bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

# Understanding the B and T cells epitopes of spike protein of severe respiratory syndrome coronavirus-2: A computational way to predict the immunogens

Yoya Vashi[†], Vipin Jagrit[†], Sachin Kumar[*]

Viral Immunology Group, Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, Guwahati 781039 Assam India.

†Contributed equally

*Corresponding author Email: sachinku@iitg.ac.in (S Kumar)

Contact number: 91-3612582229

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.


## Abstract

The 2019 novel severe respiratory syndrome coronavirus-2 (SARS-CoV-2) outbreak has caused a large number of deaths with thousands of confirmed cases worldwide. The present study followed computational approaches to identify B- and T-cell epitopes for spike glycoprotein of SARS-CoV-2 by its interactions with the human leukocyte antigen alleles. We identified twenty-four peptide stretches on the SARS-CoV-2 spike protein that are well conserved among the reported strains. The S protein structure further validated the presence of predicted peptides on the surface. Out of which twenty are surface exposed and predicted to have reasonable epitope binding efficiency. The work could be useful for understanding the immunodominant regions in the surface protein of SARS-CoV-2 and could potentially help in designing some peptide-based diagnostics.


**Keywords:** SARS-CoV-2; Spike protein; epitopes; diagnostics.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Introduction

Emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a recent pandemic and declared as a public health emergency by World Health Organization (WHO) [1]. The disease rapidly spread across the globe and caused havoc to humanity [2]. By the end of March, SARS-CoV-2 had spread to 200 countries and infected over 4,50,000 people [3]. The WHO is continuously monitoring and updating the health-related plans to curtail the disease spread. The absence of specific treatment and vaccine worsen the situation and threat the world.

International Committee on Taxonomy of Viruses (ICTV), classified SARS-CoV-2 under family *coronaviridae* of order *nidovirales*. The genomic sequence of SARS-CoV-2 isolated from the bronchoalveolar lavage fluid of a patient from the Wuhan, China showed a length of 29,903 nucleotides (GenBank accession number NC_045512) [4]. The SARS-CoV-2 contains a positive single-stranded RNA with 5′ and 3′ UTR. The genome codes for ORF1a, ORF1b, Spike (S), ORF3a, ORF3b, Envelope (E), Membrane (M), ORF6, ORF7a, ORF7b, ORF8, ORF9b, ORF14, Nucleocapsid (N), and ORF10 from 5′ to 3′ [4,5].

The S glycoprotein forms homotrimers and represents a potential target for therapeutic and vaccine design as it mediates viral entry into host cells [6,7]. S glycoprotein comprises of two functional subunits. Whereas the S1 subunit is responsible for binding to the host cell receptor, the S2 subunit is responsible for the fusion of the viral with the cell membrane. Usually, in CoVs, S is cleaved at the boundary between the S1 and S2 subunits, which remain non-covalently bound in the prefusion conformation, to activate the protein for membrane fusion via extensive irreversible conformational changes [8-10]. Setting apart from other SARS-CoVs, it is found that S glycoprotein of SARS-CoV-2 harbors a furin cleavage site at the boundary between the S1/S2 subunits [11]. By now, it is evident that SARS-CoV-2 S uses angiotensin-converting enzyme 2 (ACE2) receptor-mediated entry into cells. It is found that the receptor-

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

binding domains of S proteins of SARS-CoV-2 and SARS-CoV bind with similar affinities to human ACE2 [11,12].

As the situation worsens, there is a growing need for the development of suitable therapeutics and alternate diagnostics against SARS-CoV-2 for effective disease management strategies. Diagnostic assays based on peptides have become increasingly substantial and indispensable for its advantages over conventional methods [13]. The present study aimed to locate appropriate epitopes within a particular protein antigen, which can elicit an immune response that could be selected for the synthesis of the immunogenic peptide. Using computational approach, S glycoprotein of SARS-CoV-2 was explored to identify various immunodominant epitopes for the development of diagnostics. Besides, the results could also help us to understand the SARS-CoV-2 surface protein response towards T and B cells.

## Materials & Methods

### Collection of targeted protein sequence

We downloaded amino acid sequences (n=98) of S protein available at the time of study on targeted SARS-CoV-2 from the National Centre for Biotechnological Information (NCBI) database.

### Identification of potential peptides

To identify an immunodominant region, it is of extreme importance to select the conserved region within the S protein of SARS-CoV-2. All the sequences were compared among themselves for variability using protein variability server by Shannon method [14]. The average solvent accessibility (ASA) profile was predicted for each sequence using SABLE server [15]. BepiPred 1.0 Linear Epitope Prediction module [16-18] incorporated in Immune Epitope Database (IEDB) [19] was used to predict potential epitopes within the S protein. The FASTA sequence of the targeted protein was used as an input for all the default parameters.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

99 **Identification of B-cell epitopes**

100 We used two web-based tools for B-cell epitope prediction, viz., the IEDB, and
101 ABCpred server [20]. S protein structure from protein data bank (PDB) (6VSB) [21] was analyzed
102 for linear and discontinuous B-cell epitopes using ElliPro module [22] on IEDB server, with
103 default settings. Also, ABCpred server was used to detect B-cell epitopes using artificial neural
104 network (ann) method.

105 **Identification of T-cell epitopes**

106 T-cell epitopes having binding affinity towards MHC-I and MHC-II alleles were
107 selected to boost up both cytotoxic T-cell and helper T-cell mediated immune response. IEDB
108 server was used to predict the MHC-I and MHC-II binding epitopes for targeted protein. The
109 reference set of alleles was used for predicting the MHC-I and MHC-II T-cell epitopes. [23-27]

110

111 **Results and Discussion**

112 In our study, we targeted the S glycoprotein of SARS-CoV-2 as it is present outside the
113 virus and interacts with the host receptor. At the time of the study, there were 98 sequences
114 available for the targeted protein of SARS-CoV-2. The S glycoprotein is 1273 amino acids
115 long sequence except for the virus isolated from Kerala (India), which is 1272 amino acids
116 long spike glycoprotein (GenBank accession number MT012098). Our interest here was to
117 determine conserved regions first and then determine surface-exposed regions, which are
118 potential epitopes to generate immune response. We found that sequences among all the S
119 proteins in the analysis are least variable and highly conserved as shown in Figure 1. Regions
120 having a high value of ASA are more surface exposed as compared to others. We identified a
121 total of 24 peptides of varying length, as shown in Table 1, which are selected based on high
122 ASA values. The potential epitope regions were predicted using the sequence of S protein of
123 SARS-CoV-2, which showed the least variability (GenBank accession number NC_045512).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

124 The potential epitopes are represented by blue peaks, while green-colored slopes represent non-
125 epitopic regions (Figure 2).

126 The existence of B-cell linear and discontinuous (conformational) epitopes within the
127 identified segments could help us to identify the peptides, which can elicit immune response
128 [28]. We identified 18 linear epitopes, predicted by ElliPro (IEDB), which contains regions from
129 19 of our selected peptides highlighted in red in Table 2. These identified B-cell linear epitopes
130 are placed based on their positional value, and scores. Epitopes with high scores have more
131 potential for antibody binding. Five of our selected peptides (peptide numbers 3, 5, 19, 23, and
132 24 in Table 1) were not considered as potential linear B-cell epitopes.

133 Using the same module, B-cell discontinuous epitopes were predicted, which gave 16
134 epitope regions that contained regions from 18 of our selected peptides highlighted in red
135 (Table 3). Six peptides (peptide numbers 3, 5, 14, 19, 23, and 24 in Table 1) were not predicted
136 as discontinuous B-cell epitopes. To further confirm, we used ABCpred server to detect B-cell
137 epitopes, with default threshold of 0.51. It identified various epitopes with different length and
138 scores; out of those, the regions which contained our selected peptides are highlighted in red
139 (Table 4). A high score represents a good binding affinity with epitopes, and most of our
140 peptides scored more than 0.7 and were predicted as linear B-cell epitopes.

141 We used the IEDB server to determine the binding affinity for human leucocyte antigen
142 (HLA) with our selected peptides from Table 1. As recommended by the IEDB server,
143 reference HLA allele sets were used for the prediction of MHC-I and MHC-II T-cell epitopes,
144 as they provide comprehensive coverage of the population. All the predictions were made using
145 IEDB recommended procedures. We observed good binding affinities for our selected peptides.
146 The list of binding affinities for MHC-I T-cell epitopes is given in Table 5, where low rank
147 represents high binding affinity. The epitopes with rank <1% for very high binding affinity
148 were selected. Regions from all of our selected peptides were found to be potential T-cell

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

149 epitope(s) with high binding affinity with HLA-A and HLA-B alleles, except one. Similarly,
150 the list of binding affinities for MHC-II T-cell epitopes are given in Table 6. Regions from our
151 selected peptides are highlighted in red. The results revealed that around half of our selected
152 peptides are potential T-cell epitope(s) with high binding affinity with HLA-DRB and HLA-
153 DP/DQ alleles.

