsps). The 3' end of the genome which is around 10 kb encodes 4 structural and 8 accessory proteins. The structural proteins consist of spike (S) protein, membrane (M) protein, envelope (E) protein and the nucleocapsid (N) protein [15]. S protein allows the entry of virus into the host cell. M and E protein regulate virus assembly and N protein facilitates RNA synthesis. S protein is

---

**Figure 1.** (A, B) Structure and genome organization of SARS-CoV-2 (Figure illustrated with the help of [https://smart.servier.com/](https://smart.servier.com/)).
Hypoalbuminemia, liver dysfunction and renal abnormalities are also observed in severe cases of COVID-19 [27–29]. About 8–12% of COVID-19 patients also have acute cardiac injury due to systemic inflammation [30].

SARS-CoV-2 virus shares 79% similarity with the SARS-CoV virus. Like SARS-CoV it also binds to the human Angiotensin Converting Enzyme 2 (ACE2) that is expressed by various cells including lungs [31]. Apart from the lungs, surface expression of ACE2 is also observed on the epithelial cells of the small intestine, mucosa of the oral cavity, venous, arterial endothelial cells and arterial smooth muscles of all organs [32, 33]. This forms one of the major reason behind the multiorgan failure observed in case of SARS-CoV-2 [34]. ACE2 is homologue of ACE which produces angiotensin II from angiotensin I. Angiotensin II is the major protein of Renin Angiotensin System (RAS) and is required for vasoconstriction and for other biological functions. It has been reported that ACE2 negatively regulates RAS by suppressing the angiotensin II level, and thus, prevents from ARDS [35]. ACE2 receptor is expressed on the ciliated airway epithelial cells and on the type 2 alveolar cells of the lungs [36]. SARS-CoV-2 interacts with the ACE2 receptor with the help of S protein. S protein is cleaved into two subunits (S1 and S2) by the host protease TMPRSS2 at the boundary of the S1–S2 [37] (Figure 2). S1 subunit exhibits the receptor binding domain whereas S2 subunit facilitates the membrane fusion [38].

Interestingly, it has been observed that the RBD domains of the SARS-CoV and SARS-CoV-2 shares 72% similarity in the amino acids sequence [39]. Residues required for attachment to ACE2 in RBD are also highly conserved in both the viruses [1]. Molecular modeling and biophysical techniques revealed that RBD of the SARS-CoV-2 binds more strongly and with higher affinity to the ACE2 receptor than the SARS-CoV [40]. ACE2 binding ridge in SARS-CoV-2 is more compact [12]. Cryo-electron microscopy structure of novel CoV S protein trimer at the resolution of 3.5 Å revealed that RBD domain of the S trimer rotate in a conformation that is more accessible to the ACE2 [40]. SARS-CoV-2 also consists of different loop having flexible glycy1 residues in place of rigid propyl residues that are present in SARS-CoV [39]. It has a very different furin like cleavage site in the S protein that is not present in other SARS like CoVs [41]. Thus, we can conclude that these features may be one of the reasons responsible for increased pathogenicity of SARS-CoV-2.

3. Pathogenesis of COVID-19

SARS-CoV-2 is transmitted from infected patients to healthy individuals through direct contact or via spread of respiratory droplets from the infected patients [19]. The median incubation period of the SARS-CoV-2 infection is approximately 5.1 days with 97.5% of the population developing symptoms within 11.5 days [20]. The common symptoms associated with the infection are dry cough, fever, pain, weakness, chest tightness, loss of smell and taste, dyspnea, accompanied by acute respiratory distress syndrome (ARDS) [21, 22]. ARDS is a severe disease which leads to respiratory failure and is characterized by hypoxemia, difficulty in breathing and onset of pulmonary edema. It leads to damage in lung endothelium and alveolar epithelium [23]. RNAemia (detectable viral load of SARS-CoV-2 in serum) and acute cardiac injury has also been observed in some of the COVID-19 patients [21]. Immunocompromised individuals like those who have diabetes, cardiovascular diseases and hypertension are estimated to be at higher risk of COVID-19 infection [24]. Mortality rate is reported to be higher in elderly population. Recently, a study reported that 73.9% of the infected individuals less than the age of 60 did not develop symptoms. Thus, indicating that elderly people are highly susceptible to developing symptoms [25]. Gender difference also affects the severity and mortality of the disease with men being more prone to infection and along with enhanced mortality with respect to women [26]. Hypoalbuminemia, liver dysfunction and renal
**Figure 2.** The schematic diagram of SARS-CoV-2 invasion and replication into host cells in step wise manner: (1) SARS-CoV-2 enter into the cell by binding to the ACE2 receptor present on the host cell membrane via S protein which mediates viral-host membrane fusion and viral entry. (2) Entry of SARS-CoV-2 results in the uncoating of viral RNA into cytoplasm which then undergoes translation to produce polyproteins pp1a and pp1b, which are further processed by virus-encoded proteinases into individual nonstructural proteins (nsps). (3) Replication transcription complex is formed by some of the nsps. Replication transcription complex uses the (+) strand genomic RNA as template. Following replication (+) strand genomic RNA is produced which becomes the genome of the new viral particle. (4) Subgenomic RNAs synthesized through transcription are translated into viral structural proteins: S (spike) protein, M (membrane) protein, E (envelope) protein, N (nucleocapsid) protein. (5) N protein combined with the (+) strand genomic RNA to form the nucleoprotein complex. S, E and M proteins enter into endoplasmic reticulum (ER) and transported to Golgi apparatus. The nucleoprotein complex and S, E and M proteins are further assembled in the ER-Golgi intermediate compartment (ERGIC) to form mature virion. (6) Virions are released from the host cells by exocytosis (Figure illustrated with the help of https://smart.servier.com/).
stimulate the secretion of cytokines or chemokines like interleukin-6 (IL-6), interferon-γ (IFN-γ), interferon-γ-inducible protein-10 (IP-10) and monocyte chemoattractant protein-1 (MCP)-1. These cytokines and chemokines promote the influx of monocytes/macrophages and neutrophils from the blood to the site of infection. These cells secrete the cytotoxic substances to clear the viral infection. Normally, this response is capable of eliminating the virus but sometimes immune system is dysregulated which leads to the disruption of the immune homeostasis [42]. In case of SARS-CoV-2 infection intense inflammatory response against the virus leads to excessive production of pro-inflammatory cytokines which thereby promote lung pathogenesis and respiratory failure (Figure 3). Pro-inflammatory cytokines induce uncontrolled accumulation of monocytes/macrophages and neutrophils at the site of infection. These cells then produce cytotoxic substances like reactive oxygen species (ROS) which leads to cell death and tissue damage [42]. In further sections, we have discussed the immunopathogenesis of SARS-CoV-2 infection in more details (Figure 4 and Table 1).

4.1. Immune evasion by SARS-CoV-2

Innate immune cells are activated when they recognize pathogen associated molecular patterns (PAMPs) and damage associated molecular patterns (DAMPs) with the help of pattern recognition receptors present on them (PRRs). PRRs include the toll like receptors (TLRs), NOD like receptor (NLR), RIG-1 like receptor (RLR), C-type lectin-like receptors (CLR) and melanoma differentiation associated protein 5 (MDA5). Recognition of viral genes activate the interferon regulatory (IRF)-3, IRF-7 and nuclear factor kappa-light-chain enhancer of activated B cells (NF-κB) thereby promoting their nuclear translocation. Generally, activation of TLR3/7 promotes nuclear translocation of IRF-3 and NF-κB and activation of RIG/MDA5 stimulate the nuclear translocation of IRF-3 [56]. Type 1 IFNs provides the first line of defence by preventing the spread of viral infection. It induces the expression of IFN stimulated genes (ISGs) by activating the JAK-STAT signaling pathway [4]. Successful activation of type 1 IFN response inhibits viral replication and prevents spread of viral infection. Innate immune response depends heavily on the type 1 IFN responses for eliminating the virus.

SARS-CoV-2 virus evades the innate immune response by suppressing the antiviral type 1 IFN responses. It is observed that SARS-CoV-2 infection stimulate low levels of antiviral cytokines IFN-α and IFN-β. Several studies on the SARS-CoV virus revealed that it antagonizes the type 1 IFN responses through various mechanisms (Figure 5). Versteeg et al. demonstrated that SARS-CoV prevented IFN production by shielding the viral RNA from the host cellular sensory molecules [57]. Furthermore, it was observed that SARS-CoV nsp1 inhibited host immune response by modulating the production of type 1 IFNs. nsp 1 mutant showed higher expressions of type 1 IFN [58]. Other viral proteins viz. nsp1, nsp7, nsp15 and ORFs like ORF3b, ORF6 and ORF9b also suppressed the activation of type 1 IFN [59]. Another study reported that SARS-CoV prevented IFN production by disrupting the stimulator of IFN genes (STING) signaling which is required for activation of IRF3 that induces the production of IFN [60]. IFN production is also inhibited via first transmembrane domain (TM1) of the M protein which is present in the N-terminal. TM1 can bind to the RIG-I, TNF receptor associated factor 3 (TRAF3), TANK-binding kinase1 (TBK1) and homolog IκB kinase epsilon (IKKe). By binding to them M protein inhibits the binding of these molecules with other downstream effectors, and thus, preventing induction of type 1 IFN [61]. Similarly, from in vitro studies it has been reported that SARS-CoV inhibits the expression of IFN-β in macrophages and stimulates the production of chemokines like IP-10 and MCP-1 [62]. Channappanavar and Perlman showed that SARS-CoV infection delayed type 1 INFα/β responses inducing the accumulation of inflammatory monocytes and macrophages along with inhibiting T cells mediated immune response against the virus [63]. Delayed type 1 IFN response and monocytes/macrophages recruitment are found to be the major reasons for the lethal pneumonia and ARDS observed in SARS patients [63]. Melo et al. showed that SARS-CoV-2 like SARS-CoV stimulate lower levels of type 1 and type 3 IFN responses. This leads to restrained ISG expression and induction of high expression of cytokines like IL-6 and IL-1Ra and chemokines like C-X-C motif ligand 2 (CXCL2) and CXCL8. CXCL2 and CXCL8 promote immune cells infiltration [64]. Hadjadj et al. demonstrated that immune cells in peripheral blood of critically ill COVID-19 patients showed reduced type 1 IFN response. These patients also had higher levels of IL-6 and TNF-α [65]. Furthermore, Lucas et al. showed that patients with mild symptoms had decreased levels of type 1 and type 3 IFN responses [66]. Thus, it can be concluded that SARS-CoV-2 cause dysregulation of innate immunity by modulating the type 1 IFN responses.

