Research paper

Recognition of plausible therapeutic agents to combat COVID-19: An omics data based combined approach

Mohammad Uzzal Hossain [a,1], Arittra Bhattacharjee [a,b,1], Md. Tabassum Hossain Emon [c,1], Zeshan Mahmud Chowdhury [b,1], Md. Golam Mosaib [d], Muntahi Mourin [b,f], Keshob Chandra Das [e], Chaman Ara Keya [b], Md. Salimullah [e,*]

[a] Bioinformatics Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh
[b] Department of Biochemistry and Microbiology, North South University, Bashundhara, Dhaka 1229, Bangladesh
[c] Department of Biotechnology and Genetic Engineering, Life Science Faculty, Mawlana Bhashani Science and Technology University, Santosh, Tangail 1902, Bangladesh
[d] Department of Biochemistry and Molecular Biology, Faculty of Health & Medical Sciences, Gono Bishwabidyaloy, Ashulia, Savar, Dhaka 1344, Bangladesh
[e] Molecular Biotechnology Division, National Institute of Biotechnology, Ganakbari, Ashulia, Savar, Dhaka 1349, Bangladesh
[f] Department of Microbiology, University of Manitoba, 66 Chancellors Cir, Winnipeg, MB R3T 2N2, Canada

ARTICLE INFO

ABSTRACT

Coronavirus disease-2019 (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), has become an immense threat to global public health. In this study, we performed complete genome sequencing of a SARS-CoV-2 isolate. More than 67,000 genome sequences were further inspected from Global Initiative on Sharing All Influenza Data (GISAID). Using several *in silico* techniques, we proposed prospective therapeutics against this virus. Through meticulous analysis, several conserved and therapeutically suitable regions of SARS-CoV-2 such as RNA-dependent RNA polymerase (RdRp), Spike (S) and Membrane glycoprotein (M) coding genes were selected. Both S and M were chosen for the development of a chimeric vaccine that can generate memory B and T cells. siRNAs were also designed for S and M gene silencing. Moreover, six new drug candidates were suggested that might inhibit the activity of RdRp. Since SARS-CoV-2 and SARS-CoV-1 have 82.30% sequence identity, a Gene Expression Omnibus (GEO) dataset of Severe Acute Respiratory Syndrome (SARS) patients were analyzed. In this analysis, 13 immunoregulatory genes were found that can be used to develop type 1 interferon (IFN) based therapy. The proposed vaccine, siRNAs, drugs and IFN based analysis of this study will accelerate the development of new treatments.

1. Introduction

Coronavirus Disease-2019 (COVID-19) has drastically spread throughout the world and become a massive threat to humanity (WHO, 2019). Most healthcare systems and governing bodies of developed countries are struggling to control the spread of this pandemic (Lai et al., 2020). Quick preventative actions are challenging in curbing its spread since the causative agent of this disease is a novel SARS-related Coronavirus (CoV) or SARS-CoV-2. This particular strain of the virus was unknown before the outbreak that took place in Wuhan, China; therefore, its ambiguous pattern of transmission and pathogenicity is an ongoing and unrelenting threat to public health and the global economy (Lai et al., 2020; Wu et al., 2020).

CoVs, members of the *Coronaviridae* family and the *Coronavirinae* subfamily were considered insignificant pathogens for humans before the outbreak of SARS occurred in 2002 and 2003, which originated in the Guangdong province, China (Fung and Liu, 2019). *Coronavirinae* has been divided into four categories: *Alpha*; *Beta*, *Gamma* and *Deltacoronavirus* (Fung and Liu, 2019). These pathogens were described as the "virology backwater" since very little was known about their virulence (Cavanagh et al., 2005). In the past couple of decades, they have caught the attention of the scientific community several times due to their zoonotic characteristics. The *Betacoronavirus* group especially consists of the riskiest Human CoVs (HCoVs), such as HCoV-OC43, HCoV-HKU1, SARS-CoV-1 and Middle East respiratory syndrome coronavirus (MERS-CoV) (Cui et al., 2019).

Within itself, *coronavirinae* has vast genetic diversity, but these have

---

Nomenclature

COVID-19   Coronavirus disease-2019
SARS-CoV-2 Severe Acute Respiratory Syndrome-Coronavirus-2
RdRp   RNA-dependent RNA polymerase
S   Spike
M   Membrane glycoproteins
ACE2   Angiotensin-Converting Enzyme 2
TMPRSS2 Transmembrane Serine Protease 2
GCSF   Granulocyte-colony stimulating factor
MCP1   Monocyte chemoattractant protein 1
MIP   Macrophage Inflammatory Proteins
IP-10   Interferon gamma-induced protein- 10
TNFα   Tumor necrosis factor alpha

mostly been found in bats (Lau et al., 2019). This mammalian host poses a serious threat to CoV outbreaks (Fan et al., 2019). A possible reason behind this occurrence is that members of *Coronavirinae* frequently undergo RNA recombination inside bats and often generate novel CoVs (Cui et al., 2019; Lau et al., 2019). SARS-CoV-2, a member of *Betacoronavirus*, most probably originated in bats and transmitted to humans via intermediary animals in Wuhan, China. This transmission started the COVID-19 pandemic around December 2019 (Wu and McGoogan, 2020).

SARS-CoV-2, a positive-stranded (+) RNA virus, spreads from an infected person to a healthy individual via respiratory droplets as a result of sneezing and/or coughing (Wu et al., 2020). The presence of SARS-CoV-2 is also observed in the stool; thus contaminated water supply might promote transmission (Tang et al., 2020). The virions pass through the upper respiratory tract of the host, then enter the lung cells by engaging Angiotensin-Converting Enzyme 2 (ACE2) and Transmembrane Serine Protease 2 (TMPRSS2) (Hoffmann et al., 2020). This entry triggers innate immunity; as a result, the following inflammatory cytokines become elevated: Interleukin (IL) – 2, IL-10, IL-7, Granulocyte-colony stimulating factor (GCSF), monocyte chemoattractant protein 1 (MCP1), Macrophage Inflammatory Proteins (MIP) 1A, Interferon gamma-induced protein- 10 (IP-10) and tumor necrosis factor-alpha (TNFα) (Wu and McGoogan, 2020). Consequently, patients develop indistinguishable flu-like symptoms such as cough, sore throat, headache, fatigue, myalgia, breathlessness, conjunctivitis and/or fever. In the worst cases, patients develop pneumonia leading to respiratory failure and death (Chen et al., 2020). To alleviate these symptoms, physicians are non-specifically applying antiparasitics such as Ivermectin and the antiviral drug Remdisivir, along with other antiviral therapeutics. Plasma therapy is also helping many patients to survive (Wu and McGoogan, 2020; Gao et al., 2020; Cascella et al., 2020; Caly et al., 2020; Eastman et al., 2020).

A recent *in vitro* study revealed a potential clinically proven protease inhibitor that can block TMPRSS2, and consequently the entry of SARS-CoV-2 (Hoffmann et al., 2020). Conducting similar studies will help to reveal a variety of possible therapeutics and druggable targets, which increases the probability of creating successful treatments against COVID-19. *In silico* approaches are the most suitable way to execute this type of research in a short time with nearly zero cost. Powerful computational methods such as subtractive proteomics, comparative genomics or virtual screening can determine therapeutic targets and help develop drugs against bacteria, fungi or non-communicable diseases (Hossain et al., 2018; Abadio et al., 2011; Bhattacharjee et al., 2020). Unfortunately, for viral infections, this type of combined rational approach is hardly seen. As a result, unveiling the potential targets, novel drugs, interferons (IFNs) and immunodominant vaccines against COVID-19 will be extremely challenging if the virus becomes resistant against different therapeutics.

In this study, we propose a combined *in silico* approach to recognize plausible therapeutics. Here, we have sequenced one SARS-CoV-2 isolate from Bangladesh and identified the regions that are relatively conserved in more than 67,000 isolates. Several computational techniques were then merged to develop treatments against those conserved regions. Using these techniques, a chimeric vaccine (which is not dependent only on spike protein), small druggable molecules, and Small Interfering RNAs (siRNAs) were developed. Additionally, a comparative genome analysis was carried out between SARS-CoV-2, SARS-CoV-1 and MERS-CoV to detect the genetic similarity. SARS-CoV-1 showed close genetic relation and virulence pattern with SARS-CoV-2. Thus, a transcriptomic analysis of this related virus was performed to retrieve the IFN inducing genes to implement IFN based therapy.

2. Materials and methods

2.1. Sample collection and processing

The viral isolate (SARS-CoV-2/human/BGD/NIB-BCSIR_02/2020) was collected from a 45-year-old female patient who was tested positive for COVID-19 by Reverse-Transcriptase Polymeric Chain Reaction (RT-PCR) along with the symptoms such as cough, mild fever, and throat congestion. The oropharyngeal swab of the patient was taken as a specimen on the 11th of May 2020. From there, the viral RNA was extracted using PureLink Viral RNA/DNA Mini Kit (Invitrogen). The RNA was then converted into cDNA using SuperScript™ VILO™ cDNA Synthesis Kit (Invitrogen).

2.2. Sequencing of the viral cDNA

The viral cDNA was sequenced with Illumina Nextseq 550 next-generation sequencing technology. Here, the library preparation kit of Nextera DNA Flex was used for the synthesis of the nucleotides. The reagent cartridge was NextSeq High Output Kit that covers 300 cycles. The run mode was a local run manager that generated the FASTQ data workflow from the NextSeq 4-channel chemistry. Quality estimation and analysis of the sequences were performed using a customized version of the DRAGEN RNA pipeline, which is also available on local DRAGEN server hardware. The Illumina® DRAGEN RNA Pathogen Detection App uses a combined human and virus reference to explore pathogen data. Here, the low-quality raw reads were trimmed with Trimmomatic 0.36 (-phred33, LEADING: 20, TRAILING: 20, SLID) (Bolger et al., 2014). The assembly was performed by SPAdes V. 03 using default parameters as well as used to cross-validate with the reference-based method as an internal control (Bankevich et al., 2012). A total of 1, 78, 22,898 reads were produced in the reference-based alignment. After trimming, 99% of them were mapped to the SARS-CoV-2 reference genome. The assembled genome was submitted in National Center for Biotechnology Information (NCBI) GenBank (Accession number: MT568643.1) and Global Initiative on Sharing All Influenza Data (GISAID) (Accession ID: EPI_ISL_458133) (https://www.gisaid.org/).

2.3. Phylogenetic tree construction and identification of conserved region

The phylogenetic tree was constructed as stated by Moniruzzaman et al., 2020 (Moniruzzaman et al., 2020). In short, the closely related genomes were aligned with MAFFT version 7 and the Phylogenetic tree was constructed using FastTree version 2.1.10 via Galaxy platform (Katoh and Standley, 2013; Price et al., 2010; Afgan et al., 2018). To identify conserved regions, SARS-CoV-2 genomes from different continents were aligned. From there relatively conserved regions that are suitable for therapeutics development were collected. The conservation (segments that are less diversified) of these regions were further observed in the GISAID database (Fig. 2) (Elbe and Buckland-Merrett, 2017). Afterward, these selected regions/ genes were obtained from our submitted genome to develop a chimeric vaccine, small molecules

and siRNAs.

