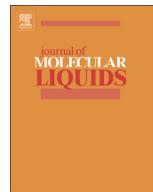




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



A computational study to disclose potential drugs and vaccine ensemble for COVID-19 conundrum



Sajjad Ahmad ^a, Yasir Waheed ^{a,*}, Saba Ismail ^a, Sumra Wajid Abbasi ^b, Muzammil Hasan Najmi ^a

^a Foundation University Medical College, Foundation University Islamabad, DHA-I, Islamabad 44000, Pakistan.

^b NUMS Department of Biological Sciences, National University of Medical Sciences, Abid Majid Rd, The Mall, Rawalpindi, Pakistan

ARTICLE INFO

Article history:

Received 16 September 2020

Received in revised form 1 November 2020

Accepted 5 November 2020

Available online 10 November 2020

Keywords:

SARS-CoV-2

Nucleocapsid (N) protein

COVID-1

Drugs

Vaccine

ABSTRACT

The nucleocapsid (N) protein of SARS-CoV-2, a virus responsible for the current COVID-19 pandemic, is considered a potential candidate for the design of new drugs and vaccines. The protein is central to several critical events in virus production, with its highly druggable nature and rich antigenic determinants making it an excellent anti-viral biomolecule. Docking-based virtual screening using the Asinex anti-viral library identified binding of drug molecules at three specific positions: loop 1 region, loop 2 region and β-sheet core pockets, the loop 2 region being the most common binding and stable site for the bulk of the molecules. In parallel, the protein was characterized by vaccine design perspective and harboured three potential B cell-derived T cell epitopes: PINTNSSPD, GVPINTNSS, and DHIGTRNPA. The epitopes are highly antigenic, virulent, non-allergic, non-toxic, bind with good affinity to the highly prevalent DRB*0101 allele and show an average population coverage of 95.04%. A multi-epitope vaccine ensemble which was 83 amino acids long was created. This was highly immunogenic, robust in generating both humoral and cellular immune responses, thermally stable, and had good physicochemical properties that could be easily analyzed in vivo and in vitro studies. Conformational dynamics of both drug and vaccine ensemble with respect to the receptors are energetically stable, shedding light on favourable conformation and chemical interactions. These facts were validated by subjecting the complexes to relative and absolute binding free energy methods of MMGB/PBSA and WaterSwap. A strong agreement on the system stability was disclosed that supported ligand high affinity potential for the receptors. Collectively, this work sought to provide preliminary experimental data of existing anti-viral drugs as a possible therapy for COVID-19 infections and a new peptide-based vaccine for protection against this pandemic virus.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

The ongoing COVID-19 pandemic caused by SARS-CoV-2 is a highly infectious disease posing a severe threat to worldwide public health [1–4]. SARS-CoV-2 has the incredible potential of human to human spreading, contributing to rapid global dissemination [5,6]. To date, the virus has been reported to cause 24,357,067 cases with 830,150 deaths and 16,890,125 recoveries. Despite tremendous efforts done on mitigating the virus transmission and developing therapeutics in almost every country around the globe, we still have no specific treatment and cure yet [6–9]. This prompted us to devise highly accepted *in silico* approaches with the ultimate aim to assist in the rapid design of new classes of drugs and vaccines for SARS-CoV-2.

SARS-CoV-2 is a single-stranded 30 kb long RNA genome virus [6,10,11]. The main region of the genome, named ORF1a/b, covers 2/3 of the length and codes nonstructural proteins (nsp) [12]. The remaining

genome encodes four essential structural proteins which include a small envelope (E) protein, a matrix (M) protein, a nucleocapsid (N) protein, and a spike (S) glycoprotein [6,13,14]. The current coronavirus anti-viral regime primarily targets 3c-like (3CL) protease, papain-like (PLP) protease, and the S protein [15,16]. As protease inhibitors may act nonspecifically on the host homologous protease, there is a risk of associated host cell toxicity and can induce severe side effects. The S protein is highly vulnerable to mutations, enabling this protein to acquire a different pattern of host cell receptor binding, which in turn, aid in the protein escape from targeted therapeutics [15]. Considering this, novel strategies are obligatory to curtail infections caused by SARS-CoV-2.

The N protein of SARS-CoV-2 binds to leader RNA and plays several pivotal roles in RNA transcription and replication, and thus is considered an attractive pharmacological target [17–19]. Its primary function is to produce a ribonucleoprotein (RNP) complex, key to the formation of highly ordered RNA conformation essential for viral RNA replication and transcription and modulating metabolism of infected cells [19]. Additionally, this protein regulates host-pathogen interactions that involve reorganization of actin, the progression of the host cell cycle and

* Corresponding author at: Foundation University Medical College, Foundation University Islamabad, Islamabad, Pakistan.

E-mail address: yasir.waheed@fui.edu.pk (Y. Waheed).

apoptosis [20]. From an architectural point of view, the protein is composed of three distinct but highly conserved portions; the C-terminal domain (CTD), the N-terminal domain (NTD), and Ser/Arg (SR) rich linker [17]. CTD functions as a dimerization domain, whereas NTD and SR are responsible for binding to RNA and direct phosphorylation, respectively [6,21]. The crystal structure of the NTD from SARS-CoV-2 (PDB ID: 6M3M) has been found to interact with 3' end of the SARS-CoV-2 genome through several key residues mediating infectivity and give specific electrostatic distribution [17]. Such structural information is essential for accelerating drug discovery against this appealing drug target to block SARS-CoV-2 production. Furthermore, the N protein has high expression during infection, is highly immunogenic and has the potential to induce protective immune responses targeting SARS-CoV-2 [22,23].

Herein, we adopted a comprehensive in silico methodology to identify potent anti-viral leads, protective antigens and diagnostic markers against SARS-CoV-2 N protein. Findings of this study may promote the discovery of new anti-viral drugs and vaccination strategies against this high priority virus.

2. Materials and methods

The complete methodology used in the present work is summarized in Fig. 1.

2.1. SARS-CoV-2 N protein preparation

The study commenced with the retrieval of the crystal structure of N protein from the PDB database available with a PDB tag of 6M3M [17]. The structure was determined by X-ray diffraction up to a resolution of 2.70 Å. The 3D structure was then treated in UCSF Chimera [24] for structure editing to optimize receptor energy. During the process, the

protein first underwent processing using a steepest descent algorithm keeping the step size to 0.02 Å and number of cycles to 100. Afterwards, the conjugate gradient algorithm was applied on the structure for the default ten steps. Both algorithms tended to clean the structure by improving local interactions, fine-tuning widespread structural errors and moving the structure towards local minima. Next, missing hydrogen atoms were added to the receptor and charges to standard and non-standard residues were assigned by mean of AMBER ff14SB [25] and AM1-BCC [26], respectively. Both initial crystal N protein and post-treated minimized structures were examined using a PDBSum PROCHECK analysis [27] to assist in the selection of starting receptors in a high throughput anti-viral scaffold screening process.

2.2. Ligands preparation

The Asinex anti-viral library delivers a meaningful starting point by arranging chemical entities of potent anti-viral activities and good safety profile for the discovery of new powerful leads. The compounds are also easy to access and purchasable for testing in experimental assays. The anti-viral library was retrieved in SDF format and subsequently filtered in PyRx software [28] to select compounds that fulfil the Lipinski rule of five [29]. As per this rule, only compounds with molecular weight ≤ 500 Da, MlogP ≤ 4.15, N or O ≤ 10, and NH or OH ≤ 5 are selected. The primary library had 6827 compounds that, after deletion of molecules violating Lipinski rule of five, reduced to 4860. This new list of compounds was minimized for energy minima and converted to .pdbqt format to be ready for docking study.

2.3. Uncovering potential epitopes for SARS-CoV-2 N protein

The amino acid sequence of the SARS-CoV-2 N protein was scanned for potential B and T cell epitopes capable of evoking strong but

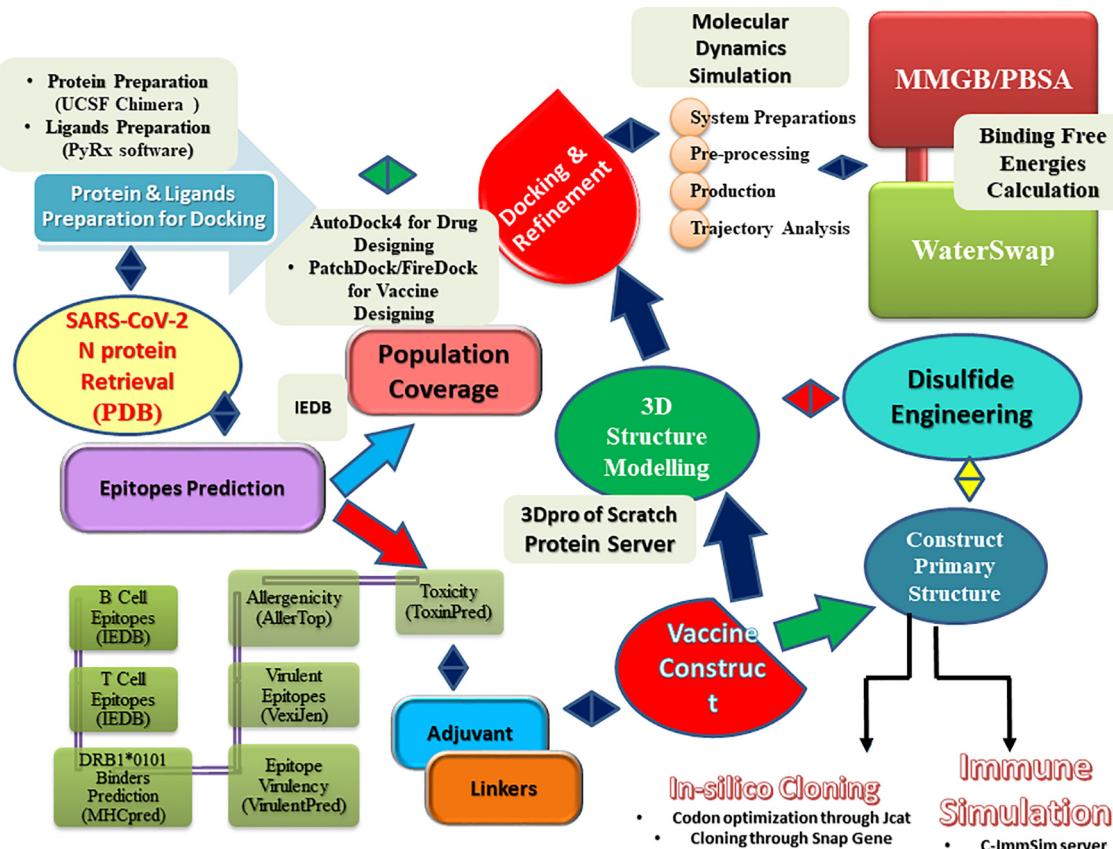


Fig. 1. Flow of steps used in the current study.

protective immunological responses. To accomplish this objective, we employed the Immune Epitope Database (IEDB) [30] where first linear B cell epitopes were predicted using Bepipred Linear Epitope Prediction 2.0 setting cutoff score of >0.5. In parallel, T cell epitopes were also predicted starting with MHC-I alleles using IEDB recommended 2020.04 (NetMHCpanEL 4.0) considering a reference set of alleles (S-Table 1). The epitopes were sorted on IC50 score basis, and those with score < 500 nM were selected. The MHC-II alleles prediction was accomplished by selecting IEDB 2.22 method and full HLA reference set (S-Table 1). Each of the predicted epitopes was then evaluated for their potential of evoking immunological response by scanning the epitopes in Vaxijen (cut off antigenic score ≥ 0.4) [31]. Next, the shortlisted immunodominant candidates were checked for allergenicity via AllerTOP version 2.0 [32]. Non-allergens were then toxicity filtered using ToxinPred [33] to opt for non-toxic epitopes. Additionally, the non-homologous and virulent nature of the epitopes was verified by using the epitopes in Blastp search against human (Homosapien: 9606) and VirulentPred [34], respectively. Lastly, IFN-gamma inducing epitopes were filtered via IFNepitope web server [35].