4.2. Cytokine storm

One of the major reasons for the lung pathogenesis in COVID-19 patients is the over-production of proinflammatory cytokines. It was observed that early enhancement in levels of cytokines lead to worst outcomes [66]. Dysregulated immune system in COVID-19 patients results in development of cytokine storm which promotes lung inflammation. Severe COVID-19 patients have
Figure 3. Schematic representation of immunopathogenesis during COVID-19 in a step wise manner: (1) SARS-CoV-2 gain entry into the lung cells through ACE2 receptor. ACE2 receptor is expressed by different organs of the body like brain, lungs, kidney, liver and intestine. SARS-CoV-2 frequently infects these organs but the pathogenesis of SARS-CoV-2 infection initiates from lungs and cause major damage to the lungs. Lung infection starts when SARS-CoV-2 enter into type 2 alveolar cells through ACE2 receptor. (2) Virus replicates into the lung cells and induce cells to undergo pyroptosis and secrete damage associated molecular patterns (DAMPs). (3) DAMPs are recognized by adjoining epithelial cells, alveolar macrophages and endothelial cells which stimulate the secretion of various pro-inflammatory cytokines and chemokines like IL-6, IL-10, MIP1-α, MIP1-β, MCP-1. (4) Release of these cytokines and chemokines induce the recruitment of monocytes, macrophages and neutrophils into lungs which further secrete the proinflammatory cytokines and form the inflammatory feedback loop. (5) During SARS-CoV-2 infection immune response is dysregulated which leads to the persistent recruitment of immune cells which promote overproduction of inflammatory cytokines which cause various lung disorders like ARDS and acute lung injury. SARS-CoV-2 infection induce delayed activation of dendritic cells that cause defective T cell response. Most of the T lymphocytes in SARS-CoV-2 infection are induced to become Th1 cells. SARS-CoV-2 infection also induce delayed antibody response (Figure illustrated with the help of https://smart.servier.com/).
Figure 4. The schematic diagram depicting immunopathogenesis of COVID-19 (Figure illustrated with the help of https://smart.servier.com/.).
increased levels of cytokines such as IFN-γ, IL-1, IL-6, IL-8, IL-12 and transforming growth factor β (TGF-β) and chemokines C-C motif ligand 2 (CCL2), CXCL9 and CXCL10 [43]. COVID-19 patients also showed elevated levels of plasma IL-2, IL-7, IL-10, granulocyte-macrophage colony stimulating factor (GM-CSF), macrophage inflammatory protein 1-α (MIP1-α), tumor necrosis factor-α (TNF-α) and MCP-1 [21]. It was observed that patients who were in non-survivor group showed higher levels of IL-6 throughout the clinical course than patients in the survivor group [67]. In severe cases, cytokine storm violently attacks the body and causes ARDS, multiple organ failure eventually leading to death [4].

Cytokine storm induce unchecked influx of immune cells especially monocytes and neutrophils into the lungs. These cells secrete various inflammatory cytokines and chemokines such as IL-1β, IL-6, TNF-α, CCL2, CCL7 and CCL12 that further enhance disease severity. Earlier studies reported that IL-1β promoted pyroptosis defined as non-programmed cell death upon pathogenic infection [68]. Uncontrolled cell infiltrations promote lung damage due to excessive secretion of toxic substances like proteases and ROS. Furthermore, it causes alveolar damage, pulmonary edema, hyaline membrane formation and desquamation of pneumocytes which are defined as the earlier signs of ARDS [69, 70].

Table 1. Response of various immune cells in COVID-19.

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Immune cells</th>
<th>Response to SARS-CoV-2 infection</th>
<th>Treatment options</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Monocytes and Macrophages</td>
<td>- Increased in number&lt;br&gt;- Macrophages produce proinflammatory response (secrete IL-6 and IL-1β)&lt;br&gt;- Infiltration into lungs</td>
<td>- Anakinra (IL-1β antagonist) and Tocilizumab (IL-6 receptor antagonist)</td>
<td>[43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Delayed or supressed type 1 IFN response</td>
<td>- Immunosuppressing drugs like Tocilizumab, Lianhuajingwen, Hydroxychloroquine</td>
<td>[44] [45]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Produce proinflammatory cytokines and chemokines (TNF-α, IL-6, RANTES, IP-10, MCP-1, MIP-1α, CCR1, CCR3 and CCR5)</td>
<td>- Depletion of alveolar macrophages in SARS infection induce the activation and migration of dendritic cells</td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Delayed type 1 IFN response</td>
<td></td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Delayed induction and migration to lymph nodes</td>
<td></td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Increased expression of TRAIL which leads to apoptosis of lymphocytes on their interaction with Dendritic cells</td>
<td></td>
<td>[44]</td>
</tr>
<tr>
<td>2.</td>
<td>Dendritic cells (DCs)</td>
<td>- Decreased counts&lt;br&gt;- Exhaustion of NK cells (High expression of NKG2A and low expression of TNF-α, IL-2, CD107α, granzyme B and IFN-γ)</td>
<td>- Antiviral therapy with chloroquine recovers the levels NK cells&lt;br&gt;- Blocking of IL-6 can prevent the decrease in cytotoxicity of NK cells</td>
<td>[46] [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Decreased cytotoxicity (reduced perforin and granzyme secretion)</td>
<td></td>
<td>[46] [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- High activation in peripheral blood</td>
<td></td>
<td>[46] [47]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Peripheral NK cells showed higher expression of markers like perforin, NKG2C and Ksp37</td>
<td></td>
<td>[46] [47]</td>
</tr>
<tr>
<td>3.</td>
<td>Natural Killer (NK) cells</td>
<td>- Increased number&lt;br&gt;- Increased neutrophil to lymphocyte ratio&lt;br&gt;- Increased NET release which leads to ARDS, thrombosis, coagulation and worst oxygenation conditions observed in COVID-19 patients</td>
<td>- Dornase Alfa is suggested to prevent NET release&lt;br&gt;- Inhibition of IL-1β secretion</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Induce proinflammatory response (secrete IL-1, IL-6 and IL-33, histamine and protease)</td>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secretion of peroxynitrite in airways</td>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secretory enzymes</td>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td>4.</td>
<td>Neutrophils</td>
<td>- Increased in number&lt;br&gt;- Increased neutrophil to lymphocyte ratio&lt;br&gt;- Increased NET release which leads to ARDS, thrombosis, coagulation and worst oxygenation conditions observed in COVID-19 patients</td>
<td></td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Upregulate RAS activity in airways</td>
<td></td>
<td>[50] [51]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secrete peroxynitrite in airways</td>
<td></td>
<td>[50] [51]</td>
</tr>
<tr>
<td>5.</td>
<td>Eosinophils</td>
<td>- Decreased in number</td>
<td></td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td>Mast cells</td>
<td>- Induce proinflammatory response (secrete IL-1, IL-6 and IL-33, histamine and protease)</td>
<td>- Lopinavir treatment&lt;br&gt;- Anti-inflammatory cytokines of IL-1&lt;br&gt;- Farnotidine&lt;br&gt;- Sodium chromo glycate and palmitoylethanolamide</td>
<td>[51] [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secretion of peroxynitrite in airways</td>
<td></td>
<td>[51] [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secretory enzymes</td>
<td></td>
<td>[51] [52]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Secretory enzymes</td>
<td></td>
<td>[51] [52]</td>
</tr>
<tr>
<td>6.</td>
<td>T cells</td>
<td>- Depletion of CD4+ and CD8+ T cells&lt;br&gt;- T helper cells are polarized into Th1 cells and Th17 cells&lt;br&gt;- Exhaustion of CD8+ T cells&lt;br&gt;- Increased in number of GM-CSF, IL-6 and TNF-α expressing CD4+ T cells&lt;br&gt;- CD8+ T cells express GM-CSF and produce more granzyme, perforin and IFN-γ&lt;br&gt;- Infiltration into lungs&lt;br&gt;- Decreased number of Tregs&lt;br&gt;- Decreased number of γδ T cells&lt;br&gt;- Phenotypic alterations of unconventional T cells&lt;br&gt;- Decreased number of TFH cells&lt;br&gt;- Exhaustion of T cells with expression of markers like Tim-3 and PD-1</td>
<td>- Blocking of IL-6 with Tocilizumab can enhance the counts of circulating lymphocytes.&lt;br&gt;- Drugs targeting GM-CSF such as Lenzilumab, Gimsilumab and Namilumab or other anti-inflammatory drugs&lt;br&gt;- Antiviral therapy with Chloroquine can recovers the level of CD8+ T cells</td>
<td>[47] [53] [54]</td>
</tr>
<tr>
<td>7.</td>
<td>B cells</td>
<td>- Delayed antibody response&lt;br&gt;- Antibody dependent enhancement</td>
<td>- Plasma therapy</td>
<td>[55]</td>
</tr>
<tr>
<td>8.</td>
<td></td>
<td></td>
<td></td>
<td>[55]</td>
</tr>
</tbody>
</table>
4.3. Complement activation

Complement proteins are the component of innate immunity and it has been reported that complement activation is also related with pathogenesis of ARDS. In case of SARS-CoV infection, activation of complement component C3 was found to be responsible for pathogenesis. Furthermore, studies reported that depletion of C3 in SARS-CoV infected mice reduced lung injury. Recently, it has been revealed that C3 deficient mice also have reduced levels of IL-6 and other chemokines and cytokines with decreased influx of neutrophils and monocytes [71]. Various studies have shown the role of complement activation in COVID-19. Significantly elevated levels of complement component C5b-9 and C5a were observed in the plasma of severe COVID-19 patients [72]. In study of Gao et al. it has been observed that N protein of SARS-CoV, MERS-CoV and SARS-CoV-2 binds to the mannan-binding lectin-associated serine protease 2 (MASP)-2 which is an enzymatic initiator of lectin pathway. N protein; MASP-2 interaction promotes abnormal complement activation and lung injury whereas blockade of this interaction suppressed the complement activation and prevented inflammation and lung injury [73]. Complement activation may also promote coagulopathy, thrombosis and microvascular injury that are widely reported in COVID-19 patients [74–76]. Moreover, in COVID-19 patients enhanced deposition of complement components C5b-9, C4d and MASP-2 are observed in the lung and are associated with microvascular injury and thrombosis [76]. Diao et al. observed that acute renal failure in COVID-19 patients is associated with increased depositions of C59-b in the kidney tubules [77].
Inhibition of C3 with anti-C3 agents like AMY-101 and with pathway inhibitors like lectin pathway can be beneficial in preventing induced lung injury in COVID-19 [71].

4.4. Dysregulation of immune cells in COVID-19

4.4.1. Monocytes and macrophages

Single stranded RNA viruses like CoVs activate the monocytes/macrophages through the TLR7/8 and stimulate type 1 IFN responses and cytokine production which leads to suppression of viral infection [78]. However, SARS-CoV-2 dysregulated the activity of monocytes and macrophages that further promotes lung injury and respiratory disorders like ARDS. Single cell RNA sequencing (scRNA) of the bronchoalveolar lavage fluid (BALF) revealed that there was abundance of monocyte derived macrophages in the lungs of severe COVID-19 patients [79]. Similarly, single cell transcriptome of peripheral blood revealed that there was increased number of circulatory monocytes in COVID-19 patients [73]. Increased numbers of monocytes were also observed during the early recovery stage (ERS). scRNA of peripheral blood during ERS revealed that there were increased counts of classical CD14+ monocytes and CD14+IL-1β+ monocytes in circulation of COVID-19 patients. These cells further enter into the pulmonary circulation and cause lung infection [14, 80]. Remarkably, increased numbers of monocyte derived FCN1+ macrophages have also been observed in BALF than the alveolar macrophages in severe patients. Alveolar macrophages are found to be the major type of macrophages in BALF [79].

In COVID-19 patients increased expression of inflammatory cytokines such as TNF-α, IL-1β, IL-2, IL-6, IL-7, GM-CSF, IP-10, MIP-1α and MCP-1 had been observed [43, 81]. Macrophages are known to play an important role in induction of these inflammatory cytokines. It has been observed that critically ill COVID-19 infected patients showed increased levels of CCL2, CCL3, CCL20, CXCL1, CXCL3, CXCL10, IL-8, IL-1β and TNF-α expressing inflammatory macrophages [82]. Sustained production of IL-1β and IL-6 from the circulating monocytes and macrophages resulted in macrophage activation syndrome (MAS) in COVID-19 patients that leads to severe respiratory failure [83]. MAS is characterized by the increased production of IL-6 along with higher levels of hemophagocytic macrophages [84]. In macrophages, NOD like receptor family, pyrin domain containing 3 (NLRP3) inflammasome regulates the release of proinflammatory cytokines IL-1β and IL-18 [85]. COVID-19 patients also have activated MAPK-interferon pathway as revealed by the single cell profiling and has very imperative role in immune response produced in case of COVID-19 [86].