2.4. Chimeric vaccine against SARS-CoV-2

2.4.1. Antigenicity analysis of the candidate proteins

The antigenicity of the selected Membrane (M) (NCBI GenBank accession: QKI36904.1) and Spike Glycoprotein (S) (NCBI GenBank accession: QKI36901.1) were determined by Vaxijen 2.0 (Doytchinova and Flower, 2007) and AntigenPro (Magnan et al., 2010). These servers predicted antigenicity through alignment independent methods.

2.4.2. Identification of B cell and T cell epitopes

The outside amino acid residues of the vaccine candidates were identified with THMM (Krogh et al., 2001). The identified sequences were uploaded in the BepiPred-2.0 server to select the most potential B cell epitopes (Jespersen et al., 2017). During this selection, surface-exposed amino acids were prioritized mostly. The Cytotoxic T cell (CTL), Helper T cell (HTL) epitopes were disclosed by NetCTL 1.2 and NetMHCII respectively (Larsen et al., 2007; Jensen et al., 2018). NetCTL 1.2 was implemented to identify the CTL epitopes for 12 classes of Major Histocompatibility Complex 1 (MHC I) supertypes. The best CTL epitopes were selected based on the combined score. NetMHCII was used to detect 15-mer HTL epitopes for Human Leukocyte Antigen (HLA)-DP, HLA-DQ, and HLA-DR alleles. Most potential epitopes were chosen by evaluating affinity, percentage ranking and binding level.

2.4.3. Construction of chimeric vaccine

To construct a chimeric or multi-epitope vaccine, all the epitopes were joined by EAAK, GPGPG and AAY linkers (Shey et al., 2019). Human Beta Defensin-2 (HBD-2) (PDB ID: 1FD3) and a recombinant viral protein were added in the N terminal and the C terminal of the vaccine respectively (Yu et al., 2010; Li et al., 2013). HBD-2 was conjugated because this protein can activate Toll-Like Receptor 4 (TLR 4) and has chemotactic activity (Yu et al., 2010; Niyonsaba et al., 2004). The recombinant viral protein was added to stimulate antiviral responses.

2.4.4. Evaluation of immune response and Interferon gamma (IFNγ)

The fasta sequence of the vaccine was uploaded to an agent-based immune system simulator C-ImmSim (http://150.146.2.1/C-IMMSIM/index.php) for measuring the immune responses (Rapin et al., 2010). The parameters were kept in default during this simulation. C-ImmSim showed an adequate secretion of IFNγ which was further evaluated by IFNepitope (Dhanda et al., 2013).

2.4.5. Allergenicity and toxicity exploration

Recombinant vaccines can initiate an allergic response or induce various types of toxicity. Therefore, evaluation of toxicity and allergenicity is an essential step for vaccine designing. Here, the allergenicity of the vaccine was calculated by AlgPred and AllerTop v.2 (Saha and Raghava, 2006; Dimitrov et al., 2014). Toxicity was measured by ToxinPred (Gupta et al., 2013).

2.4.6. Analysis of physicochemical properties and tertiary structure

The physicochemical analysis of the protein was executed via ProtParam (Gasteiger et al., 2005). The tertiary structure was generated through Contact-guided Iterative Threading ASSEmbly Refinement (C-I-TASSER) (Zheng et al., 2019). The structure was refined by GalaxyRefine and validated with Procheck and ProSAWeb (Heo et al., 2016; Laskowski et al., 1993; Wiederstein and Sippl, 2007).

2.4.7. Molecular docking analysis

The crystal structure of TLR 4 was collected from Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB) (www.rcsb.org). The PDB ID of the structure was 3FXI. For molecular docking analysis, only Chain A of TLR 4 was preserved. All

heterogeneous atoms were removed. This preparation was executed by BIOVIA Discovery Studio (https://www.3dsbiovia.com/). To determine the active site pocket of Chain A, the PDB file was uploaded in Computer Atlas of Surface Topology of Proteins (CASTp) (Tian et al., 2018). Finally, the vaccine and Chain A of TLR-4 was uploaded in ClusPro 2.0 for protein–protein docking (keeping the parameters default) (Kozakov et al., 2017).

2.4.8. Codon optimization and visualization of cloning

The vaccine sequence was reversely translated and the codons were optimized for *Escherichia coli* (strain K12) through JCat (Grote et al., 2005). The optimized DNA sequence was kept free from rho-independent transcription terminators and prokaryotic ribosome binding sites. This sequence was flanked by Nde I and Xho I restriction sites. Stop codons were added at the end of 3′OH or C terminal end. After these preparations, the DNA sequence was inserted in pET28a (+) Plasmid Vector via SnapGene tools (www.snapgene.com).

2.5. Small molecule therapeutics against COVID-19

2.5.1. Homology modelling and binding site analysis

The conserved RNA directed RNA polymerase (RdRp) sequence was retrieved from our submitted genome to develop small molecules. C-I-TASSER was employed to build the 3D model of RdRP (Zheng et al., 2019). Thereafter, CASTp was implemented to identify the drug binding sites that were allowed to recognize the critical amino acids for drug-protein interactions (Tian et al., 2018).

2.5.2. Screening the druggable compounds and molecular docking simulation

DrugBank Database (www.drugbank.ca) was utilized to search for potent drugs against RdRP. Molecular docking of the receptor and the selected compounds were performed in AutodockVina to observe the binding affinity at the drug binding site of RdRp (Trott and Olson, 2010; Agrawal et al., 2001). Here, AutoDock tools 1.5.6 was used to prepare the input pdbqt file for the receptor. A grid box parameter was set in size with X = 70, Y = 70, Z = 36 points (center grid box: x = 23.448, y = 53.587, z = -2.012, spacing = 0.5A). The molecular visualization of protein-ligands was executed by BIOVIA Discovery Studio and PyMol (Seeliger and de Groot, 2010).

2.5.3. Pharmacoinformatics illustration

Osiris property explorer, Molinspiration, ACToR (Aggregated Computational Toxicology Resource), admetSAR (absorption, distribution, metabolism, excretion, and toxicity Structure-Activity Relationship database) and ACD/I-lab were exploited for the calculation of Absorption, distribution, metabolism, excretion (ADME) properties and toxicity (T) profile (Sander, 2001; Kumar Mishra, et al., 2011; Judson et al., 2008; Cheng et al.; Masunov, 2001). ADMET properties are necessary to establish an effective drug.

2.6. Nucleic acid-based therapeutics development

2.6.1. Designing of potential siRNA molecules

To design the effective siRNAs molecule, I-Score Designer was employed (Ichihara et al., 2007). I-Score Designer computes nine different siRNA designing scores such as Ui-Tei (Ui-Tei et al., 2008), Amarzguioui (Amarzguioui and Prydz, 2004), Hsieh (Hsieh et al., 2004) Takasaki (Takasaki et al., 2004) s-Biopredsi (Huesken et al., 2005) i-Score (Ichihara et al., 2007) Reynolds (Reynolds et al., 2004) Katoh (Katoh and Suzuki, 2007) and Design of SIRna (DSIR) (Vert et al., 2006) along with other essential parameters. The server also ranks the best siRNA molecules. From there, the best siRNA sequence was taken for further analysis. The secondary structure of the siRNA was predicted via RNA structure web server (Mathews et al., 2004). RNAfold web server was applied to calculate the free energy of the thermodynamic ensemble

for the secondary structures (Mathews and Turner, 2006). Transcription and Translation Tool (http://biomodel.uah.es/en/lab/cybertory/analysis/trans.htm) was implemented to transcribe the viral DNA sequences. Finally, the designated Antisense siRNAs were hybridized against viral RNA sequences using the HNADOCK server (He et al., 2019). HNADOCK executed RNA-RNA docking and performed molecular dynamics simulations to refine the best 10 predicted siRNA-mRNA complexes. The model with the best docking score for Membrane Glycoprotein mRNA and Spike Glycoprotein mRNA was visualized with PyMol.

2.7. Interferon stimulating genes (ISGs) based Interferon therapy

2.7.1. Comparative genomics

The genome of SARS-Cov-2 was compared with SARS-CoV-1 and MERS-CoV. We Blasted SARS-CoV-2 against SARS-CoV-1 and MERS-CoV, using the Megablast and Discontiguous Megablast algorithm. The graphical representation of the side-by-side genome comparison is demonstrated in the Artemis Comparison Tool (Carver et al., 2005).

2.7.2. Exploring SARS-CoV expression profile

The Gene Expression profile (Accession Number: GSE5972) of SARS-CoV-1 was collected from NCBI. Afterward, the normalization study was performed in between 10 Healthy samples and 54 SARS patient samples, where recovered cases were excluded from this study. Limma (R package) was used to find the differentially expressed genes (Ritchie et al., 2015). Integrated Differential Expression and Pathway Analysis (IDEP) tools were utilized to create Hierarchical Clustering Heatmap and Boxplot visualization of the data for both up- and down-regulated genes (Ge et al., 2018).

2.7.3. Gene enrichment analysis

Kyoto Encyclopedia of Genes and Genomes (KEGG) and Enrichr4 were employed to identify the genes which were involved in pathways and Gene Ontology (GO) processes such as biological process (BP), molecular function (MF) and cellular component (CC) (Sherman et al., 2007; Kanehisa and Goto, 2000; Kuleshov et al., 2016; Ge et al.). The genes with GO accession ID &KEGG pathways were enlisted for further study. ShinyGO was utilized to construct a clustered tree of the top 30 GO terms for both up and down-regulated genes in BP, MF and CC (Ge et al.).

2.7.4. Identification of virus-induced genes

The down and up-regulated genes were analyzed using BP, MF and CC in gene ontology dataset to find out genes that are connected in cytokine regulation and viral production.

2.7.5. Exploration of Interferon stimulating genes (ISGs)

INTERFEROME database was analyzed to recognize the Interferon Stimulating Genes (ISGs) and potential IFNs (Samarajiwa et al., 2009). Further, we tried to explore the pathway of ISGs which modulate the immune system and IFN regulation by Reactome (Jassal et al., 2020).

