Anti-COVID-19 multi-epitope vaccine designs employing global viral genome sequences

Tahreem Zaheer[1+], Maaz Waseem[1+], Walifa Waqar[1], Kanwal Naz[1], Muhammad Shehroz[2], Zaara Ishaq[1], Hamza Arshad Dar[1], Tahir Ahmad[1], Syeda Marriam Bakhtiar[3], Aun Syed Muhammad[4], Amjad Ali*[1]

[1] Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan
[2] Department of biotechnology, Virtual University of Pakistan, Pakistan
[3] Department of Bioinformatics and Biosciences, Capital University of Science and Technology Islamabad, Pakistan
[4] Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Punjab, Pakistan

[+]These authors contributed equally for the work and should be considered as joint first authors

*Corresponding Author:
Amjad Ali
Atta-ur-Rahman School of Applied Biosciences (ASAB)
National University of Sciences and Technology (NUST)
Sector H-12, Islamabad, 44000, Pakistan
Email address: amjad.ali@asab.nust.edu.pk

Abstract

Background. The coronavirus SARS-CoV2 is an evolved strain from the Coronaviridae family that has caused a global public health emergency. Currently, there is no approved treatment or vaccine available to treat this deadly virus. The current study aimed to find the diversity in SARS-CoV2 strains reported from all over the world and to design a broad-spectrum multi-epitope vaccine using an immunoinformatics approach.

Methods. For this purpose, all available complete genomes were retrieved from GISAID and CNBC followed by genome multiple alignments to develop a global consensus sequence to compare with the reference genome. Fortunately, comparative genomics and phylogeny revealed a significantly high level of conservation between the viral strains. All Open Reading Frames (ORFs) were subjected to Cytotoxic T-lymphocyte (CTL) epitope and Helper T cell lymphocyte (HTL) epitopes using CTLpred and HLApred, respectively. The predicted CTL epitopes were then screened for antigenicity, immunogenicity and strong binding affinity with HLA superfamily alleles. HTL predicted epitopes were screened for antigenicity, interferon induction

Deleted: C
Deleted: belongs to the family Coronaviridae.
Deleted: of
Deleted: according to WHO and CDC
Deleted: FDA
Deleted: the
Deleted: and then
Deleted: d
Deleted: C
Deleted: A phylogenetic tree was constructed using the UPGMA method.
Deleted: the available
Deleted: of proteome
Deleted: server
Deleted: While
Deleted: I

potential, overlapping B cell epitopes and strong HLA DR binding potential. The shortlisted epitopes were arranged into two multi-epitope sequences, Cov-I-Vac and Cov-II-Vac, and molecular docking was performed with Toll-Like Receptor 8 (TLR8).

Results. The designed multi-epitopes were found to be antigenic and non-allergenic. Both multi epitopes were stable and predicted to be soluble in an *E. coli* expression system. The molecular docking with TLR8 also demonstrated that they have a strong binding affinity and immunogenic potential. These *in silico* analyses suggest that the proposed multi-epitope vaccine can effectively evoke an immune response.

Keywords: SARS-CoV2, COVID-19, global pandemic, Pakistan, immunoinformatics, immunity

Introduction

COVID-19 has been declared a pandemic by the World Health Organization (WHO) as of the 11th March 2020 (WHO, 2020). The outbreak nCoV-2019 starting in Wuhan, China has now affected millions of lives across the world. The common symptoms patients exhibit include fatigue, fever, dry cough, shortness of breath, and upper airway congestion (Chan, Wong and Tang, 2020). The delay in symptom appearance due to the incubation period of the virus has caused more harm as infected people without symptoms might have already infected many others even before knowing they are carrying the infection. The number of infected people is increasing daily. Scientists are working tirelessly to find a cure but as of now, the impasse continues.

Coronaviruses (CoVs) are a family of viruses causing diseases ranging from the common cold to more severe illnesses such as Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) (Ceraolo and Giorgi, 2020b; Sahin, 2020). Belonging to the family of Betacoronavirus, the latest CoV, recently named as SARS-CoV-2, is an enveloped spherical virion about 60-140nm in diameter. It has a positive-sense single-stranded RNA genome of ~30kb. The virus is believed to have originated from bats and is phylogenetically linked closest to the bat-SL-CoVZC45 and bat-SL-CoVZXC21 (Zhu *et al*., 2020). The virus is contagious and spreads from person-to-person through the respiratory droplets when an infected person sneezes or coughs or when a person encounters contaminated surfaces with the virus. The diseases state caused by viral infection is termed as COVID-19 (Chan, Wong and Tang, 2020). The genome contains six open reading frames (ORFs) including ORF1ab, ORF3a, ORF6, ORF7a, ORF8, ORF10, encoding spike glycoprotein trimer S, nucleoprotein N, membrane protein M and envelop small membrane protein pentamer E (Zhu *et al*., 2020). SARS-CoV-2 causes wide-ranging infections like mild upper respiratory tract disease, severe viral pneumonia with respiratory failure and in severe cases even death (Huang *et al*., 2020).

Considering the gravity of the situation of COVID-19 worldwide, there is a dire need for the development of an effective vaccine or antiviral drugs that are effective to treat this viral

Deleted: ers using online servers

Deleted: safe,

Deleted: ,

Deleted: of the

Deleted: exhibited

Deleted: the

Deleted: The experimental verification of these results is required, nonetheless and is a part of the prospects of our study. …

Deleted: (

Deleted:)

Deleted: It

Deleted: ,

Deleted: RNA

Deleted: serves as mRNA and

Deleted: has

Deleted: and

Deleted:

infection. Currently there is no FDA approved treatment or licensed vaccine against COVID-19. Vaccination is the most desired option available at our disposal to eradicate the diseases associated with infectious microorganisms (Greenwood, 2014). Reverse vaccinology is a valuable technique that adopts genome-level analysis to identify potential antigenic determinants of the pathogenic microorganism (Rappuoli, 2001). Previously designed multi-epitope vaccines against various emerging viruses have been reported as safe and elicit potent immune responses.

