Engineering Novel Epitope-based Subunit Vaccine against SARS-CoV-2 by Exploring the Immunoinformatics Approach

Bishajit Sarkar[1,2], Md. Asad Ullah[1,2], Yusha Araf[1,3], Mohammad Shahedur Rahman[1,2*]

[1]COVID Research Cell (CRC), Wazed Miah Science Research Center (WMSRC), Jahangirnagar University, Savar, Dhaka, Bangladesh

[2]Department of Biotechnology and Genetic Engineering, Faculty of Biological Sciences, Jahangirnagar University, Savar, Dhaka, Bangladesh

[3]Department of Genetic Engineering and Biotechnology, School of Life Sciences, Shahjalal University of Science and Technology, Sylhet, Bangladesh

*For correspondence:

Professor Dr. Mohammad Shahedur Rahman,

Deputy Director

Wazed Mia Science Research Center, Jahangirnagar University,

Savar, Dhaka-1342, Bangladesh.

Email: rahmanms@juniv.edu

Supplementary Information

Tables

Supplementary Table S1. The antigenicity and physicochemical property analyses of the selected viral proteins. AN; antigenicity, pI; theoretical pI, II; instability index, AI; aliphatic index, GRAVY; grand average of hydropathicity.

Name of the protein	AN	pI	Total number of negatively charged residues	Total number of positively charged residues	Ext. coefficient	Half-life	II	AI	GRAVY
Spike glycoprotein	Antigenic	6.32	109	103	148960	30 hours (mammalian reticulocytes, >20 hours (yeast, >10 hours (Escherichia coli	Stable (32.86)	84.67	-0.077
Nucleocapsid phosphoprotein	Antigenic	10.09	36	61	43890	30 hours (mammalian reticulocytes, >20 hours (yeast, >10 hours (Escherichia coli	Unstable (55.81)	52.53	-0.980
Membrane Glycoprotein	Antigenic	9.51	13	21	52160	30 hours (mammalian reticulocytes, >20 hours (yeast, >10 hours (Escherichia coli	Stable (39.14)	120.86	-0.446
Envelope Protein	Antigenic	8.57	3	5	6085	30 hours (mammalian reticulocytes, >20 hours (yeast, >10 hours (Escherichia coli	Stable (38.68)	144.00	-0.827

Supplementary Table S2. Prediction of MHC class-I and MHC class-II epitopes of Spike Glycoprotein, and their topology, antigenicity, allergenicity, toxicity, conservancy, and human homology analyses. AN; antigenicity, AG; allergenicity, HH; human homology, CN; conservancy, NA; not allowed.

MHC classes	Epitopes	AN	AG	Toxicity	HH	CN	IFN-gamma inducing capability	IL-4 inducing capability	IL-10 inducing capability
MHC class-I	NSFTRGVYY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	STQDLFLPF	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	VLPFNDGVY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	CNDPFLGVY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	WMESEFRVY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	YSSANNCTF	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	SANNCTFEY	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	FVFKNIDGY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	NIDGYFKIY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	WTAGAAAYY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	GAAAYYVGY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	ITDAVDCAL	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	LSETKCTLK	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	NATRFASVY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	RISNCVADY	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	CVADYSVLY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	NSASFSTFK	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	ASFSTFKCY	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	FTNVYADSF	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	VGGNYNYLY	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	ERDISTEIY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	TSNQVAVLY	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	YQDVNCTEV	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA

	QLTPTWRVY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	GAEHVNNSY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	VASQSIIAY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	KTSVDCTMY	Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	STECSNLLL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	ECSNLLLQY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	RSFIEDLLF	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	LTDEMIAQY	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	GTITSGWTF	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	RVDFCGKGY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	FVSNGTHWF	Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	VSNGTHWFV	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	MTSCCSCLK	Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	VLKGVKLHY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
MHC class-II	KTQSLLIVNNATNVV	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	Non-inducer	Non-inducer	Non-inducer
	LLIVNNATNVVIKVC	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Non-inducer	Non-inducer
	QSLLIVNNATNVVIK	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	SLLIVNNATNVVIKV	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	TQSLLIVNNATNVVI	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	Non-inducer	Non-inducer	Non-inducer
	LIVNNATNVVIKVCE	Non-Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	Non-inducer	Inducer	Non-inducer
	LSFELLHAPATVCGP	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	Inducer	Inducer	Non-inducer
	SKTQSLLIVNNATNV	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	VLSFELLHAPATVCG	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Inducer
	VVLSFELLHAPATVC	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Inducer

Supplementary Table S3. Prediction of MHC class-I and MHC class-II epitopes of Nucleocapsid phosphoprotein, and their topology, antigenicity, allergenicity, toxicity, conservancy, and human homology analyses. AN; antigenicity, AG; allergenicity, HH; human homology, CN; conservancy, NA; not allowed.

MHC classes	Epitopes	AN	AG	Toxicity	HH	CN	IFN-gamma inducing capability	IL-4 inducing capability	IL-10 inducing capability
MHC class-I	NTASWFTAL	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	SSPDDQIGY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	SPDDQIGYY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	MKDLSPRWY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	DLSPRWYFY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	LSPRWYFYY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	GTTLPKGFY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	GTDYKHWPQ	Antigen	Allergen	Non-toxic	Non-homolog	Non-conserved	NA	NA	NA
	LLNKHIDAY	Non-Antigen	Allergen	Non-toxic	Non-homolog	conserved	NA	NA	NA
MHC class-II	AQFAPSASAFFGMSR	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Non-inducer
	IAQFAPSASAFFGMS	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Non-inducer	Non-inducer
	PQIAQFAPSASAFFG	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	Inducer	Non-inducer	Inducer
	QIAQFAPSASAFFGM	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	Inducer	Inducer	Inducer
	WPQIAQFAPSASAFF	Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	Inducer	Inducer	Inducer
	NPANNAAIVLQLPQG	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	Non-inducer	Non-inducer	Non-inducer
	PANNAAIVLQLPQGT	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Non-inducer	Non-inducer
	RNPANNAAIVLQLPQ	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	TRNPANNAAIVLQLP	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Non-conserved	Non-inducer	Non-inducer	Non-inducer
	GTRNPANNAAIVLQL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Non-inducer	Non-inducer

Supplementary Table S4. Prediction of MHC class-I and MHC class-II epitopes of Membrane Glycoprotein, and their topology, antigenicity, allergenicity, toxicity, conservancy, and human homology analyses. AN; antigenicity, AG; allergenicity, HH; human homology, CN; conservancy, NA; not allowed.