3. Results

The workflow of the current study was shown in Fig. 1

3.1. Selection of conserved regions for therapeutics development

The isolated and assembled genome was 29737 bp long. According to GISAID (www.gisaid.org/) this virus is a member of L clade. Through alignment and diversity observation, 3 proteins such as S, M and RdRP were collected. These protein-coding regions of the genome are relatively conserved and therapeutically potential (Fig. 2, Supplementary Fig. 1, Supplementary Data 1, 2, 3). These proteins were further analyzed to develop vaccine, siRNAs and drugs.
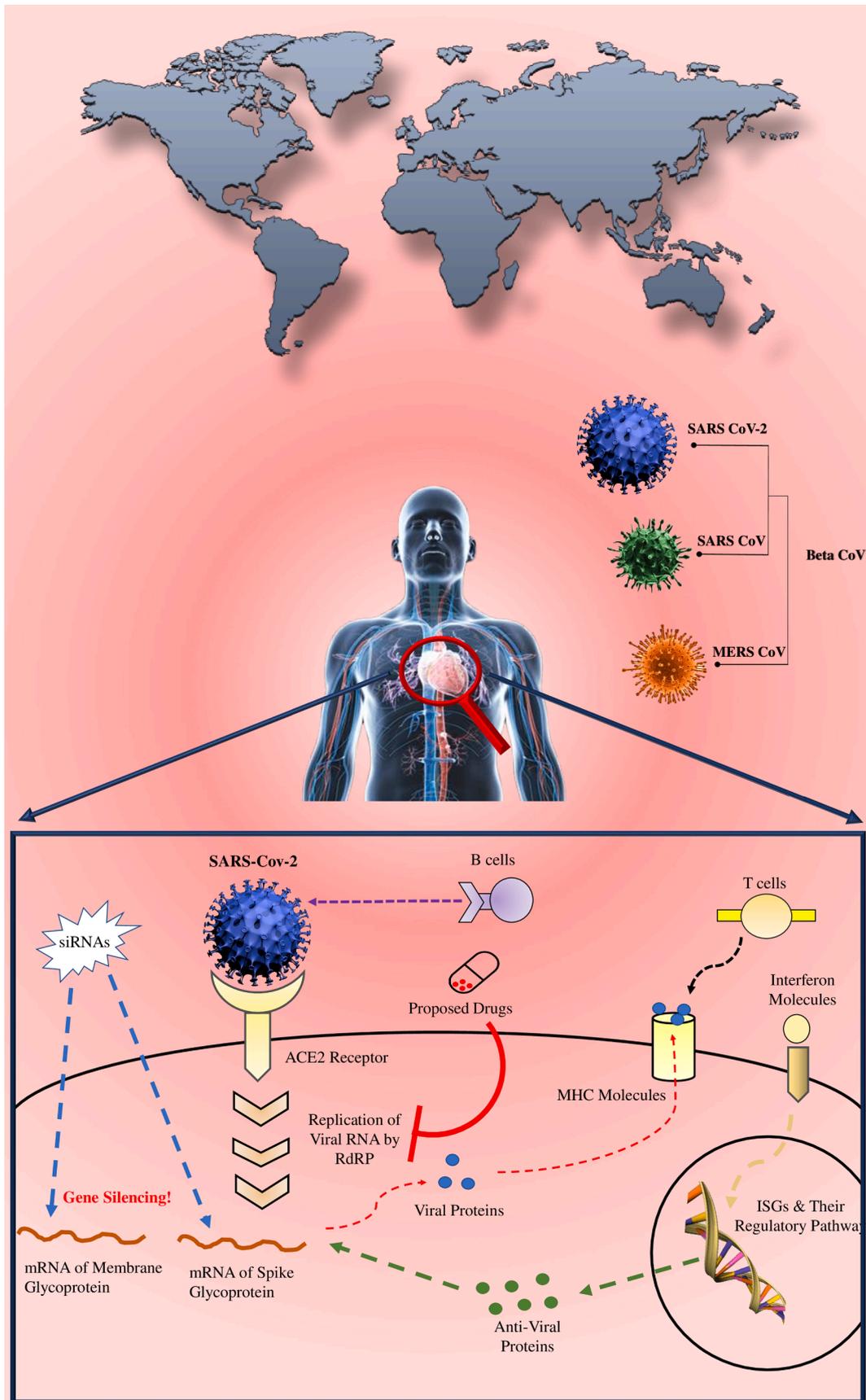
3.2. Chimeric vaccine against SARS-CoV-2

The selected M (Vaxijen score 0.55 and AntigenPro score 0.232547) and S (Vaxijen score 0.4695 and AntigenPro score 0.717053) proteins were predicted as "antigenic" by the assigned programs. The M and S have a total of 24 and 1214 outside amino acids respectively. Among those residues, one potential B cell epitope NGTITVEELKKLLEQ was identified in M (Table 1). Four possible B-cell epitopes (HAPATVCGPKKSTN, NNSYECDIP, FYEPQIITTD, and VEGFNCYFPLQ) were found in S (Table 1). The selected four Cytotoxic T-cell (CTL) epitopes from M and S cover HLA-A12, A24, A26, B58, B8 alleles. Moreover, Helper T-cell (HTL) epitopes can interact with HLA DRB1-0101, DPA10103-DPB10301, DQA10101-DQB10501 and DRB1-1501 (Supplementary Table 1). The chimeric vaccine exhibited a suitable binding with Toll-Like Receptor-4 (TLR4) (Fig. 3A). A single dose injection of the vaccine has demonstrated an adequate release of IFNγ in the C-ImmSim immune system simulator (Fig. 3B). The IFNepitope also showed 127 IFNγ inducing epitopes by different methods. Additionally, Interleukin (IL)-2 was also released significantly for lymphocyte differentiation. These differentiations are mentioned as danger signal D (Sympshon Index) (Fig. 3B). After the release of inflammatory cytokines, the immune system released anti-inflammatory IL-10 at the 2.5th day and consequently generated Immunoglobulin M (IgM) and later IgG (Fig. 3C). The generation of memory B cells was also observed (Fig. 3D). The vaccine specifically stimulated the differentiation of T helper cell-1 (Th1) with a certain level of regulatory T cell (Fig. 3E). Memory Helper T Cell was also produced within 2.5 days (Fig. 3F and 3G). The vaccine does not contain any major allergenic properties according to AlgPred and AllerTop. ToxinPred identified 343 non-toxic peptides in different combinations and only 11 peptides with toxic properties (Supplementary Table 2). The vaccine is stable according to the instability index of ProtParam (Supplementary Table 3). The estimated half-life of the protein is 30 h in mammalian reticulocytes (in vitro), >20 h in yeast (in vivo) and >10 h in Escherichia coli (in vivo) with −0.010 Grand Average of Hydropathicity (GRAVY) and 81.38 Aliphatic index (Supplementary Table 3). The refined three dimensional (3D) structure of the vaccine has a 1.905 MolProbity score (provided by GalaxyWEB), −2.62 Z-Score and 87.9% residues in the most favored regions of the Ramachandran Plot (Supplementary Figs. 2 and 3). When this protein was docked against TLR 4 (Chain A) via ClusPro 2.0, it interacted with the receptor via active site pocket and HBD-2 with the lowest energy score −1401.1.

#### 3.2.1. Codon optimization and cloning for chimeric vaccine

The DNA sequence of the vaccine has 1089 nucleotides that demonstrated 0.96 Codon Adaptation Index (CAI) with 54.26% GC content. Here, the GC content of the host organism E. coli strain K12 is 50.73 (Fig. 4). This sequence was taken under modifications for cloning.

3.3. Small molecule drugs against COVID-19

Six experimental compounds were collected from the DrugBank. The chemical properties of the compounds were suitable for the drug candidates to work against RdRp (Supplementary Table 4). RdRp and those six compounds went under molecular docking analysis to identify the binding affinities. Strong binding affinity was observed for benzyl (2-oxopropyl)carbamate (-4.2 Kcal/mol) (Fig. 5A), 2-[(2,4-dichloro-5-methylphenyl)sulfonyl]-1,3-dinitro-5-(trifluoromethyl)benzene (-7.6 Kcal/mol) (Fig. 5B), s-[5-(trifluoromethyl)-4h-1,2,4-triazol-3-yl]5-(phenylethynyl)furan-2-carbothioate (-5.9 Kcal/mol) (Fig. 5C), 5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide (-6.3 Kcal/mol) (Fig. 5D), Nalpha-[(benzyloxy)carbonyl]-n-[(1r)-4-hydroxy-1-methyl-2-oxobutyl]-l-phenylalaninamide (-7.4 Kcal/mol) (Fig. 5E), 4-(Dimethylamino)benzoic acid (-4.0 Kcal/mol) (Fig. 5F). Fifteen common amino acid residues including Val473, Glu474, Asp477, Lys478, Leu636, Val637, Arg735, Asp736, Val737, Asp304, Arg305, Arg640, His642, Arg651 and Leu648 of RdRp interacted with the druggable

Fig. 1. Overview of the possible therapeutics mechanism against the pandemic-COVID-19. Here, the designed chimeric vaccine could prompt the immune response against the SARS-CoV-2 by targeting membrane and spike glycoprotein. Proposed drug compounds can inhibit the replication of viral RNA by targeting RNA dependent RNA polymerase (RdRp) protein. The designed siRNAs can silence the activity of both membrane and spike glycoproteins. Moreover, the plausible Interferon stimulating genes (ISGs) can be targeted to induce the interferons against the viral proteins of SARS-CoV-2.
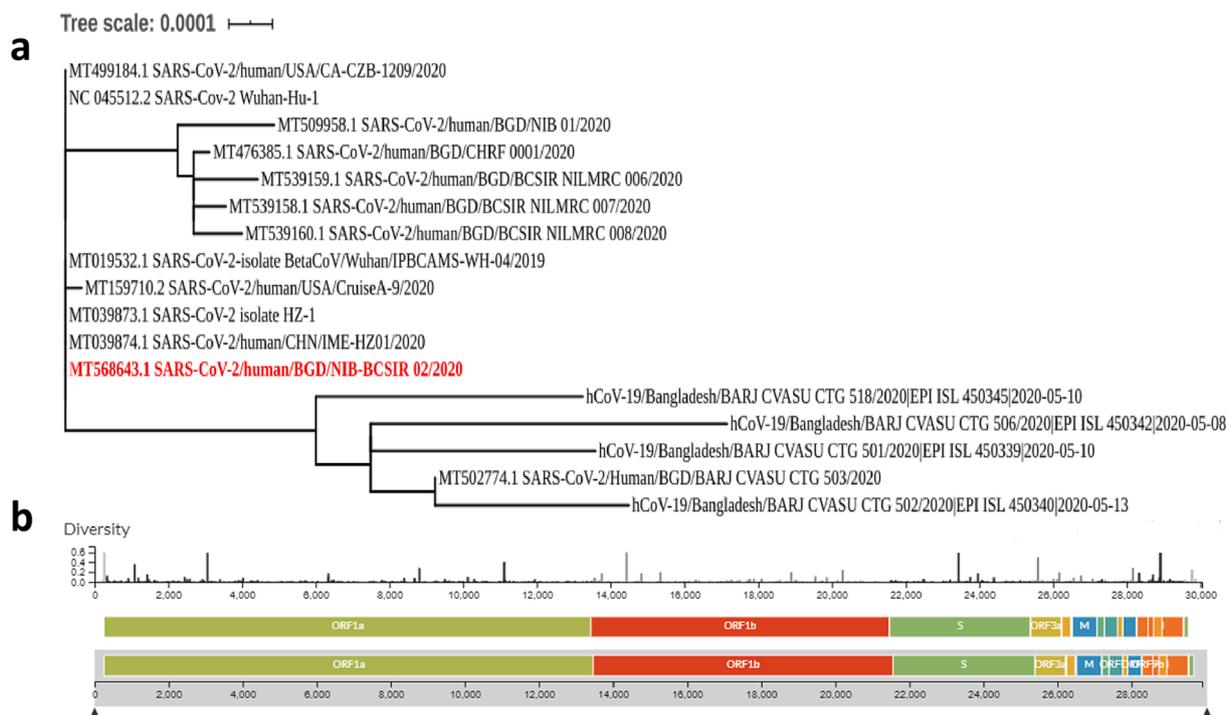
Fig. 2. Phylogenetic tree for the isolated viral genome and the conserved regions in the SARS-CoV-2 genome. (a) The position of the isolated genome is labeled with red color. (b)Genetic diversity or conservation in different genes of the SARS-CoV-2 virus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
The potential B cell, Cytotoxic T cell and Helper T cell epitopes.

