A multi-method and structure-based in silico vaccine designing against *Echinococcus granulosus* through investigating enolase protein

Mohammad Mostafa Pourseif[1,2], Mitra Yousefpour[1*], Mohammad Aminianfar[2], Gholamali Moghaddam[3], Ahmad Nematollahi[4]

[1] Department of Physiology, Faculty of Medicine, AJA University of Medical Sciences, Tehran, Iran

[2] Infectious Diseases and Tropical Medicine Research Center (IDTMRC), Department of Aerospace and Subaquatic Medicine, AJA University of Medical Sciences, Tehran, Iran

[3] Department of Animal Sciences, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

[4] Department of Pathobiology, Veterinary College, University of Tabriz, Tabriz, Iran

----
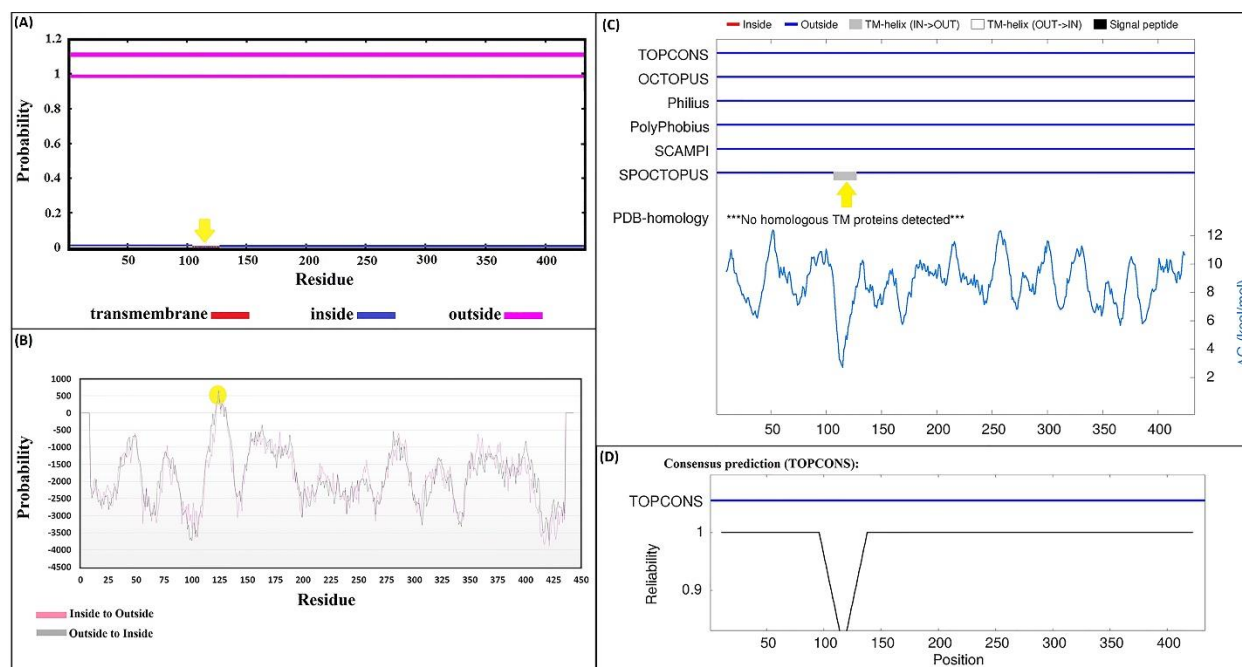
