Scrutinizing the SARS-CoV-2 protein information for designing an effective vaccine encompassing both the T-cell and B-cell epitopes

Neha Jain[¶], Uma Shankar[¶], Prativa Majee[¶], Amit Kumar*

Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Indore, Simrol, Indore, 453552, India

[¶] These authors contributed equally to this work

*To whom correspondence should be addressed.

Amit Kumar, Discipline of Biosciences and Biomedical Engineering, Indian Institute of Technology-Indore, Indore, Simrol - 453 552, India, Contact - +91-732-430-6771, Email: amitk@iiti.ac.in

SUPPLEMENTARY METHOD

Retrieval of SARS-CoV-2 proteome

The complete proteome of latest reported novel Wuhan strain of SARS Coronavirus (SARS-CoV-2) was downloaded from the Nucleotide database available at National Center for Biotechnology Information (NCBI).

Antigenicity prediction in the Coronavirus proteome

For the antigenic analysis of the proteome of COVID-19 strain, VaxiJen v2.0 server available at http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html was used[1]. For antigenicity prediction of the proteins of SARS-CoV-2 with higher accuracy, Virus model available at the VaxiJen server with a threshold of 0.4 was utilized.

Selection of MHC I and MHC II alleles

MHC class I and II alleles were selected on the basis of their occurrence worldwide. We focused specially on the countries which are severely affected by the deadly SARS-CoV-2 strain. For MHC class I, following 14 HLA alleles A01:01, A02:01, A02:03, A02:06, A03:01, A11:01, A24:02, A31:01, A33:01, B07:02, B35:01, B51:01, B54:01, B58:01 were used. For MHC-II, DRB1_0101, DRB1_0301, DRB1_0401, DRB1_0405, DRB1_0701, DRB1_0802, DRB1_0901, DRB1_1101, DRB1_1201, DRB1_1302, DRB1_1501, DPA10103-DPB10101, DPA10201-DPB10201, DPA10301-DPB10301, DQA10101-DQB10201, DQA10301-DQB10301, DQA10401-DQB10401, DQA10501-DQB10501, DQA10102-DQB10202, DQA10402-DQB10402 were used for the epitope screening.

Helper T cell epitope prediction

Various online servers were explored for the accurate prediction of Helper T Lymphocytes (HTL) epitopes. Firstly, the antigenic proteins were analysed by using NetMHCIIPanv3.2 server (http://www.cbs.dtu.dk/services/NetMHCIIpan/) [2]. The server provides the epitope predictions for three human MHC class II isotypes that includes HLA-DR, HLA-DP and HLA-DQ. We explored NetMHCIIPan server for HTL epitope predictions for the selected 20 MHC class II alleles. The length for the HTL epitopes was kept 15 mer. The server predicts the binding affinity

of the epitopes with the respective HLA allele and provides an IC_{50} (in nanoMolar) and %Rank for each MHC class II allele – epitope pair which was utilized for initial screening of the potential HTL epitopes.

The predicted strong binders were further analysed by MHCII binding prediction tool available at IEDB server (http://tools.immuneepitope.org/mhcii/)[3]..

Cytotoxic T cell epitope prediction

For Cytotoxic T cell (CTL) epitope prediction, NetMHCPan 4.0 [4]based on artificial neural network (ANN) was utilized for the 14 selected HLA class I molecules. The predicted strong binders were then checked for their antigenicity by using VaxiJen server. Further, immunogenicity of the epitopes was checked using class I Immunogenicity tool available at IEDB server (http://tools.iedb.org/immunogenicity/)[5].

Epitope Conservation analysis

The presence of selected best epitopes in all the reported human infecting SARS-coronavirus strain were checked by using Epitope Conservancy Analysis tool of IEDB (http://tools.iedb.org/conservancy/) [6].

Molecular interaction of the HLA-epitope pair

The MHC class I ad class II molecules were downloaded from RCSB PDB database (https://www.rcsb.org/). Those that were not available were in PDB database were retrieved from pHLA database (https://www.phla3d.com.br/) [7]. The structures of the SARS-CoV-2 proteins were constructed using I-TASSER (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) [8], HTL and CTL epitopes were mapped and their structures were retrieved using PyMol tool. For the molecular interaction analysis of the predicted best HLA-epitope pairs for both MHC class I and class II alleles, ClusPro protein-protein docking tool (https://cluspro.org/login.php) [9] was utilized.

B-Cell epitope prediction

ABCpred tool based on artificial neural network was explored for B-cell epitope prediction (https://webs.iiitd.edu.in/raghava/abcpred/index.html)[10]. For higher accuracy, the threshold for the prediction was kept 0.90.

Designing of Vaccine construct

For constructing a multi-epitope vaccine construct, the selected best HTL, CTL and B-cell epitopes were joined by using various linkers. Four adjuvants namely, β-defensin, universal memory T cell helper peptide (TpD), PADRE sequence and a M cell ligand were also added to the vaccine construct.

Antigenicity, allergenicity and toxicity analysis of the epitopes and vaccine construct:

The antigenicity of the selected high affinity HTL, CTL and B-cell epitopes and the vaccine construct was analyzed using VaxiJen server. For the allergenicity analysis, three different tools namely AlgPred (https://webs.iiitd.edu.in/raghava/algpred/submission.html) [11], AllerTop (https://www.ddg-pharmfac.net/AllerTOP/method.html) [12] and AllergenFP v.1.0 (http://ddg-pharmfac.net/AllergenFP/index.html)[13] were utilized. Thereafter, toxic nature of the epitopes and the vaccine construct was checked by utilizing the ToxinPred server (https://webs.iiitd.edu.in/raghava/toxinpred/motif_scan.php) [14].

Population coverage of the vaccine construct

To check the population coverage of the vaccine construct, Population coverage tool available at IEDB (http://tools.iedb.org/population/)[15] server was utilized. The HLA class I and class II alleles in the final construct were entered in the tool and the population coverage of the alleles were calculated for the top 26 countries that are severely affected by the SARS-COV-2 virus.

Physiochemical property analyses of the multi-epitope vaccine construct

Expasy's ProtParam (https://web.expasy.org/protparam/)[16] was explored for the physiochemical properties evaluation of the vaccine construct. Solubility of the construct was

calculated using SolPro tool available at SCRATCH protein predictor server (http://scratch.proteomics.ics.uci.edu/) [17].