2.4. Epitope conservation and population coverage

All entries of the SARS-CoV-2 N protein available in the NCBI COVID-19 datahub were retrieved and used to examine predicted IFN epitopes conservation through an online IEDB conservancy analysis tool [36]. It was important to include population coverage of the epitopes in the study as it gave clear directions about the percentage of a specific population likely to respond to the epitopes. This was achieved by employing the IEDB epitope conservation analysis tool [37].

2.5. Construction of vaccine ensemble

The final set of conserved epitopes that provided the greatest population coverage were fused to each other, followed by addition of an adjuvant molecule to the epitope peptide. The 3D structure of the full ensemble was created ab initio using 3Dpro [38]. Loops of the model structure were refined to strengthen structural stability using GalaxyRefine version 2 [39]. Quality assessment of the ensemble 3D model was made employing free available tools: Ramachandran plot of PDBsum [27], ERRAT [40], VERIFY-3D score [41], PROSA Z-score [42]. Then, the structure was minimized to the local energy minima and passed through downward analysis of vaccine design. Different physicochemical parameters of the vaccine were predicted using

ProtParam [43]. Solubility and aggregation-prone regions of the vaccine were predicted using Protein-Sol [44] and Aggescan3D 2.0 [45], respectively.

2.6. MHC clustering analysis and immune simulation of the vaccine ensemble

During the vaccine design process, identification of candidates showing an optimized affinity for a wide range of MHC HLA alleles is fundamental. The MHC clustering analysis was performed using MHCcluster v2.0 [46]. Further, the immune response profile of the vaccine construct was understood using an agent based immune simulator - the C-ImmSim server [47]. The server employed a position specific scoring matrix to spot immune dominant epitopes and machine learning methods to elucidate immune interactions. Most of the simulation parameters were treated as default: adjuvant = 100, number of antigen injection = 1000, random seed chosen = 12,345, and the vaccine was injected with non LPS. The 1000 units of antigen are considered suitable to induce an appropriate immune response to the viral antigen [48]. The simulation steps allowed are 1100 and volume to 110. C-ImmSim server has been successfully applied in several studies to understand host immune system dynamics in response to an antigen [49–53].

2.7. Molecular docking of antiviral ligands and vaccine ensemble

A blind docking approach was applied for docking both anti-viral ligands with the N protein and vaccine ensemble with TLR3 (PDB tag: 1ZW). In the case of anti-viral ligands, docking was performed with AutoDock4 [54] allowing the central XYZ dimensions search to be restricted to 9.9921 Å on the X-axis, −3.8536 Å on the Y-axis and −12.7487 Å on the Z-axis. This gives dimensions on the XYZ plane as 31.6231 Å, 46.5257 Å, and 43.2407 Å, respectively, and as net, the whole surface of the receptor molecule was covered to allow ligand molecules to bind freely to the hotspot points of the N protein. Each ligand molecule was docked 10 times to the receptor and the best binding pose was selected by looking for the one with best binding affinity score in kcal/mol (more negative score indicates good binding affinity). The vaccine ensemble docking with innate immune receptor (as a test case here we used TLR3) was assessed in Patchdock [55], the generated complexes were refined with FireDock [56], and the best complex with minimum global energy was considered for visualization and assay [57–60]. Both AutoDock4 and FireDock are best for performing blind docking and generating intermolecular poses that bind best to each

Table 1

Immunological analysis of SARS-CoV-2 N protein.

B cell	T Cell	MHC I	MHC II	MHCpred	Antigenicity	Allergenicity	Toxicity
GSRPQGLPNNTAS	GSRPQGLPN	1.6	48	93.54	0.3955 (Non-antigen)	Non-allergen	
	RHQGLPNNT	0.35	22	43.55	0.5758 (Antigenic)	Allergen	
	QLPNNTAS	0.28	7.4	79.8	0.0743 (Non-antigen)	Allergen	
QHGKEDLKPRGQGVINTNSSPDDQIG	PINTNSSPD	15	15	5.6	0.5204 (Antigenic)	Non-allergen	Non-toxic
	HKGEDLKFP	2	31	20.75	1.1473 (Antigenic)	Allergen	
	QGVINTNS	11	15	23.12	0.3642 (Non-antigenic)	Allergen	
	TNSSPDDQI	2.3	17	33.19	0.4559 (Antigenic)	Allergen	
	INTNSSPDD	2.6	22	34.99	0.4458 (Antigenic)	Allergen	
	KFPRQQGV	0.04	23	46.99	−0.1527 Non Antigenic	Allergen	
	KEDLKPRG	3.2	35	47.75	−0.0113 (Non-antigenic)	Allergen	
	GKEDLKFP	2	35	60.39	0.7725 (Antigenic)	Allergen	
	FPRCQQGVPI	0.04	11	61.24	0.7585 (Antigenic)	Allergen	
	GVPINTNS	9	15	62.09	0.7585 (Antigenic)	Non-allergen	Non-toxic
	DLKFPRGQG	7.6	16	69.18	−0.0028 (Non-antigenic)	Non-allergen	
RIRGGDGKMKDL	RIRGGDGKM	0.42	20	58.48	0.1803 (Non-antigenic)	Non-allergen	
TGPEAGLPHYGANK	TGPEAGLPHY	0.56	2.7	40.83	−0.0349 (Non-antigenic)	Non-allergen	
GALNTPKDHGTRNPANN	DHICTRNPA	7.8	12	20.23	0.4313 (Antigenic)	Non-allergen	Non-toxic
	IGTRNPA	8.2	12	30.69	0.3555 (Non-antigenic)	Allergen	
	LNTPKDHIG	14	52	83.75	−1.2393 (Non-antigenic)	Allergen	
	ALNTPKDH	0.87	37	86.1	−0.8423 (Non-antigenic)	Allergen	

other, hence achieving highly stable complexes. Both sets of software are the most citable forms of docking software, widely used and are freely available [56,61–65]. A blind docking approach was used in the present work by providing the complete surface of receptors (N protein in case of drug molecules identification and TLR3 in case of vaccine ensemble docking). This overcomes the limitations of the specific docking that, in majority cases gives false-positive results [66]. Additionally, we proved the docking procedure by docking the co-crystallized ligands to the receptors through the same procedure used in virtual screening of Asinex antiviral library and vaccine ensemble docking to the TLR3 [67]. Both docking protocols results revealed coherent results, thus validating the docking protocol. Further, we employed widely accepted and more accurate molecular dynamics simulation and binding free energies methods to validate the good affinity of the drug molecules and vaccine ensemble [68–70].

2.8. Molecular dynamics simulation assay

Molecular dynamics simulations for both top complexes of anti-viral ligand with N protein and vaccine ensemble with TLR3 were performed using AMBER18 package [71]. The N protein parameters were generated using a ff14SB force field [25] whereas an Amber force field (GAFF) [72] was chosen for anti-viral ligands. To record topology files of both complexes, the leap module [73] was employed. The systems were neutralized by adding an appropriate number of counterions. Next, both systems were submerged in a TIP3 water box, allowing a padding distance of 12 Å. The waterbox with submerged N protein-drug complex and TLR3-vaccine complex are depicted in Fig. 2. Systems minimization was achieved by running 1500 rounds of steepest descent and conjugate gradient to clean the complexes for unfavorable structural clashes. Heating of systems was done for 100 ps with a gradual increase from 0 K to 300 K, applying pressure of 1 atm. Afterheat, systems were equilibrated for a time period of 100 ps at a constant temperature of 300 K. A production run for each system was completed at time scale of 100 ns. In the process, SHAKE algorithm [74] was applied to constrain all covalently bonded hydrogen atoms of the systems. Periodic boundary conditions were used in the solvation box by the canonical ensemble. Temperature was kept constant at 300 K using a Langevin thermostat [75] and the non-bounded interaction threshold was treated as 8.0 Å. Ewald simulations were utilized for long range interactions. Structural dynamics of the complexes were elucidated through several statistical parameters analyzed through the CPPTRAJ module [76] of the AMBER. Visualization of the snapshots was done by means of UCSF Chimera [24] and VMD software [77].

2.9. Binding free energy estimation

Estimation of binding free energies for biomolecular complexes is a good way of validating intermolecular strength of interactions and shedding light on the dominancy of a particular chemical energy contributing to overall stability. The MMPBSA method of AMBER is an easy and relatively straight forward approach for quantifying binding free energies of ligand(s) docked to a receptor, though this method does not account for entropy contribution [78]. The MMPBSA binding free energy was estimated using equation given below,

$$\Delta G_{\text{binding}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) = \Delta G_{\text{MM}} + \Delta G_{\text{solv}}$$

where,

ΔG is the net binding free energy for a given system by subtracting the sum of receptor and ligand combine binding energy from the complex. ΔG_{MM} is the gas phase energy change estimated by molecular mechanics and consists of van der Waals and electrostatic energy. ΔG_{solv} is the solvation free energy change and is the product of polar and non-polar energy. The latter term is calculated via the solvent accessible surface area (SASA).

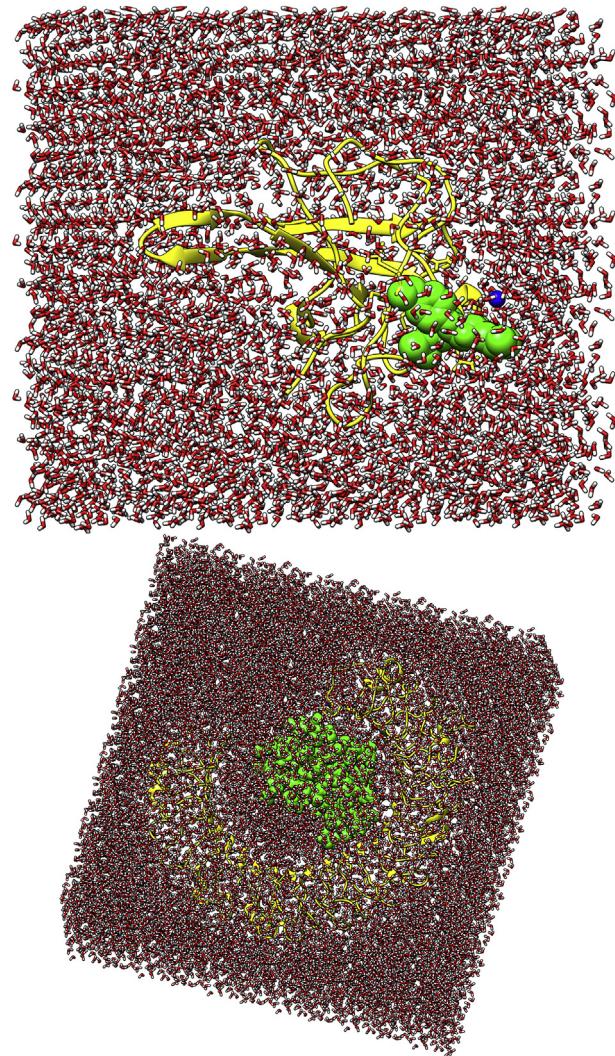


Fig. 2. Waterbox presentation of drug (top) and vaccine ensemble complex with respective receptors (colored by yellow cartoon). The drug and vaccine construct is shown by a green sphere.

2.10. Waterswap absolute protein-ligand binding free energies

The Waterswap method implemented in the Sire package permits estimation of absolute protein-ligand binding free energies [79]. The reaction coordinate is constructed during the process which swaps the protein bound ligand to an equivalent volume of water present in the binding pocket. This method uses all available processing cores of a computer node, slow in-process and converges the free energy averages that took at least five days. The average binding free energy is estimated simultaneously through thermodynamic integration (TI), free energy perturbation (FEP) and Bennetts Acceptance Ratio (BAR) methods [70]. Waterswap was run on simulation trajectories of last 10 ns to estimate the absolute binding free energy of the systems.

