

Supplementary

Immunoinformatics Study: Multi-Epitope Based Vaccine Design from SARS-CoV-2 Spike Glycoprotein in Indonesia

Ramadhita Umitaibatin¹, Azza Hanif Harisna², Muhammad Miftah Jauhar², Putri Hawa Syaifie², Adzani Gaisani Arda², Dwi Wahyu Nugroho², Donny Ramadhan³, Etik Mardliyati⁴, Wervyan Shalannanda⁵, and Isa Anshori^{1,*}

¹ Lab-on-chip Group, Department of Biomedical Engineering, School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Jawa Barat 40132, Indonesia

² Nano Center Indonesia, Jl. PUSPIPTEK, South Tangerang, Banten, 15314, Indonesia

³ Research Center for Pharmaceutical Ingredients and Traditional Medicine, National Research and Innovation Agency (BRIN), Cibinong, Bogor, Jawa Barat, 16911, Indonesia

⁴ Research Center for Vaccine and Drug, National Research and Innovation Agency (BRIN), Cibinong, Bogor, Jawa Barat, 16911, Indonesia

⁵ Department of Telecommunication Engineering, School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Jawa Barat 40132, Indonesia

* Correspondence: isaa@staff.stei.itb.ac.id



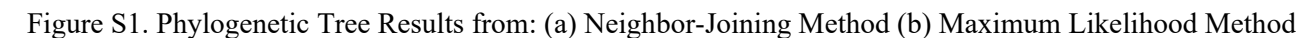


Table S1. Wuhan-Hu-1 and EIJK-61453 epitopes population coverage

Epitope	Cover age %	HLA allele (genotypic frequency (%))																			Total HLA hits
		Class I and II	HLA-A*01:01	HLA-A*02:01	HLA-A*26:01	HLA-A*32:01	HLA-B*07:02	HLA-B*15:01	HLA-B*15:02	HLA-B*40:01	HLA-B*44:01	HLA-B*51:01	HLA-B*58:01	HLA-DPA1*01:03	HLA-DPB1*04:01	HLA-DRB1*04:02	HLA-DRB1*04:03	HLA-DRB1*04:05	HLA-DRB1*14:01	HLA-DRB1*15:01	
			1	3	1	1	2	2	7	1	3	1	1	01:03	04:01	04:02	04:03	04:05	14:01	15:01	
			-2.57	-3.94	-1.21	-0.42	-0.87	-11.91	-0.9	-3.65	-9.1	-3.22	-5.81	0	-14.91	-0.57	-1.1	-1.8	-0.82	-4.53	
Epitope #1: SPRRARSVA	1.71%	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	1	
Epitope #2: KIYSKHTPI	8.48%	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	
Epitope #3: AEIRASANL	23.53%	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	2	
Epitope #4: QLTPTWRVY	22.07%	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	1	
Epitope #5: WTAGAAAYY	9.01%	+	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	3	
Epitope #6: LAIPTNFTI	18.59%	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	-	3	
Epitope #7: GINITRFQTLALHR	11.46%	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	+	3	
Epitope #8: HWFVTQRNFYEPQII	3.40%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	1	
Epitope #9: CTFEYVSQPFLMDLE	27.60%	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	2	
Epitope #10: RFQTLALHRSYLTP	1.56%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	1	
Epitope set	78.26%	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	19	

+ : restricted

- : not restricted

shaded column : genotypic frequency of this allele is 0 (zero)

