

 Open access • Posted Content • DOI:10.1101/2021.05.02.442313

## **Immunoinformatics approach for a novel multi-epitope vaccine construct against spike protein of human coronaviruses — [Source link](#)**

[Avinash Kumar](#), [Ekta Rathi](#), [Suvarna G. Kini](#)

**Institutions:** [Manipal University](#)

**Published on:** 02 May 2021 - [bioRxiv](#) (Cold Spring Harbor Laboratory)

**Topics:** [Epitope](#), [Immune receptor](#) and [Adjuvant](#)

Related papers:

- [Design an efficient multi-epitope peptide vaccine candidate against SARS-CoV-2: An in silico analysis](#)
- [Multiple epitope-based vaccine prediction against SARS-CoV-2 spike glycoprotein.](#)
- [Designing a multi-epitope vaccine candidate to combat MERS-CoV by employing an immunoinformatics approach.](#)
- [Immuno-informatics approach for multi-epitope vaccine designing against SARS-CoV-2.](#)
- [Designing spike protein \(S-Protein\) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/immunoinformatics-approach-for-a-novel-multi-epitope-vaccine-4wd38tsc2i>

## **Immunoinformatics approach for a novel multi-epitope vaccine construct against spike protein of human coronaviruses**

Avinash Kumar<sup>1</sup>, Ekta Rathi<sup>1</sup>, and Suvarna G Kini<sup>1\*</sup>

<sup>1</sup>Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India.

\*Corresponding author: Dr Suvarna G Kini, Department of Pharmaceutical Chemistry, Manipal College of Pharmaceutical Sciences, Manipal Academy of Higher Education, Manipal, Karnataka, India. E-mail: [suvarna.gk@manipal.edu](mailto:suvarna.gk@manipal.edu)

### **Abstract:**

Spike (S) proteins are an attractive target as it mediates the binding of the SARS-CoV-2 to the host through ACE-2 receptors. We hypothesize that the screening of S protein sequences of all the HCoVs would result in the identification of potential multi-epitope vaccine candidates capable of conferring immunity against various HCoVs. In the present study, several machine learning-based *in-silico* tools were employed to design a broad-spectrum multi-epitope vaccine candidate against S protein of human coronaviruses. To the best of our knowledge, it is one of the first study, where multiple B-cell epitopes and T-cell epitopes (CTL and HTL) were predicted from the S protein sequences of all seven known HCoVs and linked together with an adjuvant to construct a potential broad-spectrum vaccine candidate. Secondary and tertiary structures were predicted, validated and the refined 3D-model was docked with an immune receptor. The vaccine candidate was evaluated for antigenicity, allergenicity, solubility, and its ability to achieve high-level expression in bacterial hosts. Finally, the immune simulation was carried out to evaluate the immune response after three vaccine doses. The designed vaccine is antigenic (with or without the adjuvant), non-allergenic, binds well with TLR-3 receptor and might elicit a diverse and strong immune response.

**Abbreviations:** SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; ACE-2: angiotensin-converting enzyme 2; HCoVs: human coronaviruses, RBD: receptor binding domain, TLR-3: toll-like receptor.

**Keywords:** Covid-19, Coronaviruses, Cheminformatics, Immunoinformatic, SARS-CoV-2, Vaccine

## 1. Introduction

The increased incidence of infectious diseases, especially of zoonotic origin, might become an existential threat to human species. In 2018, WHO (world health organization) had come out with a blueprint of priority disease list which identified viruses like Nipah, Ebola and MERS (middle east respiratory syndrome) posing international biosafety threats (“Blueprint,” n.d.). It also waved a red flag on the pandemic risk potential of the infections caused by known zoonotic pathogens and collectively classified these threats as “disease X” (“Blueprint,” n.d.). The fears of 2018 came true in the early months of 2020 when Covid-19 emerged as the most devastating pandemic the world has ever seen. Around 3.15 million people have lost life with numbers rising daily and more than 150 million people reported positive to the SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) virus responsible for Covid-19 (“Home - Johns Hopkins Coronavirus Resource Center,” n.d.). A recent study suggests that coronaviruses (CoVs) and bats have co-evolved for millions of years but CoVs seldom jumped across to infect humans (Joffrin et al., 2020). They were considered as a comparatively harmless pathogen causing mild respiratory illness to humans (Song et al., 2019). But in the last two decades after the incidences of SARS infection in 2003 and MERS infection in 2013, the understanding regarding CoVs have changed. SARS-CoV-2 has been reported to target pneumocytes which cause severe respiratory distress characterized by bronchiolitis, bronchitis and pneumonia (Abraham Peele, Srihansa, Krupanidhi, Ayyagari, & Venkateswarulu, 2020). The current pandemic of Covid-19 is unprecedented in many ways and its impact on the economies around the world has been devastating. The scientific community has thrown all its resources to come up with either a drug to treat Covid-19 or a vaccine as a prophylactic measure with limited success.

SARS-CoV-2 like SARS and MERS belongs to the *Coronaviridae* family. The presence of spikes on the outer surface of CoVs has a resemblance of the crown and hence the name corona (in Latin corona means crown) (Pandey et al., 2020). SARS-CoV-2 is an enveloped, positive-sense, single-stranded RNA beta-coronavirus with 26-32 thousand base pairs (de Wilde, Snijder, Kikkert, & van Hemert, 2018). Its genome encodes both structural like E (envelop) protein, N (nucleocapsid) protein, M (membrane) protein, S (spike) protein and non-structural proteins (nsp) like proteases (Mousavizadeh & Ghasemi, 2020). Out of all the four structural proteins, S protein plays a prominent role in the binding of the SARS-CoV-2 to the host target receptors, especially through ACE-2 (angiotensin-converting enzyme 2) receptor (Wrapp et al., 2020; Zhou et al., 2020). Cleavage of S protein by furin or other

proteases has been reported in many of the avian and mammalian CoVs. The cleavage of S results into S1 which bears the receptor-binding domain (RBD) and S2 responsible for fusion (Wrapp et al., 2020). Many authors have suggested that RBD of S protein might be an interesting target for potential therapies against SARS-CoV-2 (Kadioglu1, Saeed1, Johannes Greten2, & Efferth1, n.d.; Panda et al., 2020; Prajapat et al., 2020). Scientists have achieved significant success in deciphering the key structural and non-structural proteins of the virus. Wrapp et al. have reported a Cryo-EM structure of S protein of SRAS-CoV-2 in the prefusion conformation (Wrapp et al., 2019). Zhang et al. have reported the x-ray crystal structure of the unliganded SARS-CoV-2 M<sup>pro</sup> and its complex with an  $\alpha$ -ketoamide inhibitor (L. Zhang et al., 2020). The researchers have employed various cheminformatics tools to target these proteins and develop potential drugs for Covid-19. Recently, several authors have reported SARS-CoV-2 M<sup>pro</sup> inhibitors by using molecular modeling methods like virtual screening, molecular docking, molecular dynamics simulations, fast pulling of ligand (FPL), and free energy perturbation (FEP) (Alexpandi, De Mesquita, Pandian, & Ravi, 2020; Ferraz, Gomes, S Novaes, & Goulart Trossini, 2020; Gentile et al., 2020; Ngo, Quynh Anh Pham, Thi Le, Pham, & Vu, 2020; Olubiyi, Olagunju, Keutmann, Loschwitz, & Strodel, 2020). Hagar et al. have investigated some antiviral N-heterocycles as potential Covid-19 drug by using molecular docking and DFT calculations (Wrapp et al., 2019).

But apart from the search of drugs for Covid-19, researchers have tried to develop a vaccine against it. The trust in vaccine development comes primarily from the experience of successful mitigation of epidemics caused by coronaviruses in farm animals by vaccines (F. Li & Du, 2019; Song et al., 2019). Vaccines against human coronavirus which are under development comprise of DNA and RNA vaccines, vector-based vaccine, live attenuated virus itself or protein subunit-based vaccines (Corbett et al., 2020; Mateus et al., 2020; Sahin et al., 2020; van Doremalen et al., 2020). Epitope-based vaccine approach has been employed to combat several diseases like solid tumor (Knutson, Schiffman, & Disis, 2001), malaria (R. Wang et al., 2004) and multiple sclerosis (Bourdette et al., 2005). Recently, many authors have reported epitope-based vaccine candidates for the mitigation of Covid-19 (Abdelmageed et al., 2020; Abraham Peele et al., 2020; Enayatkhani et al., 2020; Mukherjee, Tworowski, Detroja, Mukherjee, & Frenkel-Morgenstern, 2020; Naz, Shahid, Butt, Awan, & Ali, 2020; Oany, Emran, & Jyoti, 2014; Sarkar, Ullah, Tuz, Afrin, & Araf, 2020). Yang et al. have combined immunoinformatic methods with the deep learning algorithm to directly predict potential vaccine subunits from the S protein sequence of SARS-CoV-2 and constructed a multi-epitope vaccine for SARS-CoV-2 virus (Z. Yang, Bogdan, & Nazarian, 2020).

Similarly, Singh et al. Have designed multi-epitope vaccine peptides by targeting four different structural proteins (Spike, Envelop, Membrane and Nucleocapsid) of SARS-CoV-2(A. Singh, Thakur, Sharma, & Chandra, 2020). Li et al. have also employed several immunoinformatic tools to design a multi-epitope vaccine by targeting three different structural proteins of SARS-CoV-2(Lin et al., 2020).

When we looked into the recent literature, we found most of the designed multi-epitope vaccine were against the S protein of SARS-CoV-2 virus(Dar et al., 2020; Kar et al., 2020; Naz, Shahid, Butt, Awan, Ali, et al., 2020; Sanami et al., 2020; H. Singh, Jakhar, & Sehrawat, 2020; Tahir ul Qamar et al., 2020). Hence, in the present study, we also decided to target the S protein but not of only SARS-CoV-2 but all the seven known strains of human coronaviruses (HCoVs) i.e. both alpha (HCoV-NL63 and HCoV-229E) and beta coronaviruses (HCoV-OC43, HCoV-HKU1, MERS, SARS-CoV and SARS-CoV-2). As S protein is a common structural protein among the HCoVs, hence, we have taken up the S protein sequence to design a multi-epitope vaccine candidate which might confer immunity against all these HCoVs. Our hypothesis of identifying an epitope capable of providing immunity against various CoVs is also supported by the recent work of Mateus et al., where they have identified cross-reactive T-cell epitopes of SARS-CoV-2 in unexposed humans. They have concluded that pre-existing immunity could be derived from the exposure to other HCoVs like HCoV-OC43, HCoV-229E, HCoV-NL63, or HCoV-HKU1, which cause the common cold(Mateus et al., 2020). Yang et al have reported that a vaccine candidate targeted towards the RBD of the S protein of SARS-CoV-2 led to the induction of protective immunity(J. Yang et al., 2020). To the best of our knowledge, it is the first such work where S proteins of all the seven known HCoVs were screened to design broad-spectrum multiepitope vaccine candidate which might confer immunity from all the seven HCoVs. We have taken extra care to select T-cell epitopes and hence, the predicted T-cell epitopes from different web servers were further filtered based on epitope conservancy analysis and molecular docking analysis with HLA-A\*24:02 protein. Further, the best docked epitope-protein complex was also analyzed using MD simulations study for 50 ns. Thus, for the final vaccine construct, not only top-ranked predicted T-cell epitopes were considered but also based on conservation across different HCoV strains and peptide docking analysis. Finally, our designed vaccine candidate was analyzed for its potential to induce immune response through immune simulation studies and when compared to other recent publications, our designed vaccine candidate induced a higher level of immunoglobins (IgM + IgG),

immunocomplexes and interleukins (IL-2). The vaccine candidate showed a minimum conservancy of 73.97% across the seven HCoV's and hence, it might provide protection against all these HCoV's. A flowchart summarizing the methodology and tools employed for the current work has been depicted in Fig. 1.

## **2. Materials and Methods**

### **2.1 Experimental**

All the immunoinformatic exercises were carried out by using free web-based tools. For cheminformatics experiments like molecular docking and molecular dynamics simulations were carried out by using Maestro (Release 2018-3) which is a commercial small-molecule drug discovery suite from Schrodinger Inc. (USA) on an HP computer with Linux Ubuntu 18.04.1 LTS operating system. It had Intel Haswell graphics card, 8GB RAM and Intel Core i3 processor.

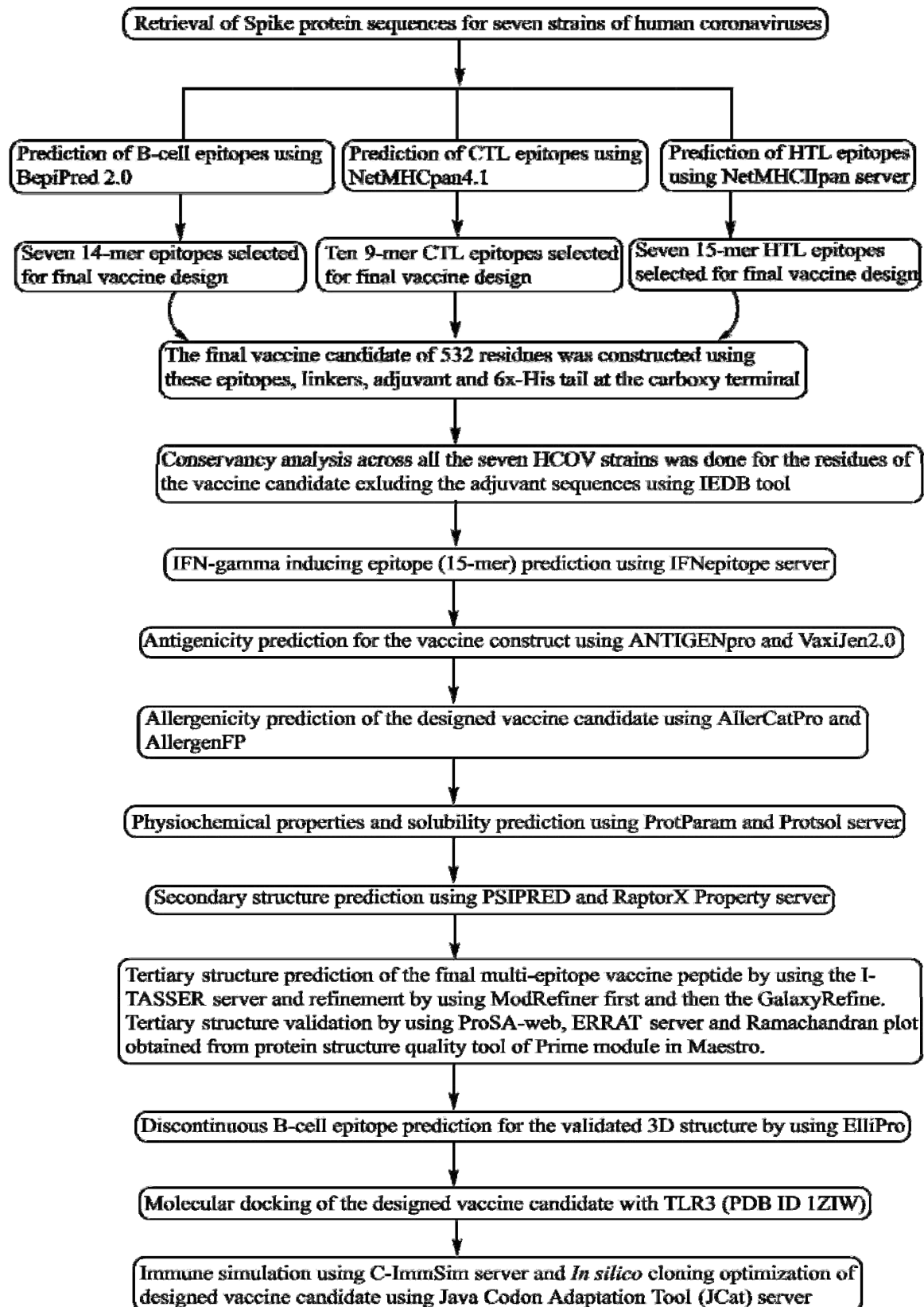


Fig. 1 A flowchart summarizing the methodology and tools employed for the current work.



