

Supplementary Information

Immunoinformatics method to design universal multi-epitope nanoparticle vaccine for TGEV S protein

Shinian Li^{1,3}, Jingjing Yu², Chencheng Xiao^{2*}, Yaling Li^{1*}

Affiliations :

1: College of Animal Science and Technology & College of Veterinary Medicine of Zhejiang A&F University, Hangzhou, China, 311300

2: College of Animal Science and Technology, Shihezi University, Shihezi, China, 832003

3: Shenzhen New Industries Biomedical Engineering Co., Ltd., Shenzhen ,China, 518122

*** Correspondence:**

Chencheng Xiao

Email address: 305639447@qq.com

Affiliations: College of Animal Science and Technology, Shihezi University, Shihezi, China, 832003

Yaling Li

Email address: 4578187@qq.com;

Affiliations: College of Animal Science and Technology & College of Veterinary Medicine of Zhejiang A&F University, Hangzhou, China, 311300

Supplemental Table S1. Amino acid sequence of TGEV-S protein for candidate vaccine analysis

ID	Length	Amino acid sequence
ACN71 196.1	1448 aa	MKKLFVVLVVMPLIYGDNFPCSKLTNRTIGNHWNLIETFLNYSRLPPNSDAVLGDYFPTVQPWFNCIRNNSNDLYVT LENLKALYWDAIENITSNHKQRLNVVNGYPYSITVTTRNFNSAEGAIHCICKGSPPTTTTESSLTCNWGSECLRNHK FPICPSNSEANCGNMLYGLQWFADAVVAYLHGASYRISFENQWSGTVTLGDMRATTLETAGTLVDLWWFNPVYDVS YYRVNNKNGTTVVSINCTQCASYVANVFTTQPGGFIPSDFSFNNWFLLTNSSTLVSGKLVTQPLLNVCLWPVPSFEEA ASTFCFEGAGFDQCNGAVLNNTVDVIRFNLNFTTNVQSGKGATVFSLNTTGGVTLTLEISCYNDTVSDSSFSYGEMPFGV TDGPRYCYVLYNGTALKYLGTLPPVKEIAISKWGHFYINGYNFFSTFPIDCISFNLTTGSDVFWTIAYSYTEALVQV ENTAITKVITYCNSYVNNIKCSQLTANLNGFYVSSSEVGLVNKSVVLLPSFYTHITIVNITIGLGMKRSGYGQPIASTLSN ITLPMQDDNTDVYCIQSDQFSVYVHSTCKSALWDNVFKRNCTDVLDAVIAVKTGTCPFSAKLNNYLTFNKFCLSLSPV GANCKFDVAARTRTNDQVVRSLYVIYEEDNIVGVPSDNLHDLVSLHLDSCDTYNIYGRGTGVGIRQTNRITLLSGLY YTSLSGDLGFKNVSDGVIYSVTPCDVIAQAVIDGTIVGAITSINSELLGLTHWTTTNPFYYSIYNYTNDTRGTAIDS NDDCEPVITYSNIGVCKNGALVFINVTHSDGDVQPISTGNVTIPTNFTISVQVEYIQVYTTTPSIDCSRYVCNGNPRCNKL LTQYVSACQTIEQALAMGARLENMEVDSMLFVSENALKLASVEAFNSSETLDPIYKEWPNIGGSWLEGLKYILPSDNSK RKYRSAIEDLLFAKVVTSGLGTVDDEDYKRCTGGYDIADLVCAQYYNGIMVLPGVANADKMTMYTASLAGGITLGALG GGAVAIPFAVAVQARLNYVALQTDVLNKNQQLASAFNQAINITQSFQKVNDAIHQTSRGLATVAKALAKVQDVVN AQGQALSHLTVQLQNNFQAISISSIDIYNRLDELSADAQVDRLITGRLTALNAFVSQTLTRQAEVRASRLAKDKVNEC VRSQSQRFGFCGNGTHLFLSLANAAPNGMIFHTLLPTAYETVTAWAGICALDGRFTGLVVKDVQLTLFRNLDDKFY LTPRTMYQPRVATSSDFVQIEGCDVLFVNATVSDLPISHPDYIDINQTVQDILENFRPNWTVPELTFDIFNATYLNLTGEID DLEFRSEKLHNTTVELAILDINNTLVNLEWLNRIETYVKWPWYVWLLIGLVVIFCIPLLFFCCSTGCCGCGICLGGSCC HSICSRQFENYEPIEKVHVH