To elicit strong immunogenic responses and have long-term efficacy, vaccines against coronavirus must activate both humoral and cell-mediated immune responses (Amanna and Slifka, 2011). Virus-specific T cells, the effector T cells, provide vaccine-mediated protection at the peak of antiviral response and the antiviral cytokines production is also increased (Sallusto, Geginat and Lanzavecchia, 2004), but a successful vaccine would activate B-cell mediated humoral immunity as well. In the case of re-infection, memory B cells can be re-activated via B cell receptors and provide protection even when antibody titers have gone below detection levels (Amanna, Carlson and Slifka, 2007). Therefore, it is a challenge to activate both humoral as well as cell mediated immunity of required levels to effectively provide effective protection against COVID-19. Around 1800+ cases of COVID-19 have been reported in Pakistan so far, and an increasing trend is feared in the upcoming days (Worldometers.info, 2020).

In this study, we have utilized the 475 genomes of 2019-nCoV available in CNBC and GISAID (available at https://www.epicov.org/epi3/cfrontend) database as of 15th March 2020 to identify conserved immunogenic epitopes representing the novel coronavirus strains (collection) and to design a potent multi-epitope vaccine against the COVID-19 infection.

Materials & Methods

Retrieval of Sequences and Multiple alignments

GISAID (available at https://www.epicov.org/epi3/cfrontend) and CNBC databases were mined to retrieve updated complete genomes of nCOV-19. A complete genome of SARS-CoV2 from a Pakistani patient was also added in the study (Accession number GWHACDD01000001). The retrieved genomes were then aligned using the MAFFT tool (**M**ultiple **A**lignment using **F**ast **F**ourier **T**ransform) (available at: https://www.ebi.ac.uk/Tools/msa/mafft/) (Katoh *et al*., 2002). Default parameters were used encompassing 200PAM/κ=2 scoring matrix and a gap penalty of 1.53. These aligned sequences were visualized and analyzed in Unipro UGENE v.34 software (Okonechnikov *et al*., 2012) to generate a consensus sequence of all genomes and later nBLAST was performed to analyze its diversity with Reference sequence NC_045512.2. The COVID-19 genome from Pakistan Accession number GWHACDD01000001 was also aligned with the consensus reference sequence. The ORFs from this sequence were retrieved and subjected to further analysis on predicted proteins.

CTL Epitope mapping

Deleted: . B

Deleted: C

Deleted: R

Deleted: proteome of all

Deleted: available against

Cell-mediated Immunity provides a more promising result in case of viral infection where Cytotoxic T-Lymphocytes (CTLs) kill viral infected cells and clears the infection. Hence, CTLpred (available at: http://crdd.osdd.net/raghava/ctlpred/) was used for prediction of CD8 T cell epitopes (Bhasin and Raghava, 2004). Consensus approach was used that predicted CTL epitopes based on both SVM and ANN methods. These predicted epitopes were also predicted by HLApred (available at: http://crdd.osdd.net/raghava/hlapred/) and ProPred (available at: http://crdd.osdd.net/raghava/propred/) (Singh and Raghava, 2002).

Antigenicity

All CTL epitopes were screened for antigenicity using VaxiJen (available at: http://www.ddg-pharmfac.net/vaxijen/) (Doytchinova and Flower, 2007). VaxiJen is an online server that predicts the antigenicity of peptides using the physicochemical properties of the proteins irrespective of their lengths and the need of alignments. We used VaxiJen server for our CTL epitopes using the default threshold 0.5 for viruses.

Immunogenicity

After screening for the antigenic potential of CTL epitopes, epitopes having immunogenic potential were screened using the IEDB analysis resource for MHC class I Immunogenicity (available at: http://tools.iedb.org/immunogenicity/) (Calis *et al*., 2013) This tool works best for 9-mer epitopes but larger lengths can be employed too. Only those epitopes that showed strong binding affinity with HLA super family alleles were shortlisted.

Strong binding affinity with HLA alleles

CTL Epitopes that were both antigenic and immunogenic were screened for their binding affinity with HLA superfamily alleles using Net MHC 4.0 (available at: http://www.cbs.dtu.dk/services/NetMHC/) (Nielsen *et al*., 2003; Andreatta and Nielsen, 2016). The tool predicts results on the basis of an artificial neural network that has been trained using 81 different HLA alleles. MHC binding affinity of shortlisted epitopes was checked against HLA superfamily alleles (i.e. HLA-A0101, HLA-A0201, HLA-A0301, HLA-A2402, HLA-A2601, HLA-B0702, HLA-B0801, HLA-B2705, HLA-B3901, HLA-B4001, HLA-B5801, HLA-B1501).

HTL Epitope mapping

Humoral Immunity also plays a significant role in viral clearance, so Helper T cell epitopes also known as CD4 T cell epitopes were predicted by HLApred. All alleles of MHC-II were selected, and experimental as well as predicted binders were included in the analysis. A default threshold was used, and all predicted epitopes that shared no identity with humans were selected. The predicted epitopes were screened for their antigenic potential using VaxiJen as discussed earlier. The antigenic epitopes were further screened for their IFN induction potential using IFNepitope (available at: http://crdd.osdd.net/raghava/ifnepitope/scan.php) (Dhanda, Vir and Raghava, 2013). The epitopes screened positive for IFN were checked for overlapped B cell epitopes

Deleted: the
Deleted: database
Deleted: server
Deleted: both
Deleted: D
Deleted: the
Deleted: server

predicted by ABCpred (available at: http://crdd.osdd.net/raghava/abcpred/) using a default threshold (Saha and G. P.S. Raghava, 2006). Lastly, the shortlisted epitopes were checked for their strong binding affinity with HLA DR alleles (DRB1_0101, DRB1_0103, DRB1_0301, DRB1_0401, DRB1_0402, DRB1_0403, DRB1_0404, DRB1_0405, DRB1_0701, DRB1_0801, DRB1_0802, DRB1_0901, DRB1_1001, DRB1_1101, DRB1_1201, DRB1_1301, DRB1_1302, DRB1_1501, DRB1_1602, DRB3_0101, DRB3_0202, DRB3_0301, DRB4_0101, DRB4_0103, DRB5_0101) using Net MHCII (available at: http://www.cbs.dtu.dk/services/NetMHCII/).