4.4.2. Dendritic cells

DCs are the most potent professional antigen presenting cells that also play a vital role in preventing viral infections. Many viruses use the respiratory tract to get entry into the host and cause infection in the lungs which leads to the activation of innate and adaptive immune response. DCs are activated at the initial stages of the viral infection and provide antiviral immunity [91]. Activated DCs migrate from various tissues to the lymphoid organs and present antigens to the T cells that further activate specific adaptive immune response. Various studies showed that SARS virus evade the innate immune response by impairing the activity of DCs through different mechanisms. A study reported that infection of DCs with SARS-CoV resulted in
suppression of MHC-I expression on DCs that further delayed the expression of IFN-α [92]. Another study reported that SARS-CoV induced the plasmacytoid DCs (pDCs) to produce prolonged production of type I IFNs. This leads to immune exhaustion and accumulation of monocytes, macrophages and neutrophils at the site of infection [93]. It actually suppresses the expression of antiviral genes like IFN-α, IFN-β and IFN-γ. It also induces the increased expression of proinflammatory cytokines and chemokines such as TNF-α, IL-6, CCL5, IP-10, MCP-1, MIP-1α, CCR1, CCR3 and CCR5. The SARS infection does not affect the expression of TLRs (TLR-1 to TLR-10). It stimulates the enhanced expression of TNF related apoptosis inducing ligand (TRAIL) with no effect on the expression of FAS ligand [94, 95]. The increased production of chemokines induces the migration of DCs to the lymph nodes but increased expression of TRAIL promotes the apoptosis of lymphocytes on their interaction with DCs. Thus, DCs induced apoptosis of T cells results in suppression of virus specific adaptive immune response. Zhao et al. showed that SARS infection also delayed the activation of DCs and their migration to the draining lymph nodes which leads to deficient adaptive T cell response [44]. Furthermore, a study reported by Tseng et al. showed that SARS infection leads to maturation of DCs and induced the expression of MHC-II and costimulatory molecules [96]. Impaired DC activity is also observed in MERS. MERS-CoV infects the monocyte derived DCs and induce markedly elevated expression of CCL5, IL-12, IP-10 and IFN-γ along with MHC-II molecules and co-stimulatory molecule CD86 [97].

SARS-CoV-2 like SARS-CoV and MERS-CoV affects the activity of DCs. Depletion of both pDCs and conventional DCs was observed in the severe COVID-19 patients [98]. It was observed that mammalian target of rapamycin (mTOR)-signaling and IFN-α production was diminished in the pDCs of SARS-CoV-2 infected patients [99]. In COVID-19 patients delayed activation and migration of DCs from the site of infection, i.e., from lungs to the draining lymph nodes was also observed. It can, thus, be concluded that delayed activation and migration of DCs might be responsible for the slow viral clearance observed in the COVID-19 patients [100]. Severe COVID-19 patients have lymphopenia as they have reduced counts of CD4+ and CD8+ T cells and have decreased expression of TCR. Reduced number of T cells and expression of TCR might be due to interaction of these cells with the DCs which promote their apoptosis as in case of SARS-CoV. Depletion of pDCs in COVID-19 patients is responsible for the impaired activity of IFN-α observed in these patients as these are the major producers of IFN-α [65]. From metatranscriptomic sequencing of BALF cells it was revealed that COVID-19 patients have more abundant DCs than the healthy people [101]. Chen et al. showed that cytokine storm caused by the SARS infection induced the influx of pDCs into the lungs [102]. Thus, cytokine storm induced in COVID-19 patients might promote influx of DCs into the lungs. However, ScRNA of BALF revealed that severe COVID-19 patients have reduced number of pDCs and myeloid derived DCs (mDCs) than the patients with mild symptoms.

CoVs interacts with DCs through several ways. Using mice models, it has been demonstrated that CoVs interacts with the DC in receptor dependent manner by using the murine DC marker CEACAM1a. A study reported that blockade of this receptor via anti-CEACAM1a monoclonal antibody (mAb) CC1 prevented infection of DCs [103]. DCs express DC-SIGN (CD209) which is C-type lectin present on their membrane. It is used by various viruses like human immunodeficiency virus (HIV) to interact with the DCs [103]. Jeffers et al. showed that CD209 expression induce the susceptibility of SARS-CoV infection [104]. It has been reported that single nucleotide polymorphism (SNP) in the promoter region of the CD209 enhanced the disease severity in SARS infection [100]. Thus, it might be possible that CoVs infect DCs via CD209. Through ScRNA it is revealed that DCs also express the ACE2 receptor [100], and thus, SARS-CoV-2 can directly infect the DCs through ACE2 receptor.

### 4.4.3. Natural killer cells

Natural killer (NK) cells are cytotoxic cells that play crucial role in providing immunity against the virus [105]. NK cells recognize the pathogens with the help of its receptors known as natural cytotoxicity receptors (NCRs) like NKP46 and NKP44. Activation of these markers leads to the killing of virus infected cells [106]. DCs can also activate the NK cells. During influenza virus infection it was observed that influenza infected DCs activate the NK cells by recognizing its receptors NKG2D and NKP46 [107]. The role of NK cells has been deciphered in various viral infections such as influenza, respiratory syncytial viruses and HIV.

Deficiency of NK cells lead to various viral infections and by inducing depletion of NK cells, viruses protected themselves by enhancing own survival [108]. Interestingly, it has also been observed that SARS-CoV-2 also evades the immune system by depleting the NK cell population. Yu et al. showed that the level of NK cells starts falling from day 16 of the disease onset [109]. ScRNA sequencing of BALF also revealed that COVID-19 patients have decreased levels of NK cells than their healthy counterparts [98]. Analysis of lung NK cells revealed that most of the NK cells that reside in human lungs are characterized by CD56dimCD16+ phenotype. Few NK cells also express tissue residency marker CD69 which consist of mostly CD16+ NK
cells (immature cells) and CD16+ cells (less differentiated cells). Maucourant et al. reported that highly activated NK cells were present in the peripheral blood of COVID-19 patients. They also observed that in severe cases peripheral NK cells showed higher expression of markers like perforin, NKG2C and Ksp37 [110]. NK cells in COVID-19 infected patients are found to be functionally exhausted as they highly express NKG2A. Higher expression of NKG2A is associated with the exhaustion of NK cells. NK cells of COVID-19 patients expressed lower levels of TNF-α, IL-2, CD107a, granzyme B and IFN-γ which further suggests that they are functionally exhausted. Functional exhaustion of NK cells is one of the mechanisms through which SARS-CoV-2 evades immunity [47]. Another study reported that NK cells of COVID-19 patients express higher levels of MHC-I which downregulated the cytotoxicity of NK cells by inducing their apoptosis [111]. SARS-CoV-2 is also reported to inhibit cytotoxicity of NK cells in an IL-6 dependent manner. COVID-19 patients possess higher levels of IL-6 cytokine which downregulated the cytotoxicity of NK cells by reducing the perforin and granzyme secretion [46]. Recently, it was also observed that Tocilizumab enhanced the activity of NK cells [73]. Single cell sequencing revealed that NK cells do not express the ACE2 receptor [112]. Thus, SARS-CoV-2 does not infect the NK cells directly but reduce the capability of NK cells by decreasing their cytotoxicity and inducing their exhaustion.

4.4.4. Neutrophils

Neutrophils are the most abundant polymorphonuclear leucocytes that provide first line of innate immune defence and kill microbes by phagocytosis and degranulation [113]. The role of neutrophils in providing antiviral defence has not been deciphered yet but it is observed that neutrophils can prevent viral infection by releasing various chemokines, cytokines and antimicrobial compounds [114]. One of the mechanisms by which neutrophils eradicate pathogens is through neutrophil extracellular traps (NETs) formation. NETs trap the bacteria, viruses, parasites, fungi and neutralize them to provide antimicrobial defence [115]. NET formation stimulates thrombosis by activating platelets and by inducing coagulation which promote various cardiovascular diseases like stroke and myocardial infarction [116]. NET formation is also found to be linked with various lung diseases like cystic fibrosis, obstructive pulmonary disease, acute lung injury and ARDS [116, 117]. In case of COVID-19, in vitro study revealed that sera of virus infected patients induced NET release [117]. This shows that NET formation in one of the mechanisms through which neutrophils increase the severity of COVID-19. Several studies prove the relevance of NET in COVID-19 pathogenesis [117].

It has been found that neutrophils have major role in the pathogenesis of COVID-19 that induces the proinflammatory response observed in COVID-19. In analysis of 82 patients it was observed that 74.3% of the patients showed neutrophilia whereas 94.5% of the patients showed higher neutrophil to lymphocyte ratio (NLR) [118]. ScRNA analysis of the BALF revealed that severe COVID-19 patients contain higher proportions of neutrophils than patients who had mild symptoms [79]. Interestingly, one of the study demonstrated that pregnant women infected with COVID-19 showed significant enhancement in neutrophils than in the non-pregnant counterparts [119]. Recently using machine learning algorithm Mathew et al. observed that neutrophil activation is the strongest predictor of COVID-19 severity. Activated neutrophils express resistin, lipocalin-2, IL-8, HGF (Hepatocyte Growth Factor) and G-CSF. Neutrophil activation was observed on the first day of the hospitalization only in the patients who later require ICU [120]. Altogether, from these studies it can be concluded that enhanced neutrophils population and activation and higher NLR can be utilized as biomarker to access disease severity.

In consistent to SARS-CoV infection neutrophil infiltrations have also been observed in COVID-19 patients. Cytokine storm promote the accumulation of neutrophils in lung [121]. Higher expression of neutrophil chemokine CXCR2 receptor along with delayed type 1 IFN response has been found to be responsible for accumulation of neutrophils in lungs of COVID-19 patients [63, 65]. It has been found that increased levels of IL-1β by monocytes in the circulation induce the accumulation of neutrophils and CD8+ T cells in the lung [21, 68]. Infiltrating neutrophils promote disease severity by inducing the production of chemokines and cytokines such as IL-1β, IL-6, IL-8, TNF-α, TGF-β and MCP-1 [122]. Neutrophils also produce cytotoxic substances like ROS, proteases and leukotrienes that further lead to various respiratory disorders like acute lung injury and ARDS [42, 65]. It has been observed that majority of COVID-19 patients who were diagnosed with thrombosis have higher levels of neutrophils and the respective markers of NET in the serum [117]. Furthermore, Shi et al. reported that COVID-19 patients who had higher levels of neutrophil activation marker calprotectin had higher need of mechanical ventilation than those who had lower levels of calprotectin [123]. Increased levels of IL-1β in COVID-19 patients stimulate NET release and promote lung damage [124] (Figure 6). It is reported that viable SARS-CoV-2 can directly promote the release of NET from neutrophils. NET release by
Figure 6. (A) Induction of Neutrophil Extracellular Traps (NETs) from neutrophils by SARS-CoV-2 infection. SARS-CoV-2 induce the neutrophil expansion and NET release. NET release promote the IL-1β secretion from macrophages which further leads to inflammation. SARS-CoV-2 also induce the IL-1β secretion from macrophages by activating the NLRP3 inflammasome activation. IL-1β further stimulate NET release and forms a positive feedback loop. (B) NET release causes ARDS, thrombosis and coagulation in COVID-19 patients (Figure illustrated with the help of https://smart.servier.com/).
SARS-CoV-2 is dependent on ACE2, serine protease, protein arginine deamination-4 (PAD-4) and virus replication [125]. It is observed that children suffering from Kawasaki disease (KD) which is characterized by inflammation in blood vessels and heart damage is common in children under the age 5. It is reported that KD patients have increased number of neutrophils and have dysregulated NET formation. This indicates that children might be more susceptible to COVID-19 due to higher levels of neutrophils and NET release [126]. Altogether, these studies showed that increased number of neutrophils enhances the severity of COVID-19 infection through several ways, thus, it is necessary to control the dysregulated activity of neutrophils to lessen the severity.