## 2.2 Retrieval of S protein sequences

For the present study, we have taken the sequences of S protein from seven strains of coronaviruses (CoVs): (i) NL63 (ii) 229E (iii) OC43 (iv) HKU1 (v) MERS (vi) SARS-CoV and (vii) SARS-CoV-2. The sequences were retrieved in the FASTA format from the UniProtKB database (Apweiler et al., n.d.). The sequences of S protein from all the CoVs have been shown in Table 1.

**Table 1** The sequences of the S protein in FASTA format as retrieved from the UniProtKB database.

Sr. No.	CoV strain	The sequence of S protein in FASTA format
1	HCoV-229E	<p>&gt;AWH62679.1 spike protein [Human coronavirus 229E]</p> <p>MFVLLVAYALLHIAGCQTNGTNTSHSVCNGCVGHSENVFAVESGGYIPSNFAFNNWFL            LLTNTSSVVDGVVRSFQPLLLNCLWSVSGSQFTTGFVYFNGTGRGACKGFYSNASSDVI            RYNINFEENLRRGTILFKTSYGA VVFYCTNNTLVSGDAHIPSQVTLGNFYCFVNTTIGNE            TTSAFVGALPKTVREFVISRTGHFYINGYRYFSLGNVEAVNFNVTNAATTVCTVALASY            ADVLVNVSQTAIANIYCNSVINRLRCDQLSFDVPDGFYSTSPIQPVELPVSIVSLPVYHK            HTFIVLHVNFHQRPQKCYNCRPSVINITLANFNETKGPLCVDTSHTTQFVDNVKLA            RWSASINTGNCPFSFGK VNNFVKFGSVCFSLKDIPGGCAMPIMANLVNHNKSHNIGSLYV            SWSGDGVTGVPKPEGVSSFMNVTLNKCTKYNIYDVSQVGVIRISNDTFLNGITYTSTS            GNLLGFKDVTNGTIYSITPCNPPDQLVVYQQA VVGAMLSENFYSYGFNSVEMPKFFY            ASNGTYNCTDAVLTYSFGVCADGSIIA VQPRNVSYDSVSAIVTANLSIPSNWTTSVQVE            YLQITSTPIVVDCASTYVCNGNVRCVELLKQYTSACKTIEDALRNSAMLESADVSEMLTF            DKKAFTLANVSSFGDYNLSSVIPSLPRSGSRVAGRSAIEDILFSKLVTSGLGTVDADYKK            CTKGLSIADLACAQYYNGIMVLPGVADAERMAMYTGSLIGGIALGGLTSAASIPFLAI            QSRLNYVALQTDVLQENQKILAAAFNKAMTNIVDAFTGVNDAITQTSQALQTVATALN            KIQDVVNQQGNSLNHLTSQLRQNFQAISSSIQAIYDRLDIIQADQQVDRLITGRLAALNV            FVSHLTLYTEVRASRLAQKQVNECVKSKRYGFCGNGTHIFSLVNAAPEGLVFLH            TVLLPTQYKDV EAWSGLCVDGINGYVLRQPNLALYKEGNYYRITSRIMFEPRIPTIADF            VQIENCNVTFVNISRSELQTIPEYIDVNKTLQELSYKLPNYTPDLVVEQYNQTLNLNLS            EISTLENKSAELNYTVQKLQTLIDNINSLVLDLKWLN RVETIKWPPWWVWLCISVVLIF            VVSMLLCCCGSTGCCGFFSFCFASSIKGCCCESTKLPYDVEKIHQ</p>
2	HCoV-NL63	<p>&gt;APF29063.1 spike protein [Human coronavirus NL63]</p> <p>MKLFLILLVLPLASCFFTCNSNANLSMLQLGVDPDSSSTIVTGLLPTHWFCANQSTSVYSA            NGFFYIDVGNHRSAFALHTGYD VNYQYYIYVTNEIGLNASVTLKICKFGINTTFDFLSNS            SSSFDICVNLFFTEQLGAPLGITISGETVRLHLYNVTRTFYVPAAYKLTKLSVKCYFNYS            CVFSVVNATVTVNVTTHNGRVVNYTVCDCCNGYTDNIFSVQQDGRIPNGFPFNWFL            TNGSTLVDGVSRLYQPLRLTCLWPVPLKSSSTGFVYFNATGSDVNCNGYQHNSVADV            MRYNLNFSANSVDNLKSGVIVFKTLQYDVLFYCSNSSGVLDTTIPFGPSSQPYCFINS            TINTHVSTFVGLPPTVREIVVARTGQFYINGFKYFDLGFIEAVNFNVTTASATDFWTV            AFATFVDVLVNSATKIQNLLYCDSPFEKLQCEHLQFGLQDGFYSANFLDDNVLPETY            VALPIYYQHTDINFTATASFSGSCYVCKPHQVNISLNGNTSVCVRTSHFSIRIYNRVKS            GSPGSSWHIYLKSGTCPFSSKLNNFQKFKTICFSTVEVPGSCNFPLEATWHYTSYTIV            GALYVTWSEGNISITGVPPVSGIREFNLVNNCTKYNIYDYVGTGHIRSSNQSLAGGIT            YVNSGNLLGFKNVSTGNIFVTPCNQPDQVAVYQQSIIGAMTAVNESRYGLQNLQLP            NFYYVSNNGNCTTAVMTYSNFGICADGSLIPVRPNSSDNGISAITANLSIPSNWT</p> <p>TSVQVEYLQITSTPIVVDCASTYVCNGNPRCKNLLKQYTSACKTIEDALRLSAHLETNDV            SSMLTFDSNAFSLANVTSFGDYNLSSVLPQRNIRSSRIAGRSALEDLLFSKVVTSGLGTV            DVVDYKSCTKGLSIADLACAQYYNGIMVLPGVADAERMAMYTGSLIGGMVLGGLTSA</p>



		<p>AIPFSLALQARLNYVALQTDVLEQENQKILAAAFNKAINNIVASFSSVNDAITQTAEAIHT VTIALNKIQDVVNQQGSALNHLTSQLRHNFAQISNSIQAIYDRLDSIQADQQVDRLITGR LAALNAFVSQVLNKYTEVRSSRRLAQQKVNECVKSQSNRYGFCGNGTHIFSIVN</p> <p>SAPDGLLFLHTVLLPTDYKNVKAWSGICVDGIYGYVLRQPNLVLYSDNGVFRVTSRIM FQPRLPVLSDFVQIYCNVTFVNISRVELHTVIPDYVDVNKTLQEFANLQPKYVKNPFD LTPFNLTYNLSSSELKQLEAKTASLFQTTVELQGLIDQINSTYVDLKLLNRFENYIKWPW WWWLIISVVVFLSLLVFCCLSTGCCGCCNCLTSSMRGCCDCGSKLPYYEFKVVHV Q</p>
3	HCoV-OC43	<p>&gt;AIL49503.1 spike protein [Human coronavirus OC43]</p> <p>MFLILLISLPTAFAVIGDLNCPDLPRLKGSFNRRDTGPPSISTDTVDVTNGLGTYYYVLDLDR VYLNTTLFLNGYYPTSGSTYRNMALKGTDLLSKLWFKPPFLSDFINGIFAKVKNTKVF DGVMYSEFPAITIGSTFVNTSYSVVVQPRNTINSTQDGVNKLQGLLEVSVCQYNMCEYPH TICHPKLGNHFKELEWSDTGVVSLYKRNFTYDVNATYLYHFHYQEGGTFYAYFTDTG FVTKFLFNVYLGMALESHYYVMPLTCSIRLDIGFTLEYVVTPLTPRQYLLAFNQDGIIFNA VDCMSDFMSEIKCKTQSIAPPTGVYELNGYTVQPIADVYRRKPDLPNCNIEAWLNDKSV PSPLNWERKTFSNCFNMSLSMFSIQADSFTCNNIDA AKIYGMCFSSITIDKFAIPNRRKV DLQGLNGLYLQSSNYRIDTTATSCQLYYNLPAAANVSRSRFPSTWNKRFGFIEDSVFVP QPTGVFTNHSVVYAQHCFKAPKNFCPCSSCPGKNNGLGTCPAGTNHLTCDNLCTLDLPI FKAPDTYKCPQTKSLVGIGEHCSGLAVKSDYCGNNSCTCQPQAFGLWSADSCLOQDKC NIFANFILHDVNGLTCSIDLQKANTEIELGVCVNYDLYGISGQGIFVEVNATYYNSWQ NLLYDSNGNLYGFRDYITNRTFMHSCYSGRVSAAYHANSSEPALLFRNIKCNVFNNS LTRQLQPINYSFDSYLGCVVNAYNSTAVISVQTCDLTVGSGYCVDYSKNRRSRAITTYG RFTNFEPFTVNSVNDLSEVPGGLYEIQIPSEFTIGNMEEFIQTSSPKVTIDCAAFVCGDYA ACKLQLEVEYGSFCDNINAILTEVNELLDTTQLQVANSLMNGVTLSTKLDKGVNFVDD INFSPVLGCLGSECSKASSRAIEDLLFDKVKLSDVGFVEAYNNCTGGAEIRDLCVQSY KGIKVLPPLESENQISGYTLAATSASLFPWTAAGVPLYINQYRINGLGVTMDVLSQ NQKLIANAFNNALHAIQGFDAATNSALVKIQAVVNAEALNLLQQLSNRFGAISAS LQEILSRDALEAEAQIDRLINGRLTALNAYVSQQLSDSTLVKFSAAQAMEKVNECVKS QSSRINFCGNGNHIISLVQNPYGLYFIHFNYVPTKYVTAKVSPGLCIAGNRGIAPKSGY FVNVNNTWMTYTGSGYYYPEPITENNVMSTCAVNYTKAPYVMLNTSIPNLPDFKEEL DQWFKNQTSVAPDLSLDYINVTFLDLQVEMNRLQEAIKVLNHSYINLKDIGTYEYYVK WPWYVWLLICLAGVAMLVLLFFICCTGCG</p> <p>TSCFKKCGCCDDYTG YQELVIKTSHDD</p>
4	HCoV-HKU1	<p>&gt;BBA20986.1 spike protein [Human coronavirus HKU1]</p> <p>MFLIIFILPTTLAVIGDFNCTNSFINHYNKTIPIRESDVVDVSLGLGTYYYVLRVYVYVLDLDR FTGYFPKSGANFRDLALKGSTYLSLWYKPPFLSDFNNGIFSKVKNTKLYVNNTLYSEF STIVIGSVFVNTSYTIVVQPHNGILEITACQYTMCEYPHTVCKSKGSIRNESWHIDSSEPL CLFKNFTYNVSAWLYHFHYQERGVFYAYADVGMPTTFLFSLYLGLTILSHYYVMPL TCNAISPKTDNETLEYWVTPLSRRQYLLNFDEHGVITNAVDCSSSFLSEIQCKTQSFAPN TGVDYDLGFTVKPVATVYRRIPNLPDCDIDNWLNNVSVSPNLERRIFSNCFNLSTL LRLVHVDSFSCNNLDKSKIFGSCFNSITVDKFAIPNRRRDDLLQGLSSGLFQSSNYKIDISS</p> <p>SSCQLYYSLPLLNTINNFPSSWNRRYGFSGFNVSSYDVVYSDHCFVNSDFPCADP SVVNSCVKSKPLSAICPAGTKYRHCDLDTTLYVKNWCRCSCLPDPISTYSPNTPCQKK VVVGIGEHCPGLGINEEKCQTQLNHTSCSCSPDAFLGWSFDCSISNNRKNIFSNFIENGIN SGTTCSNDLLYSNTEVSTGVCVNYDLYGITGQGIFKEVSAAYNNWQNLLYDSNGNIIG FKDFLTNKTYTILPCYSGRVSAAFYQNSSPALLYRNLKCSYVLNNSIFISQPFYFDSYLG CVLNAVNLTSYSVSSCDLRMGSGFCIDYALPSSRRKRRGISSPYRFVTFEPFNVSVFVNS VETVGGLEFIQIPTNFITAGHEEFIQTSSPKVTIDCSAFVCSNYAACHDLLSEYGTFCNDI</p>