Fig. S1. The plots show the sequence-based prediction of potential transmembrane topology and signal peptide in the EgEnolase protein sequence. **(A)** The prediction plot shows subcellular localization of EgEnolase protein by using the TMHMM server. **(B)** Transmembrane topology profile of EgEnolase protein obtained from TMPred server. **(C)** The potential transmembrane helix of EgEnolase protein predicted in the TOPCONS server and based on six different algorithms. **(D)** The consensus prediction plot of TOPCONS server.
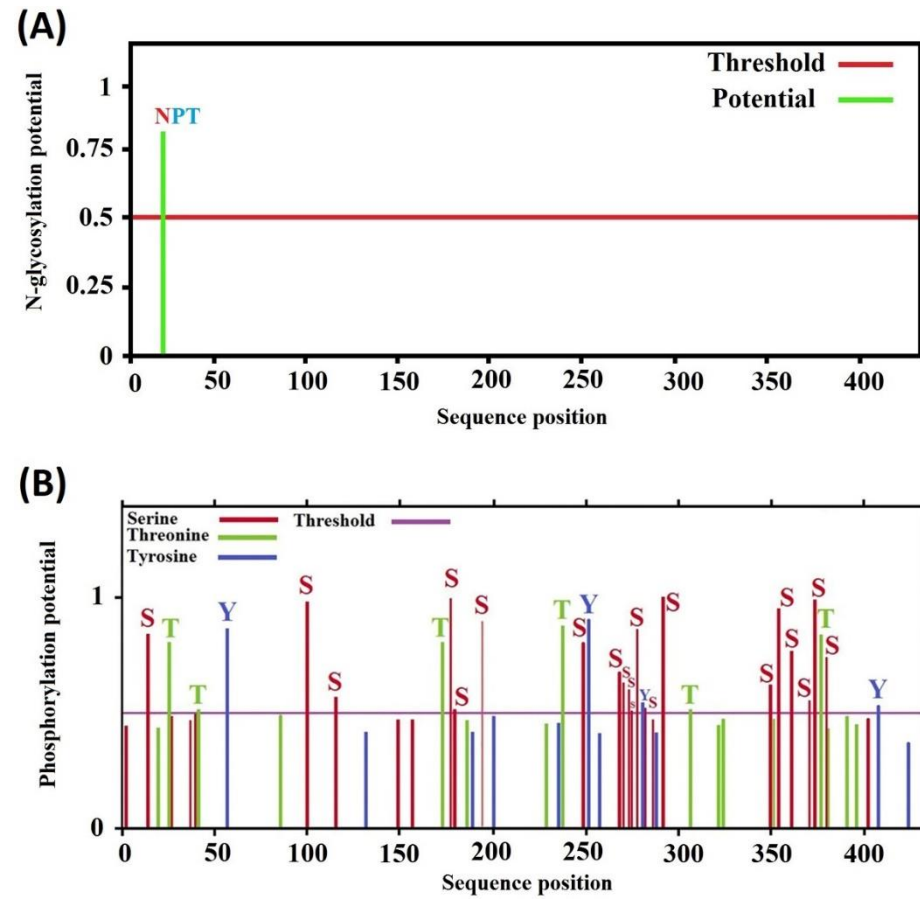
Fig. S2. Prediction of the post-translational modifications in the EgEnolase protein sequence. **(A)** The plot indicates the residue (17-NPT-20) with the potential N-glycosylation