MHC class	Epitopes	AN	AG	Toxicity	HH	CN	IFN-gamma inducing capability	IL-4 inducing capability	IL-10 inducing capability
MHC class-I	LLEQWNLVI	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	WICLLQFAY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	YANRNRFLY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	LVGLMWLSY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	VATSRTLSY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	ATSRTLSYY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	AGDSGFAAY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	YSRYRIGNY	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	SSSDNIALL	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	SSDNIALLV	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
MHC class-II	GLMWLSYFIASFRLF	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	LMWLSYFIASFRLFA	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	LSYYKLGASQRVAGD	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Non-inducer
	RTLSYYKLGASQRVA	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Non-inducer
	IKLIFLWLLWPVTLA	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Non-inducer
	SRTLSYYKLGASQRV	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Inducer	Inducer	Non-inducer
	TLSYYKLGASQRVAG	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Inducer	Non-inducer
	VGLMWLSYFIASFRL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	MWLSYFIASFRLFAR	Non-Antigen	Allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	KLIFLWLLWPVTLAC	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer

Supplementary Table S5. Prediction of MHC class-I and MHC class-II epitopes of Envelope Protein, and their topology, antigenicity, allergenicity, toxicity, conservancy, and human homology analyses. AN; antigenicity, AG; allergenicity, HH; human homology, CN; conservancy, NA; not allowed.

MHC classes	Epitopes	AN	AG	Toxicity	HH	CN	IFN-gamma inducing capability	IL-4 inducing capability	IL-10 inducing capability
MHC class-I	LTALRLCAY	Non-Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	VSLVKPSFY	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
	LVKPSFYVY	Antigen	Allergen	Non-toxic	Non-homolog	Conserved	NA	NA	NA
MHC class-II	LLFLAFVVFLLVTLA	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	VLLFLAFVVFLLVTL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	LFLAFVVFLLVTLAI	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	LLFLAFVVFLLVTLA	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	LFLAFVVFLLVTLAI	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	NSVLLFLAFVVFLLV	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	SVLLFLAFVVFLLVT	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	VLLFLAFVVFLLVTL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	FLAFVVFLLVTLAIL	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer
	LAFVVFLLVTLAILT	Antigen	Non-allergen	Non-toxic	Non-homolog	Conserved	Non-inducer	Non-inducer	Inducer

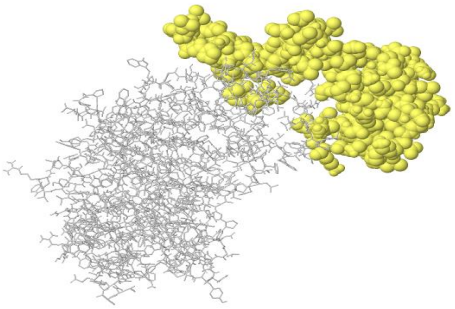
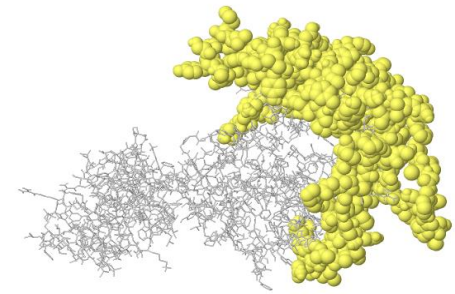
Supplementary Table S6. List of the four selected proteins from different SARS-CoV-2 isolates of different countries around the world for the conservancy analysis.

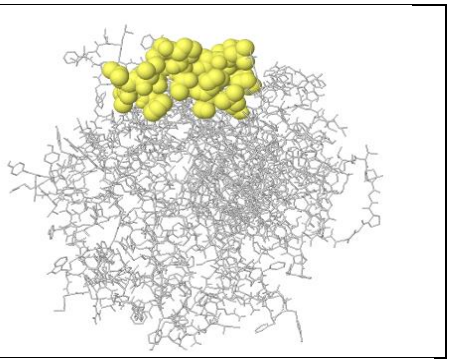
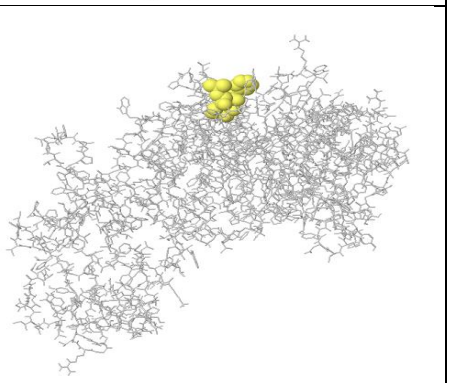
Country	Spike glycoprotein (GenBank accession number)	Nucleocapsid phosphoprotein (GenBank accession number)	Membrane Glycoprotein (GenBank accession number)	Envelope Protein (GenBank accession number)
Australia	QJR90981.1	QLG75373.1	QLG75212	QLG75187
India	QLF97939.1	QKG91282.1	QLF98110	QLF98097
South Korea	QLD99677.1	QIV15004.1	QLA46615	QIV15010
China	QLG69199.1	QKI36837.1	QKI36832	QKI36843
Italy	QLB38609.1	QKE43723.1	QKE43718	QKE43705
Japan	BCI48738.1	BCI50552.1	BCI48796	BCI48806
New Zealand	QLC90935.1	QLC90943.1	QLC90938	QLC90937
Morocco	QLE00003.1	QLE10676.1	QLE10924	QLE10670
France	QLD29172.1	QJT72118.1	QJT72125	QJT72112
Spain	QKX47197.1	QKX47229.1	QKJ68403	QKJ68402
USA	QLF99990.1	QLG00047.1	QLF95259	QLG00029
Brazil	QLF80216.1	QLD32036.1	QJA41644	QLD32030
Egypt	QKS74794.1	QKT20962.1	QKT20981	QKT21004
Poland	QLF78309.1	QJZ28151.1	QJZ28170	QJZ28277
Chile	QKY74786.1	QKY74842.1	QKY74873	QKY74824

Supplementary Table S7. Results of the secondary structure analysis of the vaccine construct.