### 4.4.5. Basophils

Basophils are the least common circulating granulocytes. Generally, basophils secrete Th2 cell type cytokines like IL-4 and IL-13 and provide a link between the innate and adaptive immunity. Basophils are responsible for various Th2 cytokine mediated inflammatory diseases like allergy [127]. Degranulation of activated basophils release histamine, proteolytic enzymes, proteoglycans, leukotrienes, prostaglandins, etc. Several cytokines and chemokines such as IL-3, IL-4, IL-6, IL-9, IL-13, IL-25, CCL5, MIP-1A, MCP-1 and GM-CSF are also released during degranulation [128]. Their role in various viral diseases like influenza and HIV has been well established. Several studies revealed that influenza virus enhances the release of histamine from basophils [129, 130]. Recently, basophil dysregulation has also been observed in case of COVID-19 where it was observed that severe COVID-19 patients have reduced number of basophils as compared to non-severe patients [131]. Thus, it can be inferred that SARS-CoV-2 impairs the immune system by reducing the level of basophils and decreased counts of basophils can be correlated with the severity of COVID-19.

### 4.4.6. Eosinophils

Eosinophils are the circulatory granulocytes but apart from circulation granulocytes are also localized in the gastrointestinal tract and in the lungs. Single stranded RNA viruses induce the eosinophils through TLR7 and promote the cytokine production and degranulation of eosinophils. Furthermore, activation of eosinophils also induces the production of superoxide and nitric oxide in eosinophils and improves their survival. A study revealed that eosinophils provide the antiviral immunity by inducing the expression of MHC-II and costimulatory molecule CD86 as reported in case of influenza virus [132].

Various studies highlighted the relevance of eosinophils in SARS-CoV-2 infection. Severe COVID-19 patients have decreased eosinophil count and enhancement in the eosinophil levels can be used as a recovery marker [133]. Analysis of routine blood of hospitalized COVID-19 patient for 26 days revealed that eosinophil levels start decreasing within one week of the hospitalization. Lowest value was observed on the seventh day of the hospitalization. There were very low counts of eosinophils on tenth day and after approximately 12 days of the hospitalization levels of the eosinophils start recovering [134]. Liu et al. also reported that eosinophil levels were low at the time of hospitalization and recover before discharge [49]. These studies showed that enhanced eosinophil count can be used as markers of improvement in COVID-19. Chen et al. reported that non-survivors have reduced levels of eosinophil than the survivors and thus eosinophils can be employed as an indicator of disease severity [135]. It was earlier hypothesized that asthma patients must be at higher risk of getting infected with SARS-CoV-2 due to lack of antiviral immunity in them. Surprisingly, it has been observed that asthma patients showed lesser susceptibility to COVID-19 because of higher eosinophils count and type 2 cytokines IL-4 and IL-13. This shows the significance of eosinophils in preventing COVID-19. One of the mechanisms through which SARS-CoV-2 impaired the host immune response is via depleting eosinophils count. Altogether, corticosteroids, allergen immunotherapy and monoclonal antibodies against IgE used for the treatment of asthma might also decrease the risk of asthmatic patients to get infected with SARS-CoV-2 [136].

### 4.4.7. Mast cells

Mast cells are distributed throughout the body but predominantly they are localized at the body surfaces that are in contact with the external environment like skin, the nasal cavity, submucosa of the respiratory tract and intestine. In these body parts mast cells reside under the epithelium in connective tissue and form the barrier against the microbes [137]. Mast cells are known regulators of both innate and adaptive immune responses and various other physiological processes like vasodilation, vascular homeostasis, angiogenesis and venom detoxification. It has been reported that mast cells are also responsible for development of various diseases like asthma, allergy, anaphylaxis, gastrointestinal and cardiovascular disorders [137]. The role of mast cells in various bacterial, parasitic and bacterial infections like HIV, dengue virus has also been reported [138]. Several studies showed that viruses stimulate mast cells through TLRs that further induce inflammatory response by stimulating the secretion of inflammatory chemical compounds like TNF-α, tryptase, IL-1, IL-6 and CCL3 [53, 138]. Mast cells were also reported to increase lung injury in H5N1 influenza viral...
infection by secreting proinflammatory compounds like histamine, IFN-γ and tryptase. Inhibition of mast cell degranulation by Ketotifen prevented lung pathogenesis and viral infection [138]. Mast cells produce type 1 and 3 IFNs in response to viral infection but unsuitable inflammatory response of mast cells is responsible for vascular leakage and fibrosis observed in viral infections [139].

Few studies are reported till date on the role of mast cells in COVID-19 but it is observed that mast cells participate in the inflammatory response that is produced in COVID-19. SARS-CoV-2 infection induces mast cells to secrete early inflammatory compounds histamine and protease whereas late infection induces production of inflammatory cytokines IL-1, IL-3 and IL-6. It was observed that suppression of production of these cytokines by using anti-inflammatory cytokines of IL-1 family like IL-37 and IL-38 prevented SARS-CoV-2 infection [50, 51]. This shows that suppression of inflammation due to mast cells can be an effective treatment for the prevention of SARS-CoV-2 infection and various drugs are considered for this. It was observed that Famotidine can prevent the SARS-CoV-2 infection by targeting the activity of histamine receptor H2 [52]. Mast cells stabilizers were also considered which can control the serum TNF levels [140].

Gigante et al. hypothesized that Sodium chromo glycate and Palmitoylethanolamide can prevent the SARS-CoV-2 infection. Sodium chromo glycate is a well-known mast cell stabilizer that prevents inflammation by suppressing the secretion of inflammatory compounds from the mast cells. It prevented lung inflammation in viral infected mice and improved their survival. Conversely, Palmitoylethanolamide is an endogenous fatty acid and nuclear factor agonist. It regulates mast cell homeostasis by downregulating the secretion of TNF-α and histamine and is reported to prevent respiratory tract infection during common cold and influenza [53]. Recently, a study suggested a novel mast cell stabilizer “chromones” which are cheap, anti-inflammatory and safest treatment option against COVID-19 [141].

Flavonoid Luteolin is also considered for preventing inflammation induced by mast cells. Luteolin exhibits antiviral and anti-inflammatory properties and it also suppresses the activity of serine proteases and prevents the entry of SARS-CoV-2 into host cells by binding to the S protein. An analogue of Luteolin called Tetramethoxyluteolin suppresses the release of proinflammatory cytokines and chemokines such as TNF, IL-1β, CCL2 and CCL5 from human mast cells [54].

Apart from producing inflammatory response mast cells can also induce lung injury through other mechanism. It is observed that mast cells produce and store enzymatically active renin which is homologous to the renin produced in human kidneys. Release of renin by the degranulation of bronchial mast cells lead to the formation of angiotensin II that further promotes bronchoconstriction and also upregulated the RAS activity in airways [142]. A study revealed that α and β chymases secreted by the mast cells also generate angiotensin II from angiotensin I [143]. Tryptases and chymases are serine proteases and are chief constituents of mast cell granules that induce the influx of inflammatory cells [144]. Serine proteases are also required for priming of S protein. Mast cells thus have very important role in the pathogenesis of COVID-19 but their role need to be further explored.

4.4.8. T helper cells

CD4+ and CD8+ T cells play a vital role in virus clearance. Activated CD4+ T cells produce cytokines that further induce the activation of other immune cells especially CD8+ T cells and B cells for killing the pathogens. CD8+ T cells exhibit potential to directly kill the virus infected target cells. Among T helper cells, Th1 cells generally play major role in preventing viral infections. It has been observed that like SARS-CoV, SARS-CoV-2 also impairs the host adaptive immune response. In similarity to SARS-CoV infected patients, reduction in CD4+ and CD8+ T lymphocytes (lymphopenia) has also been observed in COVID-19 patients [131, 145]. T cells of COVID-19 patients also showed significantly higher expression of exhaustion markers PD-1 and Tim-3 during symptomatic stages [146]. In COVID-19 patients, CD4+ T cells predominantly express IFN-γ, TNF-α and IL-2 cytokines and it has been found that in severe patients, levels of these cytokines get reduced in comparison to mildly infected patients [47].

There are various reasons responsible for the depletion of T cells in the COVID-19 patients. It is observed that the number of T cells (CD4+ and CD8+) in peripheral blood is negatively interrelated with the serum levels of inflammatory cytokines IL-6, IL-10 and TNF-α. Severe SARS-CoV-2 infected patients have higher levels of these cytokines along with reduced T cells population. Recovered patients have decreased level of IL-6, IL-10 and TNF-α and have restored T cell population [90]. Type 1 IFN is also found to be induced in response to SARS-CoV-2 infection and it is observed that induction of type 1 IFN and TNF-α directly cause the observed lymphopenia. Previous reported studies demonstrated that type 1 IFN and TNF-α cytokines control the recirculation of lymphocytes by preventing the egress of lymphocytes from lymphoid organs [147, 148]. Postmortem analysis of six deceased COVID-19 patients revealed that SARS-CoV-2 induce the apoptosis of lymphocytes. It does this by upregulating the Fas expression and induces the macrophages to secrete IL-6.
frequently after a year of the infection [155]. MAS and reduced expression of Human Leukocyte Antigen-DR isotype (HLA-DR) is observed in critically ill COVID-19 infected patients and is found to be associated with the depletion of CD4+ T cells. MAS are recognized by higher production of IL-6 and thus blocking of IL-6 with Tocilizumab enhanced the counts of circulating lymphocytes. This shows that IL-6 secreted by macrophages might be responsible for lymphopenia observed in severe COVID-19 patients [83]. Lymphocytes infiltration can also be responsible for their reduce levels in peripheral blood. ScRNA of BALF revealed the infiltration of lymphocytes into lungs in COVID-19 patients [79]. Similarly, infiltration of lymphocytes into lungs is found in postmortem biopsy reports of deceased critically ill COVID-19 patients [69].

Various studies analyzed the adaptive immune response against the SARS-CoV-2 and it is observed that large number of SARS-CoV-2 epitopes are recognized by T cells [149]. Epitope mapping of SARS-CoV revealed that most of the adaptive immune response was found to be raised against the structural proteins viz. S, M, E and N proteins [150]. In SARS-CoV-2 infection majority of the T cell response observed are mounted against the S protein [151]. S protein primarily induces CD4+ cells which are mainly characterized by CD69+CD137+ markers. Majority of these CD4+ T cells are of central memory type as they express CD45RA and CCR7. S protein also activates the CD8+ T cells. Like CD4+ T cells most of these activated CD8+ T cells exhibits CD69+CD137+ phenotype and are of effector memory type (CCR7-) or of terminally differentiated effector T cell type and produce IFN-γ [151]. Huang et al. showed that all SARS-CoV-2 infected patients exhibit structural peptide specific CD4+ T cells response and 80% of the patients also harbored peptides specific CD8+ T cells [86]. Surprisingly, it has been observed that 20% of the healthy controls those were never exposed to SARS-CoV-2 also possessed virus specific T cells [86]. Furthermore, Maters et al. showed that some of these pre-existing memory CD4+ T cells found to be reactive against SARS-CoV-2 also possess cross reactivity against common cold coronaviruses. This showed that these memory CD4+ T cells are raised in the healthy donors due to their previous exposure with the common cold coronaviruses and would also provide immunity against SARS-CoV-2 [152]. But recent studies have shown that protective immunity due to seasonal CoVs is short lived and provide little relief against SARS-CoV-2 infection [115, 153, 154]. It was observed that reinfecion due to seasonal CoVs occurred frequently after a year of the infection [155].

