Structural basis to design multi-epitope vaccines against Novel Coronavirus 19 (COVID19) infection, the ongoing pandemic emergency: an in silico approach

Sukrit Srivastava*[a,b], Sonia Verma[c], Mohit Kamthania[d], Rupinder Kaur[e], Ruchi Kiran Badyal[f], Ajay Kumar Saxena[b], Ho-Joon Shin[g], Kailash Pandey[c], Michael Kolbe[h,i]

a Department of Biotechnology, Institute of Bio-Medical Education and Research, Mangalayatan University, 202146 Aligarh, India.
b Molecular Medicine Lab., School of Life Science, Jawaharlal Nehru University, 110067 New Delhi, India.
c National Institute of Malaria Research-Indian Council of Medical Research, New Delhi, India
d Department of Biotechnology, Institute of Applied Medicines and Research, 201206 Ghaziabad, India.
e Department of Chemistry, Guru Nanak Dev University, Amritsar, India
f Department of economics, Mangalayata University, India
g TIDCL, Ajou University, South Korea
h Department for Structural Infection Biology, Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research, Notkestraße 85, 22607 Hamburg, Germany.
i Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Rothenbaumchaussee 19, 20148 Hamburg, Germany.

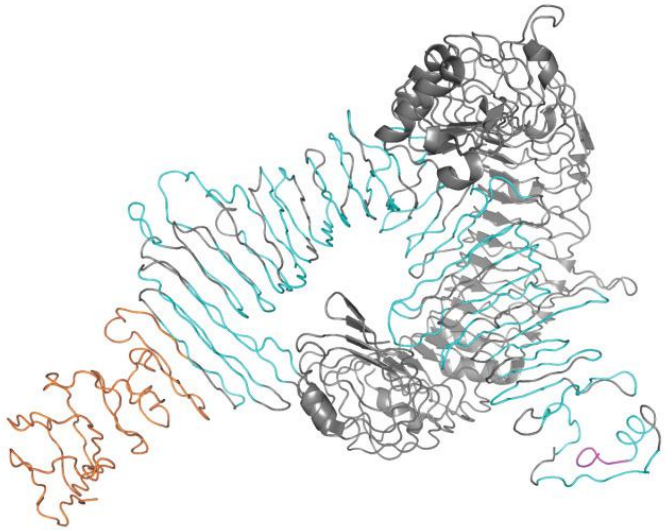
*Correspondence should be addressed to S.S. (sukrit.srivastav@mangalayatan.edu.in; srivastav.sukrit@gmail.com)

Abstract

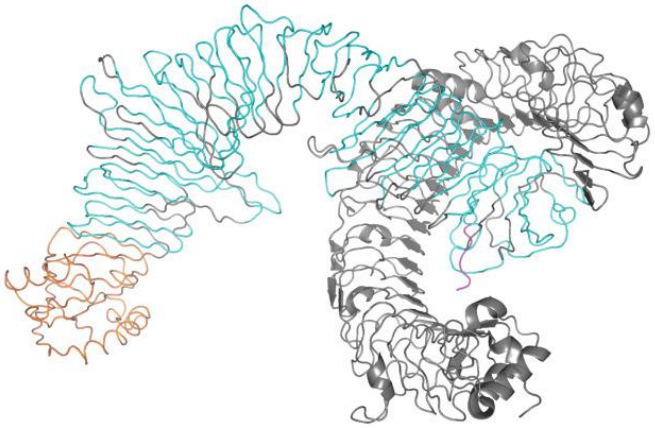
The 2019 novel coronavirus (COVID19 / Wuhan coronavirus), officially named as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), is a positive-sense single-stranded RNA coronavirus. SARS-CoV-2 causes the contagious COVID19 disease also known as 2019-nCoV acute respiratory disease and has led to the ongoing 2019–20 pandemic COVID19 outbreak. The effective

countermeasures against SARS-CoV-2 infection requires the design and development of specific and effective vaccine candidate. In the present study we have screened 38 CTL, 42 HTL and 12 B cell epitopes by utilizing different in silico tools. The screened epitopes were further validated for their binding with their respective HLA allele binders and TAP (Transporter associated with antigen processing) molecule by molecular docking. The screened epitopes were further utilized to design novel two multi-epitope vaccines (MEVs) composed of CTL, HTL and B cell epitopes overlaps with potential to elicit humoral as well as cellular immune response against SARS-CoV-2. To enhance the immune response for our vaccine design, truncated (residues 10-153) Onchocerca volvulus activation-associated secreted protein-1 (Ov-ASP-1) has been utilized as an adjuvant at N terminal of both the MEVs. Further molecular models for both the MEVs were prepared and validated for their stable molecular interaction with Toll-Like Receptor 3 (TLR 3). The codon-optimized cDNA of both the MEVs were further analyzed for their potential of high level of expression in Human cell line. The present study proposes *in-silico* validated design of two MEVs against SARS-CoV-2 to elicit humoral as well as cellular immune response.

Key words: COVID19, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), Coronavirus, Human Transporter associated with antigen processing (TAP), Toll-Like Receptor (TLR), Epitope, Immunoinformatics, Molecular Docking, Multi-epitope Vaccine



CTL Multi-epitope COVID19 vaccine - Toll Like Receptor 3 complex



CTL Multi-epitope COVID19 vaccine - Toll Like Receptor 3 complex

Graphical abstract: The designed CTL (Cytotoxic T lymphocyte) and HTL (Helper T lymphocyte) multi-epitope vaccine (MEV) against COVID19 infection. Both the CTL and HTL MEVs models show a very stable and very fitting conformational complex formation tendency with the Toll like receptor 3. CTL and HTL MEVs: ribbon; Toll like receptor 3: gray cartoon; Adjuvant [truncated (residues 10-153) Onchocerca volvulus activation-associated secreted protein-1]: orange ribbon; Epitopes: cyan ribbons regions; 6xHis Tag: magenta ribbon region

58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80

Abbreviations: Antigen-presenting cell (APCs), Codon Adaptation Index (CAI), Coronavirus ID 19 (COVID19), Coronavirus (CoV), Coverage (Cov), Cryo-Electron Microscopy (Cryo-EM), Cytotoxic T lymphocyte (CTL), ectodomain (ECD), Envelope Protein (E Protein), endoplasmic reticulum (ER), European Bioinformatics Institute (EBI), Nucleocapsid Protein (N Protein), Ope reading Frame (ORF), global distance test (GDT), Grand average of hydropathicity (GRAVY), Surface protein (S Protein), half maximal inhibitory concentration (IC50), Helper T lymphocyte (HTL), Human Leukocyte Antigen (HLA), Immune Epitope Database (IEDB), Immuno Polymorphism Database (IPD), Interferon Gama (IFN-γ), International ImMunoGeneTics project (IMGT), major histocompatibility complex (MHC), Membrane Protein (M Protein), Motif-EmeRging with ClassesIdentification (MERCI), Multi-epitope Vaccine (MEV), Multiple Sequence Alignment (MSA), non-structural protein (nsp), National Center for Biotechnology Information (NCBI), peptide-MHC (pMHC), Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), Stabilization Matrix alignment method (SMM), Transmembrane (TM), Protein Data Bank (PDB), Qualitative Model Energy ANalysis (QMEAN), root mean square deviation (RMSD), Support Vector Machine (SVM), Toll-Like Receptor 3 (TLR3), Transporter associated with antigen processing (TAP), un-normalized global distance test (uGDT)

Introduction

The novel coronavirus (COVID19), officially named as Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused the ongoing outbreak of

severe form of flu leading to death with mortality rate of 3.4 %. The SARS-CoV-2 is a novel coronavirus associated with a respiratory disease initiated from the Wuhan of Hubei province, China. The disease is highly contagious and has spread to 182 countries/ Territories since its outbreak in China in December 2019 till 21st of march 2020. Worldwide, as of 21st March 2020 the total confirmed cases has been reported to be 2,66,073 and total death count reported is 11,184 (WHO Situation report 21st March 2020). Over all the SARS-CoV-2 infection has put a global emergency condition. The economic impact of COVID-19 is even more harsh and has put the world on economic risk. As of 9th March, the downside scenario sees a $2 trillion shortfall in global income with a $220 billion hit to developing countries. The COVID-19 shock will cause a recession in several countries and depress global annual growth this year to 2.5 per cent lower, the recessionary threshold for the world economy (UNCTAD report 9th March 2020).