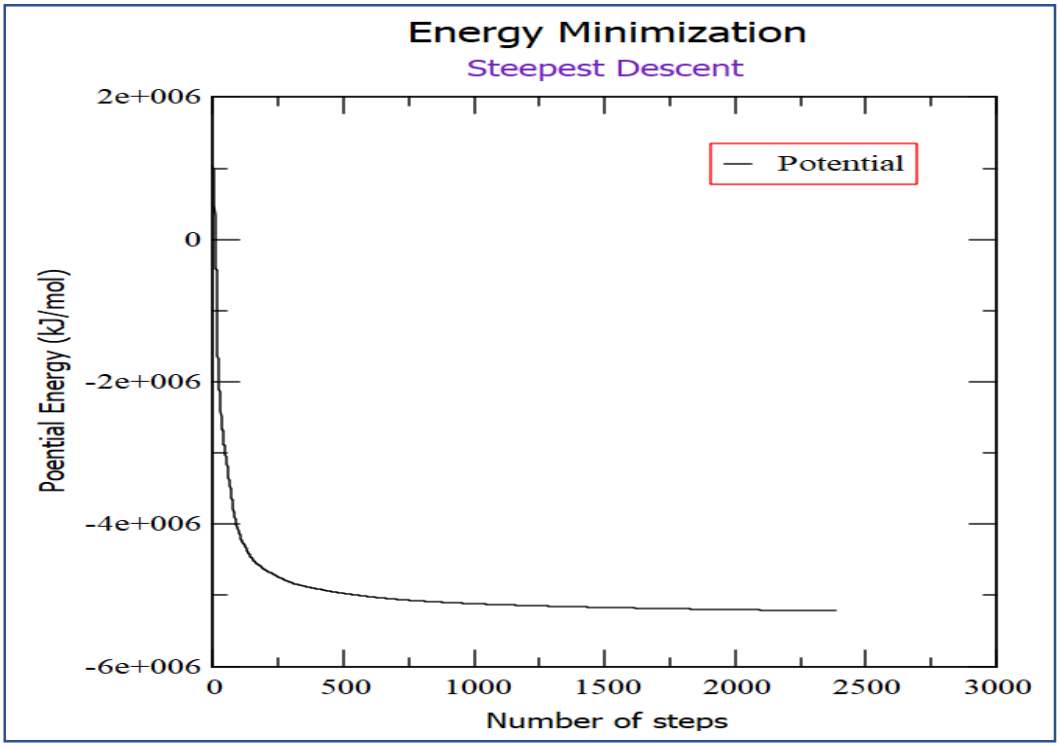
Vaccines	Secondary structure elements	PRISPRED	GOR IV	SOPMA	SIMPA96
CV	α-helix	22.34%	24.63%	27.29%	21.59%
	β-strand	36.67%	32.28%	34.44%	32.72%
	Coil structure	40.97%	43.09%	38.27%	45.51%

Supplementary Table S8. The list of the predicted conformational B-cell epitopes of the vaccine, CV with their scores.