It is reported that CD4+ T cells plays a vital role in preventing the SARS infection and deficiency of these cells hamper viral clearance that further promote lung inflammation. Majority of the CD4+ T cells in COVID-19 are induced to become Th1 that produce GM-CSF and other cytokines which stimulate the IL-6 producing CD14+CD16+ monocytes. These monocytes and Th1 cells accumulate in the pulmonary circulation and cause lung damage [151, 156]. GM-CSF producing CD4+ T cells are indicators of disease severity and are responsible for various inflammatory diseases like juvenile arthritis, multiple sclerosis and sepsis [157]. Within the CD4+ T cells, higher proportion of CCR6+ Th17 cells has also been observed in COVID-19 patients but the role of Th17 cells in COVID-19 is yet unknown [69, 151]. Th2 response was also observed in severe cases as there were higher levels of IL-2, IL-6 and IL-10 present in critically ill COVID-19 patients [151, 158]. CD4+ T cells of the severe COVID-19 patients produce low levels of IFN-γ than the non-severe patients (Chen et al., 2020). CD8+ T cells of COVID-19 patients are found to be functionally exhausted and showed enhanced expression of IFN-γ, TNF-α along with high levels of granzyme B and perforin (degranulated state). IFN-γ and TNF-α expressing CD8+ T cells were found to be present more in severe COVID-19 patients than the patients who have mild infection [159]. CD8+ T cells of COVID-19 patients admitted in hospitals express higher levels of GM-CSF than non-ICU patients [160]. Reduced number of T regulatory cells (Tregs) and naive T cells are also observed in COVID-19 patients [131]. Earlier a study reported that Tregs induced recovery in mice suffering from fatal pneumonia [161] and thus reduction in their numbers in COVID-19 patients might promote lethal pneumonia. Previous studies reported that γδ T cells exhibits strong antiviral activity and protected from pneumonia caused by influenza virus [159, 162]. In case of COVID-19, reduction in number of γδ T cells had been observed in the peripheral blood and showed higher expression of activation marker CD25 and CD4 [73, 163]. Bcl-6+ T follicular helper (TFH) cells are also found to be diminished in COVID-19 [164]. Phenotypic alterations were also observed in the circulating unconventional T cells (MAIT, γδ T cells and iNKT cells) of the COVID-19 patients. Highly activated unconventional T cells were detected in the airways of COVID-19 patients. It was also observed that the expression of CD69 activation marker on the blood MAIT and iNKT cells is the indicator of COVID-19 severity [165].

Stimulation of T cell response is essential for controlling the spread of infection but proper induction of memory T cell response is also required for preventing further reinfections. SARS-CoV-2 infection is very recent and thus it is not possible to study the memory T cell response in COVID-19 patients. But there are studies on memory T cell response in SARS-CoV patients and can be referred to understand the memory T cell response in SARS-CoV-2.
Neutralizing antibodies (NAbs) mostly comprise of IgG antibody response take place in SARS-CoV-2 infection. Analysis of blood of SARS-CoV-2 infected patients 2 weeks after discharge revealed that there was decreased level of virus specific T cell response than the IgG specific antibody response [162]. Recently it was observed that memory CD4+ and CD8+ T cells induced against SARS-CoV-2 have half-life of 3–5 months [170]. Further studies are needed to understand the virus specific T cell memory response in the recovered patients.

4.4.9. B cells

Antibodies against SARS-CoV-2 are produced by highly varied TCR and BCR V (D) J recombination [86]. In SARS-CoV infection seroconversion take place after 10–11 days of the infection [171]. Like SARS-CoV delayed antibody response take place in SARS-CoV-2 infection. Neutralizing antibodies (NAbs) mostly comprise of IgG and IgM antibodies against the SARS-CoV-2 and start appearing after 7–14 days of the disease onset [172]. Most of the NAbs are produced against the RBD of the S protein and nucleocapsid protein [173]. RBD specific monoclonal antibodies are isolated from the COVID-19 patients. It has been observed that these antibodies have great neutralizing ability and possess specificity only for SARS-CoV-2 and does not cross react with the RBD of SARS-CoV and MERS-CoV [174].

B cells not only protect from the primary infection but also provide immunity against further challenges by producing memory B cell responses. Memory B cells produce secondary immune response and protect from reinfections by producing an immediate response against the antigen. Knowledge of the memory B cells response against the SARS-CoV-2 is of the great interest. It will be very helpful in designing vaccine strategies for combating COVID-19. Memory B cell response in SARS-CoV and MERS-CoV was studied by various groups. During SARS infection it was observed that recovered patients had higher and sustainable NAbs against the S and N proteins of SARS-CoV [175]. IgG antibody and NAbs against the SARS-CoV were recovered in the patients even after 2 years of the disease onset [176]. Liu et al. analyzed IgG and NAbs at regular interval for 2 years in SARS-CoV infected patients. They observed that levels of the IgG antibody and NAbs were higher at four months after the infection and then decreased thereafter. IgG and NAbs decreased markedly after the 16 months [177]. In another study however it was reported that IgG+ memory B cells were higher after 2 months of the infection and then decreased by 6th and 8th months [178]. Memory IgG antibodies specific for SARS-CoV were not detected 6 years after the infection which shows that memory B cell response does not protect after a longer period [168]. NAbs were detected against the MERS-CoV in the 86% of the patients even nearly 3 years after the outbreak [179]. But as SARS-CoV-2 infection is very recent there is no substantial data related to the memory B cell response in COVID-19. But there are some studies which may be helpful in understanding the memory B cell response against SARS-CoV-2. In study of Thevarajan et al. it had been observed that stimulation of CD19+CD27+CD38hi antibody secreting B cells is concomitant with the increase in TFH cells [180]. TFH cells help in the development of memory B cells response. ScRNA of peripheral blood mononuclear cells revealed that population of plasma B cells exists in recovered patients [73]. It had been observed that primary infection of SARS-CoV-2 can protect from the further reinfections [89]. RBD specific memory IgG antibodies were observed in the COVID-19 patients [174]. Study from Addetia et al. also had revealed that NAbs produced against SARS-CoV-2 can protect from reinfection. They showed that crew members of US fishing vessel who had NAbs against CoVs before departure were protected against SARS-CoV-2 infection caused during the departure. Conversely, crew members who did not have the NAbs against SARS-CoV-2 got infected [181]. In another study, Gudbjartsson et al. tested the levels of SARS-CoV-2 antibodies in around 30 000 people including more than 1200 people who had recovered from the COVID-19 infection. They observed that more than 90% of the recovered patients had antibodies against SARS-CoV-2. Antibodies levels peaked after 2 months of diagnosis and did not decline until 4 months [182]. These studies have shown that memory response exists against SARS-CoV-2 but different studies provide different perspective of memory response in COVID-19. Seow et al. analyzed the antibody responses in 65 infected people for 94 days. They observed that among them 19 patients had higher levels of antibodies than people with mild disease. They also observed that in most of the people antibody levels started falling about a month after the infection [183]. This study declined the possibility of designing a vaccine based on NAbs. Recently, a study by To et al. provided the first evidence of reinfection. They observed that a Hong Kong man who was first infected with SARS-CoV-2 in April was found infected again four months later (142 days) with different variant of
SARS-CoV-2. This study also shows that immunity against SARS-CoV-2 can wane after sometime [184]. But again two recent studies have declined the possibility that memory response against SARS-CoV-2 diminish after sometime. In study of Wajnberg et al. which included more than 30 000 SARS-CoV-2 infected individuals it was observed that in 90% of the infected people antibodies against SARS-CoV-2 persisted for at least 5 months after the infection [185]. Similar results were reported by Dan et al. They analyzed the circulating memory response against SARS-CoV-2 and observed that spike IgG antibody remained for at least 6 months after the infections [170].

Antibodies usually protect from the infections but sometimes can also enhance the severity of the viral infection by mediating the entry of virus into the host cells, referred as antibody dependent enhancement of infection (ADE). ADE occurs when complex of virus and antibody comes in contact with cells having Fc receptors. Antibodies bind to the virus through the S protein on one end and to the Fc receptor on another end. Binding of antibody-virus complex to the Fc receptor bearing cells induces its internalization and entry of virus into the cell [186, 187]. ADE phenomenon hampers the activation of proper immune response against SARS-CoV-2. It is observed that IgG antibody against the S protein promote polarization of macrophages into proinflammatory phenotype which leads to acute lung injury. Blocking of Fcγ receptors inhibited the inflammatory response of macrophages and prevented lung injury. This shows that ADE might be responsible for dysregulation of macrophages activity [188]. SARS-CoV is also reported to infect macrophages in IgG dependent ADE by binding to the FcγRII receptors [189]. It was observed that monoclonal antibodies against the RBD of MERS also mediated the entry of MERS-CoV into host cell. Thus, ADE can create a huge problem in designing a vaccine against the virus. However, in contrast to the above discussion a recent study showed that NAbs against the SARS-CoV-2 RBD has not promoted ADE mediated viral entry [190].

5. Age and immune responses against COVID-19

It has been suggested that hyperinflammatory immune response is one of the reasons for higher mortality rate in aged population [191]. A study demonstrated that older people have higher expression of prostaglandin D2 which impairs the migration of respiratory DCs to draining lymph nodes and thus the induction of specific T cell response [192]. As age advances, lymph nodes undergo some changes that further effect the formation and maintenance of naive T and B cells. Recently, a study reported that with age activation, proliferation and differentiation of T cells is also impaired [193]. Thus, it is plausible to suggest that impairment in immune response during aging enhance the severity of COVID-19 in aged people.

In contrast to aged population children rarely gets SARS-CoV-2 viral infection. Children infected with virus showed mild symptoms and rarely infection progresses in them into the severe stage [194]. It has been observed that only 6% of the children below the age of 18 have severe symptoms whereas 50% of them developed only mild symptoms [195]. Furthermore, a recent study has shown that people below the age of 20 are less susceptible to household transmission whereas people above the age 60 are highly susceptible [196]. Surprisingly, it has been observed that immune system plays crucial role in protection of children from COVID-19. It is observed that children have decreased levels of proinflammatory cytokines TNF-α and IL-6 that further reduces the neutrophil infiltration and lung injury. Along with the reduced proinflammatory cytokine levels, higher levels of immunomodulatory cytokines such as IL-10 and IL-13 has also been observed in children. Moreover, children have decreased expression of ACE2 and TMPRSS2 is lung epithelium and exhibits better capability of repairing the lung infection [194]. Thus, it can be concluded that multiple factors work in synergistic manner to protect children from developing severe COVID-19 disease. But the results of recent research have shown that viral load of SARS-CoV-2 infected children is significantly higher than the adults with severe disease [197]. This study thus contradicts with the previous results which claim that children are at low risk of COVID-19.

6. Immunotherapies for COVID-19

Currently, there is no effective therapy against SARS-CoV-2 infection. Inhibition of inflammatory response caused by dysregulated immune response can be an effective therapy in preventing COVID-19. Below we discussed some of the immunotherapeutic strategies that can be employed as potential treatments option for COVID-19 (Figures 7 and 8).

