 doi:10.20944/preprints202003.0242.v1

A putative prophylactic solution for COVID-19: Development of novel multiepitope vaccine candidate against SARS-COV-2 by comprehensive immunoinformatic and molecular modelling approach

Hafiz Muzzammel Rehman [1, 2], Muhammad Usman Mirza [3], Mahjabeen Saleem [1]*, Matheus Froeyen [3], Sarfraz Ahmad [4], Roquyya Gul [5], Muhammad Shahbaz Aslam [1], Muhammad Sajjad [6], Munir Ahmad Bhinder [2]

[1] Institute of Biochemistry and Biotechnology, University of the Punjab, Lahore

[2] Department of Human Genetics and Molecular Biology, University of Health Sciences, Lahore.

[3] Department of Pharmaceutical and Pharmacological Sciences, Rega Institute for Medical Research, Medicinal Chemistry, University of Leuven, B-3000 Leuven, Belgium.

[4] Department of Chemistry, Faculty of Science, University of Malaya, 50603 Kuala Lumpur, Malaysia.

[5] Faculty of Life Sciences, Gulab Devi Educational Complex, Lahore.

[6] School of Biological Sciences, University of the Punjab, Quaid e Azam Campus, Lahore.

Corresponding author

Dr. Mahjabeen Saleem

Email: mahjabeensaleem1@hotmail.com

Institute of Biochemistry and Biotechnology, University of the Punjab, Lahore

 doi:10.20944/preprints202003.0242.v1

Abstract

The outbreak of 2019-novel coronavirus (SARS-CoV-2) that causes severe respiratory infection (COVID-19) has spread in China, and the world health organization declared it pandemic. However, no approved drug or vaccines are available, and treatment is mainly supportive and through a few repurposed drugs. In this urgency situation, development of SARS-CoV-2 based vaccines is immediately required. Immunoinformatic and molecular modelling are generally used time-efficient methods to accelerate the discovery and design of the candidate peptides for vaccine development. In recent years, the use of multiepitope vaccines is proved to be a promising immunization strategy against viruses and pathogens, which induce more comprehensive protective immunity. The current study demonstrated a comprehensive in-silico strategy to design stable multiepitope vaccine construct (MVC) from B-cell and T-cell epitopes of essential SARS-CoV-2 proteins with the help of adjuvants and linkers. The integrated molecular dynamics simulations analysis revealed the stability of MVC and its interaction with human Toll-like receptors (TLRs), which trigger an innate and adaptive immune response. Later, the in-silico cloning in a known pET28a vector system also estimated the possibility of MVC expression in E. Coli. Despite this study lacks validation of this vaccine construct in terms of its efficacy, the current integrated strategy encompasses the initial multiple epitope vaccine design concepts. After validation, this MVC can present to be a better prophylactic solution against COVID-19.

Keywords: COVID-19, SARS-CoV-2, Spike protein, Multiepitope vaccine, Molecular Modeling

Introduction:

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) enveloped, non-segmented positive-sense RNA virus, causes severe respiratory infection [1]. The ongoing 2019–20 outbreak of coronavirus disease 2019 (COVID-19) [2-4] has led to 4012 deaths with 113702 confirmed cases globally, and World Health Organization (WHO) has declared COVID-19 a global health emergency [5]. Coronaviruses are highly pathogenic viruses and are known to be contagious, which was revealed by the SARS and MERS (Middle East respiratory syndrome) outbreak in 2002 and

 doi:10.20944/preprints202003.0242.v1

2012 [6,7]. The recent SARS-CoV-2 is considered as a seventh known human coronavirus (HCoV) from the same family after 229E, NL63, OC43, HKU1, MERS-CoV, and SARS-CoV [8].

Like other coronaviruses, SARS-CoV-2 is spherical having a diameter of about 125nm and genome (~30kb) contains at least six open reading frames which encode 16 non-structural proteins and ORFs near 3'end of the genome encodes four major structural proteins named as spike protein (S) a form of glycoprotein, a membrane protein (M) which consist of the membrane and an envelope protein (E), and a nucleocapsid (N) protein. Among these structural proteins, the spike (S) glycoprotein binds to the cellular receptor angiotensin-converting enzyme 2 (ACE2), and responsible for causing viral infection [9]. The S precursor protein of SARS-CoV-2 can be proteolytically cleaved into S1 (685 amino acids) and S2 (588 amino acids) subunits [10]. Due to the integral role of S protein between viral and host cell membrane, it could be a potential target for developing new SARS-CoV-2 vaccines. Previous studies related to the development of anti-SARS-CoV vaccines and therapeutics that target S protein have already been reported [11-14]. Together with, most of the non-structural proteins play an essential role in viral replication, including SARS-CoV-2 main protease (Mpro) or chymotrypsin-like protease (3CLpro) [15,16], Nsp13 helicase [17] and, the Nsp12 RNA-dependent RNA polymerase [18], which are highly conserved among coronaviruses [19].

Owing to the high mortality rate of patients, there is an urgent need to develop vaccines and anti-viral drugs to combat with COVID-19 outbreak. Although phenomenal efforts are in process in developing vaccines and through repurposing studies [20-22]. With the advancement in computational biology, it is now possible to accelerate the drug discovery pipeline and vaccine development [23], and these methods have surpassed the conventional methods [24,25]. Numerous studies have been published related to the B and T-cell epitope-based vaccine development using *in silico* immunoinformatics methods [26-31].

Keeping in view the urgency situation of the COVID-19 outbreak, developing an effective vaccine is, therefore, a prime research priority. The current study deals with the modelling of novel multiepitope vaccine for SARS-CoV-2 by using a cost-effective integrated immunoinformatics approach. This approach has been proved to be promising against viral diseases caused by viruses like yellow fever [32], SARS-CoV [33], influenza [34], Zika [35], Congo Virus [36] and pathogens including

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

L. donovani [37], *S. pneumoniae* [38]. Few multiple epitope vaccine strategies proved to be effective against *H. pylori* infection in BALB/c mice model [39,40], chronic hepatitis B virus infection [41] and foot-and-mouth disease virus (serotype A) in pigs [42].

In the present study, we effectively designed the multiepitope subunit vaccine construct (MVC) by considering the potential B- and T-cell epitopes of SARS-CoV-2 Spike, Mpro, Nsp-12 polymerase and Nsp13 helicase proteins. The antigenicity, allergenicity and physiochemical properties of B- and T-cell epitopes were also measured. Later, the structural analysis of MVC interaction with Toll-like receptors (TLRs) was analyzed through molecular dynamics (MD) simulations and binding free energies were estimated. TLRs establish an important link between innate and adaptive immunity. Engagement of TLR signalling pathways is a promising mechanism for accelerating vaccine responses and involved in therapeutic immunization against infectious diseases [43]. Thus, the interaction of a multiepitope vaccine construct designed through an integrated modelling approach may trigger innate and specific adaptive immunity by activating TLR signalling pathways and may produce a promising immune response against SARS-CoV-2.

Results:

In the present research, plausible T-cells and B-cell epitopes (discontinuous & continuous) have been recognized to design peptide vaccines to counter SARS-CoV-2 S protein, Mpro, Nsp 12 RNA polymerase (RdRp) and Nsp13 helicase. Most potential epitopes were selected and joined together with appropriate linkers and adjuvant. The 3D model was generated using various online servers, and a reliable model was used for docking and MD simulation studies. Docking and immuno-informatics method are helpful for prediction of the binding interaction and between TLRs and ligand (multiepitope vaccine) complexes and analysis was done as these are proved useful tools in identifying novel multiepitope vaccines [23,29].

Antigenic B-Cell Epitope Prediction:

Depending on the physicochemical properties of amino acids, that have already been observed in practically determined antigen-based epitopes, Kolaskar & Tongaonkar's approach was used for predicting antigenic epitopes of provided sequences. Seventy-five percent

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

experimental precision has been reported for this approach [44]. By using this method, 11 antigenic peptides with 9 - 14 amino acid length, including two heptapeptides from SARS-CoV-2 Mpro (**Table 1**). Likewise, out of 932 amino acid, 37 antigenic peptides were predicted in SARS-CoV-2 Nsp12 polymerase. For RdRp, the length of the antigenic peptides was 6 - 29 amino acid along with ten heptapeptides and nine octapeptides (**Table 2**). For Nsp13 helicase, 18 antigenic peptides were predicted, and the length of the antigenic peptide was 6 - 38 with seven hexapeptides (**Table 3**). For the spike protein, 46 antigenic peptide were predicted from 1273 amino acids **(Table 4)**. Moreover, Kolaskar & Tongaonkar's approach also projected the highest residual score of each amino acid in all investigated proteins. In SARS-CoV-2 Mpro, 211 out of 306 amino acids have greater than 1.000 residual scores. From position 85 to 91, the antigenic peptide (CVLKLKV) having lysine at position 88th was identified with a maximal residual score of 1.22. The Nsp12-RdRp has 686 residues out of 932 with a residual score above 1.000 and valine at position 473, in the peptide (LFVVEVV) from 470 to 476, has a maximum residual score of 1.246.

Likewise, for Nsp13 helicase protein, 479 out of 601 amino acids were predicted with a residual score greater than 1.000, and lysine present at 28th position in an antigenic peptide from 25 to 31 (LCCKCCY) has a maximum residual score of 1.284. For SARS-CoV-2 S protein, 958 out of 1273 amino acids were predicted to have a residual score higher than 1.000 and leucine at position 8 of the antigenic peptide from position 5 to 11 (LVLLPLV) showed a maximal residual score of 1.261.