Multi-epitope vaccine design and construction

Epitopes were arranged as per their arrangement in the reference sequence and linked between the CTL epitopes using a flexible linker GGGGS as recommended by Chen, Zaro and Shen (2013), or between HTL epitopes with GPGPG to construct a multi-epitope (Saadi, Karkhah and Nouri, 2017). Linkers between epitopes provide efficient separation and proper functioning of individual epitopes. Moreover, a second multi-epitope construct was also designed where β-defensin, an adjuvant - was added at the N terminus of multi-epitope. β-defensin generates an innate as well as an adaptive immune response (Pandey, Bhatt and Prajapati, 2018). β-defensin is responsible for recruiting immature dendritic cells and naïve T-cell at the site of an infection (Allaker, 2008).

Sequence-based physio-chemical properties of the multi-epitope design

Different physicochemical properties of multi epitopes were predicted using ProtParam tool including prediction of *in vivo* half-life, theoretical isoelectric point (pI), molecular weight, instability index, aliphatic index and GRAVY (Grand Average of Hydropathy) index of peptides (Wilkins *et al*., 1999). The degradation of peptides was based on amino acid at N-terminus hence N-end rule in Protparam was employed to predict *in vivo* half-life of both peptides. Instability index with a threshold of 40 was used in order to determine the stability of the protein in a cellular environment. The aliphatic and GRAVY index of both peptides was calculated based on their amino acid profiles.

Allergenicity of multi epitope

Both multi-epitope constructs (with and without adjuvant) were screened for their allergenicity using two online servers; AlgPred (available at: http://crdd.osdd.net/raghava/algpred/) (Saha and G P S Raghava, 2006) and Allergen FP (available at: http://ddg-pharmfac.net/AllergenFP/) (Dimitrov *et al*., 2014). In AlgPred, allergenicity prediction is based on similarity with experimental IgE epitopes while Allergen FP predicts the allergenicity of peptide using an alignment-free method by implementing a four-step algorithm and compared by Tanimoto coefficient.

Antigenicity of multi-epitope design and modeling

Deleted: R

Deleted: between CTL epitopes

Deleted: (

Deleted: ,

Deleted: and

Deleted: between HTL epitopes

Deleted: It

Deleted: in form of

The antigenicity profiles of both multi-epitope constructs were predicted again to confirm if the epitopes were antigenic or not. We used two servers this time; VaxiJen (available at: http://www.ddg-pharmfac.net/vaxijen/) (Doytchinova and Flower, 2007) and AntigenPro (available at: http://scratch.proteomics.ics.uci.edu/) (Magnan *et al*., 2010). AntigenPro like VaxiJen predicts antigenicity using physicochemical properties, irrespective of the length of peptides as well as it's alignment.

Secondary structure prediction

For evaluation of further structural characteristics of designed multi epitopes, PDBsum (available at http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/) (Laskowski *et al*., 2018) was used. PDBsum provides insight about the unique structural aspects about protein, peptides and their ligands.

Deleted: helps in getting

Tertiary structure prediction and refinement

Tertiary structures of both vaccines were predicted by 3Dpro (available at: http://scratch.proteomics.ics.uci.edu/) (Cheng *et al*., 2005) and refined by Galaxy refine (available at: http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) (Ko *et al*., 2012). The model showing highest Ramachandran favored residues and minimum poor rotamers were selected. The refined structures were checked for their stability and flexibility using molecular dynamics (MD) simulation studies. Structural flexibility of a protein/peptide is important for its molecular recognition and its function, so a coarse-grained protein model implemented in a webserver CABS-Flex 2.0 (available at: http://biocomp.chem.uw.edu.pl/CABSflex/) (Kuriata *et al*., 2018) was used for near-native dynamics of both vaccines. Default distance restrained parameters were used, number of cycles and cycles between trajectory frames were raised to 100. Temperature of simulation was also kept default.

Molecular Docking with Immunological Receptors

The models were then checked for their interaction with TLR8. Molecular docking was performed by using guru level interface of HADDOCK 2.2 (available at: http://haddock.science.uu.nl/services/HADDOCK/haddockserver-guru.html) (Van Zundert *et al*., 2016) with default parameters and a representative structure from the top-ranked docked cluster with minimum HADDOCK score was refined using Refinement Interface (available at: http://haddock.science.uu.nl/services/HADDOCK/haddockserver-refinement.html) (Van Zundert *et al*., 2016). Moreover, NMA analysis of refined complexes was also performed using iMOD in order to determine their deformation potential (Kremer, Mastronarde and McIntosh, 1996) (available at: https://bio3d.colorado.edu/imod/paper/). The interacting residues between each complex were predicted using PDBsum analysis (Laskowski *et al*., 2018).

Codon Optimization and *in silico* cloning in *Escherichia coli*

Codon optimization of both multi epitopes was performed using JCAT (available at: http://www.jcat.de/) (Grote *et al*., 2005) and codons were adapted as per *E. coli* K12 strain. Plasmid UC19 was used for *in silico* cloning of both multi epitopes Moreover, a His6 tag was added at the both ends of sequences for the purification of multiepitope proteins.

Results

Retrieval of Sequences and Multiple Sequence Alignments (MSA)

The recruited 474 complete COVID-19 genomes showed a high level of conservancy upon MSA alignment and in the phylogenetic tree (Fig 1). The Pakistani strain shares a clad with China and India. Moreover, the consensus sequence was found to be 99% identical, with 10 gaps and no mismatches with Ref seq NC_045512.2 from Wuhan. Hence, the protein coding sequences from ORFs available under NC_045512.2 were ORF1ab YP_009724389.1, Spike glycoprotein YP_009724390.1 ORF3a YP_009724391.1, Envelope protein YP_009724392.1, membrane glycoprotein YP_009724393.1, ORF6 YP_009724393.1, ORF7a YP_009724393.1, ORF7b YP_009725318.1, ORF8 YP_009724396.1, ORF9 YP_009724397.2, ORF10 YP_009725255.1. All the ORFs from Ref Seq were compared with CDS of MT240479 and found to be 99% similar.

CTL Epitope mapping

The maximum number of epitopes from . . . as predicted from ORF1ab were 311. However, from surface glycoproteins and membrane glycoprotein, ORF3a, ORF9, ORF7a, ORF6, ORF8, ORF7b, ORF10 and envelope proteins the epitopes obtained are 50, 18, 13, 12, 7, 6, 5, 4, 4, and 2 respectively. For antigenicity a total of 204 out of 432 epitopes were found to be antigenic as determined by . . . .. Using . . . , 193 epitopes were screened, but only 124 epitopes were found to be potentially immunogenic. The shortlisted epitopes are listed in Table 1.