Vaccine construct structure prediction and validation

Secondary structure of the multi-epitope vaccine construct was predicted using SOPMA(https://npsa-prabi.ibcp.fr/cgibin/npsa_automat.pl?page=/NPSA/npsa_sopma.html) [18] and PSI-PRED (http://bioinf.cs.ucl.ac.uk/psipred/) server [19]. For the tertiary structure prediction of the vaccine construct Robetta server (http://robetta.bakerlab.org/) based on *ab-initio* and homology modelling was utilized [20]. The predicted structure was refined by using 3D Refine (http://sysbio.rnet.missouri.edu/3Drefine/) [21]and further by GalaxyRefine (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE) [22]. The structures were evaluated by using RAMPAGE (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) and ERRAT server (https://servicesn.mbi.ucla.edu/ERRAT/).

Standard molecular dynamics of the vaccine construct

The refined modelled structure of the multi-epitope vaccine construct was further evaluated for its stability in the real environment by simulating it in a water sphere using NAMD-standard molecular dynamics tool (https://www.ks.uiuc.edu/Research/namd/). The required structure files (.psf) were generated by psfgen using Visual Molecular Dynamics (VMD) tool v.1.9.3 by utilizing CHARMM force fields for proteins. Initially, a 10000 steps energy minimization was performed followed by subsequent heating the system from 0 K to 310 K. Thereafter, a 10 ns standard molecular dynamics was performed and trajectory DCD file generated was evaluated.

Interaction analysis of Vaccine construct with immune system molecules and there MD analysis

To check the interaction of the multi-epitope vaccine construct with two immuno-receptors, TLR-3 and TLR-8, ClusPro docking server was used and the resultant best complexes were then simulated for 10 ns in a water sphere using NAMD.

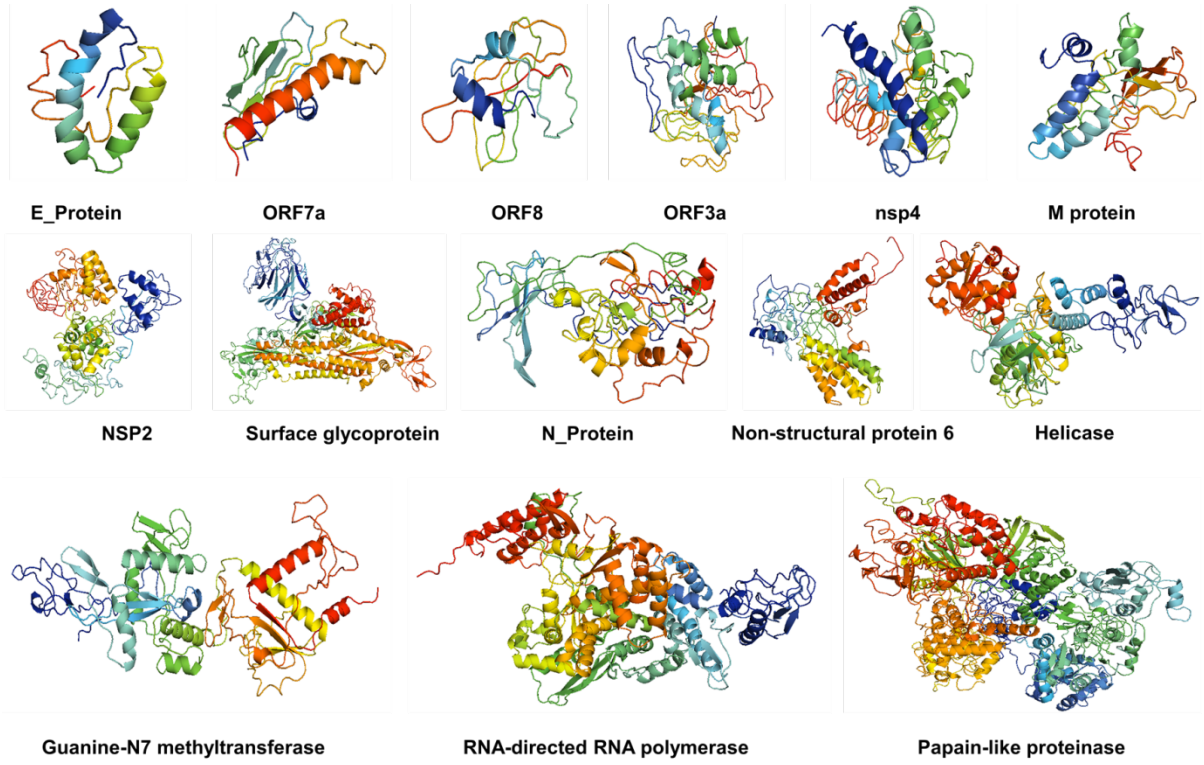
Codon optimization and *In-silico* cloning

For the expression and isolation of the constructed multi-epitope vaccine in *Escherichia coli* K12 strain, the construct was first converted to cDNA using Reverse translate tool available at Expasy server. The resultant DNA was further optimized using JCAT server [23]. Finally, the cDNA construct was inserted into the pET28a (+) vector using *HindIII* and *BamHI* restriction sites.

