

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^{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- 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

Abbreviations: SARS-CoV-2, Severe Acute Respiratory Syndrome – Coronavirus; MERS-CoV, Middle East Respiratory Syndrome - Coronavirus; ACE2, angiotensin-converting enzyme 2; PLpro, papain like protease; M^{pro} , Main Protease; 3CLpro, 3-chymotrypsin-like protease; RTC, replicase-transcriptase complex; RBD, receptor binding domain; ML, Machine Learning; DEE, Dead End Elimination; MD, Molecular Dynamics; RMSD, root mean square deviation; RMSF, root mean square fluctuation

* Corresponding author.

E-mail address: singhs@nccs.res.in (S. Singh).

<https://doi.org/10.1016/j.bbadis.2020.165978>

Received 9 July 2020; Received in revised form 10 September 2020; Accepted 21 September 2020

Available online 24 September 2020

0925-4439/ © 2020 Elsevier B.V. All rights reserved.

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 in 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 upto 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. Coronaviruses 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 M^{PRO} 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 proteinases. 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-

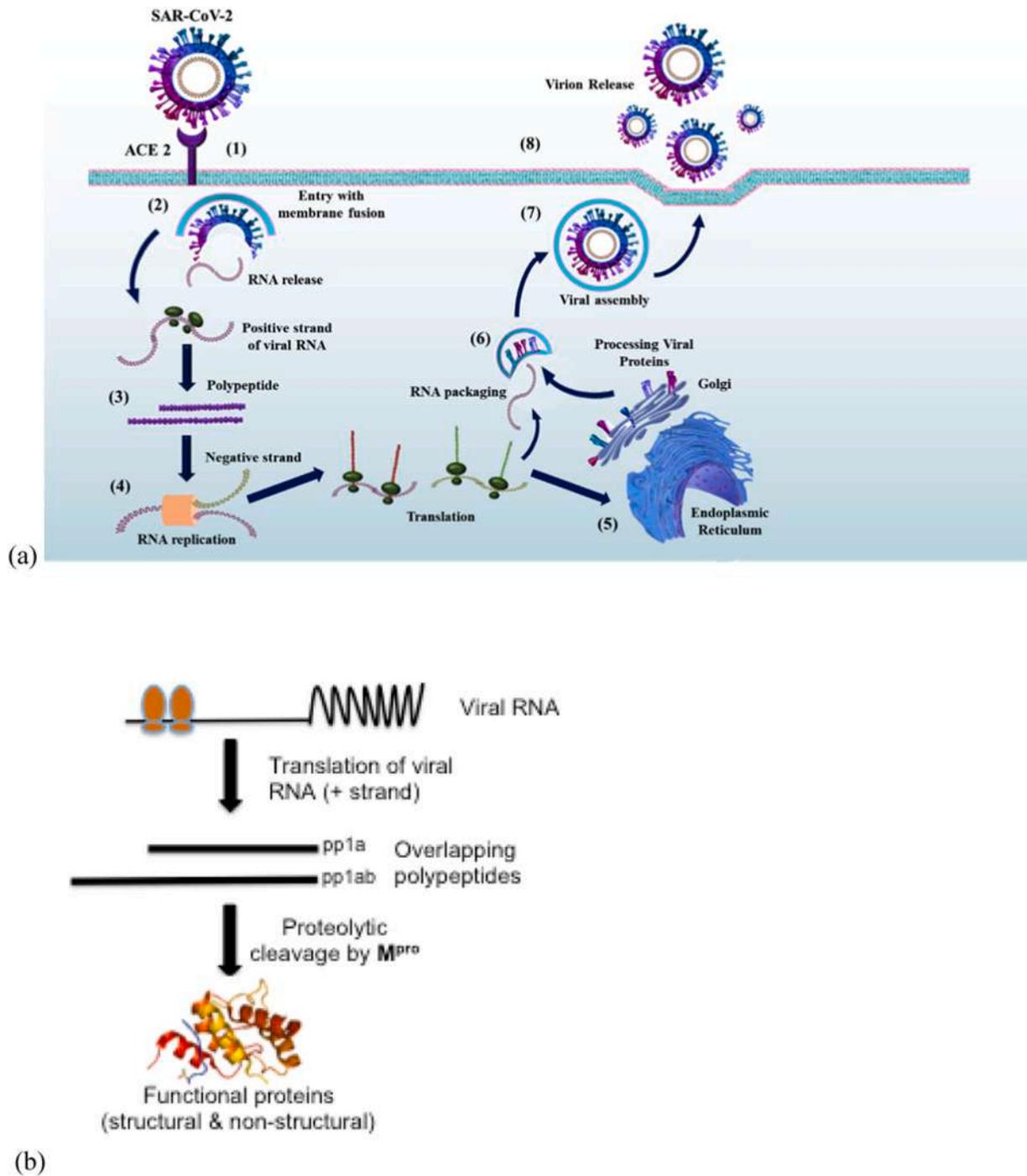


Fig. 1. (a) Life cycle of SARS-CoV-2. (1) The replication of SARS-CoV-2 begins with the binding of its spike protein (S) with receptor angiotensin-converting enzyme 2 (ACE2) on the cell surface molecules of the host. (2) The viral S protein undergoes conformational changes leading to a fusion between the viral envelope with the plasma or endosomal membranes, entering the cell's cytoplasm, and releasing RNA. (3) The positive-strand viral RNA serves as mRNA for translation of ORF 1a and 1b, encoding units of the RNA-dependent RNA polymerase. (4) RNA polymerase transcribes full-length complementary negative-sense RNA. Viral proteins (S, M, E, and N) get translated (5) The proteins are further processed in the Endoplasmic Reticulum and Golgi Apparatus. (6) Packaging of viral proteins and positive-strand RNA. (7) Assembling complete virion surrounded by the plasma membrane. (8) Apoptosis of cells with the release of the virion. (b) Role of Main protease M^{pro} in SARS-CoV-2.

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 bio-manufacturing 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].