site. **(B)** Serine, Threonine, and Tyrosine phosphorylation plot obtained from the NetPhos v2.0 web-server. The scores more than the threshold value (0.5) were predicted as a phosphorylation site.
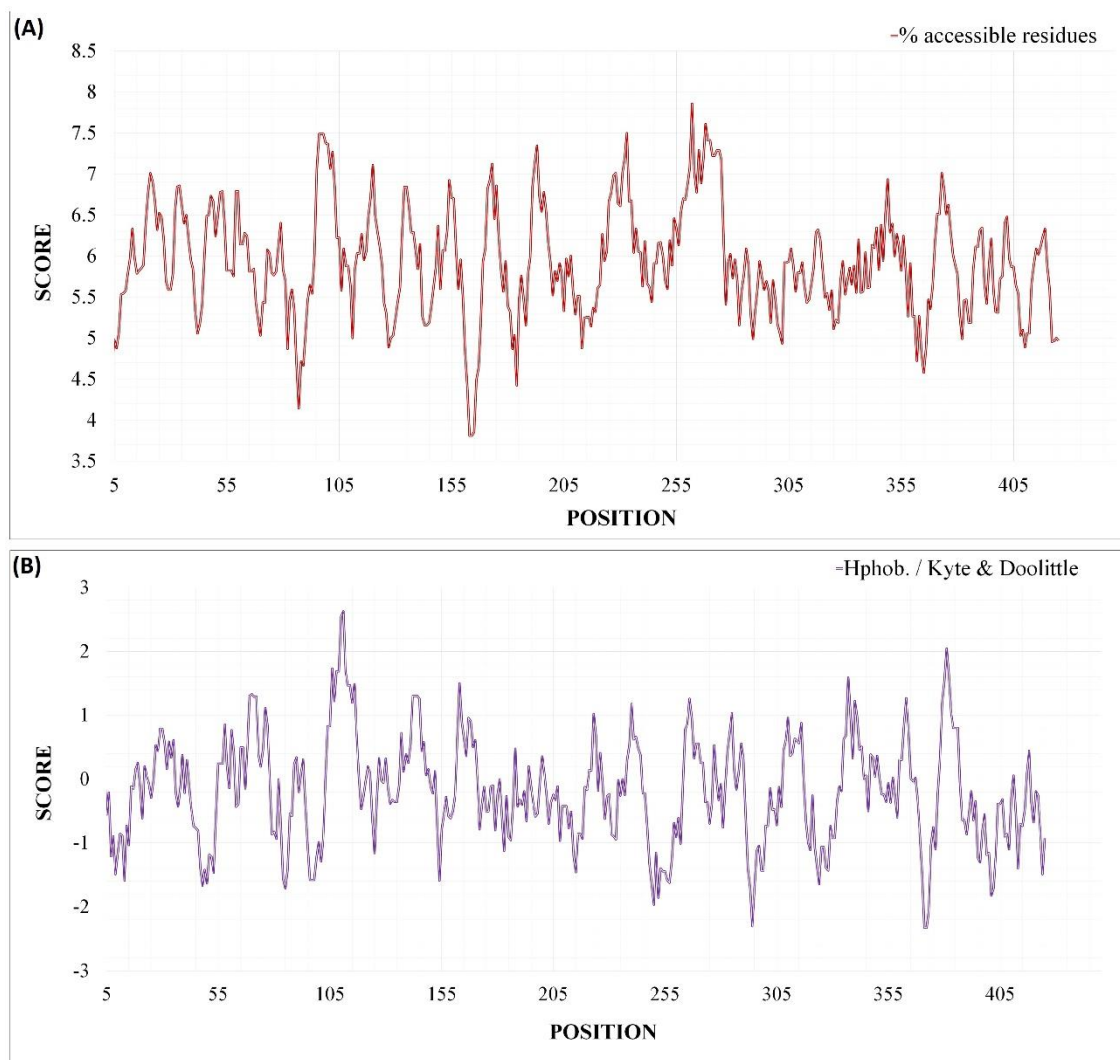
Fig. S3. Solvent accessible and hydrophobic regions of the EgEnolase protein sequence. **(A)** The plot illustrates surface-accessible regions of the EgEnolase protein sequence. **(B)** The hydrophobic residues are shown as plot and based on the Kyte and Doolittle algorithm. The negative values are related to the hydrophilic amino acids.
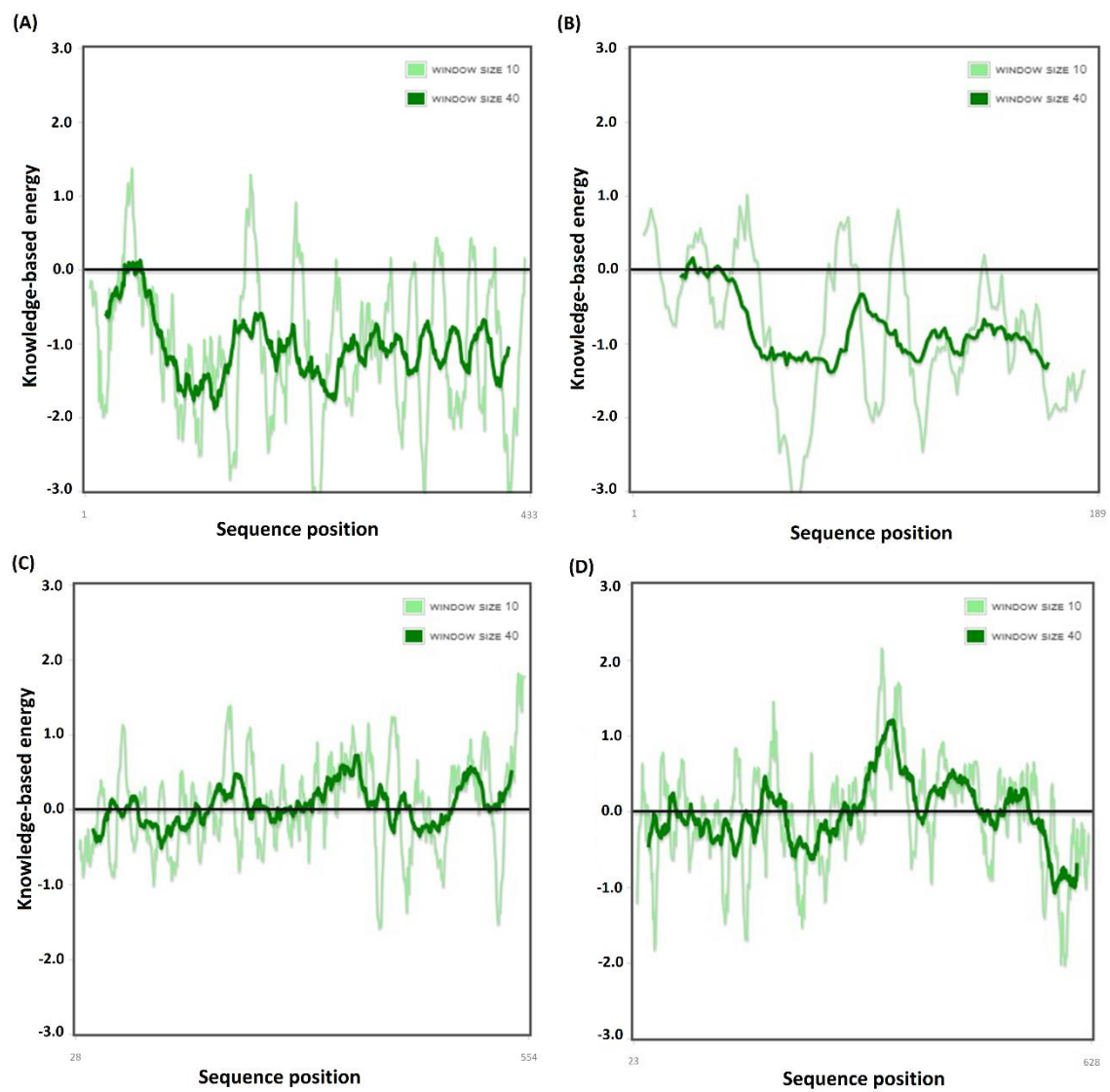
Fig. S4. ProSA-based energy plots for the 3D models. The energy plot for EgEnolase (**A**), DRB1*01101 (**B**), *C. lupus* Toll-like receptor