154 Overall, it was found that the regions identified in Table 1 not only had good B-cell and
155 T-cell affinities, but the majority of them had overlaps with discontinuous epitopes also (Table
156 3). The peptide segments identified from the set of 98 sequences of the SARS-CoV-2 S
157 glycoprotein appear to hold reasonable potential to act as immunogens. Peptide-based
158 diagnostics and vaccines have been proposed against virus outbreaks earlier [29-33]. The
159 availability of a 3D structure (6VSB) of the SARS-CoV-2 S glycoprotein provided an
160 opportunity to inspect the predicted peptides. Placement of the peptide segments identified by
161 ASA and conserved sequence analysis on the S glycoprotein showed that 20 regions that we
162 identified lie on the surface (Figure 3). In order to limit recognition and evade from immune
163 response of host, coronaviruses use conformational masking and glycan shielding [34,35]. SARS-
164 CoV-2 S trimer also exists in multiple distinct conformational states, which is necessary for
165 receptor engagement leading to initiation of fusogenic conformational changes [11]. A
166 considerable good number of peptides at the surface region of the S glycoprotein allows the
167 potential use of those peptide regions as immunogens.

168 The emergence of new viral diseases like SARS-CoV-2 represents a substantial global
169 disease burden. There is an urgent need for diagnostics, therapeutics, and vaccines against
170 newly emerged SARS-CoV-2. Facilitated by high mutation rates, traditional vaccines based on
171 antibody-mediated protection are often poor inducers of T cell responses and can have limited
172 success [36]. In our study, we predicted both B-cell and T-cell epitopes for conferring immunity
173 in different ways. We speculate that the identified epitopes with considerably good epitope

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

174 binding efficiency have the potential to be an immunodominant peptide. Peptide-based
175 sensitive and rapid diagnostic kits are considered as a better alternative to the conventional
176 serological tests including whole antigenic protein [13]. The study could help us to use the
177 predicted peptide as an immunogen for the development of diagnostics against SARS-CoV-2.
178
179 **Conclusion**
180 In the present study, peptide segments were identified on S proteins for the development
181 of diagnostics against SARS-CoV-2. The recent availability of 3D data on 2019-CoV spike
182 glycoprotein has helped the search. SARS-CoV-2, being an RNA virus, has high mutations rate
183 and undergoing active recombination [37]. Although the peptides identified are ideal candidates
184 as immunogens for peptide-based diagnostics development, more refinement and lab trials are
185 essential steps that are yet to be undertaken for early development of diagnostics before the
186 identified epitopes are rendered obsolete.
187
188

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**References:**


1. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 2020.
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China. *JAMA.* 2020.
3. WHO. Coronavirus disease 2019 (COVID-19) Situation Report - 66. 2020.
4. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. *Nature.* 2020;579(7798):265-269.
5. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-733.
6. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol.* 2016;3(1):237-261.
7. Tortorici MA, Walls AC, Lang Y, et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol.* 2019;26(6):481-489.
8. Park JE, Li K, Barlan A, et al. Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands virus tropism. *Proc Natl Acad Sci U S A.* 2016;113(43):12262-12267.
9. Burkard C, Verheije MH, Wicht O, et al. Coronavirus cell entry occurs through the endo-/lysosomal pathway in a proteolysis-dependent manner. *PLoS Pathog.* 2014;10(11):e1004502.
10. Walls AC, Tortorici MA, Snijder J, et al. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A.* 2017;114(42):11157-11162.
11. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

12. Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol.* 2020.

13. Mohanraj U, Chander S, Chavan YG. Peptide Based Viral Detection Systems for Effective Diagnosis of Common Viral Infections in India. *Curr Protein Pept Sci.* 2017;18(9):939-945.

14. Garcia-Boronat M, Diez-Rivero CM, Reinherz EL, Reche PA. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res.* 2008;36(Web Server issue):W35-41.

15. Adamczak R, Porollo A, Meller J. Accurate prediction of solvent accessibility using neural networks-based regression. *Proteins.* 2004;56(4):753-767.

16. Larsen JE, Lund O, Nielsen M. Improved method for predicting linear B-cell epitopes. *Immunome Res.* 2006;2:2.

17. Ponomarenko JV, Bourne PE. Antibody-protein interactions: benchmark datasets and prediction tools evaluation. *BMC Struct Biol.* 2007;7:64.

18. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* 2006;15(11):2558-2567.

19. Vita R, Mahajan S, Overton JA, et al. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2019;47(D1):D339-D343.

20. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins.* 2006;65(1):40-48.

21. Wrapp D, Wang N, Corbett KS, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* 2020;367(6483):1260-1263.

22. Ponomarenko J, Bui HH, Li W, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics.* 2008;9:514.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

238 23. Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of T-cell epitopes using
239 neural networks with novel sequence representations. *Protein Sci.* 2003;12(5):1007-
240 1017.
241 24. Peters B, Sette A. Generating quantitative models describing the sequence specificity
242 of biological processes with the stabilized matrix method. *BMC Bioinformatics.*
243 2005;6:132.
244 25. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method
245 for the major histocompatibility complex class I predictions. *Immunogenetics.*
246 2012;64(3):177-186.
247 26. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using
248 SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics.*
249 2007;8:238.
250 27. Sturniolo T, Bono E, Ding J, et al. Generation of tissue-specific and promiscuous HLA
251 ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat*
252 *Biotechnol.* 1999;17(6):555-561.
253 28. Purcell AW, McCluskey J, Rossjohn J. More than one reason to rethink the use of
254 peptides in vaccine design. *Nat Rev Drug Discov.* 2007;6(5):404-414.
255 29. Oany AR, Emran AA, Jyoti TP. Design of an epitope-based peptide vaccine against
256 spike protein of human coronavirus: an in silico approach. *Drug Des Devel Ther.*
257 2014;8:1139-1149.
258 30. Dey S, Nandy A, Basak SC, Nandy P, Das S. A Bioinformatics approach to designing
259 a Zika virus vaccine. *Comput Biol Chem.* 2017;68:143-152.
260 31. Ichihashi T, Yoshida R, Sugimoto C, Takada A, Kajino K. Cross-protective peptide
261 vaccine against influenza A viruses developed in HLA-A*2402 human immunity
262 model. *PLoS One.* 2011;6(9):e24626.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

263 32. Navalkar KA, Johnston SA, Stafford P. Peptide based diagnostics: are random-
264 sequence peptides more useful than tiling proteome sequences? *J Immunol Methods.*
265 2015;417:10-21.
266 33. Zhao K, Liu Q, Yu R, et al. Screening of specific diagnostic peptides of swine hepatitis
267 E virus. *Virol J.* 2009;6:186.
268 34. Walls AC, Xiong X, Park YJ, et al. Unexpected Receptor Functional Mimicry
269 Elucidates Activation of Coronavirus Fusion. *Cell.* 2019;176(5):1026-1039 e1015.
270 35. Xiong X, Tortorici MA, Snijder J, et al. Glycan Shield and Fusion Activation of a
271 Deltacoronavirus Spike Glycoprotein Fine-Tuned for Enteric Infections. *J Virol.*
272 2018;92(4).
273 36. Rosendahl Huber S, van Beek J, de Jonge J, Luytjes W, van Baarle D. T cell responses
274 to viral infections - opportunities for Peptide vaccination. *Front Immunol.* 2014;5:171.
275 37. Yi H. 2019 novel coronavirus is undergoing active recombination. *Clin Infect Dis.*
276 2020.
277
278
279
280
281
282
283
284
285
286
287

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.013516; this version posted April 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Figure legends

**Figure 1.** Profiles of average solvent accessibility (blue) in % and amino acid sequence variability (green) in numbers of the 98 SARS-CoV-2 protein plotted against amino acid numbers.

**Figure 2.** Graphical representation of B-cell linear epitopes of spike protein of SARS-CoV-2. B-cell linear epitopes predicted using BepiPred 1.0 module incorporated in IEDB server using default threshold value (0.35).

**Figure 3.** Our selected peptides are highlighted on spike protein of SARS-CoV-2 protein structure downloaded from PDB (ID: 6VSB).