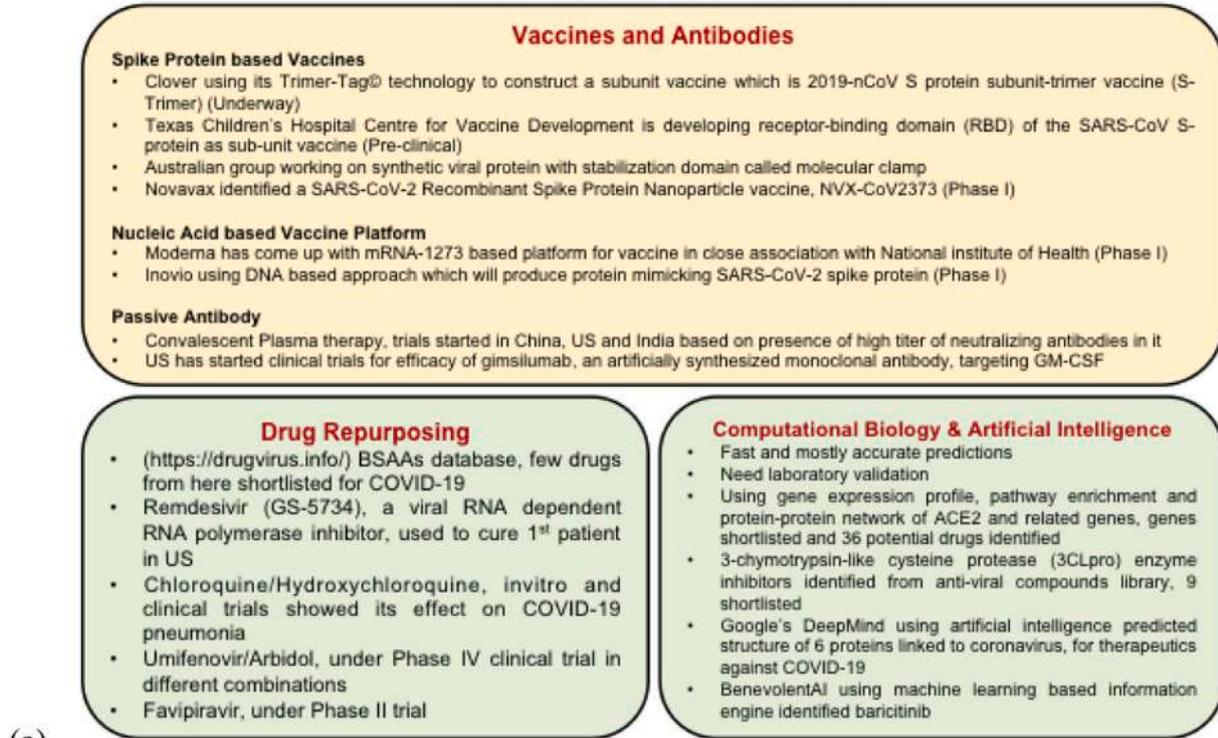
FDA has initiated guidelines for using convalescent plasma therapy and Methodist Hospital in Houston recruited plasma donors a few days back and initiated the first plasma transfusions to COVID-19 patient the following day [42-44]. The Drug Controller General of India (DCGI) has also approved clinical 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

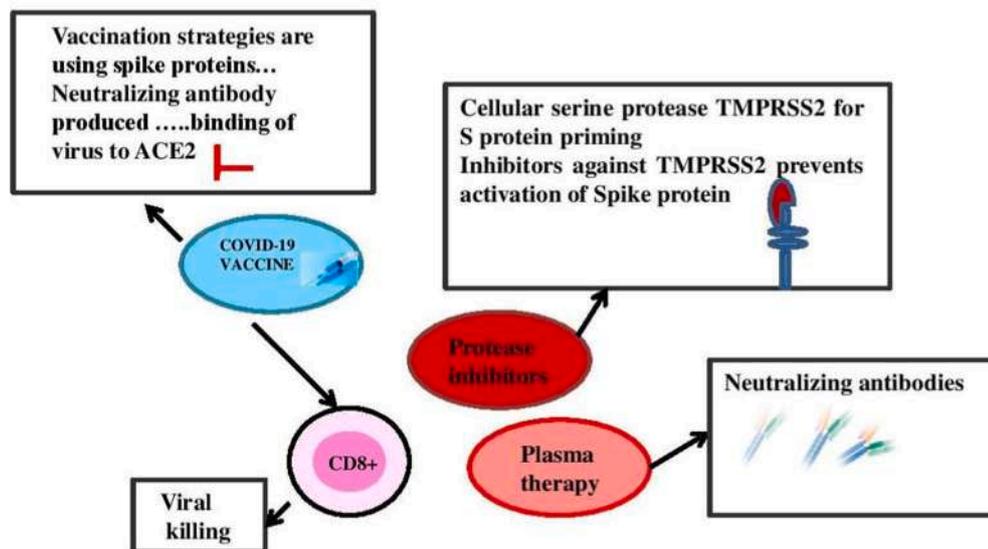
This arm of treatment against COVID-19 is bifurcated into two strategies, one is drug repurposing and other is search for novel drug molecules based on structure-based drug designing using computational biology and Artificial intelligence. FDA and many other big organizations are insisting researchers for long, for the repurposing of drugs. Broad-spectrum antiviral agents (BSAAs), 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

COVID-19 TREATMENT STRATEGIES



(a)



(b)

Fig. 2. Various therapeutic strategies underway against COVID-19 (a) and their proposed outcomes (b).

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

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](https://clinicaltrials.gov/ct2/show/study/NCT04255017) 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 drug target is the main protease of SARS-CoV-2, M^{Pro}. M^{Pro} 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 M^{Pro}, also known as 3CLpro (3-chymotrypsin-like protease) cleaves polypeptides at 11 distinct sites. M^{Pro} 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 M^{Pro} [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-M^{Pro}, 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 $m \times k$ and $m \times l$ matrices respectively. Let R and S be $m \times m$ and $l \times k$ matrices, such that R and extension of S to $n \times n$ matrices,

$n = \max(m, k, l)$, which is a positive semi definite matrix. Then, $(P, Q) = \text{tr}(P^T R Q) S$ 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:

$$D = \{(A_i T_i)\}_i^n = 1$$

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 $S \in S'$ and for each A in the training dataset, we used f_s to define the feature vector. For two sets of sequences, A, A', we constructed the complete weighted bipartite sequences and computed the kernel $K_a(A, A')$ using equation:

$$K_M(AA') = \frac{\max}{\pi} \sum_{\gamma \in V[A]} K_n(f(V) f(\pi(V)))$$

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 $S \in S'$ and for each A in the testing data set, f_s 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 $K_a(A, A')$ 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 buildups to a most extreme length of 20 buildups 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 in consideration with physicochemical 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 doesn't fit in global minimum energy conformation. The total energy of a given pair is represented as:

$$E_{Tot} = E_r + \sum_i E(i_r) + \sum_i \sum_{j>i} E(i_r j_s)$$

where, E_{Tot} is the total energy of the system; E_r is the initial energy of

Dead End Elimination (DEE)

Algorithm Overview

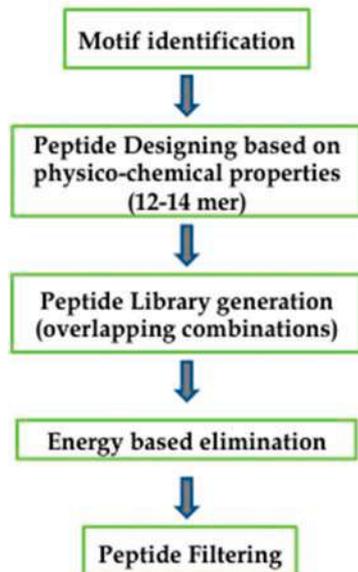


Fig. 3. Overview of the algorithm used for peptide designing and filtering.