The infection mechanism and pathogenesis of SARS-CoV-2 is largely unknown yet. The proteome of SARS-CoV-2 is composed of 11 structural and non-structural proteins. These includes polyprotein (ORF1ab), Surface protein (S Protein), ORF3, Envelope Protein (E Protein), Membrane Protein (M Protein), ORF6, ORF7a, ORF7b, ORF8, Nucleocapsid Protein (N Protein), and ORF10 (NCBI protein sequence database). The actual function and pathogenic or proliferative role of these SARS-CoV-2 coronavirus proteins are largely not known yet.

126 The SARS-CoV-2 coronavirus polyprotein (ORF1ab) with length of 7,096
127 amino acid (aa) is composed of 16 different expressed protein viz. leader protein
128 (nsp1, location: 1-180 aa); nsp2 (location: 181-818); nsp3 (former nsp1, carry
129 conserved domains - N-terminal acidic, predicted phosphoesterase, papain-like
130 proteinase, Y-domain, transmembrane domain 1 (TM1) and adenosine
131 diphosphate-ribose 1''-phosphatase, location: 819-2,763); nsp4 (contains
132 transmembrane domain 2 (TM2), location: 2,764-3,263); 3C-like proteinase (nsp5,
133 main proteinase, mediates cleavages downstream of nsp4, location: 3,264-
134 3,569); nsp6 (putative transmembrane domain, location: 3,570-3,859); nsp7
135 (location: 3,860-3,942); nsp8 (location: 3,943-4,140); nsp9 (ssRNA-binding
136 protein, location: 4,141-4,253); nsp10 (formerly known as growth-factor-like
137 protein, location: 4,254-4,392); nsp11 (location: 4,393-4,405); RNA-dependent
138 RNA polymerase (nsp12, location: 4,393-5,324); Helicase (nsp13; zinc-
139 binding domain, NTPase/helicase domain, RNA 5'-triphosphatase, location:
140 5,325-5,925); 3'-to-5' exonuclease (nsp14, location: 5,926-6,452); endo RNAse
141 (nsp15, location: 6,453-6,798); and 2'-O-ribose methyltransferase (nsp16;
142 location: 6,799-7,096).
143 The SARS-CoV-2 coronavirus surface glycoprotein (S Protein) is a
144 structural protein and acts as spike protein, its location is 21,563-25,384 aa, and
145 length is 1273 aa); The SARS-CoV-2 coronavirus ORF3a protein has location
146 25,393-26,220 aa and length 275 aa. The SARS-CoV-2 coronavirus envelope
147 protein (E Protein) (ORF4) is a structural protein and has location 26,245-26,472
148 aa and length 75 aa. The SARS-CoV-2 coronavirus membrane glycoprotein (M

149 Protein) (ORF5) is a structural protein and has location 26,523-27,191aa and
150 length 222 aa. The SARS-CoV-2 coronavirus ORF6 protein has location 27,202-
151 27,387 aa and length 61 aa. The SARS-CoV-2 coronavirus ORF7a protein has
152 location 27,394-27,759 aa and length 121 aa. The SARS-CoV-2 coronavirus
153 ORF7b protein has location 27,756-27,887 aa and length 43 aa. The SARS-CoV-
154 2 coronavirus ORF8 protein has location 27,894-28,259 aa and length 121 aa.
155 The SARS-CoV-2 coronavirus nucleocapsid phosphoprotein (N Protein) (ORF9)
156 is a structural protein with location 28,274-29,533 aa and has length of 419 aa.
157 The SARS-CoV-2 coronavirus ORF10 protein has location 29,558-29,674 aa and
158 length 38 aa (NCBI protein sequence database).

159 Although the exact mechanism and roles of all the above-mentioned
160 proteins of SARS-CoV-2 coronavirus proteome is not well known yet, but these
161 proteins could act as potential vaccine candidate against the SARS-CoV-2
162 coronavirus infection. In the present study we have screened highly potential
163 epitopes from all the above-mentioned protein and further we have also designed
164 and proposed CTL (Cytotoxic T lymphocyte) and HTL (Helper T lymphocyte)
165 multi-epitope based vaccine candidates against the SARS-CoV-2 coronavirus
166 infection.

167

168

169

170

171

172 **Methodology**

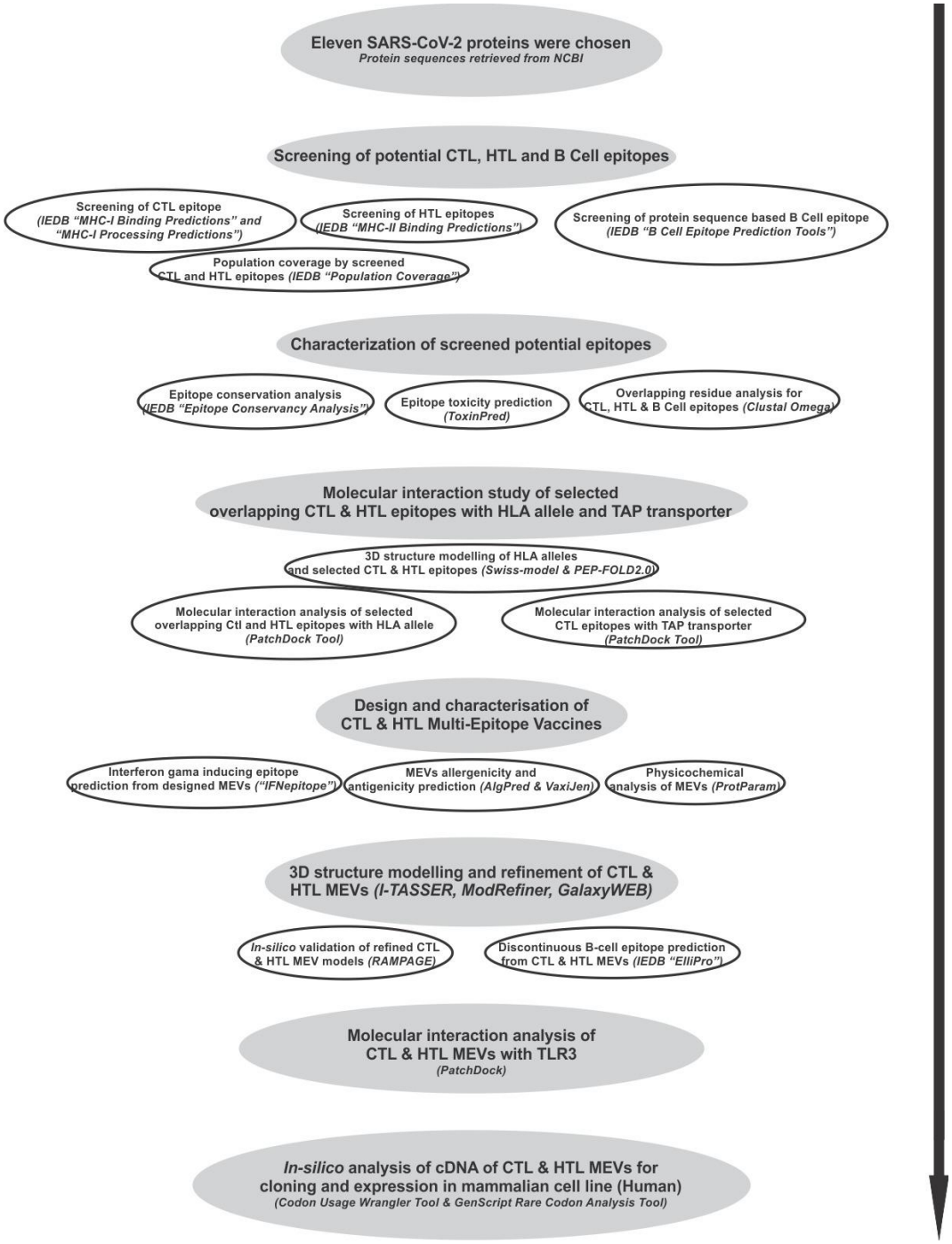
173 In the present study on SARS-CoV-2 coronavirus we have screened
174 potential epitopes and have designed and proposed two multi-epitope vaccines
175 (MEVs) composed of screened CTL (Cytotoxic T lymphocyte) and HTL (Helper T
176 lymphocyte) epitopes with overlapping regions of B cell epitopes. Hence the
177 proposed MEVs are supposed to have potential to elicit both the humoral as
178 well as cellular immune response. To enhance immune response, truncated
179 (residues 10-153) Onchocerca volvulus activation-associated secreted protein-1
180 (Ov-ASP-1) has been utilized as adjuvant at N-terminal of both the MEVs. The
181 truncated Ov-ASP-1 was chosen due to its potential to activate antigen-
182 processing cells (APCs) (MacDonald et al., 2005; Guo et al., 2015; He et al.,
183 2009). All SARS-CoV-2 coronavirus proteins mentioned in the introduction, were
184 utilized to screen the potential CTL, HTL and B cell epitopes. The screened
185 epitopes were further studied for overlapping consensus regions amongst them.
186 The epitopes showing overlapping regions in partial or complete were chosen for
187 detailed further studies.