**Figure 1.**

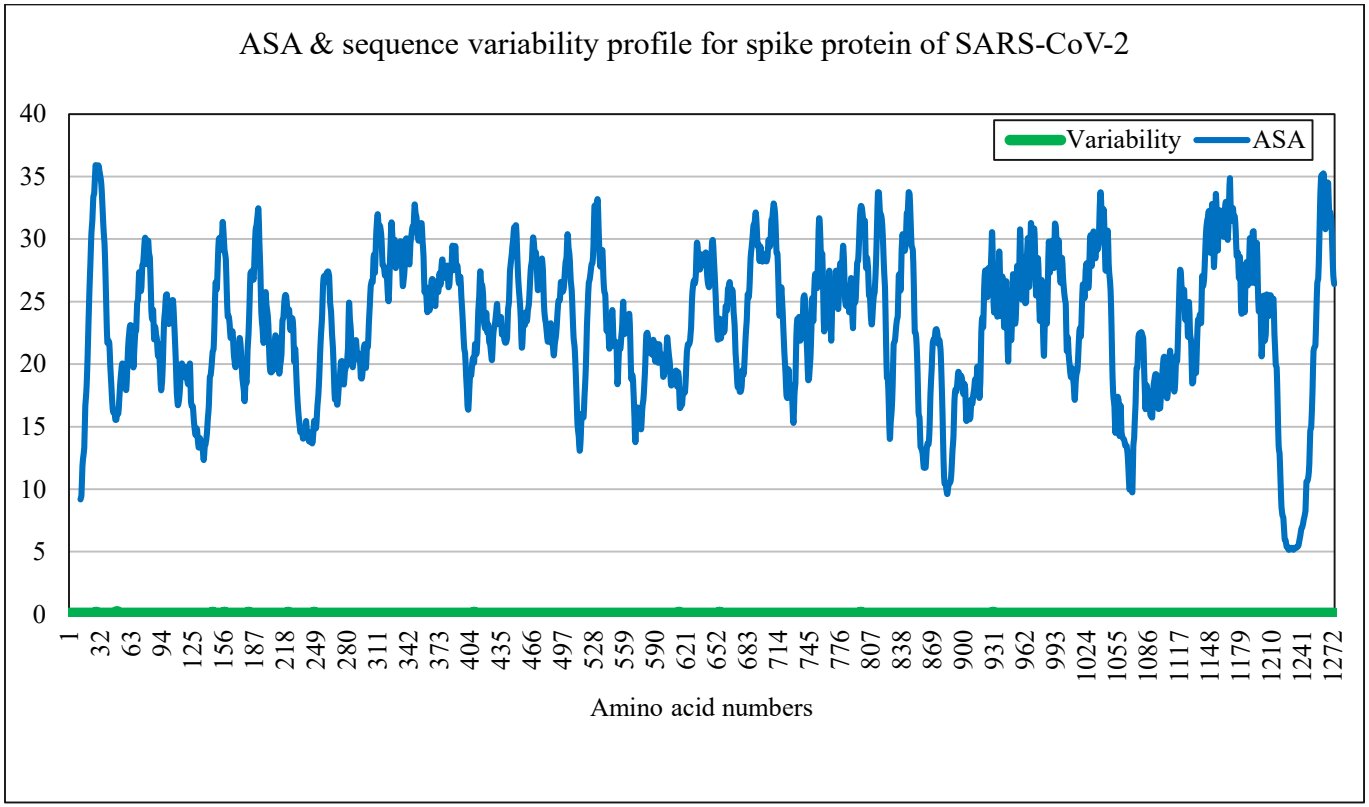
ASA & sequence variability profile for spike protein of SARS-CoV-2
Variability
ASA
40
35
30
25
20
15
10
5
0
1
32
63
94
125
156
187
218
249
280
311
342
373
404
435
466
497
528
559
590
621
652
683
714
745
776
807
838
869
900
931
962
993
1024
1055
1086
1117
1148
1179
1210
1241
1272
Amino acid numbers

**Figure 2.**

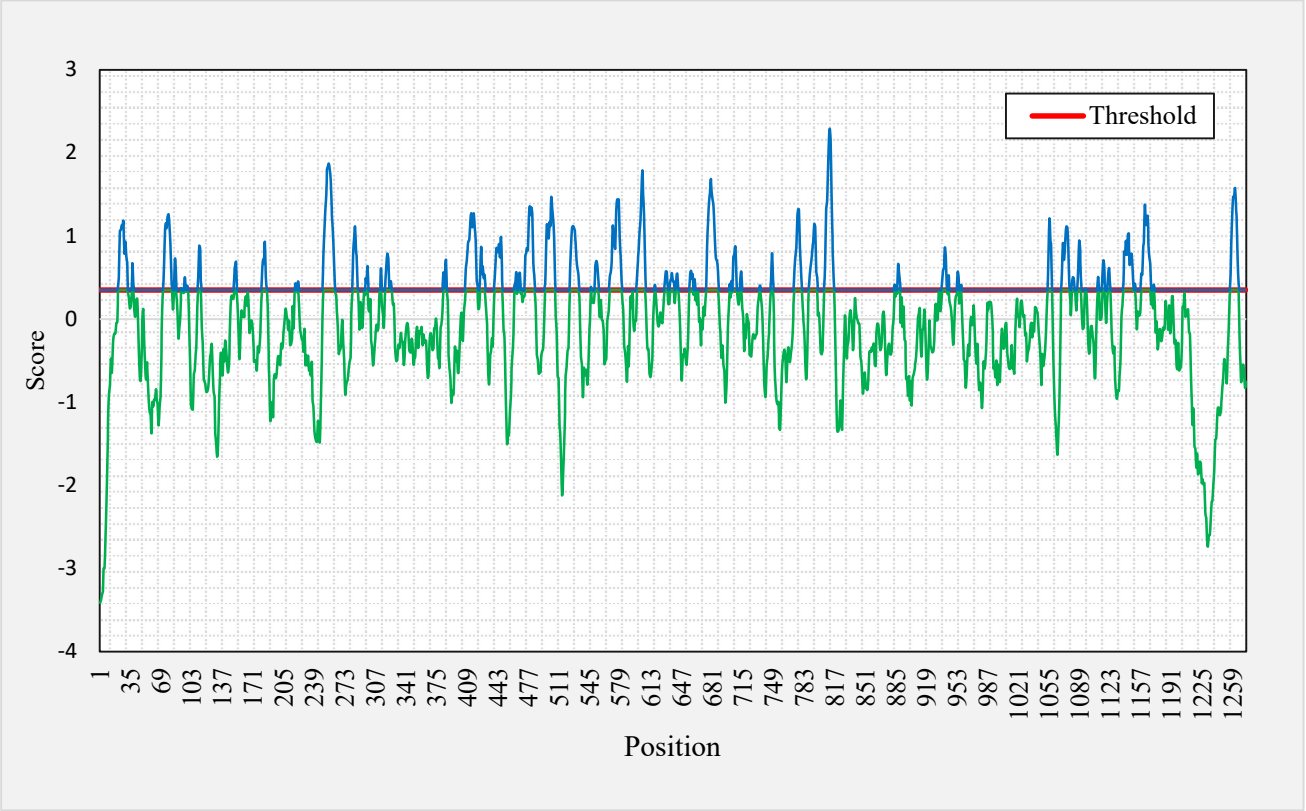
Threshold
Score
Position
3
2
1
0
-1
-2
-3
-4
1
35
69
103
137
171
205
239
273
307
341
375
409
443
477
511
545
579
613
647
681
715
749
783
817
851
885
919
953
987
1021
1055
1089
1123
1157
1191
1225
1259

**Figure 3.**

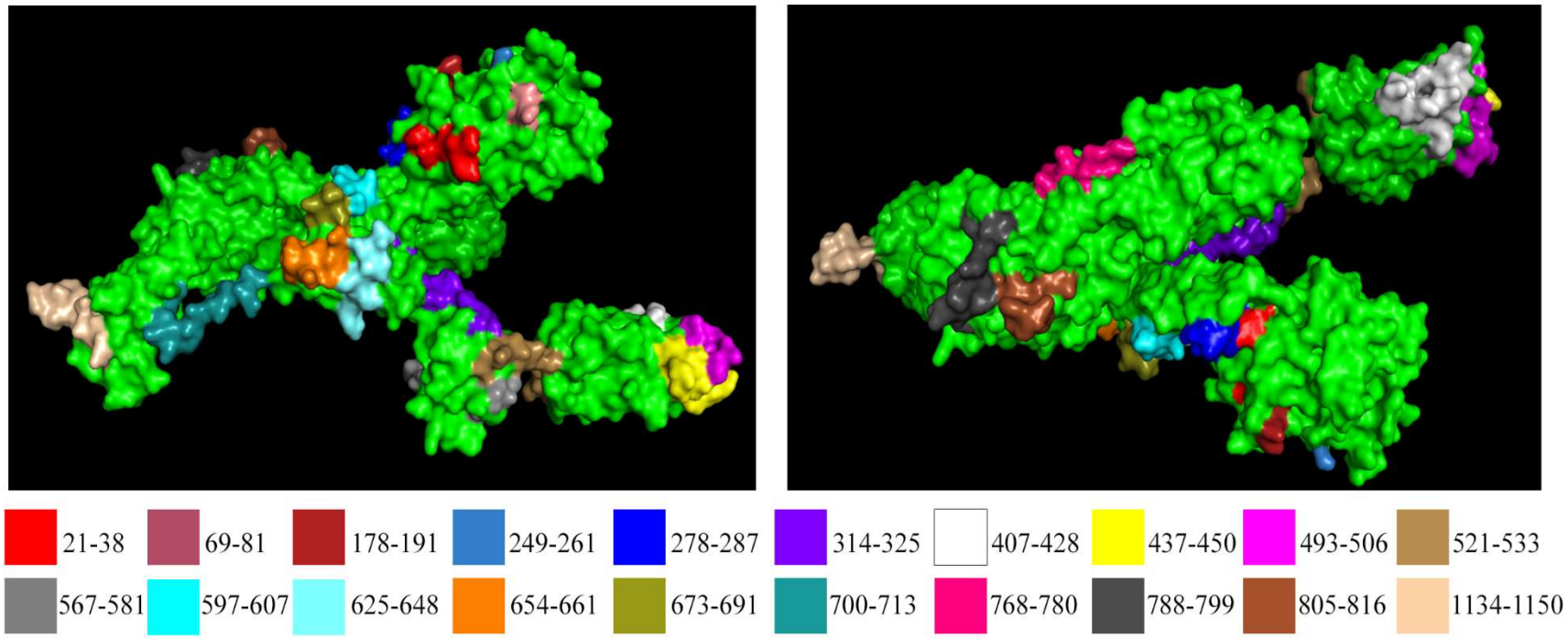
21-38
69-81
178-191
249-261
278-287
314-325
407-428
437-450
493-506
521-533
567-581
597-607
625-648
654-661
673-691
700-713
768-780
788-799
805-816
1134-1150