Peptide Filtering

- Geometric and chemical complementarity
- Stearic hindrance
- Structural stability
- Binding affinity
- Specific interactions
- Reverse Docking

the template; i_r is the rotamer state at position i ; $E(i_r)$ is the self-energy of rotamer at position i ; and $E(i_r, j_s)$ 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, hydrophobicity 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:

$$\Delta G_{\text{bind}} = \Delta G_{\text{vdw}} + \Delta G_{\text{hbond}} + \Delta G_{\text{elect}} + \Delta G_{\text{conform}} + \Delta G_{\text{tor}} + \Delta G_{\text{sol}}$$

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 Diagrams. 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:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r'_i(t_x) - r_i(t_{ref}))^2}$$

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:

$$RMSF = \sqrt{\frac{1}{T} \sum_{i=1}^T \langle (r'_i(t) - r_i(t_{ref}))^2 \rangle}$$

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:

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{oop}} + E_{\text{cross}} + E_{\text{nonbond}}$$

where, E_{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_{angle} is the bond angle energy and is defined as:

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

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

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

E_{oop} is the out of plane interactions and is defined as:

$$E_{\text{oop}} = \sum_x H_x X^2$$

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

E_{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^{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^{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^{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^{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.

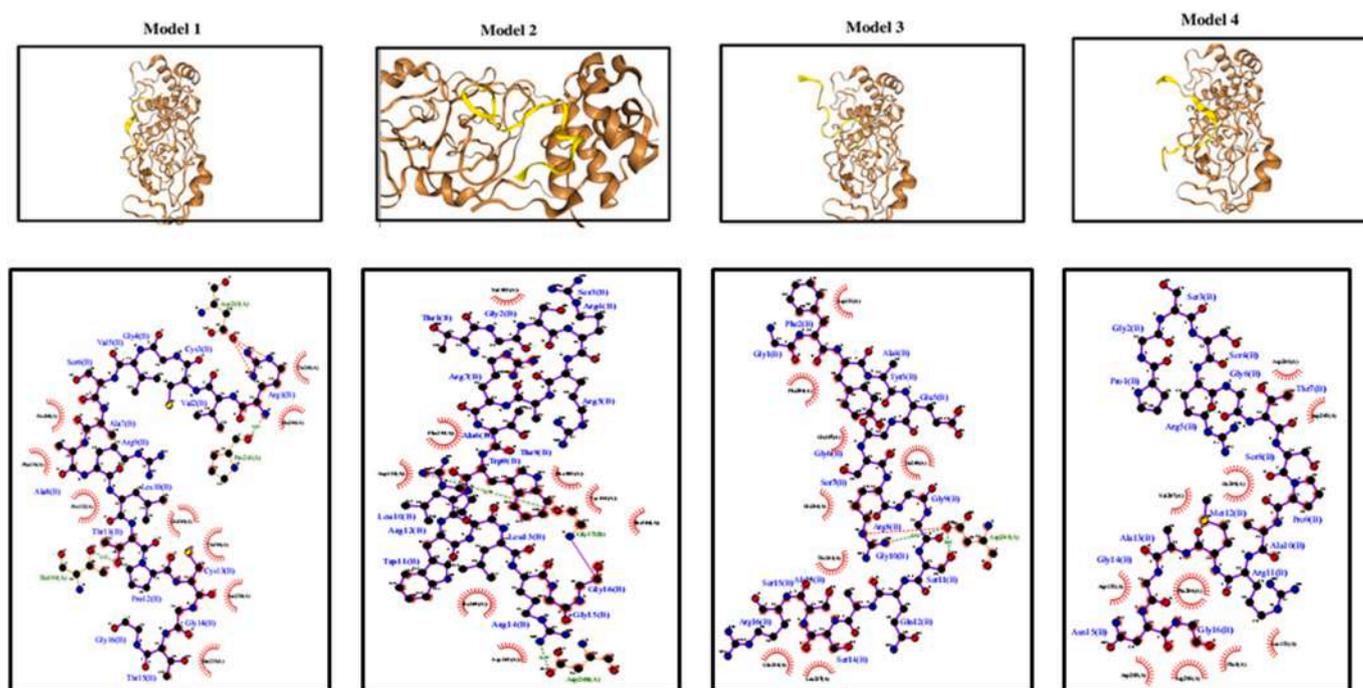


Fig. 4. Docking results of identified sequences with proteases wherein peptide is in yellow color and receptor in brown. The upper panel shows the docked poses with the 4 shortlisted peptides (protein represented in brown ribbon form while peptide in yellow color) and the lower panel represents their respective protein-peptide interactions.