2 **(C)**, and *C. lupus* Toll-like receptor 4 **(D)** are represented. The negative values of the plots are related to the stable residues.

beta-enolase isoform X1 [Canis lupus familiaris]
Sequence ID: XP_536606.4   Length: 440   Number of Matches: 5
▶ See 1 more title(s)

Range 1: 26 to 143 GenPept Graphics   ▼ Next Match ▲ Previous Matcl

Score	Expect	Method	Identities	Positives	Gaps
77.4 bits(189)	2e-14	Compositional matrix adjust.	43/118(35%)	66/118(55%)	15/118(12%)

```
Query  421  VKIGMDVAFFRKGPSLPSGASTGVHEAVELRDADKNAYMGKGG-------GGSGGGGSIK  473
                 V++ +  A  R    ++PSGAST ++EA+ELRD DK+ Y+GKG        + G   ++
Sbjct  26   VEVDLHTAKGRFRAAVPSGASTLIYEALELRDGDKSRYLGKGVLKAVEHINKTLGPALLE  85

Query  474  EKFVVTDQQRIDEFMIKLDGSPNKGKLGGGGGSG--------GGGSMGTEVYHHLKSV  523
                 +K  V DQ+++D+FMI+LDG+ NK K G      G          G   G +Y H+  +
Sbjct  86   KKLSVVDQEKVDKFMIELDGTENKSKFGANAILGVSLAVCKAGAAEKGVPLYRHIADL  143
```

Fig. S5. The alignment output between the vaccine protein sequence and the most similar protein based on the NCBI's blastp (protein-protein BLAST) algorithm. The similar proteins were searched among *Canis lupus familiaris* proteome information.