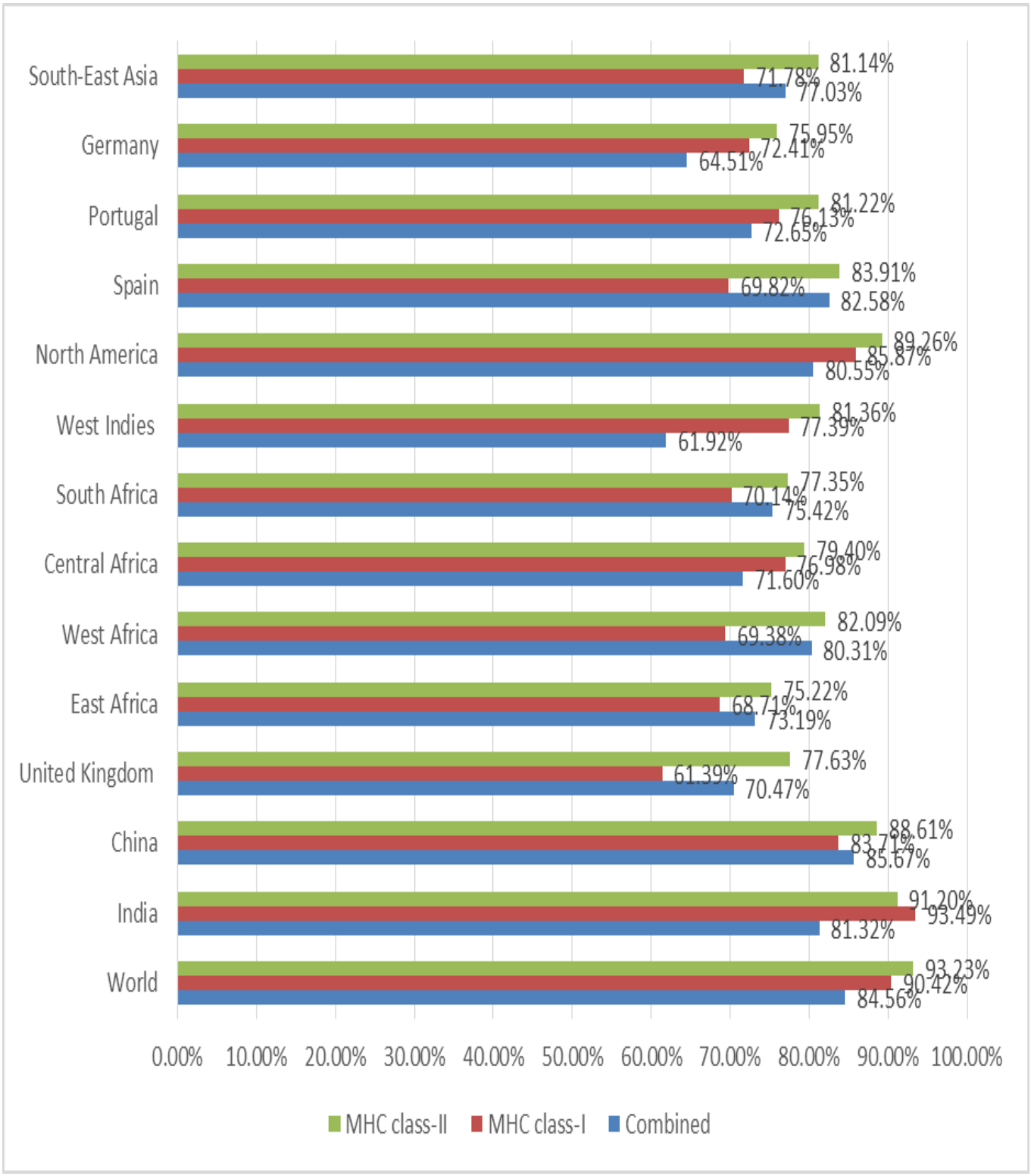
Residues	Number of residues	Score	Figure
E1, A2, A3, A4, K5, G6, I7, I8, N9, T10, _L11, Q12, K13, Y14, Y15, C16, R17, V18, R19, G20, G21, R22, C23, A24, V25, L26, S27, C28, L29, P30, K31, E32, E33, Q34, I35, G36, K37, C38, S39, T40, R41, G42, R43, K44, C45, C46, R47, R48, K49, K50, E51, A52, K55, L508, A509, F510, V511, V512, F513, L514, L515, V516, G517, P518, G519, P520, G521, S522, G537, P538, G539, P540, G541, V542, L543, L544, F545, L546, A547, F548, V549, F551, L552, L553, T555, L556, G557, P558, G559, P560, G561, F562, L563, A564, F565, V566, V567, F568, L569, L570, V571, T572, L573, A574, I575, L576, G577, P578, G579, P580, V590, T591, L592, A593, I594, L595, T596, G597, P598, G599, P600, G601	122	0.774	
A98, A99, Y100, Y101, V102, G103, Y104, A105, A106, Y107, Q108, L109, T110, P111, T112, W113, Y131, V132, L133, K134, G135, V136, K137, H139, Y140, A141, A142, Y143, S144, S145, P146, D147, D148, Q149, D158, D159, Q160, I161, G162, Y163, Y164, A165, A166, Y167, D168, L169, S170, P171, R172, W173, Y174, S181, P182, R183, W184, F186, Y188, A189, A190, Y191, L192, V193, G194, L195, M196, W197, L198, S199, Y200, A201, A202, Y203, A204, T205, S206, R207, T208, L209, S210, Y211, Y212, A213, A214, Y215, A216, G217, D218, S219, G220, F221, A222, K232, P233, S234, F235, Y236, G237, P238, G239, P240, G241, V247, N248, N249, A250, N252, K256, G257, P258, G259, P260, G261, V262, L263, S264, F265, E266, L267, L268, H269, A270, P271, A272, T273, V274, C275, G276, G277, P278, G279, P280, G281, V282, V283, L284, S285, F286, E287, L288, L289, H290, P307, S308, A309, S310, A311, M315, S316, G317, P318, G319, P320, G321, G322, T323, N325, P326, A327, N328, N329, A330, A331, I332, V333, L334, Q335, L336, G337, P338, G339, P340, G341, I342, K343, L344, I345, S368, F370, G377, P378, G379, P380, G381, K382, L383	185	0.668	

L367, P391, V392, T393, L394, A395, G397, P398, G399, A416, G417, P418, G419, P420, G421, V422, L423, L424, F425	19	0.597	
V488, F489, L490	03	0.537	