HTL Epitope mapping

A total of 289 CD4 T cell epitopes from . . . were predicted by HTL to lack homology with the human proteome. Of these epitopes, 162 were antigenic while 62 had IFN induction potential. However, only 11 had strong binding affinity with HLA DR alleles. The list of screened epitopes are provided in Table 2.

Multi epitope structure design and modeling

Two multi-epitope constructs are designed with and without β-defensin adjuvant to analyze and compare whether the designed multi-epitope is efficacious only in presence of adjuvant or has the ability to elicit immune response solely. Cov-I-Vac was the multi-epitope construct without adjuvant and Cov-II-Vac was with adjuvant (Fig). Sequences of Cov-I-Vac and Cov-II-Vac had been shown below; the sequence of β-defensin is shown in bold letters in Cov-II-Vac.

Deleted: server

Deleted: . The conservancy can also be seen

Deleted: having

Deleted: proteome of all

Deleted: (

Deleted: ,

Commented [BO1]: This is not a sentence, and need to rephrase so that this is easier to undertand

Deleted: ;

Deleted: epitopes are predicted from ORF1ab

Deleted: ,

Commented [BO2]: Not sure what you are trying to say here, needs to be rephrased

Deleted: ¶
Antigenicity¶

Deleted: A

Deleted: epitopes

Deleted: ¶
Immunogenicity↵
Out of

Deleted: having no

Deleted: among them

Deleted: found to be

Deleted: among them

Deleted: of them

Deleted: can be seen

Deleted: it

Deleted: as well

Deleted: A

Commented [BO3]: Move to figures and cite

Commented [BO4]: Move to figures

>Cov-I-Vac

TLSEQLDFIGGGGSSLREVRTIKGGGGSTRFFYVLGLGGGGSFSYFAVHFIGGGGSWLM
WLIINLGGGGSAIDAYPLTKGGGGSYVFCTVNALGGGGSVYAWNRKRIGGGGSVVFLH
VTYVGGGGSSELVIGAVIGGGGSHLVDFQVTIGGGGSFLIVAAIVFGGGGSIIFWFSLELG
GGGSFLLVTLAILGPGPGLGIITTVAAGPGPGWYIRVGARKGPGPGIGYYRRATRGPGPG
VILLNKHIDGPGPGVRATATIPIGPGPGIITLKKRWQGPGPGITLKKRWQLGPGPGLLLVA
AGLEGPGPGFLCWHTNCY

>Cov-II-Vac

GIINTLQKYYCRVRGGRCAVLSCLPKEEQIGKCSTRGRKCCRRKKEAAAKTLSEQL
DFIGGGGSSLREVRTIKGGGGSTRFFYVLGLGGGGSFSYFAVHFIGGGGSWLMWLIINLG
GGGSAIDAYPLTKGGGGSYVFCTVNALGGGGSVYAWNRKRIGGGGSVVFLHVTYVGG
GGSSELVIGAVIGGGGSHLVDFQVTIGGGGSFLIVAAIVFGGGGSIIFWFSLELGGGGSFL
LVTLAILGPGPGLGIITTVAAGPGPGWYIRVGARKGPGPGIGYYRRATRGPGPGVILLNK
HIDGPGPGVRATATIPIGPGPGIITLKKRWQGPGPGITLKKRWQLGPGPGLLLVAAGLEGP
GPGFLCWHTNCY

Sequence-based physio-chemical properties of multi-epitope

Physio-chemical analysis of Cov-I-Vac and Cov-II-Vac revealed that both constructs are reasonably stable (38.48 and 39.91 respectively), have molecular weight <50kDa, lower GRAVY indexes (i.e., 0.394 and 0.249, respectively), with a pI of 9.97 and 10.03, respectively. The half-life of Cov-I-Vac was predicted to 7.5 h (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*), and >10 hours (*Escherichia coli*, *in vivo*). However, the half-life of Cov-II-Vac was predicted to be 30 hours (mammalian reticulocytes, *in vitro*), >20 hours (yeast, *in vivo*), and >10 hours (*Escherichia coli*, *in vivo*).

Antigenicity, Allergenicity and Immunogenicity of Cov-I-Vac and Cov-II-Vac

The antigenicity potential was found to be retained in both multi epitopes and was estimated as 0.6153 and 0.5947 for Cov-I-Vac and Cov-II-Vac, respectively. Both constructs were reported as antigenic by AntigenPro as well with an antigenicity score of 0.516406 and 0.386605. Moreover, both vaccines were soluble upon overexpression in *E. coli*. The safety profile also showed that constructs do not exhibit any experimentally reported allergens and thus projected to be non-allergens.

Secondary structure of Cov-I-Vac and Cov-II-Vac

The secondary structure prediction of Cov-I-Vac shows that it has a sheet and beta hairpin between two strands (Ala160-Il161, Gly164-Gly165), six helices, 42 beta turns and two gamma turns (Fig 2). The secondary structure of Cov-II-Vac shows that it has a sheet and two beta hairpin among three strands (Val193-Thr198, Gly201-Val208, Gly216-Gly217), 10 helices, five helix-helix interactions, and 37 beta turns (Fig 3).

Commented [BO5]: Move to figures

Deleted: the

Deleted: es

Deleted: s

Deleted: (isoelectric point)

Deleted: ours

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Formatted: Font: Italic

Commented [BO6]: If these are predictions, they should all be in situ, right? Because you didn't evaluate these in vivo or in vitro?

Formatted: Font: Italic

Deleted: found to be

Deleted: S

Deleted: 1

Deleted: , 1

Deleted: 6

Deleted: 2

Deleted: as can be seen in

Deleted: ¶
¶
¶
¶

Deleted: S

Deleted: 1

Deleted: , 2

Deleted: ,

Deleted: 3

Deleted: 5

Deleted: . The illustration of Secondary structure of Cov_II-Vac can be seen in

Deleted:

Tertiary structure prediction and refinement

The tertiary structure of Cov-I-Vac had 10.2% poor rotamers with 90.5% Rama favored residues; however, the refined structure selected for further analysis had an RMSD of 0.541 with 0% poor rotamers and 93% Rama favored residues. Similarly, the initial structure of Cov-II-Vac had 5.9% poor rotamers with 89% Rama favored residues, but the refined shortlisted structure with RMSD 0.548 had 92.9% Rama favored residues and 0% poor rotamers.