3. Results and discussion

3.1. SARS-CoV-2 N protein and anti-viral library minimization

A primary phase of protein minimization was applied to SARS-CoV-2 N protein to lower its overall potential energy. This was necessary to make protein conformation as close as possible to natural biological systems that are dynamic and low in potential energy to ease spontaneous

interactions. However, a minimization event may introduce bad contacts in the structure and disturb the conformation of the molecule, which in turn might affect the compound ranking in the docking-based virtual screening process. The minimization process revealed that minimized N protein is a bit better compared to the pre-minimized original N protein structure. Ramachandran plot investigation showed that both original and energy minimized structures have the same percentage of residue distribution in all four quadrants as can be seen in S-Fig. 1. Statistically, both structures secured 89.4% and 10.6% of residues in the most favoured and additionally allowed regions whereas no residue was plotted in generously allowed and disallowed regions. According to the G-factor assay, the minimized protein had a better overall G-factor score of 0.06, reflecting no unusual features in the structure opposed to the original structure (G-factor score of -0.14). Secondly, the PROSA Z-score was calculated which indicated overall good quality and a non-erroneous structure of the minimized N protein. The Z-score for pre-minimized N protein was -5.06 while it was -5.16 for the minimized form. Based on this evidence, we used the minimized N protein structure in the downward framework.

3.2. Virtual screening

High throughput structure-based virtual screening was performed using SARS-CoV-2 N protein as a receptor and drug-like molecules from the Asinex anti-viral library as ligands in the process. As a blind docking strategy was employed, the entire surface of the protein was exposed for binding of the drug molecules. The molecules bound to three different binding pockets of loop region 1, β -sheet core, and loop region 2 of the enzyme as shown in Fig. 3. The majority of the molecules tended to interact with the pockets of the loop 2 region compared to the loop 1 region and β -sheet core pockets. Top 5 complexes were picked based on the binding affinity of the molecules for the pockets and analyzed. The 2D structures of the compounds are provided in S-Fig. 2. The high affinity top 1 compound with binding energy of -8.9 kcal/mol demonstrated binding to the front pocket of a loop 2 region. This pocket is the result of the long loop of β 6 connecting to β 7, and β 2 ending at the N-terminal site through β 1. The bulk of the compound

structure (2,5,5,7,9-pentamethyl-2,3,4,4a,5,10b-hexahydropyrano[3,2-c]chromene-3-carboxylic acid) aligned horizontally inside the long loop originating from β 6 of the β -sheet core, covering β 10 helix at the acidic wrist of the protein and ending at the carboxyl site. This chemical moiety formed weak van der Waals, pi-pi stacked, and pi-alkyl interactions with the loop of β 6. The 4-isopropylphenol moiety was deep in the cavity, making both hydrogen and hydrophobic contacts with the loops. Compound 3 (binding energy: -8.6 kcal/mol) and 5 (binding energy: -8.2 kcal/mol) followed the same pattern of binding with minimal interaction from strong hydrogen bonding and maximum interaction from weak hydrophobic binding. Compound 2 (binding energy: -8.7 kcal/mol) binding mode was stable at loop region 1 and produced a rich network of hydrogen and van der Waals contacts. The 1-(3-(1,2,4-triazolidin-4-yl)phenyl)ethanone part of the molecule mainly drove the interactions of strong hydrogen bonding with Thr92 and Arg94 of β 2 and Asp129 of β 6 loop. Compound 4 (binding energy: -8.4 kcal/mol) favoured binding to the back pocket of loop region 2, collectively formed by loops of β 2, β 5 and β 6. The 4-formyl-6-methyl-3-methylene-2,3-dihydroisoxazolo[5,4-b]pyridin-7-iun chemical moiety made hydrogen bonds with Pro74 and Asn76 of β 2. The rest of the chemical portion reacted with all three loops through van der Waals bonding. The chemical interactions of the compounds are illustrated in Fig. 4.

3.3. SARS-CoV-2 N protein epitopes prediction

A complete protein sequence was first analyzed for continuous B cell epitopes that predicted 6 epitopes ranging in length from 7mer to 28mer of score greater than default cut off (0.5). Such B cell epitopes are the recognition and binding sites of the adaptive immunity B cells. Once activated, the B cell transformed into a mature form, differentiate and produce soluble antibodies. Afterward, the antibodies bind to the epitopes and activate humoral adaptive immunity by triggering formation of neutralizing toxins, labeling the pathogen for destruction via T cell immunity. Therefore, B cell epitopes mapping holds a central role in vaccine designing [80]. The predicted B cell epitopes were examined for T cell epitopes. T cell vaccine induces protective cellular immunity and are significant in targeting mutating viruses like SARS-CoV-2.

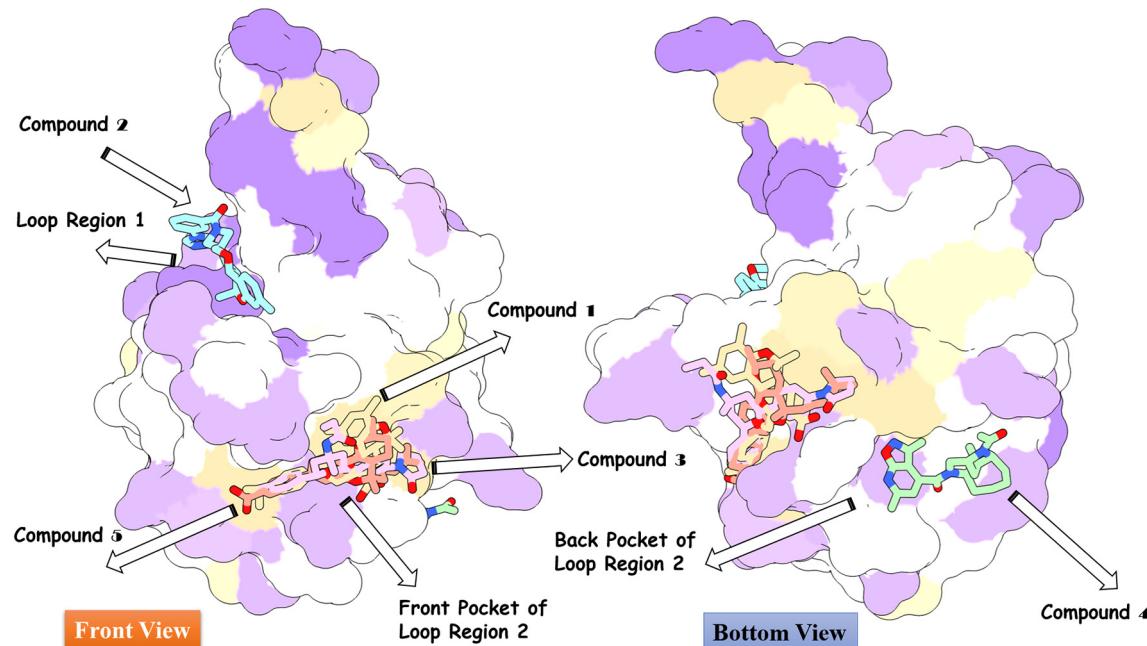


Fig. 3. Docked pose of different classes of anti-viral inhibitors at different pockets of the SARS-CoV-2 N protein. The color order of the top 5 ligands is in the following order: top 1 (tan), top 2 (cyan), top 3 (pink), top 4 (aquamarine), and top 5 (deep pink). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

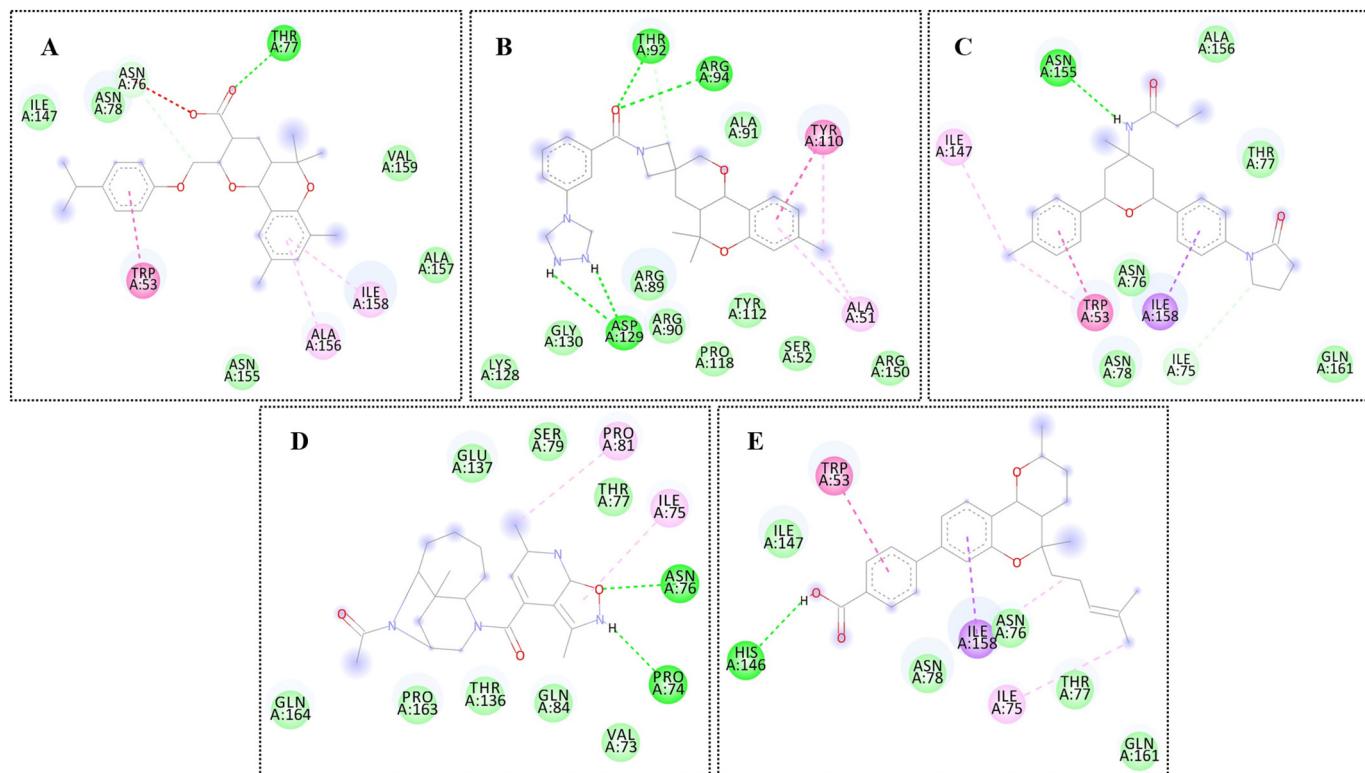


Fig. 4. Chemical network of interaction of the compounds at respective docked sites. From A to E tags represents compound 1 to 5, respectively. The different color of discs can be interpreted as residues of the N protein and can be understand as: dark green discs (hydrogen bonding residues), light green discs (van der Waals residues), pink discs (pi-pi stacked residues), purple discs (pi-sigma residues), and cream discs (alky and pi-alkyl residues). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Furthermore, T cell vaccines are regarded more effective than conventional B cell epitopes [81,82]. In the present investigation, both B and T cell epitopes were used to design highly efficacious vaccines. Antigens that are displayed on the antigen presenting cells bound to the major histocompatibility complex molecules are recognized by the T cell receptors present on the surface of T cells. T cell epitopes are surface epitopes via two classes: MHC-I and MHC-II, identified by two distinct T

cytotoxic and T helper cells. Several T cell epitopes are reported for each B cell epitopes with IC_{50} values <50 nM depicting high affinity for the reference set of MHC alleles. Next, the T cell epitopes were subjected to MHCphred assay to filter only those epitopes that associated with great affinity with the DRB*0101 allele, a prominent allele in the human population [83]. All T cell epitopes showed great binding with the DRB*0101 allele with an IC_{50} value of <100 nM. T cell epitopes

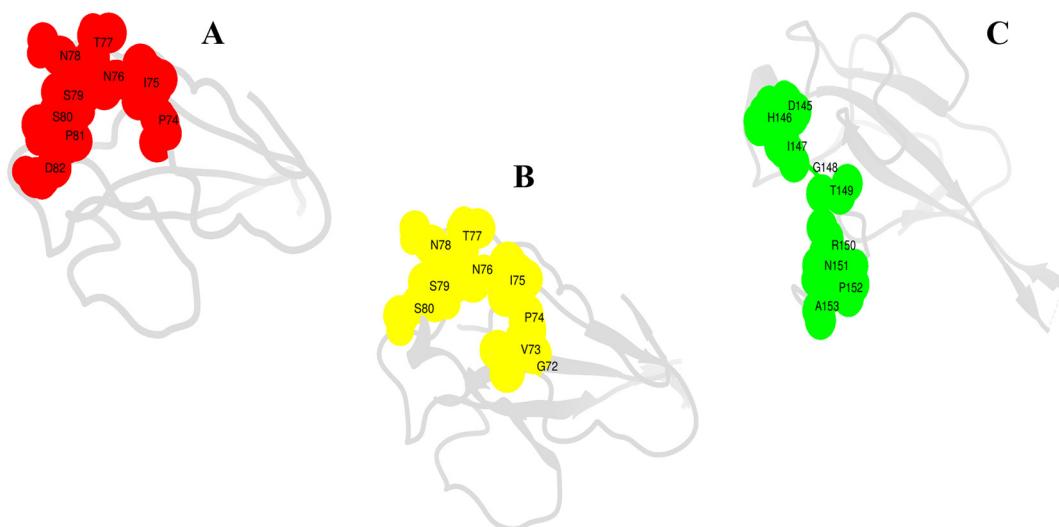


Fig. 5. Surface analysis of final set of T cell epitopes localized at the surface of N protein. A (PINTNSSPD), B (GVPINTNSS) and C (DHIGTRNPA). The amino acids are presented by single letter code.

then analyzed for their ability of provoking the host immune system. This was achieved by passing the T cell epitopes through the Vaxijen server and selecting those with predicted score higher than the default 0.4. Antigenic T cell epitopes were subsequently filtered through allergenic and toxicity check, and only non-allergenic and non-toxic epitopes were picked. The final set of potential epitopes selected for vaccine ensemble design is listed in Table 1 and their *exo*-membrane topology is shown in Fig. 5.