		<p>NSILNEVNDLLDITQLQVANALMQGVTLSSNLNTNLHSDVDNIDFKSLLGCLGSQCQSS  SRSLEEDLLFNKVKLSDVGFVEAYNNCTGGSEIRDLLCVQSFNGIKVLPILSETQISGYT  TAATVAAMFPPWSAAAGVPFSLNVQYRINGLGVTMDVLNKNQKLIANAFNKALLSIQ  NGFTATNSALAKIQSVVNANAQALNSLLQQLFNKFGAIISSSLQEILSRDLNLEAQVQIDR  LINGRLTALNAYVSQQLSDITLIKAGASRAIEKVNCEVKSQSPRINFCGNGNHLSLVQN  APYGLLFHFSYKPTSFKTVLVSPLGCLSGDRGIAPKQGYFIKQNDSWMFTGSSYYYPEP  ISDKNVVMNSCSVNFTKAPFTYLNNIPNLSDFEAEFSLWFKNHTSIAPNLTFNSHINAT  FLDLYYEMNVIQESIKSLNSSFINLKEIGTYEMYVKWPWYIWLIVLFIHFLMILFFICCC  TGCGSACFSKCHNCCDEYGGHNDFFVIKASHDD</p>
5	MERS-CoV	<p>&gt;APB87319.1 MERS-CoV spike protein [synthetic construct]</p> <p>MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFDFKTWPRPIDVSKADGIIYPQ  GRTYSNITITYQGLFPYQGDHGDYVYSAGHATGTPQKLFVANYSQDVQKQFANGFV  VRIGAAANSTGTVIISPTSATIRKIYPAFMLGSSVGNFSDGKMGRRFNHTLVLLPDGCG  TLLRAFYCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRN  CTFMYTYNITEDEILEWFGITQTAQGVHLFSSRYVDLYGGNMFQFATLPVYDTIKIYSII  PHSIRSIQSDRKAWAAFYVYKQLPLTFLDLSVDGYIRRAIDCGFNDSLQHLCSYESFDV  ESGVYSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPVYVNFKRLVFTNCNYNLTKLL  SLFSVNDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNKQSFNSPT  CLILATVPHNLTTITKPLKYSYINKCSRLSDDRTEVPQLVNANQYSPCVSIVPSTVWED  GDYYRKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDT  KIASQLGNCVEYSLYGVSGRGVFQNCATAVGVRQQRVYDAYQNLVGYYSDDGNYYC  LRACVSVVSVIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQT  PVGCVLGLVNSLFFVEDCKLPLGQSLCALPDTPTSTLTPRSVRSVPGEMRLASIAFNHPIQ  VDQLNSSYFKLSIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYQFC  SKINQALHGANLRQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSRARS  EDLLFDKVTIADPGYMQGYDDCMQQGPASARDLCAQYVAGYKVLPLMDVNMEAA  YTSSLLGSIAGVGTAGLSSFAAIPFAQSIFYRLNGVGITQQVLSNQKLIANKFNQALG  AMQTGFTTTNEAFQKQVQDAVNNAQALSCLASELSNTFGAISASIGDIIQRDLVLEQDA  QIDRLINGRLTTLNFAVAAQLVRSESAALSAQLAKDKVNECVKAQSKRSFGCGGTHI  VSFVFNAPNGLYFMHVGYYPSNHIEVVSAYGLCDAANPTNCIAPVNGYFIKTNTRIV  DEWYSYGSSFYAPEPITSLNTKYVAPQVTYQNISTNLPPPLGNSTGIDFQDELDEFFKN  VSTSIPNFGSLTQINTLLDLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIW  LGFIAGLVALALCVFFILCCTGCGTNCMGLKCNRCDDRYEEDLEPHKVVHVH</p>
6	SARS-CoV	<p>&gt;ABD75332.1 spike protein [Bat SARS CoV Rm1/2004]</p> <p>MKVLIFALLFSLAKAQEGCGIISRKPQPKMEKVSSRRGVYYNDDIFRSDVLHLTQDYF  LPFDSNLTQYFSLNIDSNKYTYFDNPILDFGDGVYFAATEKSNVIRGWIFGSSFDNTTQS  AIIVNNSTHIIIRVCNFNLCKEPMYTVSKGTQSSWVYQSAFNCTYDRVEKSFQLDTAPK  TGNFKDLREYVFNKKGFLRVYQTYTAVNLRGFPAGFSLRPLKLPFGINITSYRVV  MTMFSQFNSNPLPESAAYYVGNLKYTTFMLSFNENGTITDAVDCSQNPLAELKCTIKNF  NVSKGIYQTSNFRVPTQEVVRFNITNRCPPDKVFNASRFPNVYA WERTKISDCVADY  TVLYNSTSFSTFKCYGVSPSKLIDLCTS VYADTFLIRSEVRQVAPGETGVIADYNYKL  PDDFTGCVIAWNTAQDQGGQYYRSYRKEKLPFERDLSSDENGVYTLSTYDFYPSIP  VEYQATR VVLSFELLNAPATVCGPKLSTQLVKNQC VNFNFNGLRGTGVLTSSKRFQ  SFQQFGRDTSDFDTSVRDPQTEILDISPSCFSGGVSVITPGTNASSEVAVLYDVNCTDV  PTSIHADQLTPAWRVYSTGVNVFQQTQAGCLIGAEHVNASYECDIPIGAGICASYHTASV  LRSTGQKSIVAYTMSLGAENSIAYANNSIAIPTNFSISVTTEVMPSVIAKTSVDCCTMYICG  DSLECSNLLLQYGSFCTQLNRALTGIAIEQDKNTQEVFAQVKQMYKTPAIKDFGGFNF</p> <p>SQILPDPSPKTRSFIEDLLFNKVTLADAGFMKQYGECLGDISARDLCAQKFNGLTVLP  PLLDEMIAAYTAALVSGTATAGWTFGAGSALQIPFAMQMA YRFNGIGVTQNVLYEN  QKQIANQFNKAISIQIESLTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAIISSVLN</p>

		DILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSK RVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGV SNGTSWFITQRNFYSPQIITDNTFVAGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFK NHTSPDVLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYV WLGFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLYHT
7	SARS-CoV-2	>sp P0DTC2 SPIKE_SARS2 Spike glycoprotein OS=Severe acute respiratory syndrome coronavirus 2 OX=2697049 GN=S PE=1 SV=1  MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS  NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTEKSNIRGWIFGTTLDSKTQSLIV  NNATNVVIKVECFQFCNDPFLGVYYHKNNKSWMESEFRVYSSANNCTFEYVSQPFLM DLEGKQGNFKNLREFVFKNIDGYFKIYKHTPINLVRDLPGQFSALEPLVDLPIGINITRF QTLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPL SETKCTLSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNPK RISNCVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVRQIAPGQT GKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRFLRKSNLKPFERDISTEIQ AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVLSFELLHAPATVCGPKKSTN LVKNKCVNFNENGLTGTGVLTESNKKFLPFQGFGRDIADTTDAVRDPQTLLELDITPCSF GGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPTWRVYSTGNSVVFQTRAGCL IGAEHVNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNN SIAIPTNFTISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAV EQDKNTQEVFAQVKQYKTPPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGF IKQYGDCLGDIAARDLCAQKFNGLTVLPLLTDEMIAQYTSALLAGTITSGWTFGAGA ALQIPFAMQMA YRFNGIGVTQNVLYENQKLIANQFNASIGKIQDLSSTASALGKLQDV VNQNAQALNTLVKQLSSNFGAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQ QLIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQSAPHGVVFLHVTYV PAQEKNTTAPAICHGDKAHFPREGVVFVSNGTHWVFTQRNFYEPQIITDNTFVSGNCD VVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVLGDISGINASVVNIQKEIDRLN EVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCLKG CCSCGSCCKFDEDDSEPVLKGVKLYHT

### 2.3 Prediction of T-cell epitopes

Herein, we have predicted 9-mer CTL epitopes for each of the HCoV strain employing NetMHCpan4.1 web server which is also recommended by IEDB (Immune Epitope Database). The NetMHCpan-4.1 server (<https://services.healthtech.dtu.dk/service.php?NetMHCpan-4.1>) employs artificial neural networks (ANNs) to predict the binding of epitopes to any MHC molecule of known sequence. Its prediction method has been developed using a combined training set of >850,000 quantitative BA (Binding Affinity) and Mass-Spectrometry EL (Eluted Ligands) peptides (Reynisson, Alvarez, Paul, Peters, & Nielsen, 2020). In the present study, HLA-A\*24:02 supertype of human leukocyte antigen (HLA) allele has been employed for all the predictions. The rationale behind selecting this supertype allele was the availability of high-resolution X-ray crystal structure (PDB ID 3I6L) of HLA-A\*24:02 in complex with an

epitope N1 derived from SARS-CoV N protein. The other reason behind the selection of this supertype was that according to Lu et al., one of the most common HLA-A allele found in the Asian population is HLA-A\*24:02(Lu et al., 2005).

Best two 9-mer peptides from each HCoV strain was selected based on their binding affinity (nM) and percentile rank. Thus, a total of 14 CTL epitopes were selected for further study. But we wanted to refine our selection of CTL epitopes further and hence these epitopes were put for epitope conservancy analysis by employing the IEDB tool for the epitope conservancy analysis(Bui, Sidney, Li, Füsseder, & Sette, 2007a). For the present work linear epitope conservancy analysis was done with a sequence identity threshold of 100%. The epitopes with the highest minimum sequence identity were taken up for further analysis. Further, these 14 epitopes were put for molecular docking studies using peptide docking under the Biologics suite of Maestro (Release 2018-3). The X-ray crystal structure of HLA-A\*24:02 protein (PDB ID 3I6L) was retrieved from the protein data bank (PDB)(Berman et al., 2000; Liu et al., 2010). The retrieved protein structure was prepared using the protein preparation wizard of Maestro and the optimised protein structure was energetically minimized by employing OPLS3e force field(Kumar, Rathi, & Kini, 2019; Madhavi Sastry, Adzhigirey, Day, Annabhimoju, & Sherman, 2013). The epitope sequences were imported in the workspace and using peptide builder peptides were generated. The generated peptides were energetically minimized and docked against the receptor grid generated at the binding site of epitope N1 derived from SARS-CoV N protein on the HLA-A\*24:02 protein. The number of poses to return for each docking run was fixed at 10 as peptides are very flexible hence, several docking runs were used for each peptide. Each docking run employed different conformation of the epitope to enhance the sampling(Tubert-Brohman, Sherman, Repasky, & Beuming, 2013). The MM-GBSA (molecular mechanics generalized Born model and solvent accessibility) was used as the scoring method which is more accurate than the Glide score method but less fast(J. Li et al., 2011).

The epitopes were ranked based on the docking score and the highest-ranked CTL epitope was further put for molecular dynamics (MD) simulations analysis by using Desmond module of Maestro suite(Bowers et al., 2006). The docked complex of the selected epitopes with the HLA-A\*24:02 protein (PDB ID 3I6L) was first put for system builder where the Na<sup>+</sup> and Cl<sup>-</sup> were added to neutralize the charges. SPC (simple point charge) was used as the solvent system in an orthorhombic boundary box shape. The solvated system of epitope and the HLA-A\*24:02 protein was put for minimization by using SD (steepest descent) method.

The convergence threshold was fixed at 1 kcal/mol/Å with a maximum iteration of 2000. Slow relaxation protocol was followed for equilibration at a temperature of 300 K and pressure of 1.01325 bar. Nose–Hoover thermostat was used to maintain the temperature while Martyna–Tobias–Klein was the barostat employed (Kumar, Rai, Rathi, Agarwal, & Kini, 2020). The minimized protein-epitope complex was simulated for 50 ns to analyse the stability of the epitope-protein complex under the biological conditions. Based on the above methodologies i.e. percentile rank, binding affinity, minimum percentage conservancy and docking score, a total of ten CTL epitopes were selected for the design of vaccine candidate

#### **2.4 Prediction of linear B-cell epitope and Helper T-cell (HTL) epitopes**

B-cell epitopes are a specific part of the antigen recognized by the B lymphocytes of our immune system. For the present work, BepiPred-2.0 webserver (<http://www.cbs.dtu.dk/services/BepiPred/>) has been employed to predict linear B-cell epitopes of 14 sequence length for all the seven HCoV strains. BepiPred-2.0 is superior to other sequence-based linear epitope prediction tools and employs RF (random forest) algorithm trained on peptides annotated from antigen-antibody protein complexes (Jespersen, Peters, Nielsen, & Marcatili, 2017). HTL epitopes of 15 sequence length were predicted using the NetMHCII 2.2 Server (<http://www.cbs.dtu.dk/services/NetMHCII/>) for all the seven strains. The NetMHCII 2.2 server employs artificial neuron networks to predict the binding of epitopes to HLA-DQ, HLA-DP, and HLA-DR alleles (Nielsen & Lund, 2009).

#### **2.5 Multi-epitope vaccine candidate construction and its conservancy analysis**

The multi-epitope vaccine candidate was designed by linking different epitopes and adjuvants. CTLs epitopes were linked through an AAY linker while HTL and B-cell epitope was linked using GPGPG linker. At the N-terminal, a 50 S ribosomal protein L7/L12 (Locus RL7\_MYCTU) was added as an adjuvant (Accession no. P9WHE3) through an EAAAK linker. The adjuvant sequence was retrieved from the UniProt database (<http://www.uniprot.org/>). At the carboxy-terminal end, a 6x-His tail was also added to generate a final vaccine construct. The final designed vaccine peptide was put for conservancy analysis without the adjuvant across the seven strains of the selected S-protein sequences of the HCoVs (Bui, Sidney, Li, Füsseder, & Sette, 2007b).

#### **2.6 Prediction of IFN-gamma inducing epitope**

Cytokines like Interferon-gamma (IFN- $\gamma$ ) play a crucial role in the immune responses by stimulating NK (natural killer) cells and macrophages and provide an enhanced response to MHC antigen. Herein we have employed the IFNepitope server (<http://crdd.osdd.net/raghava/ifnepitope/scan.php>) to predict 15-mer IFN- $\gamma$  epitopes for the designed vaccine candidate. IFN- $\gamma$  epitopes were predicted separately for the adjuvant and the main vaccine peptide due to the limitation on the number of sequences that can be used as input in the server. The server employs support vector machine (SVM) hybrid approach for the prediction of IFN- $\gamma$  epitopes(Dhanda, Vir, & Raghava, 2013).

## **2.7 Antigenicity, allergenicity and assessment of physicochemical properties of the vaccine construct**

Antigenicity prediction for the vaccine construct was carried out using ANTIGENpro (<http://scratch.proteomics.ics.uci.edu/>) and VaxiJen2.0 server (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>). Like ANTIGENpro the antigenicity prediction by VaxiJen v2.0 is alignment-free and employs the physicochemical parameters of the protein(Doytchinova & Flower, 2007). Allergenicity of the designed vaccine construct was predicted using AllerCatPro (<https://allercatpro.bii.a-star.edu.sg/>) and AllergenFP (<http://ddg-pharmfac.net/AllergenFP/>) servers. AllerCatPro predicts the allergenicity of the query proteins by comparing their 3D structure along with their amino acid sequence with a dataset of 4180 unique protein sequences which have been reported to be allergenic(Maurer-Stroh et al., 2019). To predict the allergenicity of proteins, AllergenFP employs a descriptor-based fingerprint approach and thus is an alignment-free tool(Dimitrov, Naneva, Doytchinova, & Bangov, 2014). The physicochemical parameters like theoretical pI, in vitro and in vivo half-life, amino acid composition, instability index, molecular weight, aliphatic index and GRAVY (grand average of hydropathicity) of the vaccine construct was analysed by the ProtParam (<http://web.expasy.org/protparam/>) web server(Gasteiger et al., 2005). The solubility of the vaccine construct was analysed by the Protein-Sol (<https://protein-sol.manchester.ac.uk/>) web server(Hebditch, Carballo-Amador, Charonis, Curtis, & Warwicker, 2017).