Graphical depiction of peptides predicted for B cell from investigated SARS-CoV-2 proteins based on sequence position along the x-axis, and antigenic propensity (AP) as y-axis are shown in **Figure SI – S4**. Divergence in AP is related to the length of the sequence. Minimum AP score for Mpro was 0.844, and maximum AP score was 1.22 (A) while the maximum and minimum AP scores of Nsp12, Nsp13 and S protein were 1.246,1.284, and1.261 and 0.858,0.893, and 0.866 respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 1: Predicted antigenic B-cell epitopes of SARS-CoV-2 Main protease

No.	Start	End	Peptide	Length
1	15	23	GCMVQVTCG	9
2	32	45	LDDVVYCPRHVICT	14
3	65	72	NFLVQAGN	8
4	83	91	QNCVLKLKV	9
5	101	107	**YK**FVRIQ	7
6	111	120	**TFSVLACY**NG	10
7	123	129	SGVYQCA	7
8	153	162	**DYDCVSFCY**M	10
9	201	212	**TVNVLAWLY**AAV	12
10	244	253	QDHVDILGPL	10
11	258	271	GIAVLDMCASLKEL	14

***** 11 antigenic sites were predicted from protease. Residues (in bold) were also predicted as CTL epitope.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 2: Predicted antigenic B-cell epitopes of SARS-CoV-2 Nsp12 RNA polymerase

No.	Start	End	Peptide	Length
17	395	400	CFSVAA	6
3	50	56	KTNCCRF	7
8	171	177	ILRVYAN	7
10	201	207	IVGVLTL	7
13	327	333	GPLVRKI	7
20	557	563	VAGVSIC	7
21	573	579	QKLLKSI	7
22	585	591	ATVVIGT	7
26	670	676	**GGSLYVK**	7
28	725	731	HRLYECL	7
31	773	779	QGLVASI	7
2	28	35	**TDVVYRAF**	8
6	125	132	ADLVYALR	8
15	350	357	ELGVVHNQ	8
16	369	376	**KELLVY**AA	8
18	435	442	**VELKHFFF**	8
23	633	640	MASLVLAR	8
29	744	751	**EFY**AYLRK	8
32	783	790	KSVLYYQN	8
34	825	832	**DYVYLPY**P	8
4	67	75	DSYFVVKRH	9
25	658	666	ECAQVLSEM	9
30	760	768	**DDAVVCFNS**	9
35	839	847	GAGCFVDDI	9
36	859	867	**FVSLAIDAY**	9
1	8	17	LNRVCGVSAA	10
27	694	703	FNICQA**VTAN**	10
33	810	819	HEFCSQHT**ML**	10
7	144	154	EILVTYNCCDD	11
9	183	193	RQALLKTVQFC	11
14	335	345	**VDGVPFVV**STG	11
24	643	653	TT**CCSLSHRFY**	11
5	87	99	YNLLKDCPAVAKH	13
37	878	890	**ADVFHLYLQYIRK**	13
19	466	482	IRQLLFVVEVVDKYFDC	17
11	230	248	**GVPVVDSYYSLL**MPILTLT	19
12	295	323	HPNCVNCLDDRCILHCANFNVLFSTVFPP	29

***** 37 antigenic sites were predicted. Residues (in bold) were also predicted as CTL epitope.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 3: Predicted antigenic B-cell epitopes of SARS-CoV-2 Nsp13 helicase protein

No.	Start	End	Peptide	Length
3	70	75	YYCKSH	6
5	207	212	DAVVYR	6
10	369	374	DIVVFD	6
11	384	389	LSVVNA	6
13	423	428	NSVCRL	6
15	493	498	IGVVRE	6
17	542	547	**DYVIF**T	6
12	394	400	KHYVYIG	7
16	522	528	ASKILGL	7
18	570	576	VGIL**CIM**	7
1	4	11	ACVLCNSQ	8
6	222	230	GDYFVLTSH	9
9	353	361	**EQY**VFCTVN	9
4	78	87	PISFPLC**ANG**	10
14	449	458	**VDTVSALV**YD	10
7	237	250	A**PTLVPQEHYVRIT**	14
8	292	325	AIGLALYYPSARIVYTACSHAAVDALCEKALKYL	34
2	21	58	RRPFLCCKCCYDHVISTSHKLVLSVNPYVCNAPGC**DVT**	38

***** 18 antigenic sites were predicted. Residues (in bold) were also predicted as CTL epitopes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 4: Predicted antigenic B-cell epitopes of SARS-CoV-2 spike protein

No.	Start	End	Peptide	Length
1	4	18	FLVLLPLVSSQCVNL	15
2	34	41	**RGVYYP**DK	8
3	44	51	RSSVLH**ST**	8
4	53	60	**DLFLPF**FS	8
5	65	70	FHAIHV	6
6	81	87	NP**VLPFN**	7
7	115	121	QSLLIVN	7
8	125	134	NVVIKVCEFQ	10
9	136	146	**CNDPFLGVY**YH	11
10	168	174	**F**EYVSQP	7
11	210	216	INLVRDL	7
12	223	230	LEPLVDLP	8
13	239	248	QTLLALHRSY	10
14	263	270	**AAYY**VGYL	8
15	272	278	PRTFLLK	7
16	288	295	**AVDCAL**DP	8
17	333	339	TNLCPFG	7
18	359	371	SN**CVADYSVLYNS**	13
19	376	385	**TFKCY**GVSPT	10
20	430	435	TGCVIA	6
21	488	495	CYFPLQSY	8
22	505	527	YQPYRVVVLSFELLHAPATVCGP	23
23	592	599	FGGVSVIT	8
24	607	615	**QVAVLYQDV**	9
25	617	627	**CTEV**PVAIHAD	11
26	647	653	AGCLIGA	7
27	667	674	GAGICASY	8
28	687	693	**VASQSII**	7
29	723	730	TTEILPVS	8
30	735	741	**SVDCTMY**	7
31	750	763	**SNLLLQY**GSFCTQL	14
32	781	788	VFAQVKQI	8
33	803	808	SQILPD	6
34	837	843	YGDCLGD	7
35	847	853	RDLICAQ	7
36	858	864	LTVLPPL	7
37	873	880	**Y**TSALLA**G**	8
38	959	966	LNTLVKQL	8
39	973	979	ISSVLND	7
40	1003	1011	SLQTYVTQQ	9
41	1030	1037	SECVLGQS	8
42	1057	1070	PHGVVFLHVTYVPA	14
43	1079	1085	PAICHDG	7
44	1123	1132	SGNCDVVIGI	10
45	1174	1179	ASVVNI	6
46	1221	1256	IAGLIAIVMVTIMLCC**MTSCCSCLK**GCCSCGSCCKF	36

***** 46 antigenic sites were predicted. Residues (in bold) were also predicted as CTL epitope.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Prediction of Cytotoxic T-lymphocyte (CTL) epitopes:

An infected cell having antigen-presentation triggers the T-cell to turn out as an effector cell and kill the infected cells. Cell death or self-destruction is detected after the attack of CTLs on effected cells. The pathogen's peptide fragment and molecule of MHC interact and expose on the cell surface of infected cells. CTLs identify the complex of peptide-protein; also, as a consequence, infected cells are killed. The processing of fragment of the peptide (antigen), along with its appearance to the T-cell, is achieved through different steps. Peptides are treated in the cytoplasm through proteasome and transferred to the endoplasmic reticulum (ER) later on where MHC is produced. Here, the peptide is transported to MHC I molecule by the transporter associated with antigen processing (TAP). Afterwards, a complex of peptide-MHC-I is transferred to the surface of the cell. A varied array of peptides is attached to each allelic type of MHC-I protein. The molecule of MHC can interact with peptides strongly as the pathogens attempt to mutate the MHC molecule's epitope. Therefore, MHC molecule displays strong binding with a diversity of peptides[45].

Prediction of CTL epitope is significant *in silico* tool in designing the vaccine as it decreases the time and necessity for *in vitro* trials. NetCTL 1.2 server [46] was used for the prediction of CTL epitope. For all investigated SARS-CoV-2 protein, the peptide sequences were predicted as CTL epitopes based on three main factors, which include their MHC binding capacity, proteasomal cleavage of the peptide from C-terminal and affinity for the TAP transporter with the default threshold prediction score was >0.75000. Among all the peptides, 11 peptide sequences from S protein, 4 peptides from Mpro, 19 from Nsp12 and 10 from Nsp13 were selected as CTL epitopes. These CTL epitopes were also predicted as an antigenic site. Hence these peptides can be considered as potential vaccine candidates **(Table 5 to 8)**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 5: Predicted CTL from SARS-CoV-2 S protein

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score	MHC ligand
604	TSN**QVAVLY**	**0.6559**	**2.7847**	**0.9440**	**2.9910**	**3.0758**	**yes**
361	**CVADYSVLY**	**0.5348**	**2.2705**	**0.9764**	**3.1800**	**2.5759**	**yes**
733	KT**SVDCTMY**	**0.4908**	**2.0840**	**0.9649**	**3.0160**	**2.3795**	**yes**
687	**VASQSII**AY	**0.3529**	**1.4986**	**0.9656**	**3.0890**	**1.7978136**	**yes**
136	**CNDPFLGVY**	**0.2613**	**1.1095**	**0.6900**	**2.4500**	**1.3355**	**yes**
261	GA**AAYYVGY**	**0.2253**	**0.9568**	**0.7608**	**2.9690**	**1.2194**	**yes**
357	RI**SNCVADY**	**0.2106**	**0.8941**	**0.9292**	**3.3940**	**1.2032**	**yes**
285	ITD**AVDCAL**	**0.2350**	**0.9979**	**0.8708**	**0.7900**	**1.1680**	**yes**
1237	**MTSCCSCLK**	**0.2260**	**0.9595**	**0.7525**	**0.4790**	**1.0963**	**yes**
50	**ST**Q**DLFLPF**	**0.1974**	**0.8383**	**0.5530**	**2.5110**	**1.0468**	**yes**
748	EC**SNLLLQY**	**0.1413**	**0.6000**	**0.5316**	**2.7470**	**0.8171**	**yes**

* Threshold was set >0.75000, Bold shows the amino acids which were also predicted as antigenic sites

Table 6: CTL prediction from SARS-CoV-2 Main protease*.

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score	MHC ligand
201	**TVNVLAWLY**	**0.6255**	**2.6559**	**0.8852**	**2.9570**	**2.9365**	**yes**
110	Q**TFSVLACY**	**0.2625**	**1.1146**	**0.9725**	**2.9980**	**1.4104**	**yes**
153	**DYDCVSFCY**	**0.2097**	**0.8905**	**0.9722**	**0.9722**	**1.1717**	**yes**
93	TANPKTP**KY**	**0.1676**	**0.7118**	**0.9755**	**2.6760**	**0.9088**	**yes**

***** Threshold was set >0.75000, Bold shows the amino acids which were also predicted as antigenic sites.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 7: Predicted CTL from SARS-CoV-2 Nsp12 RdRp.