Table S2. UGM0002 epitopes population coverage

Epitope	Coverag e	HLA allele (genotypic frequency (%))																			Total HLA hits	
		HLA- A*01:0 1	HLA- A*02:0 3	HLA- A*03:0 1	HLA- A*11:0 1	HLA- A*26:0 1	HLA- A*30:0 1	HLA- A*32:0 1	HLA- B*07:0 2	HLA- B*15:0 2	HLA- B*15:1 7	HLA- B*40:0 1	HLA- B*44:0 3	HLA- B*51:0 1	HLA- B*58:0 1	HLA- DPA1*0 1:03	HLA- DPB1*0 4:01	HLA- DRB1* 04:02	HLA- DRB1* 04:03	HLA- DRB1* 04:05		HLA- DRB1* 15:01
		-2.57	-3.94	-2.37	-16.21	-1.21	-1.06	-0.42	-0.87	-11.91	-0.9	-3.65	-9.1	-3.22	-5.81	0	-14.91	-0.57	-1.1	-1.8		-4.53
Epitope #1: SPRRARSVA	1.71%	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	1
Epitope #2: KIYSKHTPI	8.48%	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	2
Epitope #3: AEIRASANL	23.53%	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-	-	-	2
Epitope #4: GVYFASTEK	35.25%	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
Epitope #5: QLTPTWVRVY	22.07%	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	1
Epitope #6: WTAGAAAYY	9.01%	+	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	3
Epitope #7: IAIPTNFTI	18.53%	-	-	-	-	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	-	3
Epitope #8: GINITRFQTLALHR	11.46%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	+	3
Epitope #9: HWFVTQRNFYEPQII	3.40%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	1
Epitope #10: CTFEYVSQPFLMDLE	27.60%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	2
Epitope #11: NCTFEYVSQPFLMDL	27.60%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	2
Epitope set	86.29%	1	1	1	1	1	1	1	1	1	2	1	1	1	1	2	2	1	1	1	1	23

+ : restricted

- : not restricted

shaded column : genotypic frequency of this allele is 0 (zero)

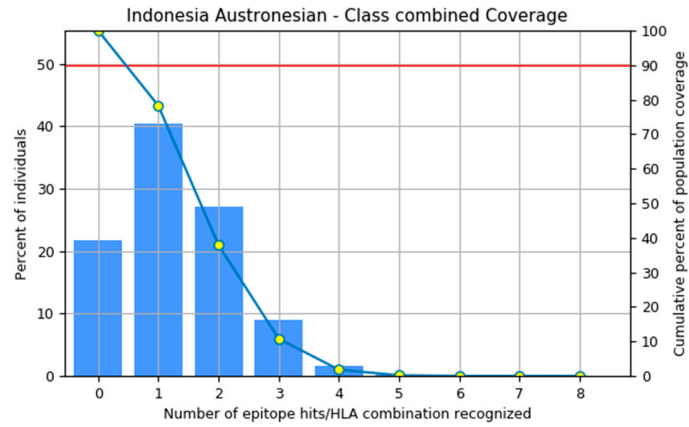
Table S3. B.1.1.7 epitopes population coverage

Epitope	Coverage	HLA allele (genotypic frequency (%))												HLA- DPA1*0 2:01	HLA- DPB1*0 5:01	HLA- DRB1*0 4:02	Total HLA hits
		Class I and II	HLA- A*01:01	HLA- A*02:03	HLA- A*03:01	HLA- A*11:01	HLA- A*26:01	HLA- A*30:01	HLA- A*32:01	HLA- B*15:02	HLA- B*15:17	HLA- B*40:01	HLA- B*44:03				
		-2.57	-3.94	-2.37	-16.21	-1.21	-1.06	-0.42	-11.91	-0.9	-3.65	-9.1					
Epitope #1: KIYSKHTPI	8.48%	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	2
Epitope #2: AEIRASANL	23.53%	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	2
Epitope #3: WTAGAAAYY	9.01%	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	3
Epitope #4: GVYFASTEK	35.25%	-	-	+	+	-	+	-	-	-	-	-	-	-	-	-	3
Epitope #5: QLTPTWRVY	22.07%	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	1
Epitope #6: RAAEIRASANLAATK	1.08%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	1
Epitope #7: KHTPINLVRDLPQGF	1.08%	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	1
Epitope #8: KGIYQTSNFRVQPTE	45.82%	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	2
Epitope set	84.28%	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	15