CABS-flex analysis

The trajectory of 10 models was generated after simulation (Fig 4). The Root Mean Square Fluctuation (RMSF) of this simulation was found to be between 3.9 - 5.8 Å.

The trajectory of Cov-II-Vac after fast simulation

Trajectory analysis of fast simulation of Cov-II-Vac was analyzed and the RMSF of the structure was found to be 0.75-7.25 Å. The maximum fluctuation was observed at residue 227 while minimum fluctuation was residue 39-59 (Fig 5).

The molecular docking of Cov-I-Vac with TLR8 resulted in the top-ranked cluster with a HADDOCK score of -40.8 +/- 5.5. Low values of HADDOCK score, especially those that are negative, usually imply significantly high interaction between proteins. HADDOCK refined the representative structure of this complex and clustered the resulting 20 structures into one complex that represented 100% of the water refined models generated. After undergoing molecular refinements, the HADDOCK score of the Cov-I-Vac-TLR8 complex improved significantly to reach -251.3 +/- 2.3. The details of the refined interaction along with statistical parameters are provided in Table 3. Cov-II-Vac also showed significant interaction with TLR8 that was -83.5 +/- 6.3 which was a higher interaction score than Cov-I-Vac. However, after refinements, the interaction of Cov-II-Vac with TLR8 was lower compared to Cov-I-Vac-TLR8 interaction, i.e. -217.4 +/- 3.5. The details of HADDOCK score with all its parameter are mentioned in Table 4.

For an in depth analysis of protein-protein interactions, PDBSum analysis revealed that a total of 36 residues of Cov-I-Vac interact with 43 residues of TLR8, with one salt bridge, 15 hydrogen bonds and 210 non bonded interactions. The detail of the interaction is illustrated in Fig 6 and atomic level interaction is provided in Supplementary Table 1. Similarly, 36 residues of Cov-II-Vac interacted with 41 residues of TLR8, prominent among them were residues of multi-epitopes of 12 hydrogen bonds and 168 non bonded contacts. The interacting residues of the complex are illustrated in Fig 7 while distances of detailed atomic interaction are provided in Supplementary table 2.

NMA analysis

Deleted: ¶

Deleted: can be seen in Fig 4

Deleted: RMSF (

Deleted: is

Deleted: ¶ ¶

Deleted: A

Deleted: although high

Deleted: ¶

Deleted: making 1

Deleted: H

Deleted: mentioned

Deleted: ¶ ¶

Deleted: a

Deleted: making

Deleted: H

Deleted: mentioned

NMA analysis of both multi epitopes showed that Cov-I-Vac has higher stability than Cov-II-Vac, as the Eigen value, the energy required to deform the structure, was higher for Cov-I-Vas as can be seen in Fig 8.

Codon Optimization and *E. coli* expression

Codon optimization results of both vaccines Cov-I-Vac and Cov-II-Vac with GC content 58.99% and 58.49% were within the optimum limits (Pandey, Bhatt and Prajapati, 2018). The CAI value was predicted to be 1.0. The total length of the Cov-I-Vac clone was 3.6 kbp while Cov- II-Vac clone was 3.8 kbp. Both sequences were designed to be added to the plasmid between restriction sites AfiIII-pciI and BspQI-sapI (Fig 9.

Discussion

The COVID-19 pandemic has affected more than 205 countries around the world with almost 783,000 patients and more than 37 thousand deaths. The current situation urges scientists all over the world to find an urgent solution to stop this pandemic and develop effective therapeutics (WHO, 2020a). To ensure viral clearance, cell mediated and humoral responses must be induced by the action of CD8 and CD4 T cells (Ikram *et al*., 2018). The reliability of such responses has enabled the use immunoinformatics approaches for vaccine development against viral diseases. The present study encompassing 475 complete genomes from all around the world suggests that while there is emerging variation in SARS-CoV2, the rate with which this variation arises is quite slow (Ceraolo and Giorgi, 2020a). The consensus sequence generated from the alignment of all complete genomes was found to be 99% identical to the reference sequence reported from Wuhan. Moreover, it also exhibited significant similarity with the strain sequenced from a Pakistani patient.

The overall structure of of 475 genomes of COVID-19 revealed that SARS-CoV2 strains of China belong to clades containing those from almost all of the countries of the world and covering the whole span of a phylogenetic tree. The relatedness of COVID-19 strains was expected considering that the outbreak originated from China. Pakistani strain GWHACDD01000001 also was found in clades with strains isolated from China and India. Pakistan claims that the virus was introduced in the country by Iran, which may be supported after the availability of a complete genome sequence/s of SARS-CoV2 from Iran anticipated at a later stage (Saqlain *et al*., 2020).

The present study was mainly conducted to propose multi-epitope based vaccines against COVID-19 by an immunoinformatics approach similar to other studies (Tomar and De, 2010; Dar *et al*., 2016; Ikram *et al*., 2018). These vaccines are advantageous compared to other monovalent vaccines because they provide superior protection by eliciting strong humoral and cell mediated immunity (Amanna and Slifka, 2011). Three subunit vaccines have been

Deleted: it

Deleted: as can be seen in

Deleted: C

Deleted: now urged the researchers to

Deleted: P

Deleted: the

Deleted: phylogenetic tree (Fig 1)

Deleted: make

Deleted: with

Deleted: the

Deleted: after persistent intervals

Deleted: is

Deleted: makes

Deleted: direct

Deleted: this could be checked

Deleted: Cov-I-Vac and Cov-II-Vac

Deleted: using

Deleted: previous

Deleted: it gives the best

previously reported that were based on one, three, and 10 proteins (Abdelmageed *et al*., 2020; Qamar *et al*., 2020). The current vaccine used all predicted ORFs and therefore covers the complete proteome. Epitopes selected were 12 CTL based and 10 HTL based that elicit cell mediated and humoral immune responses, respectively. Our subunit vaccines are therefore projected to be effective and immunogenic against COVID-19. Coding sequences of the proteins of SARS-CoV-2 from NCBI were evaluated for their antigenicity. Highly antigenic sequences were used to predict B and T cell epitopes, the key step in vaccine development. Antigenic T cell epitopes were screened to overlap with B-cell and IFN-γ epitopes (Van Regenmortel, 1996). The vaccines were designed with the help of linkers; epitopes were linked with GGGGS (CTL based) and GPGPG (HLA based) linkers. Linkers were added to help maintain the proper functioning of each epitope after being imported to the human body independently (Pandey, Bhatt and Prajapati, 2018).