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score	MHC ligand
738	DTDFVN**EFY**	**0.7922**	**3.3634**	**0.8873**	**2.4580**	**3.6194**	**yes**
336	LSF**KELLVY**	**0.3898**	**1.6552**	**0.9676**	**3.2130**	**1.9610**	**yes**
27	S**TDVVYRAF**	**0.4019**	**1.7065**	**0.6174**	**2.4000**	**1.9191**	**yes**
859	**FVSLAIDAY**	**0.3709**	**1.5746**	**0.7669**	**3.0960**	**1.8444**	**yes**
666	MVMC**GGSLY**	**0.3637**	**1.5441**	**0.9482**	**3.0080**	**1.8368**	**yes**
758	LS**DDAVVCF**	**0.3143**	**1.3345**	**0.9556**	**2.4120**	**1.5985**	**yes**
686	TTAYANSV**F**	**0.2963**	**1.2580**	**0.4772**	**2.6630**	**1.4627**	**yes**
762	**AVVCFNS**TY	**0.2435**	**1.0339**	**0.9754**	**3.1460**	**1.3375**	**yes**
463	MCD**IRQLLF**	**0.2518**	**1.0691**	**0.1005**	**2.4360**	**1.2060**	**yes**
233	**VVDSYYSLL**	**0.2332**	**0.9901**	**0.7134**	**0.8340**	**1.1388**	**yes**
700	**VTAN**VNALL	**0.2007**	**0.8523**	**0.9705**	**1.1660**	**1.0562**	**yes**
818	**ML**VKQGD**DY**	**0.1793**	**0.7614**	**0.8328**	**3.0790**	**1.0403**	**yes**
823	**GDDYVYLPY**	**0.1821**	**0.7733**	**0.8456**	**2.2130**	**1.0108**	**yes**
879	**DVFHLYLQY**	**0.1677**	**0.7119**	**0.9529**	**3.0130**	**1.0055**	**yes**
876	**EYADVFHLY**	**0.1624**	**0.6894**	**0.9603**	**2.9530**	**0.9811**	**yes**
230	**GVPVVDSYY**	**0.1504**	**0.6386**	**0.9521**	**2.9230**	**0.9276**	**yes**
434	S**VELKHFFF**	**0.1454**	**0.6176**	**0.9285**	**2.6360**	**0.8886**	**yes**
334	**FVDGVPFVV**	**0.1739**	**0.7382**	**0.8437**	**0.1910**	**0.8743**	**yes**
645	**CCSLSHRFY**	**0.1586**	**0.6732**	**0.2740**	**2.9100**	**0.8598**	**yes**

* Threshold was set >0.75000, Bold shows the amino acids which were also predicted as antigenic sites.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 8: Predicted CTL from SARS-CoV-2 Nsp13 helicase.

Residue number	Peptide sequence	Predicted MHC binding affinity	Rescale binding affinity	C-terminal cleavage affinity	Transport affinity	Prediction score	MHC ligand
57	**VT**DVTQLYL	**0.4708**	**1.9988**	**0.6073**	**0.6800**	**2.1239**	**yes**
56	**DVT**DVTQLY	**0.2890**	**1.2271**	**0.9651**	**2.7040**	**1.5071**	**yes**
535	SSQGSEY**DY**	**0.2761**	**1.1724**	**0.8149**	**2.8470**	**1.4370**	**yes**
238	**PTLVPQEHY**	**0.1794**	**0.7617**	**0.8719**	**2.5950**	**1.0222**	**yes**
448	I**VDTVSALV**	**0.1991**	**0.8453**	**0.8977**	**0.1330**	**0.9866**	**yes**
574	**CIM**SDRDLY	**0.1634**	**0.6937**	**0.1836**	**3.1250**	**0.8775**	**yes**
347	KVNSTL**EQY**	**0.1391**	**0.5907**	**0.8156**	**2.9710**	**0.8616**	**yes**
245	**HYVRIT**GLY	**0.1102**	**0.4678**	**0.9598**	**3.0090**	**0.7622**	**yes**
85	**ANG**QVFGLY	**0.1141**	**0.4845**	**0.9132**	**2.7460**	**0.7588**	**yes**
538	GSEY**DYVIF**	**0.1401**	**0.5947**	**0.3528**	**2.2030**	**0.7578**	**yes**

* Threshold was set >0.75000, Bold shows the amino acids which were also predicted as antigenic sites.

Structure-based epitope prediction

ElliPro was used to predict the epitopes from the 3D structure of proteins [47]. This advanced program is web-based and used to study the correlation among antigenicity, flexibility, solvent accessibility of proteins' structure. Furthermore, differentiation of predicted epitopes based on interactions of protein-antibody is an essential property of this program. ElliPro measures the PI score (Protrusion Index) that shows the percent of atoms of proteins that spread beyond the molecular mass/bulk as well as responsible for binding antibodies. Based on the PI score (> 0.7), 5, 3 and 2 discontinuous epitopes were selected for SARS-CoV-2 S, Nsp13 helicase and Nsp12 polymerase, while only one epitope was identified for Mpro that showed PI > 0.7. The graphical illustration of discontinuous epitopes is shown in **Figure 1**, while residues number and epitope scores are tabulated in **Table 9 - 12**.

Table 9: Conformational epitopes of SARS-CoV-2 Mpro

No.	Residues	Number of residues	Score
1	A:S1, A:G2, A:F3, A:A211, A:V212, A:I213, A:N214, A:G215, A:D216, A:R217, A:W218, A:F219, A:L220, A:N221, A:R222, A:F223, A:T224, A:T225, A:T226, A:L227, A:N228, A:D229, A:F230, A:N231, A:L232, A:V233, A:A234, A:M235, A:K236, A:Y237, A:Y239, A:E240, A:P241, A:L242, A:T243, A:Q244, A:D245, A:V247, A:D248, A:L250, A:G251, A:P252, A:S254, A:A255, A:Q256, A:T257, A:G258, A:I259, A:A260, A:V261, A:L262, A:D263, A:A266, A:S267, A:K269, A:E270, A:L271, A:L272, A:Q273, A:N274, A:G275, A:M276, A:N277, A:G278, A:R279, A:T280, A:I281, A:L282, A:G283, A:S284, A:A285, A:L286, A:C300, A:S301, A:G302	75	0.716

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table 10: Conformational epitopes of SARS-CoV-2 S protein

No.	Residues	Number of residues	Score
1	A:L119, A:T120, A:K121, A:Y122, A:T123, A:D126, A:D135, A:E136, A:G137, A:N138, A:C139, A:D140, A:T141, A:K143, A:E144, A:I145, A:L146, A:V147, A:T148, A:Y149, A:N150, A:C151, A:C152, A:D153, A:D154, A:D155, A:Y156, A:F157, A:N158, A:K159, A:W162, A:Y163, A:N168, A:P169, A:D170, A:R173, A:V174, A:N177, A:L178, A:E180, A:R181, A:R183, A:Q184, A:A185, A:L187, A:K188, A:T189, A:V190, A:Q191, A:F192, A:C193, A:D194, A:A195, A:M196, A:R197, A:N198, A:A199, A:G200, A:I201, A:V202, A:G203, A:V204, A:L205, A:T206, A:D208, A:N209, A:Q210, A:D211, A:L212, A:N213, A:G214, A:N215, A:W216, A:Y217, A:D218, A:F219, A:G220, A:D221, A:F222, A:I223, A:Q224, A:T225, A:T226, A:P227, A:G228, A:S229, A:G230, A:V231, A:P232, A:V233, A:V234, A:A250, A:D284, A:K288, A:Y289	95	0.728
2	A:D269, A:L270, A:L271, A:K272, A:Y273, A:D274, A:F275, A:E277, A:E278, A:K281, A:T324, A:L329, A:V330, A:R331, A:K332, A:I333, A:F334, A:V335, A:D336, A:G337, A:V338, A:P339, A:F340, A:V341, A:V342, A:S343, A:T344, A:H355, A:N356, A:Q357, A:D358, A:V359, A:N360, A:L361, A:H362, A:S363, A:S364, A:R365, A:L366, A:S367, A:F368, A:K369, A:E370, A:L371, A:L372, A:V373, A:Y374, A:D377, A:P378, A:A379, A:M380, A:H381, A:A382, A:A383, A:S384, A:G385, A:N386, A:L387, A:L388, A:L389, A:D390, A:K391, A:R392, A:T393, A:A399, A:A400, A:L401, A:T402, A:N403, A:N404, A:V405, A:A406, A:F407, A:Q408, A:T409, A:V410, A:K411, A:P412, A:G413, A:N414, A:F415, A:N416, A:K417, A:D418, A:F419, A:Y420, A:D421, A:F422, A:A423, A:V424, A:S425, A:K426, A:G427, A:F428, A:F429, A:K430, A:E431, A:G432, A:S433, A:S434, A:V435, A:E436, A:L437, A:K438, A:H439, A:F440, A:F441, A:F442, A:A443, A:Q444, A:D445, A:G446, A:N447, A:C487, A:I488, A:N489, A:A490, A:N491, A:Q492, A:V493, A:D517, A:S518, A:M519, A:S520, A:Y521, A:E522, A:D523, A:Q524, A:D525, A:A526, A:L527, A:A529, A:Y530, A:T531, A:K532, A:R533, A:N534, A:V535, A:I536, A:Y546, A:A550, A:F594, A:Y595, A:G596, A:H599, A:N600, A:K603, A:S607, A:D608, A:V609, A:E610, A:N611, A:P612, A:H613, A:H642, A:T643, A:T644, A:C645, A:C646, A:S647, A:H650, A:G670, A:G671, A:T710, A:D711, A:G712, A:N713, A:K714, A:I715, A:A716, A:D717, A:K718, A:Y719, A:V720, A:R721, A:N722, A:L723, A:R726, A:C730, A:V737, A:D738, A:T739, A:D740, A:F741, A:N743, A:E744, A:K751, A:H752, A:N767, A:S768, A:T769, A:Y770, A:S772, A:Q773, A:G774, A:L775, A:V776, A:T801, A:E802, A:T803, A:D804, A:L805, A:T806, A:K807, A:G808, A:M818, A:L819, A:V820, A:K821, A:Q822, A:G823, A:D824, A:D825, A:Y826, A:V827, A:Y828, A:L829, A:P832, A:D833, A:P834, A:L838, A:G839, A:G841, A:C842, A:F843, A:V844, A:D845, A:D846, A:I847, A:V848, A:K849, A:T850, A:D851, A:G852, A:T853, A:L854, A:M855, A:I856, A:E857, A:F859, A:V860, A:A863, A:I864, A:A866, A:Y867, A:P868, A:L869, A:T870, A:K871, A:H872, A:P873, A:N874, A:Q875, A:E876, A:Y877, A:A878, A:D879, A:V880, A:F881, A:H882, A:L883, A:Y884, A:L885, A:Q886, A:Y887, A:I888, A:R889, A:K890, A:L891, A:H892, A:D893, A:E894, A:L895, A:T896, A:G897, A:H898, A:M899, A:L900, A:D901, A:M902, A:Y903, A:S904, A:V905, A:M906, A:L907, A:T908, A:N909, A:D910, A:N911, A:T912, A:S913, A:R914, A:Y915, A:W916, A:E917, A:P918, A:E919	297	0.719