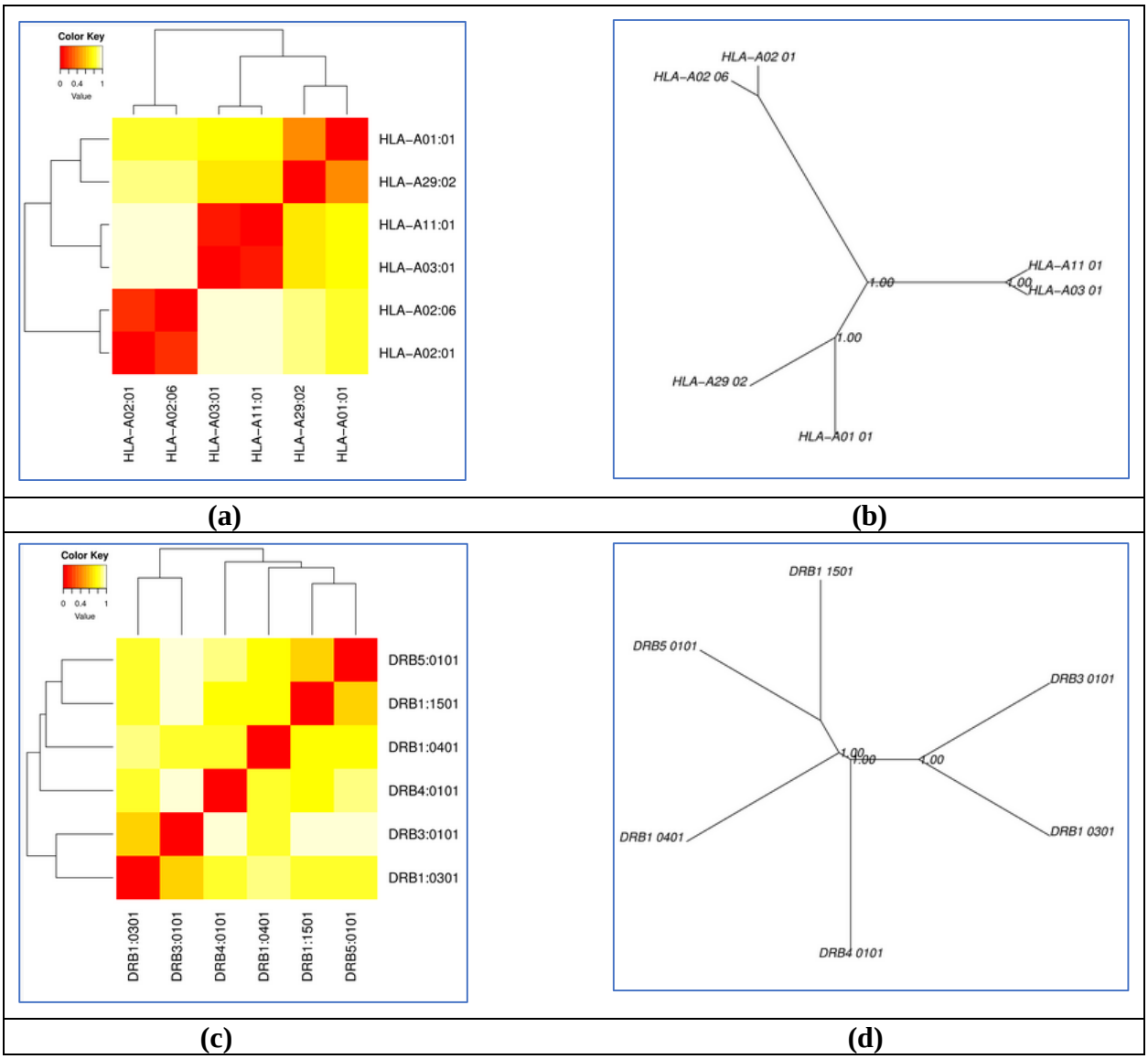
Figures



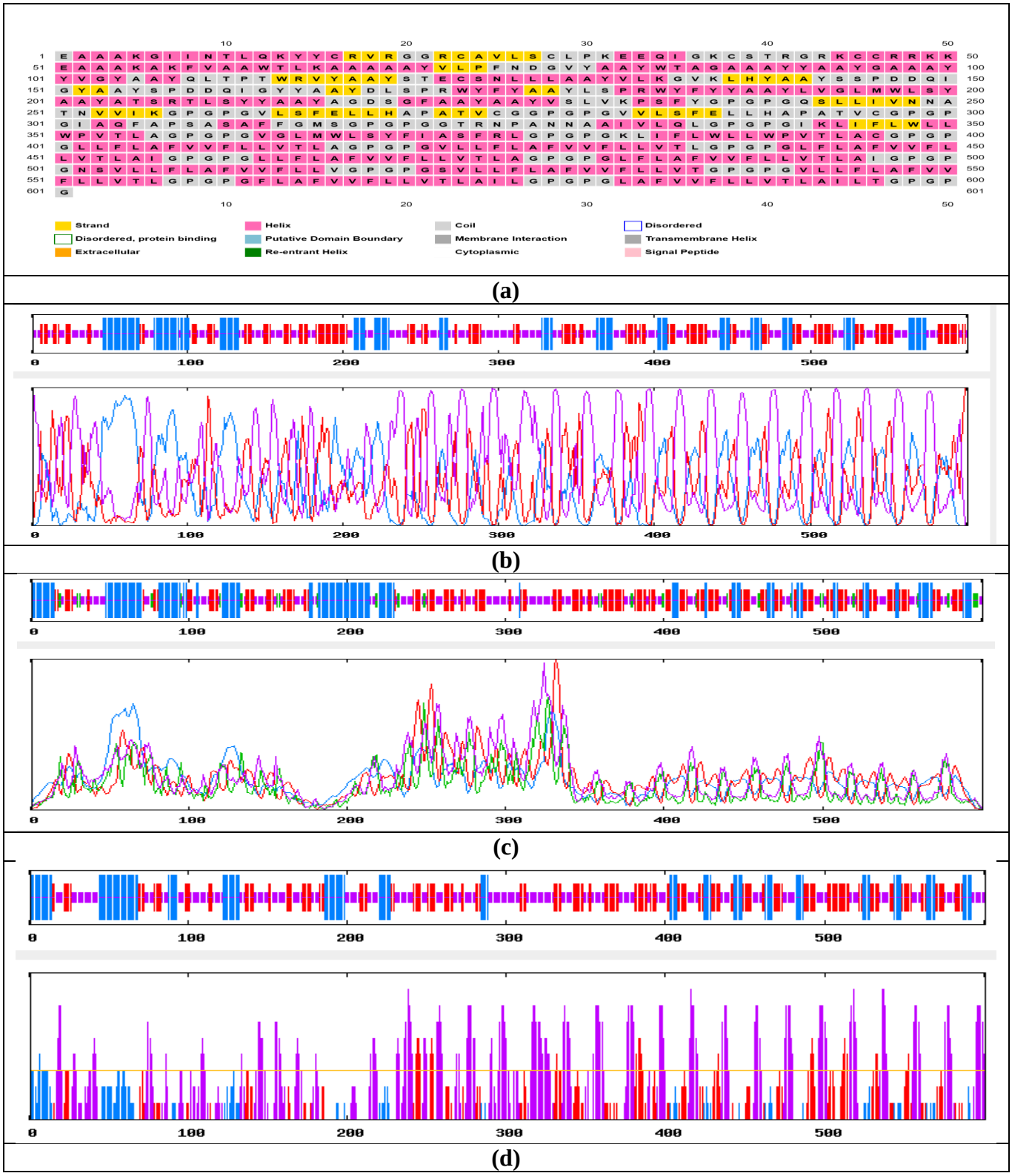
Supplementary Figure S1. Potential Energy of the docked complex calculated at the steepest descent.