(A)

IFIIVVLINATTQYDCVTSSEVVSDSYNKTTITFENKPQYYNSPSGNVVPKAIMPILIKKGQTIQVSSITTNVKYEATNQD
LTFLFRKDGCHGTNSEIATYAGATNTNVFLGNTNTVSLTQFKFTADYNGIILIVGKNLGASLPGDIRVNVFEAAAKEAA
AKAPPHALSEAAAKKLAMQEFMILPTGFFRKGPGPGGALIIHARQIFDSFFRKGPGPGAMQEFMILPTGAKFFRKGP
GPGEFMILPTGAKSFSFFRKGPGPGLIIHARQIFDSRFFRKGPGPGMSRAAGWGVMVSHFFRKGPGPGAGWGVM
VSHRSGEFFRKGPGPGLRIEEEELGPKAVYFFRKGPGPGKAVYAGEHFRNPLFFRKGPGPGYPIVSIEDPFDQDFFRKG
PGPGVLPVPSFNVLNGGFFRKGPGPGGYTGKVKIGMDVAFFRKGPSLPSGASTGVHEAVELRDADKNAYMGKGG
GGSGGGGSIKEKFVVTDQQRIDEFMIKLDGSPNKGKLGGGGGSGGGGSMGTEVYHHLKSVIKGKYGLDACNVGG
GGSGGGGSKTAIDKAGYTGKVKGGGGSGGGGSSEFYQDGNYNLDFKNPKAAASSIVSGSKLSDI

Adjuvants: IFIIV...NVF and APPHALS
Linkers: EAAAK, FFRK, GPGPG, GPSL, and GGGGSGGGGS
Helper T-cell Epitope: 13-mer epitopes

(B)

ATTTTTATTATTGTGGTGCTGATTAACGCGACCACCCAGTATGATTGCGTGACCAGCAGCGAAGTGGT
GAGCGATAGCTATAACAAAACCACCATTACCTTTGAAAACAAACCGCAGTATTATAACAGCCCGAGCG
GCAACGTGGTGCCGAAAGCGATTATGCCGATTCTGATTAAAAAAGGCCAGACCATTCAGGTGAGCAGC
ATTACCACCAACGTGAAATATGAAGCGACCAACCAGGATCTGACCTTTCTGTTTCGCAAAGATGGCTG
CCATGGCACCAACAGCGAAATTGCGACCTATGCGGGCGGCGACCAACACCAACGTGTTTCTGGGCAACA
CCAACACCGTGAGCCTGACCCAGTTTAAATTTACCGCGGATTATAACGGCATTATTCTGATTGTGGGC
AAAAACCTGGGCGCGAGCCTGCCGGGCGATATTCGCGTGAACGTGTTTGAAGCGGCGGCGAAAGAAGC
GGCGGCGAAAGCGCCGCCGCATGCGCTGAGCGAAGCGGCGGCGAAAAAACTGGCGATGCAGGAATTTA
TGATTCTGCCGACCGGCTTTTTTCGCAAAGGCCCGGGCCCGGGCGGCGCGCTGATTATTCATCGCGCC
AGATTTTTGATAGCTTTTTTCGCAAAGGCCCGGGCCCGGGCGGCGATGCAGGAATTTATGATTCTGCCG
ACCGGCGCGAAATTTTTTCGCAAAGGCCCGGGCCCGGGCGAATTTATGATTCTGCCGACCGGCGCGAA
AAGCTTTAGCTTTTTTCGCAAAGGCCCGGGCCCGGGCCTGATTATTCATCGCGCGCAGATTTTTGATA
GCCGCTTTTTTCGCAAAGGCCCGGGCCCGGGCGCGGGCTGGGGCGTGATGGTGAGCCATCGCAGCGGCGAA
ATTTTTTTCGCAAAGGCCCGGGCCCGGGCCTGCGCATTGAAGAAGAACTGGGCCCGAAAGCGGTGTATTT
TTTTCGCAAAGGCCCGGGCCCGGGCAAAGCGGTGTATGCGGGCGAACATTTTCGCAACCCGCTGTTTTTT
TCGCAAAGGCCCGGGCCCGGGCTATCCGATTGTGAGCATTGAAGATCCGTTTGATCAGGATTTTTTTC
GCAAAGGCCCGGGCCCGGGCGTGCTGCCGGTGCCGAGCTTTAACGTGCTGAACGGCGGCTTTTTTCGCA
AAGGCCCGGGCCCGGGCGGCTATACCGGCAAAGTGAAAATTGGCATGGATGTGGCGTTTTTTCGCAAA
GGCCCGAGCCTGCCGAGCGGCGCGAGCACCGGCGTGCATGAAGCGGTGGAACTGCGCGATCGCGGATAA
AAACGCGTATATGGGCAAAGCGGCGGCGGCAGCGGCGGCGGCGGCAGCATTAAAGAAAAATTTGTGG
TGACCGATCAGCAGCGCCATTGATGAATTTATGATTAAACTGGATGGCAGCCCGAACAAAGGCAAACTG
GGCGGCGGCGGCGGCAGCGGCGGCGGCGGCAGCATGGGCACCGAAGTGTATCATCATCTGAAAAGCCGT
GATTAAAGGCAAATATGGCCTGGATGCGTGCAACGTGGGCGGCGGCGGCGGCAGCGGCGGCGGCGGCAGCA
AAACCGCGATTGATAAAGCGGGCTATACCGGCAAAGTGAAAGGCGGCGGCGGCGGCAGCGGCGGCGGCGGC
AGCAGCGAATTTTATCAGGATGGCAACTATAACCTGGATTTTAAAAACCCGAAAGCGGCGGCGGCGAGCAG
CATTGTGAGCGGCAGCAAACTGAGCGATATTTAA