Supplementary Table S2. GenBankID of 45 TGEV-S proteins for conservative analysis

>WBU87469.1	>UOZ96074.1	>QQY96587.1	>AAB30949.1	>AAM18703.1
>AAA46909.1	>sp P07946.2	>sp P33470.2	>sp Q01977.1	>ASV64429.1
>ASV64420.1	>ASV64411.1	>ASV64393.1	>ASV64375.1	>ASV64384.1
>ASV64358.1	>ASV64340.1	>ASV64349.1	>ASV64322.1	>ASV64331.1
>ASV64304.1	>ASV64295.1	>ASV64286.1	>ASV64277.1	>WOV89918.1
>WGG26263.1	>UC94274.1	>QK94055.1	>AKA60054.1	>UV20557.1
>UVJ89042.1	>AAC96004.1	>AFZ88844.1	>AAT00645.1	>AAQ02624.1
>AAY22404.1	>AAG30227.1	>ACJ63242.1	>AAA47108.1	>ASV64438.1
>ASV64367.1	>ASV64313.1	>CAB91145.1	>WCF76184.1	>AAQ02623.1

Supplementary Table S3. Immunogenicity and conserved analysis of T and B cell dominant epitopes of S protein

Name	Sequences	Immunogenicity	Toxicity	Length	Minimum identity	Maximum identity
CTL-1	EMPFGVTDGPRYCY	0.76475	Non-Toxin	16	85.71%	100.00%
CTL-2	NAAPNGMIFFHTLL	0.66629	Non-Toxin	16	92.86%	100.00%
CTL-3	TVPELTFDIFNATY	0.55296	Non-Toxin	14	85.71%	100.00%
Th-1	THTIVNITIGLGMKR	0.45483	Non-Toxin	9	80%	100.00%
Th-2	LDPIYKEWPNIGGSW	0.3673	Non-Toxin	16	86.67%	100.00%
Th-3	RQFENYEPIEKVHVH	0.32293	Non-Toxin	15	93.33%	100.00%
B-1	VLNNTVDVIRF	0.30626	Non-Toxin	15	90.91%	100.00%
B-2	DVFWTIAYT	0.28351	Non-Toxin	10	77.78%	100.00%
B-3	AVAIPFAVAVQARLNY	0.2831	Non-Toxin	14	93.75%	100.00%
B-4	LPTAYETVTAWAGICA	0.27241	Non-Toxin	11	93.75%	100.00%
B-5	NTTVELAILI	0.27157	Non-Toxin	14	90%	100.00%
B-6	YVKWPWYVWLLIGLVV	0.25732	Non-Toxin	15	100%	100.00%