188 The chosen CTL and HTL epitopes were analyzed for their molecular
189 interaction with their respective HLA allele binders. Moreover, the chosen CTL
190 epitopes were also analyzed for their molecular interaction with TAP (Transporter
191 associated with antigen processing) transporter cavity to observe their smooth
192 passage from cytoplasm to endoplasmic reticulum lumen (Oldham et al., 2016;
193 Abele et al., 2004). The tertiary model for both MEVs were generated and refined.

Both the MEVs models were further utilized to screened B Cell linear and discontinuous epitopes as well as IFN-γ inducing epitopes.

The molecular signaling by multiple TLRs, is an essential component of the innate immune response against SARS-CoV-2 coronavirus. Since the rOv-ASP-1 primarily binds APCs among human PBMCs and trigger pro-inflammatory cytokine production via Toll-like receptor 3 (TLR3), hence both the CTL and HTL MEV models were further analyzed for their molecular interaction with the TLR-3 by molecular docking studies (Antoniou et al., 2003; Delneste et al., 2007; Totura et al., 2015; Farina et al., 2005). Further the codon optimized cDNA of both the MEVs were analyzed to have high level of expression in mammalian cell line (human), which would facilitate *in-vivo* expression, experimentation and trials (Fig.1).

Figure 1. Workflow chart.



240 **Screening of Potential Epitopes**

241 ***T cell Epitope Prediction***

242 *Screening of Cytotoxic T lymphocyte (CTL) Epitope.* The screening of Cytotoxic
243 T lymphocyte epitopes was performed by the IEDB (Immune Epitope Database)
244 tools "MHC-I Binding Predictions" (http://tools.iedb.org/mhci/) and "MHC-I
245 Processing Predictions" (http://tools.iedb.org/processing/) (Tenzer et al., 2005;
246 Peters et al., 2003; Hoof et al., 2009). Both the tools use six different methods
247 (viz. Consensus, NN-align, SMM-align, Combinatorial library, Sturniolo and
248 NetMHCIIpan) and generate "Percentile rank" and a "total score" respectively.

249 The screening is based on the total amount of cleavage site in the protein.
250 TAP score estimates an effective log -(IC50) values (half maximal inhibitory
251 concentration (IC50)) for binding to TAP of a peptide or its N-terminal prolonged
252 precursors. The MHC binding prediction score is -log(IC50) values for binding to
253 MHC of a peptide (Calis et al., 2013). The IC(50) (nM) value for each epitope and
254 MHC allele binding pairs were also obtained by this IEDB tool. Epitopes having
255 high, intermediate, and least affinity of binding to their HLA allele binders have
256 IC50 values < 50 nM, < 500 nM and < 5000 nM respectively.

257 Immunogenicity of all the screened CTL epitopes was also obtained by
258 using "MHC I Immunogenicity" tool of IEDB (http://tools.iedb.org/immunogenicity/)
259 with all the parameters set to default analyzing 1st, 2nd, and C-terminus amino
260 acids of the given screened epitope (Calis et al., 2013). The tool predicts the
261 immunogenicity of a given peptide-MHC (pMHC) complex on the basis of the

262 physiochemical properties of constituting amino acid and their position within the
263 peptide sequence.
264
265 *Screening of Helper T lymphocyte (HTL) Epitopes.* To screen out the Helper T
266 lymphocyte epitopes from SARS-CoV-2 proteins, the IEDB tool "MHC-II Binding
267 Predictions" (http://tools.iedb.org/mhcii/) was used. The tool generates
268 "Percentile rank" for each potential peptide. The lower the value of percentile,
269 higher would be the affinity. This percentile rank is generated by the combination
270 of three different methods viz. combinatorial library, SMM_align & Sturniolo; and
271 by comparing the score of the peptide against the scores of other random five
272 million 15-mer peptides of SWISSPROT database (Wang et al., 2010; Sidney et
273 al., 2008; Nielsen et al., 2007; Sturniolo et al., 1999). The rank from the
274 consensus of all three methods was generated by the median percentile rank of
275 the three methods.
276
277 *Population Coverage by CTL and HTL epitopes.* The "Population Coverage" tool
278 of IEDB (http://tools.iedb.org/population/) was used to elucidate the world human
279 population coverage by the shortlisted 38 CTL and 42 HTL epitopes derived from
280 nine SARS-CoV-2 proteins (Bui et al., 2006). The T cells recognize complex
281 between a specific major MHC molecule and a particular pathogen-derived
282 epitope. The given epitope will elicit a response only in an individual that express
283 an MHC molecule, which is capable of binding that particular epitope. This
284 denominated MHC restriction of T cell responses and the MHC polymorphism

285 provide basis for population coverage study. The MHC types are expressed at
286 dramatically different frequencies in different ethnicities. Hence a vaccine with
287 larger population coverage could be of greater importance (Sturniolo et al., 1999).
288 Clinical administration of multiple-epitopes involving both the CTL and the HTL
289 epitopes are predicted here to have a greater probability of larger human
290 population coverage worldwide.
291
292 ***B Cell Epitope Prediction***
293 *Sequence-based B Cell epitope prediction.* Protein sequence-based method
294 "Bepipred Linear Epitope Prediction" was utilized to screen linear B cell epitopes
295 from eleven different SARS-CoV-2 proteins. The tool "B Cell Epitope Prediction
296 Tools" of IEDB server (http://tools.iedb.org/bcell/) was utilized. In this screening
297 the parameters such as hydrophilicity, flexibility, accessibility, turns, exposed
298 surface, polarity and antigenic propensity of the polypeptides is correlated with its
299 location in the protein. This allows the search for continuous epitopes prediction
300 from protein sequence. The prediction is base on the propensity scales for each
301 of the 20 amino acids. For a window size n, the i - (n-1)/2 neighboring residues
302 on each side of residue i are used to compute the score for the residue i. The
303 method "Bepipred Linear Epitope Prediction" utilized here is based on the
304 propensity scale method as well as the physiochemical properties of the given
305 antigenic sequence to screen potential epitopes (Larsen et al., 2006).
306
307

308 **Characterisation of potential epitopes**

309 *Epitope conservation analysis.* The shortlisted CTL, HTL and B cell epitopes
310 screened from eleven SARS-CoV-2 proteins were analysed for the conservancy
311 of their amino acid sequence by "Epitope Conservancy Analysis" tool
312 (http://tools.iedb.org/conservancy/) of IEDB. The epitope conservancy is the
313 number of protein sequences (retrieved from NCBI) that contain that particular
314 epitope. The analysis was done against their entire respective source protein
315 sequences of SARS-CoV-2 proteins retrieved from the NCBI protein database
316 (Bui et al., 2007).