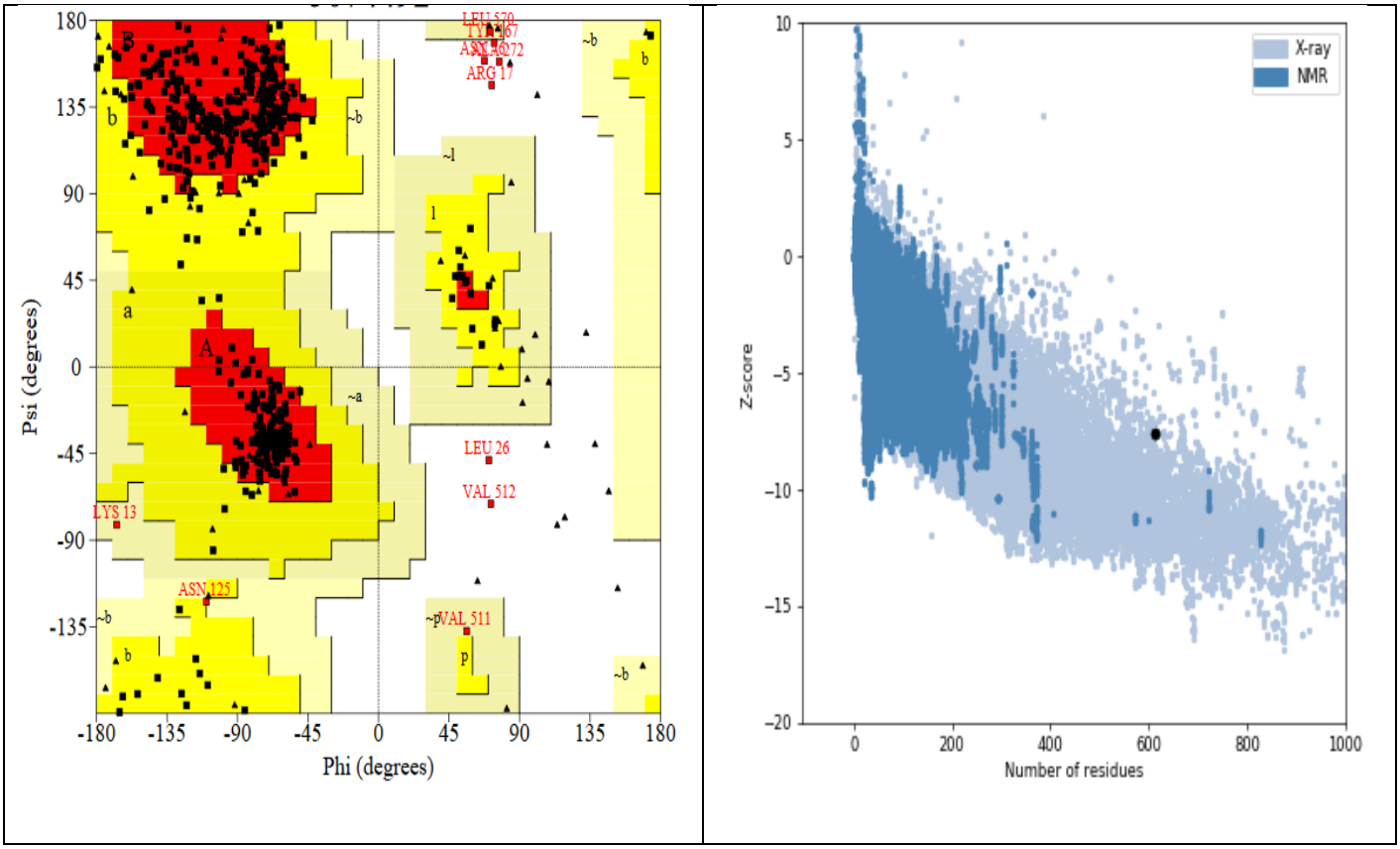
Supplementary Figure S2. The result of the population coverage analysis of the most promising epitopes and their selected MHC alleles.



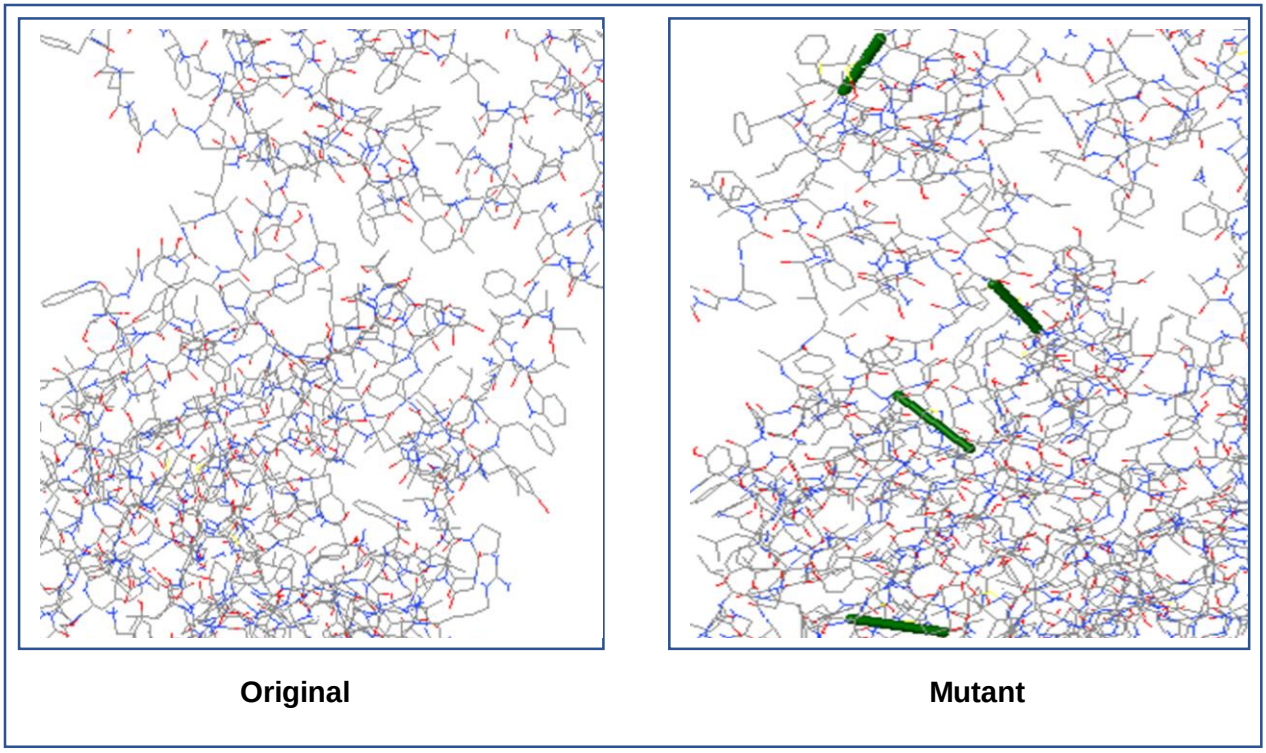
Supplementary Figure S3. The results of the MHC cluster analysis. Here, (a) is the heat map (left) and (b) is the tree map (right) of MHC class-I cluster analysis, (c) is the heat map (left) and (d) is the tree map (right) of MHC class-II cluster analysis.