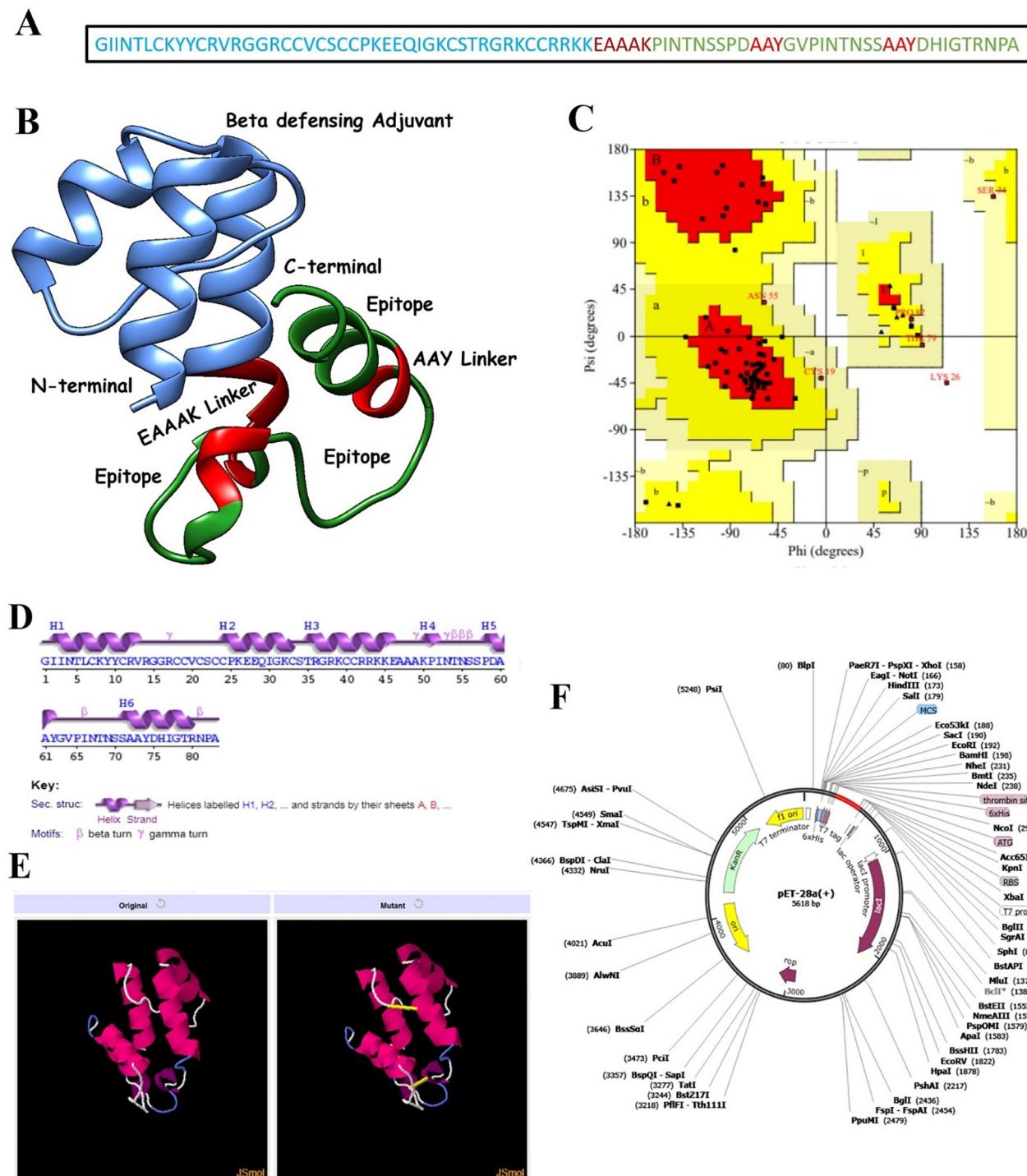


Fig. 6. A. Schematic representation of the vaccine, B. 3D structure of the vaccine, C. Ramachandran plot of the vaccine, D. Secondary structure elements of the vaccine, E. Disulfide engineering of the vaccine; both original and mutated structures are presented, F. In silico cloning of the vaccine (shown as red) into pET-28a(+) vector.

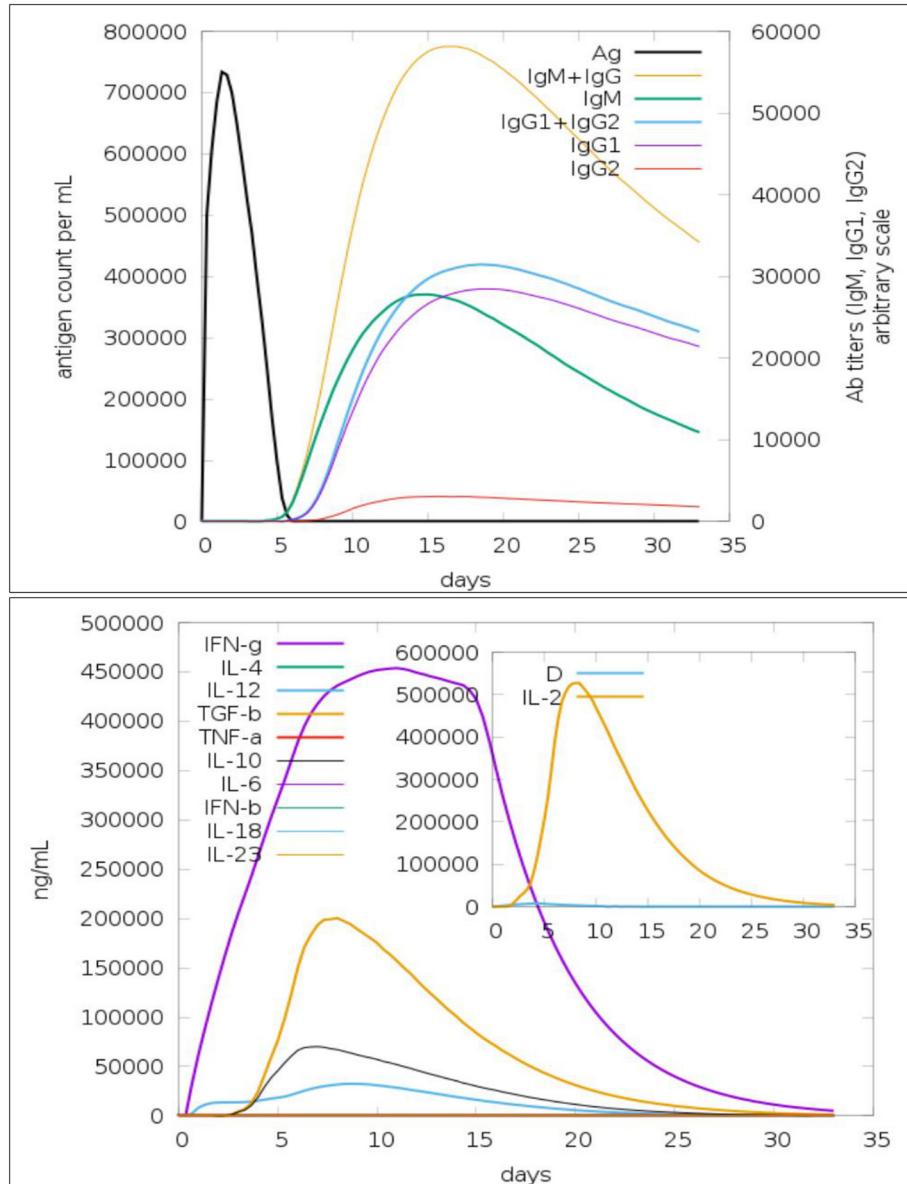


Fig. 7. Host immune system simulation in response to the vaccine ensemble antigen; production of different immunoglobulin response (top) and cytokines and interleukins with Simpson index (bottom).

cell epitopes are showing average population coverage of 95.04%. This can be divided up as: Central Africa (94.79%), Central America (53.8%), East Africa (97.08%), East Asia (97.08%), Europe (99.96%), North Africa (99.01%), North America (99.89%), Northeast Asia (97.88%), Oceania (97.88%), South Africa (95.27%), South America (95.15%), South Asia (98.7%), Southeast Asia (97.66%), Southwest Asia (95.79%), West Africa (98.43%), and West Indies (99.69%). The number of combined HLA alleles recognized and cumulative percent of population coverage of the epitopes are given in S-Fig. 3.

3.5. Construction of a chimeric multi-epitope peptide vaccine

Peptide-based vaccines are weakly immunogenic, therefore, the final set of epitopes were fused with the help of rigid AYY linkers to produce a multi-epitope peptide [84,85] (Fig. 6A). AYY linkers aid in keeping the epitopes separated and aid in easy recognition and processing of the epitopes by the host immune system. Also, an adjuvant in the form of β -defensin was added to the N-terminal site of the epitope peptide for further strengthening of the immune provoking ability of the

epitope's peptide. β -defensin activates lymphokines production which in turn results in Ig production specific to the antigen along with activation of cellular immunity [86]. Ligation of the adjuvant with the multi-epitope peptide was done through a rigid EAAAK linker. This linker is also rigid and promotes easy immune recognition and processing. The complete construct structure was modelled ab initio as no suitable template was available for homology-based modelling (Fig. 6B). The structure has the majority of its residues (80%) plotted in the most favoured region, 12.90% in the additional allowed region, 5.70% in the generously allowed region, and 1.40% in the disallowed regions (Fig. 6C). From secondary structure point of view, the structure maximum of alpha helix (53.0%), 9.6% of 3–10 helix, and 37.3% of beta turns, and gamma turns (Fig. 6D). The overall vaccine ensemble is antigenic with a score of 0.5060, soluble with a probability score of 0.940 and does not harbour any transmembrane helices; thus it is a good candidate for experimental follow up. The weight of the construct is 8.9 KDa, theoretical PI of 9.36, stability index of 39.9, GRAVY score of -0.602 and aliphatic index of 53.01. The vaccine ensemble is also predicted to be highly soluble with predicted scaled solubility value of

0.775 (threshold, 0.45) and does not contain any aggregation-prone regions. The average aggregation score of the vaccine is -1.11 , which is far less than the cutoff score of 0.0.

3.6. MHC restricted HLA alleles cluster analysis

The vaccine ensemble molecule interacted with a large numbers of HLA alleles in MHC I class compared to MHC II class. This infers the increased potential of the designed vaccine ensemble to be recognized by the vast majority of polymorphic HLA alleles, maximizing many fold epitopes presentation the immune cells leading to providing strong immune protection capabilities. The MHC I and MHC II HLA clusters are tagged as S-Fig. 4 and S-Fig. 5, respectively.