## **2.8 Secondary and tertiary structure prediction of the vaccine construct**

The secondary structure of the designed vaccine construct was predicted by employing PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) and RaptorX Property (<http://raptorx.uchicago.edu/StructurePropertyPred/predict/>) web servers. PSIPRED 4.0



performed Position-specific iterated BLAST (Psi-Blast), for the identification of sequences which showed considerable homology to the vaccine peptide. These sequences were selected to build a position-specific scoring matrix. The RaptorX Property web server employs a template-free approach to predict the secondary structure properties of the query protein. Its algorithm is based on DeepCNF (Deep Convolutional Neural Fields) to simultaneously predict secondary structure (SS), disorder regions (DISO) and solvent accessibility (ACC)(S. Wang, Peng, Ma, & Xu, 2016). The tertiary structure of the designed multi-epitope vaccine candidate was predicted by the I-TASSER server (<https://zhanglab.ccmb.med.umich.edu/I-TASSER/>) webserver. The I-TASSER (Iterative Threading Assembly Refinement) server is an integrated platform to predict 3D structure from a peptide sequence. It generates 3D atomic models from multiple threading alignments and iterative structural assembly simulations and has been recognized as the best server for the prediction of protein structure(Roy, Kucukural, & Zhang, 2010). The best 3D-model for the vaccine construct obtained from I-TASSER web server was refined first by the ModRefiner (<https://zhanglab.ccmb.med.umich.edu/ModRefiner/>) and then by using the GalaxyRefine server (<http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE>). The refinement of protein structures by the ModRefiner server is based on a two-step, atomic-level energy minimization which leads to the improvements in both local and global structures(Xu & Zhang, 2011). GalaxyRefine method first rebuilds side chains and then carries out side-chain repacking and finally employs MD simulation for overall structural relaxation(Heo, Park, & Seok, 2013). Finally, the refined 3D model of the vaccine construct was validated using ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) and ERRAT server (<http://services.mbi.ucla.edu/ERRAT/>). Ramachandran plot was obtained from protein structure quality tool of the Prime module in Maestro11.4.

## **2.9 Protein-protein docking studies**

Molecular docking was carried out to understand the binding of the vaccine candidate with TLR-3 receptor as the immune response depends upon the interaction between an antigenic molecule (here the designed vaccine candidate) and an immune receptor (in this case TLR-3). The first step in this milieu was the prediction of the binding sites of the proteins which were carried out using the CASTp (<http://sts.bioe.uic.edu/castp/>) server(Binkowski, Naghibzadeh, & Liang, 2003). Molecular docking of the vaccine candidate with the TLR3 (PDB ID: 1ZIW) receptor was carried out using the HADDOCK 2.4 web server(Van Zundert et al., 2016) (<https://wenmr.science.uu.nl/haddock2.4/>) and GRAMM-X Simulation web



server(Tovchigrechko & Vakser, 2006) (<http://vakser.compbio.ku.edu/resources/gramm/grammx/>). The interactions were visualized using LIGPLOT v.4.5.3 (<https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/>)(Wallace, Laskowski, & Thornton, 1995). Finally, the binding energy of the top-ranked docking pose of the vaccine candidate-TLR4 complex was predicted using the PRODIGY webserver (<https://nestor.science.uu.nl/prodigy/>)(Xue, Rodrigues, Kastitis, Bonvin, & Vangone, 2016).

## **2.10 Prediction of discontinuous B-cell epitopes**

It has been reported that more than 90% of B-cell epitopes are discontinuous, i.e. they comprise of distantly separated segments in the pathogen protein sequence which are brought close by the folding of the protein(Barlow, Edwards, & Thornton, 1986). Herein, we employed ElliPro (<http://tools.iedb.org/ellipro/>) for the prediction of discontinuous B-cell epitopes of the validated 3D structure of the vaccine candidate(Ponomarenko et al., 2008).

## **2.11 *In silico* cloning and codon optimization of the vaccine construct**

Java Codon Adaptation Tool (JCat) server (<http://www.prodoric.de/JCat>) was employed for reverse translation and codon optimization so that the designed multi-epitope vaccine candidate could be expressed in a selected expression vector(Grote et al., 2005). Codon optimization was carried out to express the final vaccine candidate in the *E. coli* (strain K12) host. The output of the JCat server comprises of the codon adaptation index (CAI) and percentage GC content. CAI score reflects codon usage biases and ideally, the value of CAI should be 1.0 but a score above 0.8 is a good score(Grote et al., 2005). The gene sequences of the designed vaccine candidate were optimized and two restriction sites (*Nde* I and *Xho*) were introduced at the C and N-terminals of the sequence. In the last step, the optimized gene sequence of the vaccine construct along with the inserted restriction sites was inserted into the pET-28a (+) vector using SnapGene programme.

## **2.12 Immune simulation**

*In silico* immune simulations were carried out for further characterization of the immunogenicity and immune response profile of the designed multi-epitope vaccine candidate. Herein, we employed the C-ImmSim server (<http://150.146.2.1/C-IMMSIM/index.php>) for immune simulations studies which is an agent-based model that employs a position-specific

scoring matrix (PSSM)(Rapin, Lund, Bernaschi, & Castiglione, 2010). Three injections were given at four weeks intervals by keeping all simulation parameters at default setting with time steps set at 1, 84, and 168 (each time step is 8 hours and time step 1 is injection at time = 0).

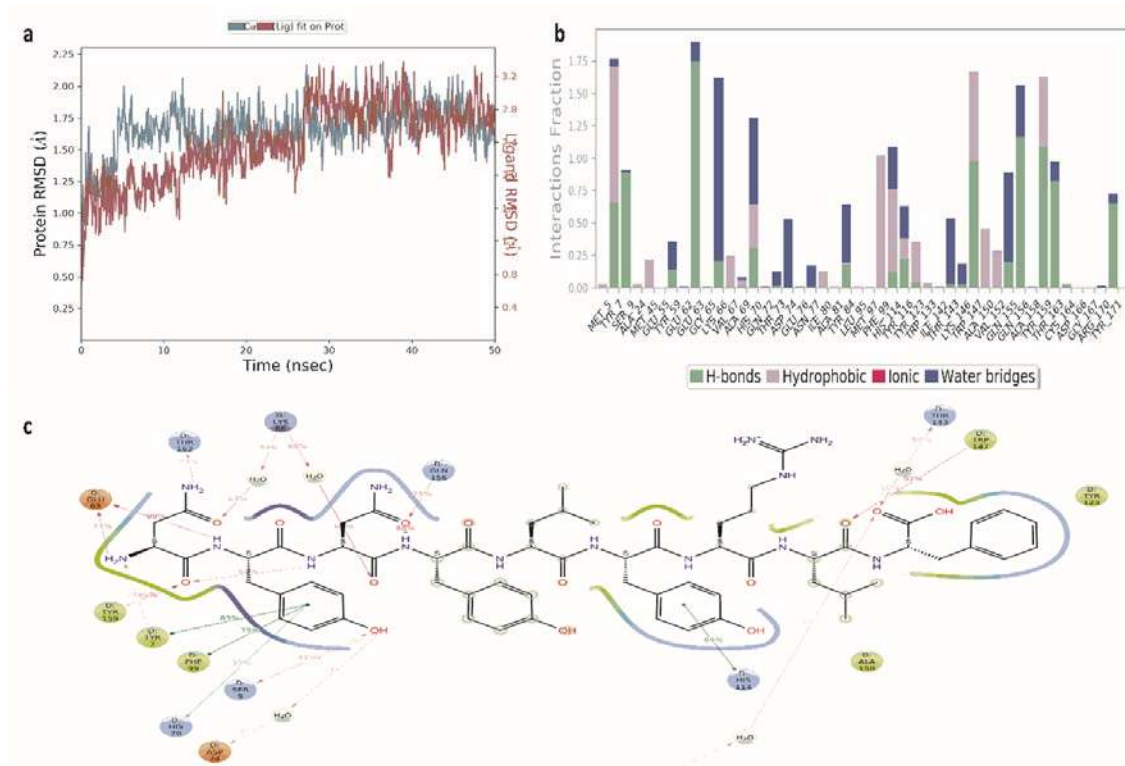
### **3. Results and Discussion**

#### **3.1 Prediction of T-cell epitopes**

The goal of T-cell epitope prediction is to identify peptides with shortest sequence length within an antigen which is capable of stimulating either CD4 or CD8 T-cells(Ahmed & Maeurer, 2009). Antigen-presenting cells (APCs) present T-cell epitopes on the surface where they bind to either MHC class I protein or MHC-II proteins. It has been reported that MHC-I protein present T-cell epitopes of 8-11 amino acid residue length while MHC-II presents longer peptides in the range of 13-17 amino acid residues(Steers et al., 2014). T-cell epitopes bound to MHC I protein are recognized by CD8 T-cells which become CTL (cytotoxic T lymphocytes), while those presented by MHC-II are recognized by CD4 T-cells and become helper T-cells(Sanchez-Trincado, Gomez-Perosanz, & Reche, 2017). Many authors have reported these tools for CTL epitope prediction not only against SARS-CoV-2 but also against SARS-CoV(Abraham Peele et al., 2020; Enayatkhani et al., 2020; Naz, Shahid, Butt, Awan, & Ali, 2020; Oany et al., 2014; Panda et al., 2020). But this is the first such analysis involving seven strains of the HCoV. Our present hypothesis to screen the S protein sequences of all the seven reported HCoVs for identifying CTL epitopes as a potential vaccine candidate against SARS-CoV-2 is also supported by the recent findings of Barun et al., where they have demonstrated the presence of S-cross-reactive T cells in unexposed healthy individuals probably because of previous exposure to endemic coronaviruses like 229E or OC43(Braun et al., 2020).

A total of 142 CTL (9-mer) epitopes were predicted to be strong binders for the seven selected S protein of different HCoVs using NetMHCpan4.1 server with default settings. We selected the best two epitopes based on their percentile rank, binding affinity (nM) and binding level for each of the HCoV strain. Thus, a total of fourteen 9-mer CTL epitopes were selected for further study. The conservancy analysis of these fourteen CTL epitopes showed that QYIKWPWYI epitope from SARS-CoV-2 strain had a minimum conservancy of 66.67% across all the strains. Another epitope with sequence YYNKWPWYI from MERS strain showed a minimum conservancy of 55.56% across all the strains. Hence, these two CTL epitopes were selected for the final vaccine design. Peptide docking analysis of these epitopes

with HLA-A\*24:02 protein showed that these peptides showed favourable binding with the protein. But we selected only one CTL epitope (NYNYLYRLF) which showed the highest docking score of -12.759. MD simulations study for this epitope showed that the epitope-protein complex was stable throughout the simulations period as reflected by their RMSD (root mean square deviation) plot in Fig. 2a. A histogram representing the interaction of the epitopes with the amino acid residues of the protein was also plotted as shown in Fig. 2b. The non-bonding interactions were similar to the ones observed during peptide docking studies, but water bridges were formed with LYS66, HIS70, ASP74, TYR84, THR143 and GLN155 residues. A 2D interaction diagram (Fig. 2c) suggested that apart from intermolecular H-bond interactions, there were quite several intramolecular H-bond interactions which could make the epitope-protein complex more stable. The importance of intramolecular H-bonding for the epitopes binding to HLA-A\*24:02 protein has also been highlighted by Liu et al (Liu et al., 2010). Several other authors like Peele et al. and Enayatkhani et al. have also employed MD simulations studies to analyse the stability of the designed epitopes and protein complex (Abraham Peele et al., 2020; Enayatkhani et al., 2020). Hence, for the final vaccine design, a total of ten CTL epitopes (Table 2) were selected out of which seven were the best-ranked epitope from each strain, two were the best-conserved epitope across all the strains and one was the best-docked epitope with HLA-A\*24:02 protein.



**Fig. 2** MD simulations report of the docked complex of CTL-epitope “NYNYLYRLF” with HLA-A\*24:02 protein (PDB ID 3I6L). (a) RMSD plot of CTL-epitope 2 in complex with HLA-A\*24:02 protein. (b) Histogram of the CTL-epitope contact with HLA-A\*24:02 protein. (c) CTL-epitope contact summary with HLA-A\*24:02 protein (PDB ID 3I6L) in 2D.

### 3.2 Prediction of linear B-cell epitope and Helper T-cell (HTL) epitopes

A total of seven 14-mer B-cell epitopes (Table 2) were selected based on the epitope threshold and relative surface accessibility as predicted by BepiPred2.0 server. A total of seven 15-mer high-binding MHC-II epitopes (Table 2) for human alleles HLA-DR, HLA-DQ and HLA-DP as predicted by the NetMHCII 2.2 web server was selected. Some of the B-cell epitopes and HTL epitopes overlapped.

**Table 2** List of the selected CTL, HTL and linear B-cell epitopes from S protein sequences of different HCoV strains for constructing the multi-epitope vaccine candidate.

HCoV strains	CTL epitopes (9-mer)	B-cell epitopes (14-mer)	HTL epitopes (15-mer)
229E	FYINGYRYF	PDGFYSTSPIQPVE	VQVEYLQITSTPIVV
NL63	FYINGFKYF	ITGVPPYPVSGIREF	VSTFVGILPPTVREI
OC43	TYYNSWQNL	KNRRSRRAITTYR	QSYKGIKVLPLLSE
HKU1	VYLNTTLF	LGINEEKCGTQLNH,	QSFNGIKVLPPILSE
MERS	LYGGNMFQF YYNKWPWYI*	IADPGYMQGYDDCM	TIKYYSIIPHSIRSI
SARS	VYSTGVNVF	NVSKGIYQTSNFRV	IVAYTMSLGAENSIA
SARS-CoV-2	VYSTGSNVF QYIKWPWYI* NYNYLYRLF#	APATVCGPKKSTNL	IVAYTMSLGAENSIA

\*CTL epitopes selected based on the epitope conservancy analysis.

#CTL epitope selected based on the docking score.

### 3.3 Multi-epitope vaccine candidate construction and its conservancy analysis

The final multi-epitope vaccine construct comprised of 10 CTL epitopes, 7 linear B-cell epitopes, and 7 HTL epitopes. The predicted CTL epitopes were joined by using AAY linker while B-cell epitope and TL epitopes were fused with GPGPG linkers. The 50S ribosomal L7/L12 (Locus RL7\_MYCTU) was selected as the adjuvant to accentuate antigen-specific immune responses and fused at the N-terminal by using EAAAK linker. Finally, to help in protein identification and purification a 6xHis tail was inserted at the carboxy-terminal end of the vaccine peptide. Thus, the final multi-epitope vaccine candidate was constructed with a total of 532 amino acid residues derived from 24 merged multi-epitopes. The final designed vaccine peptide was put for conservancy analysis without the adjuvant across the seven strains of the selected S-protein sequences of the HCoV strains. The sequences of the vaccine candidate showed a minimum and maximum conservancy of 73.97% across all the strains of the HCoVs. A higher degree of conservation across all the HCoV strains might confer broader protection across multiple strains by recognition of the conserved epitopes by our immune system. Hence, we hypothesize that this vaccine construct which has more than 73% of sequences conserved across the seven HCoV strains might be a potential broad-spectrum vaccine candidate against HCoVs.