Supplement Table S4. The amino acid sequence of the designed vaccine candidate

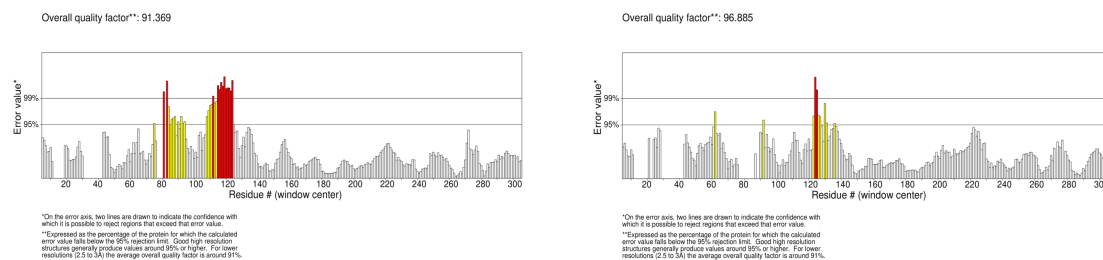
Vaccine construct	Adjuvant	Amino acid sequence
Th-AAY-CTL-KK- B-EAAAK-Ferritin	Ferritin	THTIVNITIGLGMKRGPGPLDPIYKEWPNIGGSWGPG PGRQFENYEPIEKVHVHAAYEMPFGVTDGPRYCYAAY NAAPNGMIFFHTLLAAYTVPELTFDIFNATYKKVLNNT VDVIRFKKDVFWTIAYTKKAVAIPFAVAVQARLNYKK LPTAYETVTAWAGICAKKNTTVELAILIKKYVKWPWY VWLLIGLVVEAAAKMLSKDIKLLNEQVNKEMNSSNL YMSMSSWCYTHSLDGAGLFLFDHAAEEYEHAKKLIIF LNENNVPVQLTSISAPEHKFEGLTQIFQKAYEHEQHISE SINNIVDHAISKDHATFNFLQWYVAEQHEEEVLFKDI LDKIELIGNENHGLYLADQYVKGIKSRKS

Supplementary Table S5. Galaxy Refinement Scores of predicted structure models of candidates vaccine

	Model	GDT-HA	RMSD	MolProbity	Clash score	Poor rotamers	Rama favored
	Initial	1	0	1.639	4.2	0	93.2
	MODEL 1	0.9626	0.408	1.691	7	0	95.6
Candidates vaccine	MODEL 2	0.9565	0.417	1.741	8	0.3	95.6
	MODEL 3	0.964	0.382	1.798	8.9	0	95.4
	MODEL 4	0.9552	0.413	1.704	7.7	0.3	95.9
	MODEL 5	0.9538	0.418	1.727	8.7	0	96.2
Fnal choice is MODEL 5							

Supplement Table S6. Supplement table S3. Nucleotide sequence of candidate vaccine amino acid sequence after reverse translation and optimization of codon bias in Escherichia coli K12

Strain	Length	Nucleotide sequence
E. coli K12	1104 bp	ACCCACACCATCGTTAACATCACCATCGGTCTGGGTATGAAACGTGGTCCGGGTCCGGGTCTGGA CCCGATCTACAAAGAATGGCCGAACATCGGTGTTCTTGGGGTCCGGGTCCGGGTCTGTCAGTTTCG AAAACTACGAACCGATCGAAAAAGTTCACGTTACGCTGCTTACGAAATGCCGTTCCGGTGTACC GACGGTCCGCGTTACTGCTACGCTGCTTACAACGCTGCTCCGAACGGTATGATCTTCTCCACACC CTGCTGGCTGCTTACACGTTCCGGAACGTGACCTTCGACATCTTCAACGCTACCTACAAAAAAGTT CTGAACAACACCGTTGACGTTATCCGTTTCAAAAAAGACGTTTTCTGGACCATCGCTTACACCAA AAAAGCTGTTGCTATCCCGTTTCGCTGTTGCTGTTTCAGGCTCGTCTGAACTACAAAAAAGTCCGA CCGCTTACGAAACCGTTACCGCTTGGGCTGGTATCTGCGCTAAAAAAGACACCGTTGAACTG GCTATCCTGATCAAAAAATACGTTAAATGGCCGTGGTACGTTTGGCTGCTGATCGGTCTGGTTGTT GAAGCTGCTGCTAAAATGCTGTCTAAAGACATCATCAAAGTCTGAAACGAACAGGTTAACAAAG AAATGAACTCTTCTAACCTGTACATGTCTATGTCTTCTTGGTGCTACACCCACTCTCTGGACGGTG CTGGTCTGTTCTGTTTCGACCACGCTGCTGAAGAATACGAACACGCTAAAAAAGTATCATCTTC CTGAACGAAAAACGTTCCGGTTCAGCTGACCTCTATCTCTGCTCCGGAACACAAATTCGAAGG TCTGACCCAGATCTTCCAGAAAGCTTACGAACACGAACAGCACATCTCTGAATCTATCAACAACA TCGTTGACCACGCTATCAAATCTAAAGACCACGCTACCTTCAACTTCCTGCAGTGGTACGTTGCTG AACAGCACGAAGAAGAAGTTCTGTTCAAAGACATCCTGGACAAAAATCGAACTGATCGGTAACGA AAACCACGGTCTGTACCTGGCTGACCAGTACGTTAAAGGTATCGCTAAATCTCGTAAATCT