**Table 1.** Conserved region with good average solvent accessibility selected for further analysis.

| Sl. No. | Start | End | Length | Peptide |
|---|---|---|---|---|
| 1 | 21 | 38 | 18 | RTQLPPAYTNSFTRGVYY |
| 2 | 69 | 81 | 13 | HVSGTNGTKRFDN |
| 3 | 144 | 155 | 12 | YYHKNNKSWMES |
| 4 | 178 | 191 | 14 | DLEGKQGNFKNLRE |
| 5 | 249 | 261 | 13 | LTPGDSSSGWTAG |
| 6 | 278 | 287 | 10 | KYNENGTITD |
| 7 | 314 | 325 | 12 | QTSNFRVQPTES |
| 8 | 407 | 428 | 22 | VRQIAPGQTGKIADYNYKLPDD |
| 9 | 437 | 450 | 14 | NSNNLDSKVGGNYN |
| 10 | 461 | 485 | 25 | LKPFERDISTEIYQAGSTPCNGVEG |
| 11 | 493 | 506 | 14 | QSYGFQPTNGVGYQ |
| 12 | 521 | 533 | 13 | PATVCGPKKSTNL |
| 13 | 567 | 581 | 15 | RDIADTTDAVRDPQT |
| 14 | 597 | 607 | 11 | VITPGTNTSNQ |
| 15 | 625 | 648 | 24 | HADQLTPTWRVYSTGSNVFQTRAG |
| 16 | 654 | 661 | 8 | EHVNNSYE |
| 17 | 673 | 691 | 19 | SYQTQTNSPRRARSVASQS |
| 18 | 700 | 713 | 16 | GAENSVAYSNNSIA |
| 19 | 768 | 780 | 13 | TGIAVEQDKNTQE |
| 20 | 788 | 799 | 14 | IYKTPPIKDFGG |
| 21 | 805 | 816 | 12 | ILPDPSKPSKRS |
| 22 | 1134 | 1150 | 17 | NNTVYDPLQPELDSFKE |
| 23 | 1153 | 1171 | 19 | DKYFKNHTSPDVDLGDISG |
| 24 | 1255 | 1267 | 13 | KFDEDDSEPVLKG |

**Table 2.** IEDB ElliPro predicted linear epitopes for spike protein of SARS-CoV-2. Sequences that match our selected peptides are marked in red.

| Sl. No. | Start | End | Peptide | No. of residues | Score |
|---|---|---|---|---|---|
| 1 | 27 | 37 | AYTNSFTRGVY | 11 | 0.701 |
| 2 | 56 | 196 | LPFFSNVTWFHFDNPVLPFNDGVYFASTNIIRGWIFGTTLDSKTQSLLIVNNAT NVVIKVCEFQFCNDPFLGEFRVYSSANNCTFEYVSQPFLKNLREFVFKN | 103 | 0.851 |
| 3 | 280 | 286 | NENGTIT | 7 | 0.521 |
| 4 | 322 | 375 | PTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLYNSASFS | 54 | 0.646 |
| 5 | 393 | 515 | TNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSY NYLYRPLQSYGFQPTVGYQPYRVVVLSF | 80 | 0.842 |
| 6 | 464 | 511 | FERDISTEIYNCYFPLQSYGFQPTVGYQPYRVV | 33 | 0.707 |
| 7 | 465 | 509 | ERDISTENCYFPLQSYGFQVGYQPYR | 26 | 0.663 |
| 8 | 520 | 537 | APATVCGPKKSTNLVKNK | 18 | 0.617 |
| 9 | 577 | 585 | RDPQTLEIL | 9 | 0.665 |
| 10 | 603 | 608 | NTSNQV | 6 | 0.522 |
| 11 | 616 | 643 | NCTEVTGSNVF | 11 | 0.578 |
| 12 | 652 | 661 | GAEHVNNSYE | 10 | 0.594 |
| 13 | 687 | 691 | VASQS | 5 | 0.612 |
| 14 | 700 | 719 | GAENSVAYSNNSIAIPTNFT | 20 | 0.659 |
| 15 | 789 | 805 | YKTPPIKDFGGFNFSQI | 17 | 0.621 |
| 16 | 789 | 815 | YKTPPIKDFGGFNFSQILPDPSKR | 24 | 0.609 |
| 17 | 807 | 815 | PDPSKR | 6 | 0.558 |
| 18 | 1069 | 1146 | PAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTF VSGNCDVVIGIVNNTVYDPLQPELD | 78 | 0.832 |

**Table 3.** IEDB ElliPro predicted discontinuous epitopes for spike protein of SARS-CoV-2. Sequences that match our selected peptides are marked in red.

| Sl. No. | Residues | No. of residues | Score |
|---|---|---|---|
| 1 | A:D1139, A:P1140, A:L1141, A:Q1142, A:P1143, A:E1144, A:L1145, A:D1146 | 8 | 0.962 |

| | | | |
|---|---|---|---|
| 2 | A:T323, A:E324, A:S325, A:I326, A:V327, A:R328, A:L335, A:C336, A:P337, A:F338, A:G339, A:E340, A:V341, A:F342, A:N343, A:A344, A:T345, A:R346, A:F347, A:A348, A:S349, A:V350, A:Y351, A:A352, A:W353, A:N354, A:R355, A:K356, A:R357, A:I358, A:S359, A:N360, A:C361, A:V362, A:A363, A:D364, A:Y365, A:S366, A:V367, A:L368, A:Y369, A:N370, A:S371, A:A372, A:S373, A:F374, A:S375, A:T376, A:F377, A:K378, A:C379, A:Y380, A:V382, A:S383, A:P384, A:T385, A:L387, A:N388, A:L390, A:T393, A:N394, A:V395, A:Y396, A:A397, A:D398, A:S399, A:F400, A:V401, A:I402, A:R403, A:G404, A:D405, A:E406, A:V407, A:R408, A:Q409, A:I410, A:A411, A:P412, A:G413, A:Q414, A:T415, A:G416, A:K417, A:I418, A:A419, A:D420, A:Y421, A:N422, A:Y423, A:K424, A:L425, A:P426, A:D427, A:D428, A:F429, A:T430, A:G431, A:C432, A:V433, A:I434, A:A435, A:W436, A:N437, A:S438, A:N439, A:N440, A:L441, A:D442, A:S443, A:Y449, A:N450, A:Y451, A:L452, A:Y453, A:R454, A:P491, A:L492, A:Q493, A:S494, A:Y495, A:G496, A:F497, A:Q498, A:P499, A:T500, A:V503, A:G504, A:Y505, A:Q506, A:P507, A:Y508, A:R509, A:V510, A:V511, A:V512, A:L513, A:S514, A:F515, A:P521, A:A522, A:T523, A:V524, A:G526, A:P527, A:K528, A:K529, A:S530, A:T531, A:N532, A:L533, A:V534, A:K535, A:N536, A:K537, A:E554, A:S555, A:N556, A:K557, A:F559, A:L560, A:P561, A:F562, A:Q563, A:Q564, A:R577, A:D578, A:P579, A:Q580, A:T581, A:L582, A:E583, A:I584, B:A27, B:Y28, B:T29, B:N30, B:S31, B:F32, B:T33, B:R34, B:G35, B:V36, B:Y37, B:Y38, B:K41, B:L56, B:P57, B:F58, B:F59, B:S60, B:N61, B:V62, B:T63, B:W64, B:F65, B:H66, B:N81, B:P82, B:V83, B:L84, B:P85, B:F86, B:N87, B:D88, B:G89, B:V90, B:Y91, B:F92, B:A93, B:S94, B:T95, B:E96, B:K97, B:S98, B:N99, B:I100, B:I101, B:R102, B:G103, B:W104, B:I105, B:F106, B:G107, B:T108, B:T109, B:L110, B:D111, B:S112, B:K113, B:T114, B:Q115, B:S116, B:L117, B:L118, B:I119, B:V120, B:N121, B:N122, B:A123, B:T124, B:N125, B:V126, B:V127, B:I128, B:K129, B:V130, B:C131, B:E132, B:F133, B:Q134, B:F135, B:C136, B:N137, B:D138, B:P139, B:F140, B:L141, B:S155, B:E156, B:F157, B:V159, B:Y160, B:S161, B:S162, B:A163, B:N164, B:N165, B:C166, B:T167, B:F168, B:E169, B:Y170, B:V171, B:S172, B:Q173, B:P174, B:F175, B:L176, B:K187, B:N188, B:L189, B:R190, B:E191, B:F192, B:V193, B:F194, B:K195, B:G199, B:F201, B:I203, B:Y204, B:S205, B:K206, B:H207, B:T208, B:P209, B:P217, B:Q218, B:G219, B:F220, B:S221, B:A222, B:L223, B:E224, B:P225, B:L226, B:V227, B:D228, B:L229, B:P230, B:I231, B:G232, B:I233, B:N234, B:I235, B:T236, B:R237, B:F238, B:Q239, B:T240, B:L241, B:L242, B:A263, B:A264, B:Y265, B:Y266, B:V267, B:G268, B:Y269, B:L270 | 331 | 0.774 |