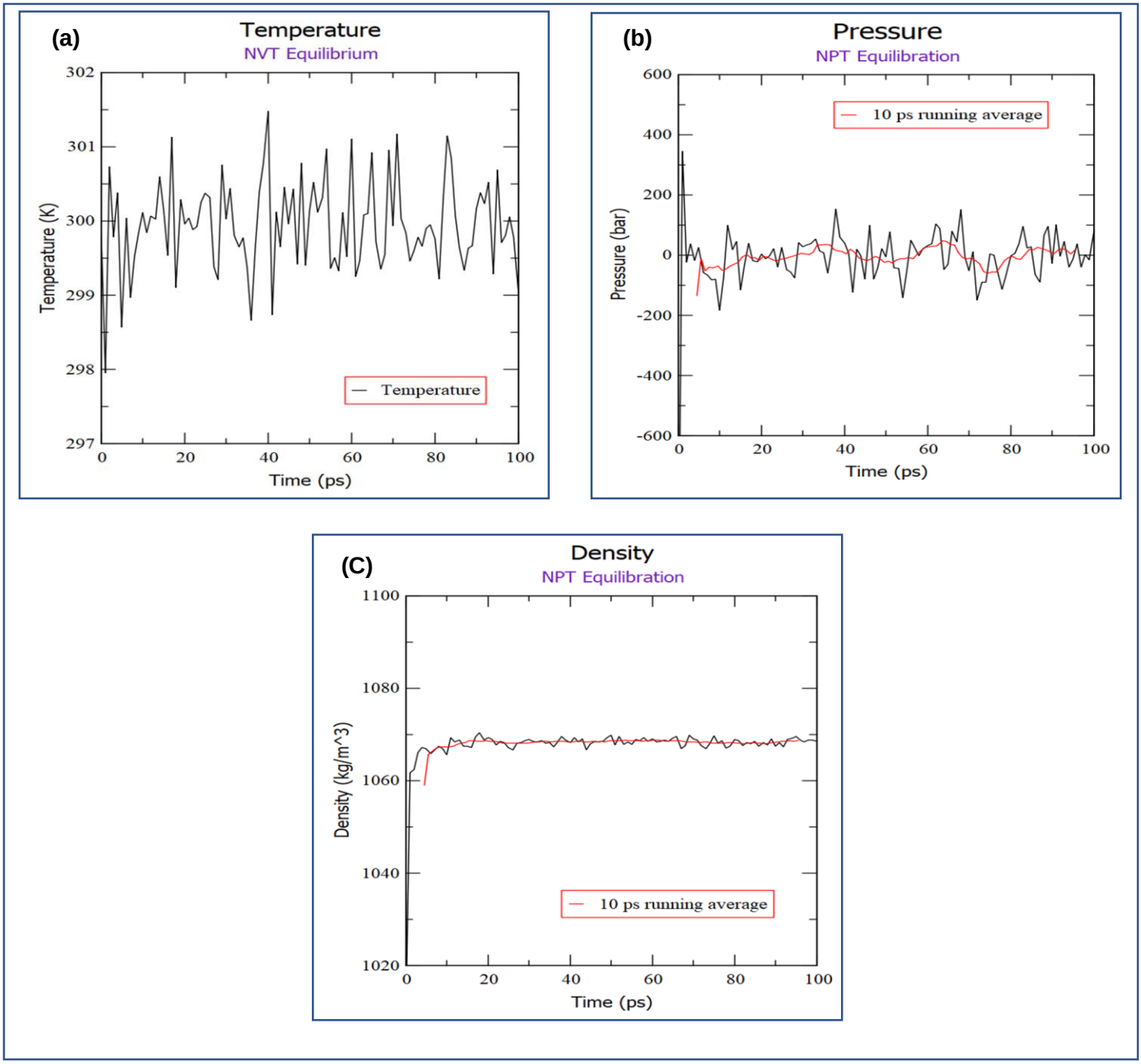
Supplementary Figure S4. The results of the secondary structure prediction of the constructed CV vaccine. (a) PRISPRED prediction, (b) GOR IV prediction, (c) SOPMA prediction, (d) SIMPA96 prediction.