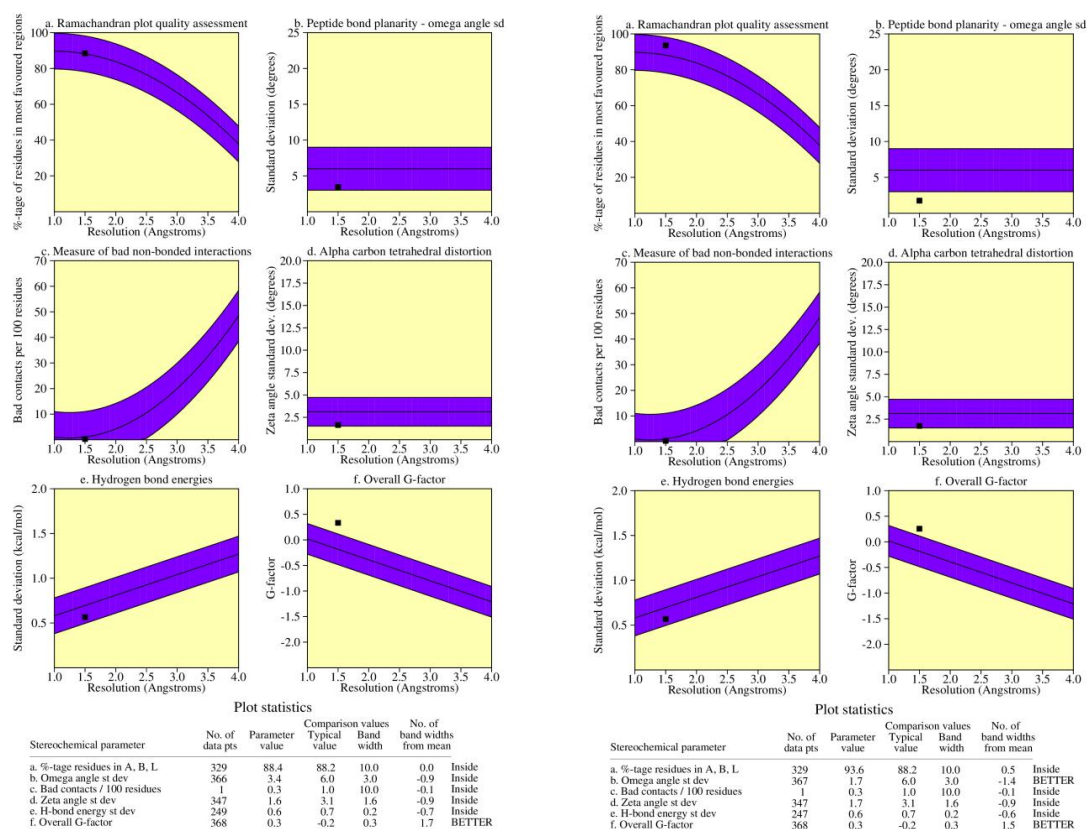
Note: The left side is the crude model; On the right is the refined model (The ERRAT score increased from 91% to 96%, revealing that the model is very close to the ideal state and has high confidence).