Type of Epitope	Serial No.	Epitope Sequence	Protein Name
B Cell epitopes	1	NGTITVEELKKLLEQ	Membrane Glycoprotein
	2	HAPATVCGPKKSTN	Spike Glycoprotein
	3	NNSYECDIP	Spike Glycoprotein
	4	FYEPQIITTD	Spike Glycoprotein
	5	VEGFNCYFPLQ	Spike Glycoprotein
CTL Epitopes	6	SSDNIALLV	Membrane Glycoprotein
	7	YFIASFRLF	Membrane Glycoprotein
	8	YIIKLIFLW	Membrane Glycoprotein
	9	LAAVYRINW	Membrane Glycoprotein
	10	LTDEMIAQY	Spike Glycoprotein
	11	FVFKNIDGY	Spike Glycoprotein
	12	YLQPRTFLL	Spike Glycoprotein
	13	RSFIEDLLF	Spike Glycoprotein
HTL Epitopes	14	TLSYYKLGASQRVAG	Membrane Glycoprotein
	15	ASFRLFARTRSMWSF	Membrane Glycoprotein
	16	SNLLLQYGSFCTQLN	Spike Glycoprotein
	17	LIRAAEIRASANLAA	Spike Glycoprotein
	18	WFVTQRNFYEPQIIT	Spike Glycoprotein

molecules. Later, ADMET analysis of the experimental compounds were enlisted (Supplementary Table 5).

3.4. Silencing the glycoproteins of SARS-CoV-2

The Sense and Antisense RNA molecules of both M and S were 19mer and 21mer respectively (Supplementary Tables 6 and 7). The siRNA generated a hairpin-like secondary structure at 310.15 K temperature. The free energy of the thermodynamic ensembles was −18.30 kcal/mol and −17.90 kcal/mol respectively for M and S siRNA secondary structures. Results of HNADOCK server demonstrate that the antisense siR-NAs can interact with the three dimensional (3D) viral mRNA molecules with around −260 docking score (Fig. 6).

### 3.5. Interferon based therapeutic for SARS-CoV-2

The whole genome of SARS-CoV-1 and MERS-CoV were compared with SARS-CoV-2. SARS-CoV-1 and SARS-CoV-2 share 88% query sequence coverage with an 82.30% sequence identity. Besides, MERS-CoV shares only 29% query coverage with a 67.06% sequence identity (Supplementary Fig. 4). Genome alignment showed higher homology between SARS-CoV-1 and SARS-CoV-2 with a very few synteny break region. Whereas, MERS-CoV shares low scoring homology with SARS-CoV-2 with a larger synteny breakage area. The entry of SARS-CoV-2 and SARS-CoV-1 are similar so differential gene expression investigation was run on SARS-CoV-1 (Fig. 7).

64 different samples were normalized and the median was centralized (Fig. 8A). 19,200 differentially expressed genes were found (Fig. 8B and Supplementary Fig. 5). Genes with official human gene symbols (HGNC) and significant p-value (<0.05) were selected (Supplementary Data 4 and 5). 1265 genes were considered for further analysis, among them, 616 genes were up-regulated (Supplementary Data 6) and 649 genes were down-regulated (Supplementary Data 7). In the hierarchical clustering heatmap, genes based on LogFC, t-stat and adjusted p-value were clustered together (Fig. 8C, D, E and F). The genes that we have considered as up-regulated seem to have a major frequency around LogFC$\approx$ 2.1 and t value$\approx$ 2.8, which dictates significant up-regulation (Supplementary Data 8). On the other side, the down-regulated genes seem to have a major frequency around LogFC$\approx$ −1.7 and t value$\approx$ −2.4 (Supplementary Data 9). 582 up-regulated and 579 down-regulated genes were mapped through the GO and KEGG Pathway (Supplementary Data 8 and 9). For overexpression, 231 BP (Supplementary Data 10), 61 MF (Supplementary Data 11) and 44 CC (Supplementary Data 12) genes with GO accession and 77 KEGG human pathways (Supplementary Data 13) were enlisted. For under-expression, 97 BP (Supplementary Data 14), 60 MF (Supplementary Data 15) and 54 CC (Supplementary Data 16) genes with GO accession and 101 KEGG pathways (Supplementary Data 17) were enlisted as well.