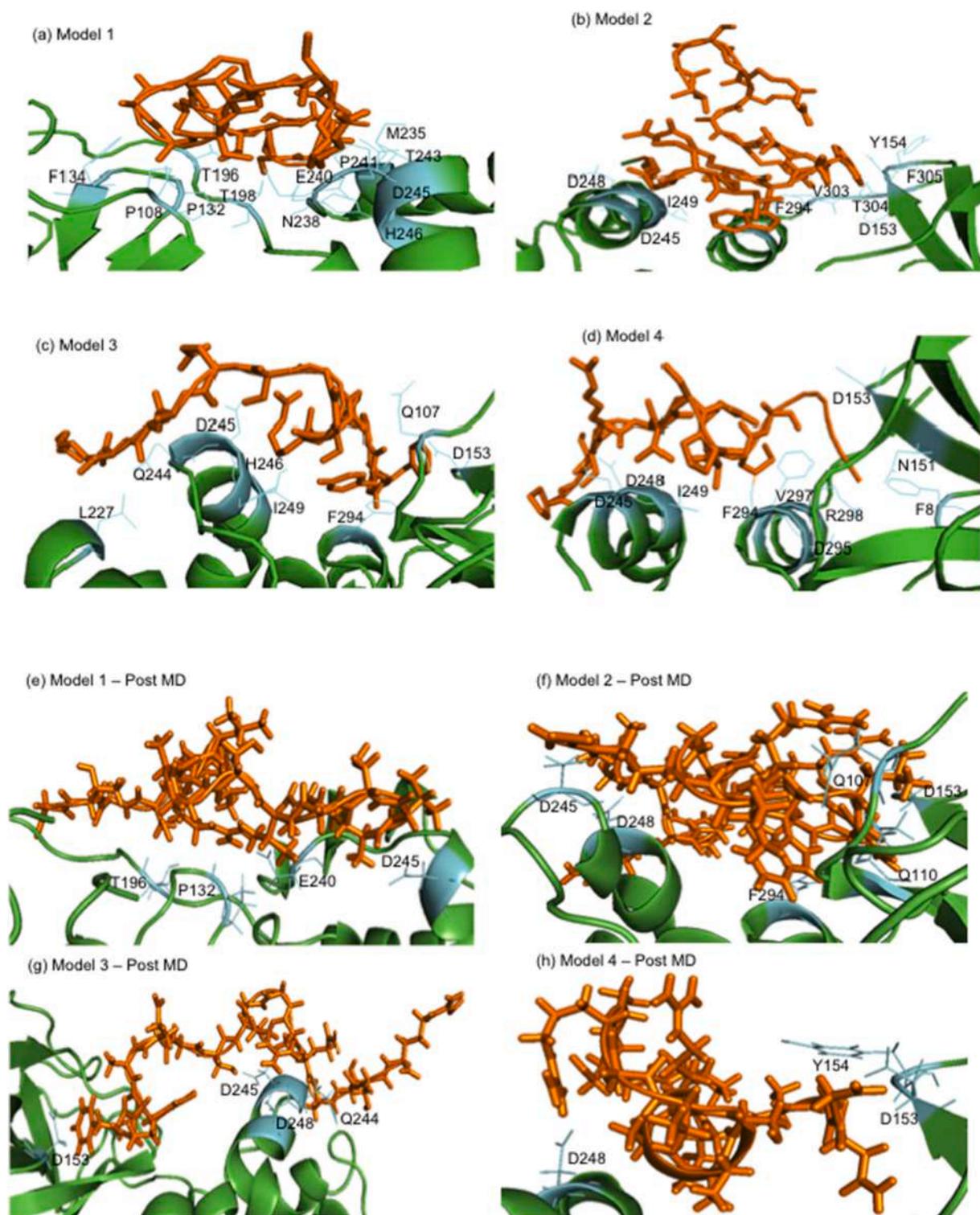


Fig. 5. M^{pro} residues interacting with peptides. (a) to (d) are protein-peptide complexes before MD Simulations and (e) to (h) are protein-peptide complexes post MD Simulations. Peptide is represented in orange stick form while residues of M^{pro} protein interacting with peptides are shown in cyan color.

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

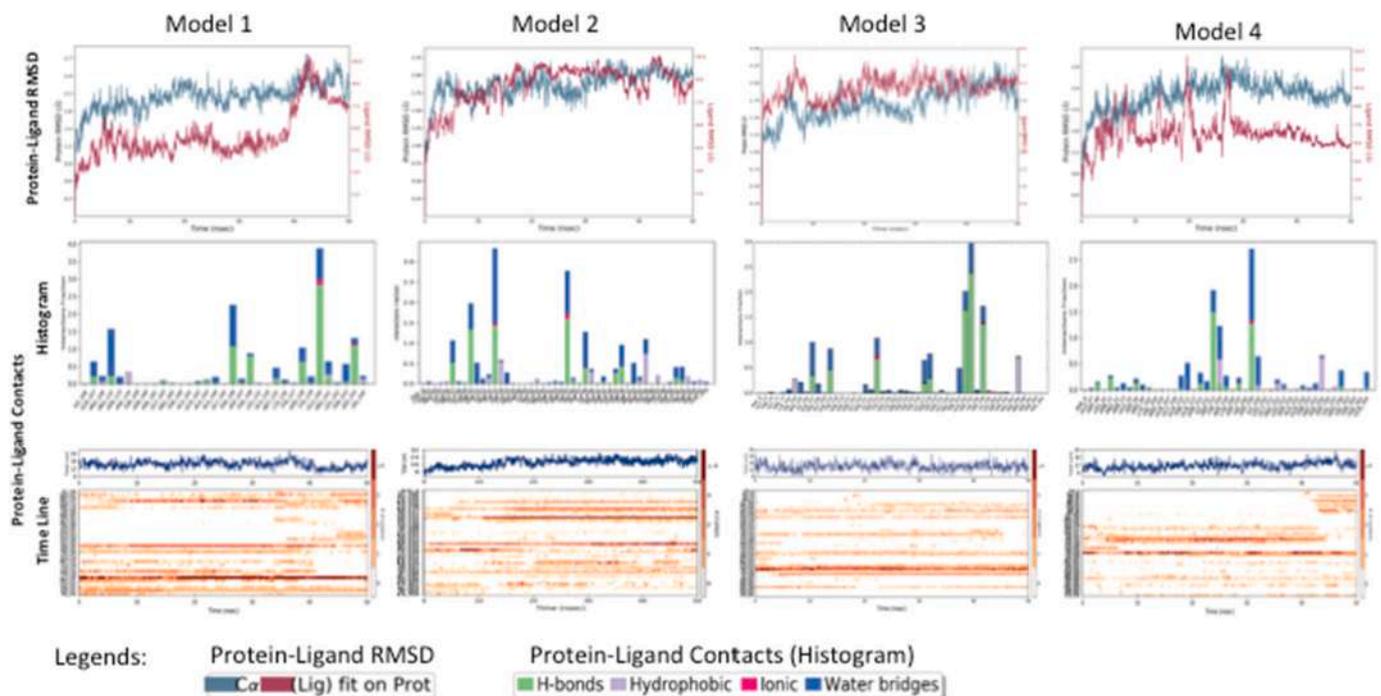


Fig. 6. Short scale Molecular Dynamics Simulation studies to check the stability, strength and amount of interactions being maintained between protein and peptide. The first row represents the protein – ligand (here peptide) RMSDs with protein in blue and ligand in red color. The middle row represents 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.

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 outcores 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 discrete design 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

Table 1
Identified nucleotide sequences for peptide designing and Total Energy score.

S. No	Motif sequences	Motif E-value	Translated sequence	Total energy (kcal/mol)
Model 1	CGGGTTTGCGGTGTAAGTGCAGCCCGTCTTACACCGTGCGGCACAGGCAC	1.0e-038	RVCGVSAARLTPCGTG	-23.060
Model 2	ACAGGTTTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGG	5.5e-038	TGSRRTWLWRLRGG	-38.704
Model 3	GGCTTCTACGCAGAAGGGAGCAGAGGGCGCAGTCAAGCCTCTTCTCGTTC	3.0e-037	GFYAEGSRGGSQASSR	-37.035
Model 4	CCAGGCAGCAGTAGGGGAACCTTCTCTGCTAGAAATGGTGGCAATGGGGG	3.0e-037	PGSSRGTSPPARMAGN	-21.124

Table 2
List of interacting residues with the shortlisted peptides.

Model	Interacting residues
1	P108, P132, F134, T196, T198, M235, N238, E240, P241, T243, D245, H246
2	D153, Y154, D245, D248, I249, F294, V303, T304, F305
3	Q107, D153, L227, T243, Q244, D245, H246, I249, F294
4	F8, N151, D153, D245, D248, I249, F294, D295, V297, R298

Table 3
Number of interactions between protein-peptide pre-MD and during MD simulations.