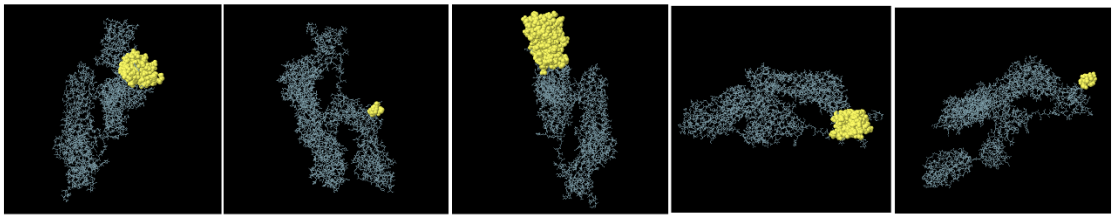
Table11. Conformational epitopes from SARS-CoV-2 Nsp12 polymerase

No.	Residues	Number of residues	Score
1	A:D1139, A:P1140, A:L1141, A:Q1142, A:P1143, A:E1144, A:L1145, A:D1146	8	0.975
2	A:Y707, A:S708, A:N709, A:N710, A:S711, A:I712, A:A713, A:I714, A:P715, A:T716, A:N717, A:Q1071, A:K1073, A:N1074, A:F1075, A:T1076, A:T1077, A:A1078, A:P1079, A:A1080, A:I1081, A:C1082, A:H1083, A:D1084, A:G1085, A:K1086, A:A1087, A:H1088, A:F1089, A:P1090, A:R1091, A:E1092, A:G1093, A:V1094, A:F1095, A:V1096, A:S1097, A:N1098, A:G1099, A:T1100, A:H1101, A:W1102, A:F1103, A:V1104, A:T1105, A:Q1106, A:R1107, A:F1109, A:Y1110, A:E1111, A:P1112, A:Q1113, A:I1114, A:I1115, A:T1116, A:T1117, A:D1118, A:N1119, A:T1120, A:F1121, A:V1122, A:S1123, A:G1124, A:N1125, A:C1126, A:D1127, A:V1128, A:V1129, A:I1130, A:G1131, A:I1132, A:V1133, A:N1134, A:N1135, A:T1136, A:V1137, A:Y1138	77	0.845
3	A:L335, A:C336, A:P337, A:F338, A:G339, A:E340, A:V341, A:F342, A:N343, A:A344, A:T345, A:R346, A:F347, A:A348, A:S349, A:V350, A:Y351, A:A352, A:W353, A:N354, A:R355, A:K356, A:R357, A:I358, A:S359, A:N360, A:C361, A:V362, A:A363, A:D364, A:Y365, A:S366, A:V367, A:L368, A:Y369, A:N370, A:S371, A:A372, A:S373, A:F374, A:S375, A:T376, A:F377, A:K378, A:C379, A:Y380, A:L390, A:C391, A:F392, A:T393, A:N394, A:V395, A:Y396, A:A397, A:D398, A:S399, A:F400, A:V401, A:I402, A:R403, A:G404, A:D405, A:E406, A:V407, A:R408, A:Q409, A:I410, A:A411, A:P412, A:G413, A:Q414, A:T415, A:G416, A:K417, A:I418, A:A419, A:D420, A:Y421, A:N422, A:Y423, A:K424, A:L425, A:P426, A:D427, A:D428, A:F429, A:T430, A:G431, A:C432, A:V433, A:I434, A:A435, A:W436, A:N437, A:S438, A:N439, A:N440, A:L441, A:D442, A:S443, A:Y449, A:N450, A:Y451, A:L452, A:Y453, A:R454, A:P491, A:L492, A:Q493, A:S494, A:Y495, A:G496, A:F497, A:Q498, A:P499, A:T500, A:V503, A:G504, A:Y505, A:Q506, A:P507, A:Y508, A:R509, A:V510, A:V511, A:V512, A:L513, A:S514, A:F515, A:E516, A:L517, A:L518, A:H519, A:A520, A:P521, A:A522, A:T523, A:V524, A:C525, A:G526, A:P527, A:K528	142	0.799
4	A:F559, A:L560, A:P561, A:F562, A:Q563	5	0.789
5	A:F79, A:D80, A:N81, A:P82, A:V83, A:L84, A:P85, A:I100, A:I101, A:R102, A:G103, A:W104, A:I105, A:T108, A:T109, A:L110, A:D111, A:S112, A:K113, A:T114, A:Q115, A:S116, A:L117, A:L118, A:I119, A:V120, A:N121, A:N122, A:A123, A:T124, A:N125, A:V126, A:V127, A:I128, A:K129, A:V130, A:C131, A:E132, A:F133, A:Q134, A:F135, A:C136, A:N137, A:D138, A:P139, A:F140, A:L141, A:G142, A:E156, A:F157, A:R158, A:V159, A:Y160, A:S161, A:S162, A:A163, A:N164, A:N165, A:C166, A:T167, A:F168, A:E169, A:Y170, A:V171, A:S172, A:Q173, A:P174, A:F175, A:L176, A:T236, A:R237, A:F238, A:Q239, A:T240, A:L241, A:L242, A:A243, A:L244, A:H245, A:R246	80	0.756

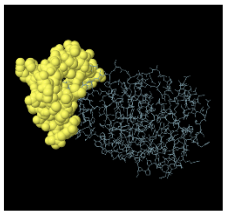
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Table12. Conformational epitopes from SARS-CoV-2 Nsp13 helicase

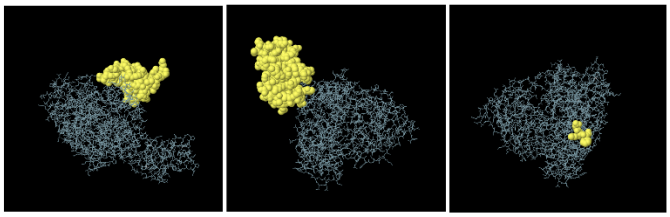
No.	Residues	Number of residues	Score
1	A:A1, A:V2, A:G3, A:A4, A:C5, A:L7, A:C8, A:N9, A:S10, A:Q11, A:T12, A:S13, A:L14, A:R15, A:C16, A:G17, A:F24, A:L25, A:C26, A:C27, A:K28, A:C29, A:C30, A:Y31, A:D32, A:V34, A:I35, A:S36, A:T37, A:S38, A:H39, A:K40, A:L41, A:V42, A:L43, A:S44, A:V45, A:N46, A:P47, A:Y48, A:V49, A:C50, A:N51, A:A52, A:P53, A:G54, A:C55, A:D56, A:V57, A:T58, A:D59, A:V60, A:T61, A:Q62, A:L63, A:Y64, A:L65, A:G66, A:G67, A:M68, A:S69, A:Y70, A:Y71, A:C72, A:K73, A:S74, A:H75, A:K76, A:P77, A:P78, A:I79, A:S80, A:F81, A:P82, A:L83, A:C84, A:A85, A:N86, A:G87, A:Q88, A:V89, A:F90, A:G91, A:L92, A:Y93, A:K94, A:N95, A:T96, A:C97, A:V98, A:G99, A:S100, A:D101, A:N102, A:V103, A:T104	96	0.761
2	A:D344, A:K345, A:F346	3	0.74
3	A:G150, A:I151, A:A152, A:T153, A:V154, A:R155, A:E156, A:V157, A:L158, A:S159, A:D160, A:R161, A:E162, A:L163, A:H164, A:L165, A:S166, A:W167, A:E168, A:V169, A:G170, A:K171, A:P172, A:R173, A:G184, A:Y185, A:R186, A:V187, A:T188, A:K189, A:N190, A:S191, A:K192, A:V193, A:Q194, A:I195, A:G203, A:D204, A:Y205, A:G206, A:D207, A:A208, A:V209, A:Y217, A:K218, A:L219, A:N220, A:V221, A:G222, A:D223, A:Y224, A:F225	52	0.738



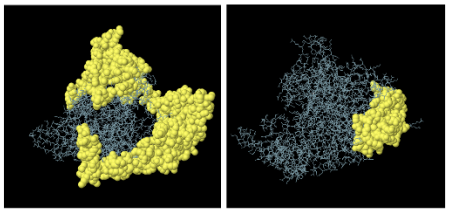
SARS-CoV-2 S protein



SARS-CoV-2 Mpro



SARS-CoV-2 Nsp12 RNA polymerase



SARS-CoV-2 Nsp13 helicase

Figure 1: 3D representation of discontinuous epitopes of SARS-CoV-2 Spike, Mpro, Nsp12 RNA polymerase and Nsp13 helicase

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Epitope prediction for (HTL) helper T Lymphocytes:

MHC class II epitope, which shows high binding affinity were predicted for human alleles HLA-DP, HLA-DQ and HLA-DR based on their IC50 values from Net MHC II 2.2 server. These epitopes were described as HTL epitopes. The epitope with similar sequences were overlapped to get single epitope**.** The total 21 number of high binding HTL epitope were selected for a spike, protease, RdRp and helicase for the novel multiepitope vaccine **(Table S1 – S4)**

Design and construction of final multiepitope vaccine:

The overlapped and high scoring CTL and HTL epitopes found from the SARS-CoV-2 S, Mpro, Nsp12 polymerase and Nsp13 helicase were combined to form the multiepitope vaccine construct (MVC). To increase the immune response, human β-defensin 2 (hβD-2) (PDB ID: 1FD3), and sequence of GIGDPVTCLKSGAICHPVFCPRRYKQIGTCGLPGTKCCKKP, and hBD-3 (PDB ID:1KJ6) and sequence of GIINTLQKYYCRVR GGRCAVLSCLPKEEQIGKCSTRGRKCCRRKK were selected as adjuvants at N and C-terminals sequence of the vaccine construct, respectively with linker EAAK [48,49]. After the adjuvant CTL epitopes were combined using appropriate AAY linkers, HTL epitopes were joined together with GPGPG linkers [50] as displayed in **Figure S5**. By combining potential CTLs, HTLs epitopes and adjuvants, a multiepitope vaccine construct of 1057 amino acids was constructed.