Supplementary Fig S1. Detailed comparison of ERRAT scores in tertiary structure models of candidate vaccines



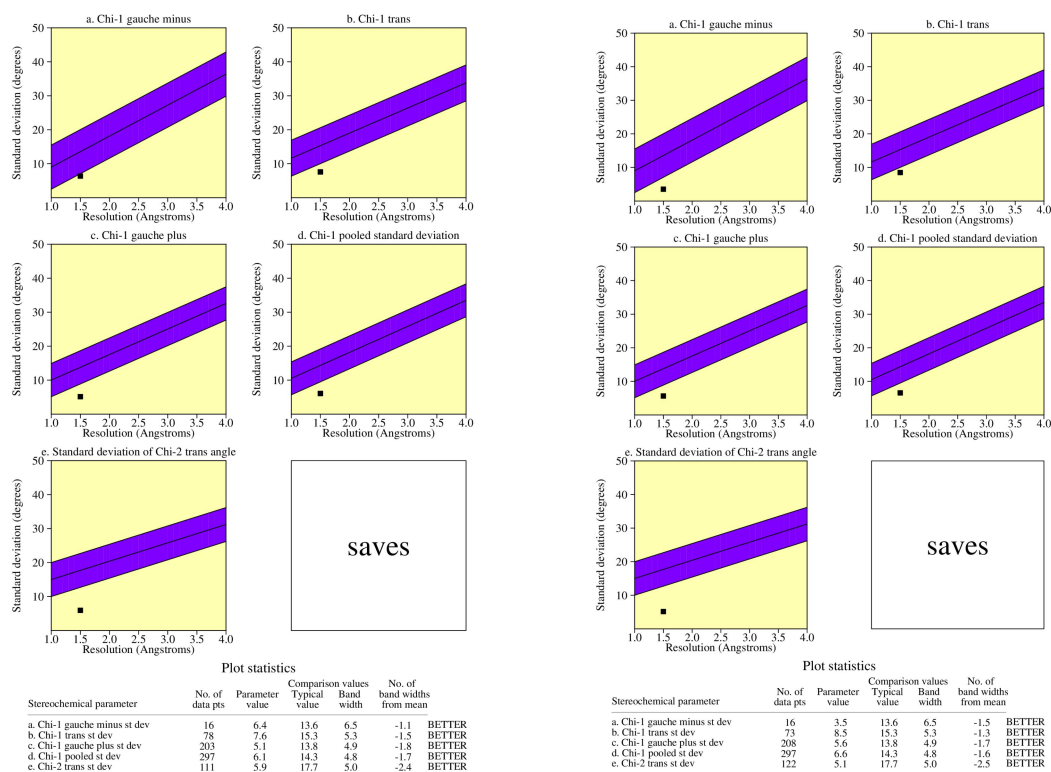
Note: The left side is the crude model; On the right is the refined model (Amino acid residues decreased from 39% to 37%).

Supplementary Fig S2. Evaluation of compatibility between tertiary structure models and amino acid sequences of candidate vaccines



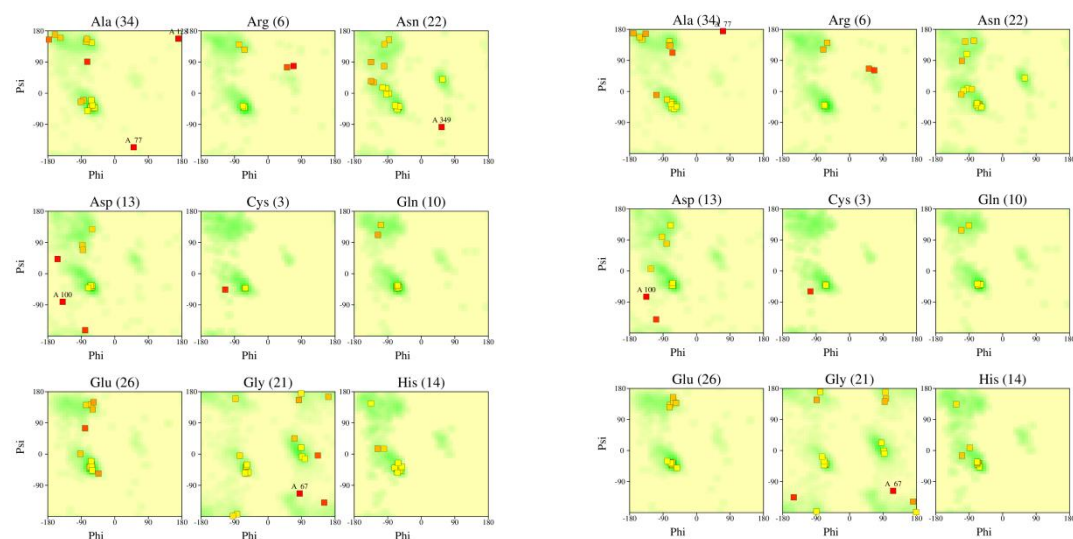
Note: The left side is the crude model; On the right is the refined model (The Parameter value rose from 88.4% to 93.6%).