Model	Interactions pre-MD	Interactions during MD
1	12	33
2	9	48
3	9	49
4	10	45

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 short-listed. These 4 protein-peptide complexes were subsequently subjected to MD Simulation studies to ensure their stability in physiological

conditions as well confirmed which peptide has maintained their potency. Of the 4 most efficient peptides obtained post screening of peptide library, four common interacting residues were found in model 2, 3 and 4. These residues were D153, D245, I249 and F294. Among these also D245 showed strong interaction with all the four peptides suggesting that these residues might be crucial for its functionality. During docking simulations we have observed that the hydrophobic residues viz., I106, H246, I249, P293, F204 and F305 showed good interactions strength with peptides on one side and on the other side the polar residues includes D153, N228, D245, D248, R298, Q306 of the amphiphatic structure. These observed structural implications accounts for peptides selectivity towards the target and apparently their antiviral effect. Docking as well as MD simulation analysis shows that model2 and model3 with peptides 2 and 3 respectively, are the best candidates. However, further experimental validation is required to assess the effect of these peptides on the progression of SARS-CoV-2. Machine learning is far more precise than the traditional conventional approaches since it takes into consideration the important factors required for biological activity and selectivity of the peptides. It can also be perceived that the computational designed set of peptides can serve as antiviral therapeutics as well as immuno-modulatory agents because of its versatility as well as tenability. These peptides may also be used as templates for the design of other molecules guided by its hydrophobicity, net positive charge and amphiphilicity.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bbadis.2020.165978>.

Credit authorship contribution statement

All authors contributed to study design and conception. **Ritika Kabra (RK)**: data curation, methodology implementation and writing the original draft. **Shailza Singh (SS)**: Initialization and conceptualization, data curation, supervision, writing and editing the original draft. All authors agree with the submission of 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.

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

- [1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet*. 395 (2020) 497–506, [https://doi.org/10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5).
- [2] P. Zhai, Y. Ding, X. Wu, J. Long, Y. Zhong, Y. Li, The epidemiology, diagnosis and treatment of COVID-19, *Int. J. Antimicrob. Agents* 55 (2020), <https://doi.org/10.1016/j.ijantimicag.2020.105955>.
- [3] K. Dhama, K. Sharun, R. Tiwari, M. Dadar, Y.S. Malik, K.P. Singh, W. Chaicumpa, COVID-19, an emerging coronavirus infection: advances and prospects in designing and developing vaccines, immunotherapeutics, and therapeutics, *Hum. Vaccines Immunother.* 16 (2020) 1232–1238, <https://doi.org/10.1080/21645515.2020.1735227>.
- [4] Manli Wang, Ruiyuan Cao, Leike Zhang, Xinglou Yang, Jia Liu, Mingyue Xu, Zhengli Shi, Zhihong Hu, Zhong Wu, Gengfu Xiao, Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro, *Cell Res.* 30 (2020) 269–271 <https://www.nature.com/articles/s41422-020-0282-0>.
- [5] A.A.T. Naqvi, K. Fatima, T. Mohammad, U. Fatima, I.K. Singh, A. Singh, S.M. Atif, G. Hariprasad, G.M. Hasan, M.I. Hassan, Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach, *Biochim. Biophys. Acta Mol. basis Dis.* 2020 (1866), <https://doi.org/10.1016/j.bbadis.2020.165878>.
- [6] J. Xu, S. Zhao, T. Teng, A.E. Abdalla, W. Zhu, L. Xie, Y. Wang, X. Guo, Systematic comparison of two animal-to-human transmitted human coronaviruses: SARS-CoV-2 and SARS-CoV, *Viruses* 12 (2020), <https://doi.org/10.3390/v12020244>.
- [7] S. Mandal, T. Bhatnagar, N. Arinaminpathy, A. Agarwal, A. Chowdhury, M. Murhekar, R. Gangakhedkar, S. Sarkar, Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India: a mathematical model-based approach, *Indian J. Med. Res.* 151 (2020) 190–199, https://doi.org/10.4103/ijmr.IJMR_504_20.
- [8] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K.S.M. Leung, E.H.Y. Lau, J.Y. Wong, X. Xing, N. Xiang, Y. Wu, C. Li, Q. Chen, D. Li, T. Liu, J. Zhao, M. Liu, W. Tu, C. Chen, L. Jin, R. Yang, Q. Wang, S. Zhou, R. Wang, H. Liu, Y. Luo, Y. Liu, G. Shao, H. Li, Z. Tao, Y. Yang, Z. Deng, B. Liu, Z. Ma, Y. Zhang, G. Shi, T.T.Y. Lam, J.T. Wu, G.F. Gao, B.J. Cowling, B. Yang, G.M. Leung, Z. Feng, Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia, *N. Engl. J. Med.* 382 (2020) 1199–1207, <https://doi.org/10.1056/NEJMoa2001316>.
- [9] N. Yu, W. Li, Q. Kang, Z. Xiong, S. Wang, X. Lin, Y. Liu, J. Xiao, H. Liu, D. Deng, S. Chen, W. Zeng, L. Feng, J. Wu, Clinical features and obstetric and neonatal outcomes of pregnant patients with COVID-19 in Wuhan, China: a retrospective, single-centre, descriptive study, *Lancet Infect. Dis.* 20 (2020) 559–564, [https://doi.org/10.1016/S1473-3099\(20\)30176-6](https://doi.org/10.1016/S1473-3099(20)30176-6).
- [10] B. Prateek, C. Doolan, C. de Silva, A.A. Chughtai, L. Bourouiba, R.C. MacIntyre, Airborne or droplet precautions for health workers treating COVID-19? *J. Infect. Dis.* 189 (2020) (in press).
- [11] W. World Health Organization, Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations, *Sci. Br.* (2020) 1–3, <https://doi.org/10.1056/NEJMoa2001316.5>.
- [12] W. Li, Z. Shi, M. Yu, W. Ren, C. Smith, J.H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B.T. Eaton, S. Zhang, L.F. Wang, Bats are natural reservoirs of SARS-like coronaviruses, *Science* (80-) 310 (2005) 676–679, <https://doi.org/10.1126/science.1118391>.
- [13] S.K.P. Lau, P.C.Y. Woo, K.S.M. Li, Y. Huang, H.W. Tsoi, B.H.L. Wong, S.S.Y. Wong, S.Y. Leung, K.H. Chan, K.Y. Yuen, Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 14040–14045, <https://doi.org/10.1073/pnas.0506735102>.
- [14] J. Gu, B. Han, J. Wang, COVID-19: gastrointestinal manifestations and potential fecal–oral transmission, *Gastroenterology*. 158 (2020) 1518–1519, <https://doi.org/10.1053/j.gastro.2020.02.054>.
- [15] F. Colavita, D. Lapa, F. Carletti, E. Lalle, L. Bordini, P. Marsella, E. Nicastri, N. Bevilacqua, M.L. Giancola, A. Corpolongo, G. Ippolito, M.R. Capobianchi, C. Castilletti, SARS-CoV-2 isolation from ocular secretions of a patient with COVID-19 in Italy with prolonged viral RNA detection, *Ann. Intern. Med.* 173 (2020) 242–243, <https://doi.org/10.7326/M20-1176>.
- [16] L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.L. Yen, M. Peiris, J. Wu, SARS-CoV-2 viral load in upper respiratory specimens of infected patients, *N. Engl. J. Med.* 382 (2020) 1177–1179, <https://doi.org/10.1056/NEJMc2001737>.
- [17] K.K.W. To, O.T.Y. Tsang, W.S. Leung, A.R. Tam, T.C. Wu, D.C. Lung, C.C.Y. Yip, J.P. Cai, J.M.C. Chan, T.S.H. Chik, D.P.L. Lau, C.Y.C. Choi, L.L. Chen, W.M. Chan, K.H. Chan, J.D. Ip, A.C.K. Ng, R.W.S. Poon, C.T. Luo, V.C.C. Cheng, J.F.W. Chan, I.F.N. Hung, Z. Chen, H. Chen, K.Y. Yuen, Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study, *Lancet Infect. Dis.* 20 (2020) 565–574, [https://doi.org/10.1016/S1473-3099\(20\)30196-1](https://doi.org/10.1016/S1473-3099(20)30196-1).
- [18] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: genome structure, replication, and pathogenesis, *J. Med. Virol.* 92 (2020) 418–423, <https://doi.org/10.1002/jmv.25681>.
- [19] X. He, E.H.Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y.C. Lau, J.Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B.J. Cowling, F. Li, G.M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19, *Nat. Med.* 26 (2020) 672–675, <https://doi.org/10.1038/s41591-020-0869-5>.
- [20] L. Zhao, B.K. Jha, A. Wu, R. Elliott, J. Ziebuhr, A.E. Gorbalenya, R.H. Silverman, S.R. Weiss, Antagonism of the interferon-induced OAS-RNase L pathway by murine coronavirus ns2 protein is required for virus replication and liver pathology, *Cell Host Microbe* 11 (2012) 607–616, <https://doi.org/10.1016/j.chom.2012.04.011>.
- [21] P.K.C. Cheng, D.A. Wong, L.K.L. Tong, S.M. Ip, A.C.T. Lo, C.S. Lau, E.Y.H. Yeung, W.W.L. Lim, Viral shedding patterns of coronavirus in patients with probable severe acute respiratory syndrome, *Lancet*. 363 (2004) 1699–1700, [https://doi.org/10.1016/S0140-6736\(04\)16255-7](https://doi.org/10.1016/S0140-6736(04)16255-7).
- [22] S. Belouzard, V.C. Chu, G.R. Whittaker, Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 5871–5876, <https://doi.org/10.1073/pnas.0809524106>.
- [23] P.V. Baranov, C.M. Henderson, C.B. Anderson, R.F. Gesteland, J.F. Atkins, M.T. Howard, Programmed ribosomal frameshifting in decoding the SARS-CoV genome, *Virology*. 332 (2005) 498–510, <https://doi.org/10.1016/j.virol.2004.11.038>.
- [24] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales, *J. Gen. Virol.* 81 (2000) 853–879, <https://doi.org/10.1099/0022-1317-81-4-853>.
- [25] E.J. Snijder, P.J. Bredenbeek, J.C. Dobbe, V. Thiel, J. Ziebuhr, L.L.M. Poon, Y. Guan, M. Rozanov, W.J.M. Spaan, A.E. Gorbalenya, Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage, *J. Mol. Biol.* 331 (2003) 991–1004, <https://doi.org/10.1016/>