Parametric evaluation of physiochemical properties:

By estimating the multiepitope vaccine construct (MVC) using ProtParam server to estimate physicochemical properties [51] it was found that MVC weighed 114.6 KDa. The hypothetical protrusion index remained 8.15, displaying the basic nature of the MVC and assessed in vitro half-life was 30 hours in mammals' reticulocytes [52]. Assessed half-life indicates the time acquired by the protein to remain half of the quantity as originally producing in the cell. Instability index was also predicted to be 35.56 and classified the MVC as stable in nature. Aliphatic index [53] was also examined, which display the relative volume retained by the aliphatic side chain. It might be reflected as a positive variable for the extension of thermo-stability of globular proteins. The attained values of the aliphatic index were found to be 80.93; indicates at varied temperature the protein is thermostable. The Grand Average value of Hydropathy [54] signifies the summation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

of the hydropathy rate, along with sequence of amino acid it indicates the hydrophilic and hydrophobic nature of the protein. The found Grand Average value of Hydropathy for the vaccine protein was found to be 0.158.

Assessment of allergenicity and immunogenicity:

The designed subunit of the vaccine was assessed on the allergenic parameter through AllergenFP 1.0 server and AllerTOP 2.0 server. Non-allergenic nature was shown by both servers have shown for FMVC. The antigenicity connected to the vaccine subunit was projected through VaxiJen v2.0 servers. According to the outcome of vaxiJen, the antigenicity of vaccine was 0.4259 displaying it as a plausible antigen. Thus, the attained outcome from servers exhibited high possibility of subunit vaccine's antigenic and non-allergenic nature.

Structure prediction and validation of MVC

In order to analysis the 3D confirmation of MVC, the 1073 amino acid peptide sequence was utilized for the prediction of the 3D model. Multiple software were used for modeling, including RaptorX and I-TASSER, to avoid biases. Information of secondary structure result showed of 34% helical, 19% E, and 45% coiled assembly. For homology modeling, P-value is a good parameter to describe the relative quality of the model and lower P-value describes the quality is good for the modeled structure. P-value obtained for the MVC structure was 1.29e-06 that is lower and significant. For I-TASSER modeling, the model with the highest C-score was selected. The models generated from both servers were compared and validated through the MolProbity server [55]. The structure with better M-score (which combines the clash score, rotamer, and Ramachandran estimations into a single score) was utilized for extensive refined through MD simulations. The refinement included optimizing bond lengths and angles and removing clashes in geometry [56-58]. The Root Mean Square Deviation (RMSD) was calculated for 50ns. **Figure 2A** displayed the MD refined 3D-model of MVC and all-backbone RMSD trajectory are showed in **Figure 2B**. Initially, the RMSD trajectory of MVC model gradually expanding till 30ns and reached a value of ~9.5 Å. Later, the RMSD value remained to converge till 50ns with a deviation < 1 Å. This higher RMSD of the simulated model indicated protein expansion during simulation to attain more stable conformation. The averaged conformation of MVC model was extracted from the trajectory and compared with the initial model through Ramachandran evaluations. The MD optimized MVC

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

model showed that 86.2% (924/1073) of all residues were in Ramachandran favoured (>98%) regions, while the initial homology model showed only 72.49% (778/1073 residues) in Ramachandran favoured regions. Moreover, residues placed in Ramachandran allowed region (>99.8%) increased from 83.8 (901/1073 residues) with 172 (16.02 %) outliers to 95.5% (1024/1073) with 49 outliers 4.1%) (**Figure 2C**).

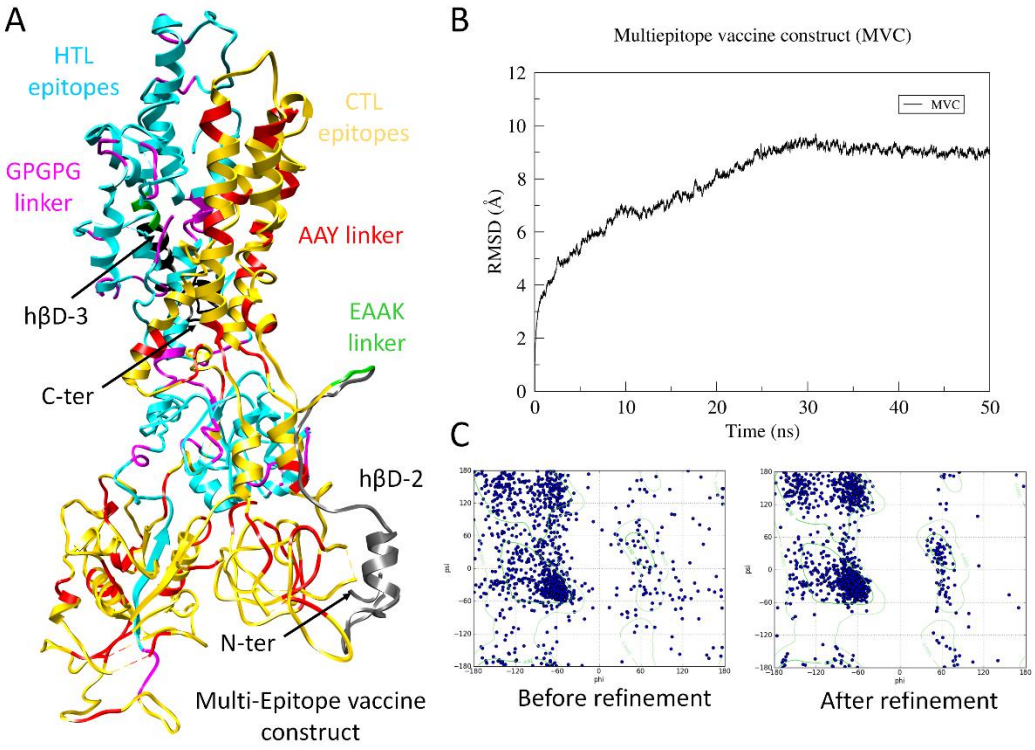


Figure 2: Molecular modeling of vaccine construct. **(A)** Structural representation of multiepitope vaccine construct (MVC) is displayed with regions (HTL, CTP epitopes, linkers and adjuvants) are highlighted accordingly. **(B)** Root-mean-square-deviation trajectory of MVC analyzed over a period of 50 ns MD simulations. **(C)** Ramachandhran evaluations of MVC before and after refinement through MD simulations.

Disulfide engineering for vaccine stability

Disulfide engineering was done to stabilize the modelled structure of FVC, by disulfide by design v2.0 server [59]. Evaluation based on other parameters like Chi3 and energy value, only 07 residues pairs were selected as their value came under the permissible range, i.e. energy value must be smaller than 2.2 and Chi3 must be between −87 & +97 degrees. Hence, total of eight mutations

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

were formed at the pairs of residues named VAL6-ALA157, TYR138-ALA163, VAL360-GLY730, LEU462-TYR474, ALA499-ARG519, SER814-GLY923, GLY816-SER927 and THR934-GLY946.

Molecular docking of vaccine constructs with TLR4

Molecular docking is the best *in silico* approach to finding out interactions between protein-protein and protein-ligand complexes [60-63]. Molecular docking of MVC with TLR-4 and TLR3 receptors was performed using the ClusPro 2.0, and 30 models were produced [64]. Among these, the complex with the lowest energy was selected. Energy score attained for TLR3 and TLR4 were -1327.2 and -1270.2 respectively and subjected to MD simulations to analyze complex stability.

Molecular dynamics simulation for TLRs/MVC complex

The stability of the TLRs/MVC docked complexes was further investigated by performing MD simulation for a period of 50 ns in an explicit solvent environment at 300 K. The potential energy of the simulation system was also found stable throughout simulation period (data not provided). The MD refined MVC was utilized for docking, and both complexes showed relatively stable RMSD as compared to MVC alone **(Figure 2)**. In the beginning, the MVC experienced small fluctuations but remained interacted with the hydrophobic groove of TLR4 and TLR3 and showed consistent stability in last ~25 ns **(Figure 3)**.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

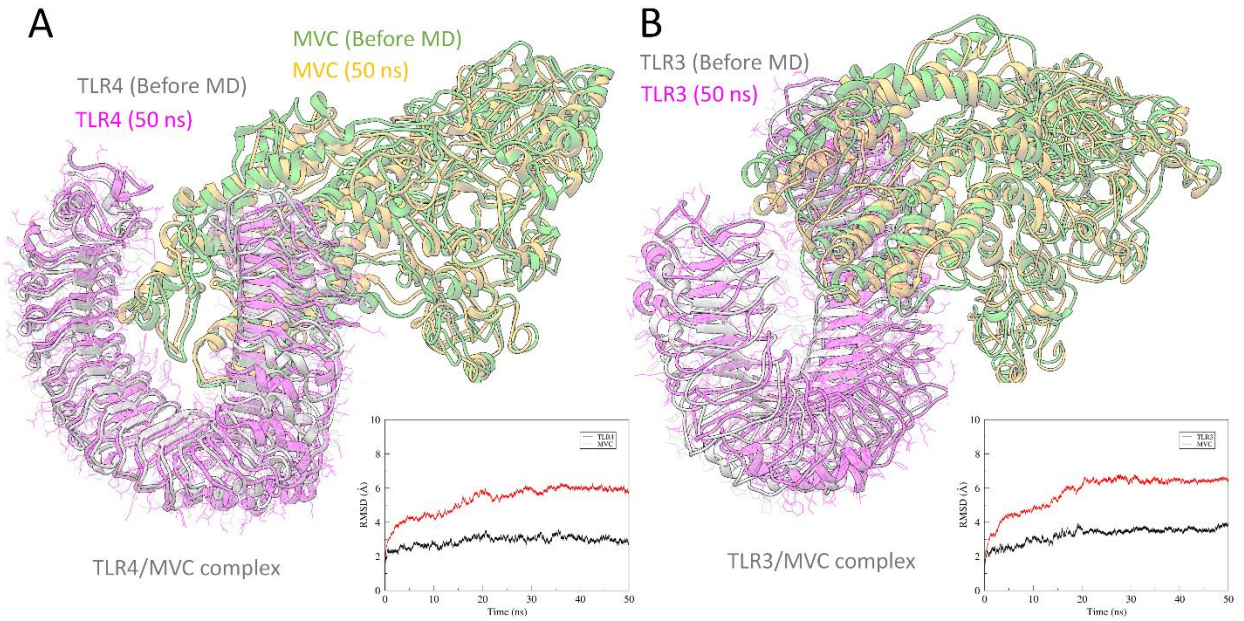


Figure 3: Toll-like Receptor (TLR) complexed with a multiepitope vaccine construct (MVC). **(A)** Conformation of **(A)** TLR4/MVC and **(B)** TLR3/MVC complex before and after 50 ns MD simulations together with RMSD plot at the bottom indicated all-atom backbone deviation of TLR (in black) and MVC (in red).