### **3.4 Prediction of IFN-gamma inducing epitope**

Due to the restriction on the number of residues that can be used as input in the IFNepitope server, IFN- $\gamma$  inducing epitopes were predicted separately for the adjuvant and main vaccine peptide. For the adjuvant, a total of 127 potential IFN- $\gamma$  inducing epitopes (15-mer) were predicted. For the main vaccine peptide, 389 potential 15-mer epitopes were predicted out of which 82 had a positive prediction score.

### **3.5 Antigenicity, allergenicity and assessment of physicochemical properties of the vaccine construct**

As per the predictions of the VaxiJen 2.0 server, the antigenicity of the vaccine construct along with the adjuvant sequence was found to be 0.7106 in a bacteria model at a threshold of 0.5 while the main vaccine sequence without the adjuvant showed a score of 0.8135. As per the predictions of ANTIGENpro server the probable antigenicity of the vaccine candidate with the adjuvant was 0.829721 while without adjuvant was 0.869507. This result indicates that the designed vaccine candidates are antigenic (with or without adjuvant) in nature. The main vaccine sequence seems to be more antigenic without the adjuvants as per the predicted

scores. The allergenicity of the vaccine candidate was predicted by the AllerCatPro and AllergenFP servers and both predicted it to be non-allergenic. The molecular weight (MW) of the final vaccine candidate was predicted to be 56.5 kDa. The theoretical isoelectric point (pI) value was predicted to be 7.11. The half-life was predicted to be 30 hours in mammalian reticulocytes (*in vitro*), and >10 hours in *E. coli* and >20 hours in yeast (*in vivo*). The instability index (II) was computed to be 22.81 which classifies the vaccine candidate to be stable. It had an aliphatic index of 79.29 which suggests it to be thermostable (Ikai, 1980). The GRAVY score for the vaccine candidate was computed to be -0.407 which indicates that it could be hydrophilic in nature. (Ali et al., 2017) The predicted scaled solubility value was computed to be 0.432 which indicates that the vaccine construct would have an acceptable solubility profile.

### **3.6 Secondary and tertiary structure prediction of the vaccine construct**

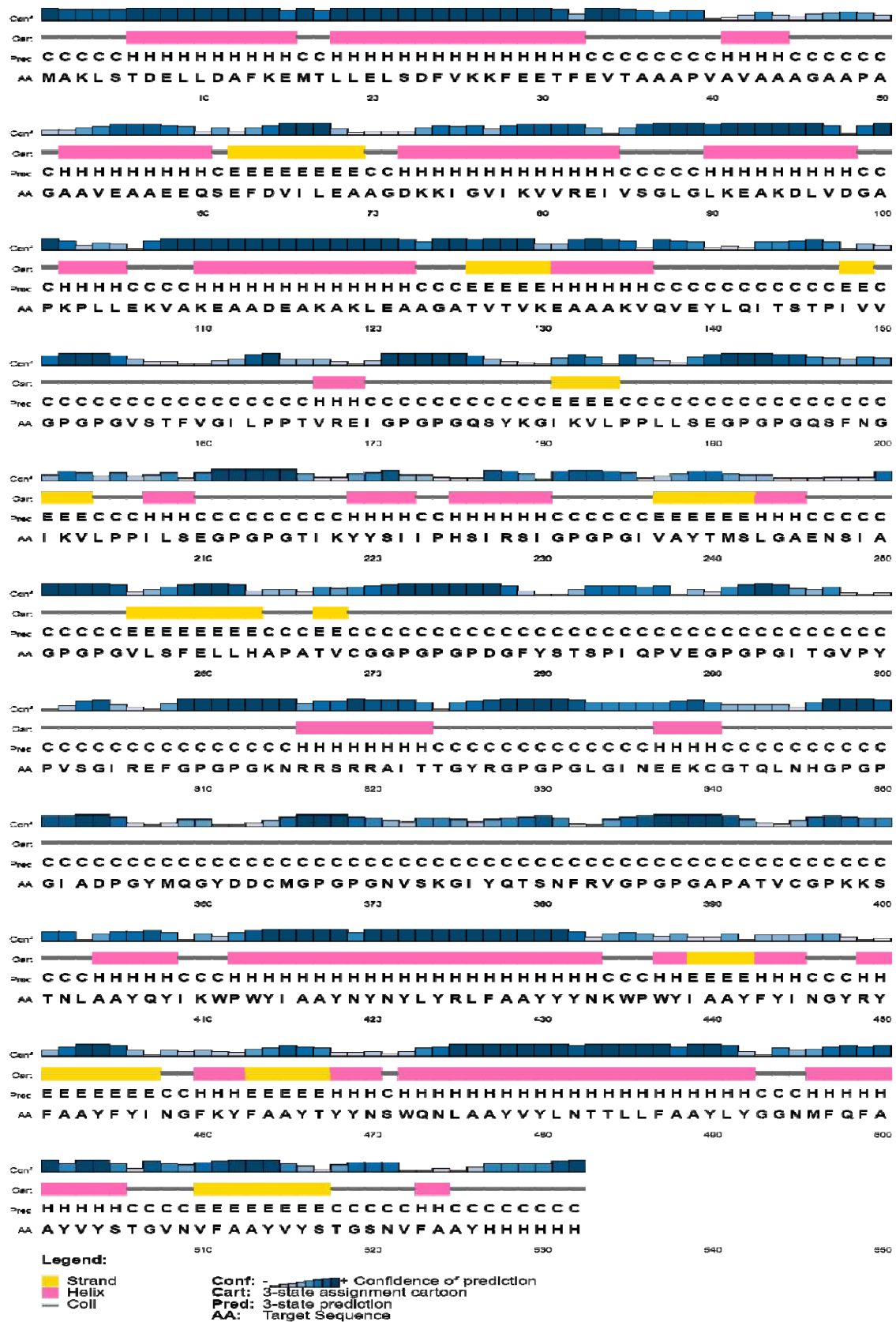
The secondary structure of the final multi-epitope vaccine peptide was computed, and it was predicted to have 20% alpha-helix, 20% beta-sheet and 59% coil (Fig. 3). The solvent accessibility of amino acid residues as predicted by RaptorX property server showed that 49% of the total residues were exposed, 24% were medium exposed and 26% was buried. A total of 42 residues i.e. 7% of residues were computed to be in the disordered region. Five tertiary structure models of the final vaccine construct were predicted by the I-TASSER server based on 10 threading templates, of which PDB hits 1dd4A, 5tsjN, 1rquA, 2nbiA, 1dd3A, 2ocwA and 2ftc were best hits. All of the 10 selected templates showed good alignment which was evident from their *Z-score* values which ranged 1.02 to 5.35. The first model (Fig. 4a) was selected for further refining which had an estimated C-score of 0.77, TM score of  $0.62 \pm 0.14$  and RMSD value  $9.2 \pm 4.6 \text{ \AA}$ . The C-score range is between -5 and 2, where higher values towards 2 indicate higher confidence. The TM-score is an indication of the structural similarity between two structures and unlike RMSD, it is not sensitive to local error (Y. Zhang & Skolnick, 2004). From the TM value of the selected model, we can infer that the model has correct topology. This selected model was further put for refining first by ModRefiner and then by GalaxyRefine server which generated five models. According to the model quality scores (Table 3) for all the five refined models, model 3 (Fig. 4b) was found to be the best model. It had a GDT-HA score of 0.9408, RMSD value 0.443, MolProbity score of 2.072, clash score of 11.3 and the poor rotamers score was 0.2. The Ramachandran plot score for model 3 was predicted to be 91.5% which was less when it was put for validation by the Ramachandran plot analysis. The Ramachandran plot of the selected model 3 (Fig. 4c)

suggests that it's 94.528% of the residues are in the favourable region with 2.64% of residues in the allowed region. The overall quality factor was computed by the ERRAT server and for the selected model 3, it was predicted to be 86.235. The ProSA-web computed the Z-score for model 3 and it was found to be -2.41 (Fig. 4d).

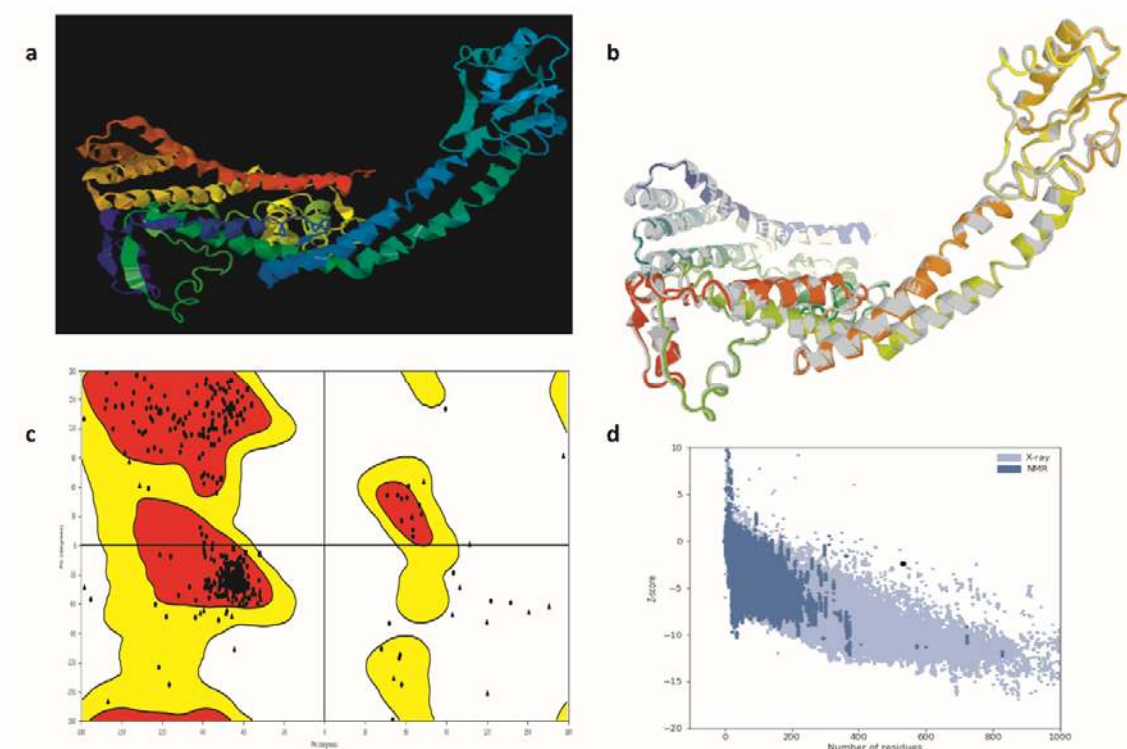
**Table 3** Detailed result of the five models refined by the GalaxyRefine server.

Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favoured
Initial	1.0000	0.000	3.083	61.4	2.0	88.1
MODEL 1	0.9314	0.467	2.126	12.6	0.5	91.1
MODEL 2	0.9305	0.476	2.111	11.6	0.5	90.6
MODEL 3	0.9408	0.443	2.072	11.3	0.2	91.5
MODEL 4	0.9375	0.465	2.180	14.2	0.5	90.9
MODEL 5	0.9337	0.457	2.122	12.1	0.5	90.8





**Fig. 3** Cartoon representation of the secondary structure of the vaccine sequence as predicted by the PSIPRED.

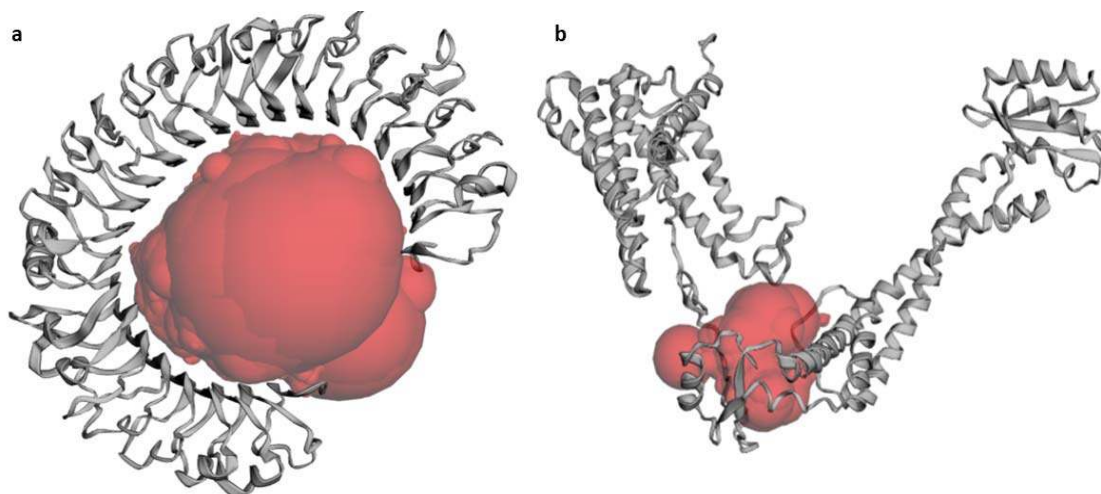


**Fig.4** Tertiary structure modelling, refining and validation. (a) The final homology 3D model of the vaccine construct as computed by the I-TASSER. (b) Refinement of the selected I-TASSER 3D model by GalaxyRefine where the refined 3D model (coloured) has been superimposed on the initial crude model (grey). (c) Ramachandran plot of the refined 3D model of the vaccine candidate (d) Plot by ProSA-web showing the Z-score.