| | | | |
|---|---|---|---|
| 3 | B:P322, B:T323, B:E324, B:S325, B:I326, B:V327, B:R328, B:F329, B:P330, B:N331, B:I332, B:T333, B:N334, B:L335, B:C336, B:P337, B:F338, B:G339, B:E340, B:V341, B:F342, B:N343, B:A344, B:T345, B:R346, B:F347, B:A348, B:S349, B:V350, B:Y351, B:A352, B:W353, B:N354, B:R355, B:K356, B:R357, B:I358, B:S359, B:N360, B:C361, B:V362, B:A363, B:D364, B:S366, B:V367, B:L368, B:N370, B:S371, B:A372, B:S373, B:F374, B:C391, B:F392, B:T393, B:A397, B:D398, B:S399, B:F400, B:V401, B:I402, B:R403, B:G404, B:Y421, B:N422, B:Y423, B:W436, B:N437, B:S438, B:N439, B:N440, B:L441, B:D442, B:S443, B:Y449, B:N450, B:Y451, B:L452, B:Y453, B:R454, B:N460, B:L461, B:K462, B:E465, B:R466, B:D467, B:I468, B:S469, B:T470, B:E471, B:I472, B:Y473, B:N487, B:C488, B:Y489, B:F490, B:P491, B:L492, B:Q493, B:S494, B:Y495, B:G496, B:F497, B:Q498, B:P499, B:T500, B:V503, B:G504, B:Y505, B:Q506, B:P507, B:Y508, B:R509, B:V510, B:V511, B:A520, B:P521, B:A522, B:T523, B:V524, B:C525, B:G526, B:P527, B:K528, B:K529, B:S530, B:T531, B:N532, B:L533, B:V534, B:K535, B:N536, B:K537, B:N544, B:T553, B:E554, B:S555, B:N556, B:K557, B:F559, B:L560, B:P561, B:F562, B:Q563, B:Q564, B:R577, B:D578, B:P579, B:Q580, B:T581, B:L582, B:E583, B:I584, B:L585, C:A27, C:Y28, C:T29, C:N30, C:S31, C:F32, C:T33, C:R34, C:G35, C:V36, C:Y37, C:K41, C:F43, C:L56, C:P57, C:F58, C:F59, C:S60, C:N61, C:V62, C:T63, C:W64, C:F65, C:H66, C:D80, C:N81, C:P82, C:V83, C:L84, C:P85, C:F86, C:N87, C:D88, C:G89, C:V90, C:Y91, C:F92, C:A93, C:S94, C:T95, C:N99, C:I100, C:I101, C:R102, C:G103, C:W104, C:I105, C:F106, C:G107, C:T108, C:T109, C:L110, C:D111, C:S112, C:K113, C:T114, C:Q115, C:S116, C:L117, C:L118, C:I119, C:V120, C:N121, C:N122, C:A123, C:T124, C:N125, C:V126, C:V127, C:I128, C:K129, C:V130, C:C131, C:E132, C:F133, C:Q134, C:F135, C:C136, C:N137, C:D138, C:P139, C:F140, C:F157, C:R158, C:V159, C:Y160, C:S161, C:S162, C:A163, C:N164, C:N165, C:C166, C:T167, C:F168, C:E169, C:Y170, C:V171, C:S172, C:Q173, C:P174, C:F175, C:L176, C:K187, C:N188, C:L189, C:R190, C:E191, C:F192, C:V193, C:F194, C:K195, C:N196, C:D198, C:G199, C:F201, C:K202, C:I203, C:Y204, C:S205, C:K206, C:H207, C:T208, C:P209, C:I210, C:N211, C:L212, C:V213, C:R214, C:D215, C:L216, C:P217, C:Q218, C:G219, C:F220, C:S221, C:A222, C:L223, C:E224, C:P225, C:L226, C:V227, C:D228, C:L229, C:P230, C:I231, C:G232, C:I233, C:N234, C:I235, C:T236, C:R237, C:F238, C:Q239, C:T240, C:L241, C:L242, C:A243, C:L244, C:H245, C:G261, C:A262, C:A263, C:A264, C:Y265, C:Y266, C:V267, C:G268, C:Y269, C:L270, C:T286 | 323 | 0.737 |

| | | | |
|---|---|---|---|
| 4 | A:A701, A:E702, A:N703, A:S704, A:V705, A:A706, A:Y707, A:S708, A:N709, A:N710, A:S711, A:I712, A:A713, A:I714, A:P715, A:T716, A:N717, A:F718, A:T719, A:Q787, A:Y789, A:K790, A:T791, A:P792, A:P793, A:I794, A:K795, A:D796, A:F797, A:G798, A:G799, A:F800, A:N801, A:F802, A:S803, A:Q804, A:I805, A:L806, A:T883, A:G891, A:A892, A:A893, A:L894, A:Q895, A:I896, A:P897, A:F898, A:A899, A:M900, A:A903, A:F906, A:N907, A:G910, A:V911, A:T912, A:Q913, A:N914, A:V915, A:L916, A:Y917, A:E918, A:N919, A:Q920, A:L922, A:I923, A:A924, A:N925, A:Q926, A:F927, A:N928, A:S929, A:G932, A:D936, A:P1069, A:A1070, A:Q1071, A:E1072, A:K1073, A:N1074, A:F1075, A:T1076, A:T1077, A:A1078, A:P1079, A:A1080, A:I1081, A:C1082, A:H1083, A:D1084, A:G1085, A:K1086, A:A1087, A:H1088, A:F1089, A:P1090, A:R1091, A:E1092, A:G1093, A:V1094, A:F1095, A:V1096, A:S1097, A:N1098, A:G1099, A:T1100, A:H1101, A:W1102, A:F1103, A:V1104, A:T1105, A:Q1106, A:R1107, A:N1108, A:F1109, A:Y1110, A:E1111, A:P1112, A:Q1113, A:I1114, A:I1115, A:T1116, A:T1117, A:D1118, A:N1119, A:T1120, A:F1121, A:V1122, A:S1123, A:G1124, A:N1125, A:C1126, A:D1127, A:V1128, A:V1129, A:I1130, A:G1131, A:I1132, A:V1133, A:N1134, A:N1135, A:T1136, A:V1137, A:Y1138, B:G700, B:A701, B:E702, B:N703, B:S704, B:V705, B:A706, B:Y707, B:S708, B:N709, B:N710, B:S711, B:I712, B:A713, B:I714, B:P715, B:T716, B:N717, B:F718, B:T719, B:Q787, B:I788, B:Y789, B:K790, B:T791, B:P792, B:P793, B:I794, B:K795, B:D796, B:F797, B:G798, B:G799, B:F800, B:N801, B:F802, B:S803, B:Q804, B:I805, B:L806, B:T883, B:G891, B:A892, B:A893, B:L894, B:Q895, B:I896, B:P897, B:F898, B:A899, B:M900, B:M902, B:A903, B:F906, B:N907, B:G910, B:V911, B:T912, B:Q913, B:N914, B:V915, B:L916, B:Y917, B:E918, B:N919, B:Q920, B:K921, B:L922, B:I923, B:A924, B:N925, B:Q926, B:F927, B:N928, B:S929, B:P1069, B:A1070, B:Q1071, B:E1072, B:K1073, B:N1074, B:F1075, B:T1076, B:T1077, B:A1078, B:P1079, B:A1080, B:I1081, B:C1082, B:H1083, B:D1084, B:G1085, B:K1086, B:A1087, B:H1088, B:F1089, B:P1090, B:R1091, B:E1092, B:G1093, B:V1094, B:F1095, B:V1096, B:S1097, B:N1098, B:G1099, B:T1100, B:H1101, B:W1102, B:F1103, B:V1104, B:T1105, B:Q1106, B:R1107, B:N1108, B:F1109, B:Y1110, B:E1111, B:P1112, B:Q1113, B:I1114, B:I1115, B:T1116, B:T1117, B:D1118, B:N1119, B:T1120, B:F1121, B:V1122, B:S1123, B:G1124, B:N1125, B:C1126, B:D1127, B:V1128, B:V1129, B:I1130, B:G1131, B:I1132, B:V1133, B:N1134, B:N1135, B:T1136, B:V1137, B:Y1138, B:D1139, B:P1140, B:L1141, B:Q1142, B:P1143, B:E1144, B:L1145, B:D1146, C:A701, C:E702, C:N703, C:S704, C:V705, C:A706, C:Y707, C:S708, C:N709, C:N710, C:S711, C:I712, C:A713, C:I714, C:P715, C:T716, C:N717, C:F718, C:T719, C:Q787, C:I788, C:Y789, C:K790, C:T791, C:P792, C:P793, C:I794, C:K795, C:D796, C:F797, C:G798, C:G799, C:F800, C:N801, C:F802, C:S803, C:Q804, C:I805, C:L806, C:P807, C:D808, C:P809, C:S810, | 452 | 0.724 |