3.7. Loop modelling and refinement of the vaccine ensemble

Loop regions of the vaccine ensemble were modelled/refined to present its closest possible structure for the downward analysis. The structure was refined in consecutive rounds covering regions: Arg14-Ser34, Asn53-Ser57, Gly63-Ser70, Tyr74-Ala83, and Ala47-Ala48. The output structure was further refined for global and local conformations. An improved vaccine ensemble structure was revealed in which the percent of residues in the Rama favoured regions was increased to 96.3% compared to 93.8% in the original model. The clash error score in the original structure was 29.4 and was found only 2.2 in the refined model. The MolProbity score is relatively low (1.241) illustrating the good quality of the refined model. Overall, the galaxy energy of the refined vaccine ensemble is -1744.48 which is significantly less than the initial structure (132.01), a confirmation of good stability.

3.8. Immune responses simulation of the vaccine ensemble

An in silico immune response model of the vaccine ensemble was generated that illustrated primary, secondary and tertiary immune responses are credible in clearing the pathogen (Fig. 7). The IgM + IgG is critical in showing an immune protection response to the antigen whereas IgG2 is the least productive. Similarly, significant interleukin and cytokine reactions are witnessed, IFN-gamma being the most key immune protective factor (Fig. 7). Formation of different isotypes of the B cell immunity to the antigen demonstrates a fundamental role of humoral immunity against the pathogen and subsequent creation of antigen memory. Increased production of cytotoxic and helper T cell populations with corresponding memory formation further affirm the role of T cell immunity complementing B cell immunity in protecting the host from the pathogen. A high population of dendritic cells, as well as that of macrophages, makes it evident that the set of epitopes used in the construction of multi-epitope vaccine ensembles is attractive for activation of all components of the host immune system aiding in combatting the pathogen.

Table 2
Top 10 refined TLR3-vaccine solutions ranked on basis of global energy in kJ/mol.

Rank	Solution Number	Global Energy	Attractive Van der Waals Energy	Repulsive Van der Waals Energy	Atomic Contact Energy	Hydrogen bonding Energy
1	9	-7.93	-33.10	14.31	15.51	-3.66
2	7	-2.54	-31.03	22.05	14.68	-6.93
3	3	0.83	-7.43	1.96	2.17	-1.19
4	2	1.88	-14.95	2.48	14.70	-5.13
5	6	6.71	-2.19	0.00	3.38	0.00
6	10	17.93	-7.56	13.93	9.32	-2.48
7	5	32.79	-43.26	90.77	14.74	-3.88
8	8	45.32	-18.23	6.96	15.53	-1.49
9	4	123.17	-11.32	142.99	4.43	-2.45
10	1	200.68	-14.32	256.72	5.08	-0.91

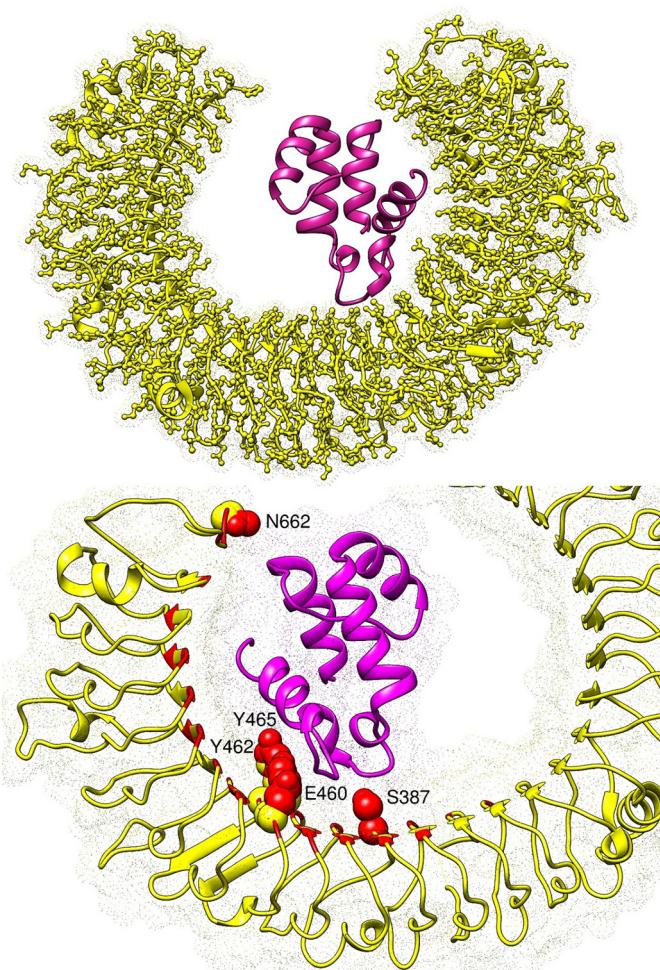


Fig. 8. The top figure presents docked conformation of the vaccine ensemble (shown in dark maroon) at the active pocket of TLR3 (yellow cartoon surface as dot). The bottom figure illustrates closed view of the vaccine interacting with TLR3 residues (shown in red bubbles) via hydrogen bonding. The rest of red highlighted regions in sticks are those involved in van der Waals interactions with the TLR3 residues.

3.9. Vaccine ensemble docked conformation with TLR3

The vaccine ensemble was docked with TLR3, an immune receptor from innate immunity that prompts viral recognition and induces type I interferon production. Blinded protein-peptide docking was performed to predict the predominant orientation of the vaccine ensemble with respect to the TLR3 receptor. In total, 100 conformations of the vaccine ensemble were generated, followed by refinement of each complex to opt for the highly stable conformation. Different energy values for the top 10 refined solutions of the docked vaccine ensemble and TLR3 complexes are presented in Table 2. Solution 9 was ranked as the top solution due to global energy of -7.93 kJ/mol (attractive van der Waals energy: -33.10 kJ/mol, repulsive van der Waals energy: 14.31 kJ/mol, atomic contact energy: 15.51 kJ/mol, and hydrogen bonding energy: -3.66 kJ/mol). Visual interpretation of solution 9 unveils deep central docking of the vaccine ensemble at the TLR3 cavity. The vaccine generates a network of hydrophobic and hydrophilic chemical interactions with TLR3 residues. Ser387, Glu460, Tyr462, Tyr465, and Asn662 are hydrogen bonding contributing residues from the TLR3 with Ser56, Thr68, Asp59, Tyr62, and Arg12 of the vaccine, respectively. Van der Waals interactions from the TLR3 that played a significant role in binding and anchoring of the vaccine are: Lys330, Glu363, Tyr383, Ile411, Thr415, Lys416, Val435, Phe459, Arg484, Arg488, Ile534, His565, Ile566,

Phe657, and Asn659. From the vaccine, residues Val13, Arg14, Lys50, Asn55, Ser57, Pro58, Ile66, Ala73, Tyr74, Ile77, and Asn81 participated in van der Waals interactions. The docked conformation of the vaccine ensemble with TLR3 can be seen in Fig. 8.

3.10. Disulfide engineering and codon optimization of the vaccine ensemble

Disulfide engineering is a directed approach of introducing disulfide bonds in a protein, and is a logical approach of emulating molecular interactions stability required in many industrial and biomedical applications [87]. The chimeric vaccine sequence was subjected to Design 2.0 server which highlighted 5 pairs of residues with high bond energy ranging from 1.26 kcal/mo to 6.75 kcal/mol. The pairs of residues are: Ile2-Ala61, Arg17-Cys40, Val20-Cys41, Cys21-Lys26, and Pro51-Pro62 were mutated to enhanced vaccine stability (S-Table 2). The original and mutated vaccine structures are shown in Fig. 6E. Furthermore, the vaccine sequence was reverse translated to improve the codon usage of the sequence as per *Escherichia coli* expression system to get maximum expression of the vaccine during wet lab experimentation. The codon adaptation index of the vaccine is 1 which is an ideal value for enhance expression. Lastly, the vaccine sequence was cloned into pET28a (+) expression vector (Fig. 6F).

3.11. Evaluation of conformational stability

The conformational stability of the SARS-CoV-2 N protein in complex with the top drug molecule and vaccine ensemble was deciphered by running 50-ns of MD simulations. Structural stability of the systems was monitored first by calculating carbon alpha distance of superimposed

50,000 snapshots of the MD simulation (Fig. 9 (top left)). For the drug complex, a small number of minor structure fluctuations were noticed in the protein suggesting the structure has endured less conformational changes during the course of simulation. The average rmsd estimated for the drug complex is 1.38 Å with maximum observed is 2.53 Å. The variations were investigated by visualizing snapshots at regular intervals (0-ns, 10-ns, 20-ns, 30-ns, 40-ns, and 50-ns) and superimposed in UCSF Chimera [88] (Fig. 10A). The 3D alignment revealed an rmsd of 1.3 Å that clearly showed the drug complex stability. In the process, the compound was seen in different conformations where 4-isopropylphenol region of the molecule is showing stability at the initial docked site but the 2,5,5,7,9-pentamethyl-2,3,4,4a,5,10b-hexahydropyran[3,2-c]chromene-3-carboxylic acid is stretching along the length allowing the molecule entry deep in the cavity with formation of extra van der Waals contacts as depicted by the binding free energy estimation in section. Further, these movements are the result of flexible loop regions of the N protein that propel the molecule to alter its conformation by adjusting the molecule along the channel of the receptor. This depicts binding free energy methods, with the compound enjoying binding at the cavity site and is an attempt at showing stable binding as simulations progress. In the same way, rmsd analysis of the vaccine ensemble with TLR3 was carried out (Fig. 9 (top left)). The average rmsd of the TLR3-vaccine ensemble is 3.34 Å. Variations in the rmsd are mainly because of the ligand structure move in an attempt to get a highly stable and immune cell recognizable pose, resulting in force applied on the flexible regions of the TLR3. This was affirmed by taking regular snapshots at interval of 10-ns and superimposition in UCSF Chimera that revealed an rmsd of 1.124 Å (Fig. 10B). The receptor rmsd was followed by

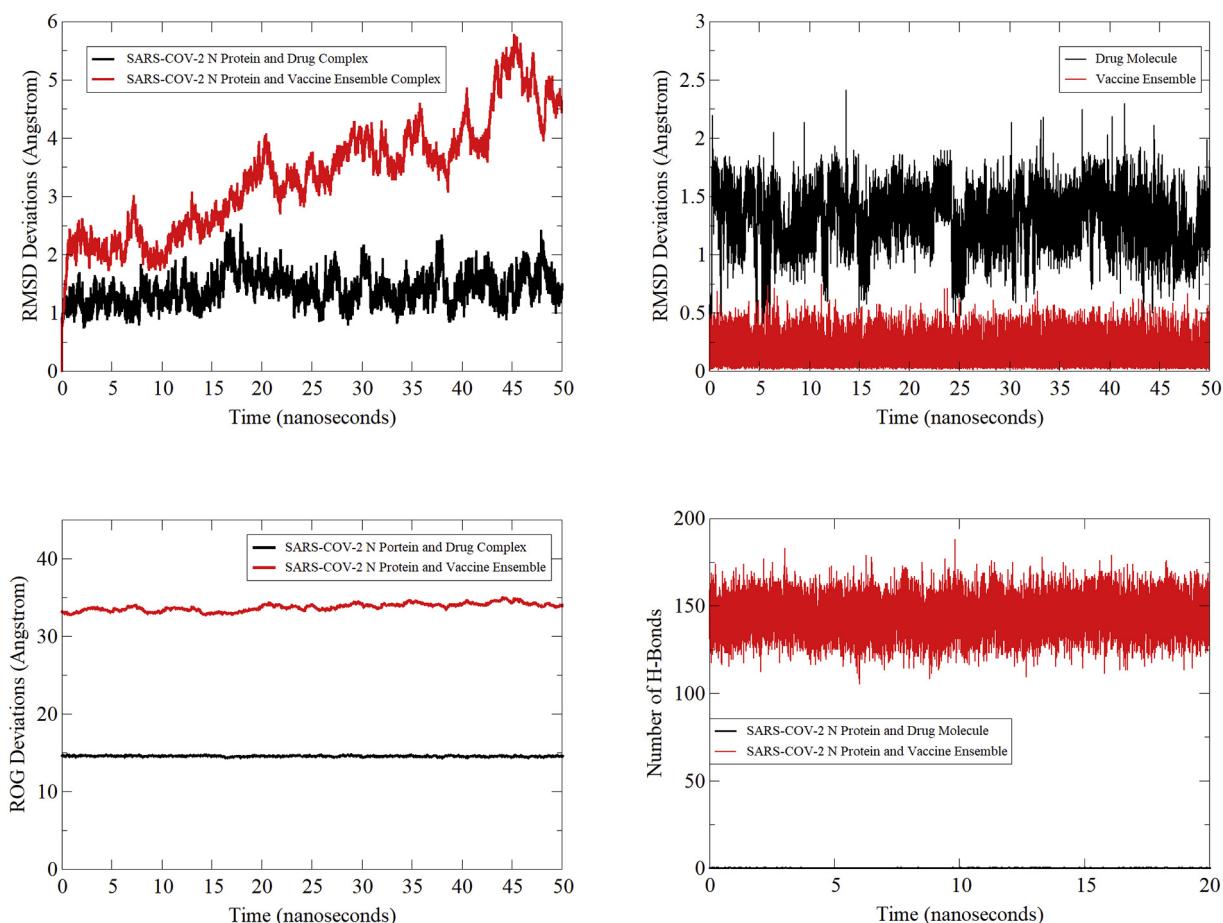


Fig. 9. Different MD based analysis: protein rmsd (top left), ligand rmsd (top right), protein ROG (bottom left) and hydrogen bonds plotting (bottom right).