An adjuvant, β-Defensin was added to the N-terminal end of the Cov-II-Vac to increase the immunogenicity of the vaccine. β-defensins are effector antimicrobial peptides having antimicrobial activity including antiviral, antibacterial and antifungal activity (Lehrer and Lu, 2012). Kim et al. (2018) have demonstrated that use of human β-defensins can initiate both innate and adaptive immune response against the conjugated antigen. Defensins use both Th1 and Th2 dependent pathways to boost the synthesis of antigen-specific immunoglobulins (Tani *et al*., 2000). Several studies suggest that β-defensins efficiently induce a prolonged humoral as well as cellular immune response against a pathogen (Allaker, 2008) which makes them a potent antiviral vaccine adjuvant.

The two multi epitopes Cov-I-Vac and Cov-II-Vac are 317 and 367 amino acids long respectively with a molecular mass of 31.6 kDa and 37.2 kDa, respectively. Both predicted proteins are basic according to theoretical pI values. The calculated instability index and aliphatic index scores indicated that the vaccine protein is stable and thermostable. The GRAVY score for our vaccines was positive and suggests that both are hydrophobic, which may necessitate the use of micelles for better interaction of the vaccine protein within the polar environment of the body (Pandey *et al*., 2018). Cov-I-Vac and Cov-II-Vac were found to be antigenic, immunogenic, non-allergenic and strong MHC binders. These factors suggest that the multi-epitopic vaccines will hae a robust immune response without allergic reaction. The RMSF obtained from CABS flex analysis is comparable to the RMSF of NMR ensembles (Jamroz, Kolinski and Kmiecik, 2014). The efficacy of a multi-epitope vaccine depends on the population for which the vaccine is prepared. The human CTL and HTL based T cell epitopes included in our multi-epitope vaccines were 99.29% similar to those of the world population and 100% similar to Pakistan based sequences, indicating that it would be effective not only for Pakistan but also for the world population. The strong binding of our vaccine products with immune receptors is necessary for triggering effective immunological responses in our body. The designed multi epitope-based vaccine showed significant interaction with an immune receptor

Deleted: ten
Deleted: respectively
Deleted: available
Deleted: (Table. 1)
Deleted: is
Deleted: (https://www.ncbi.nlm.nih.gov/genbank/SARS-CoV-2 -seqs/) were retrieved
Deleted: and checked
Deleted: then
Deleted: cell
Deleted: . Predicting and screening of B cell and T cell epitopes… is
is
Deleted: is
Deleted: then
Deleted: V
Deleted: as
Deleted: . GGGGS linkers were used to link epitopes that were CTL based while GPGPG linkers were used for HLA [1]
Commented [BO7]: Explain why you added to Cov-II- [2]
Deleted: for
Deleted: ing
Deleted: (Kim *et al*., 2018)
Deleted: designed
Deleted: having
Deleted: molecular weight
Deleted: Moreover, t
Commented [BO8]: In the human bloodstream? [3]
Deleted: showed
Deleted: that
Deleted: our protein's
Deleted: nature
Deleted: . As these proteins show hydrophobic nature it
Deleted: s
Deleted: indicate
Deleted: '
Deleted: any
Deleted: CABS dynamics used for this study gives similar [4]
Deleted: Moreover t
Commented [BO10]: What does this mean?
Deleted: showed
Deleted: with

further predicting the efficacy of both vaccines. However, the vaccine Cov-I-Vac showed a stronger interaction with TLR8, indicating that multi epitope, even without adjuvant, is able to elicit immune response.

The CAI value after *in silico* cloning in vector suggests that translational efficacy, mRNA codons, of both multi epitope vaccines is compatible with the host system. Moreover, it also indicates that synthesis of both vaccines is experimentally possible with higher expression within the E. *coli* K-12 system.

We believe that immunoinformatics approaches are useful to design effective vaccine candidates that when tested in experimental studies yield proposed outcomes (Ikram *et al*., 2018). All the ORFs of COVID-19 were subjected to epitope mapping; however, it is noteworthy to mention that no epitope from 3' UTRs had strong binding affinity with HLA super family alleles. Therefore, no epitopes from this region are part of either CovI-Vac and CovII-Vac vaccines. However, the multi epitopes do contain strong HLA superfamily allele binding regions from structural as well as non-structural proteins. These structural as well as nonstructural proteins play an important part in viral assembly and attachment and specific immune responses against them are required to combat the infection. Despite the numerous benefits associated with this computational study, experimental validation is required to verify these results.

Conclusions

The study was aimed to predict a vaccine against SARS-CoV-2 using immunoinformatics approaches. Two multi epitopes were designed, one without adjuvant (Cov-I-Vac) and the other with β-defensin adjuvant (Cov-II-Vac). The *in situ* designed multi epitopes Cov-I-Vac and Cov-II-Vac are predicted to be antigenic, immunogenic, thermostable and safe for human applications. The multi epitopes cover the predicted proteome of SARS-CoV-2 (Fig. 10). Cov-I-Vac demonstrates better interaction with TLR8 and more energy is required to deform the structure as compared to Cov-II-Vac. Future prospects of the study involve experimental validation of these results.

.

Acknowledgements

We would like to thank administration of Atta ur Rahman School of Applied Biosciences (ASAB) for the administrative support needed to conduct this study.

References

Abdelmageed, M. I. *et al*. (2020) 'Design of multi epitope-based peptide vaccine against E protein of human 2019-nCoV: An immunoinformatics approach', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.02.04.934232. doi: 10.1101/2020.02.04.934232.

Deleted: validating
Deleted: that
Deleted: shows
Commented [BO11]: Didn't Cov-II-Vac have more epitopes?
Deleted: both
Deleted: ,
Deleted: h
Deleted: ers
Deleted: p
Deleted: whole
Deleted:
Deleted: shows
Deleted: ¶

Allaker, R. P. (2008) 'Host defence peptides—a bridge between the innate and adaptive immune responses', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. Oxford Academic, 102(1), pp. 3–4. doi: 10.1016/j.trstmh.2007.07.005.