The biological process of GO reported that 91 up-regulated genes and

Fig. 3. Innate and adaptive immune responses from the proposed vaccine. (A) The predicted interaction of the vaccine with Toll-Like Receptor-4 (TLR4). Protein-Protein docking analysis demonstrated that the vaccine can interact with TLR4. The constructed vaccine was injected into the immune system simulator. (B) The interaction consequently released inflammatory and anti-inflammatory cytokines. Here, the D (Simpson index) represents the diversity of T cell population. Moreover, (C) IgM and IgG containing B cells produced with (D) generation of Memory B cell (E) Population of Cytotoxic T cell spiked and (F) Memory Helper T cell produced (G) The immune response stimulated Th$_1$ cells with Regulatory T cells. The immune system produced memory B cell, memory T cell, IgM and IgG. Moreover, the induction of Th$_1$ cells and regulatory T cells were observed.
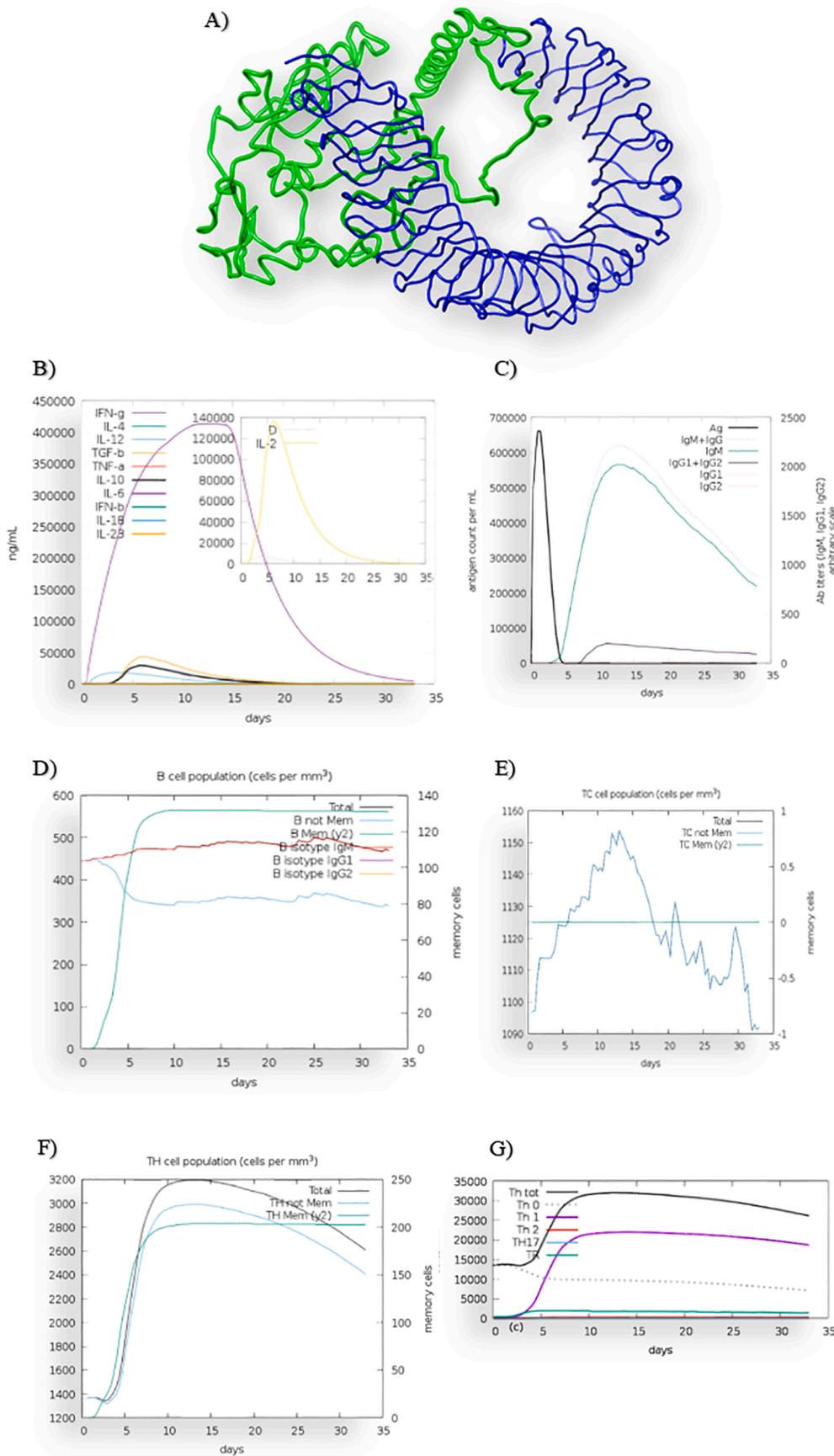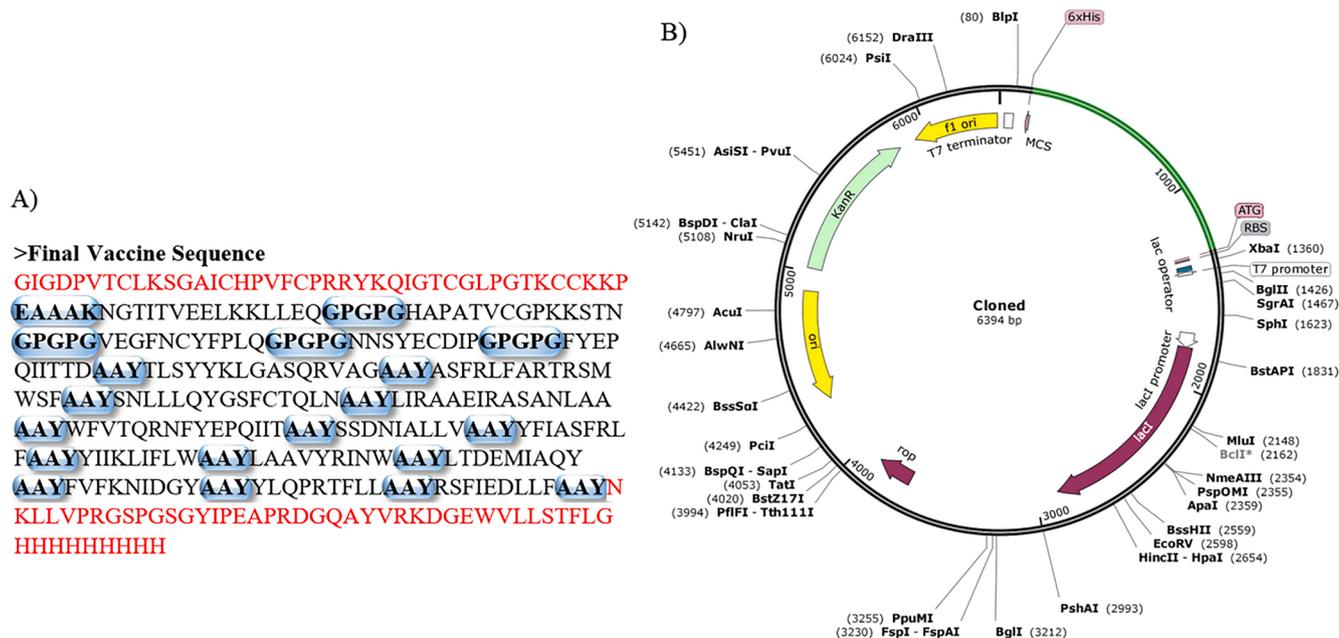
Fig. 4. Production of the vaccine can be achieved through recombinant DNA technology. (A) The primary sequence of the chimeric vaccine. The red letters represent adjuvants and the bold capsules depict the peptide linkers. The adjuvants were conjugated to elicit TLR 4 mediated Th_1 specific antiviral responses. The epitopes were joined with amino acid linkers. This sequence was reversely translated and (B) The DNA sequence of was inserted in pET-28a(+) expression vector for vaccine production in *E. coli* strain K12. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
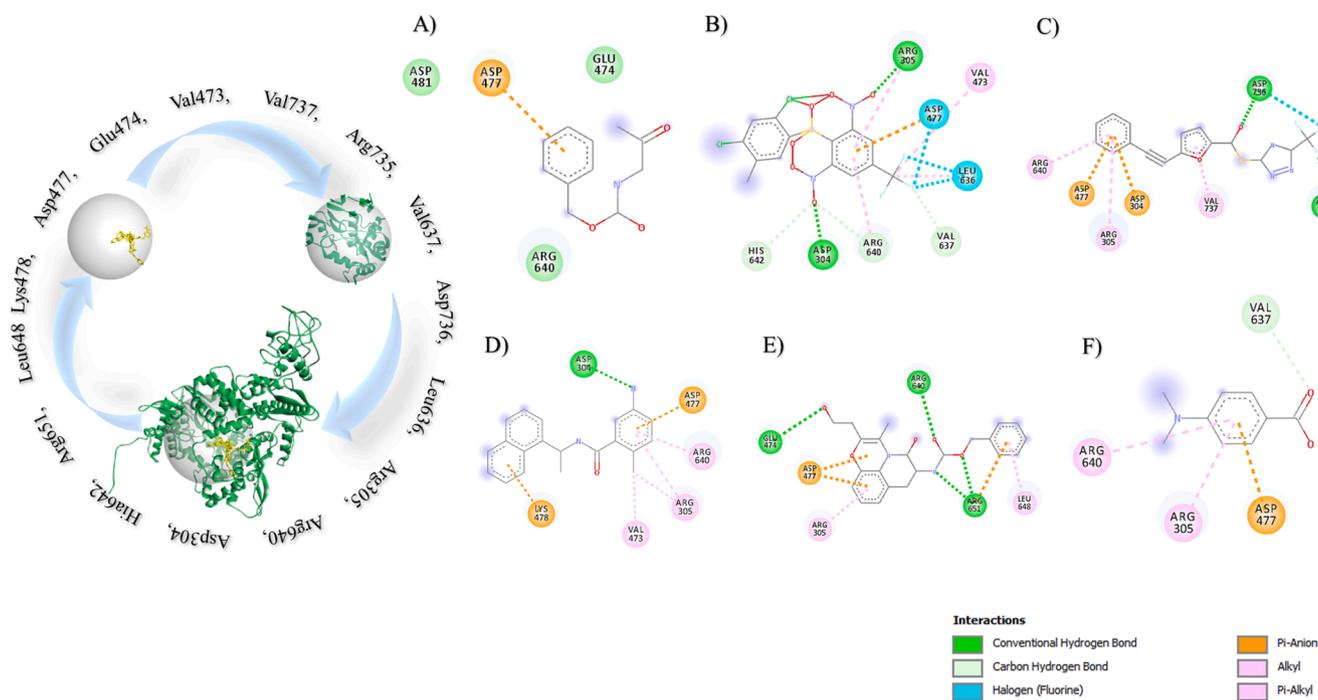
Fig. 5. Drug-receptor interactions. A) benzyl (2-oxopropyl) carbamate showed the binding energy −4.2 Kcal/mol and their interacting amino acid residues are ASP481, ASP477, GLU474, ARG640; B) 2-[(2,4-dichloro-5-methylphenyl)sulfonyl]-1,3-dinitro-5-(trifluoromethyl)benzene showed the binding energy −7.6 Kcal/mol with ASP304, ARG305, VAL473, ASP477, LEU636, VAL637, ARG640, HIS642 amino acid interactions ;C) s-[5-(trifluoromethyl)-4h-1,2,4-triazol-3-yl]5-(phenyl-ethynyl)furan-2-carbothioate docked RdRp protein with the binding energy −5.9 Kcal/mol and their interacting residues are ASP304, ARG305, ASP477, ARG735, ASP736, ARG640, VAL737; D) 5-amino-2-methyl-N-[(1R)-1-naphthalen-1-ylethyl]benzamide and RdRp protein had the docking energy about −6.3 Kcal/mol with the interacting amino acid residues ASP304, ARG305, ASP477, ARG640, VAL473, LYS478; E) Nalpha-[(benzyloxy)carbonyl]-n-[(1r)-4-hydroxy-1-methyl-2-oxobutyl]-l-phenylalaninamide had the docking enrgy-7.4 Kcal/mol and interacting residues ARG305, ASP477, GLU474, ARG640, ARG651, LEU648; F) 4-(Dime-thylamino)benzoic acid also showed the significant binding energy −4.0 Kcal/mol with ARG305, ASP477, ARG640, VAL637. The most common interacting residues are Val473, Glu474, Asp477, Lys478, Leu636, Val637, Arg735, Asp736, Val737, Asp304, Arg305, Arg640, His642, Arg651 and Leu648 and all of these docked into the binding site of the RdRp protein. Here, the interaction type has been marked with different color codes.
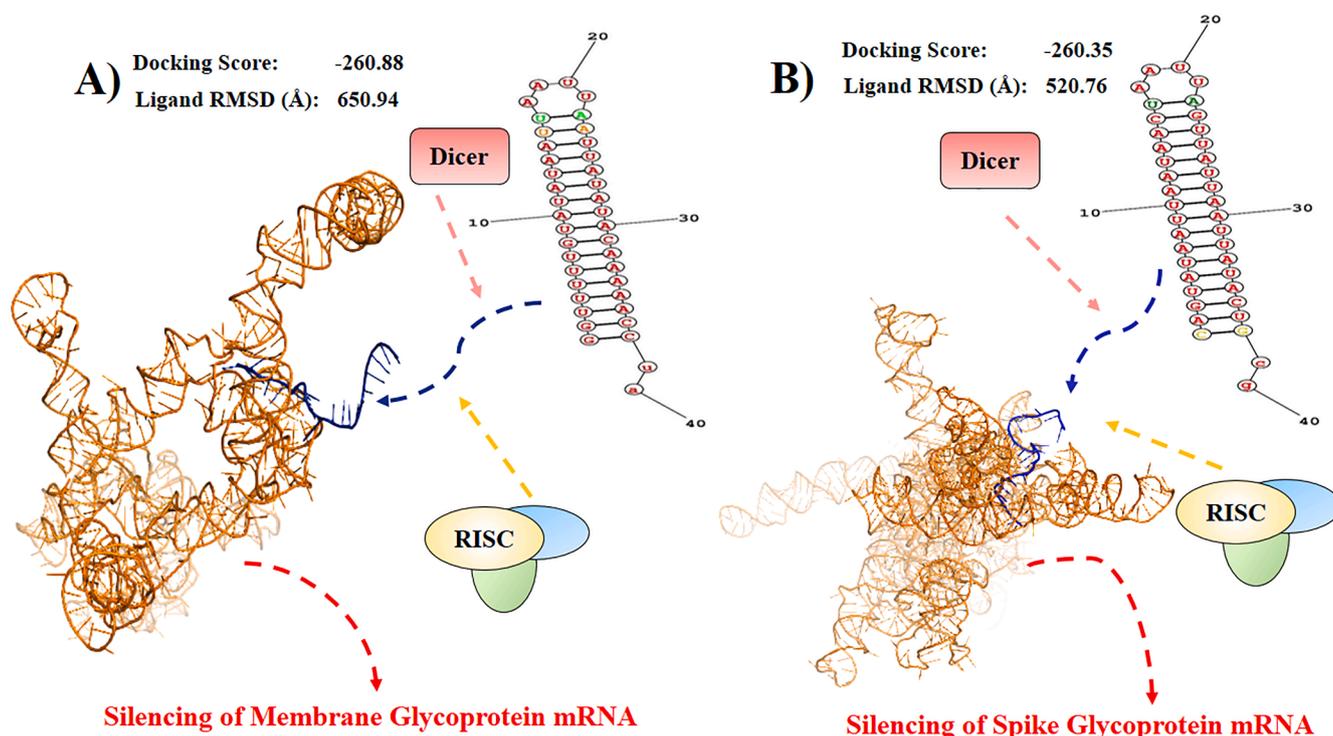
Fig. 6. Predicted pathway for siRNA based viral gene silencing. In the cytosolic environment, the Dicer and RNA-induced silencing complex (RISC) might degrade the viral mRNAs. Here, (A) the guiding antisense strand bonded with membrane glycoprotein mRNA (B) and spike glycoprotein mRNA. The guiding antisense strand is very specific to the glycoproteins to silence their activity.

76 down-regulated genes have a connection with the viral regulation process (Supplementary Data 18, 19, 20 and 21). The up-regulated (KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2) and the down-regulated (TRIM8, VCAM1, IFNGR2 and JAK1) genes were found in IFN regulated pathways (Fig. 9, Supplementary Fig. 6, and Supplementary Data 21–25).

4. Discussion

COVID-19 pandemic has destabilized the social, political and economic spheres globally (Ren et al., 2020; Berera and Zambon, 2013; Andersen et al., 2020). SARS-CoV-2, the causative agent of this pandemic, can undergo frequent mutations, adaptations and transmission, which will, unfortunately, pose a continuous risk of outbreaks (Gurwitz et al., 2020). To confront these future risks, the sustainable development of therapeutics against this deadly virus is necessary. In this study, we have exploited a large number of SARS-CoV-2 genome sequences including one from our group and merged several computational approaches to propose therapeutics that could be effective against all the strains of SARS-CoV-2 (Fig. 1). To achieve this goal of therapeutic efficacy, we have identified some therapeutic regions that have a decent level of conservation in the genome. Alongside this, whole genomes of SARS-CoV-2 from different countries were explored to find out their conserved region (Supplementary Data 1). Among all the strains, several consensus regions were obtained (Supplementary Data 2, 3 and Supplementary Fig. 1). The proposed chimeric vaccine, drug compounds and siRNA therapeutics were developed against those conserved regions (Supplementary Data 3).