vaccine construct.

Table S1. Antigenic scores for eight EgEnolase protein sequences in terms of two different predictor tools.

NCBI accession number	Prediction method	
	ANTIGENpro	**VaxiJen *v*2.0**
ACY30465	0.5309	0.4814
XP_024346720	0.4791	0.3705
EUB55526	0.3309	0.2935
XP_024346722	0.3309	0.2935
EUB55524	0.4791	0.3705
CDS19796	0.5903	0.4814
CDS21390	0.4791	0.3705

Table S2. The six high rank homologous PDB structures that were used as template for homology modeling.

Query proteins	PDB Templates	Identity (%)	Total score
TLR-2	2Z7X	75	768
	5D3I	64	744
	3A79	59	657
	2Z81	61	645
	2Z80	70	403
	1O77	90	284
TLR-4	4G8A	68	795
	3FXI	68	791
	3VQ1	59	696
	2Z64	59	692
	2Z63	67	672
	5IJB	55	605
	2PSN	75	682

EgEnolase	3B97	75	679
	2XSX	73	670
	4ZA0	72	658
	1TE6	71	655
	3UCC	71	655
DRB1*01101	4AH2	82	74
	3PDO	82	74
	1AQD	82	74
	4X5X	81	74
	2WBJ	79	74
	1YMM	80	74

Table S3. Binding energy between eighty 13-mer peptides and DLA-DRB1*01101 allele obtained from the molecular docking method.