6.1. Immunosuppressive drugs

Thalidomide is an anti-inflammatory, anti-angiogenic and anti-fibrotic immune modulator. Interestingly, it has been observed that thalidomide in combination with low dose glucocorticoid prevented pneumonia in SARS-CoV-2 infected patients [198]. Currently clinical trials are in progress for Thalidomide to further explore its role in prevention of COVID-19 (ClinicalTrials.gov https://clinicaltrials.gov/ct2/show/NCT04273529, 2020). Recently, a study
demonstrated that Lianhuaqingwen, a traditional Chinese medicine prevented SARS-CoV-2 infection by significantly reducing the pro-inflammatory cytokines and chemokines such as TNF-α, IL-6, CXCL10 and MCP-1 levels [199]. Clinical trials on Tocilizumab and on anti-TNF-α antibody are also in progress [45]. Clinical trials for drugs against GM-CSF such as Lenzilumab, Gimsilumab and Namlumab are also under way [200–202] (Figure 9). Several studies reported the immunomodulatory properties of a common antimalarial drug Chloroquine. Clinical applications of chloroquine are increasing day by day. For years, Chloroquine is being used in chronic kidney disease, oncology, cardiovascular and rheumatological disorders. Currently, Chloroquine is suggested for the treatment of COVID-19. Antiviral therapy with Chloroquine recovers the levels of NK cells and CD8+ T cells during convalescent period in COVID-19 patients [47]. A systemic review on Chloroquine which includes six studies and various ongoing trials suggested that there is evidence of effectiveness of Chloroquine in preventing COVID-19. But for better understanding of anti-viral properties of Chloroquine some high-quality effective trials are recommended [203]. Hydroxychloroquine is also an anti-inflammatory drug that works by same mechanism as Chloroquine. Hydroxychloroquine has been well established as an immunomodulatory drug. Furthermore, Hydroxychloroquine prevents cellular autophagy, an important step for innate and adaptive immune system activation [204]. Hydroxychloroquine is preferred over Chloroquine for treatment of malaria as it is safer. It is reported to be three times more effective than Chloroquine in treatment of SARS-CoV-2 infection at a dose of 400 mg twice along with maintenance dose of 200 mg twice daily for 4 days [205]. Hydroxychloroquine reduced viral load in COVID-19 patients [206]. Chloroquine and Hydroxychloroquine recovered the levels of CD8+ T cells and NK cells in convalescent period in COVID-19 patients [47].
patients and Azithromycin in addition to Hydroxychloroquine reinforced its anti-viral potential and makes it more efficient in eliminating the virus [206]. Zheng et al. suggested that cross reactive antibodies are responsible for cytokine storm and their inhibition can prevent the SARS infection. They proposed that inhibition of activated memory B cells in early stage of COVID-19 patients can prevent the production of cross-reactive antibodies. mTOR inhibitors can potentially inhibit activation of memory B cells and can be used for the treatment of COVID-19 [207].

6.2. Plasma therapy

In passive antibody therapy, antibodies against a particular agent are given to the infected person that provides immediate relief to infected person. In case of COVID-19, the source of antibodies against the SARS-CoV-2 is the serum of patient who recovered from COVID-19, i.e., convalescent sera [208]. Earlier, convalescent plasma therapy has also been employed for the treatment of SARS, MERS and influenza A virus pandemic in 2009 [209–211]. Recently, several studies evidenced that convalescent sera can be an effective therapy for the treatment of COVID-19. It is reported that treatment of 10 severe COVID-19 patients with single dose of 200 mL of convalescent sera markedly enhanced the NAbs against the SARS-CoV-2 virus [55]. Furthermore, Zhang et al. reported that all severely ill patients who received convalescent sera along with supportive care recovered from COVID-19 infection [212]. Convalescent sera administration improved health of the patients [213]. Food and drug administration (FDA) also approved the use of convalescent sera for treatment of critically ill COVID-19 patients [214].

6.3. Immunosuppression by corticosteroids and nonsteroidal drugs

Corticosteroids are steroid hormones and classified into glucocorticoids and mineralocorticoids [215]. Glucocorticoids are involved in various essential biological processes like metabolism, development, reproduction, cardiovascular
Figure 9. The schematic representation of SARS-CoV-2 life cycle in stepwise manner depicting the potential therapeutic approaches against SARS-CoV-2: (1) SARS-CoV-2 enter into the cell by binding to the ACE2 receptor present on the host cell membrane via S protein which mediates viral-host membrane fusion and viral entry. (2) Entry of SARS-CoV-2 results in the uncoating of viral RNA into cytoplasm which then undergoes translation to produce polyproteins pp1a and pp1b, which are further processed by virus-encoded proteinases into individual nonstructural proteins (nsps). (3) Replication transcription complex is formed by some of the nsps. Replication transcription complex uses the (+) strand genomic RNA as template. Following replication (+) strand genomic RNA is produced which becomes the genome of the new viral particle. (4) Subgenomic RNAs synthesized through transcription are translated into viral structural proteins: S (Spike) protein, M (Membrane) protein, E (Envelope) protein, N (Nucleocapsid) protein. (5) N protein combined with the (+) strand genomic RNA to form the nucleoprotein complex. S, E and M proteins enter into endoplasmic reticulum (ER) and transported to Golgi apparatus. The nucleoprotein complex and S, E and M proteins are further assembled in the ER-Golgi intermediate compartment (ERGIC) to form mature virion. (6) Virions are released from the host cells by exocytosis. Many potential therapeutic approaches are considered for the management of COVID-19. These therapeutics act on various steps of the SARS-CoV-2 life cycle. Monoclonal antibodies and Convalescent plasma targets the S proteins of SARS-CoV-2 from interacting with the ACE2 receptor. Camostat mesylate and Nafomostat targets the serine protease (TMPRSS2) and prevents the S protein cleavage which is required for the viral fusion to the host cells. Lopinavir and Ritonavir targets the proteolysis of polypeptide chains. Remdesivir prevents the replication of SARS-CoV-2 and drugs like Tocilizumab, Lenzilumab, Gimsilumab, Namilumab and Lianhuaqingwen inhibit the pro-inflammatory response.
functions, water and electrolyte balance, growth and development [216]. Glucocorticoids exhibits immunomodulatory properties and are being increasingly used for the treatment of various immune related disorders like inflammatory bowel disease (IBD), asthma, rheumatoid arthritis (RA), allergy, septic shock and multiple sclerosis [215]. Corticosteroids were previously considered for the treatment of MERS, SARS and influenza virus [217, 218].

Now, their use is also being suggested for the treatment of severe COVID-19 cases to suppress lung inflammation [21]. Recently, a study revealed that no association exists between corticosteroids treatment and virus clearance in case of COVID-19 [219]. Clinical outcomes from other studies also do not support the theory that corticosteroids can prevent lung inflammation in COVID-19 patients [220]. But recently in a randomized clinical trial (2100 participants) in United Kingdom (UK) it has been observed that synthetic glucocorticoid dexamethasone suppressed the deaths by one third in the COVID-19 patients who were on ventilators. But dexamethasone showed no impact on the non-severe patients but it was very impactful in case of severe patients that require ventilators and reduced their chance of dying by 20% [221]. Moreover, another study reported that short term treatment with dexamethasone significantly reduced the C-reactive protein levels and the length of stay of COVID-19 patients in hospitals [222].

Nonsteroidal drugs (NSDs) are the common prescribed drugs indorsed for reducing pain and inflammation. NSDs inhibit cyclooxygenase-2 (COX-2) enzyme which are required for prostaglandins induced pain and inflammation [223]. It has been observed that NSDs like ibuprofen and aspirin suppressed antibody production and prevented MHC restricted antigen presentation in DCs [224]. Because of their immunosuppressive properties NSDs are recommended for the inhibition of SARS-CoV-2 infection. Naproxen and Indomethacin are the two NSDs that are mainly suggested for COVID-19 infection. Naproxen is an antiviral drug that prevented infection due to influenza A and B virus [225] and can be an effective therapy for COVID-19 [226]. Indomethacin has anti-inflammatory and antiviral activities and reported to inhibit SARS-CoV infection by suppressing RNA synthesis in vitro in monkey VERO cells [227]. In vivo studies showed reduced level of canine CoVs in dogs, and thus, can be considered for suppressing SARS-CoV-2 infection [228].

### 6.4. Intravenous immunoglobulins

Intravenous immunoglobulins (IVIG) is the mixture of immunoglobulins pooled from thousands of normal individuals [229]. IVIG has been employing for the treatment of various neurological, hematological, nephrological, dermatological, rheumatological and immune related disorders [230]. Several mechanisms are being involved in preventive activities of IVIG. A study found that IVIG promoted cytokine balance, inhibited auto-reactive T cells, reduced antibody production from B cells and also reduced macrophage activity [231]. Recently, the relevance of IVIG has also been explored in case of COVID-19. It has been observed that administration of IVIG at a dose of 0.3–0.5g/kg for 5 days to critically ill COVID-19 patients those were not responding to other treatment options reduced the progression of pulmonary lesions and improved their clinical and respiratory conditions [232]. Furthermore, another study reported that usage of IVIG for treatment of severely ill COVID-19 patients decreased the need of mechanical ventilation, reduced the length of hospital stay and decreased the mortality rate [233]. It was observed that all IVIG treated patients had satisfactory recovery with no adverse effects [234]. These studies raise the possibility that IVIG may provide a better way in the treatment of COVID-19.

### 6.5. Vaccines

There is an urgent requirement of vaccine for the prevention of SARS-CoV-2 infection because of its highly transmissible nature and pathogenesis. Before SARS-CoV-2 there were epidemics due to SARS-CoV and MERS-CoV. Several vaccines like inactivated virus vaccines, live attenuated virus vaccines, viral vector vaccines, subunit vaccines and DNA vaccines have been designed against SARS-CoV and MERS-CoV. But so far, they have been tested only on animals and no data on their safety and efficacy in case of human is available [235, 236]. Thus, these vaccines cannot be recommended for the treatment of SARS-CoV-2 infection. But the knowledge that has been gained in making these vaccines is now being used for designing a better SARS-CoV-2 vaccine. All over the world within few weeks of SARS-CoV-2 outbreak many research groups started making vaccines. After 2 months of the outbreak it was observed that over 37 pharmaceutical companies or research institutes were in the race of making COVID-19 vaccine [237]. Right now, various biopharmaceutical companies in collaboration with research institutes are designing several types of vaccines like DNA vaccines, RNA vaccines, inactivated vaccines, viral vector-based vaccines and recombinant vaccines for COVID-19 (Table 2). Some of these vaccines have shown positive results in clinical trials. Pfizer and BioNtech vaccine candidate, BNT162b2 has found to be more than 90% effective in preventing COVID-19. These companies have received emergency use authorization (EUA) in various countries like UK, America, Europe, Australia and have already
started vaccinating people [238–241]. Moderna vaccine candidate, mRNA-1273 was found to be 94.5% effective in phase 3 clinical trials and the vaccine is already approved by America [241, 242]. Phase 3 clinical trials of Oxford-AstraZeneca COVID-19 vaccine, ChAdOx1 nCoV-19 or Covidshield (in India) has shown that the vaccine is 70% effective in preventing COVID-19 and have already started vaccinating people in UK and India [241, 243]. The Russian Direct Investment Fund (RDIF) and Gameleya center have also announced their vaccine candidate, SPUTNIK-V with 92% efficacy in phase 3 clinical trials and is already approved by countries like Argentina and Belarus [244]. Bharat biotech vaccine candidate Covaxin is also approved by Indian government and the Indian government has already started vaccinating its citizens [245]. Although the process of vaccination has started in various countries it is still early to estimate whether these vaccines will be able to provide effective immunity against COVID-19 as postvaccination data is still awaited.