Most other studies focused only on the S protein for vaccine development. This study targets both S and M proteins for vaccine and siRNA development. The vaccine candidates contain several well-conserved B and T cell epitopes (Table 1, Supplementary Table 1). The epitopes were joined to design a chimeric vaccine which is expected to elicit TLR4 mediated Th1 antiviral response. Immune simulation analysis showing that this vaccine can produce memory B and T cells (Fig. 3D & 3F). The

vaccine is expected to induce inflammatory and anti-inflammatory cytokines in sufficient amounts with non-allergen activities. Since the vaccine contains very few toxic peptides (Supplementary Table 2), *in vitro* and *in vivo* validation are required before the actual application. Again, the constructed expression vector should successfully produce adequate vaccine molecules in *E. coli* strain K12, considering that the half-life of the protein is >10 h inside the bacterial host with an acceptable stability index (Fig. 4).

A universal drug is necessary to reduce the post-infection consequences of COVID-19. For coronaviruses, RdRp is an attractive therapeutic target that catalyzes the replication of RNA from the RNA template. Thus, we targeted the conserved RNA directed RdRP protein from SARS-CoV-2 (Supplementary Data 3). The virtual screening directed by the RdRP protein identified six experimental compounds from the DrugBank server (Supplementary Table 4). Later, the molecular docking analysis revealed a higher binding affinity between the compounds and the RdRp receptor. All the interacting residues between the drug compounds and the RdRp receptor were located in the predicted drug binding site (Fig. 5). Among the ADMET properties, all the experimental compounds exhibited sufficient human intestinal absorption, a higher metabolic rate, no blood–brain barrier permeability (Supplementary Table 5). Moreover, the toxicity levels of these compounds appeared to be safe for humans (Supplementary Table 5). These characteristics are significant for drug development since poor ADMET properties could exclude a drug in phase III clinical trial.

RNA interference (RNAi) pathway is very efficient and specific in viral gene silencing, but only in post-transcription. Here, siRNAs (small interfering RNAs) can play a therapeutic role (Elbashir et al., 2001). The mRNA sequences have potential sense siRNA strands CAGUAUAAUUAAUAACUAA and GGUUUUUGUAUAUAAUUAA from S and M, respectively (Supplementary Tables 6 and 7). Therefore, the 19mer and 21mer nucleotides of the sense and antisense strands were designed based on siRNA universal rules. The thermodynamic value and strong interaction between antisense and viral mRNA strongly support its prospective application (Fig. 6A and 6B). These siRNAs might be
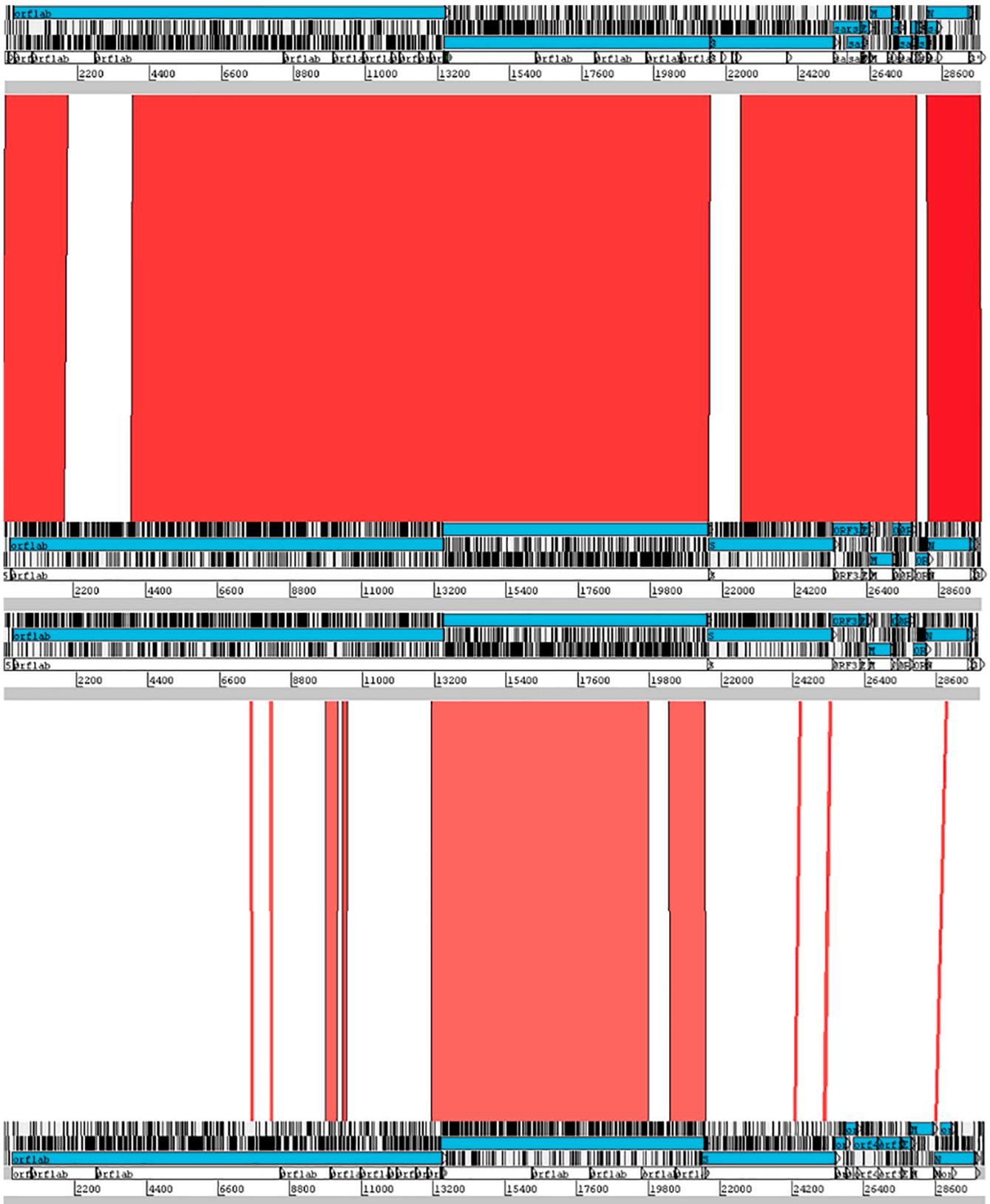
Fig. 7. The whole-genome comparison among the SARS-CoV-1, SARS-CoV-2 and MERS CoV. MERS seems to be distantly related as it showed only 29% similarity with the SARS-CoV-2. Besides, SARS-CoV shared 88% similarity with the newly evolved SARS-CoV-2 virus. The variation found only in ORF1ab, Spike and ORF8 region of SARS-CoV-2 with SARS-CoV. Alternatively, the orf1ab region from MERS is similar to SARS-CoV-2.
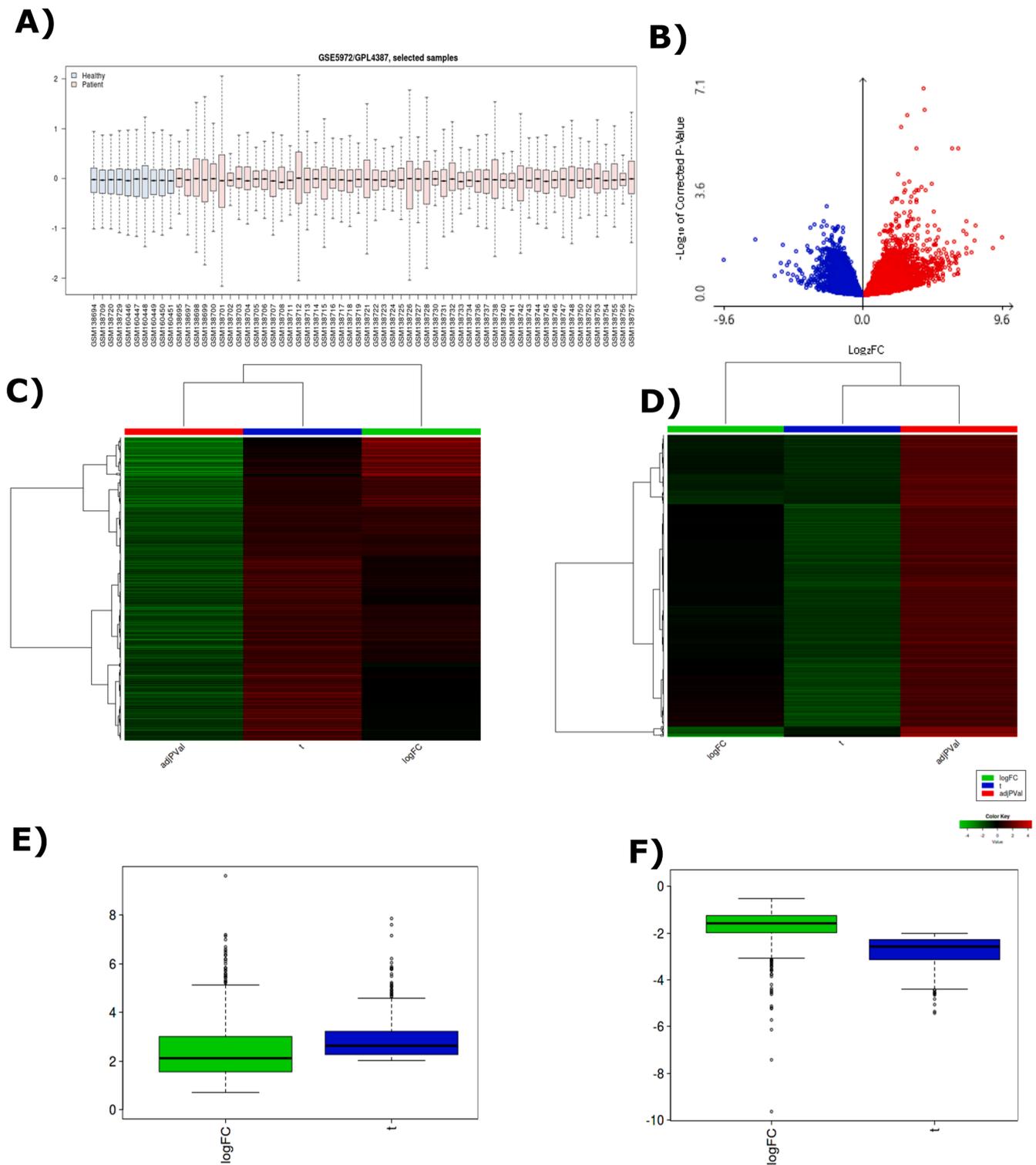
Fig. 8. Analysis of Differential Gene Expression. A) Median centralized 10 healthy and 54 SARS patient samples for comparison with the Limma normalization method. B) LogFC against adjusted normalized p-value modeled for Gene distribution in Volcano Plot. Hierarchical Clustering Heatmap of C) Upregulated and D) Downregulated Genes. Distance was based on Correlation and Z score in where cut off was set to 4. Distance-based on Correlation & Z score Cut off was set to 4. Frequency distribution of LogFC and T value of E) upregulated and F) downregulated genes.

conjugated with aptamer or delivered through various delivery systems to inhibit the intracellular growth of SARS-CoV-2.