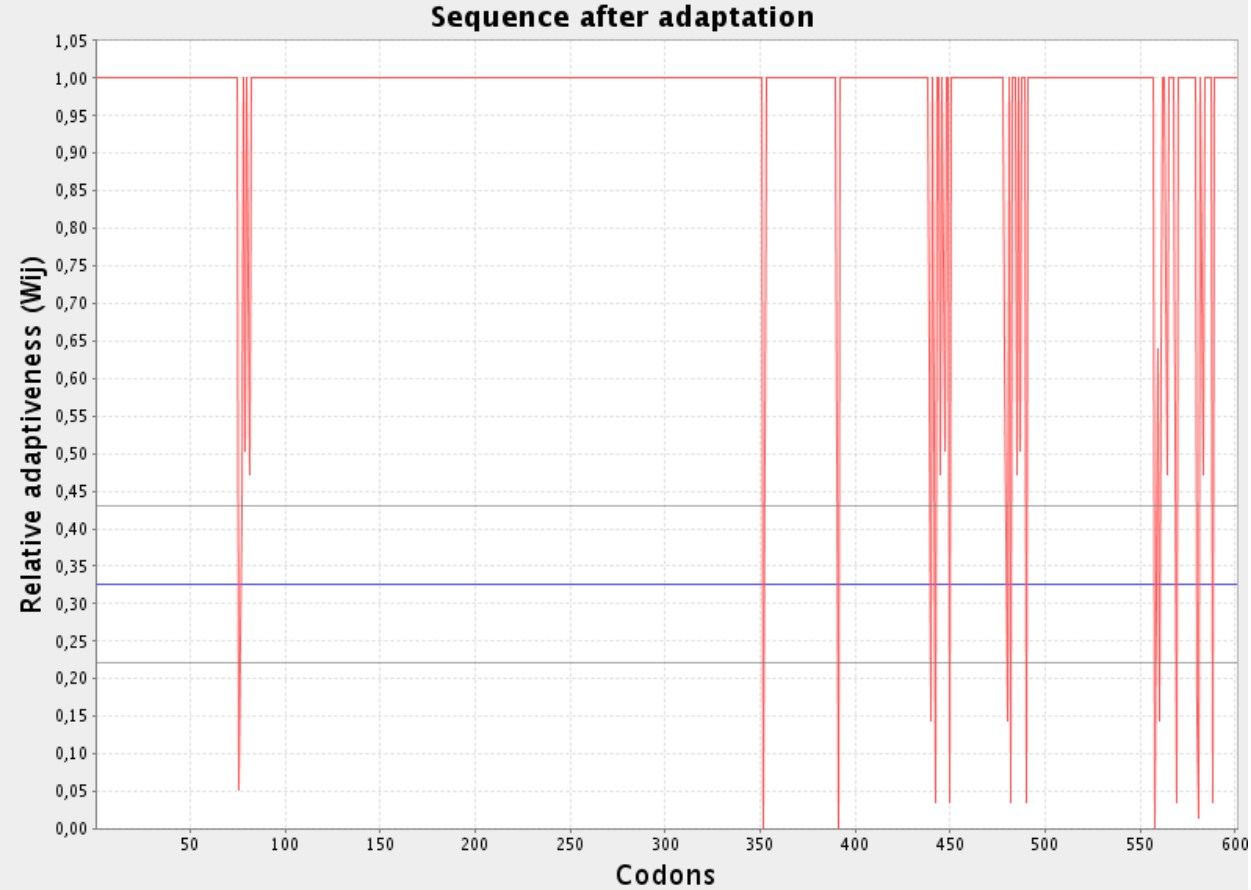
Supplementary Figure S5. The results of the Ramachandran plot analysis generated by PROCHECK server (left) and quality score or z-score graph (right) generated by the ProSA-web server of the refined vaccine construct, CV. In the Ramachandran plots, the orange and deep yellow colored regions are the allowed regions, the light yellow regions are the generously allowed regions and the white regions are the outlier regions and the glycine residues are represented as triangles. All the three vaccine constructs were predicted to have quite good structural quality.



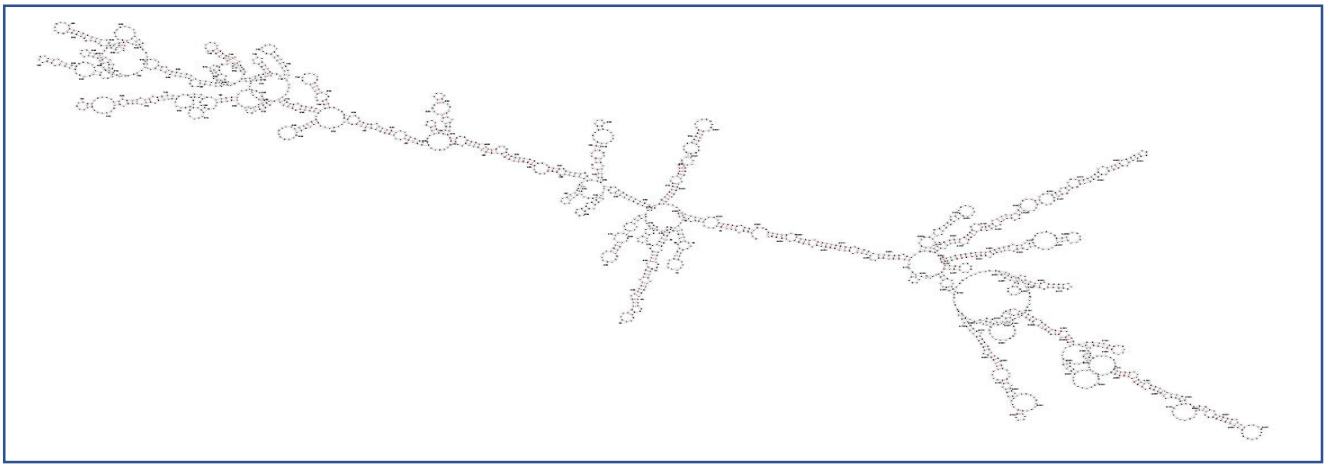
Supplementary Figure S6. The disulfide engineering of the vaccine construct, CV, both the original (left) and mutant (right) forms are shown.



Supplementary Figure S7. (a) Temperature variations during simulation. Temperature had reached 300K and showed minimum fluctuations afterwards. (b) Pressure variations along with a running average with 10 ps window. (c) Density variations over 100 ps with a running average plot with 10 ps window.



Supplementary Figure S8. Figure showing the codon adaptation graph of the CV vaccine.



Supplementary Figure S9. The vaccine mRNA secondary structure predicted by RNAfold server.