### 3.7 Protein-protein docking studies

The CASTp 3.0 server was employed to compute the protein binding and hydrophobic interaction sites on the surface of the TLR-3 receptor and the final refined 3D model of the vaccine construct. A total of 105 binding pockets were predicted for the vaccine candidate with different molecular surface area and volume. Similarly, for the TLR-3 protein (PDB ID 1ZIW), a total of 77 binding pockets were predicted. For the present work, a pocket (Fig. 5a) with a molecular surface area of  $2125.000 \text{ \AA}^2$  and volume of  $13438.169 \text{ \AA}^3$  was selected for the TLR-3 receptor. For the vaccine candidate, the selected pocket (Fig. 5b) had a molecular surface area of  $917.628 \text{ \AA}^2$  and volume of  $2414.204 \text{ \AA}^3$ . Both the proteins were put for molecular docking simulations first using the GRAMM-X server which led to the generation of ten binding poses. The poses were visualised using the DIMPLOT of LIGPLUS version 4.5.3 and the best-docked pose of the vaccine-TLR-3 complex in 3D has been depicted in Fig. 6a. The non-bonding interactions between the docked protein complex were also visualized in 2D and have been shown in Fig. 6b. It was observed that ARG306, TYR526,

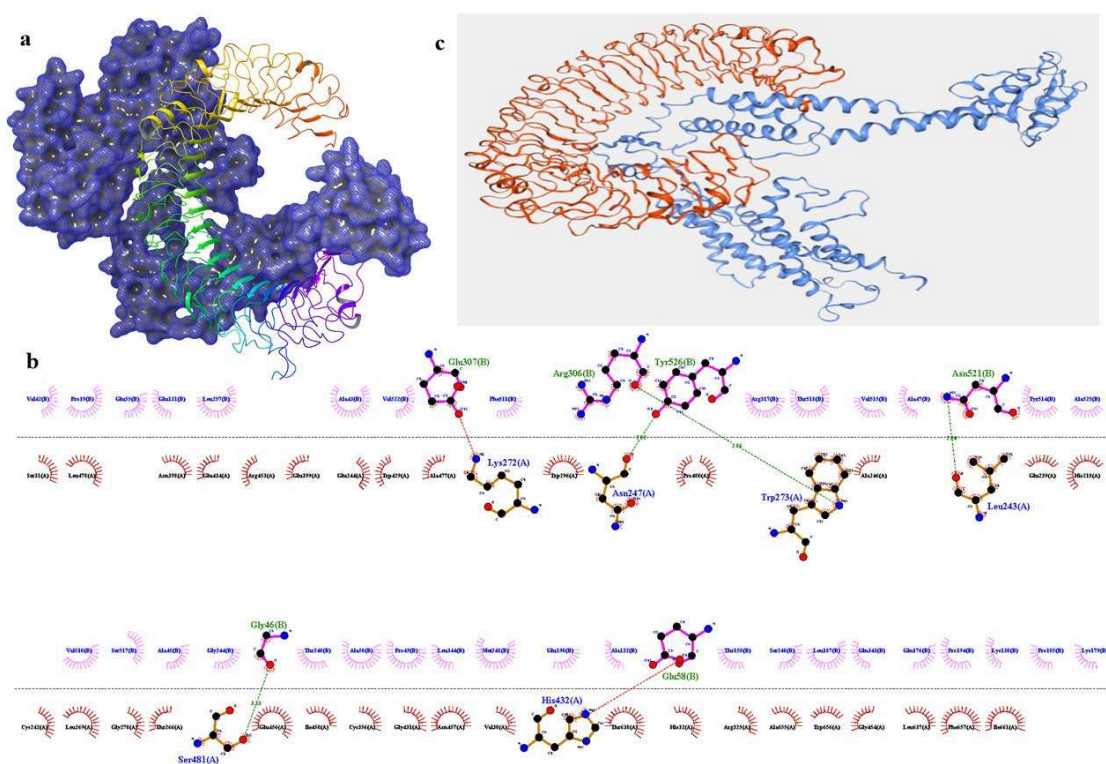
ANS521 and GLY46 residues of the vaccine construct showed hydrogen bonding with TRP273, ANS247, LEU243 and HIS432 residues of the TLR-3 protein respectively. The 2D interaction plot also showed possible salt bridge between GLU307 and GLY46 of the vaccine construct with the LYS272 and SER481 of the TLR-3 protein. Further, the best-docked pose was put for the prediction of binding affinity ( $\Delta G$ ) and dissociation constant ( $K_d$ ). The binding energy ( $\Delta G$ ) for the vaccine-TLR-3 complex was computed to be -20.8 kcal/mol with a  $K_d$  value of  $5.9 \times 10^{16}$  M at 25.0 °C. The Prodigy server also calculated the number of interfacial contacts (ICs) per property within the threshold distance of 5.5 Å and its ICs charged-charged, ICs polar-polar, and ICs apolar-apolar was computed to be 14, 7 and 64 respectively. It suggests that most of the interfacial contacts were hydrophobic in nature.



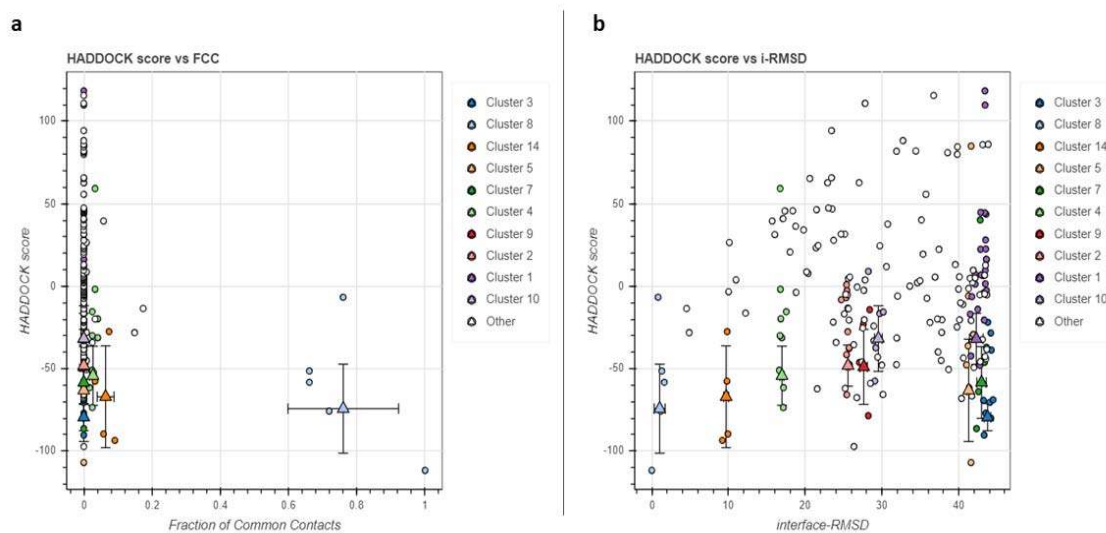
**Fig. 5** The selected binding pocket of the proteins as predicted by the CASTp3.0 server. (a) The predicted binding pocket of the TLR-3 protein. (b) The predicted binding pocket of the vaccine candidate.

The HADDOCK2.4 server clustered 103 structures in 14 clusters, which represent 51% of the water-refined model generated by the HADDOCK. The clusters were ranked on several parameters and out of the 14 clusters, cluster 3 (cluster size 10) was reported to be the best cluster. It had a HADDOCK score of -79.4 +/- 7.1, RMSD from the overall lowest-energy structure was 43.8 +/- 0.4, Van der Waals energy -100.6 +/- 15.7 kcal/mol, Electrostatic energy -415.7 +/- 78.8 kcal/mol, Desolvation energy -27.2 +/- 4.0 kcal/mol and Z-Score -1.5. The detailed result for all the clusters has been depicted in Fig. 7 and Fig. 8. The best pose from cluster 3 was also put for the prediction of binding affinity ( $\Delta G$ ) and dissociation constant ( $K_d$ ). The binding energy ( $\Delta G$ ) for was computed to be -20.3 kcal/mol with a  $K_d$  value of  $1.3 \times 10^{15}$  M at 25.0 °C which was similar to that of the best docked pose generated

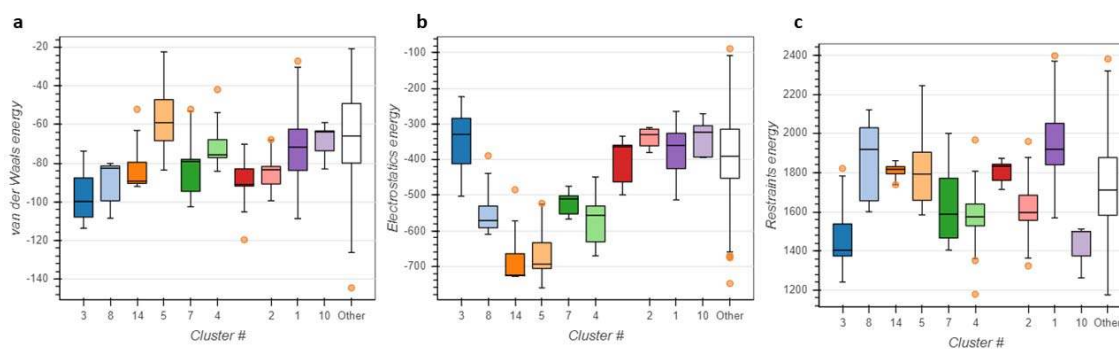
by the GRAMM-X server. The ICs charged-charged, ICs polar-polar, and ICs apolar-apolar were computed to be 16, 7 and 25 respectively.



**Fig. 6** Visualization of the best-docked poses for the vaccine-TLR-3 receptor complex. (a) 3D-interaction diagram of the protein-protein complex generated by the GRAMM-X server where blue mesh structure represents the vaccine candidate while coloured ribbon shows the TLR-3 receptor. (b) 2D-interaction diagram of the protein-protein complex generated by the GRAMM-X server where chain A is from the TLR-3 receptor and chain B is the vaccine construct. (c) 3D-interaction diagram of the protein-protein complex generated by the HADDOCK2.4 server where blue ribbon represents the vaccine candidate while orange ribbon shows the TLR-3 receptor.



**Fig. 7** Bokeh plot showing the result of various parameters for all the clusters by the HADDOCK2.4 server. (a) Plot between HADDOCK score and FCC (fraction of common contacts). (b) Plot depicting HADDOCK score vs interface-RMSD.



**Fig. 8** Bokeh plot depicting the analysis of various energy parameters for all the clusters as predicted by the HADDOCK2.4 server. (a) Van der Waals energy (b) Electrostatic energy (c) Restraints energy.

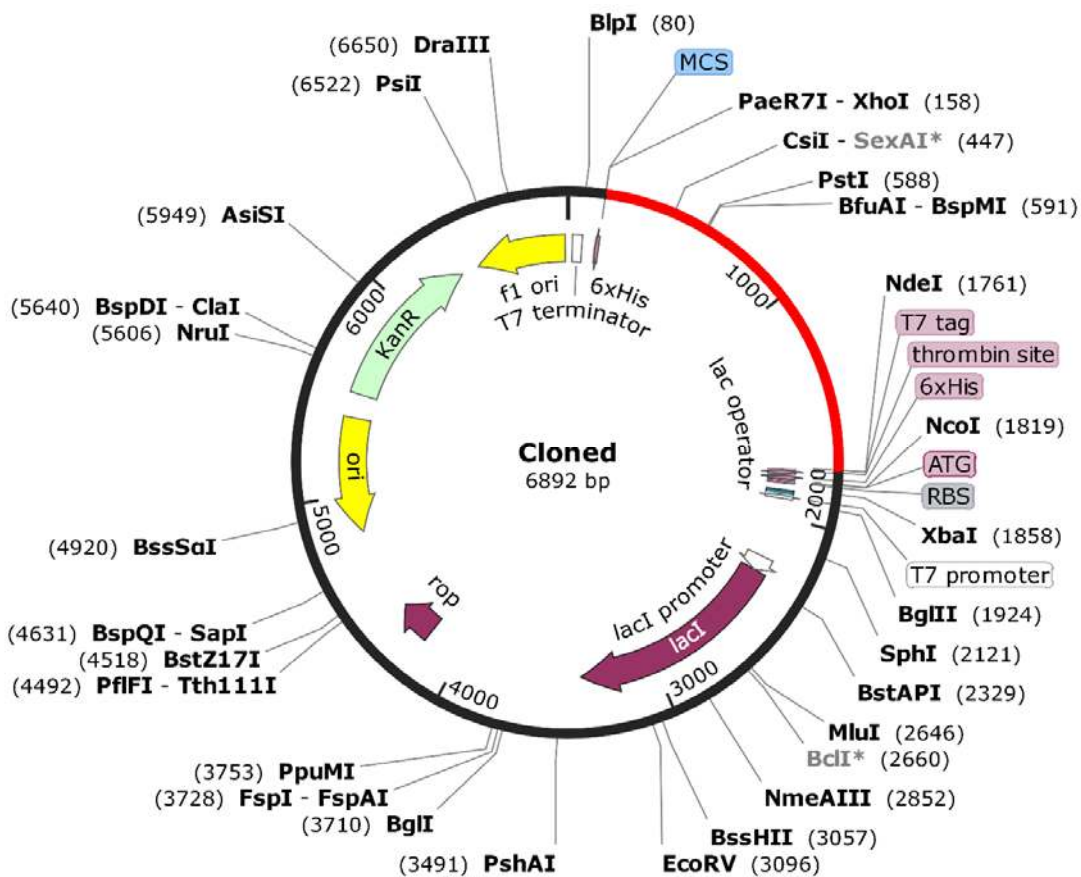
### 3.8 Prediction of discontinuous B-cell epitopes

The ElliPro server predicted a total of 7 discontinuous B-cell epitopes of the validated 3D structure of the vaccine candidate comprising of 270 residues. The conformational B-cell epitopes had scores ranging from 0.598 to .875 where the best epitope had only three residues while the last ranked epitope comprised of 8 residues. This study suggests that the designed multi-epitope vaccine might lead to heightened immune response though the conformational B-cell epitopes also.

### 3.9 *In silico* cloning and codon optimization of the vaccine construct



Immunoreactivity through serological assays are needed to validate a designed vaccine candidate and the first step for it is the expression of the multi-epitope vaccine candidate in a suitable host (Gori, Longhi, Peri, & Colombo, 2013). *Escherichia coli* expression systems have been widely used to produce recombinant proteins and hence to achieve maximal protein expression, we selected *E. coli* (strain K12) system for the codon optimization by the Java Codon Adaptation Tool (JCat) (Chen, 2012; Rosano & Ceccarelli, 2014). The optimized codon sequence for the final vaccine candidate was 1596 with a CAI value of 0.9237 and 53.57% of GC content. Ideally, the GC content should be in the range of 30% to 70% and hence, our vaccine candidate might show high-level expression in the bacterial host system. Finally, the optimized gene sequences of the designed vaccine candidate with two restriction sites (*Nde* I and *Xho*) at the C and N-terminals of the sequence were inserted into the pET-28a (+) vector using SnapGene software to generate the sequences of the recombinant plasmid (6892 bp) as shown in Fig. 9.