IFNs instigate an array of antiviral effectors. Though every virus contains a distinctive property, partly overlapping antiviral IFN Stimulating Genes (also known as "ISG profile") could be a potential target (Jassal et al., 2020). IFN effectors target various phases of the virus life cycle to interfere with viral infection and proliferation. To identify the ISG and IFN of COVID-19 cases, the gene expression profile of COVID-19 patients was needed, but it was not available. Therefore, three genomes of *Betacoronavirus* that have common virulence patterns, signs and symptoms were taken, including SARS-CoV-1, MERS-CoV and SARS-CoV-2. A comparison of their genome was performed. The whole-
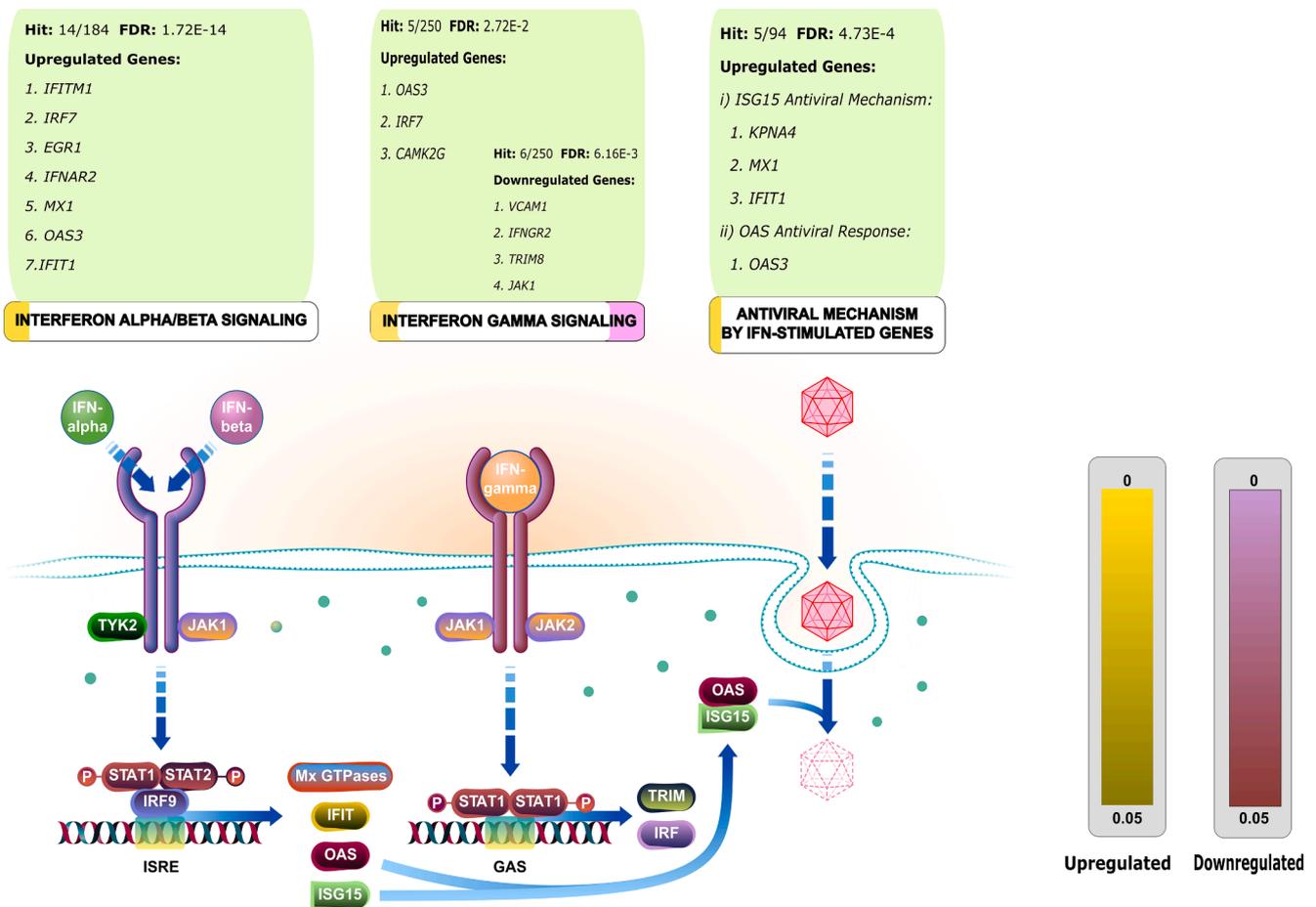
Fig. 9. Interferon stimulating genes (ISGs) in the pathway. Overrepresented pathway of interferon regulation by overexpressed and under expressed ISGs. The gene list including KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2 from the upregulated genes and TRIM8, VCAM1, IFNGR2 and JAK1 from the downregulated genes might stimulate the Interferon against the SARS-CoV-2. Here, the possible pathway mechanism has been explored to induce the immune response by the Interferon stimulating genes (ISGs).

genome comparison between SARS-CoV-1 and SARS-CoV-2 showed 88% similarity with each other (Fig. 7 and Supplementary Fig. 4). Therefore, the SARS-CoV-1 expression profile was investigated to identify ISG and IFNs. Based on this connection, we analyzed the genome-wide expression profile of SARS patients which unveiled suitable genes/receptors for the IFN-based therapeutics (Supplementary Data 4–9). A total of 167 immunoregulatory genes (91 up-regulated genes and 76 down-regulated genes) were found from the biological process of GO (Fig. 8, Supplementary Fig. 5, Supplementary Data 9–17). We then identified the IFN stimulating genes KPNA4, IRF7, OAS3, CAMK2G, IFITM1, IFIT1, EGR1, MX1 and IFNAR2 from the up-regulated immuno-regulatory genes and TRIM8, VCAM1, IFNGR2, JAK1 from the down-regulated immuno-regulatory genes (Fig. 9). The functions of these immune-regulatory genes can be modulated by the interferons IFN alpha, IFN beta and IFN gamma which can directly induce the antiviral pathways and immunological responses (Fig. 9, Supplementary Fig. 6, Supplementary Data 17–25). This analysis provides deep insight into the type 1 IFN based therapy for COVID-19 patients.

5. Conclusion

The present study has proposed several therapeutics including chimeric vaccine, potent drugs, siRNAs and ISG targets against COVID-19. This study can be used as a convenient source by the researchers and pharmaceutical companies for developing novel therapeutics against SARS-CoV-2. We hope this *in silico* pipeline will also help many researchers to combat against other viral outbreaks in the future.

Author contributions

MS: Conceived, designed, and guided the study, analyzed the data, drafting the manuscript and performed critical revision. CAK and KCD: Guided the study, acquisition and analyzed the data, helped in drafting the manuscript. MM and MGM: Helped in Bioinformatics analysis and helped in drafting the manuscript. MUH, AB, MTHE, ZMC: Helped to design the study, performed bioinformatics analysis, drafted, and developed the manuscript and performed critical revision. All authors have approved the manuscript.

Ethical statement

All applicable international, national, and/or institutional guidelines for the care and use of human patients were followed. The Ethical approval number is NIBREC2020-02.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2020.145368.

References

Abadio, A.K., Kioshima, E.S., Teixeira, M.M., Martins, N.F., Maigret, B., Felipe, M.S., et al., 2011. Comparative genomics allowed the identification of drug targets against human fungal pathogens. BMC Genomics 12 (1), 75.

Afgan, E., Baker, D., Batut, B., et al., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucl. Acids Res. 46 (W1), W537–W544.

Agrawal, S., Gupta, D., Panda, S.K., et al., 2001. The 3′ end of hepatitis E virus (HEV) genome binds specifically to the viral RNA-dependent RNA polymerase (RdRp). Virology 282 (1), 87–101.

Amarzguioui, M., Prydz, H., et al., 2004. An algorithm for selection of functional siRNA sequences. Biochem. Biophys. Res. Commun. 316 (4), 1050–1058.

Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., et al., 2020. The proximal origin of SARS-CoV-2. Nat. Med. 26 (4), 450–452.

Bankevich, A., Nurk, S., Antipov, D., et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 19 (5), 455–477.

Berera, D., Zambon, M., et al., 2013. Antivirals in the 2009 pandemic–lessons and implications for future strategies. Influenza Other Respir. Viruses 7 (Suppl 3), 72–79.

Bhattacharjee, A., Hossain, M.U., Chowdhury, Z.M., Rahman, S.A., Bhuyan, Z.A., Salimullah, M., Keya, C.A., et al., 2020. Insight of druggable cannabinoids against estrogen receptor β in breast cancer. J. Biomol. Struct. Dyn. 1–0.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30 (15), 2114–2120.

Caly, L., Druce, J.D., Catton, M.G., Jans, D.A., Wagstaff, K.M., 2020. The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 in vitro. Antiviral Res. 178, 104787.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., et al., 2005. ACT: the Artemis comparison tool. Bioinformatics 21 (16), 3422–3423.

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S.C., Di Napoli, R., et al., Features, evaluation and treatment coronavirus (COVID-19). InStatpearls [internet] 2020 Mar 8. StatPearls Publishing.

Cavanagh, D., et al. Coronaviridae: a review of coronaviruses and toroviruses. InCoronaviruses with Special Emphasis on First Insights Concerning SARS 2005 (pp. 1-54). Birkhäuser Basel.

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., Yu, T., et al., 2020. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The Lancet. 395 (10223), 507–513.

Cheng F, Li W, Zhou Y, Shen J, Wu Z, Liu G, Lee PW, Tang Y, et al. admetSAR: a comprehensive source and free tool for assessment of chemical ADMET properties.

Cui, J., Li, F., Shi, Z.L., et al., 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17 (3), 181–192.

Dhanda, S.K., Vir, P., Raghava, G.P., et al., 2013. Designing of interferon-gamma inducing MHC class-II binders. Biol. Direct. 8 (1), 30.

Dimitrov, I., Bangov, I., Flower, D.R., Doytchinova, I., et al., 2014. AllerTOP v. 2—a server for in silico prediction of allergens. J. Mol. Model. 20 (6), 2278.

Doytchinova, I.A., Flower, D.R., 2007. Identifying candidate subunit vaccines using an alignment-independent method based on principal amino acid properties. Vaccine 25 (5), 856–866.

Eastman, R.T., Roth, J.S., Brimacombe, K.R., et al., 2020. A review of its discovery and development leading to emergency use authorization for treatment of COVID-19 [published correction appears in ACS Cent Sci. 2020 Jun 24;6(6):1009]. ACS Cent Sci. 6 (5), 672–683.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., Tuschl, T., et al., 2001. Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature 411 (6836), 494–498.