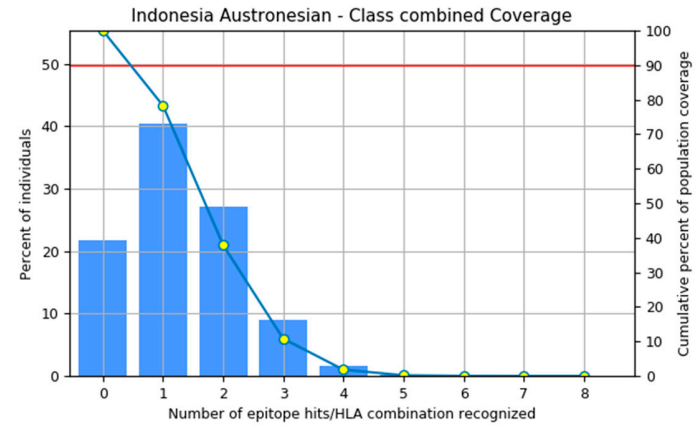
+ : restricted

- : not restricted

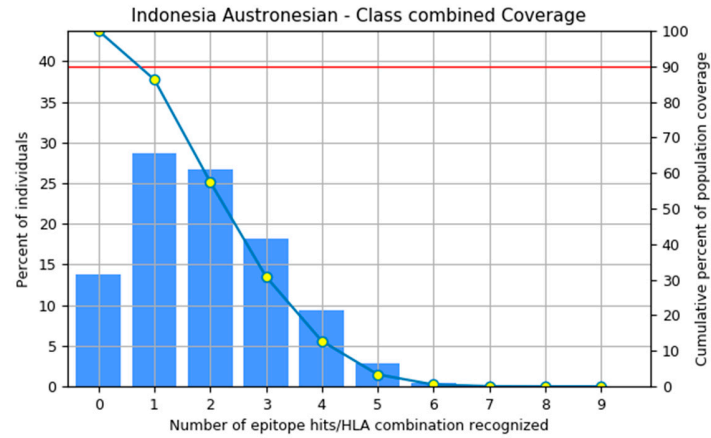
shaded column: genotypic frequency of this
allele is 0 (zero)



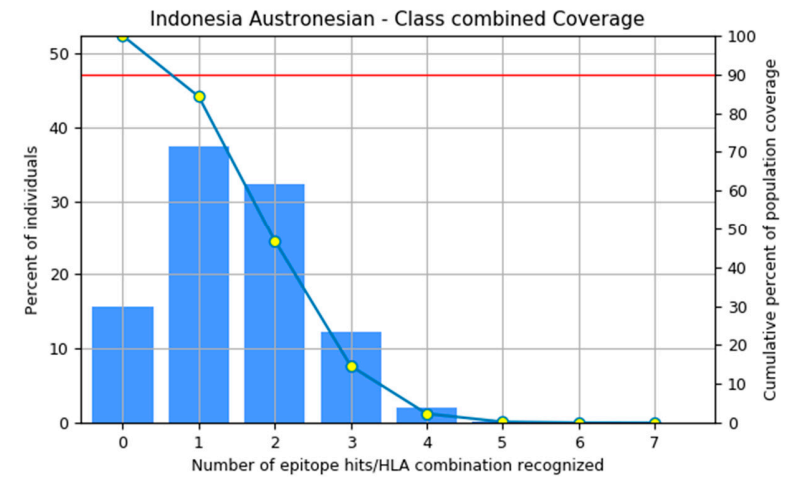
(a)



(b)



(c)



(d)

Figure S2. Number of epitopes recognized by HLA from Indonesia population for Wuhan-Hu-1 (a), EIJK-61453 (b), UGM0002 (c), and B.1.1.7 (d)

Table S4. The predicted discontinuous B-cell epitopes

Epitope	Epitope position
Wuhan-Hu-1	
GSVASQSI	C: G685; C: S686; C: V687; C: A688; C: S689; C: Q690; C: S691; C: I692
FHVNNSYA Y	A: F643; A: H655; A: V656; A: N657; A: N658. A: S659; A: Y660; A: A694; A: Y695
EIJK-61453	
GSVASQSI	C: G685; C: S686; C: V687; C: A688; C: S689. C: Q690; C: S691; C: I692
FHVNNSYA Y	A: F643; A: H655; A: V656; A: N657; A: N658. A: S659; A: Y660; A: A694; A: Y695
UGM0002	
IRSTGSNVG AEHVNNSY	B: I624; B: R634; B: S637; B: T638; B: G639; B: S640; B: N641; B: V642; B: G652; B: A653; B: E654; B: H655; B: V656. B: N657; B: N658; B: S659; B: Y660
HVNNSY	A: H655; A: V656; A: N657; A: N658; A: S659; A: Y660
B.1.1.7	
IRSTGSNVG AEHVNNSY	B: I624; B: R634; B: S637. B: T638; B: G639; B: S640; B: N641; B: V642. B: G652; B: A653; B: E654; B: H655; B: V656; B: N657; B: N658; B: S659; B: Y660
HVNNSY	A: H655; A: V656; A: N657; A: N658; A: S659; A: Y660