### 6.6. Monoclonal antibodies

Designing antibodies against the RBD of the S protein is one of the strategies used to combat COVID-19. RBD of S protein binds to the ACE2 receptor and mediate viral entry. Blockade of RBD binding to the S protein by antibodies against it can prevent the viral entry into host cells and thus prevents SARS-CoV-2 infection [246]. In 2010, Berry et al. successfully designed mAbs against the SARS-CoV RBD which cross neutralized and prevented its binding to the ACE2 receptor [247]. Study of Berry et al. showed the potential and feasibility of mAbs as vaccine for combating COVID-19. SARS-CoV RBD specific human mAbs can also be explored against the SARS-CoV-2. After screening of various mAbs, it has been observed that one SARS-CoV specific mAb CR3022 mAb bound effectively with the SARS-CoV-2 RBD thus provide a treatment option for COVID-19 [248]. Moreover, 47D11 mAb designed against the conserved epitope in the SARS2-S-S1b domain was also found to neutralize SARS-CoV-2 [5]. Lei et al. designed fusion protein comprised of extracellular domain of the ACE2 receptor and fc region of the human IgG1 antibody. The fusion protein showed high binding affinity for both SARS-CoV and SARS-CoV-2. They observed that fusion protein was capable of neutralizing pseudotyped SARS-CoV and SARS-CoV-2 viruses [13]. Not only the RBD but other parts of the S protein can also be targeted for vaccine preparation. Chi et al. isolated the mAbs from the convalescent SARS-CoV-2 infected patients and observed that a mAb named as 4A8 showed high potency of neutralizing the SARS-CoV-2. 4A8 showed no binding with RBD but to the N terminal domain of the S protein. This showed that mAbs targeting N terminal domain can also acts as promising therapeutic against COVID-19 [249]. Recently Liu et al. isolated 61 mAbs against SARS-CoV-2 from 5 severe hospitalized patients. They observed that 19 antibodies among these potently neutralized SARS-CoV-2 in vitro. Epitope mapping revealed that these antibodies were directed against the RBD.

### Table 2. Status of different types of COVID-19 vaccines.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Type of vaccine</th>
<th>Name</th>
<th>Developer</th>
<th>Trial phase</th>
<th>Registration number/Clinical trial number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>RNA-based vaccine</td>
<td>BNT162</td>
<td>Pfizer/BioNTech</td>
<td>Phase 3</td>
<td>NCT04368728</td>
</tr>
<tr>
<td>2.</td>
<td>mRNA-based vaccine</td>
<td>mRNA 1273</td>
<td>Moderna</td>
<td>Phase 3</td>
<td>NCT04283461</td>
</tr>
<tr>
<td>3.</td>
<td>Viral vectored vaccine</td>
<td>Inactivated vaccine</td>
<td>Wuhan Institute of Biological Products, China National Pharmaceutical group (Sinopharm)</td>
<td>Phase 3</td>
<td>ChiCTR2000031809</td>
</tr>
<tr>
<td>4.</td>
<td>Adenovirus-based vaccine</td>
<td>Ad5-cCoV</td>
<td>CanSino Biologics</td>
<td>Phase 3</td>
<td>NCT04313127</td>
</tr>
<tr>
<td>5.</td>
<td>Inactivated vaccine</td>
<td>CoronaVac</td>
<td>Sinovac</td>
<td>Phase 3</td>
<td>NCT04456595</td>
</tr>
<tr>
<td>6.</td>
<td>Adenovirus vectored vaccine</td>
<td>ChAdOx1 nCoV-19 or Covidshield</td>
<td>The University of Oxford</td>
<td>Phase 3</td>
<td>NCT04324406</td>
</tr>
<tr>
<td>7.</td>
<td>Adenovirus vectored vaccine</td>
<td>Sputnik V</td>
<td>Gamaleya Research Institute of epidemiology and Microbiology, Health Ministry of the Russian Federation</td>
<td>Phase 3</td>
<td>NCT04530396</td>
</tr>
<tr>
<td>8.</td>
<td>Recombinant vaccine</td>
<td>NVX-CoV2373</td>
<td>Novavax</td>
<td>Phase 3</td>
<td>NCT04533399</td>
</tr>
<tr>
<td>9.</td>
<td>Whole virion inactivated vaccine</td>
<td>Covaxin</td>
<td>Bharat Biotech</td>
<td>Phase 3</td>
<td>NCT04471519</td>
</tr>
<tr>
<td>10.</td>
<td>Inactivated vaccine</td>
<td>BBIBP-CoV</td>
<td>Beijing Institute of Biological Products, China National Pharmaceutical group (Sinopharm)</td>
<td>Phase 1/2</td>
<td>–</td>
</tr>
<tr>
<td>11.</td>
<td>DNA vaccine</td>
<td>GX-19</td>
<td>Genexine, Arcturus Therapeutics and Duke-NUS Medical School</td>
<td>Phase 1/2</td>
<td>NCT04445389</td>
</tr>
<tr>
<td>12.</td>
<td>mRNA-based vaccine</td>
<td>LUNAR-COV19</td>
<td>Arcturus Therapeutics and Duke-NUS Medical School</td>
<td>Phase 1/2</td>
<td>–</td>
</tr>
<tr>
<td>13.</td>
<td>DNA vaccine</td>
<td>INO-4800</td>
<td>Inovio Pharmaceuticals</td>
<td>Phase 1/2</td>
<td>–</td>
</tr>
<tr>
<td>14.</td>
<td>mRNA-based vaccine</td>
<td>CVnCoV</td>
<td>CureVac</td>
<td>Phase 1</td>
<td>NCT04515147</td>
</tr>
<tr>
<td>15.</td>
<td>Recombinant subunit vaccine</td>
<td>SCB-2019</td>
<td>Clover Biopharmaceuticals</td>
<td>Phase 1</td>
<td>NCT04405908</td>
</tr>
<tr>
<td>16.</td>
<td>Recombinant vaccine</td>
<td>COVAX-19</td>
<td>Vaxine Pty Ltd.</td>
<td>Phase 1</td>
<td>NCT04453852</td>
</tr>
<tr>
<td>17.</td>
<td>Recombinant vaccine</td>
<td>Plant-based adjutant COVID-19 vaccine</td>
<td>Medicago</td>
<td>Phase 1</td>
<td>NCT04450004</td>
</tr>
<tr>
<td>18.</td>
<td>Lentivirus vector vaccine</td>
<td>COVID-19 aAPC</td>
<td>Shenzhen Geno-Immune Medical Institute</td>
<td>Phase 1</td>
<td>NCT04299724</td>
</tr>
</tbody>
</table>
and NTD of the S protein [250]. Mercado et al. also showed the efficacy of adenovirus serotype 26 (AD26) vector-based vaccine in 52 rhesus macaques. They showed that AD26 induced NAbs and protected against SARS-CoV-2 challenge [251]. Furthermore, Baum et al. made a cocktail of two potent mAbs (REGN10987 + REGN10933) against the SARS-CoV-2 S protein. They demonstrated that the antibody cocktail reduced the viral load in rhesus macaques and golden hamsters [252].

### 6.7. Herbal drugs

A number of traditional Chinese medicines (TCM) and herbal drugs like Gypsum, Semen armeniacaeamarae, Liquorice, Scutellariabaiacalenishas been reported for the treatment of COVID-19 [253]. On 14 April 2020 China approved the usage of herbal drugs Xuebijing for severe SARS-CoV-2 infection and Lianhuaqingwen and Jinhuaqinggan for the treatment of mild conditions. TCM and herbal drugs are observed beneficial in preventing SARS-CoV-2 infection as they have antiviral and anti-inflammatory properties. It has been reported that approximately 60 107 COVID-19 patients were treated with TCM or herbal drugs [254]. But some of these herbal drugs have been found to shown adverse effects which limit their use in COVID-19 treatment [255]. TCMs are sometimes adulterated and are not processed properly [256].

Indian Ayurveda also has great potential of treating deadly diseases and can replace the TCM in treatment of COVID-19. Ayurveda which means “The science of life” is the oldest medicinal science was originated in India dating back 5000 years. Ayurveda recommends the use of traditional Indian medicines, medicinal plants and spices for the treatment of diseases. Medicinal herbs promote immune system development and exhibits anti-inflammatory, anti-cancer and anti-oxidant properties. Medicinal herbs modulate the immune system and maintain immune homeostasis [257]. Several ayurvedic plants exhibits immunomodulators like: *Curcuma longa* (Cucumin), *Tinosporacordifolia* (Guduchi), *Andographispaniculate*, *Dioscorea japonica*, *Boerhaviaadiiffusa* (Punarnavine), *Withaniasominfera* (Ashwagandha), *Allium sativum* (Garlic), *Terminaliaarjuna* and *Mangiferaindica* [258]. Ministry of AYUSH, Government of India also proposed guidelines for use of ayurvedic herbs during COVID-19 pandemic. Recently, a group published an intervention plan based on ayurvedic plants for the treatment of COVID-19 [254]. A recent study which collected data from 163 countries also demonstrated the role of spices in COVID-19. The study had shown that correlation exits between the total cases of COVID-19/million and the gram of spices consumed/capita/day. They observed that nations which consumed less amount of spices/capita had more cases of COVID-19/million than high spice consuming nations [259]. But till now there is no study which directly proves the role of Indian Ayurvedic herbs or medicines in the treatment of COVID-19. Nevertheless, Ayurvedic herbs have the potential of curing various diseases and thus there should be clinical trials on further investigating their role in preventing COVID-19. Ayurvedic herbs can also provide a safe and better option in place of TCM.

### 7. Inhibitors of viral entry and replication

Apart from immunosuppression, blocking of viral entry into cells can be an effective way of preventing the SARS-CoV-2 infection. SARS-CoV-2 enters into the cells with the help of S protein and required host protease TMPRSS2 and cathepsin L for S protein priming. Hoffman et al. showed that viral infection can be prevented by blocking the viral entry using the inhibitors for TMPSS2 protease. They used camostat mesylate to block TMPSS2 and observed that it had prevented viral entry and thus viral infection. Hoffman et al. also showed that NAbs produced in the convalescent patients against SARS-CoV-2 S proteins also blocked S protein mediated viral entry [37]. Earlier Yamamoto et al. showed that Nafomostat, an inhibitor of TMPSS2 had inhibited S protein mediated entry of MERS virus [260]. Two antiviral drugs remdesivir and chloroquine has also been reported to prevent SARS-CoV-2 infection. Remdesivir acts on post virus entry. It is an adenosine analogue and gets incorporated into the viral RNA chains which cause premature termination. Beigel et al. had shown that Remdesivir decreased the recovery time in hospitalized COVID-19 patients to 11 days from 15 days [261]. Remdesivir is widely used for the treatment of COVID-19 and on May 1st FDA EUA to the Remdesivir for COVID-19 treatment. On August 28 FDA expanded the EUA to treat hospitalized COVID-19 patients [262]. Chloroquine which is common antimalarial drug inhibits viral entry into the host cell. Chloroquine interferes with the glycosylation of S protein [263]. SARS virus enter into the host cell by endocytosis also and thus SARS infection can also be prevented by inhibiting the regulators of endocytosis like AP-2 associated protein kinase 1 (AAK1). Richardson et al. suggested through computational studies that Baricitinib which is one of the best AAK1 inhibitors can also be used for the treatment of COVID-19. It also binds to G-associated kinase which also regulates endocytosis [264]. Some studies target the SARS-CoV-2 main protease (CoV M\(^{\text{pro}}\)). CoV M\(^{\text{pro}}\) cleaves poly-proteins required for the replication and transcription of COVID-19 and inhibition of CoV M\(^{\text{pro}}\) will prevent viral multiplication. Computer aided drug designing and bioinformatics tools are used for designing the inhibitors for CoV M\(^{\text{pro}}\). Jin et al. designed inhibitor M3 which inhibits the activity CoV M\(^{\text{pro}}\) by fitting into the substrate binding
patients have altered GM during hospitalization. In one of the most important functions of the GM is that it promotes immune system development. A study reported that germ free mice have reduced number of CD4+ T cells in the lamina propria and IgA producing B cells [269]. They have smaller number of payer's patches with reduced size, poorly developed B and T cells zones in spleen [270]. Germ free mice also possessed imbalanced Th1/Th2 response [271]. GM is mainly dominated by *Firmicutes* and *Bacteroidetes* phyla and constitutes over 90% of the total bacterial species localized in gut. Other microbial phyla like *Proteobacteria, Actinobacteria, Fusobacteria* and *Verrucomicrobiae* are present in smaller proportions [272]. GM regulates various body functions. But alterations in GM composition can cause various gastrointestinal, metabolic, respiratory, autoimmune, psychological and liver diseases [273]. Alternation is GM composition is known as dysbiosis. A study reported that excessive intake of medications; antibiotics and irradiation altered the GM [274]. As altered GM leads to several diseases it might also be possible that GM gets altered during SARS-CoV-2 infection. Several studies showed that elderly people and immunocompromised people having cardiovascular diseases, hypertension and diabetes are at much higher risk of COVID-19. Dysbiosis is also one of the reasons for cardiovascular diseases, diabetes and hypertension and elderly people also have altered GM [275]. This further suggests the role of dysbiosis in COVID-19. Dysbiosis leads to dysregulated immune response and promote inflammatory response. Dysbiosis induced enhanced production of Th17 cells and proinflammatory cytokines like TNF-α, IL-1β and IL-6. Several studies revealed that dysbiosis lead to development of various inflammatory diseases like IBD [276], colorectal cancer [277] and RA [278]. Thus, it might be possible that altered GM is behind the inflammatory response reported in COVID-19.