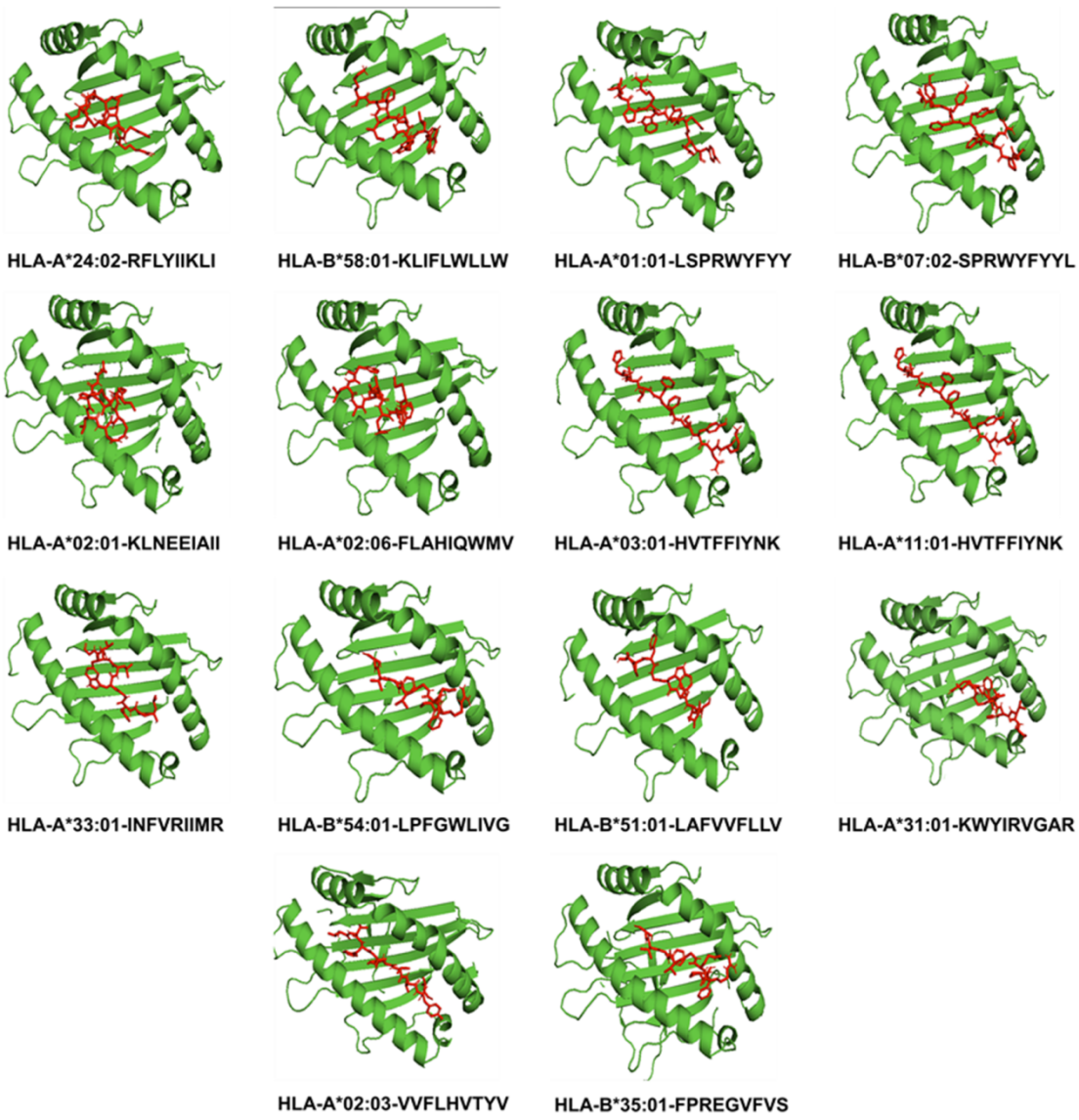
Supplementary References

[1] Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC bioinformatics. 2007;8:4.
[2] Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3. 0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics. 2013;65:711-24.
[3] Sidney J, Assarsson E, Moore C, Ngo S, Pinilla C, Sette A, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome research. 2008;4:2.
[4] Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. The Journal of Immunology. 2017;199:3360-8.
[5] Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. PLoS computational biology. 2013;9.
[6] Bui H-H, Sidney J, Li W, Fusseder N, Sette A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. BMC bioinformatics. 2007;8:361.
[7] Liu T, Pan X, Chao L, Tan W, Qu S, Yang L, et al. Subangstrom accuracy in pHLA-I modeling by Rosetta FlexPepDock refinement protocol. Journal of chemical information and modeling. 2014;54:2233-42.
[8] Zhang Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics. 2008;9:40.
[9] Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: a fully automated algorithm for protein–protein docking. Nucleic acids research. 2004;32:W96-W9.
[10] Saha S, Raghava GPS. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins: Structure, Function, and Bioinformatics. 2006;65:40-8.
[11] Saha S, Raghava G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic acids research. 2006;34:W202-W9.
[12] Dimitrov I, Flower DR, Doytchinova I. AllerTOP-a server for in silico prediction of allergens. BMC bioinformatics: BioMed Central; 2013. p. S4.
[13] Dimitrov I, Naneva L, Doytchinova I, Bangov I. AllergenFP: allergenicity prediction by descriptor fingerprints. Bioinformatics. 2014;30:846-51.
[14] Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Open Source Drug Discovery C, et al. In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLOS ONE. 2013;8:e73957.
[15] Bui H-H, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC bioinformatics. 2006;7:153.
[16] Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD, Bairoch A. Protein identification and analysis tools on the ExPASy server. The proteomics protocols handbook: Springer; 2005. p. 571-607.
[17] Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic acids research. 2005;33:W72-W6.
[18] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Bioinformatics. 1995;11:681-4.

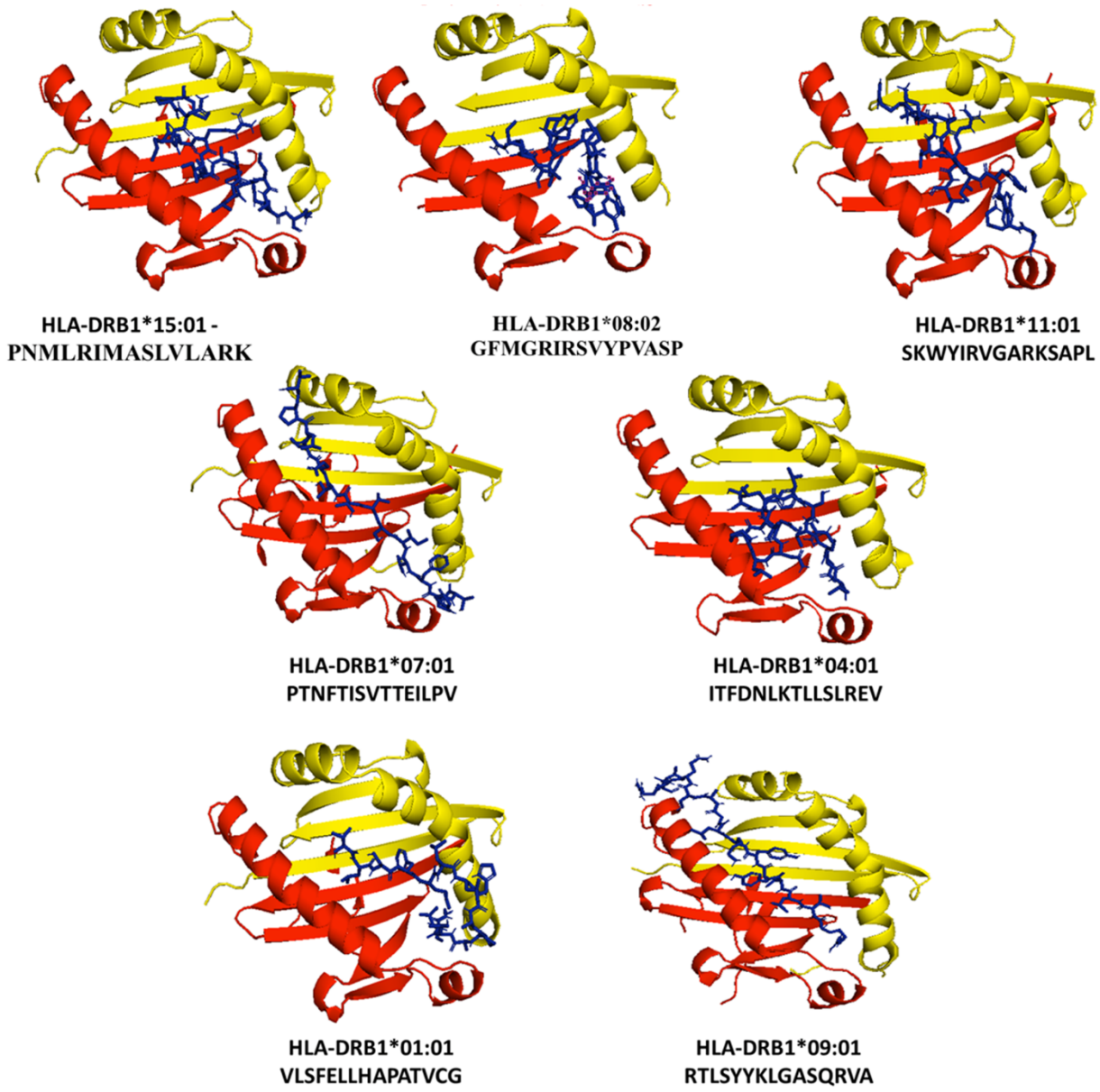
[19] McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000;16:404-5.
[20] Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic acids research. 2004;32:W526-W31.
[21] Bhattacharya D, Nowotny J, Cao R, Cheng J. 3Drefine: an interactive web server for efficient protein structure refinement. Nucleic acids research. 2016;44:W406-W9.
[22] Heo L, Park H, Seok C. GalaxyRefine: protein structure refinement driven by side-chain repacking. Nucleic acids research. 2013;41:W384-W8.
[23] Grote A, Hiller K, Scheer M, Münch R, Nörtemann B, Hempel DC, et al. JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. Nucleic acids research. 2005;33:W526-W31.