- 10022-2836(03)00865-9.
- [26] A.R. Fehr, S. Perlman, Coronaviruses: an overview of their replication and pathogenesis, *Coronaviruses Methods Protoc*, 2015, pp. 1–23, https://doi.org/10.1007/978-1-4939-2438-7_1.
- [27] S. Skariyachan, S.B. Challapilli, S. Packiramsay, S.T. Kumargowda, V.S. Sridhar, Recent aspects on the pathogenesis mechanism, animal models and novel therapeutic interventions for middle east respiratory syndrome coronavirus infections, *Front. Microbiol.* 10 (2019), <https://doi.org/10.3389/fmicb.2019.00569>.
- [28] T. Thanh Le, Z. Andreadakis, A. Kumar, R. Gómez Román, S. Tollefsen, M. Saville, S. Mayhew, The COVID-19 vaccine development landscape, *Nat. Rev. Drug Discov.* 19 (2020) 305–306, <https://doi.org/10.1038/d41573-020-00073-5>.
- [29] C. Wu, Y. Liu, Y. Yang, P. Zhang, W. Zhong, Y. Wang, Q. Wang, Y. Xu, M. Li, X. Li, M. Zheng, L. Chen, H. Li, Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods, *Acta Pharm. Sin. B* 10 (2020) 766–788, <https://doi.org/10.1016/j.apsb.2020.02.008>.
- [30] L. Peeples, Avoiding pitfalls in the pursuit of a COVID-19 vaccine, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 8218–8221, <https://doi.org/10.1073/pnas.2005456117>.
- [31] G. Yamey, M. Schäferhoff, R. Hatchett, M. Pate, F. Zhao, K.K. McDade, Ensuring global access to COVID-19 vaccines, *Lancet*. 395 (2020) 1405–1406, [https://doi.org/10.1016/S0140-6736\(20\)30763-7](https://doi.org/10.1016/S0140-6736(20)30763-7).
- [32] M.Z. Tay, C.M. Poh, L. Rénia, P.A. MacAry, L.F.P. Ng, The trinity of COVID-19: immunity, inflammation and intervention, *Nat. Rev. Immunol.* 20 (2020) 363–374, <https://doi.org/10.1038/s41577-020-0311-8>.
- [33] W.H. Chen, U. Strych, P.J. Hotez, M.E. Bottazzi, The SARS-CoV-2 vaccine pipeline: an overview, *Curr. Trop. Med. Rep.* 7 (2020) 61–64, <https://doi.org/10.1007/s40475-020-00201-6>.
- [34] L. Corey, J.R. Mascola, A.S. Fauci, F.S. Collins, A strategic approach to COVID-19 vaccine R&D, *Science* (80-) 368 (2020) 948–950, <https://doi.org/10.1126/science.abc5312>.
- [35] B.S. Graham, Rapid COVID-19 vaccine development, *Science* (80-) 368 (2020) 945–946, <https://doi.org/10.1126/science.abb8923>.
- [36] N. Lurie, M. Saville, R. Hatchett, J. Halton, Developing covid-19 vaccines at pandemic speed, *N. Engl. J. Med.* 382 (2020) 1969–1973, <https://doi.org/10.1056/NEJMp2005630>.
- [37] E. Prompetchara, C. Ketloy, T. Palaga, Immune responses in COVID-19 and potential vaccines: lessons learned from SARS and MERS epidemic, *Asian Pac. J. Allergy Immunol.* 38 (2020) 1–9, <https://doi.org/10.12932/AP-200220-0772>.
- [38] K. Duan, B. Liu, C. Li, H. Zhang, T. Yu, J. Qu, M. Zhou, L. Chen, S. Meng, Y. Hu, C. Peng, M. Yuan, J. Huang, Z. Wang, J. Yu, X. Gao, D. Wang, X. Yu, L. Li, J. Zhang, X. Wu, B. Li, Y. Xu, W. Chen, Y. Peng, Y. Hu, L. Lin, X. Liu, S. Huang, Z. Zhou, L. Zhang, Y. Wang, Z. Zhang, K. Deng, Z. Xia, Q. Gong, W. Zhang, X. Zheng, Y. Liu, H. Yang, D. Zhou, D. Yu, J. Hou, Z. Shi, S. Chen, Z. Chen, X. Zhang, X. Yang, Effectiveness of convalescent plasma therapy in severe COVID-19 patients, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 9490–9496, <https://doi.org/10.1073/pnas.2004168117>.
- [39] B. Zhou, N. Zhong, Y. Guan, Treatment with convalescent plasma for influenza A (H5N1) infection [9], *N. Engl. J. Med.* 357 (2007) 1450–1451, <https://doi.org/10.1056/NEJMc070359>.
- [40] J.H. Tanne, Covid-19: FDA approves use of convalescent plasma to treat critically ill patients, *BMJ*. 368 (2020) m1256, <https://doi.org/10.1136/bmj.m1256>.
- [41] J.H. Ko, H. Seok, S.Y. Cho, Y.E. Ha, J.Y. Baek, S.H. Kim, Y.J. Kim, J.K. Park, C.R. Chung, E.S. Kang, D. Cho, M.A. Müller, C. Drosten, C.I. Kang, D.R. Chung, J.H. Song, K.R. Peck, Challenges of convalescent plasma infusion therapy in Middle East respiratory coronavirus infection: a single centre experience, *Antivir. Ther.* 23 (2018) 617–622, <https://doi.org/10.3851/IMP3243>.
- [42] Y. Cheng, R. Wong, Y.O.Y. Soo, W.S. Wong, C.K. Lee, M.H.L. Ng, P. Chan, K.C. Wong, C.B. Leung, G. Cheng, Use of convalescent plasma therapy in SARS patients in Hong Kong, *Eur. J. Clin. Microbiol. Infect. Dis.* 24 (2005) 44–46, <https://doi.org/10.1007/s10096-004-1271-9>.
- [43] I.F.N. Hung, K.K.W. To, C.K. Lee, K.L. Lee, K. Chan, W.W. Yan, R. Liu, C.L. Watt, W.M. Chan, K.Y. Lai, C.K. Koo, T. Buckley, F.L. Chow, K.K. Wong, H.S. Chan, C.K. Ching, B.S.F. Tang, C.C.Y. Lau, I.W.S. Li, S.H. Liu, K.H. Chan, C.K. Lin, K.Y. Yuen, Convalescent plasma treatment reduced mortality in patients with severe pandemic influenza A (H1N1) 2009 virus infection, *Clin. Infect. Dis.* 52 (2011) 447–456, <https://doi.org/10.1093/cid/ciq106>.
- [44] B. Zhang, S. Liu, T. Tan, W. Huang, Y. Dong, L. Chen, Q. Chen, L. Zhang, Q. Zhong, X. Zhang, Y. Zou, S. Zhang, Treatment with convalescent plasma for critically ill patients with severe acute respiratory syndrome coronavirus 2 infection, *Chest*. 158 (2020) e9–e13, <https://doi.org/10.1016/j.chest.2020.03.039>.
- [45] L. Chen, J. Xiong, L. Bao, Y. Shi, Convalescent plasma as a potential therapy for COVID-19, *Lancet Infect. Dis.* 20 (2020) 398–400, [https://doi.org/10.1016/S1473-3099\(20\)30141-9](https://doi.org/10.1016/S1473-3099(20)30141-9).
- [46] Q.-L. Zeng, Z.-J. Yu, J.-J. Gou, G.-M. Li, S.-H. Ma, G.-F. Zhang, J.-H. Xu, W.-B. Lin, G.-L. Cui, M.-M. Zhang, C. Li, Z.-S. Wang, Z.-H. Zhang, Z.-S. Liu, Effect of convalescent plasma therapy on viral shedding and survival in COVID-19 patients, *J. Infect. Dis.* 222 (1) (2020) 38–43, <https://doi.org/10.1093/infdis/jiaa228>.
- [47] Y. Gao, L. Yan, Y. Huang, F. Liu, Y. Zhao, L. Cao, T. Wang, Q. Sun, Z. Ming, L. Zhang, J. Ge, L. Zheng, Y. Zhang, H. Wang, Y. Zhu, C. Zhu, T. Hu, T. Hua, B. Zhang, X. Yang, J. Li, H. Yang, Z. Liu, W. Xu, L.W. Guddat, Q. Wang, Z. Lou, Z. Rao, Structure of the RNA-dependent RNA polymerase from COVID-19 virus, *Science* (80-) 368 (2020) 779–782, <https://doi.org/10.1126/science.abb7498>.