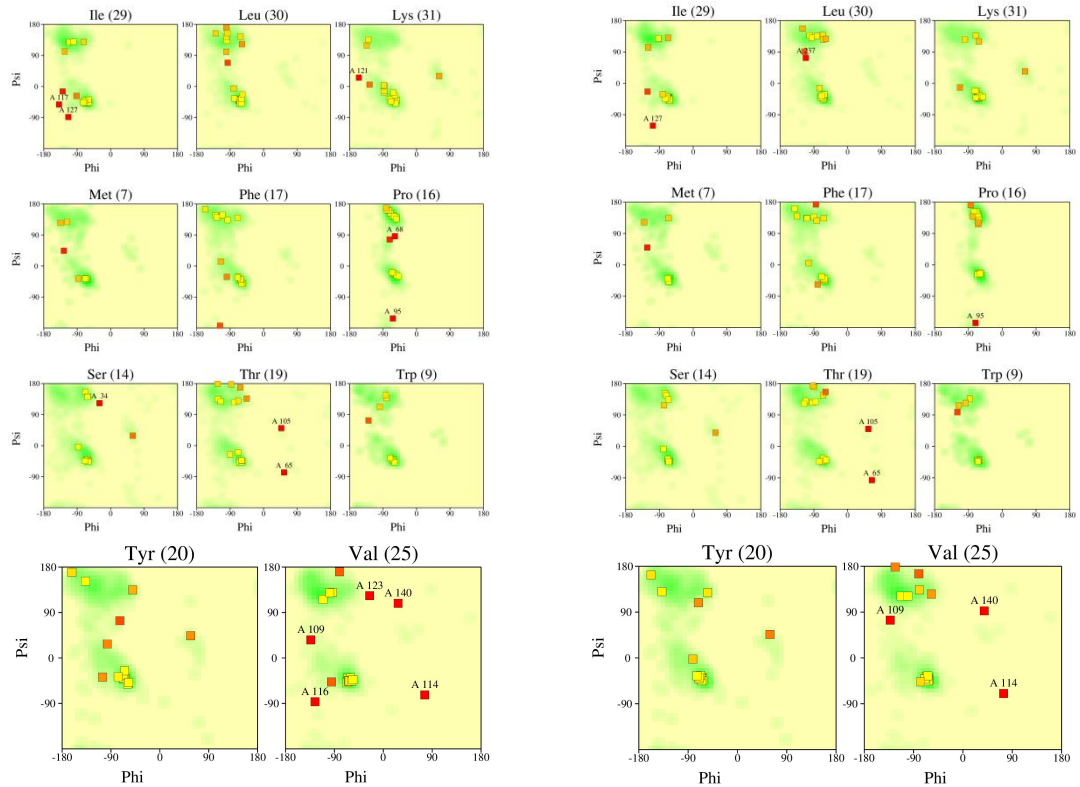
Supplementary Fig S3. Evaluation of main chain parameters of crude and refined tertiary structure models of candidate vaccines



Note: The left side is the coarse model; On the right is the refined model (parameter values decreased from 6.4% to 3.5%).