317

318 *Epitope Toxicity prediction.* The tool ToxinPred
319 (http://crdd.osdd.net/raghava/toxinpred/multi_submit.php) was used to analyse
320 the toxicity of shortlisted CTL, HTL and B cell epitopes. The tool allows to
321 identifying highly toxic or non-toxic short peptides. The toxicity check analysis
322 was done by the "SVM (Swiss-Prot) based" (support vector machine) method
323 utilizing dataset of 1805 sequences as positive, 3593 negative sequences from
324 Swissprot as well as an alternative dataset comprises the same 1805 positive
325 sequences and 12541 negative sequences from TrEMBLE (Gupta et al., 2013).

326

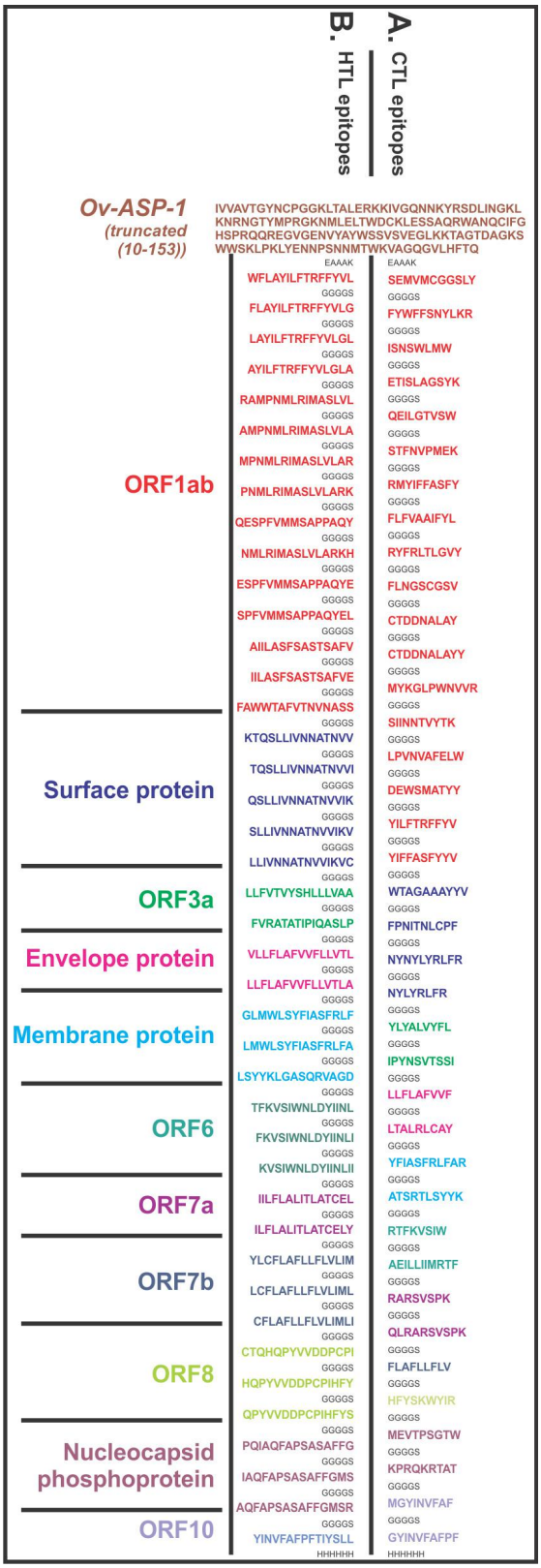
327 *Overlapping residue analysis.* The overlapping residue analysis for the shortlisted
328 38 CTL, 42 HTL and the 12 B cell linear epitopes was performed by the Multiple
329 Sequence Alignment (MSA) analysis by Clustal Omega tool
330 (https://www.ebi.ac.uk/Tools/msa/clustalo/) of EBI (European Bioinformatics

331 Institute) (Sievers et al., 2011). The Clustal Omega multiple sequence alignment
332 tool virtually aligns any number of protein sequences and delivers an accurate
333 alignment.
334
335 *Epitope selected for molecular interaction study with HLA allele and TAP*
336 *transporter.* On the basis of the overlapping residue analysis of shortlisted CTL,
337 HTL and linear B cell epitopes few numbers of CTL and HTL epitopes were
338 chosen for further analysis. The chosen epitopes are encircled shown in Fig.2.
339 These epitopes were chosen on the basis of them having partial or full
340 overlapping sequence region amongst all three types of epitopes (CTL, HTL and
341 B Cell). The chosen epitopes were further analyzed for their interaction with their
342 respective HLA allele binders and TAP cavity interaction.
343
344 **Molecular interaction analysis of selected epitopes with HLA allele and TAP**
345 **transporter.**
346 *Tertiary structure modelling of HLA alleles and selected T cell epitopes.* The
347 Swiss-model was used for homology modelling of the HLA class I and II allele
348 binders of the chosen epitopes (Arnold et al., 2006). The amino acid sequence of
349 the HLA allele binders were retrieved from Immuno Polymorphism Database
350 (IPD-IMGT/HLA) (https://www.ebi.ac.uk/ipd/imgt/hla/allele.html). Templates for
351 homology modelling were chosen on the basis of highest amino acid sequence
352 similarity. All the generated HLA allele models had acceptable QMEAN value
353 (cutoff -4.0) (Supplementary table S1). The QMEAN value gives a composite

354 quality estimate involving both global as well as local analysis of the model
355 (Benkert et al., 2008).
356 The PEP-FOLD 2.0 a de novo structure prediction tool at RPBS Web
357 Portal (https://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-
358 FOLD) was utilized to generate tertiary structures for the chosen CTL and HTL
359 epitopes (Shen et al., 2014).
360
361 *Molecular interaction analysis of chosen CTL and HTL epitopes with HLA alleles.*
362 The PatchDock tool (http://bioinfo3d.cs.tau.ac.il/PatchDock/) was utilized for *in-*
363 *silico* molecular docking study of the selected CTL and HTL epitopes with their
364 respective HLA class I and II allele binders (Bell et al., 2005; Duhovny et al.,
365 2002; Schneidman-Duhovny et al., 2005). PatchDock utilizes algorithm for
366 unbound (real life) docking of molecules for protein-protein complex formation.
367 The algorithm carries out the rigid docking, with the surface variability/flexibility
368 implicitly addressed through liberal intermolecular penetration. The algorithm
369 focuses on the (i) initial molecular surface fitting on localized, curvature based
370 surface patches (ii) use of Geometric Hashing and Pose Clustering for initial
371 transformation detection (iii) computation of shape complementarity utilizing the
372 Distance Transform (iv) efficient steric clash detection and geometric fit scoring
373 based on a multi-resolution shape representation and (v) utilization of biological
374 information by focusing on hot spot rich surface patches (Bell et al., 2005;
375 Duhovny et al., 2002; Schneidman-Duhovny et al., 2005).
376

377 *Molecular interaction analysis of selected CTL epitopes with TAP transporter.*
378 TAP transporter plays an important role in the presentation of CTL epitope. From
379 the cytosol after proteasome processing, the fragmented peptide of foreign
380 protein gets transported to endoplasmic reticulum (ER) through the TAP
381 transporter. From the ER these short peptides reach to the golgi bodies and then
382 get presented on the cell surface (Abele et al., 2004). Molecular interaction study
383 of the chosen CTL epitopes within the TAP transporter cavity was performed by
384 molecular docking study utilizing the PatchDock tool. For accurate prediction, the
385 cryo-EM structure of TAP transporter (PDB ID: 5u1d) was used by removing the
386 antigen from TAP cavity of the original structure (Oldham et al., 2016).
387
388 **Design, characterisation and molecular interaction analysis of Multi-**
389 **Epitope Vaccines with immunological receptor.**
390 *Design of Multi-Epitope Vaccines.* The screened and shortlisted high scoring 38
391 CTL and 42 HTL epitopes were utilized to design CTL and HTL Multi-Epitope
392 vaccines (Table 1 and 2). Short peptide EAAAK and GGGGS were used as rigid
393 and flexible linkers respectively (Fig.2). The GGGGS linker provides proper
394 conformational flexibility to the vaccine tertiary structure and hence facilitates
395 stable conformation to the vaccine. The EAAAK linker facilitates in domain
396 formation and hence facilitates the vaccine to obtain its final stable structure.
397 Truncated (residues 10-153) Onchocerca volvulus activation-associated secreted
398 protein-1 (Ov-ASP-1) has been utilized as an adjuvant at N terminal of both the
399 CTL and HTL MEVs (MacDonald et al., 2005; Guo et al., 2015; He et al., 2009;

Hu et al., 2004; Hajighahramani et al., 2017; Chen et al., 2013; Srivastava et al., 2019; Srivastava et al., 2018).