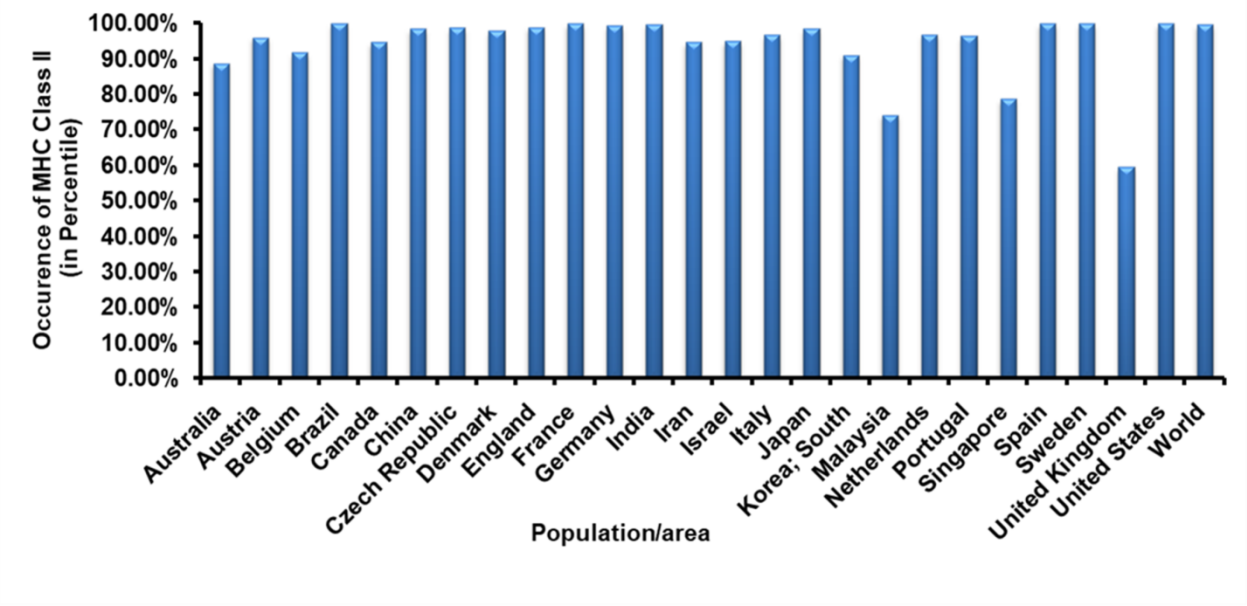
Supplementary Figure S1. Three dimensional structures in ribbon illustration of the modelled SARS-CoV-2 proteins as predicted by I-TASSER. The Figures are generated using Pymol visualization tool.



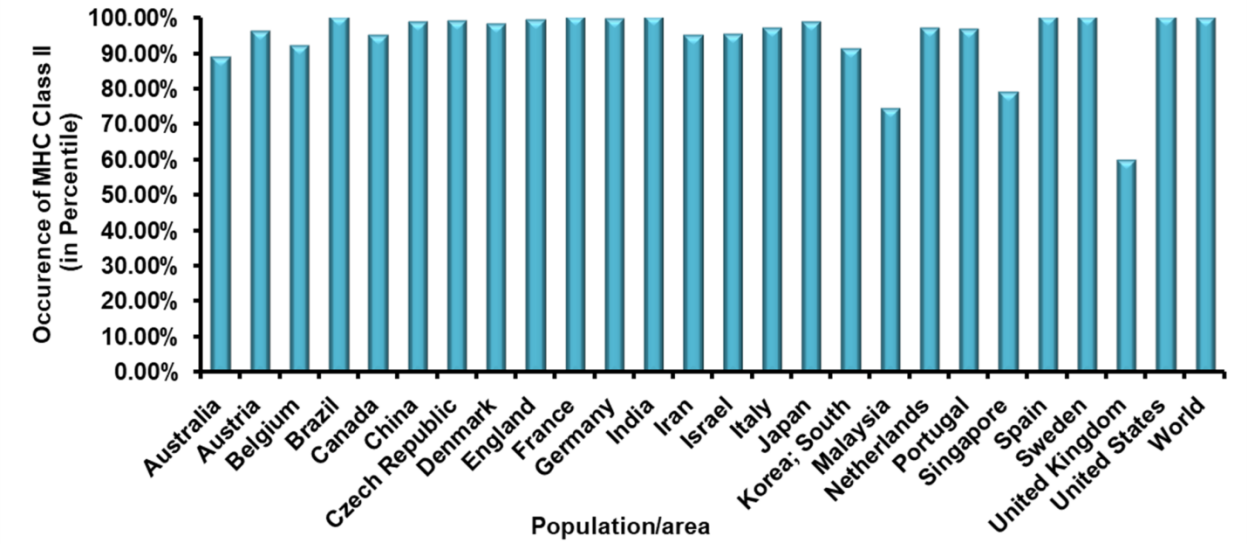
Supplementary Figure S2a: The molecular interaction of the screened CTL epitopes with the epitope binding groove of HLA-I molecules.