13-mer peptides	weighted score¶		13-mer peptides	weighted score		13-mer peptides	weighted score	
	center	lowest E		center	lowest E		center	lowest E
GALIIHARQIFDS	-825.7	-999.1	GVSLAVCKAGAAE	-602.3	-633.7	SSIVSGSKLSDIY	-626.0	-685.1
LIIHARQIFDSR	-793.3	-945.2	CKAGAAEKGVPLY	-587.5	-670.3	SKLSDIYSEMISK	-603.1	-678.8
HARQIFDSRGNPT	-616.4	-809.5	AEKGVPLYRHVAD	-623.5	-711.1	YSEMISKYPIVSI	-675.1	-819.4
ARQIFDSRGNPTV	-580.5	-721.6	PLYRHVADLAGNK	-605.8	-746.5	YPIVSIEDPFDQD	-876.3	-896.0
FDSRGNPTVEVDL	-786.6	-786.6	VADLAGNKDVVLP	-571.3	-675.7	DPFDQDDWAAWTE	-762.9	-821.4
DSRGNPTVEVDLT	-691.1	-711.3	GNKDVVLPVPSFN	-708.8	-830.6	DWAAWTEFNAKAG	-718.3	-805.3
NPTVEVDLTTSKG	-676.0	-712.7	VLPVPSFNVLNGG	-844.0	-896.6	FNAKAGIQIVGDD	-623.8	-769.6
VEVDLTTSKGLFR	-748.5	-763.1	FNVLNGGSHAGNK	-635.1	-727.1	IQIVGDDLTVTNP	-652.9	-812.7
DLTTSKGLFRAAV	-634.1	-710.4	GSHAGNKLAMQEF	-613.0	-702.8	LTVTNPERVQQAI	-611.1	-678.6
TSKGLFRAAVPSG	-675.9	-812.7	KLAMQEFMILPTG	-906.4	-1016.2	RVQQAIDRKACNA	-604.3	-667.7
LFRAAVPSGASTG	-750.5	-848.3	AMQEFMILPTGAK	-882.7	-983.0	DRKACNALLLKVN	-684.5	-815.0
AVPSGASTGVHEA	-574.4	-713.7	EFMILPTGAKSFS	-801.7	-946.9	ALLLKVNQIGSVT	-716.3	-719.2
ASTGVHEAVELRD	-627.7	-813.8	TGAKSFSEAMKMG	-581.1	-736.1	IGSVTESIKACKM	-687.7	-687.7
VHEAVELRDADKN	-720.5	-770.4	FSEAMKMGTEVYH	-673.6	-730.0	KACKMSRAAGWGV	-658.5	-746.5
VELRDADKNAYMG	-636.3	-747.7	MKMGTEVYHHLKS	-715.1	-715.1	MSRAAGWGVMVSH	-922.9	-922.9
DADKNAYMGKGVL	-595.1	-712.6	GTEVYHHLKSVIK	-528.5	-638.2	AGWGVMVSHRSGE	-819.7	-916.8
NAYMGKGVLNAVK	-594.5	-664.3	LKSVIKGKYGLDA	-720.3	-773.5	VSHRSGETEDSTI	-618.1	-729.4
GKGVLNAVKNVNE	-499.4	-556.7	GKYGLDACNVGDE	-669.4	-858.5	ETEDSTIADIVVG	-554.9	-632.4
LNAVKNVNEVIAP	-505.6	-597.1	LDACNVGDEGGFA	-746.1	-813.1	STIADIVVGLRTG	-603.1	-703.3
KNVNEVIAPALIK	-569.6	-578.6	GDEGGFAPNIQDN	-626.0	-712.3	IADIVVGLRTGQI	-669.7	-794.5

EVIAPALIKEKFV	-743.7	-804.9	APNIQDNMEGLEL	-579.1	-645.3	GLRTGQIKTGAPC	-738.8	-833.7
ALIKEKFVVTDQQ	-687.0	-823.2	NMEGLELLKTAID	-513.2	-619.5	IKTGAPCRSERLA	-656.9	-711.1
VVTDQQRIDEFMI	-709.4	-774.6	LKTAIDKAGYTGK	-723.0	-776.9	CRSERLAKYNQLL	-715.0	-756.8
QQRIDEFMIKLDG	-635.2	-779.2	GYTGKVKIGMDVA	-859.7	-859.7	SERLAKYNQLLRI	-639.9	-716.5
EFMIKLDGSPNKG	-578.3	-634.6	KIGMDVAASEFYQ	-612.2	-773.3	AKYNQLLRIEEEL	-629.2	-702.1
DGSPNKGKLGANA	-365.8	-470.3	ASEFYQDGNYNLD	-626.6	-743.8	LRIEEELGPKAVY	-710.7	-909.9
KGKLGANAILGVS	-463.4	-592.5	FYQDGNYNLDFKN	-778.0	-778.6	LGPKAVYAGEHFR	-792.8	-885.1
ANAILGVSLAVCK	-601.5	-665.7	KNPKAAASSIVSG	-490.5	-555.1	KAVYAGEHFRNPL	-743.9	-903.3

¶ binding energy unit is kJ/mol. * The lowest binding energy value between the 13-mer peptides of EgEnolase and DRB1*01101 allele was selected as final CD4[+] T-helper epitope.