422 **Figure 2. Design of SARS-CoV-2 Multi-Epitope Vaccine (MEVs). (A)** CTL and
423 **(B)** HTL epitopes were linked by the short peptide linker 'GGGGS'. Truncated
424 (residues 10-153) Onchocerca volvulus activation-associated secreted protein-1
425 (Ov-ASP-1) has been utilized at the N terminal of both the MEVs. The short
426 peptide EAAAK was used to link Ov-ASP-1 at N terminal. Epitopes from different
427 proteins were colored in different colors. C terminal 6xHis is designed as His tag.
428

429 ***Characterisation of designed Multi-Epitope Vaccines***

430 *Physicochemical property analysis of designed MEVs.* The ProtParam
431 (https://web.expasy.org/protparam/) tool was utilized to analyse the
432 physiochemical properties of the amino acid sequence of the designed CTL and
433 HTL MEVs (Gasteiger et al., 2005). The ProtParam analysis performs an
434 empirical investigation for the given query amino acid sequence. ProtParam
435 computes various physicochemical properties derived from a given protein
436 sequence.
437

438 *Interferon gama inducing epitope prediction.* From the designed amino acid
439 sequence of both the MEVs potential interferon gama (IFN-γ) epitopes were
440 screened by "IFNepitope" server
441 (http://crdd.osdd.net/raghava/ifnepitope/scan.php) using "Motif and SVM hybrid",
442 (MERCI: Motif-EmeRging and with Classes-Identification, and SVM: support
443 vector machine) method. The tool predicts peptides from protein sequences
444 having the capacity to induce IFN-gamma release from CD4+ T cells. This

445 module generates overlapping peptides from query sequence and predicts IFN-
446 gamma inducing peptides. For the screening, IEDB database with 3705 IFN-
447 gamma inducing and 6728 non-inducing MHC class II binders is utilized (Nagpal
448 et al., 2015; Dhanda et al., 2013).
449
450 *MEVs allergenicity and antigenicity prediction.* Both the designed MEVs were
451 further analysed for allergenicity and antigenicity prediction by utilizing the
452 AlgPred (http://crdd.osdd.net/raghava/algpred/submission.html) and the Vaxigen
453 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html) tools respectively
454 (Saha et al., 2006; Doytchinova and Flower 2007). The AlgPred prediction is
455 based on the similarity of already known epitope with any region of the submitted
456 protein. For the screening of allergenicity, the Swiss-prot dataset consisting of
457 101725 non-allergens and 323 allergens is utilized. The VaxiJen utilizes an
458 alignment-free approach, solely based on the physicochemical properties of the
459 query amino acid sequence. For prediction of antigenicity, the Bacterial, viral and
460 the tumour protein datasets are used to derive models for the prediction of whole
461 protein antigenicity. Every set consisted of known 100 antigens and 100 non-
462 antigens.
463
464 *Tertiary structure modelling and refinement of MEVs.* The tertiary structure of
465 both the designed CTL and HTL MEVs were generated by homology modelling
466 utilizing the I-TASSER modeling tool (https://zhanglab.ccmb.med.umich.edu/I-
467 TASSER/). The I-TASSER is a tool for protein structure prediction based on the

468 sequence-to-structure-to-function paradigm (Roy et al., 2010). The tool
469 generates three-dimensional (3D) atomic models from multiple threading
470 alignments and iterative structural assembly simulations for a submitted amino
471 acid sequence. I-TASSER works on the basis of the structure templates
472 identified by LOMETS, a meta-server, from the PDB library. I-TASSER only uses
473 the templates of the highest Z-score which is the difference between the raw and
474 average scores in the unit of standard deviation. For each target model the I-
475 TASSER simulations generate a large ensemble of structural conformations,
476 called decoys. To select the final models, I-TASSER uses the SPICKER program
477 to cluster all the decoys based on the pair-wise structure similarity, and reports
478 up to five models. The Normalized Z-score >1 mean a good alignment and vice
479 versa. The Cov represents the coverage of the threading alignment and is equal
480 to the number of aligned residues divided by the length of query protein. Ranking
481 of templet proteins is based on TM-score of the structural alignment between the
482 query structure model and known structures. The RMSD is the RMSD between
483 templet residues and query residues that are structurally aligned by TM-align.

484 The refinement of both the generated MEV models was performed by
485 ModRefiner (https://zhanglab.ccmb.med.umich.edu/ModRefiner/) and
486 GalaxyRefine tool (http://galaxy.seoklab.org/cgi-bin/submit.cgi?type=REFINE)
487 (Dong et al., 2011). TM-score generated by ModRefiner indicates the structural
488 similarity of the refined model with the original input model. Closer the TM-Score
489 to 1, higher would be the similarity of original and the refined model. RMSD value

490 of the refined model shows the conformational deviation from the initial input
491 models.

492 The GalaxyRefine tool refines the query tertiary structure by repeated
493 structure perturbation as well as by utilizing the subsequent structural relaxation
494 by the molecular dynamics simulation. The tool GalaxyRefine generates reliable
495 core structures from multiple templates and then re-builds unreliable loops or
496 termini by using an optimization-based refinement method (Ko et al., 2012; Wang
497 et al., 2013; Shin et al., 2014). To avoid any breaks in the 3D model
498 GalaxyRefine uses triaxial loop closure method. The MolProbity score generated
499 for a given refined model indicates the log-weighted combination of the clash
500 score, the percentage of Ramachandran not favoured residues and the
501 percentage of bad side-chain rotamers.

502

503 *Validation of CTL and HTL MEVs refined models.* Both the refined CTL and HTL
504 MEV 3D models were further validated by RAMPAGE analysis tool
505 (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) (S.C. Lovell et al., 2008;
506 Ramakrishnan, C. et al., 1965). The generated Ramachandran plots for the MEV
507 models show the sterically allowed and disallowed residues along with their
508 dihedral psi (ψ) and phi (φ) angles.

509

510 *Linear and Discontinuous B-cell epitope prediction from MEVs.* The Ellipro
511 (ElliPro: Antibody Epitope Prediction tool; http://tools.iedb.org/ellipro/) method
512 available at IEDB, was used to screen the linear and the discontinuous B cell

513 epitopes from the MEVs vaccine models. The ElliPro method analyses on the
514 basis of the location of residue in the protein's 3D structure. The residues lying
515 outside of an ellipsoid covering 90% of the inner core protein residues score
516 highest Protrusion Index (PI) of 0.9; and so on. The discontinuous epitopes
517 predicted by the ElliPro tool are clustered on the basis of the distance "R" in Å
518 between two residue's centers of mass lying outside of the largest possible
519 ellipsoid. The larger value of R indicates larger distant residues (residue
520 discontinuity) are screened in the epitopes (Kringelum et al., 2012; Ponomarenko
521 et al., 2008).
522
523 ***Molecular interaction analysis of MEVs with immunological receptor.***
524 *Molecular docking and dynamics simulation study of MEVs and TLR-3.* Molecular
525 interaction analysis of both the designed MEVs with Toll-Like receptor-3 (TLR-3),
526 was performed by molecular docking and molecular dynamics simulation.
527 Molecular docking was performed by PatchDock server
528 (http://bioinfo3d.cs.tau.ac.il/PatchDock/) (Schneidman-Duhovny et al., 2005).
529 PatchDock utilizes algorithm for unbound (real life) docking of molecules for
530 protein-protein complex formation as explained earlier (Bell et al., 2005; Duhovny
531 et al., 2002). For molecular docking, the 3D structure of human TLR-3
532 ectodomain (ECD) was retrieved from PDB databank (PDB ID: 2A0Z). The study
533 provides dynamical properties of the designed system with MEVs-TLR3
534 complexes with a guess at the interactions between the molecules, and also it