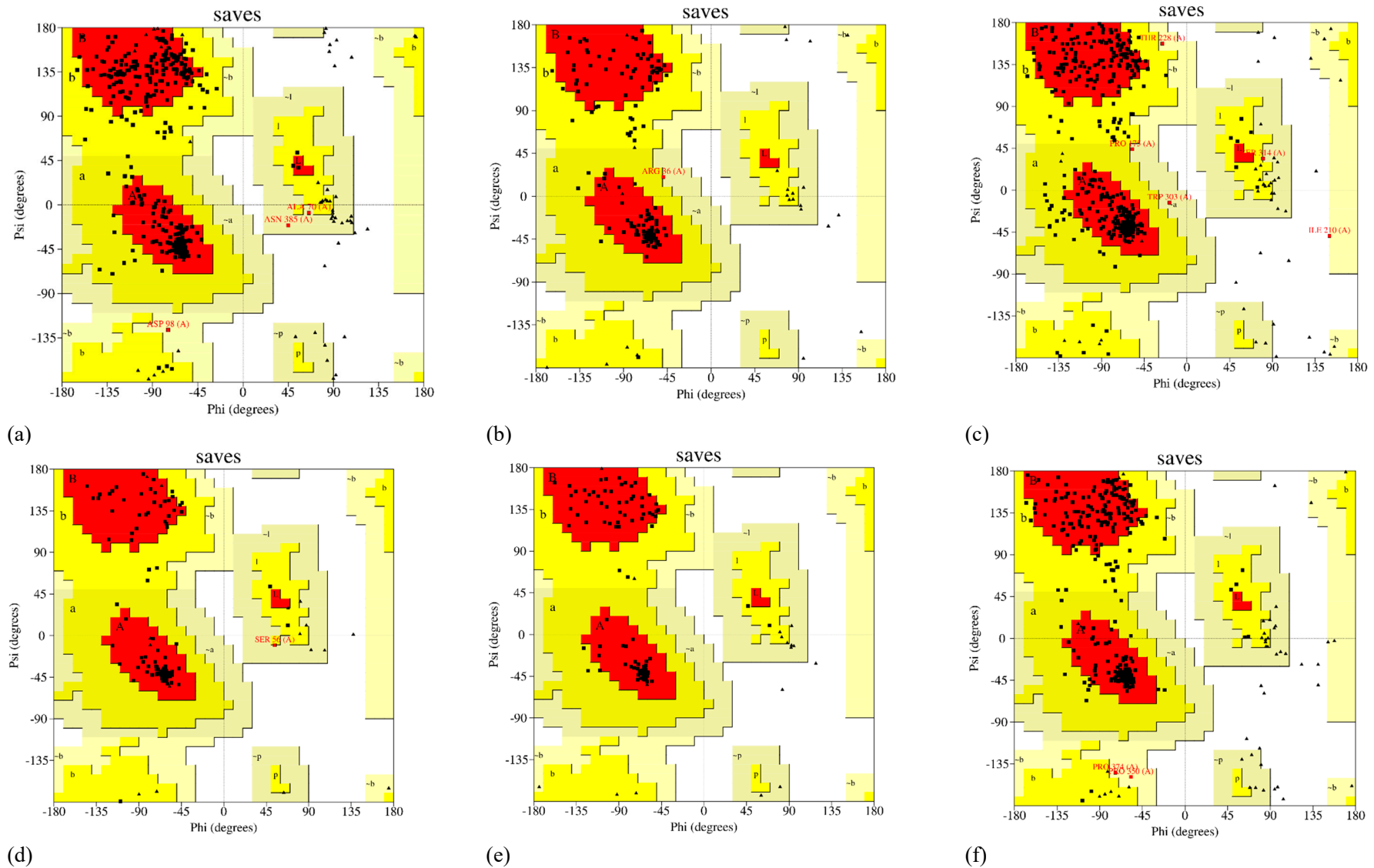


Figure S3. Ramachandran plot for (a) Wuhan-Hu-1-2 and EIJK-61453-2 with 454 aa, 88.2% in the most-favorable region, 59 Glycine residues, and 38 Proline residues.