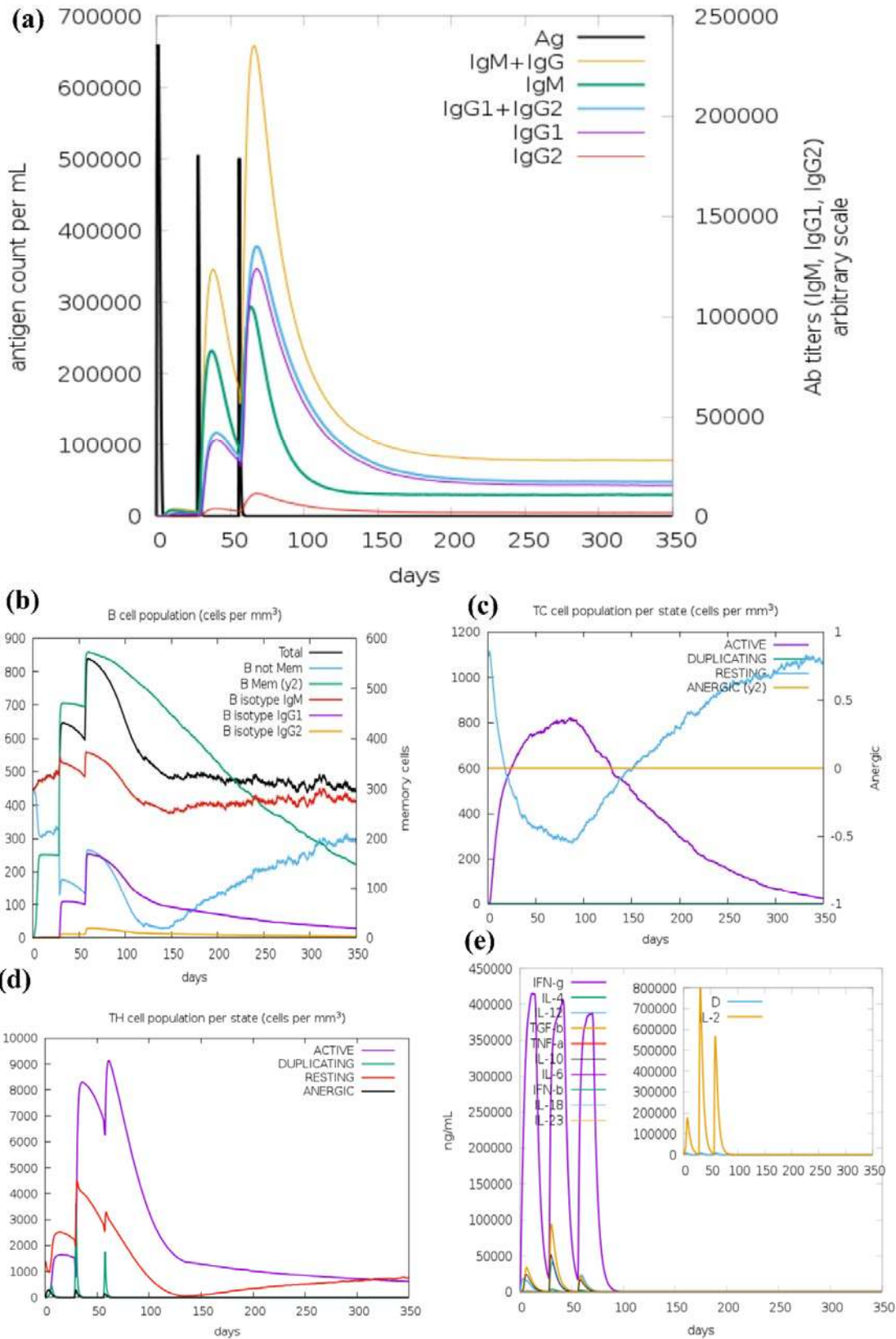


**Fig. 9** *In silico* cloning of the final multi-epitope vaccine candidate into the pET28a (+) expression vector where the red portion depicts the gene coding region for the vaccine and the black circle shows the vector backbone.

### 3.10 Immune simulation

The immune simulation by the C-ImmSim server showed results which are consistent with actual immune responses. The primary response after the first dose of the vaccine was characterized by the increased levels of IgM as shown in Fig. 10a. After the second and third dose of vaccine, there was a remarkable increase in the levels of IgG1, IgG1 + IgG2, IgM, and IgG + IgM antibodies (Fig. 10a). The level of IgM peaked after the third dose to around 2,30,000 which suggests a strong humoral immune response which is better than reported by Peele et al. for multi-epitope vaccine design against SARS-CoV-2 (Abraham Peele et al., 2020) and Shey et al for multi-epitope vaccine candidate against onchocerciasis (Shey et al., 2019). It was also observed that after the second exposure the population of the antigen was also falling which indicates increased clearance by the immune system. B-cell population (Fig. 10b) and especially the B-memory cell population ( $>800$  cells per  $\text{mm}^3$ ) also increased and peaked after the third dose. Similarly, a greater response was observed in the cytotoxic T-cells (Fig. 10c) and helper-T cell populations (Fig. 10d) with corresponding memory development which lasted for many months. It was also observed that the levels of IFN- $\gamma$  and IL-2 (Fig. 10e) increased significantly after the first dose of the vaccine and maintained their peak levels after multiple exposures to the antigen. The Simpson index, D which indicates a measure of diversity, predicted this vaccine candidate to have a diverse immune response. This diverse immune response might be possible as the vaccine comprises of different epitopes.





**Fig. 10** Results of immune simulation by the C-ImmSim server. (a) Immunoglobulin and immunocomplexes production in response to antigen injections at different time intervals (black vertical lines); specific subclasses have been shown as coloured peaks. (b) The increase in the populations of B-cell post three vaccine doses of the vaccine. (c) The evolution of T-cytotoxic cell population per state post three vaccine doses. (d) The evolution of T-helper cell population per state post three vaccine doses. (e) The main plot depicts the levels of cytokine levels post three vaccine doses. The insert plot depicts IL-2 level with the Simpson index, D which is indicated by the green dotted line.

#### 4. Conclusions

In the present study, several machine learning based in silico tools were used to design a potential broad-spectrum multi-epitope vaccine candidate against spike protein of human coronaviruses. To the best of our knowledge, it is one of the first study, where multiple B-cell epitopes and T-cell epitopes (CTL and HTL) were predicted from the spike protein sequences of all the seven known human coronaviruses. The final vaccine candidate was found to have a minimum conservancy of 73.97% across all the seven HCoV strains and it might provide protection from SARS-CoV-2 as well as other HCoVs. The designed vaccine is predicted to be antigenic and non-allergenic. Its physicochemical properties are also in the acceptable range. Molecular docking of the refined tertiary structure of the vaccine candidate with the TLR-3 protein also indicated its favourable binding with the TLR-3 receptor. In silico cloning study indicated that the designed vaccine candidate might show high-level expression in the bacterial host system. The immune simulation analysis suggests that the vaccine might elicit a strong immune response and when compared to other similar publications which have reported immune simulation for the designed vaccines, our designed vaccine candidate induced a higher level of immunoglobins (IgM + IgG), immunocomplexes and interleukins (IL-2). The Simpson index, D which indicates a measure of diversity, predicted this vaccine candidate to have a diverse immune response. The findings of this in silico studies need to be validated through various immunological assays.

**Acknowledgements:** We are grateful to Manipal-Schrödinger Centre for Molecular Simulations, Manipal Academy of Higher Education and Manipal College of Pharmaceutical Sciences for providing necessary supports and facilities to carry out the present research work.

**Conflicts of interest:** Authors declare no conflict of interest.

#### References

Abdelmageed, M. I., Abdelmoneim, A. H., Mustafa, M. I., Elfadol, N. M., Murshed, N. S., Shantier, S. W., & Makhawi, A. M. (2020). Design of a Multiepitope-Based Peptide Vaccine against the e Protein of Human COVID-19: An Immunoinformatics Approach.

*BioMed Research International*, 2020. <https://doi.org/10.1155/2020/2683286>

- Abraham Peele, K., Srihansa, T., Krupanidhi, S., Ayyagari, V. S., & Venkateswarulu, T. C. (2020). Design of multi-epitope vaccine candidate against SARS-CoV-2: a *in-silico* study. *Journal of Biomolecular Structure and Dynamics*, 1–9. <https://doi.org/10.1080/07391102.2020.1770127>
- Ahmed, R. K. S., & Maeurer, M. J. (2009). T-cell epitope mapping. *Methods in Molecular Biology*, 524, 427–438. [https://doi.org/10.1007/978-1-59745-450-6\\_31](https://doi.org/10.1007/978-1-59745-450-6_31)
- Alexpandi, R., De Mesquita, J. F., Pandian, S. K., & Ravi, A. V. (2020). Quinolines-Based SARS-CoV-2 3CLpro and RdRp Inhibitors and Spike-RBD-ACE2 Inhibitor for Drug-Repurposing Against COVID-19: An in silico Analysis. *Frontiers in Microbiology*, 11, 1796. <https://doi.org/10.3389/fmicb.2020.01796>
- Ali, M., Pandey, R. K., Khatoon, N., Narula, A., Mishra, A., & Prajapati, V. K. (2017). Exploring dengue genome to construct a multi-epitope based subunit vaccine by utilizing immunoinformatics approach to battle against dengue infection. *Scientific Reports*, 7(1), 1–13. <https://doi.org/10.1038/s41598-017-09199-w>
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., ... Yeh, L.-S. L. (n.d.). UniProt: the Universal Protein knowledgebase. <https://doi.org/10.1093/nar/gkh131>
- Barlow, D. J., Edwards, M. S., & Thornton, J. M. (1986). Continuous and discontinuous protein antigenic determinants. *Nature*, 322(6081), 747–748. <https://doi.org/10.1038/322747a0>
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., ... Bourne, P. E. (2000). *The Protein Data Bank*. *Nucleic Acids Research* (Vol. 28). Retrieved from <http://www.rcsb.org/pdb/status.html>
- Binkowski, T. A., Naghibzadeh, S., & Liang, J. (2003). CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Research*, 31(13), 3352–3355. <https://doi.org/10.1093/nar/gkg512>
- Blueprint. (n.d.). Retrieved August 4, 2020, from <https://www.who.int/teams/blueprint>
- Bourdette, D. N., Edmonds, E., Smith, C., Bowen, J. D., Guttmann, C. R. G., Nagy, Z. P., ... Vandembark, A. A. (2005). A highly immunogenic trivalent T cell receptor peptide vaccine for multiple sclerosis. *Multiple Sclerosis*, 11(5), 552–561. <https://doi.org/10.1191/1352458505ms1225oa>
- Bowers, K. J., Chow, D. E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., ... Shaw, D. E. (2006). Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. In *ACM/IEEE SC 2006 Conference (SC'06)*. <https://doi.org/10.1109/SC.2006.54>
- Braun, J., Loyal, L., Frensch, M., Wendisch, D., Georg, P., Kurth, F., ... Thiel, A. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature*, 1–8. <https://doi.org/10.1038/s41586-020-2598-9>
- Bui, H. H., Sidney, J., Li, W., Fusseder, N., & Sette, A. (2007a). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and

- vaccines. *BMC Bioinformatics*, 8. <https://doi.org/10.1186/1471-2105-8-361>
- Bui, H. H., Sidney, J., Li, W., Füsseder, N., & Sette, A. (2007b). Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. *BMC Bioinformatics*, 8. <https://doi.org/10.1186/1471-2105-8-361>
- Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*, 30(5), 1102–1107. <https://doi.org/10.1016/j.biotechadv.2011.09.013>
- Corbett, K. S., Edwards, D. K., Leist, S. R., Abiona, O. M., Boyoglu-Barnum, S., Gillespie, R. A., ... Graham, B. S. (2020). SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*, 1–8. <https://doi.org/10.1038/s41586-020-2622-0>
- Dar, H. A., Waheed, Y., Najmi, M. H., Ismail, S., Hetta, H. F., Ali, A., & Muhammad, K. (2020). Multiepitope Subunit Vaccine Design against COVID-19 Based on the Spike Protein of SARS-CoV-2: An in Silico Analysis. *Journal of Immunology Research*, 2020. <https://doi.org/10.1155/2020/8893483>
- de Wilde, A. H., Snijder, E. J., Kikkert, M., & van Hemert, M. J. (2018). Host factors in coronavirus replication. In *Current Topics in Microbiology and Immunology* (Vol. 419, pp. 1–42). Springer Verlag. [https://doi.org/10.1007/82\\_2017\\_25](https://doi.org/10.1007/82_2017_25)
- Dhanda, S. K., Vir, P., & Raghava, G. P. S. (2013). Designing of interferon-gamma inducing MHC class-II binders. *Biology Direct*, 8(1), 30. <https://doi.org/10.1186/1745-6150-8-30>
- Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: allergenicity prediction by descriptor fingerprints. *Bioinformatics*, 30(6), 846–851. <https://doi.org/10.1093/bioinformatics/btt619>
- Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, 8(1), 4. <https://doi.org/10.1186/1471-2105-8-4>
- Enayatkhani, M., Hasaniazad, M., Faezi, S., Guklani, H., Davoodian, P., Ahmadi, N., ... Ahmadi, K. (2020). Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *Journal of Biomolecular Structure and Dynamics*, 0(0), 1–16. <https://doi.org/10.1080/07391102.2020.1756411>
- Ferraz, W. R., Gomes, R. A., S Novaes, A. L., & Goulart Trossini, G. H. (2020). Ligand and structure-based virtual screening applied to the SARS-CoV-2 main protease: an *in silico* repurposing study. *Future Medicinal Chemistry*, 12(20), 1815–1828. <https://doi.org/10.4155/fmc-2020-0165>
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook* (pp. 571–607). Humana Press. <https://doi.org/10.1385/1-59259-890-0:571>
- Gentile, D., Patamia, V., Scala, A., Sciortino, M. T., Piperno, A., & Rescifina, A. (2020). Putative Inhibitors of SARS-CoV-2 Main Protease from A Library of Marine Natural Products: A Virtual Screening and Molecular Modeling Study. *Marine Drugs*, 18(4),

225. <https://doi.org/10.3390/md18040225>

- Gori, A., Longhi, R., Peri, C., & Colombo, G. (2013, August). Peptides for immunological purposes: Design, strategies and applications. *Amino Acids*. Amino Acids. <https://doi.org/10.1007/s00726-013-1526-9>
- Grote, A., Hiller, K., Scheer, M., Münch, R., Nörtemann, B., Hempel, D. C., & Jahn, D. (2005). JCat: A novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*, 33(SUPPL. 2), W526. <https://doi.org/10.1093/nar/gki376>
- Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R., & Warwicker, J. (2017). Protein-Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics*, 33(19), 3098–3100. <https://doi.org/10.1093/bioinformatics/btx345>
- Heo, L., Park, H., & Seok, C. (2013). GalaxyRefine: Protein structure refinement driven by side-chain repacking. *Nucleic Acids Research*, 41(Web Server issue), W384–W388. <https://doi.org/10.1093/nar/gkt458>
- Home - Johns Hopkins Coronavirus Resource Center. (n.d.). Retrieved April 26, 2021, from <https://coronavirus.jhu.edu/>
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *Journal of Biochemistry*, 88(6), 1895–1898.
- Jespersen, M. C., Peters, B., Nielsen, M., & Marcatili, P. (2017). BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Research*, 45(W1), W24–W29. <https://doi.org/10.1093/nar/gkx346>
- Joffrin, L., Goodman, S. M., Wilkinson, D. A., Ramasindrazana, B., Lagadec, E., Gomard, Y., ... Lebarbenchon, C. (2020). Bat coronavirus phylogeography in the Western Indian Ocean. *Scientific Reports*, 10(1), 1–11. <https://doi.org/10.1038/s41598-020-63799-7>
- Kadioglu1, O., Saeed1, M., Johannes Greten2, H., & Efferth1, T. (n.d.). Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervised machine learning. <https://doi.org/10.2471/BLT.20.251561>
- Kar, T., Narsaria, U., Basak, S., Deb, D., Castiglione, F., Mueller, D. M., & Srivastava, A. P. (2020). A candidate multi-epitope vaccine against SARS-CoV-2. *Scientific Reports*, 10(1), 1–24. <https://doi.org/10.1038/s41598-020-67749-1>
- Knutson, K. L., Schiffman, K., & Disis, M. L. (2001). Immunization with a HER-2/neu helper peptide vaccine generates HER-2/neu CD8 T-cell immunity in cancer patients. *Journal of Clinical Investigation*, 107(4), 477–484. <https://doi.org/10.1172/JCI11752>
- Kumar, A., Rai, S., Rathi, E., Agarwal, P., & Kini, S. G. (2020). Pharmacophore-guided fragment-based design of novel mammalian target of rapamycin inhibitors: extra precision docking, fingerprint-based 2D and atom-based 3D-QSAR modelling. *Journal of Biomolecular Structure and Dynamics*. <https://doi.org/10.1080/07391102.2020.1726816>
- Kumar, A., Rathi, E., & Kini, S. G. (2019). Identification of E6 Inhibitors Employing Energetically Optimized Structure-Based Pharmacophore Modelling, Ligand Docking and Molecular Dynamics Simulations Studies. *ChemistrySelect*, 4(36), 10701–10708.