535 gives ‘exact’ predictions of bulk properties including the hydrogen bond formation
536 and conformation of the molecules forming the complex.
537
538 ***In-silico* analysis of MEVs for cloning and expression potency**
539 *Analysis of cDNA of both the MEVs for cloning and expression in mammalian*
540 *host cell line.* Complementary DNA of both the MEVs, codon optimized for
541 expression in Mammalian cell line (Human) were generated by Java Codon
542 Adaptation Tool (http://www.jcat.de/). The generated cDNA of both the MEVs
543 were further analysed by GenScript Rare Codon Analysis Tool
544 (https://www.genscript.com/tools/rare-codon-analysis). The tool analyses the GC
545 content, Codon Adaptation Index (CAI) and the Tandem rare codon frequency for
546 a given cDNA (Morla et al., 2016; Wu et al., 2010). The CAI indicates the
547 possibility of cDNA expression in a chosen expression system. The tandem rare
548 codon frequency indicates the presence of low-frequency codons in the given
549 cDNA.
550
551
552 **RESULTS & DISCUSSION**
553 **Screening of potential epitopes**
554 ***T cell Epitope Prediction***
555 *Screening of Cytotoxic T lymphocyte (CTL) Epitope.* Cytotoxic T lymphocyte
556 (CTL) epitopes were screened by “MHC-I Binding Predictions” and “MHC-I
557 Binding Predictions” IEDB tools. These epitopes are shortlisted on the basis of

558 total amount of cleavage site in the protein, low IC(50) (nM) value for epitope-
559 HLA class I allele pairs, and for binding to TAP cavity.

560 The 38 epitopes predicted by "MHC-I Binding Predictions" tool with
561 highest "Percentile Rank" were shortlisted for multi-epitope vaccine design and
562 are listed in table 1. Rest 101 epitopes are listed in Supplementary table S8. The
563 67 epitopes predicted by "MHC-I Processing Predictions" tool with highest "Total
564 score" are listed in Supplementary table 9.

565 The immunogenicity of the shortlisted CTL epitopes was also determined
566 and are mentioned in table 1, Supplementary table S8, and S9. The higher
567 immunogenicity score indicates the greater immunogenic potential of the given
568 epitope.

569

570 **Table 1. Shortlisted high percentile ranking SARS-CoV-2 CTL epitopes.** Selected
571 high percentile CTL epitopes and their respective HLA alleles binders are listed. *In-silico*
572 analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they
573 show significant conservancy and high immunogenicity. Six epitope shown with ** show
574 exact match with the epitopes published by Grifoni et al., 2020, indicate consensus in
575 results.
576
577

SARS-CoV - 19 Proteins	Peptide	Conservancy	Immuno genicity	Toxicity	Allele	Position	Length	Method used	Percentile Rank
E Protein	LLFLAFVVF	100.00% (35/35)	0.2341	Non-Toxin	B*15:01	18-26	9	Consensus (ann/comblib_sidney2008/smm)	0.1
E Protein	LTALRLCAY	97.14% (34/35)	0.01886	Non-Toxin	A*01:01	34-42	9	Consensus (ann/smm)	0.12
M Protein	ATSRTLSYYK**	100.00% (41/41)	-0.13563	Non-Toxin	A*11:01	31-40	10		0.06
M Protein	YFIASFRLFAR	100.00% (41/41)	0.19709	Non-Toxin	A*33:01	25-35	11	ann	0.03
N protein	KPRQKRTAT	100.00% (40/40)	-0.20542	Non-Toxin	B*07:02	47-55	9	Consensus (ann/comblib_sidney2008/smm)	0.1
N protein	MEVTPSGTW	100.00% (40/40)	-0.06279	Non-Toxin	B*44:02	42-50	9	Consensus (ann/smm)	0.06
ORF-1ab	GYINVFAFPF**	100.00% (28/28)	0.32004	Non-Toxin	A*23:01	2-11	10		0.11
ORF-1ab	MGYINVFAF	100.00% (28/28)	0.28694	Non-Toxin	B*35:01	1-9	9	Consensus (ann/comblib_sidney2008/smm)	0.1
ORF-1ab	CTDDNALAY	100.00% (36/36)	0.07355	Non-Toxin	A*01:01	33-41	9	Consensus (ann/smm)	0.06
ORF-1ab	CTDDNALAYY**	100.00% (36/36)	0.08174	Non-Toxin	A*01:01	33-42	10		0.06

ORF-1ab	DEWSMATYY**	97.22% (35/36)	-0.19814	Non-Toxin	B*44:03	58-66	9		0.07
ORF-1ab	ETISLAGSYK	100.00% (36/36)	-0.20585	Non-Toxin	A*68:01	33-42	10		0.06
ORF-1ab	FLFVAAIFYL	100.00% (36/36)	0.37766	Non-Toxin	A*02:01	50-59	10		0.06
ORF-1ab	FLNGSCGSV	100.00% (36/36)	-0.24791	Non-Toxin	A*02:03	43-51	9		0.06
ORF-1ab	FYWFFSNYLKR	100.00% (36/36)	-0.00254	Non-Toxin	A*33:01	3-13	11	ann	0.04
ORF-1ab	ISNSWLMW	100.00% (36/36)	-0.11151	Non-Toxin	B*58:01	49-56	8		0.05
ORF-1ab	LPVNVAFELW	97.22% (35/36)	0.27341	Non-Toxin	B*53:01	61-70	10	Consensus (ann/smm)	0.06
ORF-1ab	MYKGLPWNVVR	94.44% (34/36)	0.21107	Non-Toxin	A*33:01	58-68	11	ann	0.06
ORF-1ab	QEILGTVSW	100.00% (36/36)	0.03976	Non-Toxin	B*44:02	35-43	9	Consensus (ann/smm)	0.06
ORF-1ab	RMYIFFASFY	100.00% (36/36)	0.32633	Non-Toxin	A*30:02	2-11	10		0.06
ORF-1ab	RYFRLTLGVY	100.00% (36/36)	0.15936	Non-Toxin	A*30:02	22-31	10		0.06
ORF-1ab	SEMVMCGGSLY	97.22% (35/36)	-0.32016	Non-Toxin	B*44:02	16-26	11	ann	0.03
ORF-1ab	SIINNTVYTK**	97.22% (35/36)	0.12661	Non-Toxin	A*11:01	37-46	10	Consensus (ann/smm)	0.06
ORF-1ab	STFNVPMEK	100.00% (36/36)	-0.02845	Non-Toxin	A*11:01	10-18	9		0.06
ORF-1ab	YIFFASFYYV	100.00% (36/36)	0.13772	Non-Toxin	A*02:06	4-13	10		0.07
ORF-1ab	YILFTRFFYV	100.00% (36/36)	0.40924	Non-Toxin	A*02:06	21-30	10		0.07
ORF3a	IPYNSVTSSI	100.00% (28/28)	-0.32835	Non-Toxin	B*51:01	18-27	10		0.11
ORF3a	YLYALVYFL**	100.00% (28/28)	0.13151	Non-Toxin	A*02:01	37-45	9	Consensus (ann/comblib_sidney2008/smm)	0.1
ORF6	AEILLIIMRTF	100.00% (34/34)	0.1815	Non-Toxin	B*44:02	12-22	11	ann	0.06
ORF6	RTFKVSIW	100.00% (34/34)	-0.18221	Non-Toxin	B*57:01	20-27	8		0.05
ORF7a	QLRARSVSPK	100.00% (28/28)	-0.16177	Non-Toxin	A*03:01	6-15	10	Consensus (ann/smm)	0.16
ORF7a	RARSVSPK	100.00% (28/28)	-0.27456	Non-Toxin	A*30:01	8-15	8	ann	0.11
ORF7b	FLAFLLFLV	100.00% (2/2)	0.20158	Non-Toxin	A*02:03	13-21	9	Consensus (ann/smm)	0.07
ORF8	HFYSKWYIR	100.00% (34/34)	-0.09452	Non-Toxin	A*31:01	40-48	9		0.11
S Protein	FPNITNLCPF	100.00% (41/41)	0.1009	Non-Toxin	B*53:01	49-58	10		0.06
S Protein	NYLYRLFR	100.00% (41/41)	0.13144	Non-Toxin	A*33:01	30-37	8	ann	0.07
S Protein	NYNYLYRLFR	100.00% (41/41)	0.08754	Non-Toxin	A*33:01	28-37	10	Consensus (ann/smm)	0.07
S Protein	WTAGAAAYYV	100.00% (41/41)	0.15455	Non-Toxin	A*68:02	48-57	10		0.06

Screening of Helper T lymphocyte (HTL) epitopes. The screening of helper T lymphocyte (HTL) epitopes from eleven different proteins of SARS-CoV-2 was performed on the basis of “Percentile rank”. The smaller the value of percentile rank the higher would be the affinity of the peptide with its respective HLA allele binders. The 42 epitopes with high percentile ranking were shortlisted (Table 2). Another 171 potential HTL cell epitopes with high Percentile rank, screened in our study are listed in Supplementary table S10.