- [48] E. de Wit, F. Feldmann, J. Cronin, R. Jordan, A. Okumura, T. Thomas, D. Scott, T. Cihlar, H. Feldmann, Prophylactic and therapeutic remdesivir (GS-5734) treatment in the rhesus macaque model of MERS-CoV infection, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 6771–6776, <https://doi.org/10.1073/pnas.1922083117>.
- [49] J. Grein, N. Ohmagari, D. Shin, G. Diaz, E. Asperges, A. Castagna, T. Feldt, G. Green, M.L. Green, F.X. Lescure, E. Nicastro, R. Oda, K. Yo, E. Quiros-Roldan, A. Studemeister, J. Redinski, S. Ahmed, J. Bennett, D. Chelliah, D. Chen, S. Chihara, S.H. Cohen, J. Cunningham, A. D'Arminio Monforte, S. Ismail, H. Kato, G. Lapadula, E. L'Her, T. Maeno, S. Majumder, M. Massari, M. Mora-Rillo, Y. Mutoh, D. Nguyen, E. Verweij, A. Zoufaly, A.O. Osinusi, A. DeZure, Y. Zhao, L. Zhong, A. Chokkalingam, E. Elboudwarej, L. Telep, L. Timbs, I. Henne, S. Sellers, H. Cao, S.K. Tan, L. Winterbourne, P. Desai, R. Mera, A. Gaggari, R.P. Myers, D.M. Brainard, R. Childs, T. Flanigan, Compassionate use of remdesivir for patients with severe Covid-19, *N. Engl. J. Med.* 382 (2020) 2327–2336, <https://doi.org/10.1056/NEJMoa2007016>.
- [50] P. Colson, J.M. Rolain, J.C. Lagier, P. Brouqui, D. Raoult, Chloroquine and hydroxychloroquine as available weapons to fight COVID-19, *Int. J. Antimicrob. Agents* 55 (2020), <https://doi.org/10.1016/j.ijantimicag.2020.105932>.
- [51] J. Liu, R. Cao, M. Xu, X. Wang, H. Zhang, H. Hu, Y. Li, Z. Hu, W. Zhong, M. Wang, Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro, *Cell Discov.* 6 (2020), <https://doi.org/10.1038/s41421-020-0156-0>.
- [52] Z. Zhu, Z. Lu, T. Xu, C. Chen, G. Yang, T. Zha, J. Lu, Y. Xue, Arbidol monotherapy is superior to lopinavir/ritonavir in treating COVID-19, *J. Inf. Secur.* 81 (2020) e21–e23, <https://doi.org/10.1016/j.jinf.2020.03.060>.
- [53] N. Lian, H. Xie, S. Lin, J. Huang, J. Zhao, Q. Lin, Umifenovir treatment is not associated with improved outcomes in patients with coronavirus disease 2019: a retrospective study, *Clin. Microbiol. Infect.* 26 (2020) 917–921, <https://doi.org/10.1016/j.cmi.2020.04.026>.
- [54] Q. Cai, M. Yang, D. Liu, J. Chen, D. Shu, J. Xia, X. Liao, Y. Gu, Q. Cai, Y. Yang, C. Shen, X. Li, L. Peng, D. Huang, J. Zhang, S. Zhang, F. Wang, J. Liu, L. Chen, S. Chen, Z. Wang, Z. Zhang, R. Cao, W. Zhong, Y. Liu, L. Liu, Experimental treatment with favipiravir for COVID-19: an open-label control study, *Engineering* (2020), <https://doi.org/10.1016/j.eng.2020.03.007>.
- [55] D.P. Misra, V. Agarwal, A.Y. Gasparyan, O. Zimba, Rheumatologists' perspective on coronavirus disease 19 (COVID-19) and potential therapeutic targets, *Clin. Rheumatol.* 39 (2020) 2055–2062, <https://doi.org/10.1007/s10067-020-05073-9>.
- [56] X. Cao, COVID-19: immunopathology and its implications for therapy, *Nat. Rev. Immunol.* 20 (2020) 269–270, <https://doi.org/10.1038/s41577-020-0308-3>.
- [57] X. Xu, M. Han, T. Li, W. Sun, D. Wang, B. Fu, Y. Zhou, X. Zheng, Y. Yang, X. Li, X. Zhang, A. Pan, H. Wei, Effective treatment of severe COVID-19 patients with tocilizumab, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 10970–10975, <https://doi.org/10.1073/pnas.2005615117>.
- [58] V. Kumar, T. Kashav, I. Hassan, Amyotrophic lateral sclerosis: current therapeutic perspectives, *Pathol. Prev. Ther. Neurodegener. Dis.*, 2018, pp. 207–224, https://doi.org/10.1007/978-981-13-0944-1_19.
- [59] J. Lung, Y.S. Lin, Y.H. Yang, Y.L. Chou, L.H. Shu, Y.C. Cheng, H. Te Liu, C.Y. Wu, The potential chemical structure of anti-SARS-CoV-2 RNA-dependent RNA polymerase, *J. Med. Virol.* 92 (2020) 693–697, <https://doi.org/10.1002/jmv.25761>.
- [60] A. Shamsi, T. Mohammad, S. Anwar, M.F. AlAjmi, A. Hussain, R. Md Tabish, A. Islam, H. Md Imtaiyaz, Glecaprevir and Maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy, *Biosci. Rep.* 40 (2020), <https://doi.org/10.1042/BSR20201256>.
- [61] A.M. Kanhed, D.V. Patel, D.M. Teli, N.R. Patel, M.T. Chhabria, M.R. Yadav, Identification of potential Mpro inhibitors for the treatment of COVID-19 by using systematic virtual screening approach, *Mol. Divers.* (2020), <https://doi.org/10.1007/s11030-020-10130-1>.
- [62] Z. Jin, X. Du, Y. Xu, Y. Deng, M. Liu, Y. Zhao, B. Zhang, X. Li, L. Zhang, C. Peng, Y. Duan, J. Yu, L. Wang, K. Yang, F. Liu, R. Jiang, X. Yang, T. You, X. Liu, X. Yang, F. Bai, H. Liu, X. Liu, L.W. Guddat, W. Xu, G. Xiao, C. Qin, Z. Shi, H. Jiang, Z. Rao, H. Yang, Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors, *Nature*. 582 (2020) 289–293, <https://doi.org/10.1038/s41586-020-2223-y>.
- [63] S. Khaerunnisa, H. Kurniawan, R. Awaluddin, S. Suhartati, Potential inhibitor of COVID-19 main protease (M pro) from several medicinal plant compounds by molecular docking study, *Preprints*. (2020) 1–14, <https://doi.org/10.20944/preprints202003.0226.v1>.
- [64] Q. Du, S. Wang, D. Wei, S. Sirois, K.C. Chou, Molecular modeling and chemical modification for finding peptide inhibitor against severe acute respiratory syndrome coronavirus main proteinase, *Anal. Biochem.* 337 (2005) 262–270, <https://doi.org/10.1016/j.ab.2004.10.003>.
- [65] Y. Han, P. Král, Computational design of ACE2-based peptide inhibitors of SARS-CoV-2, *ACS Nano* 14 (2020) 5143–5147, <https://doi.org/10.1021/acsnano.0c02857>.
- [66] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health, *Glob. Challenges* (Hoboken, NJ) 1 (2017) 33–46, <https://doi.org/10.1002/gch2.1018>.
- [67] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.* 7 (2011) 539, <https://doi.org/10.1038/msb.2011.75>.
- [68] T.L. Bailey, M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, W.S. Noble, MEME suite: tools for motif discovery and searching, *Nucleic Acids Res.* 37 (2009) 202–208, <https://doi.org/10.1093/nar/gkp335>.
- [69] J. Desmet, M. De Maeyer, B. Hazes, I. Lasters, The dead-end elimination theorem and its use in protein side-chain positioning, *Nature*. 356 (1992) 539–542, <https://doi.org/10.1038/356539a0>.
- [70] H. Kaur, A. Garg, G.P.S. Raghava, PEPstr: a de novo method for tertiary structure prediction of small bioactive peptides, *Protein Pept. Lett.* 14 (2007) 626–631, <https://doi.org/10.2174/0929886607781483859>.
- [71] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput.*