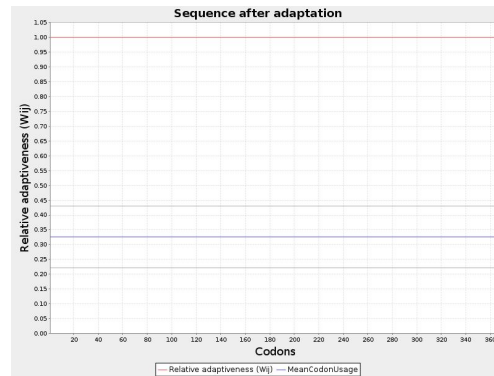
Supplement Fig S4. Evaluation of side chain parameters of crude and refined tertiary structure models of candidate vaccines





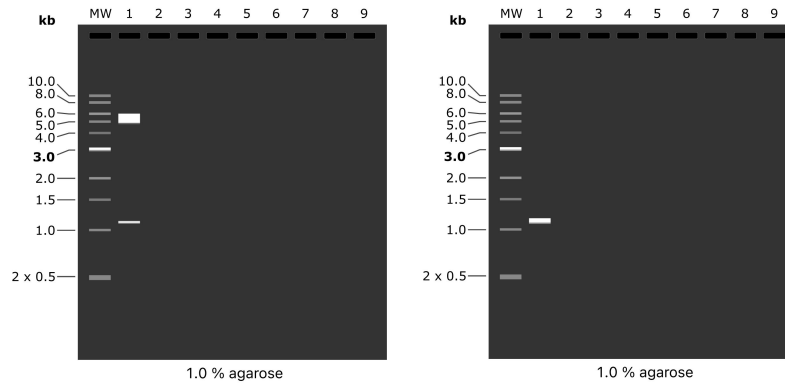
Note: Numbers of residues are shown in brackets, Those in unfavourable conformations (score < -3.00) are labelled. Shading shows favourable conformations as obtained from an analysis of 163 structures at resolution 2.0 Å or better.

Supplementary Fig S5. Ramachandran plots of all remaining types of crude and refined tertiary structure models of candidate vaccines



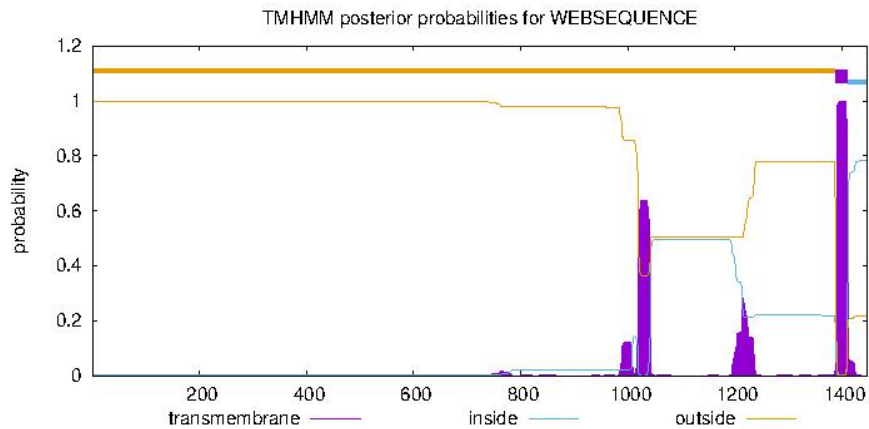
Note: Relative adaptiveness (w_{ij}) is 1.0; MeanCodonUsage ranges from 0.3 to 0.35.

Supplementary Fig S6. JCat server reverse translation of candidate vaccine amino acid sequence results



Note: The left is Double enzymatic digestion of pET-28a-Vaccine plasmid;The right is PCR simulation of the candidate nanoparticle vaccine Analysis.

Supplementary Fig S7. The SnapGene software simulates the gel electrophoresis results of double enzyme digestion and PCR amplification in the molecular cloning of candidate vaccines



Note: Outside: 1-1387 aa; transmembrane: 1388-1410 aa; inside: 1411-1448aa, Except B-6 dominant epitopes, all the other dominant epitopes were located on the surface of S protein.

Supplementary Fig S8. TMHMM2.0 server analyzed the transmembrane domain results of full-length TGEV S protein