Elbe, S., Buckland-Merrett, G., Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1(1):33-46. Published 2017 Jan 10. doi: 10.1002/gch2.1018.

Fan, Y., Zhao, K., Shi, Z.L., Zhou, P., et al., 2019. Bat Coronaviruses in China. Viruses 11 (3), 210.

Fung, T.S., Liu, D.X., et al., 2019. Human coronavirus: host-pathogen interaction. Annu. Rev. Microbiol. 73, 529–557.

Gao, J., Tian, Z., Breakthrough, Y.X., 2020. Chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. Bioscience trends.

Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M.R., Appel, R.D., Bairoch, A., et al. Protein identification and analysis tools on the ExPASy server. InThe proteomics protocols handbook 2005 (pp. 571-607). Humana press.

Ge SX, Jung D. ShinyGO: a graphical enrichment tool for animals and plants. bioRxiv, 315150.

Ge, S.X., Son, E.W., Yao, R., et al., 2018. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. BMC Bioinf. 19 (1), 534.

Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D.C., Jahn, D., et al., 2005. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucl. Acids Res. 33 (suppl_2). W526–31.

Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P., et al., 2013. Open Source Drug Discovery Consortium. In silico approach for predicting toxicity of peptides and proteins. PloS One 8 (9), e73957.

Gurwitz, D., et al. Angiotensin receptor blockers as tentative SARS-CoV-2 therapeutics. Drug development research. 2020 Mar 4.

He, J., Wang, J., Tao, H., Xiao, Y., Huang, S.Y., et al., 2019. HNADOCK: a nucleic acid docking server for modeling RNA/DNA–RNA/DNA 3D complex structures. Nucl. Acids Res. 47 (W1), W35–W42.

Heo, L., Lee, H., Seok, C., et al., 2016. GalaxyRefineComplex: Refinement of protein-protein complex model structures driven by interface repacking. Sci. Rep. 6, 32153.

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Müller, M.A., et al., 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell.

Hossain, M.U., Omar, T.M., Alam, I., Das, K.C., Mohiuddin, A.K., Keya, C.A., Salimullah, M., et al., 2018. Pathway based therapeutic targets identification and development of an interactive database CampyNIBase of Campylobacter jejuni RM1221 through non-redundant protein dataset. PLoS ONE 13 (6), e0198170.

Hsieh, A.C., Bo, R., Manola, J., Vazquez, F., Bare, O., Khvorova, A., Scaringe, S., Sellers, W.R., et al., 2004. A library of sirna duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. Nucl. Acids Res. 32 (3), 893–901.

Huesken, J., Lange, J., Mickanin, C., Weiler, J., Asselbergs, F., Warner, J., Meloon, B., Engel, S., Rosenberg, A., Cohen, D., Labow, M., et al., 2005. Design of a genome-wide siRNA library using an artificial neural network. Nat. Biotechnol. 23 (8), 995–1001.

Ichihara, M., Murakumo, Y., Masuda, A., Matsuura, T., Asai, N., Jijiwa, M., Ishida, M., Shinmi, J., Yatsuya, H., Qiao, S., Takahashi, M., et al., 2007. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. Nucl. Acids Res. 35 (18), e123.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., Loney, F., et al., 2020. The reactome pathway knowledgebase. Nucl. Acids Res. 48 (D1), D498–D503.

Jensen, K.K., Andreatta, M., Marcatili, P., et al., 2018. Improved methods for predicting peptide binding affinity to MHC class II molecules. Immunology 154 (3), 394–406.

Jespersen, M.C., Peters, B., Nielsen, M., Marcatili, P., et al., 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucl. Acids Res. 45 (W1), W24–W29.

Judson, R., Richard, A., Dix, D., Houck, K., Elloumi, F., Martin, M., Cathey, T., Transue, T.R., Spencer, R., Wolf, M., et al., 2008. ACToR—aggregated computational toxicology resource. Toxicol. Appl. Pharmacol. 233 (1), 7–13.

Kanehisa, M., Goto, S., et al., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucl. Acids Res. 28 (1), 27–30.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30 (4), 772–780.

Katoh, T., Suzuki, T., et al., 2007. Specific residues at every third position of siRNA shape its efficient RNAi activity. Nucleic Acids Res. 35 (4), e27.

Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S., et al., 2017. The ClusPro web server for protein–protein docking. Nat. Protoc. 12 (2), 255.

Krogh, A., Larsson, B., Von Heijne, G., Sonnhammer, E.L., et al., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305 (3), 567–580.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., McDermott, M.G., et al., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucl. Acids Res. 44 (W1), W90–W97.

Kumar Mishra, N., Raghava, P.S., et al., 2011. Prediction of specificity and cross-reactivity of kinase inhibitors. Lett. Drug Design Discovery 8 (3), 223–228.

Lai, C.C., Shih, T.P., Ko, W.C., Tang, H.J., Hsueh, P.R., et al., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Int. J. Antimicrob. Agents 17, 105924.

Larsen, M.V., Lundegaard, C., Lamberth, K., Buus, S., Lund, O., Nielsen, M., et al., 2007. Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. BMC Bioinf. 8 (1), 424.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26 (2), 283–291.

Lau, S.K., Wong, A.C., Zhang, L., Luk, H.K., Kwok, J.S., Ahmed, S.S., Cai, J.P., Zhao, P.S., Teng, J.L., Tsui, S.K., Yuen, K.Y., et al., 2019. Novel bat alphacoronaviruses in Southern China support chinese horseshoe bats as an important reservoir for potential novel coronaviruses. Viruses 11 (5), 423.

Li, J., Ulitzky, L., Silberstein, E., Taylor, D.R., Viscidi, R., 2013. Immunogenicity and protection efficacy of monomeric and trimeric recombinant SARS coronavirus spike protein subunit vaccine candidates. Viral Immunol. 26 (2), 126–132.

Magnan, C.N., Zeller, M., Kayala, M.A., Vigil, A., Randall, A., Felgner, P.L., Baldi, P., et al., 2010. High-throughput prediction of protein antigenicity using protein microarray data. Bioinformatics 26 (23), 2936–2943.

Masunov, A., et al., 2001. ACD/I-Lab 4.5: an internet service review. J. Chem. Inf. Comput. Sci. 41 (4), 1093–1095.

Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., Turner, D.H., et al., 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. 101 (19), 7287–7292.

Mathews, D.H., Turner, D.H., et al., 2006. Prediction of RNA secondary structure by free energy minimization. Curr. Opin. Struct. Biol. 16 (3), 270–278.

Moniruzzaman, M., Hossain, M.U., Islam, M.N., et al., Coding-Complete Genome Sequence of SARS-CoV-2 Isolate from Bangladesh by Sanger Sequencing. Microbiol Resour Announc. 2020;9(28):e00626-20. Published 2020 Jul 9.

Niyonsaba, F., Ogawa, H., Nagaoka, I., et al., 2004. Human β-defensin-2 functions as a chemotactic agent for tumour necrosis factor-α-treated human neutrophils. Immunology 111 (3), 273–281.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS ONE 5 (3), e9490.

Rapin, N., Lund, O., Bernaschi, M., Castiglione, F., et al., 2010. Computational immunology meets bioinformatics: the use of prediction tools for molecular binding in the simulation of the immune system. PLoS ONE 5 (4), e9862.

Ren, S.Y., Gao, R.D., Chen, Y.L., et al., 2020. Fear can be more harmful than the severe acute respiratory syndrome coronavirus 2 in controlling the corona virus disease 2019 epidemic. World J. Clin. Cases 8 (4), 652.

Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W.S., Khvorova, A., et al., 2004. Rational siRNA design for RNA interference. Nat. Biotechnol. 22 (3), 326–330.

Ritchie, M.E., Phipson, B., Wu, D.I., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., et al., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucl. Acids Res. 43 (7) e47-.

Saha, S., Raghava, G.P., et al., 2006. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Res. 34 (suppl_2). W202–9.

Samarajiwa, S.A., Forster, S., Auchettl, K., Hertzog, P.J., et al., 2009. INTERFEROME: the database of interferon regulated genes. Nucl. Acids Res. 37 (suppl_1). D852–7.

Sander, T., et al., 2001. OSIRIS property explorer. Organic Chem. Portal.

Seeliger, D., de Groot, B.L., et al., 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J. Comput. Aided Mol. Des. 24 (5), 417–422.

Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M. W., Lane, H.C., Lempicki, R.A., et al., 2007. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8 (9), R183.

Shey, R.A., Ghogomu, S.M., Esoh, K.K., Nebangwa, N.D., Shintouo, C.M., Nongley, N.F., Asa, B.F., Ngale, F.N., Vanhamme, L., Souopgui, J., et al., 2019. In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. Sci. Rep. 9 (1), 1–8.

Takasaki, S., Kotani, S., Konagaya, A., et al., 2004. An effective method for selecting siRNA target sequences in mammalian cells. Cell Cycle 3 (6), 788–793.

Tang, A., Tong, Z.D., Wang, H.L., Dai, Y.X., Li, K.F., Liu, J.N., Wu, W.J., Yuan, C., Yu, M. L., Li, P., Yan, J.B., et al., 2020. Detection of novel Coronavirus by RT-PCR in stool specimen from asymptomatic child, China. Emerg. Infect. Dis. 26 (6).

Tian, W., Chen, C., Lei, X., Zhao, J., Liang, J., et al., 2018. CASTp 3.0: computed atlas of surface topography of proteins. Nucl. Acids Res. 46 (W1), W363–W367.

Trott, O., Olson, A.J., 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 31 (2), 455–461.

Ui-Tei, K., Naito, Y., Zenno, S., Nishi, K., Yamato, K., Takahashi, F., Juni, A., Saigo, K., et al., 2008. Functional dissection of siRNA sequence by systematic DNA substitution: modified siRNA with a DNA seed arm is a powerful tool for mammalian gene silencing with significantly reduced off-target effect. Nucl. Acids Res. 36 (7), 2136–2151.

Vert, J.P., Foveau, N., Lajaunie, C., Vandenbrouck, Y., et al., 2006. An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinf. 7 (1), 520.

WHO. Naming the coronavirus disease (COVID-19) and the virus that causes it 2029 (2019).

Wiederstein, M., Sippl, M.J., et al., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucl. Acids Res. 35 (suppl_2). W407–10.

Wu, Y.C., Chen, C.S., Chan, Y.J., et al., 2020. The outbreak of COVID-19: an overview. J. Chin. Med. Assoc. 83 (3), 217.

Wu, Z., McGoogan, J.M., et al., 2020. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 323 (13), 1239–1242.

Yu, L., Wang, L., Chen, S., et al., 2010. Endogenous toll-like receptor ligands and their biological significance. J. Cell Mol. Med. 14 (11), 2592–2603.

Zheng, W., Li, Y., Zhang, C., Pearce, R., Mortuza, S.M., Zhang, Y., 2019. Deep-learning contact-map guided protein structure prediction in CASP13. Proteins Struct. Funct. Bioinf. 87 (12), 1149–1164.