Codon adaptation and *In silico* cloning of the MVC

The reverse translation and codon optimization were performed for the sequence of MVC by the online JCat server [65]. The GC content and codon adaptation index (CAI) were determined out as output from the server. The GC content obtained for MVC was 54.39% which lies in the acceptable range i.e., from 30% to 70%. While, CAI value of MVC was 1, which indicates high level of expression in K12 strain of the *E. coli*. Later, the restrictions sites of *NdeI* and *XhoI* were generated and the MVC sequence was cloned in pET28a (+) vector (**Figure 4**). The MVC sequence is represented in yellow with the restriction sites. The sequence of multiepitope vaccine construct has been cloned between 6-histidine residue on both sites. Hence it will help in purification of MVC.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

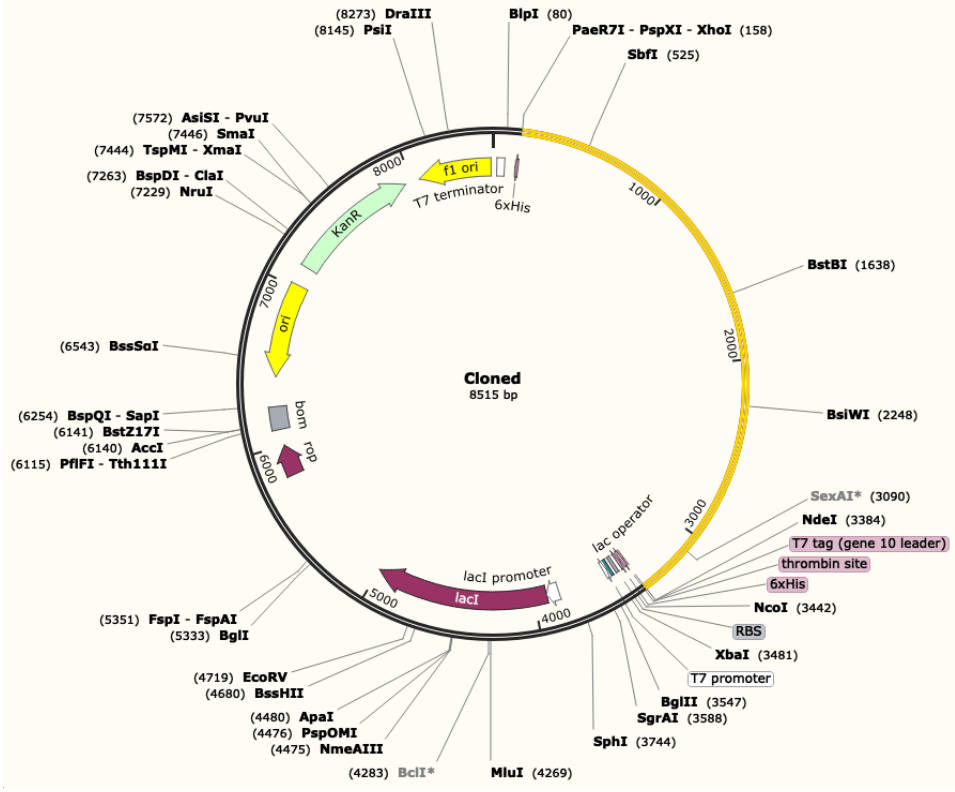


Figure 4: ***In silico*** cloning of the multiepitope vaccine construct (MVC). The cDNA of the MVC (yellow) was inserted at the upstream of the T7 promoter.

DISCUSSION:

The announcement of emergency by world health organization (WHO) on COVID-19 outbreak urged the researchers to develop a therapeutics, mainly identification of drug candidates or vaccines [20]. The use of cost effective and less time-consuming methods especially immunoinformatics approaches have already assisted the researchers to predict potential antigenic epitope for the multiepitope based vaccine [32-34,38,66-69]. Multiple epitope vaccine has a distinctive design concept, compared to classical single-epitope based vaccines [68,70-72]. The concept behind the scanning of the viral genome to find immunogenic epitopes leads to an elicited immune response without any reversal of viral pathogenesis [73].

To design a multiepitope vaccine, the research focused on the identification of epitopes for potential B and T cells using the immune-informatics approach. *In silico* method can be employed using patho-genomics analysis on the genome on a vast scale to identify new vaccines

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

[73,74]. Various limitations are there in the context of appropriate candidate antigens, their immunodominant epitopes and experimental methods, which include the development of an effective delivery system[75,76]. Investigation of the whole spectrum of probable antigens is achievable through immunoinformatics and with the aid of molecular modelling to analyze the potential binding with host proteins [27,29,32,35,76]. Besides, the difficulty of culturing the pathogens as well as *in vitro* antigen expression problems can be avoided [69,77]. Although some multiple epitope vaccines showed *in vivo* efficacy with promising protective immunity [39,40,42,70], while some have entered in phase-I clinical trials including, H2NVAC in patients with HER2-expressing ductal carcinoma in situ (DCIS) (NCT03793829), E1602 for patients with Metastatic Melanoma [78], EMD640744 in patients with advanced solid tumors [71], and TAB9 in non-HIV-1 infected human volunteers [79]. However, the designing effective multi-epitope vaccine remains a great challenge. Hence, estimation of B cell and CTL cell epitopes by different immune-informatics methods considered to be a vital tool for designing multi-epitope construct.

In the present research, potential T-cells and B-cell epitopes (discontinuous & continuous) have been recognized from SARS-CoV-2 main protease, Nsp12 RNA polymerase, Spike and Nsp13 helicase proteins to design multi-epitope construct (MVC) by using adjuvants (hβ defensins) and appropriate linkers. The employed linkers (GPGPG and AAY) were carefully selected because their length, composition and structure may affect the activity of the domains and overall characteristics of the molecule [80]. For example, as being somewhat basic antigenic domains (isoelectric point pI > 8), a linker that contains more basic amino acids may increase the pI, such as KK [81]. Therefore, basic linker were avoided and glycine-rich linker, i.e., GPGPG was chosen for joining potential epitopes which usually improve solubility and allow the adjoining domains to be accessible and act freely [82]. Following this, a reliable MVC model was generated through molecular modeling and optimized accordingly. All-atom backbone stability of MVC was analyzed through molecular dynamics simulation over a period of 50 ns because the optimal structural stability of MVC is considered a prime aspect in its efficacy [83] and to trigger immune response by interacting Toll-like receptors (TLRs) signaling as successful immunization results is accomplished through stimulation of the TLRs [43]. The resulted model showed fewer outliers, while rotamers were adjusted during simulation. Molecular docking with TLR3 and TLR4 followed by 50 ns MD simulation revealed stability in the overall complex in last ~20 ns. The designed MVC

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

interacted with TLR3 and TLR4 directly and their molecular interactions were strengthened during MD simulation which led to reduce backbone RMSD fluctuation in both TLR/MVC complexes (**Figure 5**). However, the epitopes were estimated as non-allergenic, showed antigenicity and predicted cloning in vector pET28a (+) of E. coli, but given the limitation of *in silico* tools, the expression and efficacy of the designed multiple vaccine construct should be further proven through *in vitro* and *in vivo* experiments.

Materials and Methods

Coronavirus protein sequences and structural information

The primary amino acid sequences of SARS-CoV-2 main protease (Mpro) (306 amino acids), Nsp12 RNA dependent RNA polymerase (932 amino acids) , spike (1237 amino acids) and Nsp13 helicase (601 amino acids) proteins were retrieved from GenBank ID: AHZ13508.1.. For structural studies, the crystal structures of recently deposited SARS-CoV-2 Mpro (PDB ID: 6LU7) and Spike (PDB ID: 6VYB) protein were obtained from PDB, while the homology models of SARS-CoV-2 Nsp12 RNA polymerase and Nsp13 helicase were obtained from our recent study [84]. These homology models were generated from templates that showed 99.83% and 96.08% identities with SARS-CoV Nsp12 (PDB ID: 6NUR) [18], Nsp13 (PDB ID: 6JYT) [17]. These models showed strikingly similar domain architecture with SARS-CoV and found reliable enough to use in epitope identification studies.

Prediction of Linear and Conformational B-Cell epitopes

The interaction between the Antigen B-cell epitope and B-lymphocyte causes the B-lymphocytes to differentiate into memory cells and antibody-secreting plasma [85]. B-cell epitope has two significant features such as, accessibility to the flexible region and hydrophilic nature of an immunogen[86]. As per prediction of Parker hydrophilicity, for surface accessibility, Emini prediction [87], antigenicity scale for Kolaskar and Tongaonkar [44], flexibility prediction for Karplus and Schulz [88], the analysis was employed arithmetically at IEDB (http://www.iedb.org/). Discontinuous (conformational) epitopes prediction for B-cell was predicted by using Ellipro from IEDB (http:// tools.immuneepitope.org/toolsElliPro/) [47] which used three diverse algorithms such

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

as residues' Protrusion Index (PI) [89], adjoining clustering residues liable upon PI and approximation of protein shape [90].

Prediction of Potential CTL epitopes

NetCTL.1.2 server (http://www.cbs.dtu.dk/services/Net CTL) was used to predicted the CTL epitopes[46], The eliciting of CTLs happen on the surface of antigen-presenting MHC molecules. To assimilate the MHC class I binding, efficiency of TAP transport and the cleavage of proteasomal C-terminal NetCTL 1.2 server was employed. HLA (Human Leukocyte Antigen) alleles and peptide lengths both were selected and submitted. T-cell epitopes were predicted as an output. For predicting the TAP transport efficiency weight matrix was utilized while for cleavage of proteasomal C-terminal and MHC class I binding, the ANN (Artificial Neural Network) was employed.

Epitope prediction of helper T-cell

For the prediction of epitope of helper T-cell NetMHCII 2.2 Server was used which give 15-mer epitope for human alleles. NetMHCII 2.2 Server uses artificial neuron network for the prediction of a peptide with human alleles i.e HLA-DP, HLA-DR and HLA-DQ [91]. Based on receptor interaction, MHC II epitopes were predicted and deduced from IC50 values as well as they assigned percentile ranks. Those peptides which show strong interaction have IC50 value <50 nM while those having intermediate and low affinity have IC50 <500 and IC50 <5000 respectively. It is, therefore, the percentile rank has direct relation with IC50 and invers to the affinity for epitope.