(b) Wuhan-Hu-1-3 and EIJK-61453-3 with 189 AA, 86.7% in the most favorable region, 26 Glycine residues, and 18 Proline residues

(c) UGM0002-2 with 493 AA, 84.8% in the most favorable region, 17 Glycine residues, and 9 Proline residues

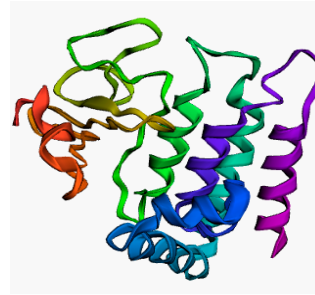
(d) UGM0002-3 with 454 AA, 88.2% in the most favorable region, 59 Glycine residues, and 38 Proline residues

(e) B.1.1.7-1 with 140 AA, 93.8% in the most favorable region, 17 Glycine residues, and 9 Proline residues

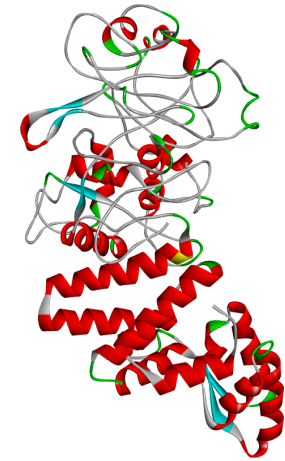
(f) B.1.1.7-1-3 with 395 AA, 86.4% in the most favorable region, 49 Glycine residues, and 35 Proline residues



(a)



(b)



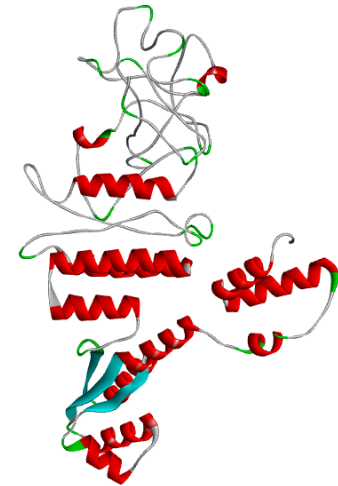
(c)



(d)



(e)



(f)

Figure S4. 3D structure prediction of (a) Wuhan-Hu-1-2 and EIJK-61453-2 vaccine, (b) Wuhan-Hu-1-3 and EIJK-61453-3 vaccine (c) UGM0002-2, (d) UGM0002-3, (e) B.1.1.7-1-1, and (e) B.1.1.7-1-3