Amanna, I. J., Carlson, N. E. and Slifka, M. K. (2007) 'Duration of humoral immunity to common viral and vaccine antigens', *New England Journal of Medicine*. Massachussetts Medical Society, 357(19), pp. 1903–1915. doi: 10.1056/NEJMoa066092.

Amanna, I. J. and Slifka, M. K. (2011) 'Contributions of humoral and cellular immunity to vaccine-induced protection in humans', *Virology*. NIH Public Access, pp. 206–215. doi: 10.1016/j.virol.2010.12.016.

Andreatta, M. and Nielsen, M. (2016) 'Gapped sequence alignment using artificial neural networks: Application to the MHC class i system', *Bioinformatics*. Oxford University Press, 32(4), pp. 511–517. doi: 10.1093/bioinformatics/btv639.

Bhasin, M. and Raghava, G. P. S. (2004) 'Prediction of CTL epitopes using QM, SVM and ANN techniques', *Vaccine*, 22(23–24), pp. 3195–3204. doi: 10.1016/j.vaccine.2004.02.005.

Calis, J. J. A. *et al*. (2013) 'Properties of MHC Class I Presented Peptides That Enhance Immunogenicity', *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1003266.

Ceraolo, C. and Giorgi, F. M. (2020a) 'Genomic variance of the 2019-nCoV coronavirus', *Journal of Medical Virology*. John Wiley and Sons Inc., 92(5), pp. 522–528. doi: 10.1002/jmv.25700.

Ceraolo, C. and Giorgi, F. M. (2020b) 'Phylogenomic analysis of the 2019-nCoV coronavirus', *bioRxiv*. Cold Spring Harbor Laboratory, p. 2020.02.02.931162. doi: 10.1101/2020.02.02.931162.

Chan, K. W., Wong, V. T. and Tang, S. C. W. (2020) 'COVID-19: An Update on the Epidemiological, Clinical, Preventive and Therapeutic Evidence and Guidelines of Integrative Chinese–Western Medicine for the Management of 2019 Novel Coronavirus Disease', *The American Journal of Chinese Medicine*. World Scientific Pub Co Pte Lt, pp. 1–26. doi: 10.1142/s0192415x20500378.

Chen, X., Zaro, J. L. and Shen, W. C. (2013) 'Fusion protein linkers: Property, design and functionality', *Advanced Drug Delivery Reviews*. NIH Public Access, pp. 1357–1369. doi: 10.1016/j.addr.2012.09.039.

Cheng, J. *et al*. (2005) 'SCRATCH: a protein structure and structural feature prediction server', *Nucleic Acids Research*, 33(suppl_2), pp. W72–W76. doi: 10.1093/nar/gki396.

Dar, H. *et al*. (2016) 'Prediction of promiscuous T-cell epitopes in the Zika virus polyprotein: An in silico approach', *Asian Pacific Journal of Tropical Medicine*. Elsevier (Singapore) Pte Ltd, 9(9), pp. 844–850. doi: 10.1016/j.apjtm.2016.07.004.

Dhanda, S. K., Vir, P. and Raghava, G. P. S. (2013) 'Designing of interferon-gamma inducing MHC class-II binders', *Biology Direct*. BioMed Central, 8(1), p. 30. doi: 10.1186/1745-6150-8-30.

Dimitrov, I. *et al*. (2014) 'AllergenFP: allergenicity prediction by descriptor fingerprints.', *Bioinformatics (Oxford, England)*, 30(6), pp. 846–51. doi: 10.1093/bioinformatics/btt619.

716 Doytchinova, I. A. and Flower, D. R. (2007) 'VaxiJen: A server for prediction of protective
717 antigens, tumour antigens and subunit vaccines', *BMC Bioinformatics*. BioMed Central, 8, p. 4.
718 doi: 10.1186/1471-2105-8-4.
719 Greenwood, B. (2014) 'The contribution of vaccination to global health: Past, present and
720 future', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society of
721 London. doi: 10.1098/rstb.2013.0433.
722 Grote, A. *et al*. (2005) 'JCat: a novel tool to adapt codon usage of a target gene to its potential
723 expression host', *Nucleic Acids Research*, 33(suppl_2), pp. W526–W531. doi:
724 10.1093/nar/gki376.
725 Huang, C. *et al*. (2020) 'Clinical features of patients infected with 2019 novel coronavirus in
726 Wuhan, China', *The Lancet*. Lancet Publishing Group, 395(10223), pp. 497–506. doi:
727 10.1016/S0140-6736(20)30183-5.
728 Ikram, A. *et al*. (2018) 'Exploring NS3/4A, NS5A and NS5B proteins to design conserved
729 subunit multi-epitope vaccine against HCV utilizing immunoinformatics approaches', *Scientific*
730 *Reports*. Nature Publishing Group, 8(1), pp. 1–14. doi: 10.1038/s41598-018-34254-5.
731 Jamroz, M., Kolinski, A. and Kmiecik, S. (2014) 'CABS-flex predictions of protein flexibility
732 compared with NMR ensembles.', *Bioinformatics (Oxford, England)*, 30(15), pp. 2150–4. doi:
733 10.1093/bioinformatics/btu184.
734 Katoh, K. *et al*. (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on
735 fast Fourier transform', *Nucleic Acids Research*, 30(14), pp. 3059–3066. doi:
736 10.1093/nar/gkf436.
737 Kim, J. *et al*. (2018) 'Human β-defensin 2 plays a regulatory role in innate antiviral immunity
738 and is capable of potentiating the induction of antigen-specific immunity', *Virology Journal*.
739 BioMed Central Ltd., 15(1), p. 124. doi: 10.1186/s12985-018-1035-2.
740 Ko, J. *et al*. (2012) 'GalaxyWEB server for protein structure prediction and refinement', *Nucleic*
741 *Acids Research*, 40(W1), pp. W294–W297. doi: 10.1093/nar/gks493.
742 Kremer, J. R., Mastronarde, D. N. and McIntosh, J. R. (1996) 'Computer visualization of three-
743 dimensional image data using IMOD', *Journal of Structural Biology*. doi:
744 10.1006/jsbi.1996.0013.
745 Kuriata, A. *et al*. (2018) 'CABS-flex 2.0: a web server for fast simulations of flexibility of
746 protein structures', *Nucleic Acids Research*, 46(W1), pp. W338–W343. doi: 10.1093/nar/gky356.
747 Laskowski, R. A. *et al*. (2018) 'PDBsum: Structural summaries of PDB entries', *Protein Science*.
748 Blackwell Publishing Ltd, 27(1), pp. 129–134. doi: 10.1002/pro.3289.
749 Lehrer, R. I. and Lu, W. (2012) 'α-Defensins in human innate immunity', *Immunological*
750 *Reviews*, pp. 84–112. doi: 10.1111/j.1600-065X.2011.01082.x.
751 Magnan, C. N. *et al*. (2010) 'High-throughput prediction of protein antigenicity using protein
752 microarray data.', *Bioinformatics (Oxford, England)*, 26(23), pp. 2936–43. doi:
753 10.1093/bioinformatics/btq551.