Multiepitope vaccine designing

The MVC was designed by connecting the peptide sequences in a successive manner with the help of suitable linkers. The occurrence of overlapping residues amid the B-cell (BCL), HTL and CTL epitopes were unwavering and epitopes with overlapping regions were used for multiepitope vaccine design. In recent, mammals8' β-defensin was documented to have a possible role to confer HIV infection as a mucosal adjuvant; consequently, due to its adjuvant characteristics against viral infection, therefore it was chosen and added to the at N and C-terminals

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

sequence of the vaccine construct[33]. Adjuvant were joined with epitopes at N and C-terminal by using the EAAAK linker, whereas intra-CTL epitopes were joined together by using the AAY linker. After CTL epitope, HTL-epitopes were added next to the CTL epitope using the GPGPG linkers.

Antigenicity and allergenicity estimation of the MVC

To be an effective and safe vaccine candidate, the vaccine candidate should be nonallergic with minimum off-targets effects. The nonallergenic and allergenic behavior of the VC were assessed by 2 servers AllerTOP V2.0 (http://www.ddg-pharmfac.net/AllerTOP/) and AlgPred (http://www.imtech.res.in/raghava/algpred/) [92].Out of these, the later categorizes the protein sequence (input) by a k-nearest neighbour algorithm (kNN; k = 3) on the basis of the training set comprising 2210 already known allergens from diverse species and nonallergens (2210) from the similar species. The former assimilates the SVM module for the prediction of the allergenic nature of protein with high accuracy. The MAST/MEME allergen motif was examined with the help of MAST, and the nature's allergenic nature was allocated if an identical motif was determined.

Evaluation of antigenicity of the FVC was done using 2 freely available servers VaxiJen v2.0 (http://www.ddg-pharmfac.net/vaxijen/ VaxiJen/VaxiJen.html) [93] and ANTIGENpro (http://scratch.proteomics.ics.uci.edu/), The latter categorizes the antigen-based only on the physio-chemical characteristics of the input protein sequence instead sequence placement algorithm. The correctness of the server is relatively high and differs between 70% & 89% based on the target organism. While the former envisages the entire protein antigenicity based on results obtained by the protein microarray data analysis. It predicts the antigenicity being independent of the pathogen, but this approach is sequence-based.

Physiochemical parameters evaluation

The vaccine construct was employed to ProtParam (http://web.expasy.org/protparam/) [51] to examine its physio-chemical properties. The criteria on which the sequences of multiepitope vaccine were examined are theoretic PI, half-life, instability index, aliphatic index, grand average and stability profiling of Hydropathy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Tertiary structure prediction and refinement of MVC

The final vaccine construct was submitted structure prediction server is known as RaptorX (http://raptorx.uchicago.edu/StructurePrediction/ predict/) [94] and I-TASSER [95]. It is an exceptional server for protein 3D structures predictions on *ab initio* method and able to generate from template lie in twilight zone (<30%). It utilizes an exclusive nonlinear context-specific alignment and prospective consistency algorithm. The generated models were evaluated through Molprobity for all-atom contacts and geometry [55]. The model was selected for further refinement through molecular dynamics simulations that showed reliable Ramachandran evaluations.

Stability enhancement of MVC by disulfide engineering

Before moving towards the docking protocol, it was essential to enhance the stability of the model through disulfide engineering. It is a novel concept to introducing disulfide bonds to the modelled protein structure. Consequently, the multiepitope model was employed to the Disulfide by Design 2.0 server [59] to achieve disulfide engineering. The protein model was uploaded to identify the residue pair which can be utilized for disulfide engineering. To create the disulfide bonds, four residues were selected to mutated them with cysteine residue by the Disulfide using Design 2.0 server.

Molecular docking of vaccine constructs with TLR4:

In order to analyze the binary interaction of MVC with TLRs, protein-protein docking was performed using Cluspro. The PDB structures of TLR-4 (PDB: 4G8A) and TLR-3 (PDB: 1ZIW) were retrieved from PDB. The multiepitope vaccine model was used as a ligand. Both proteins were prepared accordingly, by removing heteroatoms, and addition of hydrogens and charges. ClusPro is an automated protein-protein docking server which generates RMSD-based clustering of 1000 docked conformations. The most representative docked conformation from the largest cluster were used for further structural analyses.

Molecular dynamics simulation for TLRs/MVC complex:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 March 2020 doi:10.20944/preprints202003.0242.v1

Molecular dynamics simulations were performed in two steps: (1) a 50 ns MD simulation was performed to optimize and refine 3D model of multiepitope vaccine construct before docking, (2) a second 50 ns MD simulations to examine the backbone stability of TLRs/MVC complexes. All simulations were performed by AMBER simulation package 18 using the same protocol as described in previous immunoinformatics studies [27,29]. Briefly, stepwise minimization and equilibration procedure, the solvated system in explicit water molecules (TIP3P) was submitted to a production run at standard temperature (300 K) and pressure (1 bar). The trajectories were collected after every 2ps for a complete production run, and CPPTRAJ module was utilized to analyze trajectories. The MD simulation complexes were analyzed using Chimera 1.14.

Codon adaptation and *in silico* cloning

The sequence of multiepitope vaccine construct was employed to the online server JCat for reverse translation, and cDNA was obtained, which was submitted for codon optimization [65]. The cDNA was evaluated by codon optimization according to the Codon Adaptation Index (CAI) and GC content of the sequence. The acceptable range of the GC content is 30% - 70% and the value of CAI varies from 0 – 1. The higher the value of CAI indicate a higher level of gene expression [67]. The maximum value of CAI is 1 and considered ideal, whereas the value of more than 0.8 is also acceptable. After this step, the adapted and optimized sequence of the nucleotides consistent to the design of multiepitope vaccine construct was cloned by using the restriction cloning module of SnapGene toll in the vector pET28a (+) of E. coli.

Contributions

H.M.R and M.U.M. conceived and designed the experiments. H.M.R, M.S, M.S.A performed the immunoinformatic and molecular docking analyses. M.U.M performed the molecular dynamics simulations analysis. M.U.M prepared the Figures. M.S. supervised the study. H.M.R and M.U.M. wrote the main manuscript text. M.F, S.A, R.G, M.A.B and M.U.M critically reviewed the manuscript. All the authors reviewed and approved the final manuscript.

Competing interests

The author(s) declare no competing interests.

REFERENCES:

1 Parry, J. (British Medical Journal Publishing Group, 2020).

2 Benvenuto, D. *et al.* The 2019-new Coronavirus epidemic: evidence for virus evolution. (2020).

3 Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. J. T. L. A novel coronavirus outbreak of global health concern. (2020).

4 Perlman, S. Another Decade, Another Coronavirus. doi:10.1056/NEJMe2001126 (2020).

5 WHO. Situation report - 52 (https://www.who.int/docs/default-source/coronaviruse/20200312-sitrep-52-covid-19.pdf?sfvrsn=e2bfc9c0_2). (2020).

6 Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. **348**, 1967-1976 (2003).

7 Azhar, E. I., Hui, D. S., Memish, Z. A., Drosten, C. & Zumla, A. J. I. D. C. The Middle East Respiratory Syndrome (MERS). **33**, 891-905 (2019).

8 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. (2020).

9 Chen, Y., Liu, Q. & Guo, D. J. J. o. M. V. Coronaviruses: genome structure, replication, and pathogenesis. (2020).

10 Chan, J. F.-W. *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. **9**, 221-236 (2020).

11 Buchholz, U. J. *et al.* Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. **101**, 9804-9809 (2004).

12 Li, T. *et al.* Long-term persistence of robust antibody and cytotoxic T cell responses in recovered patients infected with SARS coronavirus. **1** (2006).

13 Zakhartchouk, A. N. *et al.* Immunogenicity of a receptor-binding domain of SARS coronavirus spike protein in mice: implications for a subunit vaccine. **25**, 136-143 (2007).

14 Zhi, Y. *et al.* Identification of murine CD8 T cell epitopes in codon-optimized SARS-associated coronavirus spike protein. **335**, 34-45 (2005).

15 Yang, H. *et al.* Design of wide-spectrum inhibitors targeting coronavirus main proteases. **3** (2005).

16 Bacha, U., Barrila, J., Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. J. B. Identification of novel inhibitors of the SARS coronavirus main protease 3CLpro. **43**, 4906-4912 (2004).

17 Jia, Z. *et al.* Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. **47**, 6538-6550 (2019).

18 Kirchdoerfer, R. N. & Ward, A. B. J. N. c. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. **10**, 1-9 (2019).

19 Xu, J. *et al.* Systematic Comparison of Two Animal-to-Human Transmitted Human Coronaviruses: SARS-CoV-2 and SARS-CoV. **12**, 244 (2020).

20 Morse, J. S., Lalonde, T., Xu, S. & Liu, W. Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV. (2020).

21 Li, G. & De Clercq, E. (Nature Publishing Group, 2020).

22 Shang, W., Yang, Y., Rao, Y. & Rao, X. J. n. V. The outbreak of SARS-CoV-2 pneumonia calls for viral vaccines. **5**, 1-3 (2020).
23 Verma, S., Sajid, A., Singh, Y., Shukla, P. J. H. v. & immunotherapeutics. Computational tools for modern vaccine development. (2019).
24 Arnon, R. & Ben-Yedidia, T. J. I. i. Old and new vaccine approaches. **3**, 1195-1204 (2003).
25 Hughes, J. P., Rees, S., Kalindjian, S. B. & Philpott, K. L. J. B. j. o. p. Principles of early drug discovery. **162**, 1239-1249 (2011).
26 Damfo, S. A., Reche, P., Gatherer, D., Flower, D. R. J. J. o. M. G. & Modelling. In silico design of knowledge-based Plasmodium falciparum epitope ensemble vaccines. **78**, 195-205 (2017).
27 Jabbar, B. *et al.* Antigenic Peptide Prediction From E6 and E7 Oncoproteins of HPV Types 16 and 18 for Therapeutic Vaccine Design Using Immunoinformatics and MD Simulation Analysis. **9**, 3000 (2018).
28 Michel-Todó, L. *et al.* In silico design of an epitope-based vaccine ensemble for Chagas disease. **10**, 2698 (2019).
29 Mirza, M. U. *et al.* Towards peptide vaccines against Zika virus: Immunoinformatics combined with molecular dynamics simulations to predict antigenic epitopes of Zika viral proteins. **6**, 37313 (2016).
30 Sabetian, S. *et al.* Exploring dengue proteome to design an effective epitope-based vaccine against dengue virus. **37**, 2546-2563 (2019).
31 Ul Qamar, M. T. *et al.* Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. **17**, 362 (2019).
32 de Oliveira Tosta, S. F. *et al.* Multi-epitope based vaccine against Yellow fever virus applying immunoinformatics approaches. 1-28 (2019).
33 Srivastava, S. *et al.* Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. **37**, 4345-4360 (2019).
34 Goodman, A. G. *et al.* A human multi-epitope recombinant vaccinia virus as a universal T cell vaccine candidate against influenza virus. **6** (2011).
35 Shahid, F., Ashfaq, U. A., Javaid, A., Khalid, H. J. I., Genetics & Evolution. Immunoinformatics guided rational design of a next generation multi epitope based peptide (MEBP) vaccine by exploring Zika virus proteome. **80**, 104199 (2020).
36 Nosrati, M., Behbahani, M. & Mohabatkar, H. J. J. o. b. i. Towards the first multi-epitope recombinant vaccine against Crimean-Congo hemorrhagic fever virus: A computer-aided vaccine design approach. **93**, 103160 (2019).
37 Khatoon, N. *et al.* Exploratory algorithm to devise multi-epitope subunit vaccine by investigating Leishmania donovani membrane proteins. **37**, 2381-2393 (2019).
38 Dorosti, H. *et al.* Vaccinomics approach for developing multi-epitope peptide pneumococcal vaccine. **37**, 3524-3535 (2019).
39 Guo, L. *et al.* Immunological features and efficacy of a multi-epitope vaccine CTB-UE against H. pylori in BALB/c mice model. **98**, 3495-3507 (2014).
40 Zhou, W.-Y. *et al.* Therapeutic efficacy of a multi-epitope vaccine against Helicobacter pylori infection in BALB/c mice model. **27**, 5013-5019 (2009).