Table S5. Top three list of vaccine construction sequences

Model	Vaccine Sequence	Score	Explanation
Wuhan-Hu-1			
1	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAA KSPRRARSVAKKWTAGAAAYYGPGPGGINITRFQTLALHRGPGPGCTFEYVSQPFLMDLEGPGPGQS KRVDFC	Quality Factor ERRAT = 91,73 TM-Score = 0,383 Template = 1KJ6_A Confidence = 97,0%	Using the template protein 1KJ6_A with 97.0% confidence Vaccine length 146 aa
2	MAKLSTDELDDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAA GDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLEKVAKAADEAKAKLEAAGATVTVKEAAAKKIYSK HTPIAAYWTAGAAAYYAAYQLTPTWRVYAAYSPRRARSVAAYAEIRASANLPGPGPGCTFEYVSQPFLM DLEGPGPGGINITRFQTLALHRGPGPGIIAYTMSLGAENSVGPGPGHWFVTQRNFYEPQIIGPGPGFSNVT WFHAIHVS GTNGTKRFDNGPGPGNITNLCPGGEVFNATRFASVYAWNRRKPGPGPGNSASFSTFKCYGVSP KLNDLCFTNVGPGPGGDEV RQIAPGQTGKIADYNYKPGPGPGNNLDSKVGGNYNYGPGPGFQPTNGPG PGGQSKRVDFCGPGPGSCCKFDEDDSEPVLKGVKL	Quality Factor ERRAT = 83,18 TM-Score = 0,148 Template = 6YLA Confidence = 100%	Using the 6YLA protein template with 100% confidence Vaccine length 454 aa
3	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAA KSPRRARSVAAYKIYSKHTPIGPGPGGINITRFQTLALHRGPGPGCTFEYVSQPFLMDLEGPGPGFQPTN GGPGPGAYTMSLGAENSVAYSNGPGPGSCCKFDEDDSEPVLKGVKL	Quality Factor ERRAT = 85,48 TM-Score = 0,197	Using a de novo folding approach Vaccine length 189 aa
EIJK-61453			
1	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAA KSPRRARSVAKKWTAGAAAYYGPGPGGINITRFQTLALHRGPGPGCTFEYVSQPFLMDLEGPGPGQS KRVDFC	Quality Factor ERRAT = 91,73 TM-Score = 0,383 Template = 1KJ6_A Confidence = 97,0%	Using the template protein 1KJ6_A with 97.0% confidence Vaccine length 146 aa
2	MAKLSTDELDDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAA GDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLEKVAKAADEAKAKLEAAGATVTVKEAAAKKIYSK HTPIAAYWTAGAAAYYAAYQLTPTWRVYAAYSPRRARSVAAYAEIRASANLPGPGPGCTFEYVSQPFLM DLEGPGPGGINITRFQTLALHRGPGPGIIAYTMSLGAENSVGPGPGHWFVTQRNFYEPQIIGPGPGFSNVT WFHAIHVS GTNGTKRFDNGPGPGNITNLCPGGEVFNATRFASVYAWNRRKPGPGPGNSASFSTFKCYGVSP KLNDLCFTNVGPGPGGDEV RQIAPGQTGKIADYNYKPGPGPGNNLDSKVGGNYNYGPGPGFQPTNGPG PGGQSKRVDFCGPGPGSCCKFDEDDSEPVLKGVKL	Quality Factor ERRAT = 83,16 TM-Score = 0,148 Template = 6YLA Confidence = 100%	Using the 6YLA protein template with 100% confidence Vaccine length 454 aa
3	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAA KSPRRARSVAAYKIYSKHTPIGPGPGGINITRFQTLALHRGPGPGCTFEYVSQPFLMDLEGPGPGFQPTN GGPGPGAYTMSLGAENSVAYSNGPGPGSCCKFDEDDSEPVLKGVKL	Quality Factor ERRAT = 85,48 TM-Score = 0,197	Using a de novo folding approach Vaccine length 189 aa
UGM0002			
1	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKKEAAA KSPRRARSVAAYKIYSKHTPIAAYGVYFASTEKAAYWTAGAAAYYGPGPGGINIFQTLALHRTRGPGP GGQSKRVDFC	Quality Factor ERRAT = 99,27 TM-Score = 0,358	Using a de novo folding approach Vaccine length 151 aa