A recent study published has shown that COVID-19 patients have altered GM during hospitalization. In one of the studied COVID-19 patient, GM was dominated by pathogens and was depleted of beneficial microbes. It was observed that there was positive correlation between the abundance of *Clostridium ramosum*, *Clostridium hathewayi* and *Coprobacillus* and COVID-19 severity. Abundance of *Facalibacterium prausnitzii* was negatively correlated with COVID-19 severity [279]. This study thus has shown the role of GM in COVID-19, but further research is required to establish the relation between dysbiosis and SARS-CoV-2 infection.

It is also reported that like gut, lung also has its own microbiota denoted as lung microbiota. Lung microbiota is dominated by *Firmicutes*, *Bacteroidetes* and *Proteobacteria*. Surprisingly, it has been revealed that GM regulates pulmonary health and there is effect of lung inflammation on GM also. Thus, GM and lung interconnect with each other and affect each other's development. This bidirectional interconnection between the GM and lung is referred as gut-lung axis. Gut-lung axis increases the likelihood that SARS-CoV-2 infection might affect the GM. It was also observed that GM also has role in ARDS which further increase the possibility that GM has role in COVID-19 [280]. Interestingly, a study has reported that GM colonization in germ free mice reduced the expression of colonic ACE2 [281]. This again confirms the role of GM in SARS-CoV-2 pathogenesis.

To further test the role of GM in COVID-19, a clinical trial (NCT04355741) is under progress. The aim of the trial is to test if there is any difference of GM composition between the patients those having mild disease and those having severe disease. The study will assess is there any relation between the GM composition and the rate of mortality, length of stay in hospitals and the duration of mechanical ventilation. The outcome of this study will explain further the role of GM in COVID-19. There should be more studies performed to find out if there is any correlation of dysbiosis found in patients having cardiovascular diseases, hypertension and diabetes and their increased chances of getting infected with SARS-CoV-2.

### 8. Gut microbiota and COVID-19

Microbiota is the collection of microbes like bacteria, protozoa, archaea, fungi, viruses and eukaryotes that colonize in or on the host [267]. The aggregate of microbes that reside in the gastrointestinal tract is termed as gut microbiota (GM). Human GM consists of over 100 trillion microbes encoding more than three million genes. GM is considered as an essential organ of our body [268]. GM regulates various biological processes in our body like they increase nutrient absorption, enhance extraction of energy from food and maintain integrity of gut. One of the most important functions of the GM is that it performs immune system development. A study reported that germ free mice have reduced number of CD4+ T cells in the lamina propria and IgA producing B cells [269]. They have smaller number of payer's patches with reduced size, poorly developed B and T cells zones in spleen [270]. Germ free mice also possessed imbalanced Th1/Th2 response [271]. GM is mainly dominated by *Firmicutes* and *Bacteroidetes* phyla and constitutes over 90% of the total bacterial species localized in gut. Other microbial phyla like *Proteobacteria, Actinobacteria, Fusobacteria* and *Verrucomicrobiae* are present in smaller proportions [272]. GM regulates various body functions. But alterations in GM composition can cause various gastrointestinal, metabolic, respiratory, autoimmune, psychological and liver diseases [273]. Alternation is GM composition is known as dysbiosis. A study reported that excessive intake of medications; antibiotics and irradiation altered the GM [274]. As altered GM leads to several diseases it might also be possible that GM gets altered during SARS-CoV-2 infection. Several studies showed that elderly people and immunocompromised people having cardiovascular diseases, hypertension and diabetes are at much higher risk of COVID-19. Dysbiosis is also one of the reasons for cardiovascular diseases, diabetes and hypertension and elderly people also have altered GM [275]. This further suggests the role of dysbiosis in COVID-19. Dysbiosis leads to dysregulated immune response and promote inflammatory response. Dysbiosis induced enhanced production of Th17 cells and proinflammatory cytokines like TNF-α, IL-1β and IL-6. Several studies revealed that dysbiosis lead to development of various inflammatory diseases like IBD [276], colorectal cancer [277] and RA [278]. Thus, it might be possible that altered GM is behind the inflammatory response reported in COVID-19.

A recent study published has shown that COVID-19 patients have altered GM during hospitalization. In one of the studied COVID-19 patient, GM was dominated by pathogens and was depleted of beneficial microbes. It was observed that there was positive correlation between the abundance of *Clostridium ramosum*, *Clostridium hathewayi* and *Coprobacillus* and COVID-19 severity. Abundance of *Facalibacterium prausnitzii* was negatively correlated with COVID-19 severity [279]. This study thus has shown the role of GM in COVID-19, but further research is required to establish the relation between dysbiosis and SARS-CoV-2 infection.

It is also reported that like gut, lung also has its own microbiota denoted as lung microbiota. Lung microbiota is dominated by *Firmicutes*, *Bacteroidetes* and *Proteobacteria*. Surprisingly, it has been revealed that GM regulates pulmonary health and there is effect of lung inflammation on GM also. Thus, GM and lung interconnect with each other and affect each other's development. This bidirectional interconnection between the GM and lung is referred as gut-lung axis. Gut-lung axis increases the likelihood that SARS-CoV-2 infection might affect the GM. It was also observed that GM also has role in ARDS which further increase the possibility that GM has role in COVID-19 [280]. Interestingly, a study has reported that GM colonization in germ free mice reduced the expression of colonic ACE2 [281]. This again confirms the role of GM in SARS-CoV-2 pathogenesis.

To further test the role of GM in COVID-19, a clinical trial (NCT04355741) is under progress. The aim of the trial is to test if there is any difference of GM composition between the patients those having mild disease and those having severe disease. The study will assess is there any relation between the GM composition and the rate of mortality, length of stay in hospitals and the duration of mechanical ventilation. The outcome of this study will explain further the role of GM in COVID-19. There should be more studies performed to find out if there is any correlation of dysbiosis found in patients having cardiovascular diseases, hypertension and diabetes and their increased chances of getting infected with SARS-CoV-2.


Probiotics which are defined as “live organisms that when administered in adequate amounts confer health benefits on the host.” Probiotics are used as therapeutics for the treatment of various gastrointestinal and metabolic disorders like IBD, irritable bowel syndrome, antibiotic associated diarrhoea, obesity, osteoporosis and diabetes [282, 283]. Probiotics modulate the GM, regulate immune system and maintain the gut integrity [284–286]. Probiotics modify GM and suppress inflammation by modulating the immune system. We propose here that probiotics can
also be an effective therapy for the treatment and management of COVID-19. China’s National Administration of Traditional Chinese Medicine and China’s National Health Commission also proposed the use of probiotics for the treatment of SARS-CoV-2 infection. Administration of probiotics- Enterococcus faecalis, Lactobacillus rhamnosus GG and Bacillus subtilis to the severely ill pneumonia patients (on ventilator support) significantly reduced the need of ventilators [287]. Thus, probiotics can be used to reduce the demand of mechanical ventilation in COVID-19 patients. In China it was observed that 2–36% of the COVID-19 patients who were given antibiotics had diarrhoea [287]. Probiotics replace the harmful bacteria in gut and prevent gut dysbiosis in several diseases like IBD, colorectal cancer, diabetes and osteoporosis. Thus, probiotics can also be used for preventing dysbiosis in COVID-19 [279].

Another mechanism through which probiotics can prevent SARS-CoV-2 infection is by modulating the immune system. Probiotics showed anti-inflammatory properties [284, 285]. Probiotic combination of L. rhamnosus, B. lactis and B. longum decreased the production of IL-β and IL-6 and unregulated the expression of IL-10 [288]. L. fermentum and L. salivarius downregulated the expression of inflammatory cytokine IL-1β in DSS induced colitis mouse model [289]. Probiotics mixture VSL#3 decreased the level of TNF-α in peripheral blood and prevented sickness behavior [290]. A study reported that L. rhamnosus suppressed the production of TNF-α from macrophages [291]. Many other studies proved the role of probiotics in preventing inflammation [292, 293]. Thus, the anti-inflammatory properties of probiotics and their role in preventing dysbiosis should be exploited for preventing SARS-CoV-2 infection. More future research should be done on the pathogenesis and efficacy of probiotics in humans. Probiotics role in COVID-19 should be explored because probiotics can provide a cheap and effective way of treating and managing novel CoV 2019.

10. Conclusion

SARS-CoV-2 enforced public health crisis with an unprecedented challenge for the development of successful therapeutics. SARS-CoV-2 infection impairs the immune system leading to various inflammatory responses and in severe cases death, but till now there is no specific therapy available for COVID-19. Thus, there is an exigent need to identify a potential therapeutic compound that can be employed against COVID-19. This review summarizes various mechanisms of SARS-CoV-2 immunopathogenesis and available therapeutic interventions for the cure of COVID-19. Control of inflammatory response is vital for targeting the viral infection, and therefore, it is imperative that the mechanisms behind the hyperinflammation are further elucidated to design better therapeutic strategy to restrict the viral spread. In the last section, we also summarized the hidden potential of probiotics that can provide a safe and inexpensive option in the treatment and management of COVID-19.

Acknowledgements

AB, LS, CS, ZA, BV, PKM, GCM and RKS acknowledge the Department of Biotechnology, AIIMS, New Delhi, India for providing infrastructural facilities. ZA thanks Dr. Harisingh Gour Central University, Sagar (MP). All figures (except figure 7) are created with the help of https://smart.servier.com/.

Authors’ contributions

RKS contributed in conceptualization and writing of the manuscript. AB, LS, CS, ZA, PKM BV, GCM participated in writing and editing of the review. RKS suggested and AB and LS created the illustrations.

Conflicts of interest

The authors declare no conflicts of interest.

Funding

This work was financially supported by projects: DST-SERB (EMR/2016/007158), Govt. of India, Intramural project from All India Institute of Medical Sciences (AIIMS, A-596), New Delhi-India and AIIMS-IITD (AI-15) collaborative project sanctioned to RKS. AB thank DST SERB project for research fellowship, LS and ZA thank UGC for research fellowship.

References


37. Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is


251. Baum A, Fulton BO, Wlopa E, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational es-