41 Depla, E. *et al.* Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. **82**, 435-450 (2008).

42 Cao, Y. *et al.* Rational design and efficacy of a multi-epitope recombinant protein vaccine against foot-and-mouth disease virus serotype A in pigs. **140**, 133-141 (2017).

43 Van Duin, D., Medzhitov, R. & Shaw, A. C. J. T. i. i. Triggering TLR signaling in vaccination. **27**, 49-55 (2006).

44 Kolaskar, A. & Tongaonkar, P. C. J. F. l. A semi-empirical method for prediction of antigenic determinants on protein antigens. **276**, 172-174 (1990).

45 Alberts, B. *et al.* Molecular biology of the cell, 6th edn New York. (2002).

46 Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. **8**, 424 (2007).

47 Ponomarenko, J. *et al.* ElliPro: a new structure-based tool for the prediction of antibody epitopes. **9**, 514 (2008).

48 Nielsen, M., Lundegaard, C. & Lund, O. J. B. b. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. **8**, 238 (2007).

49 Arai, R., Ueda, H., Kitayama, A., Kamiya, N. & Nagamune, T. J. P. e. Design of the linkers which effectively separate domains of a bifunctional fusion protein. **14**, 529-532 (2001).

50 Bergmann, C. C., Yao, Q., Ho, C.-K. & Buckwold, S. L. J. T. J. o. I. Flanking residues alter antigenicity and immunogenicity of multi-unit CTL epitopes. **157**, 3242-3249 (1996).

51 Gasteiger, E. *et al.* in *The proteomics protocols handbook* 571-607 (Springer, 2005).

52 Bachmair, A., Finley, D. & Varshavsky, A. J. S. In vivo half-life of a protein is a function of its amino-terminal residue. **234**, 179-186 (1986).

53 Ikai, A. J. T. J. o. B. Thermostability and aliphatic index of globular proteins. **88**, 1895-1898 (1980).

54 Kyte, J. & Doolittle, R. F. J. J. o. m. b. A simple method for displaying the hydropathic character of a protein. **157**, 105-132 (1982).

55 Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. **35**, W375-W383 (2007).

56 Chen, J., Brooks III, C. L. J. P. S., Function, & Bioinformatics. Can molecular dynamics simulations provide high-resolution refinement of protein structure? **67**, 922-930 (2007).

57 Mirjalili, V., Noyes, K., Feig, M. J. P. S., Function, & Bioinformatics. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. **82**, 196-207 (2014).

58 Raval, A. *et al.* Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. **80**, 2071-2079 (2012).

59 Craig, D. B. & Dombkowski, A. A. J. B. b. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. **14**, 346 (2013).

60 Ikram, N. *et al.* Inhibition of Oncogenic Kinases: An In Vitro Validated Computational Approach Identified Potential Multi-Target Anticancer Compounds. **9**, 124 (2019).

61 Iman, K. *et al.* In silico Structure-based Identification of Novel Acetylcholinesterase Inhibitors Against Alzheimer's Disease. *CNS & Neurological Disorders-Drug Targets* **17**, 54-68 (2018).

62 Mirza, M. U. & Ikram, N. J. I. j. o. m. s. Integrated computational approach for virtual hit identification against ebola viral proteins VP35 and VP40. **17**, 1748 (2016).

63 Mirza, M. U. *et al.* In silico structural elucidation of RNA-dependent RNA polymerase towards the identification of potential Crimean-Congo Hemorrhagic Fever Virus inhibitors. **9**, 1-18 (2019).

64 Comeau, S. R., Gatchell, D. W., Vajda, S. & Camacho, C. J. J. N. a. r. ClusPro: a fully automated algorithm for protein–protein docking. **32**, W96-W99 (2004).

65 Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. **33**, W526-W531 (2005).

66 Livingston, B. *et al.* A rational strategy to design multiepitope immunogens based on multiple Th lymphocyte epitopes. **168**, 5499-5506 (2002).

67 Nezafat, N., Eslami, M., Negahdaripour, M., Rahbar, M. R. & Ghasemi, Y. J. M. B. Designing an efficient multi-epitope oral vaccine against Helicobacter pylori using immunoinformatics and structural vaccinology approaches. **13**, 699-713 (2017).

68 Saadi, M., Karkhah, A., Nouri, H. R. J. I., Genetics & Evolution. Development of a multi-epitope peptide vaccine inducing robust T cell responses against brucellosis using immunoinformatics based approaches. **51**, 227-234 (2017).

69 Mahmoodi, S. *et al.* Harnessing bioinformatics for designing a novel multiepitope peptide vaccine against breast cancer. **17**, 1100-1114 (2016).

70 Jiang, P. *et al.* Evaluation of tandem Chlamydia trachomatis MOMP multi-epitopes vaccine in BALB/c mice model. **35**, 3096-3103 (2017).

71 Lennerz, V. *et al.* Immunologic response to the survivin-derived multi-epitope vaccine EMD640744 in patients with advanced solid tumors. **63**, 381-394 (2014).

72 Zhu, S. *et al.* Hepatitis B virus surface antigen as delivery vector can enhance Chlamydia trachomatis MOMP multi-epitope immune response in mice. **98**, 4107-4117 (2014).

73 Zhang, L. J. C. & immunology, m. Multi-epitope vaccines: a promising strategy against tumors and viral infections. **15**, 182-184 (2018).

74 Rappuoli, R. J. C. o. i. m. Reverse vaccinology. **3**, 445-450 (2000).

75 Channappanavar, R. & Perlman, S. in *MERS Coronavirus: Methods and Protocols* (ed Rahul Vijay) 195-204 (Springer US, 2020).

76 Yin, D. *et al.* A novel multi-epitope recombined protein for diagnosis of human brucellosis. **16**, 219 (2016).

77 Davies, M. N. & Flower, D. R. J. D. d. t. Harnessing bioinformatics to discover new vaccines. **12**, 389-395 (2007).

78 Slingluff, C. L. *et al.* A randomized phase II trial of multiepitope vaccination with melanoma peptides for cytotoxic T cells and helper T cells for patients with metastatic melanoma (E1602). **19**, 4228-4238 (2013).

79 Toledo, H. *et al.* A phase I clinical trial of a multi-epitope polypeptide TAB9 combined with Montanide ISA 720 adjuvant in non-HIV-1 infected human volunteers. **19**, 4328-4336 (2001).

80 Yu, K., Liu, C., Kim, B.-G. & Lee, D.-Y. J. B. a. Synthetic fusion protein design and applications. **33**, 155-164 (2015).

81 Yano, A. *et al.* An ingenious design for peptide vaccines. **23**, 2322-2326 (2005).

82 Kavoosi, M., Creagh, A. L., Kilburn, D. G., Haynes, C. A. J. B. & bioengineering. Strategy for selecting and characterizing linker peptides for CBM9-tagged fusion proteins expressed in Escherichia coli. **98**, 599-610 (2007).

 doi:10.20944/preprints202003.0242.v1

83 Scheiblhofer, S., Laimer, J., Machado, Y., Weiss, R. & Thalhamer, J. J. E. r. o. v. Influence of protein fold stability on immunogenicity and its implications for vaccine design. **16**, 479-489 (2017).

84 Mirza, M. U. & Froeyen, M. Structural Elucidation of SARS-CoV-2 Vital Proteins: Computational Methods Reveal Potential Drug Candidates Against Main Protease, Nsp12 RNA-dependent RNA Polymerase and Nsp13 Helicase. (2020).

85 Nair, D. T. *et al.* Epitope recognition by diverse antibodies suggests conformational convergence in an antibody response. **168**, 2371-2382 (2002).

86 Fieser, T. M., Tainer, J. A., Geysen, H. M., Houghten, R. A. & Lerner, R. A. J. P. o. t. N. A. o. S. Influence of protein flexibility and peptide conformation on reactivity of monoclonal anti-peptide antibodies with a protein alpha-helix. **84**, 8568-8572 (1987).

87 Emini, E. A., Hughes, J. V., Perlow, D. & Boger, J. J. J. o. v. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. **55**, 836-839 (1985).

88 Karplus, P. & Schulz, G. J. N. Prediction of chain flexibility in proteins. **72**, 212-213 (1985).

89 Thornton, J., Edwards, M., Taylor, W. & Barlow, D. J. T. E. j. Location of 'continuous' antigenic determinants in the protruding regions of proteins. **5**, 409-413 (1986).

90 Taylor, W., Thornton, J. t. & Turnell, W. J. J. o. M. G. An ellipsoidal approximation of protein shape. **1**, 30-38 (1983).

91 Nielsen, M. & Lund, O. J. B. b. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. **10**, 296 (2009).

92 Saha, S. & Raghava, G. J. N. a. r. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. **34**, W202-W209 (2006).

93 Doytchinova, I. A. & Flower, D. R. J. B. b. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. **8**, 4 (2007).

94 Källberg, M. *et al.* Template-based protein structure modeling using the RaptorX web server. **7**, 1511 (2012).

95 Zhang, Y. J. B. b. I-TASSER server for protein 3D structure prediction. **9**, 40 (2008).