ligand rmsd to affirm the ligand binding stability with the receptors (Fig. 9 (top right)). An average rmsd of 1.30 Å and maximum of 2.41 Å were reported for the drug compound. Again minor ups and downs in ligand rmsd were spotted, reflecting conformational moves as stated earlier, though these adjustments are in favour of the increased system stability. The vaccine ensemble rmsd is very low (0.15 Å) demonstrating very stable behavior of the vaccine with TLR3. Next, compactness of the N protein structure was tested using the radius of gyration tool (Fig. 9 (bottom left)). The average gyration value of the N protein in the presence of the drug along the simulation is 14.56 Å and maximum value observed is 14.87 Å. This clearly shows that the secondary elements are highly compact in the N protein 3D structure and show stability in the drug molecule presence. The same compact complex system was noted for the TLR3 and vaccine ensemble with a mean radius of gyration value of 33.72 Å. The number of hydrogen bonds formed between the drug/vaccine with the receptors is shown in Fig. 9 (bottom right). On average, the drug molecule is in contact with the N protein docked site via 1 hydrogen bond through the simulation period whereas the vaccine produced a robust network of hydrogen bonds with TLR3 that, on average, is higher in each simulation frame.

3.12. Binding free energy calculation

The MMPBSA is now commonly applied on biological systems to model molecular recognition and is a central focus in molecular simulations [89]. This method uses reasonable computational cost and is in routine use to estimate binding free energies of small ligand molecules bound to a large biomolecule receptor [89]. The MMGBSA analysis, which is complementary to the MMPBSA, revealed binding with considerable affinity to the N protein as depicted by the net total energy of -30.4504 kcal/mol. This net energy can be split into components as: complex ($-10.452.14$ kcal/mol), receptor ($-10.414.58$ kcal/mol), and ligand (-7.10 kcal/mol). The gas phase energy and solvation energy contribution to the net energy of the system are significant; the latter dominates by polar energy as the favourable (-15.27 kcal/mol) opposed to minor support from non-polar energy (-4.27 kcal/mol). The gas phase energy is the attribute of van der Waals energy (-37.49 kcal/mol) in contrast to the non-favourable part from electrostatic energy (26.60 kcal/mol). The detailed N protein and drug molecule MMGBSA binding energies of the complex, receptor, ligand and the net energies are provided in S-Table 3. In the MMPBSA method, the system net total binding energy is -23.59 kcal/mol (delta gas

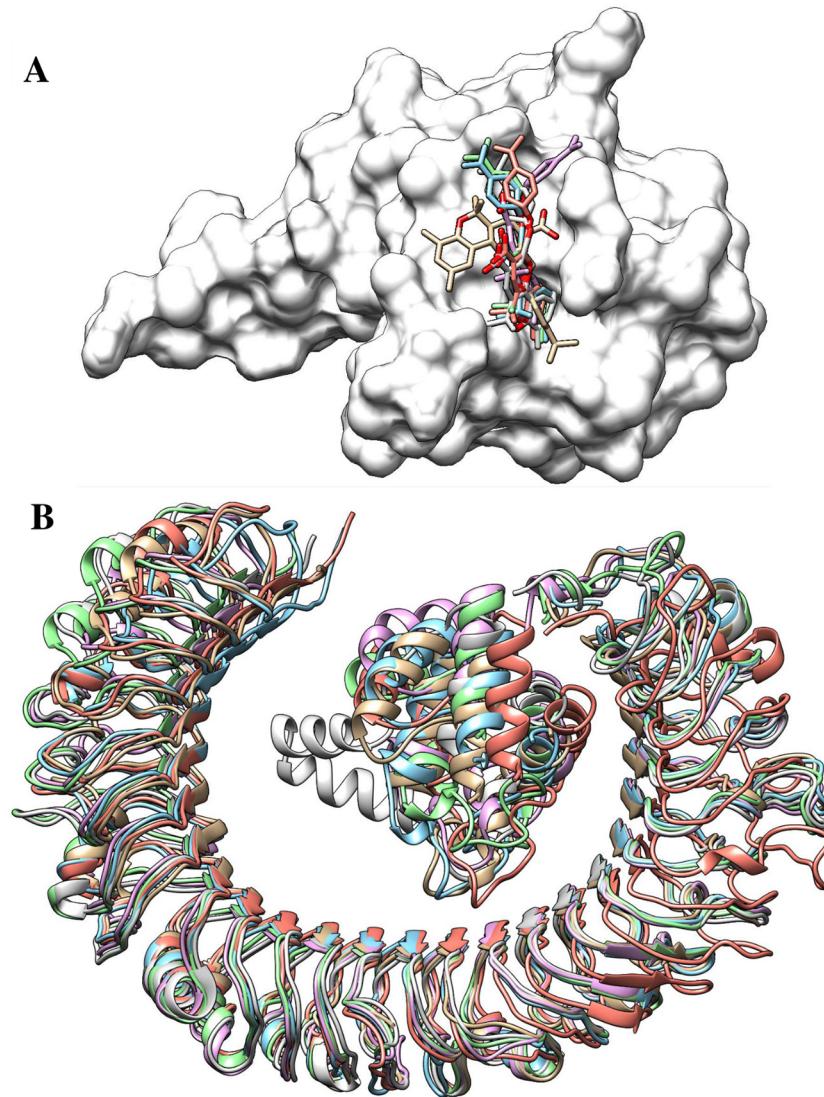


Fig. 10. A. Binding conformation adjustments of drug molecule (shown in different color sticks) at the binding pocket of SARS-CoV2 N protein (shown by gray surface), B. Binding conformation adjustments of vaccine (shown in different color cartoons) at the binding pocket of TLR3. The colouring pattern can be interpreted as: 0-ns (tan), 10-ns (cyan), 20-ns (orchid), 30-ns (light green), 40-ns (salmon), and 50-ns (gray). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Hotspot residues with binding free energy values.

Residue/Ligand	MMGBSA	MMPBSA	Residue/Ligand	MMGBSA	MMPBSA
Drug molecule	-13.3381	-10.6936	Vaccine Ensemble	-22.57	-24.57
Trp5	-3.52555	-2.19901	Ser387	-4.45	-5.67
Thr29	-1.88196	-1.01	Glu460	-3.48	-5.98
Asn30	-3.23374	-1.68084	Tyr462	-2.78	-3.14
Ile94	-1.02521	-1.12	Tyr465	-1.24	-1.78
Ile105	-3.05031	-2.58066	Asn662	-3.65	-4.14

phase (10.89 kcal/mol) + delta solvation energy (-12.70 kcal/mol)). Together, both gas and solvation contributions play a balanced role in the binding affinity of the compound for the receptor. As in MMGBSA, the solvation energy is the output of a good polar solvation role (-9.43 kcal/mol) compared to less non-polar energy (-3.27 kcal/mol) in MMPBSA. The gas phase energy van der Waals component as reported in the MMGBSA is the prime factor in the compound interaction (37.49 kcal/mol) with the protein whereas, due to the lack of formation of ionic moieties that can give the electrostatic nature of interactions, the columbic interaction is insignificant (26.60 kcal/mol). The complete data of the MMPBSA analysis for N protein and drug molecule are presented in S-Table 4. Likewise, the TLR3-vaccine ensemble was highly stable with MMPBSA net binding energy of -47.96 kcal/mol. The stabilizing factor was revealed to be the gas phase electrostatic energy (-1359.83 kcal/mol) in the system lower energy state along with -67.79 kcal/mol of van der Waals energy. The polar solvation energy is the non-favourable contributor (1390.67 kcal/mol) to the net solvation energy (1379.67 kcal/mol), in contrast to little favourable energy from non-polar energy (-10.99 kcal/mol). The full data of MMGBSA and MMPBSA for TLR3 and vaccine ensemble are given as S-Table 5 and S-Table 6, respectively.

3.13. Energy contribution of key receptor residues and ligands

To underline residues that are essential to ligand binding and are favourable in complex stability, per residue decomposition was performed. Common residues of MMGBSA and MMPBSA with negative average binding energy were categorized as essential amino acids and were vital in interactions with the ligand (Table 3). For instance, residues such Trp5, Thr29, Asn30, Ile94, and Ile105 were demonstrated as hot spot residues due to their profound contribution in complex stabilization. For the TLR3-vaccine, Ser387, Glu460, Tyr462, Tyr465, and Asn662 are hotspots in anchoring the vaccine ensemble at the docked position.

3.14. WaterSwap absolute binding free energy

To estimate the absolute binding free energy of the system, WaterSwap was used. Unlike MMPBSA, this uses an explicit water model thus does not suffer from the shortcomings of the usually applied MMPBSA method. The binding free energy is calculated by means of three highly sophisticated approaches including TI, FEP and BAR. The resultant value is considered good when the mentioned techniques are converged with difference of ≤ 1 kcal/mol. The drug complex produced a TI value of -22.314 kcal/mol, FEP of -22.145 kcal/mol, and BAR of -23.87 kcal/mol. Similarly, the vaccine ensemble in complex TLR3 was reported to have TI, FEP and BAR values of -47.45 kcal/mol, -48.74 kcal/mol, and -48.91 kcal/mol, respectively.