589 **Table 2. Shortlisted high scoring SARS-CoV-2 HTL epitopes.** Selected high scoring
590 HTL epitopes and their respective HLA alleles binders are listed. *In-silico* analysis have
591 shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show
592 significant conservancy. Six epitope shown with ** show exact match with the epitopes
593 published by Grifoni et al., 2020, indicate consensus in results.
594
595

SARS-CoV- 19 Proteins	Peptide	Conservancy	Non-Toxin	Allele	Position	Method used	Percentile Rank
E Protein	VLLFLAFVVFLLVTL	100.00% (35/35)	Non-Toxin	DPA1*03:01/DPB1*04:02	17-31	Consensus (comb.lib./smm/nn)	0.02
E Protein	LLFLAFVVFLLVTLA	100.00% (35/35)	Non-Toxin	DPA1*03:01/DPB1*04:02	18-32		0.02
M Protein	GLMWLSYFIASFRLF	100.00% (41/41)	Non-Toxin	DPA1*01:03/DPB1*02:01	19-33		0.05
M Protein	LMWLSYFIASFRLFA	100.00% (41/41)	Non-Toxin	DPA1*01:03/DPB1*02:01	20-34		0.05
M Protein	LSYYKLGASQRVAGD**	100.00% (41/41)	Non-Toxin	DRB1*09:01	36-50		0.06
N Protein	PQIAQFAPSASAFFG	100.00% (40/40)	Non-Toxin	DRB1*09:01	22-36		0.01
N Protein	IAQFAPSASAFFGMS	100.00% (40/40)	Non-Toxin	DRB1*09:01	24-38		0.01
N Protein	AQFAPSASAFFGMSR	100.00% (40/40)	Non-Toxin	DRB1*09:01	25-39		0.01
ORF1ab	WFLAYILFTRFFYVL	100.00% (36/36)	Non-Toxin	DPA1*01:03/DPB1*02:01	17-31		0.01
ORF1ab	FLAYILFTRFFYVLG	100.00% (36/36)	Non-Toxin	DPA1*01:03/DPB1*02:01	18-32		0.01
ORF1ab	LAYILFTRFFYVLGL	100.00% (36/36)	Non-Toxin	DPA1*01:03/DPB1*02:01	19-33		0.01
ORF1ab	AYILFTRFFYVLGLA	100.00% (36/36)	Non-Toxin	DPA1*01:03/DPB1*02:01	20-34		0.01
ORF1ab	RAMPNMLRIMASLVL**	97.22% (35/36)	Non-Toxin	DRB1*15:01	46-60		0.01
ORF1ab	AMPNMLRIMASLVLA	97.22% (35/36)	Non-Toxin	DRB1*15:01	47-61		0.01
ORF1ab	MPNMLRIMASLVLAR	97.22% (35/36)	Non-Toxin	DRB1*15:01	48-62		0.01
ORF1ab	PNMLRIMASLVLARK	97.22% (35/36)	Non-Toxin	DRB1*15:01	49-63		0.01
ORF1ab	QESPFVMMSAPPAQY	100.00% (36/36)	Non-Toxin	DRB1*01:01	50-64		0.01
ORF1ab	NMLRIMASLVLARKH	97.22% (35/36)	Non-Toxin	DRB1*15:01	50-64		0.01
ORF1ab	ESPFVMMSAPPAQYE**	100.00% (36/36)	Non-Toxin	DRB1*01:01	51-65		0.01
ORF1ab	SPFVMMSAPPAQYEL	100.00% (36/36)	Non-Toxin	DRB1*01:01	52-66		0.01
ORF1ab	AIILASFSASTSAFV	100.00% (36/36)	Non-Toxin	DRB1*09:01	54-68		0.01
ORF1ab	IILASFSASTSAFVE	100.00% (36/36)	Non-Toxin	DRB1*09:01	55-69		0.01
ORF1ab	FAWWTAFVTNVNASS	97.22% (35/36)	Non-Toxin	DRB3*02:02	55-69	NetMHCIIpan	0.01
ORF3a	FVRATATIPIQASLP	100.00% (28/28)	Non-Toxin	DPA1*02:01/DPB1*14:01	28-42		0.12
ORF3a	LLFVTVYSHLLLVAA	100.00% (28/28)	Non-Toxin	DRB1*01:01	15-29	Consensus (comb.lib./smm/nn)	0.1
ORF6	TFKVSIWNLDYIINL**	100.00% (34/34)	Non-Toxin	DQA1*01:01/DQB1*05:01	21-35		0.02
ORF6	FKVSIWNLDYIINLI	100.00% (34/34)	Non-Toxin	DQA1*01:01/DQB1*05:01	22-36		0.02
ORF6	KVSIWNLDYIINLII	100.00% (34/34)	Non-Toxin	DQA1*01:01/DQB1*05:01	23-37		0.02
ORF7a	IILFLALITLATCEL	100.00% (28/28)	Non-Toxin	DRB1*01:01	3-17		0.16
ORF7a	ILFLALITLATCELY	100.00% (28/28)	Non-Toxin	DRB1*01:01	4-18		0.16
ORF7b	YLCFLAFLLFLVLIM	100.00% (2/2)	Non-Toxin	DPA1*03:01/DPB1*04:02	10-24		0.02
ORF7b	LCFLAFLLFLVLIML	100.00% (2/2)	Non-Toxin	DPA1*03:01/DPB1*04:02	11-25		0.02
ORF7b	CFLAFLLFLVLIMLI	100.00% (2/2)	Non-Toxin	DPA1*03:01/DPB1*04:02	12-26		0.03

ORF8	CTQHQPYVVDDPCPI	100.00% (34/34)	Non-Toxin	DRB3*01:01	25-39		0.08
ORF8	HQPYVVDDPCPIHFY	100.00% (34/34)	Non-Toxin	DRB3*01:01	28-42		0.08
ORF8	QPYVVDDPCPIHFYS	100.00% (34/34)	Non-Toxin	DRB3*01:01	29-43		0.07
ORF10	YINVFAFPFTIYSLL	100.00% (28/28)	Non-Toxin	DPA1*01:03/DPB1*02:01	3-17		0.29
S protein	KTQSLLIVNNATNVV	100.00% (41/41)	Non-Toxin	DRB1*13:02	43-57		0.01
S protein	TQSLLIVNNATNVVI	100.00% (41/41)	Non-Toxin	DRB1*13:02	44-58		0.01
S protein	QSLLIVNNATNVVIK	100.00% (41/41)	Non-Toxin	DRB1*13:02	45-59		0.01
S protein	SLLIVNNATNVVIKV**	100.00% (41/41)	Non-Toxin	DRB1*13:02	46-60		0.01
S protein	LLIVNNATNVVIKVC	100.00% (41/41)	Non-Toxin	DRB1*13:02	47-61		0.01

Population Coverage by CTL and HTL epitopes. The population coverage by the shortlisted epitopes was also studied, in particular involving China, France, Italy, United States of America, South Asia, East Asia, Northeast Asia and the Middle East. From this study, we may conclude that the combined use of all the shortlisted CTL and HTL epitopes would have an average worldwide population coverage as high as 96.10%, with a standard deviation of 31.17 (Table 3).