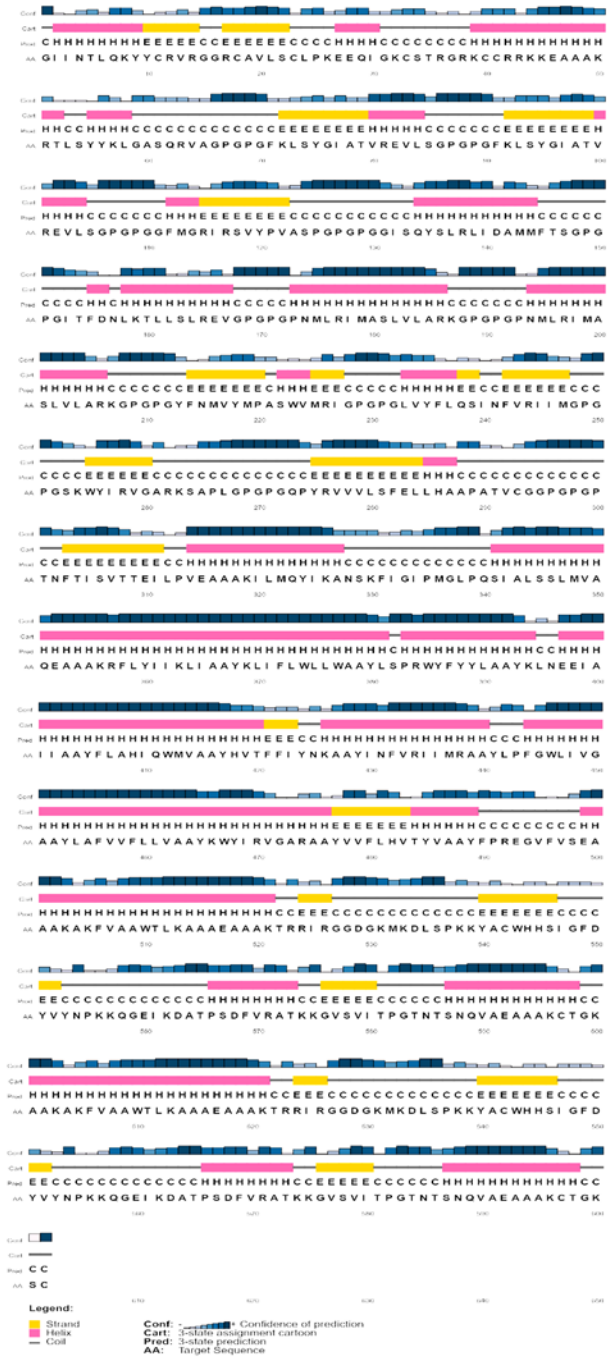
Supplementary Figure S2b: The molecular interaction of the screened CTL epitopes with the epitope binding groove of HLA-II molecules.



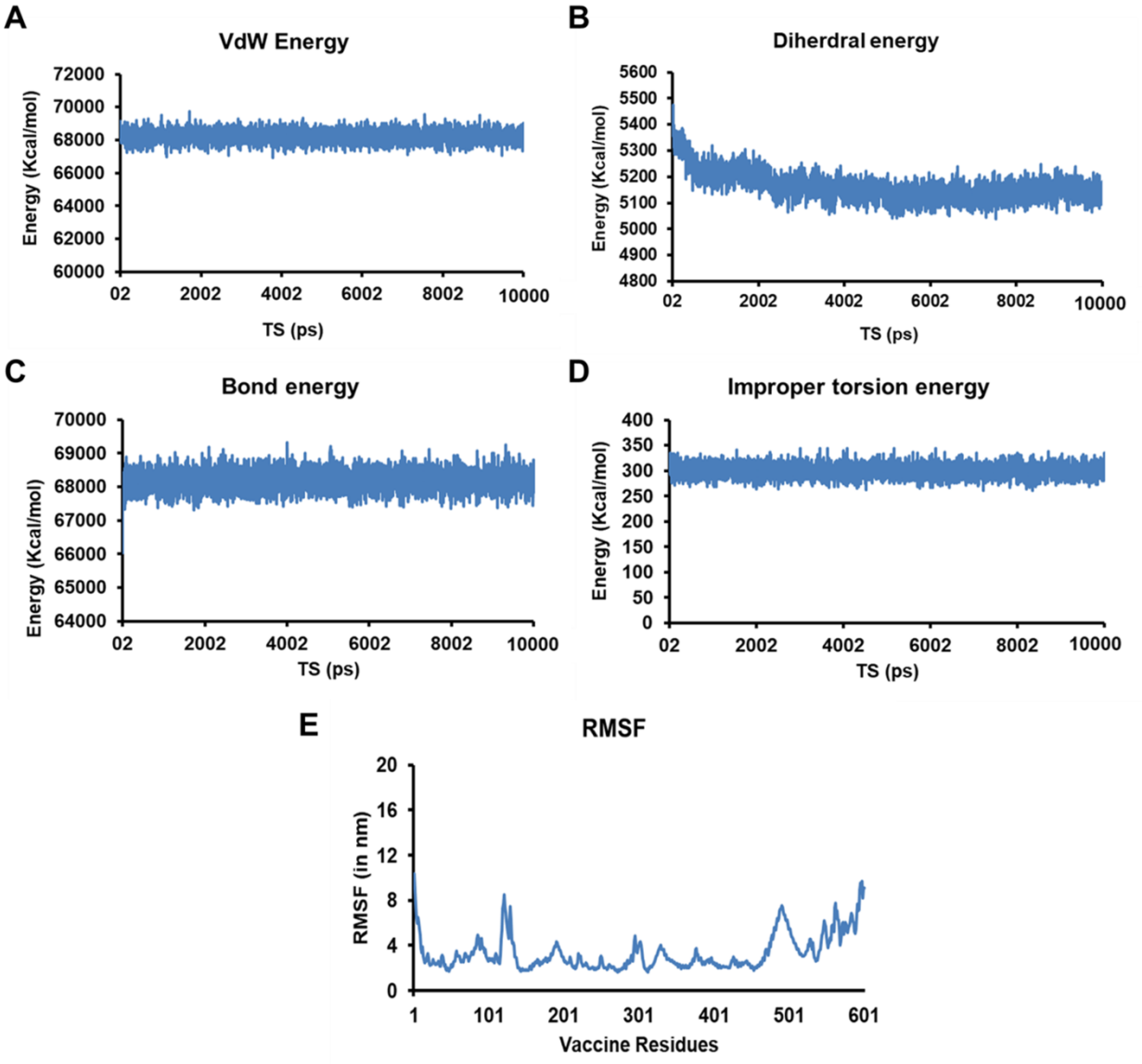
Supplementary Figure S3: Population Coverage of HLA class I for whom the best antigenic epitopes were screened in various countries severely affected by COVID-19 in the world.