| | | | |
|---|---|---|---|
| | C:K811, C:R815, C:F817, C:T883, C:G891, C:A892, C:A893, C:L894, C:Q895, C:I896, C:P897, C:F898, C:A899, C:M900, C:A903, C:F906, C:N907, C:G910, C:V911, C:T912, C:Q913, C:N914, C:V915, C:L916, C:Y917, C:E918, C:N919, C:Q920, C:K921, C:L922, C:I923, C:A924, C:N925, C:Q926, C:N928, C:S929, C:A1070, C:Q1071, C:E1072, C:K1073, C:N1074, C:F1075, C:T1076, C:T1077, C:A1078, C:P1079, C:A1080, C:I1081, C:C1082, C:H1083, C:D1084, C:G1085, C:K1086, C:A1087, C:H1088, C:F1089, C:P1090, C:R1091, C:E1092, C:G1093, C:V1094, C:F1095, C:V1096, C:S1097, C:N1098, C:G1099, C:T1100, C:H1101, C:W1102, C:F1103, C:V1104, C:T1105, C:Q1106, C:R1107, C:N1108, C:F1109, C:Y1110, C:E1111, C:P1112, C:Q1113, C:I1114, C:I1115, C:T1116, C:T1117, C:D1118, C:N1119, C:T1120, C:F1121, C:V1122, C:S1123, C:G1124, C:N1125, C:C1126, C:D1127, C:V1128, C:V1129, C:I1130, C:G1131, C:I1132, C:V1133, C:N1134, C:N1135, C:T1136, C:V1137, C:Y1138, C:D1139, C:P1140, C:L1141, C:Q1142, C:P1143, C:E1144, C:L1145, C:D1146 | | |

| | | | |
|---|---|---|---|
| 5 | A:A27, A:Y28, A:T29, A:N30, A:S31, A:F32, A:T33, A:R34, A:G35, A:V36, A:Y37, A:Y38, A:K41, A:F43, A:L56, A:P57, A:F58, A:F59, A:S60, A:N61, A:V62, A:T63, A:W64, A:F65, A:F79, A:D80, A:N81, A:P82, A:V83, A:L84, A:P85, A:F86, A:N87, A:D88, A:G89, A:V90, A:Y91, A:F92, A:A93, A:S94, A:T95, A:I100, A:I101, A:R102, A:G103, A:W104, A:I105, A:F106, A:G107, A:T108, A:T109, A:L110, A:D111, A:S112, A:K113, A:T114, A:Q115, A:S116, A:L117, A:L118, A:I119, A:V120, A:N121, A:N122, A:A123, A:T124, A:N125, A:V126, A:V127, A:I128, A:K129, A:V130, A:C131, A:E132, A:F133, A:Q134, A:F135, A:C136, A:N137, A:D138, A:P139, A:F140, A:L141, A:G142, A:E156, A:F157, A:R158, A:V159, A:Y160, A:S161, A:S162, A:A163, A:N164, A:N165, A:C166, A:T167, A:F168, A:E169, A:Y170, A:V171, A:S172, A:Q173, A:P174, A:F175, A:L176, A:K187, A:N188, A:L189, A:R190, A:E191, A:F192, A:V193, A:F194, A:N196, A:G199, A:F201, A:K202, A:I203, A:Y204, A:S205, A:K206, A:H207, A:T208, A:P209, A:I210, A:N211, A:L212, A:V213, A:R214, A:D215, A:L216, A:P217, A:Q218, A:G219, A:F220, A:S221, A:A222, A:L223, A:E224, A:P225, A:L226, A:V227, A:D228, A:L229, A:P230, A:I231, A:G232, A:I233, A:N234, A:I235, A:T236, A:R237, A:F238, A:Q239, A:T240, A:L241, A:L242, A:A243, A:L244, A:H245, A:R246, A:G261, A:A262, A:A263, A:A264, A:Y265, A:Y266, A:V267, A:G268, A:Y269, A:L270, A:R273, A:N280, A:E281, A:N282, A:G283, A:T284, A:T286, C:P322, C:T323, C:E324, C:S325, C:I326, C:V327, C:R328, C:F329, C:P330, C:N331, C:I332, C:T333, C:N334, C:L335, C:C336, C:P337, C:F338, C:G339, C:E340, C:V341, C:F342, C:N343, C:A344, C:T345, C:R346, C:F347, C:A348, C:S349, C:V350, C:Y351, C:A352, C:N354, C:R357, C:I358, C:S359, C:N360, C:C361, C:V362, C:A363, C:D364, C:S366, C:V367, C:N370, C:S371, C:A372, C:S373, C:F374, C:T393, C:S399, C:V401, C:G404, C:W436, C:N437, C:S438, C:N439, C:N440, C:L441, C:D442, C:S443, C:Y449, C:N450, C:Y451, C:L452, C:Y453, C:R454, C:E465, C:R466, C:D467, C:I468, C:S469, C:T470, C:E471, C:F490, C:P491, C:L492, C:Q493, C:S494, C:Y495, C:G496, C:F497, C:Q498, C:V503, C:G504, C:Y505, C:Q506, C:P507, C:Y508, C:R509, C:A522, C:T523, C:G526, C:P527, C:K528, C:K529, C:S530, C:T531, C:N532, C:L533, C:V534, C:K535, C:N536, C:K537, C:T553, C:E554, C:S555, C:N556, C:K557, C:K558, C:F559, C:L560, C:P561, C:F562, C:Q563, C:Q564, C:R577, C:D578, C:P579, C:Q580, C:T581, C:L582, C:E583, C:I584, C:L585 | 301 | 0.723 |
| 6 | A:V687, A:A688, A:S689, A:Q690, A:S691 | 5 | 0.612 |
| 7 | B:T638, B:G639, B:S640, B:N641, B:V642, B:F643, B:G652, B:A653, B:E654, B:H655, B:V656, B:N657, B:N658, B:S659, B:Y660, B:A672, B:V687, B:A688, B:S689, B:Q690, B:S691, B:I692, B:I693, B:A694, B:S698 | 25 | 0.607 |
| 8 | A:P807, A:D808, A:P809, A:S810, A:K811 | 5 | 0.596 |

| | | | |
|---|---|---|---|
| 9 | C:N641, C:V642, C:G652, C:A653, C:E654, C:H655, C:V656, C:N657, C:N658, C:S659, C:Y660, C:V687, C:A688, C:S689, C:Q690, C:S691, C:I693 | 17 | 0.592 |
| 10 | A:H655, A:V656, A:N657, A:N658, A:S659, A:Y660 | 6 | 0.57 |
| 11 | A:S640, A:N641, A:V642, A:G652, A:A653 | 5 | 0.551 |
| 12 | B:P807, B:D808, B:P809, B:S810, B:K811, B:R815 | 6 | 0.549 |
| 13 | B:G932, B:K933, B:D936 | 3 | 0.532 |
| 14 | B:N280, B:E281, B:N282, B:G283, B:T284, B:T286 | 6 | 0.523 |
| 15 | B:K558, C:S45, C:N280, C:E281, C:N282, C:G283, C:T284 | 7 | 0.519 |
| 16 | C:G932, C:Q935, C:D936 | 3 | 0.519 |