2	MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAA GDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVK EAAAK GVYF ASTEK AAY KIYSKHTPI AAY WTAGAAAYY AAY QLTPTWRVY AAY SPRRARSVA AAY IAIPTNFTI AAY AEI RASANL GPGPG SQCVNLTTTRTQLPPAYTNSFTRGVY GPGPG FSNVTWFHAIHVSGTNGTKRFDNP GPGPG NITNLCPFGEVFNATRFASVYAWNRK GPGPG SASFSTFKCYGVSPTKLNDLCFTNV GPGPG VIRGDEVROI APGQTGKIADYNYKL GPGPG NNLDSKVGGNYNY GPGPG SNKKFLPF GPGPG TNTSNQ GPGPG VNCTEV GPGPG GQSKRVDFC GPGPG CTFEYVSQPFLMDLE GPGPG GINITRFQTLALHR GPGPG HWFVTQRNFYE PQII	Quality Factor ERRAT = 75,74 TM Score = 0,150 Template = 6XMO Confidence = 100%	Using the 6XMO protein template with 100% confidence Vaccine length 493 aa
3	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKK EAAA K SPRRARSVA K KKIYSKHTPI K KGVYFASTEK K KWTAGAAAYY GPGPG GINIFQTLALHRTR GPGPG GQ SKRVDFC	Quality Factor ERRAT = 94,93 TM Score = 0,358	Using a de novo folding approach Vaccine length 148 aa
B.1.1.7			
1	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKK EAAA K KIYSKHTPI AAY WTAGAAAYY AAY GVYFASTEK GPGPG KGIYQTSNFRVQPTE GPGPG LGQSKRVDFC	Quality Factor ERRAT = 94,21 TM-Score = 0,361 Template = 2LWL_A Confidence = 97,7%	Using the 2LWL_A protein template with 97.7% confidence Vaccine length 140 aa
2	MRIHYLLFALLFLFLVPVPGHGGIINTLQKYYCRVRGGRCVLSCLPKEEQIGKCSTRGRKCCRRKK EAAA K KIYSKHTPI AAY WTAGAAAYY GPGPG RAAEIRASANLAATK GPGPG LGQSKRVDFC	Quality Factor ERRAT = 97,32 TM-Score = 0,405 Template = 6VSJ_C Confidence = 100%	Using the 6VSJ_C protein template with 100% confidence Vaccine length 128 aa
3	MAKLSTDELLDAFKEMTLLELSDFVKKFEETFEVTAAAPVAVAAAGAAPAGAAVEAAEEQSEFDVILEAA GDKKIGVIKVVREIVSGLGLKEAKDLVDGAPKPLLEKVAKEAADEAKAKLEAAGATVTVK EAAAK GVYF ASTEK AAY KIYSKHTPI AAY WTAGAAAYY AAY QLTPTWRVY AAY AEIRASANL GPGPG KHTPINLVRDLP QGF GPGPG KGIYQTSNFRVQPTE GPGPG RAAEIRASANLAATK GPGPG SQCVNLTTTRTQLPPAYTNSFTRG VY GPGPG YNASASFSTFKCYGVSPTKLNDLCFT GPGPG GDEVROIAPGQTGKIA GPGPG SNKKFLPF GPGP G VNCTEV GPGPG LGQSKRVDFC GPGPG SCCKFDEDDSEPVLGKVK	Quality Factor ERRAT = 87,89 TM-Score = 0,22	Using a de novo folding approach Vaccine length 395 aas