- Chem. 31 (2010) 455–461, <https://doi.org/10.1002/jcc.21334>.
- [72] G. Morris, R. Huey, AutoDock4 and AutoDockTools4: Automated Docking With Selective Receptor Flexibility, *J.* 30 (2009), pp. 2785–2791, <https://doi.org/10.1002/jcc.21256>.
- [73] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, *J. Chem. Inf. Model.* 51 (2011) 2778–2786, <https://doi.org/10.1021/ci200227u>.
- [74] K.J. Bowers, F.D. Sacerdoti, J.K. Salmon, Y. Shan, D.E. Shaw, E. Chow, H. Xu, R.O. Dror, M.P. Eastwood, B.A. Gregersen, J.L. Klepeis, I. Kolossvary, M.A. Moraes, Molecular dynamics—scalable algorithms for molecular dynamics simulations on commodity clusters, *Proc. 2006 ACM/IEEE Conf. Supercomput. - SC '06*, ACM Press, New York, New York, USA, 2006, p. 84, , <https://doi.org/10.1145/1188455.1188544>.
- [75] R. Schneider, A.R. Sharma, A. Rai, Introduction to molecular dynamics, *Lect. Notes Phys.*, 2008, pp. 3–40, , https://doi.org/10.1007/978-3-540-74686-7_1.