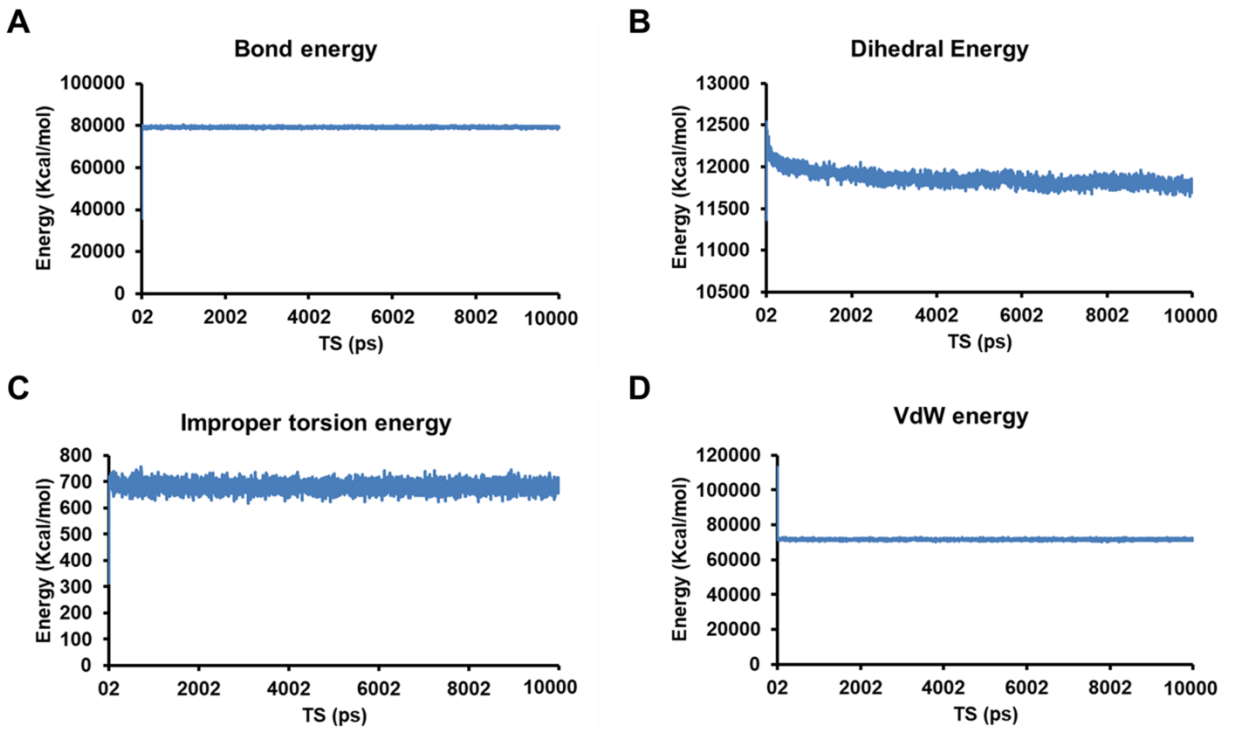
Supplementary Figure S4: Population Coverage of HLA class II for whom the best antigenic epitopes were screened in various countries severely affected by COVID-19 in the world.



Supplementary Figure S5. Secondary Structure prediction result of the SARS-CoV-2 multi-epitope vaccine as predicted by PSI-PRED server.



Supplementary Figure S6. (A-D) Energy plots depicting the vanderwaal's energy (A), dihedral angle energy (B), improper dihedral energy (C) and bond energy (D) of the SARS-CoV-2 multi-epitope vaccine during the 10 ns molecular dynamic simulation analysis. (E) Root mean square fluctuation as observed in the residues of the vaccine construct in 10ns duration.



Supplementary Figure S7. (A-D) Energy plots depicting the vanderwaal's energy (A), dihedral angle energy (B), improper dihedral energy (C) and bond energy (D) of the SARS-CoV-2 multi-epitope vaccine and TLR3 complex during the 10 ns molecular dynamic simulation analysis.

Supplementary Table S1. List of Helper T-cell epitopes of SARS-CoV-2 antigenic proteins selected for the multi-epitope vaccine construct with the percentile rank, antigenicity score and allergenicity prediction.