Table S6. Optimized vaccine sequences

EIJK
Pasted Sequence: GC=68.72%, CAI=0.58
ATGAGGATCCACTACCTGCTGTTTCGCCCTGCTGTTTCCTGTTCTGGTGCCCGTGCCCGGCCACGGCGGCATCATCAACACCCTGCAGAAGTACTAC TGCAGGGTGAGGGGCGGCAGGTGCGCCGTGCTGAGCTGCCTGCCCCAAGGAGGAGCAGATCGGCAAGTGCAGCACCAGGGGCAGGAAGTGCTGCA GGAGGAAGAAGGAGGCGCCGCCAAGAGCCCCAGGAGGGCCAGGAGCGTGGCCAAGAAGTGGACCGCCGGCGCCGCCGCCTACTACGGCCCCG GCCCCGGCGGCATCAACATCACCAGGTTCCAGACCCTGCTGGCCCTGCACAGGGGGCCCCGGCCCCGGCTGCACCTTCGAGTACGTGAGCCAGCCC TTCCTGATGGACCTGGAGGGCCCCGGCCCCGGCGGCCAGAGCAAGAGGGTGGACTTCTGC
Improved DNA[1]: GC=59.82%, CAI=0.93
ATGCGTATTCACTACCTGCTGTTTGGCCCTGCTGTTTCTGTTTCTGGTGCCGGTGCCGGGCCATGGCGGCATTATTAACACCCTGCAGAAATATTATT GCCGCGTGCGCGGTGGCCGCTGCGCCGTTCTGAGCTGTCTGCCGAAAGAAGAACAGATCGGTAAATGTAGCACCCGCGGCCGTAAATGCTGCCGT CGTAAAAAAGAAGCGGCGGCGAAATCACCGCGCCGTGCCCGTAGCGTGGCAAAAAAATGGACCGCCGGCGCCGCGGCGTACTATGGCCCCGGGTC CGGGCGGCATTAACATTACCCGCTTTCAGACCCTGCTGGCCCTGCATCGCGGTCCGGGGCCCGGGCTGCACCTTTGAATATGTGTGCGAGCCGTTTC TGATGGATCTGGAAGGCCCGGGCCCCGGGCGGCCAAAGCAAACGCGTTGATTTTTGT
UGM
Pasted Sequence: GC=68.21%, CAI=0.59
ATGAGGATCCACTACCTGCTGTTTCGCCCTGCTGTTTCCTGTTCTGGTGCCCGTGCCCGGCCACGGCGGCATCATCAACACCCTGCAGAAGTACTAC TGCAGGGTGAGGGGCGGCAGGTGCGCCGTGCTGAGCTGCCTGCCCCAAGGAGGAGCAGATCGGCAAGTGCAGCACCAGGGGCAGGAAGTGCTGCA GGAGGAAGAAGGAGGCGCCGCCAAGAGCCCCAGGAGGGCCAGGAGCGTGGCCGCCGCCTACAAGATCTACAGCAAGCACACCCCCATCGCCGC CTACGGCGTGTACTTCGCCAGCACCGAGAAGGCCGCCTACTGGACCGCCGGCGCCGCCGCCTACTACGGCCCCGGCCCCGGCGGCATCAACATCT TCCAGACCCTGCTGGCCCTGCACAGGACCAGGGGCCCCCGGCCCGGCGGCCAGAGCAAGAGGGTGGACTTCTGC
Improved DNA[1]: GC=59.82%, CAI=0.95
ATGCGCATTCATTATCTGCTGTTTGCCTGCTGTTTCTGTTTCTGGTGCCGGTGCCGGGCCATGGCGGCATTATTAACACCCTGCAGAAATACTATT GCCGCGTGCGTGCGGCCGCTGCGCCGTGCTGAGCTGCCTGCCGAAAGAAGAACAGATTGGCAAATGCAGCACCCGCGGCCGTAAATGTTGCCGC CGCAAAAAAGAAGCGGCCGCGAAAAGCCCGCGCCGCGCACGTAGCGTGGCCGCGGCATATAAAATTTACAGCAAACATACCCCGATTGCGGCCT ACGGCGTGTACTTTGCGTCAACCGAAAAAGCGGCCTATTGGACCGCGGGCGCGGCGGCCTATTATGGTCCTGGCCCCGGGCGGCATTAACATTTTTTC AGACCCTGCTGGCGCTGCATCGCACCCGCGGCCCGGGCCCCGGGTGGCCAGAGTAAACGTGTGGATTTTTGC
B117

Pasted Sequence: GC=68.75%, CAI=0.59

ATGAGGATCCACTACCTGCTGTTTCGCCCTGCTGTTTCCTGTTTCCTGGTGCCCGTGCCCGGCCACGGCGGCATCATCAACACCCTGCAGAAGTACTAC
TGCAGGGTGAGGGGCGGCAGGTGCGCCGTGCTGAGCTGCCTGCCCAAGGAGGAGCAGATCGGCAAGTGCAGCACCCAGGGGGCAGGAAGTGCTGCA
GGAGGAAGAAGGAGGCCGCCGCCAAGAAGATCTACAGCAAGCACACCCCCATCGCCGCCTACTGGACCGCCGGCGCCGCCGCTACTACGGCCC
CGGCCCCGGCAGGGCCGCCGAGATCAGGGCCAGCGCCAACCTGGCCGCCACCAAGGGCCCCGGCCCCGGCCTGGGCCAGAGCAAGAGGGTGGAC
TTCTGC

Improved DNA[1]: GC=59.64%, CAI=0.95

ATGCGTATTCATTACCTGCTGTTTTCGCTGCTGTTTCTGTTTCTGGTGCCGGTGCCGGGTCATGGCGGTATTATTAACACCCTGCAGAAATATTATT
GTCGCGTGCGTGGCGGCCGCTGTGCGGTTCTGAGCTGCCTGCCGAAAGAAGAACAGATTGGTAAATGCAGCACCCGCGGCCGCAAATGCTGCCGT
CGCAAAAAAGAAGCGGCGGCGCAAAAAAATTTATTCGAAACATACCCCGATTGCCGCCTATTGGACCGCGGGTGCGGCCGCGTACTATGGCCCGG
GCCCCGGGCCGCGCGGCGGAAATCCGCGCGAGCGCAAACCTGGCGGCGACCAAAGGCCCGGGCCCCGGGCCTGGGCCAGAGCAAACGCGTTGATTT
TTGC