**Table 4.** ABCpred determination of B-cell binding affinities. Note that high score indicates good binding affinity.

| Sl. No. | Sequence | Start | Score |
|---|---|---|---|
| 1 | PPAYTNSFTRGVYY | 25 | 0.91 |
| 2 | IHVSGTNGTKRFDNPVLPFN | 68 | 0.89 |
| 3 | VYYHKNNKSWMESEFRVYSS | 143 | 0.9 |
| 4 | DLEGKQGNFKNLREFVFKNI | 178 | 0.82 |
| 5 | YLTPGDSSSGWT | 248 | 0.7 |
| 6 | LLKYNENGTITDAVDCALDP | 276 | 0.76 |
| 7 | IYQTSNFRVQPTES | 312 | 0.68 |
| 8 | RQIAPGQTGKIADYNYKLPD | 408 | 0.75 |
| 9 | WNSNNLDSKVGGNYNYLY | 436 | 0.67 |
| 10 | SNLKPFERDISTEIYQAGST | 459 | 0.82 |
| 11 | LQSYGFQPTNGVGYQP | 492 | 0.9 |
| 12 | HAPATVCGPKKSTN | 519 | 0.72 |
| 13 | QQFGRDIADTTDAVRDPQTL | 563 | 0.82 |
| 14 | VITPGTNTSNQVAV | 597 | 0.77 |
| 15 | AIHADQLTPTWRVYSTGS | 623 | 0.67 |
| 16 | IGAEHVNNSYECDIPIGAGI | 651 | 0.9 |
| 17 | YQTQTNSPRRARSVASQS | 674 | 0.82 |
| 18 | GAENSVAYSNNSIAIPTN | 700 | 0.63 |
| 19 | AVEQDKNTQEVFAQ | 771 | 0.89 |
| 20 | IYKTPPIKDFGGFN | 788 | 0.77 |
| 21 | ILPDPSKPSKRSFIEDLL | 805 | 0.63 |
| 22 | VIGIVNNTVYDPLQPE | 1129 | 0.83 |
| 23 | DKYFKNHTSPDVDLGD | 1153 | 0.69 |
| 24 | CSCGSCCKFDEDDSEPVLKG | 1248 | 0.73 |

**Table 5.** IEDB prediction of binding affinity with MHC-I alleles, only our selected peptides with percentile rank less than 1.00 are shown here. The binding affinity is considered higher for low percentile rank. Sequences that match our selected peptides are marked in red.

| Sl. No. | Allele | Start | End | Peptide | Method | Percentile Rank |
|---|---|---|---|---|---|---|
| 1 | HLA-A*01:01 | 17 | 28 | NLTTRTQLPPAY | ann | 0.21 |
| 2 | HLA-A*30:01 | 19 | 27 | TTRTQLPPA | Consensus | 0.2 |
| 3 | HLA-A*31:01 | 21 | 34 | RTQLPPAYTNSFTR | ann | 0.47 |
| 4 | HLA-A*68:01 | 21 | 34 | RTQLPPAYTNSFTR | ann | 0.82 |
| 5 | HLA-B*15:01 | 23 | 32 | QLPPAYTNSF | Consensus | 0.43 |
| 6 | HLA-A*30:02 | 24 | 37 | LPPAYTNSFTRGVY | ann | 0.76 |
| 7 | HLA-B*07:02 | 24 | 32 | LPPAYTNSF | Consensus | 0.6 |
| 8 | HLA-B*35:01 | 24 | 32 | LPPAYTNSF | Consensus | 0.6 |
| 9 | HLA-A*01:01 | 25 | 38 | PPAYTNSFTRGVYY | ann | 0.04 |
| 10 | HLA-A*01:01 | 27 | 38 | AYTNSFTRGVYY | ann | 0.04 |
| 11 | HLA-A*68:01 | 69 | 78 | HVSGTNGTKR | Consensus | 0.41 |
| 12 | HLA-A*03:01 | 142 | 150 | GVYYHKNNK | Consensus | 0.2 |
| 13 | HLA-A*11:01 | 142 | 150 | GVYYHKNNK | Consensus | 0.43 |
| 14 | HLA-A*23:01 | 144 | 157 | YYHKNNKSWMESEF | ann | 0.62 |
| 15 | HLA-A*24:02 | 144 | 157 | YYHKNNKSWMESEF | ann | 0.38 |
| 16 | HLA-A*33:01 | 145 | 158 | YHKNNKSWMESEFR | ann | 0.56 |
| 17 | HLA-A*11:01 | 182 | 195 | KQGNFKNLREFVFK | ann | 0.99 |
| 18 | HLA-A*31:01 | 182 | 190 | KQGNFKNLR | Consensus | 0.19 |
| 19 | HLA-A*23:01 | 185 | 194 | NFKNLREFVF | Consensus | 0.71 |
| 20 | HLA-A*33:01 | 185 | 195 | NFKNLREFVFK | ann | 0.67 |
| 21 | HLA-B*44:03 | 187 | 200 | KNLREFVFKNIDGY | ann | 0.64 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 22 | HLA-B*57:01 | 246 | 259 | RSYLTPGDSSSGWT | ann | 0.38 |
| 23 | HLA-B*58:01 | 246 | 259 | RSYLTPGDSSSGWT | ann | 0.47 |
| 24 | HLA-B*53:01 | 250 | 258 | TPGDSSSGW | Consensus | 0.2 |
| 25 | HLA-A*30:02 | 252 | 265 | GDSSSGWTAGAAAY | ann | 0.28 |
| 26 | HLA-A*26:01 | 253 | 266 | DSSSGWTAGAAAYY | ann | 0.11 |
| 27 | HLA-A*01:01 | 254 | 266 | SSSGWTAGAAAYY | ann | 0.34 |
| 28 | HLA-A*68:02 | 254 | 267 | SSSGWTAGAAAYYV | ann | 0.39 |
| 29 | HLA-A*30:02 | 255 | 266 | SSGWTAGAAAYY | ann | 0.09 |
| 30 | HLA-A*68:02 | 281 | 289 | ENGTITDAV | Consensus | 0.69 |
| 31 | HLA-A*33:01 | 306 | 319 | FTVEKGIYQTSNFR | ann | 0.72 |
| 32 | HLA-A*68:01 | 306 | 319 | FTVEKGIYQTSNFR | ann | 0.37 |
| 33 | HLA-A*31:01 | 310 | 319 | KGIYQTSNFR | Consensus | 0.28 |
| 34 | HLA-A*23:01 | 312 | 320 | IYQTSNFRV | Consensus | 0.68 |
| 35 | HLA-A*24:02 | 312 | 320 | IYQTSNFRV | Consensus | 0.34 |
| 36 | HLA-A*03:01 | 408 | 417 | RQIAPGQTGK | Consensus | 0.63 |
| 37 | HLA-A*30:02 | 408 | 421 | RQIAPGQTGKIADY | ann | 0.61 |
| 38 | HLA-B*15:01 | 408 | 421 | RQIAPGQTGKIADY | ann | 0.3 |
| 39 | HLA-A*02:01 | 417 | 425 | KIADYNYKL | Consensus | 0.7 |
| 40 | HLA-A*32:01 | 417 | 425 | KIADYNYKL | Consensus | 0.5 |
| 41 | HLA-A*01:01 | 440 | 453 | NLDSKVGGNYNYLY | ann | 0.08 |
| 42 | HLA-A*30:02 | 441 | 453 | LDSKVGGNYNYLY | ann | 0.31 |
| 43 | HLA-A*31:01 | 441 | 454 | LDSKVGGNYNYLYR | ann | 0.47 |
| 44 | HLA-A*33:01 | 442 | 454 | DSKVGGNYNYLYR | ann | 0.67 |
| 45 | HLA-A*24:02 | 443 | 456 | SKVGGNYNYLYRLF | ann | 0.46 |
| 46 | HLA-A*33:01 | 444 | 457 | KVGGNYNYLYRLFR | ann | 0.16 |
| 47 | HLA-A*33:01 | 455 | 466 | LFRKSNLKPFER | ann | 0.96 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 48 | HLA-A*31:01 | 458 | 466 | KSNLKPFER | Consensus | 0.14 |
| 49 | HLA-A*01:01 | 460 | 473 | NLKPFERDISTEIY | ann | 0.63 |
| 50 | HLA-B*40:01 | 464 | 472 | FERDISTEI | Consensus | 0.88 |
| 51 | HLA-B*15:01 | 473 | 486 | YQAGSTPCNGVEGF | ann | 0.97 |
| 52 | HLA-B*35:01 | 478 | 489 | TPCNGVEGFNCY | ann | 0.81 |
| 53 | HLA-B*40:01 | 479 | 492 | PCNGVEGFNCYFPL | ann | 0.31 |
| 54 | HLA-A*23:01 | 488 | 497 | CYFPLQSYGF | Consensus | 0.13 |
| 55 | HLA-A*24:02 | 488 | 497 | CYFPLQSYGF | Consensus | 0.12 |
| 56 | HLA-A*24:02 | 488 | 501 | CYFPLQSYGFQPTN | ann | 0.7 |
| 57 | HLA-B*35:01 | 490 | 497 | FPLQSYGF | Consensus | 0.71 |
| 58 | HLA-B*53:01 | 490 | 497 | FPLQSYGF | Consensus | 0.71 |
| 59 | HLA-B*15:01 | 492 | 505 | LQSYGFQPTNGVGY | ann | 0.14 |
| 60 | HLA-A*68:02 | 495 | 503 | YGFQPTNGV | Consensus | 0.8 |
| 61 | HLA-A*01:01 | 499 | 508 | PTNGVGYQPY | Consensus | 0.8 |
| 62 | HLA-A*03:01 | 517 | 529 | LLHAPATVCGPKK | ann | 0.82 |
| 63 | HLA-B*07:02 | 520 | 533 | APATVCGPKKSTNL | ann | 0.9 |
| 64 | HLA-A*68:01 | 568 | 577 | DIADTTDAVR | Consensus | 0.34 |
| 65 | HLA-A*68:02 | 568 | 576 | DIADTTDAV | Consensus | 0.7 |
| 66 | HLA-A*01:01 | 599 | 612 | TPGTNTSNQVAVLY | ann | 0.55 |
| 67 | HLA-A*30:02 | 603 | 612 | NTSNQVAVLY | Consensus | 0.79 |
| 68 | HLA-B*53:01 | 620 | 633 | VPVAIHADQLTPTW | ann | 0.21 |
| 69 | HLA-B*57:01 | 622 | 633 | VAIHADQLTPTW | ann | 0.47 |
| 70 | HLA-B*58:01 | 622 | 635 | VAIHADQLTPTWRV | ann | 0.27 |
| 71 | HLA-B*35:01 | 625 | 636 | HADQLTPTWRVY | ann | 0.54 |
| 72 | HLA-B*53:01 | 625 | 633 | HADQLTPTW | Consensus | 0.2 |
| 73 | HLA-A*24:02 | 630 | 643 | TPTWRVYSTGSNVF | ann | 0.51 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 74 | HLA-A*31:01 | 633 | 646 | WRVYSTGSNVFQTR | ann | 0.29 |
| 75 | HLA-A*23:01 | 634 | 643 | RVYSTGSNVF | Consensus | 0.62 |
| 76 | HLA-B*15:01 | 634 | 643 | RVYSTGSNVF | Consensus | 0.21 |
| 77 | HLA-A*68:01 | 637 | 646 | STGSNVFQTR | Consensus | 0.96 |
| 78 | HLA-A*02:06 | 643 | 651 | FQTRAGCLI | Consensus | 0.64 |
| 79 | HLA-A*30:02 | 651 | 660 | IGAEHVNNSY | Consensus | 0.35 |
| 80 | HLA-A*30:01 | 683 | 691 | RARSVASQS | Consensus | 0.5 |
| 81 | HLA-A*30:02 | 683 | 695 | RARSVASQSIIAY | ann | 0.46 |
| 82 | HLA-A*31:01 | 673 | 682 | SYQTQTNSPR | Consensus | 0.64 |
| 83 | HLA-A*33:01 | 673 | 682 | SYQTQTNSPR | Consensus | 0.7 |
| 84 | HLA-A*33:01 | 677 | 685 | QTNSPRRAR | Consensus | 0.55 |
| 85 | HLA-B*07:02 | 676 | 689 | TQTNSPRRARSVAS | ann | 0.7 |
| 86 | HLA-B*07:02 | 679 | 692 | NSPRRARSVASQSI | ann | 0.09 |
| 87 | HLA-B*07:02 | 680 | 692 | SPRRARSVASQSI | ann | 0.04 |
| 88 | HLA-B*08:01 | 680 | 688 | SPRRARSVA | Consensus | 0.72 |
| 89 | HLA-B*57:01 | 685 | 697 | RSVASQSIIAYTM | ann | 0.7 |
| 90 | HLA-B*58:01 | 685 | 693 | RSVASQSII | Consensus | 0.4 |
| 91 | HLA-A*23:01 | 705 | 718 | VAYSNNSIAIPTNF | ann | 0.65 |
| 92 | HLA-A*32:01 | 704 | 712 | SVAYSNNSI | Consensus | 0.8 |
| 93 | HLA-A*68:02 | 703 | 712 | NSVAYSNNSI | Consensus | 0.99 |
| 94 | HLA-B*15:01 | 699 | 707 | LGAENSVAY | Consensus | 0.8 |
| 95 | HLA-B*35:01 | 699 | 707 | LGAENSVAY | Consensus | 0.3 |
| 96 | HLA-B*44:02 | 700 | 712 | GAENSVAYSNNSI | ann | 0.49 |
| 97 | HLA-B*44:02 | 701 | 714 | AENSVAYSNNSIAI | ann | 0.12 |
| 98 | HLA-B*44:03 | 701 | 714 | AENSVAYSNNSIAI | ann | 0.31 |
| 99 | HLA-A*68:02 | 773 | 785 | EQDKNTQEVFAQV | ann | 0.71 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 100 | HLA-A*02:06 | 786 | 794 | KQIYKTPPI | Consensus | 0.38 |
| 101 | HLA-A*03:01 | 786 | 795 | KQIYKTPPIK | Consensus | 0.47 |
| 102 | HLA-A*30:01 | 786 | 795 | KQIYKTPPIK | Consensus | 0.52 |
| 103 | HLA-A*32:01 | 786 | 794 | KQIYKTPPI | Consensus | 0.3 |
| 104 | HLA-B*15:01 | 786 | 797 | KQIYKTPPIKDF | ann | 0.73 |
| 105 | HLA-A*03:01 | 787 | 795 | QIYKTPPIK | Consensus | 0.27 |
| 106 | HLA-B*07:02 | 811 | 818 | KPSKRSFI | Consensus | 0.95 |
| 107 | HLA-B*57:01 | 811 | 823 | KPSKRSFIEDLLF | ann | 0.65 |
| 108 | HLA-A*03:01 | 1136 | 1149 | TVYDPLQPELDSFK | ann | 0.15 |
| 109 | HLA-A*11:01 | 1136 | 1149 | TVYDPLQPELDSFK | ann | 0.14 |
| 110 | HLA-A*68:01 | 1136 | 1149 | TVYDPLQPELDSFK | ann | 0.34 |
| 111 | HLA-A*01:01 | 1142 | 1155 | QPELDSFKEELDKY | ann | 0.55 |
| 112 | HLA-B*40:01 | 1257 | 1265 | DEDDSEPVL | Consensus | 0.86 |
| 113 | HLA-A*01:01 | 1259 | 1272 | DDSEPVLKGVKLHY | ann | 0.64 |