SARS-CoV-2 proteins	Position	HLA alleles	Method	Epitope Sequence	percentile_rank	adjusted_rank	Antigenicity Score	Allergenicity (AlgPred)
M glycoprotein	174	HLA-DRB1*09:01	Consensus (comb.lib./smm/nn)	RTLSYYKLGASQRVA	0.06	0.06	0.5644	NON ALLERGEN
ORF1ab [Helicase (Hel)]	5469	HLA-DQA1*04:01/DQB1*04:01	NetMHCIIpan	FKLSYGIATVREVLS	0.89	0.89	0.7783	NON ALLERGEN
ORF1ab [Helicase (Hel)]	5469	HLA-DQA1*04:02/DQB1*04:02	NetMHCIIpan	FKLSYGIATVREVLS	0.89	0.89	0.7783	NON ALLERGEN
ORF1ab (Non-structural protein 2_nsp2)	295	HLA-DRB1*08:02	Consensus (smm/nn/sturniolo)	GFMGRIRSVYPVASP	0.64	0.64	0.7131	NON ALLERGEN
ORF1ab(Non-structural protein 2_nsp2)	572	**HLA-DQA1*01:01/DQB1*02:01**	NetMHCIIpan	GISQYSLRLIDAMMF	0.66	0.66	0.7195	NON ALLERGEN
ORF1ab (Non-structural protein 2_nsp2)	574	HLA-DQA1*01:02/DQB1*02:02	NetMHCIIpan	SQYSLRLIDAMMFTS	0.33	0.33	0.7276	NON ALLERGEN
ORF1ab Non-structural protein 6 (nsp6)	3649	HLA-DQA1*05:01/DQB1*05:01	NetMHCIIpan	YFNMVYMPASWVMRI	0.19	0.19	0.7244	NON ALLERGEN
ORF1ab RNA-directed RNA polymerase (RdRp)	5019	HLA-DPA1*03:01/DPB1*03:01	NetMHCIIpan	PNMLRIMASLVLARK	0.06	0.06	0.4128	NON ALLERGEN
ORF1ab RNA-directed RNA polymerase (RdRp) **(H8 model.000.06)**	5019	HLA-DRB1*15:01	Consensus (smm/nn/sturniolo)	PNMLRIMASLVLARK	0.01	0.01	0.4128	NON ALLERGEN
ORF1ab Papain-like proteinase	1551	HLA-DRB1*04:01	Consensus (smm/nn/sturniolo)	ITFDNLKTLLSLREV	0.33	0.33	0.8234	NON ALLERGEN
ORF3a protein	111	HLA-DPA1*02:01/DPB1*02:01	NetMHCIIpan	LVYFLQSINFVRIIM	0.09	0.09	0.5186	NON ALLERGEN
ORF8 protein	43	HLA-DRB1*11:01	Consensus (smm/nn/sturniolo)	SKWYIRVGARKSAPL	0.72	0.72	0.8829	NON ALLERGEN
S protein	506	HLA-DPA1*01:03/DPB1*01:01	NetMHCIIpan	QPYRVVVLSFELLHA	0.29	0.29	0.9109	NON ALLERGEN
S protein	512	HLA-DRB1*01:01	Consensus (comb.lib./smm/nn)	VLSFELLHAPATVCG	0.03	0.03	0.4784	NON ALLERGEN
S protein	715	HLA-DRB1*07:01	Consensus (comb.lib./smm/nn)	PTNFTISVTTEILPV	0.51	0.51	1.1349	NON ALLERGEN

Supplementary Table S2. List of Cytotoxic T-cell epitopes of SARS-CoV-2 antigenic proteins selected for the multi-epitope vaccine construct with the percentile rank, binding level, immunogenicity score, antigenicity score and allergenicity prediction.

SARS-CoV-2 Protein	Start genomic point	HLA Allele	Epitope	Score	Aff(nM)	%Rank	Binding Level	Antigenicity score	Immunogenecity Score	Allergenicity
M Protein	44	HLA-A*24:02	RFLYIIKLI	0.496296	232.8	0.4035	SB	0.4257	0.24108	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number O00626 defined as non-allergen
M Protein	50	HLA-B*58:01	KLIFLWLLW	0.588682	85.7	0.4281	SB	0.4968	0.34287	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number P60140 defined as non-allergen
N Protein	104	HLA-A*01:01	LSPRWYFYY	0.59866	76.9	0.073	SB	1.2832	0.36094	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number Q0VD83 defined as non-allergen
N Protein	105	HLA-B*07:02	SPRWYFYYL	0.748063	15.3	0.0496	SB	0.734	0.34101	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number Q7XAQ6 defined as non-allergen
ORF1ab (Non-structural protein 2_nsp2)	468	HLA-A*02:01	KLNEEIAII	0.673256	34.3	0.4552	SB	0.6394	0.43221	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number A2Z5D8defined as non-allergen
ORF1ab (Non-structural protein 4 (nsp4)	3122	HLA-A*02:06	FLAHIQWMV	0.914969	2.5	0.0074	SB	0.8064	0.38891	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number P31358 defined as non-allergen
ORF3a Protein	227	HLA-A*03:01	HVTFFIYNK	0.544282	138.5	0.4821	SB	0.9862	0.36278	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number O43852 defined as non-allergen
ORF3a Protein	227	HLA-A*11:01	HVTFFIYNK	0.759484	13.5	0.0517	SB	0.9862	0.36278	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number O43852 defined as non-allergen
ORF3a Protein	118	HLA-A*33:01	INFVRIIMR	0.538132	148	0.3814	SB	0.7646	0.26494	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number O15511 defined as non-allergen

ORF3a Protein	41	HLA-B*54:01	LPFGWLIVG	0.63608	51.3	0.0762	SB	1.981	0.4138	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number Q156A1 defined as non-allergen
ORF4 (E protein)	21	HLA-B*51:01	LAFVVFLLV	0.305577	1832.6	0.3161	SB	0.7976	0.2141	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number P60140 defined as non-allergen
ORF8 Protein	44	HLA-A*31:01	KWYIRVGAR	0.72322	20	0.1425	SB	1.5772	0.27344	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number Q96PC3 defined as non-allergen
S Protein	1060	HLA-A*02:03	VVFLHVTYV	0.793601	9.3	0.1844	SB	1.5122	0.1278	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number Q6UWT4 defined as non-allergen
S Protein	1089	HLA-B*35:01	FPREGVFVS	0.56001	116.8	0.3526	SB	0.5509	0.31233	PROBABLE NON-ALLERGEN The nearest protein is: UniProtKB accession number P07711 defined as non-allergen

Supplementary Table S3. List of B cell Lymphocytes epitopes of SARS-CoV-2 antigenic proteins selected for the multi-epitope vaccine construct with the ABCPred prediction score, antigenicity score and allergenicity prediction.

SARS-CoV-2 proteins	B cell epitope Sequence	Start position	Score	Antigenicity Score	AllerTop V2
Nucleocapsid	TRRIRGGDGKMKDLSP	91	0.94	1.1467	Non-allergen
ORF1ab Guanine-N7 methyltransferase (ExoN)	YACWHHSIGFDYVYNP	6149	0.96	0.9219	Non-allergen
ORF3a	QGEIKDATPSDFVRAT	17	0.94	0.9161	Non-allergen
Surface glycoprotein	GVSVITPGTNTSNQVA	594	0.95	0.4651	Non-allergen

Supplementary Table S4. The Lowest energy score obtained for the 30 conformers of TLR3 – Vaccine construct complex as predicted by ClusPro results.