4. Concluding remarks

Efficient development of new vaccines and biologically useful drug molecules usually take years of research efforts and is a multibillion dollar gamble. The use of available pharmaceutically active and safe anti-viral agents in silico and in wet lab experiments as a substitute is a

swift approach for uncovering medication which may efficiently deal with deadly and evolving viral infections. The conventional drug discovery pipeline takes a decade for safe anti-viral therapy development [90,91]. For diseases that are highly contagious like COVID-19, we don't have enough time and require to speed up the process by screening available drugs and repurposing them against this new disease threat in a process called drug repurposing. In an NIH-funded study published in the journal "Nature", scientists screened a library of 12,000 existing drugs against SARS-CoV-2 using laboratory-grown human cell lines and non-human primates. They found 21 drugs that showed potential to thwart SARS-CoV-2, 13 out of which can be safely given to people. Most of these drugs have tested clinically against autoimmune diseases, HIV, osteoporosis and other complications [92]. Recently, an international team led by Sumit Chanda at Sanford Burnham Prebys Medical Discovery Institute, together with Yuen Kwok-Yung's team at the University of Hong Kong, employed a high throughput method to rapidly screen 1987 compounds from the REFRAME library [93] that shortlisted 100 drugs that reliably hinder the virus growth by at least 40%. Further, 21 drugs were filtered based on dose-response relationship. Out of these 21 drugs, one drug was remdesivir, which is the FDA approved drug originally developed against Ebola virus, as a possible treatment option against COVID-19 [94]. Additionally, studies on the most potent drugs in the list reduced the viral load by 65% to 85%. Most potent in the list were apilimod (a drug in clinical trials to treat rheumatoid arthritis, and Crohn's disease) and clofazimine (a 70 year old FDA drug for the treatment of leprosy) (<https://directorsblog.nih.gov/2020/08/04/exploring-drug-repurposing-for-covid-19-treatment/>). These findings suggest the use of existing and experimental drugs has the potential to treat COVID-19. Likewise, immune-informatics tools can be employed to recognize immunological active sites in the viral genome for the purpose of developing epitope-based vaccine candidates. Such vaccines have several advantages over whole organism-based vaccines as they are safe and easy to produce [95–97]. Peptide-based vaccines significantly limit reactogenic and allergenic complications, and triggers stimulation of B cells and T cells or both simultaneously [96]. There are low batch to batch differences in peptide-based vaccine production and can be easily standardized [98]. The peptide structure is well known and structure-function can be easily correlated in contrast to a traditional vaccine. Furthermore, the peptide can be easily formulated to conjugated structures and multi-epitope vaccines [98]. Despite the many benefits of using peptide vaccines, they possess lower immunogenic ability which can be overcome by fusing adjuvants in peptide vaccine formulations [96]. Such peptide-based vaccines are currently being considered and hold substantial promise to prevent human viruses like hepatitis C virus and HIV [96,97]. In the work reported herein, an integrated study of computational drugs leads to identification and peptide-based vaccine ensemble design against SARS-CoV-2 N protein is performed. The scientific and medical hunt for COVID-19 drugs and vaccines is in full bloom to stop the pandemic and subsequent waves of virus spread. This can be witnessed by an exponential number of COVID-19 publications in journals or in preprint archives (around 22,000 in PubMed and 5000 preprint in BioRxiv/MedRxiv) covering SARS-CoV-2 drugs and vaccines directly or indirectly [99]. Making the process of data sharing and joint global push is much needed for prioritizing drug and vaccine candidates, clinical trial streamlining. Coordinating on a regulatory process might be a prompt way of dealing with this deadly and ghastly virus. Finally, we acknowledge several limitations of this work due to lack of experimental support but nevertheless these results are promising and can save time and resources for scientific personnel involved directly in experimental therapeutics design. In this way, the translational distance between preclinical and clinical products might be reduced considerably and will aid in paving a path for rapid practical development of drugs and the much anticipated vaccine.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.molliq.2020.114734>.

Authors contributions

Sajjad Ahmad: Conceptualization, Data Curation, Formal Analysis, Software, Writing-Review & Editing. **Yasir Waheed:** Conceptualization, Funding acquisition, Supervision, Writing-review & editing. **Saba Ismail:** Data Curation, Formal Analysis, Methodology, Software. Sumra Wajid Abbasi: Data Curation, Formal Analysis, Methodology, Software. **Muzammil Hasan Najmi:** Investigation, Writing-review & editing.

Declaration of Competing Interest

The authors in this study have no conflict of interest.

Acknowledgment

Dr. Yasir Waheed is highly grateful to the Foundation University, Islamabad, for granting financial assistance to accomplish this work.

References

- [1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, W. Tan, A novel coronavirus from patients with pneumonia in China, 2019, *N. Engl. J. Med.* 0 (2020) <https://doi.org/10.1056/NEJMoa2001017> null.
- [2] W. Guan, Z. Ni, Y. Hu, W. Liang, C. Ou, J. He, L. Liu, H. Shan, C. Lei, D.S.C. Hui, et al., Clinical characteristics of coronavirus disease 2019 in China, *N. Engl. J. Med.* 382 (18) (2020) 1708–1720, <https://doi.org/10.1056/NEJMoa2002032>.
- [3] B. Cao, Y. Wang, D. Wen, W. Liu, J. Wang, G. Fan, L. Ruan, B. Song, Y. Cai, M. Wei, et al., A trial of lopinavir-ritonavir in adults hospitalized with severe Covid-19, *N. Engl. J. Med.* 382 (19) (2020) 1787–1799, <https://doi.org/10.1056/NEJMoa2001282>.
- [4] T. Ahmed, M. Noman, A. Almatroudi, M. Shahid, M. Khurshid, F. Tariq, M.T. ul Qamar, R. Yu, B. Li, A Novel Coronavirus 2019 Linked with Pneumonia in China: Current Status and Future Prospects, 2020.
- [5] J.-F.-W. Chan, S. Yuan, K.-H. Kok, K.K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C.-C.-Y. Yip, R.W.-S. Poon, et al., A familial cluster of pneumonia associated with the 2019 Novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet.* 395 (2020) 514–523.
- [6] C. Gil, T. Ginex, I. Maestro, V. Nozal, L. Barrado-Gil, M.A. Cuesta-Geijo, J. Urquiza, D. Ramírez, C. Alonso, N.E. Campillo, et al., COVID-19: drug targets and potential treatments, *J. Med. Chem.* (2020) <https://doi.org/10.1021/acs.jmedchem.0c00606>.
- [7] L. Wang, Y. Wang, D. Ye, Q. Liu, A review of the 2019 novel coronavirus (COVID-19) based on current evidence, *Int. J. Antimicrob. Agents* 105948 (2020).
- [8] H. Li, S.-M. Liu, X.-H. Yu, S.-L. Tang, C.-K. Tang, Coronavirus disease 2019 (COVID-19): current status and future perspective, *Int. J. Antimicrob. Agents* 105951 (2020).
- [9] I. Chakraborty, P. Maity, COVID-19 outbreak: migration, effects on society, global environment and prevention, *Sci. Total Environ.* 138882 (2020).
- [10] K. Stadler, V. Masignani, M. Eickmann, S. Becker, S. Abrignani, H.-D. Klenk, R. Rappuoli, SARS—beginning to understand a new virus, *Nat. Rev. Microbiol.* 1 (2003) 209–218.
- [11] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, et al., A new coronavirus associated with human respiratory disease in China, *Nature* 579 (2020) 265–269.
- [12] J.-F.-W. Chan, K.-H. Kok, Z. Zhu, H. Chu, K.K.-W. To, S. Yuan, K.-Y. Yuen, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, *Emerg. Microbes Infect.* 9 (2020) 221–236.
- [13] G. Li, E. De Clercq, Therapeutic Options for the 2019 Novel Coronavirus (2019-nCoV), 2020.
- [14] J.S. Morse, T. Lalonde, S. Xu, W.R. Liu, Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV, *ChemBiochem.* 21 (5) (2020) 730–738.
- [15] R. Ramayam, K.-P. Tan, P.-H. Liang, Recent Development of 3C and 3CL Protease Inhibitors for Anti-Coronavirus and Anti-Picornavirus Drug Discovery, 5th ed.39, *Biochem Soc Trans.*, 2011 1371–1375.
- [16] D. Wrapp, N. Wang, K.S. Corbett, J.A. Goldsmith, C.-L. Hsieh, O. Abiona, B.S. Graham, J.S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation, *Science* (80) 367 (2020) 1260–1263.
- [17] S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q. Chen, et al., Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites, *Acta Pharm. Sin. B* 10 (7) (2020) 1228–1238.
- [18] G.W. Nelson, S.A. Stohlman, S.M. Tahara, High affinity interaction between nucleocapsid protein and leader/intergenic sequence of mouse hepatitis virus RNA, *J. Gen. Virol.* 81 (2000) 181–188.
- [19] Y. Cong, M. Ulasli, H. Schepers, M. Mauthe, P. V'kovski, F. Kriegelburg, V. Thiel, C.A.M. de Haan, F. Reggiori, Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle, *J. Virol.* 94 (2020).
- [20] L. Du, G. Zhao, Y. Lin, C. Chan, Y. He, S. Jiang, C. Wu, D.-Y. Jin, K.-Y. Yuen, Y. Zhou, et al., Priming with rAAV encoding RBD of SARS-CoV S protein and boosting with RBD-specific peptides for T cell epitopes elevated humoral and cellular immune responses against SARS-CoV infection, *Vaccine* 26 (2008) 1644–1651.
- [21] C. Chang, S.-C. Sue, T. Yu, C.-M. Hsieh, C.-K. Tsai, Y.-C. Chiang, S. Lee, H. Hsiao, W.-J. Wu, W.-L. Chang, et al., Modular organization of SARS coronavirus nucleocapsid protein, *J. Biomed. Sci.* 13 (2006) 59–72.
- [22] S.F. Ahmed, A.A. Quadeer, M.R. McKay, Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies, *Viruses* 12 (2020) 254.
- [23] S.-J. Liu, C.-H. Leng, S. Lien, H.-Y. Chi, C.-Y. Huang, C.-L. Lin, W.-C. Lian, C.-J. Chen, S.-L. Hsieh, P. Chong, Immunological characterizations of the nucleocapsid protein based SARS vaccine candidates, *Vaccine* 24 (2006) 3100–3108.
- [24] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF chimera—a visualization system for exploratory research and analysis, *J. Comput. Chem.* 25 (2004) 1605–1612.
- [25] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham III, T.A. Darden, R.E. Duke, H. Gohlke, et al., The FF14SB force field, 14, *Amber*, 2014 29–31.
- [26] J.E. Kerrigan, AMBER 10.0 Introductory Tutorial, 2009.
- [27] R.A. Laskowski, PDBsum: summaries and analyses of PDB structures, *Nucleic Acids Res.* 29 (2001) 221–222.
- [28] S. Dallakyan, A.J. Olson, Small-molecule library screening by docking with PyRx, *Chem. Biol.* Springer 2015, pp. 243–250.
- [29] C.A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, *Drug Discov. Today Technol.* 1 (2004) 337–341, <https://doi.org/10.1016/j.ddtec.2004.11.007>.
- [30] R. Vita, S. Mahajan, J.A. Overton, S.K. Dhanda, S. Martini, J.R. Cantrell, D.K. Wheeler, A. Sette, B. Peters, The immune epitope database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (2018) D339–D343.
- [31] I.A. Doytchinova, D.R. Flower, Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines, *BMC Bioinform.* 8 (2007) 4.
- [32] I. Dimitrov, D.R. Flower, I. Doytchinova, AllerTOP-a server for in silico prediction of allergens, *BMC Bioinformatics* 2013, p. S4.
- [33] S. Gupta, P. Kapoor, K. Chaudhary, A. Gautam, R. Kumar, G.P.S. Raghava, O.S.D.D. Consortium, et al., In silico approach for predicting toxicity of peptides and proteins, *PLoS One* 8 (2013) e73957.
- [34] A. Garg, D. Gupta, VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens, *BMC Bioinform.* 9 (2008) 62.
- [35] S.K. Dhanda, P. Vir, G.P.S. Raghava, Designing of interferon-gamma inducing MHC class-II binders, *Biol. Direct* 8 (2013) 30.
- [36] H.-H. Bui, J. Sidney, W. Li, N. Fuseder, A. Sette, Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines, *BMC Bioinform.* 8 (2007) 361.
- [37] H.-H. Bui, J. Sidney, K. Dinh, S. Southwood, M.J. Newman, A. Sette, Predicting population coverage of T-cell epitope-based diagnostics and vaccines, *BMC Bioinform.* 7 (2006) 1–5.
- [38] J. Cheng, A.Z. Randall, M.J. Sweredoski, P. Baldi, SCRATCH: a protein structure and structural feature prediction server, *Nucleic Acids Res.* 33 (2005) W72–W76.
- [39] L. Heo, H. Park, C. Seok, GalaxyRefine: protein structure refinement driven by side-chain repacking, *Nucleic Acids Res.* 41 (2013) W384–W388.
- [40] C. Colovos, T.O. Yeates, ERRAT: an empirical atom-based method for validating protein structures, *Protein Sci.* 2 (1993) 1511–1519.
- [41] D. Eisenberg, R. Lüthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, *Methods Enzymol.* Elsevier 1997, pp. 396–404.
- [42] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res.* 35 (2007) W407–W410.
- [43] E. ProtParam, ExPASy-ProtParam Tool, 2017.
- [44] M. Hebditch, M.A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwicker, Protein-sol: a web tool for predicting protein solubility from sequence, *Bioinformatics*. 33 (2017) 3098–3100.
- [45] R. Zambrano, M. Jamroz, A. Szczasiuk, J. Pujols, S. Kmiecik, S. Ventura, AGGRESCAN3D (AG3D): server for prediction of aggregation properties of protein structures, *Nucleic Acids Res.* 43 (2015) W306–W313.
- [46] M. Thomsen, C. Lundsgaard, S. Buus, O. Lund, M. Nielsen, MHCcluster, a method for functional clustering of MHC molecules, *Immunogenetics*. 65 (2013) 655–665.
- [47] N. Rapin, O. Lund, F. Castiglione, C-Imm5im 10.1 Server, 2012.
- [48] M. ul Qamar, Z. Shokat, I. Muneeb, U.A. Ashfaq, H. Javed, F. Anwar, A. Bari, B. Zahid, N. Saari, Multi-epitope-based subunit vaccine design and evaluation against respiratory syncytial virus using reverse vaccinology approach, *Vaccines* 8 (2020) 288.
- [49] R.A. Shey, S.M. Ghogomu, K.K. Esoh, N.D. Nebangwa, C.M. Shintouo, N.F. Nongley, B.F. Asa, F.N. Ngale, L. Vanhamme, J. Souopgui, In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases, *Sci. Rep.* 9 (2019) 4409.
- [50] R. Dong, Z. Chu, F. Yu, Y. Zha, Contriving multi-epitope subunit of vaccine for COVID-19: immunoinformatics approaches, *Front. Immunol.* 11 (2020) 1784.
- [51] N. Kumar, D. Sood, R. Chandra, Design and optimization of a subunit vaccine targeting COVID-19 molecular shreds using an immunoinformatics framework, *RSC Adv.* 10 (2020) 35856–35872.
- [52] A.H. Abdelmoneim, M.I. Mustafa, M.I. Abdelmageed, N.S. Murshed, E. dk Dawoud, E.M. Ahmed, R.M. Kamal Eldein, N.M. Elfadol, A.O.M. Sati, A.M. Makhlawi, Immunoinformatics design of multipeptides peptide-based universal cancer vaccine using matrix metalloproteinase-9 protein as a target, *Immunol. Med.* (2020) 1–18.
- [53] D. Chaudhuri, J. Datta, S. Majumder, K. Giri, In silico designing of peptide based vaccine for hepatitis viruses using reverse vaccinology approach, *Infect. Genet. Evol.* 104388 (2020).