**Table 6.** IEDB prediction of binding affinity with MHC-II alleles, only our selected peptides with percentile rank less than 1.00 are shown here. The binding affinity is considered higher for low percentile rank. Sequences that match our selected peptides are marked in red.

| Sl. No. | Allele | Start | End | Method | Peptide | Percentile Rank |
|---|---|---|---|---|---|---|
| 1 | HLA-DPA1*01:03/DPB1*04:01 | 168 | 185 | NetMHCIIpan | FEYVSQPFLMDLEGKQGN | 0.87 |
| 2 | HLA-DPA1*02:01/DPB1*05:01 | 183 | 197 | Consensus | QGNFKNLREFVFKNI | 0.74 |
| 3 | HLA-DPA1*02:01/DPB1*05:01 | 184 | 197 | Consensus | GNFKNLREFVFKNI | 0.62 |
| 4 | HLA-DQA1*05:01/DQB1*03:01 | 255 | 268 | Consensus | SSGWTAGAAAYYVG | 0.76 |
| 5 | HLA-DQA1*05:01/DQB1*03:01 | 255 | 269 | Consensus | SSGWTAGAAAYYVGY | 0.94 |
| 6 | HLA-DRB1*04:01 | 314 | 331 | Consensus | QTSNFRVQPTESIVRFPN | 1 |
| 7 | HLA-DRB1*15:01 | 319 | 335 | Consensus | RVQPTESIVRFPNITNL | 0.95 |
| 8 | HLA-DRB3*01:01 | 400 | 413 | Consensus | FVIRGDEVRQIAPG | 0.47 |
| 9 | HLA-DRB3*01:01 | 402 | 418 | Consensus | IRGDEVRQIAPGQTGKI | 0.94 |
| 10 | HLA-DRB1*11:01 | 443 | 460 | Consensus | SKVGGNYNYLYRLFRKSN | 0.65 |
| 11 | HLA-DRB1*11:01 | 444 | 457 | Consensus | KVGGNYNYLYRLFR | 0.9 |
| 12 | HLA-DRB3*01:01 | 461 | 472 | Consensus | LKPFERDISTEI | 0.79 |
| 13 | HLA-DPA1*02:01/DPB1*01:01 | 501 | 518 | Consensus | NGVGYQPYRVVVLSFELL | 0.38 |
| 14 | HLA-DPA1*01:03/DPB1*02:01 | 501 | 518 | Consensus | NGVGYQPYRVVVLSFELL | 0.6 |
| 15 | HLA-DRB1*01:01 | 514 | 526 | Consensus | SFELLHAPATVCG | 0.02 |
| 16 | HLA-DRB1*01:01 | 515 | 528 | Consensus | FELLHAPATVCGPK | 0.49 |
| 17 | HLA-DRB1*07:01 | 684 | 701 | Consensus | ARSVASQSIIAYTMSLGA | 0.77 |
| 18 | HLA-DQA1*04:01/DQB1*04:02 | 763 | 775 | Consensus | LNRALTGIAVEQD | 0.73 |
| 19 | HLA-DQA1*03:01/DQB1*03:02 | 764 | 776 | Consensus | NRALTGIAVEQDK | 0.91 |
| 20 | HLA-DPA1*01:03/DPB1*04:01 | 809 | 826 | NetMHCIIpan | PSKPSKRSFIEDLLFNKV | 0.59 |
| 21 | HLA-DRB1*13:02 | 1127 | 1141 | Consensus | DVVIGIVNNTVYDPL | 0.7 |