Nielsen, M. *et al*. (2003) 'Reliable prediction of T-cell epitopes using neural networks with novel sequence representations', *Protein Science*. Wiley, 12(5), pp. 1007–1017. doi: 10.1110/ps.0239403.

Okonechnikov, K. *et al*. (2012) 'Unipro UGENE: a unified bioinformatics toolkit.', *Bioinformatics (Oxford, England)*, 28(8), pp. 1166–7. doi: 10.1093/bioinformatics/bts091.

Pandey, R. K. *et al*. (2018) 'Development of multi-epitope driven subunit vaccine in secretory and membrane protein of Plasmodium falciparum to convey protection against malaria infection', *Vaccine*. Elsevier Ltd, 36(30), pp. 4555–4565. doi: 10.1016/j.vaccine.2018.05.082.

Pandey, R. K., Bhatt, T. K. and Prajapati, V. K. (2018) 'Novel Immunoinformatics Approaches to Design Multi-epitope Subunit Vaccine for Malaria by Investigating Anopheles Salivary Protein', *Scientific Reports*. Nature Publishing Group, 8(1), pp. 1–11. doi: 10.1038/s41598-018-19456-1.

Qamar, M. T. ul *et al*. (2020) 'Structural modeling and conserved epitopes prediction against SARS-COV-2 structural proteins for vaccine development', *Research Square*, pp. 1–49. doi: 10.21203/rs.2.23973/v1.

Rappuoli, R. (2001) 'Reverse vaccinology, a genome-based approach to vaccine development', in *Vaccine*, pp. 2688–2691. doi: 10.1016/S0264-410X(00)00554-5.

Van Regenmortel, M. H. V. (1996) 'Mapping epitope structure and activity: From one-dimensional prediction to four-dimensional description of antigenic specificity', *Methods: A Companion to Methods in Enzymology*. Academic Press Inc., 9(3), pp. 465–472. doi: 10.1006/meth.1996.0054.

Saadi, M., Karkhah, A. and Nouri, H. R. (2017) 'Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches', *Infection, Genetics and Evolution*. Elsevier B.V., 51, pp. 227–234. doi: 10.1016/j.meegid.2017.04.009.

Saha, S. and Raghava, G P S (2006) 'AlgPred: prediction of allergenic proteins and mapping of IgE epitopes', *Nucleic Acids Research*, 34(suppl_2), pp. W202–W209. doi: 10.1093/nar/gkl343.

Saha, S. and Raghava, G. P.S. (2006) 'Prediction of continuous B-cell epitopes in an antigen using recurrent neural network', *Proteins: Structure, Function and Genetics*. doi: 10.1002/prot.21078.

Sahin, A. R. (2020) '2019 Novel Coronavirus (COVID-19) Outbreak: A Review of the Current Literature', *Eurasian Journal of Medical Investigation*, 4(1), pp. 1–7. doi: 10.14744/ejmo.2020.12220.

Sallusto, F., Geginat, J. and Lanzavecchia, A. (2004) ' C entral M emory and E ffector M emory T C ell S ubsets : Function, Generation, and Maintenance ', *Annual Review of Immunology*. Annual Reviews, 22(1), pp. 745–763. doi: 10.1146/annurev.immunol.22.012703.104702.

Saqlain, M. *et al*. (2020) 'Is Pakistan prepared to tackle the coronavirus epidemic?', *Drugs & Therapy Perspectives*. Springer, pp. 1–2. doi: 10.1007/s40267-020-00721-1.

Singh, H. and Raghava, G. P. S. (2002) 'ProPred: Prediction of HLA-DR binding sites', *Bioinformatics*, 17(12), pp. 1236–1237. doi: 10.1093/bioinformatics/17.12.1236.

Tani, K. *et al*. (2000) 'Defensins act as potent adjuvants that promote cellular and humoral immune responses in mice to a lymphoma idiotype and carrier antigens', *International Immunology*, 12(5), pp. 691–700. doi: 10.1093/intimm/12.5.691.
Tomar, N. and De, R. K. (2010) 'Immunoinformatics: An integrated scenario', *Immunology*. Wiley-Blackwell, pp. 153–168. doi: 10.1111/j.1365-2567.2010.03330.x.
WHO (2020) *Coronavirus disease (COVID-19) Pandemic*. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019.
Wilkins, M. R. *et al*. (1999) 'Protein identification and analysis tools in the ExPASy server.', *Methods in molecular biology (Clifton, N.J.)*, pp. 531–552. doi: 10.1385/1-59259-584-7:531.
Worldometers.info (2020) *Coronavirus Cases: Pakistan*. Dover, Delaware, U.S.A. Available at: https://www.worldometers.info/coronavirus/country/pakistan/.
Zhu, N. *et al*. (2020) 'A novel coronavirus from patients with pneumonia in China, 2019', *New England Journal of Medicine*. Massachussetts Medical Society, 382(8), pp. 727–733. doi: 10.1056/NEJMoa2001017.
Van Zundert, G. C. P. *et al*. (2016) 'The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes', *Journal of Molecular Biology*. Academic Press, 428(4), pp. 720–725. doi: 10.1016/j.jmb.2015.09.014.

Page 11: [1] Deleted	BRENDA OPPERT	4/11/20 11:26:00 AM

Page 11: [2] Commented [BO7]	BRENDA OPPERT	4/11/20 11:28:00 AM

Explain why you added to Cov-II-Vac instead of Cov-1-Vac

Page 11: [3] Commented [BO8]	BRENDA OPPERT	4/11/20 11:31:00 AM

In the human bloodstream? Biologically?

Page 11: [4] Deleted	BRENDA OPPERT	4/11/20 11:34:00 AM