- [54] 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.
- [55] D. Schneidman-Duhovny, Y. Inbar, R. Nussinov, H.J. Wolfson, PatchDock and SymmDock: servers for rigid and symmetric docking, *Nucleic Acids Res.* 33 (2005) W363–W367.
- [56] N. Andrusier, R. Nussinov, H.J. Wolfson, FireDock: fast interaction refinement in molecular docking, *Proteins Struct. Funct. Bioinforma.* 69 (2007) 139–159.
- [57] N. Nezafat, Z. Karimi, M. Eslami, M. Mohkam, S. Zandian, Y. Ghasemi, Designing an efficient multi-epitope peptide vaccine against *Vibrio cholerae* via combined immunoinformatics and protein interaction based approaches, *Comput. Biol. Chem.* 62 (2016) 82–95.
- [58] A.K. Jaiswal, S. Tiwari, S.B. Jamal, D. Barh, V. Azevedo, S.C. Soares, An in silico identification of common putative vaccine candidates against *Treponema pallidum*: a reverse vaccinology and subtractive genomics based approach, *Int. J. Mol. Sci.* 18 (2017) <https://doi.org/10.3390/ijms18020402>.
- [59] V. Chauhan, T. Rungta, K. Goyal, M.P. Singh, Designing a multi-epitope based vaccine to combat Kaposi Sarcoma utilizing immunoinformatics approach, *Sci. Rep.* 9 (2019) 2517.
- [60] M. Enayatkhani, M. Hasaniazad, S. Faezi, H. Gulkani, P. Davoodian, N. Ahmadi, M.A. Einakian, A. Karmostaji, K. Ahmadi, Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study, *J. Biomol. Struct. Dyn.* (2020) 1–19.
- [61] T.F. Vieira, S.F. Sousa, Comparing AutoDock and Vina in ligand/decoy discrimination for virtual screening, *Appl. Sci.* 9 (2019) 4538.
- [62] S.F. Sousa, P.A. Fernandes, M.J. Ramos, Protein-ligand docking: current status and future challenges, *Proteins Struct. Funct. Bioinforma.* 65 (2006) 15–26.
- [63] Z. Wang, H. Sun, X. Yao, D. Li, L. Xu, Y. Li, S. Tian, T. Hou, Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power, *Phys. Chem. Chem. Phys.* 18 (2016) 12964–12975.
- [64] S.S. Azam, S.W. Abbasi, Molecular docking studies for the identification of novel melatonergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines, *Theor. Biol. Med. Model.* 10 (2013) 63.
- [65] P. Kangueane, C. Nilofer, Protein-protein docking: Methods and tools, *Protein-Protein Domain-Domain Interact.*, Springer 2018, pp. 161–168.
- [66] W. Zhang, E.W. Bell, M. Yin, Y. Zhang, EDock: blind protein-ligand docking by replica-exchange Monte Carlo simulation, *Aust. J. Chem.* 12 (2020) 1–17.
- [67] C. Shivanika, D. Kumar, V. Ragunathan, P. Tiwari, A. Sumitha, et al., Molecular docking, validation, dynamics simulations, and pharmacokinetic prediction of natural compounds against the SARS-CoV-2 main-protease, *J. Biomol. Struct. Dyn.* (2020) 1.
- [68] L.H.S. Santos, R.S. Ferreira, E.R. Caffarena, Integrating molecular docking and molecular dynamics simulations, *Docking Screens Drug Discov.* Springer 2019, pp. 13–34.
- [69] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, *Expert Opin. Drug Discovery* 10 (2015) 449–461, <https://doi.org/10.1517/17460441.2015.1032936>.
- [70] C.J. Woods, M. Malaisree, J. Michel, B. Long, S. McIntosh-Smith, A.J. Mulholland, Rapid decomposition and visualisation of protein-ligand binding free energies by residue and by water, *Faraday Discuss.* 169 (2014) 477–499, <https://doi.org/10.1039/C3fd00125c>.
- [71] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham III, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, et al., AMBER 18, Univ. California, San Fr., 2018 (n.d.).
- [72] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.
- [73] C. Schafmeister, W.S. Ross, V. Romanovski, The Leap Module of AMBER, Univ. Calif., 1995.
- [74] V. Kräutler, W.F. Van Gunsteren, P.H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, *J. Comput. Chem.* 22 (2001) 501–508.
- [75] J.A. Izaguirre, D.P. Catarello, J.M. Wozniak, R.D. Skeel, Langevin stabilization of molecular dynamics, *J. Chem. Phys.* 114 (2001) 2090–2098.
- [76] D.R. Roe, T.E. Cheatham III, PTraJ and CPPTraJ: software for processing and analysis of molecular dynamics trajectory data, *J. Chem. Theory Comput.* 9 (2013) 3084–3095.
- [77] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (1996) 33–38.
- [78] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theory Comput.* 8 (2012) 3314–3321, <https://doi.org/10.1021/ct300418h>.
- [79] C.J. Woods, M. Malaisree, S. Hannongbua, A.J. Mulholland, A water-swap reaction coordinate for the calculation of absolute protein-ligand binding free energies, *J. Chem. Phys.* 134 (2011) <https://doi.org/10.1063/1.3519057>.
- [80] J.L. Sanchez-Trincado, M. Gomez-Perez, P.A. Reche, Fundamentals and methods for T-and B-cell epitope prediction, *J. Immunol. Res.* 2017 (2017).
- [81] M. Saadi, A. Karkhah, H.R. Nouri, Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches, *Infect. Genet. Evol.* 51 (2017) 227–234.
- [82] Y. Tian, R. da Silva Antunes, J. Sidney, C.S.L. Arlehamn, A. Griffoni, S.K. Dhanda, S. Paul, B. Peters, D. Weiskopf, A. Sette, A review on T cell epitopes identified using prediction and cell-mediated immune models for *Mycobacterium tuberculosis* and *Bordetella pertussis*, *Front. Immunol.* 9 (2018).
- [83] A. Naz, F.M. Awan, A. Obaid, S.A. Muhammad, R.Z. Paracha, J. Ahmad, A. Ali, Identification of putative vaccine candidates against *Helicobacter pylori* exploiting exoproteome and secretome: a reverse vaccinology based approach, *Infect. Genet. Evol.* 32 (2015) 280–291.
- [84] M.T. Ul Qamar, S. Saleem, U.A. Ashfaq, A. Bari, F. Anwar, S. Alqahtani, Epitope-based peptide vaccine design and target site depiction against Middle East respiratory syndrome coronavirus: an immune-informatics study, *J. Transl. Med.* 17 (2019) 362.
- [85] L. Zhang, Multi-epitope vaccines: a promising strategy against tumors and viral infections, *Cell. Mol. Immunol.* 15 (2018) 182.
- [86] G. Diamond, N. Beckloff, A. Weinberg, K.O. Kisich, The roles of antimicrobial peptides in innate host defense, *Curr. Pharm. Des.* 15 (2009) 2377–2392.
- [87] A.A. Dombrowski, K.Z. Sultan, D.B. Craig, Protein disulfide engineering, *FEBS Lett.* 588 (2014) 206–212.
- [88] S. Kaliappan, I.I.T. Bombay, UCSF Chimera-Superimposing and Morphing, 2016.
- [89] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the Performance of the MM_PBSA and MM_GBSA Methods. 1. The Accuracy. pdf, 2011 69–82.
- [90] M.T. ul Qamar, S.M. Alqahtani, M.A. Alamri, L.-L. Chen, Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants, *J. Pharm. Anal.* 10 (4) (2020) 313–319, <https://doi.org/10.1016/j.jpha.2020.03.009>.
- [91] M.T. ul Qamar, S. Kiran, U.A. Ashfaq, M.R. Javed, F. Anwar, M.A. Ali, A.H. ul Gilani, Discovery of novel dengue NS2B/NS3 protease inhibitors using pharmacophore modeling and molecular docking based virtual screening of the zinc database, *Int. J. Pharmacol.* 12 (2016) 621–632.
- [92] L. Riva, S. Yuan, X. Yin, L. Martin-Sancho, N. Matsunaga, L. Pache, S. Burgstaller-Muehlbacher, P.D. De Jesus, P. Teriete, M.V. Hull, et al., Discovery of SARS-CoV-2 antiviral drugs through large-scale compound repurposing, *Nature* (2020) 1–11.
- [93] J. James, M.E. Young, E. Chen, N.H. Rogers, S. Burgstaller-Muehlbacher, L.D. Hughes, M.S. Love, M.V. Hull, K.L. Kuhen, A.K. Woods, et al., The ReFRAME library as a comprehensive drug repurposing library and its application to the treatment of cryptosporidiosis, *Proc. Natl. Acad. Sci.* 115 (2018) 10750–10755.
- [94] R. Poduri, G. Joshi, G. Jagadeesh, Drugs targeting various stages of the SARS-CoV-2 life cycle: exploring promising drugs for the treatment of Covid-19, *Cell. Signal.* 74 (2020) 109721.
- [95] M. Skwarczynski, I. Toth, Peptide-based synthetic vaccines, *Chem. Sci.* 7 (2016) 842–854.
- [96] W. Li, M. Joshi, S. Singhania, K. Ramsey, A. Murthy, Peptide vaccine: progress and challenges, *Vaccines* 2 (2014) 515–536.
- [97] R.J. Malonis, J.R. Lai, O. Vergnolle, Peptide-based vaccines: current progress and future challenges, *Chem. Rev.* 120 (2019) 3210–3229.
- [98] M. Topuzogullari, T. Acar, P.P. Arayici, B. Uçar, E. Uğurel, E. Şefik Abamor, T. Arasoğlu, D.I. BALIK, S. Derman, An insight into the epitope-based peptide vaccine design strategy and studies against COVID-19, *Turk. J. Biol.* 44 (2020) 215–227.
- [99] C.D. Funk, C. Laferrière, A. Ardakani, A snapshot of the global race for vaccines targeting SARS-CoV-2 and the COVID-19 pandemic, *Front. Pharmacol.* 11 (2020) 937.