Cluster	Members	Representative	Weighted Score
0	35	Center	-1227.0
		Lowest Energy	-1275.4
1	30	Center	-1106.4
		Lowest Energy	-1171.4
2	29	Center	-1198.2
		Lowest Energy	-1198.2
3	25	Center	-1036.6
		Lowest Energy	-1158.1
4	23	Center	-1108.9
		Lowest Energy	-1108.9
5	21	Center	-999.0
		Lowest Energy	-1427.4
6	21	Center	-1243.7
		Lowest Energy	-1243.7
7	20	Center	-1030.6
		Lowest Energy	-1128.3
8	19	Center	-1101.4
		Lowest Energy	-1146.2
9	18	Center	-1252.6
		Lowest Energy	-1287.4
10	17	Center	-1202.9
		Lowest Energy	-1202.9
11	16	Center	-1017.1
		Lowest Energy	-1150.9
12	16	Center	-1218.4

		Lowest Energy	-1218.4
13	16	Center	-1052.3
		Lowest Energy	-1140.9
14	16	Center	-1012.1
		Lowest Energy	-1268.0
15	15	Center	-1032.3
		Lowest Energy	-1356.2
16	15	Center	-1156.2
		Lowest Energy	-1266.9
17	15	Center	-1131.0
		Lowest Energy	-1267.3
18	15	Center	-1069.0
		Lowest Energy	-1162.9
19	14	Center	-1014.0
		Lowest Energy	-1160.3
20	13	Center	-1189.0
		Lowest Energy	-1189.0
21	13	Center	-1130.0
		Lowest Energy	-1130.0
22	13	Center	-1084.9
		Lowest Energy	-1084.9
23	13	Center	-1055.6
		Lowest Energy	-1070.5
24	12	Center	-1084.9
		Lowest Energy	-1084.9
25	11	Center	-1036.2
		Lowest Energy	-1123.7
26	11	Center	-1205.6
		Lowest Energy	-1205.6
27	11	Center	-1119.2
		Lowest Energy	-1119.2
28	11	Center	-1098.0
		Lowest Energy	-1206.7
29	11	Center	-1019.6
		Lowest Energy	-1077.6

Supplementary Data S1. Codon optimized and adapted cDNA sequence of the SARD-CoV multi-epitope vaccine construct.

GGTATCATCAACACCCTGCAGAAATACTACTGCCGTGTTCGTGGTGGTCG 50
TTGCGCTGTTCTGTCTTGCCTGCCGAAAGAAGAACAGATCGGTAAATGCT 100
CTACCCGTGGTCGTAAATGCTGCCGTCGTAAAAAAGAAGCTGCTGCTAAA 150
CGTACCCTGTCTTACTACAAACTGGGTGCTTCTCAGCGTGTTGCGGGTCC 200
GGGCCCGGGCTTCAAACTGTCTTACGGTATCGCTACCGTTCGTGAAGTTC 250
TGTCGGGTCCGGGCCCGGGCTTCAAACTGTCTTACGGTATCGCTACCGTT 300
CGTGAAGTTCTGTCGGGTCCGGGTCCAGGTGGCTTCATGGGTCGTATACG 350
TTCTGTTTACCCGGTTGCTTCTCCGGGTCCGGGTCCGGGTGGTATCTCTC 400
AGTACTCTCTGCGTCTGATCGACGCTATGATGTTCACCTCTGGTCCGGGT 450
CCGGGTATCACCTTCGACAACCTGAAAACCCTGCTGTCTCTGCGTGAAGT 500
TGGTCCGGGTCCGGGTCCGAACATGCTGCGTATCATGGCTTCTCTGGTTC 550
TGGCTCGTAAAGGTCCGGGTCCGGGTCCGAACATGCTGCGTATCATGGCT 600
TCTCTGGTTCTGGCTCGTAAAGGTCCGGGTCCGGGTTACTTCAACATGGT 650
TTACATGCCGGCTTCTTGGGTTATGCGTATCGGTCCGGGTCCAGGGCTGG 700
TATACTTCCTGCAATCTATCAACTTCGTTCGTATCATCATGGGTCCGGGT 750
CCGGGTTCTAAATGGTACATCCGTGTTGGTGCTCGTAAATCTGCTCCGCT 800
GGGTCCGGGTCCGGGTCAGCCGTACCGTGTTGTTGTTCTGTCTTTCGAAC 850
TGCTGCACGCTGCTCCGGCTACCGTTTGCGGTGGTCCGGGTCCGGGTCCG 900
ACCAACTTCACCATCTCTGTTACCACCGAAATCCTGCCGGTTGAAGCTGC 950
TGCTAAAATCCTGATGCAGTACATCAAAGCTAACTCTAAATTCATCGGTA 1000
TCCCGATGGGTCTGCCGCAGTCTATCGCTCTGTCTTCTCTGATGGTTGCT 1050
CAGGAAGCTGCTGCTAAACGTTTCCTGTACATCATCAAACTGATCGCTGC 1100
TTACAAACTGATCTTCCTGTGGCTGCTGTGGGCTGCTTACCTGTCTCCGC 1150
GTTGGTACTTCTACTACCTGGCTGCTTACAAACTGAACGAAGAAATCGCT 1200
ATCATCGCTGCTTACTTCCTGGCTCACATCCAGTGGATGGTTGCTGCTTA 1250
CCACGTTACCTTCTTCATCTACAACAAAGCTGCTTACATCAACTTCGTTC 1300
GTATCATCATGCGTGCTGCTTACCTGCCGTTCGGTTGGCTGATCGTTGGT 1350
GCTGCTTACCTGGCTTTCGTTGTTTTCCTGCTGGTTGCTGCTTACAAATG 1400
GTACATCCGTGTTGGTGCTCGTGCTGCTTACGTTGTTTTCCTGCACGTTA 1450
CCTACGTTGCTGCTTACTTCCCGCGTGAGGGTGTGTTCGTCTCTGAAGCT 1500
GCTGCTAAAGCTAAATTCGTTGCTGCTTGGACCCTGAAAGCTGCTGCTGA 1550
AGCTGCTGCTAAAACCCGTCGTATCCGTGGTGGTGACGGTAAAATGAAAG 1600
ACCTGTCTCCGAAAAAATACGCTTGCTGGCACCACTCTATCGGTTTCGAC 1650
TACGTTTACAACCCGAAAAAACAGGGTGAAATCAAAGACGCTACCCCGTC 1700
TGACTTCGTTCGTGCTACCAAAAAAGGTGTTTCTGTTATCACCCCGGGTA 1750
CCAACACCTCTAACCAGGTTGCTGAAGCTGCTGCTAAATGCACCGGTAAA 1800
TCTTGC