<https://doi.org/10.1002/slct.201902105>

- Li, F., & Du, L. (2019, July 1). MERS Coronavirus: An emerging zoonotic virus. *Viruses*. MDPI AG. <https://doi.org/10.3390/v11070663>
- Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., & Friesner, R. A. (2011). The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins: Structure, Function and Bioinformatics*, 79(10), 2794–2812. <https://doi.org/10.1002/prot.23106>
- Lin, L., Ting, S., Yufei, H., Wendong, L., Yubo, F., & Jing, Z. (2020). Epitope-based peptide vaccines predicted against novel coronavirus disease caused by SARS-CoV-2. *Virus Research*, 288, 198082. <https://doi.org/10.1016/j.virusres.2020.198082>
- Liu, J., Wu, P., Gao, F., Qi, J., Kawana-Tachikawa, A., Xie, J., ... Gao, G. F. (2010). Novel Immunodominant Peptide Presentation Strategy: a Featured HLA-A\*2402-Restricted Cytotoxic T-Lymphocyte Epitope Stabilized by Intrachain Hydrogen Bonds from Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein. *Journal of Virology*, 84(22), 11849–11857. <https://doi.org/10.1128/jvi.01464-10>
- Lu, X., Wu, X., Liang, Z., Weng, X., Li, Q., & Gong, F. (2005). Construction and functional test of HLA-A\*2402-peptide tetramer. *Cellular & Molecular Immunology*, 2(2), 145–149.
- Madhavi Sastry, G., Adzhigirey, M., Day, T., Annabhimoju, R., & Sherman, W. (2013). Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*, 27(3), 221–234. <https://doi.org/10.1007/s10822-013-9644-8>
- Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S. I., Dan, J. M., ... Weiskopf, D. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science*, eabd3871. <https://doi.org/10.1126/science.abd3871>
- Maurer-Stroh, S., Krutz, N. L., Kern, P. S., Gunalan, V., Nguyen, M. N., Limviphuvadh, V., ... Frank Gerberick, G. (2019). AllerCatPro-prediction of protein allergenicity potential from the protein sequence. *Bioinformatics*, 35(17), 3020–3027. <https://doi.org/10.1093/bioinformatics/btz029>
- Mousavizadeh, L., & Ghasemi, S. (2020, March 31). Genotype and phenotype of COVID-19: Their roles in pathogenesis. *Journal of Microbiology, Immunology and Infection*. Elsevier Ltd. <https://doi.org/10.1016/j.jmii.2020.03.022>
- Mukherjee, S., Tworowski, D., Detroja, R., Mukherjee, S. B., & Frenkel-Morgenstern, M. (2020). Immunoinformatics and structural analysis for identification of immunodominant epitopes in SARS-CoV-2 as potential vaccine targets. *Vaccines*, 8(2), 1–17. <https://doi.org/10.3390/vaccines8020290>
- Naz, A., Shahid, F., Butt, T. T., Awan, F. M., & Ali, A. (2020). Designing Multi-Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 ( COVID-19 ) by Employing Immuno-Informatics Approach, 11(July), 1–13. <https://doi.org/10.3389/fimmu.2020.01663>
- Naz, A., Shahid, F., Butt, T. T., Awan, F. M., Ali, A., & Malik, A. (2020). Designing Multi-

- Epitope Vaccines to Combat Emerging Coronavirus Disease 2019 (COVID-19) by Employing Immuno-Informatics Approach. *Frontiers in Immunology*, *11*, 1663. <https://doi.org/10.3389/fimmu.2020.01663>
- Ngo, S. T., Quynh Anh Pham, N., Thi Le, L., Pham, D.-H., & Vu, V. V. (2020). Computational Determination of Potential Inhibitors of SARS-CoV-2 Main Protease. *Journal of Chemical Information and Modeling*, *60*(12), 5771–5780. <https://doi.org/10.1021/acs.jcim.0c00491>
- Nielsen, M., & Lund, O. (2009). NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC Bioinformatics*, *10*(1), 296. <https://doi.org/10.1186/1471-2105-10-296>
- Oany, A. R., Emran, A. Al, & Jyoti, T. P. (2014). Design of an epitope-based peptide vaccine against spike protein of human coronavirus: An in silico approach. *Drug Design, Development and Therapy*, *8*(August), 1139–1149. <https://doi.org/10.2147/DDDT.S67861>
- Olubiyi, O. O., Olagunju, M., Keutmann, M., Loschwitz, J., & Strodel, B. (2020). High Throughput Virtual Screening to Discover Inhibitors of the Main Protease of the Coronavirus SARS-CoV-2. *Molecules*, *25*(14), 3193. <https://doi.org/10.3390/molecules25143193>
- Panda, P. K., Arul, M. N., Patel, P., Verma, S. K., Luo, W., Rubahn, H.-G., ... Ahuja, R. (2020). Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Science Advances*, *6*(28), eabb8097. <https://doi.org/10.1126/sciadv.abb8097>
- Pandey, A., Nikam, A. N., Shreya, A. B., Mutalik, S. P., Gopalan, D., Kulkarni, S., ... Prassl, R. (2020, September 1). Potential therapeutic targets for combating SARS-CoV-2: Drug repurposing, clinical trials and recent advancements. *Life Sciences*. Elsevier Inc. <https://doi.org/10.1016/j.lfs.2020.117883>
- Ponomarenko, J., Bui, H. H., Li, W., Fusseder, N., Bourne, P. E., Sette, A., & Peters, B. (2008). ElliPro: A new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics*, *9*(1), 514. <https://doi.org/10.1186/1471-2105-9-514>
- Prajapat, M., Sarma, P., Shekhar, N., Avti, P., Sinha, S., Kaur, H., ... Medhi, B. (2020, January 1). Drug targets for corona virus: A systematic review. *Indian Journal of Pharmacology*. Wolters Kluwer Medknow Publications. [https://doi.org/10.4103/ijp.IJP\\_115\\_20](https://doi.org/10.4103/ijp.IJP_115_20)
- Rapin, N., Lund, O., Bernaschi, M., & Castiglione, F. (2010). Computational immunology meets bioinformatics: The use of prediction tools for molecular binding in the simulation of the immune system. *PLoS ONE*, *5*(4), 9862. <https://doi.org/10.1371/journal.pone.0009862>
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Research*, *48*(W1), W449–W454. <https://doi.org/10.1093/nar/gkaa379>
- Rosano, G. L., & Ceccarelli, E. A. (2014, April 17). Recombinant protein expression in *Escherichia coli*: Advances and challenges. *Frontiers in Microbiology*. Frontiers



- Research Foundation. <https://doi.org/10.3389/fmicb.2014.00172>
- Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nature Protocols*, 5(4), 725–738. <https://doi.org/10.1038/nprot.2010.5>
- Sahin, U., Oehm, P., Derhovanessian, E., Jabulowsky, R. A., Vormehr, M., Gold, M., ... Türeci, Ö. (2020). An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature*, 1–6. <https://doi.org/10.1038/s41586-020-2537-9>
- Sanami, S., Zandi, M., Pourhossein, B., Mobini, G. R., Safaei, M., Abed, A., ... Alizadeh, M. (2020). Design of a multi-epitope vaccine against SARS-CoV-2 using immunoinformatics approach. *International Journal of Biological Macromolecules*, 164, 871–883. <https://doi.org/10.1016/j.ijbiomac.2020.07.117>
- Sanchez-Trincado, J. L., Gomez-Perosanz, M., & Reche, P. A. (2017). Review Article Fundamentals and Methods for T-and B-Cell Epitope Prediction. <https://doi.org/10.1155/2017/2680160>
- Sarkar, B., Ullah, A., Tuz, F., Afrin, M., & Araf, Y. (2020). 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 , (January).
- Shey, R. A., Ghogomu, S. M., Esoh, K. K., Nebangwa, N. D., Shintouo, C. M., Nongley, N. F., ... Souopgui, J. (2019). In-silico design of a multi-epitope vaccine candidate against onchocerciasis and related filarial diseases. *Scientific Reports*, 9(1), 1–18. <https://doi.org/10.1038/s41598-019-40833-x>
- Singh, A., Thakur, M., Sharma, L. K., & Chandra, K. (2020). Designing a multi-epitope peptide based vaccine against SARS-CoV-2. *Scientific Reports*, 10(1), 16219. <https://doi.org/10.1038/s41598-020-73371-y>
- Singh, H., Jakhar, R., & Sehrawat, N. (2020). Designing spike protein (S-Protein) based multi-epitope peptide vaccine against SARS COVID-19 by immunoinformatics. *Heliyon*, 6(11), e05528. <https://doi.org/10.1016/j.heliyon.2020.e05528>
- Song, Z., Xu, Y., Bao, L., Zhang, L., Yu, P., Qu, Y., ... Qin, C. (2019, January 1). From SARS to MERS, thrusting coronaviruses into the spotlight. *Viruses*. MDPI AG. <https://doi.org/10.3390/v11010059>
- Steers, N. J., Currier, J. R., Jobe, O., Tovanabuttra, S., Ratto-Kim, S., Marovich, M. A., ... Rao, M. (2014). Designing the epitope flanking regions for optimal generation of CTL epitopes. *Vaccine*, 32(28), 3509–3516. <https://doi.org/10.1016/j.vaccine.2014.04.039>
- Tahir ul Qamar, M., Rehman, A., Tusleem, K., Ashfaq, U. A., Qasim, M., Zhu, X., ... Chen, L.-L. (2020). Designing of a next generation multiepitope based vaccine (MEV) against SARS-COV-2: Immunoinformatics and in silico approaches. *PLOS ONE*, 15(12), e0244176. <https://doi.org/10.1371/journal.pone.0244176>
- Tovchigrechko, A., & Vakser, I. A. (2006). GRAMM-X public web server for protein-protein docking. *Nucleic Acids Research*, 34(WEB. SERV. ISS.), W310–W314. <https://doi.org/10.1093/nar/gkl206>

- Tubert-Brohman, I., Sherman, W., Repasky, M., & Beuming, T. (2013). Improved docking of polypeptides with glide. *Journal of Chemical Information and Modeling*, 53(7), 1689–1699. <https://doi.org/10.1021/ci400128m>
- van Doremalen, N., Lambe, T., Spencer, A., Belij-Rammerstorfer, S., Purushotham, J. N., Port, J. R., ... Munster, V. J. (2020). ChAdOx1 nCoV-19 vaccine prevents SARS-CoV-2 pneumonia in rhesus macaques. *Nature*, 1–8. <https://doi.org/10.1038/s41586-020-2608-y>
- Van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastiris, P. L., Karaca, E., ... Bonvin, A. M. J. J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology*, 428(4), 720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>
- Wallace, A. C., Laskowski, R. A., & Thornton, J. M. (1995). Ligplot: A program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection*, 8(2), 127–134. <https://doi.org/10.1093/protein/8.2.127>
- Wang, R., Epstein, J., Charoenvit, Y., Baraceros, F. M., Rahardjo, N., Gay, T., ... Hoffman, S. L. (2004). Induction in humans of CD8+ and CD4+ T cell and antibody responses by sequential immunization with malaria DNA and recombinant protein. *Journal of Immunology (Baltimore, Md. : 1950)*, 172(9), 5561–5569. <https://doi.org/10.4049/jimmunol.172.9.5561>
- Wang, S., Peng, J., Ma, J., & Xu, J. (2016). Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. *Scientific Reports*, 6(1), 1–11. <https://doi.org/10.1038/srep18962>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., ... McLellan, J. S. (2019). *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. Retrieved from <http://science.sciencemag.org/>
- Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., ... McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, 367(6483), 1260–1263. <https://doi.org/10.1126/science.aax0902>
- Xu, D., & Zhang, Y. (2011). Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. *Biophysical Journal*, 101(10), 2525–2534. <https://doi.org/10.1016/j.bpj.2011.10.024>
- Xue, L. C., Rodrigues, J. P., Kastiris, P. L., Bonvin, A. M., & Vangone, A. (2016). PRODIGY: A web server for predicting the binding affinity of protein-protein complexes. *Bioinformatics*, 32(23), 3676–3678. <https://doi.org/10.1093/bioinformatics/btw514>
- Yang, J., Wang, W., Chen, Z., Lu, S., Yang, F., Bi, Z., ... Wei, X. (2020). A vaccine targeting the RBD of the S protein of SARS-CoV-2 induces protective immunity. *Nature*. <https://doi.org/10.1038/s41586-020-2599-8>
- Yang, Z., Bogdan, P., & Nazarian, S. (2020). An In-silico Deep Learning Approach to Multi-epitope Vaccine Design: A SARS-CoV-2 Case Study. <https://doi.org/10.21203/rs.3.rs-36528/v1>

- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., ... Hilgenfeld, R. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science*, 368(6489), 409–412. <https://doi.org/10.1126/science.abb3405>
- Zhang, Y., & Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function and Genetics*, 57(4), 702–710. <https://doi.org/10.1002/prot.20264>
- Zhou, P., Yang, X. Lou, Wang, X. G., Hu, B., Zhang, L., Zhang, W., ... Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>