Table 3. World population coverage by the shortlisted SARS-CoV-2 CTL and HTL epitopes combined. With a standard deviation of 31.17 on an average 96.10 % of world population could be covered by the joint administration of selected CTL and HTL epitopes (given in Table 1 and 2) as vaccine candidate.

a projected population coverage
b average number of epitope hits / HLA combinations recognized by the population
c minimum number of epitope hits / HLA combinations recognized by 90% of the population

Population/area	Class combined		
	% coverage[a]	average_hit[b]	pc90[c]
Algeria	27.85	1.51	0.69
American Samoa	70.35	1.93	0.34
Argentina	96.78	3.56	1.35
Australia	79.62	2.73	0.49
Austria	99.28	7.17	2.74
Belgium	97.21	4.63	1.71
Bolivia	35.81	2.14	0.77
Borneo	30.61	1.53	0.72
Brazil	92.28	3.69	1.15
Bulgaria	95.49	4.48	1.64

Burkina Faso	61.26	1.3	0.26
Cameroon	88.9	2.71	0.9
Canada	8.75	0.46	0.54
Cape Verde	98.02	5.71	2.07
Central Africa	87.66	3.35	0.81
Central African Republic	36.4	1.25	0.16
Central America	16.77	0.65	0.12
Chile	90.76	4.36	1.06
China	90.43	4.74	1.05
Colombia	24.9	1.18	0.27
Congo	29.65	1.58	0.64
Cook Islands	22.52	1.33	0.65
Croatia	97.46	5.28	1.97
Cuba	97.36	6.08	2.03
Czech Republic	98.68	5.99	2.31
Denmark	59.3	3.61	1.23
East Africa	90.29	3.46	1.02
East Asia	93.23	5.49	1.29
Ecuador	51.11	2.55	0.2
England	99.45	7.37	3.14
Equatorial Guinea	15.47	0.81	0.58
Ethiopia	34.14	1.72	0.71
Europe	98.35	6.12	2.32
Fiji	23.54	1.33	0.59
Finland	99.58	5.27	2.46
France	98.2	6.39	2.31
Gabon	7.02	0.42	0.64
Georgia	93.92	4.08	1.31
Germany	99.34	7.17	2.92
Greece	26.34	1.51	0.68
Guatemala	17.11	0.67	0.12
Guinea-Bissau	95.03	4.49	1.37
Hong Kong	84.82	2.7	0.66
India	86.82	3.86	0.76
Indonesia	78.44	2.93	0.46
Iran	94.14	4.47	1.45
Ireland Northern	99.55	7.7	3.3
Ireland South	99.28	7.31	3.16
Israel	84.33	3.49	0.64
Italy	96.7	4.11	1.74
Ivory Coast	65.09	1.16	0.29
Jamaica	4.55	0.23	0.52
Japan	92.78	5.3	1.23
Jordan	85.1	3.54	0.67
Kenya	86.65	2.27	0.75
Kiribati	6.3	0.31	0.53
Korea; South	95.79	6.18	1.72
Lebanon	14.26	0.8	0.58
Macedonia	43.69	2.18	0.18
Malaysia	80.94	3.79	0.52
Mali	95.75	2.67	1.3
Martinique	43.47	1.8	0.18
Mexico	93.68	3.6	1.35

Mongolia	87.03	4.25	0.77
Morocco	97.32	5.55	2.01
Nauru	18.44	0.93	0.61
Netherlands	40.49	2.35	0.84
New Caledonia	61.85	2.73	0.26
New Zealand	24.11	1.39	0.66
Niue	20.26	1.2	0.63
North Africa	94.14	4.19	1.3
North America	97.36	6	2.09
Northeast Asia	90.3	4.71	1.04
Norway	52.39	3.08	1.05
Oceania	75.64	3.29	0.41
Oman	94.27	3.2	1.35
Pakistan	85	2.54	0.67
Papua New Guinea	71.23	3.29	0.35
Paraguay	0	0	0.6
Peru	95.12	3.38	1.52
Philippines	67.23	1.92	0.61
Poland	98.46	6.18	2.36
Portugal	96.27	5.25	1.75
Romania	95.9	3.54	1.56
Russia	94.88	5.26	1.49
Rwanda	45.12	1.85	0.18
Samoa	43.64	2.65	0.89
Sao Tome and Principe	96.32	4.15	1.54
Saudi Arabia	96.53	4.87	1.65
Scotland	73.06	4.19	0.37
Senegal	94.11	3.54	1.24
Serbia	62.91	1.07	0.27
Singapore	86.92	4.37	0.76
Slovenia	43.37	2.58	0.88
South Africa	88.81	2.6	0.89
South America	81.68	2.95	0.55
South Asia	90.37	4.25	1.03
Southeast Asia	82.33	3.79	0.57
Southwest Asia	86.38	3.13	0.73
Spain	87.57	3.8	0.8
Sri Lanka	52.39	1.59	0.42
Sudan	87.98	2.69	0.83
Sweden	99.29	7.32	2.59
Taiwan	84.93	4.28	0.66
Thailand	91.26	4.81	1.17
Tokelau	7.84	0.47	0.65
Tonga	33.43	1.98	0.69
Tunisia	94.64	4.17	1.38
Turkey	61.05	2.54	0.51
Uganda	92	2.99	1.13
United Arab Emirates	2.19	0.02	0.1
United States	97.45	6.04	2.11
Venezuela	69.56	1.16	0.33
Vietnam	86.19	4.11	0.72
Wales	1	0.01	0.1
West Africa	94.34	4.44	1.32

West Indies	96.71	5.04	1.74
Zambia	97.5	3.1	1.67
Zimbabwe	93.6	3.4	1.21
World	**96.10**	**5.55**	**1.8**
Average	**70.54**	**3.35**	**1.07**
Standard deviation	**31.17**	**1.87**	**0.73**

B Cell epitope prediction

Sequence-based B Cell epitope prediction. To screen B cell epitopes we utilized the Bepipred Linear Epitope Prediction method. In our study, we screened 12 B cell epitopes, from eleven SARS-CoV-2 proteins, which show partial or complete overlap with shortlisted CTL and HTL epitope (Table 4). Another 206 B Cell epitope, with the epitope length of at least four amino acids and maximum 20 amino acid were screened and are listed in Supplementary table S11.

Table 4. Shortlisted SARS-CoV-2 B Cell epitopes. BepiPred Linear B Cell epitopes showing sequence overlap with CTL and HTL epitopes are shortlisted. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant amino acid sequence conservancy.

S.No	SARS-2-CoV-Proteins	Overlapping B Cell Epitope (BepiPred method)	Conservancy	Length	Position	Toxicity
1	Membrane Protein	KLGASQRVAGDS	100.00% (41/41)	12	180-191	Non-Toxin
2	N protein	RLNQLESKMSGKGQQQQGQTVTKKSAAEASK KPRQKRTATKA	100.00% (40/40)	42	226-267	Non-Toxin
3	ORF1ab	GTTQTACTDDNALAYYNTTK	100.00% (36/36)	20	4157-4176	Non-Toxin
4	ORF3a	QGEIKDATPSDF	100.00% (28/28)	12	17-28	Non-Toxin
5	ORF3a	PYNSVT	100.00% (28/28)	6	159-164	Non-Toxin
6	ORF7a	LYHYQECVR	100.00% (28/28)	9	17-25	Non-Toxin
7	ORF7a	VKHVYQLRARSVSPKLFIRQEEVQEL	100.00% (28/28)	26	71-96	Non-Toxin
8	ORF8	QSCTQHQPYVVDDPCPIHFYSKW	100.00% (34/34)	23	23-45	Non-Toxin
9	ORF8	RVGARKSAP	100.00% (34/34)	9	48-56	Non-Toxin
10	S protein	TPGDSSSGWTA	100.00% (41/41)	11	250-260	Non-Toxin
11	S protein	FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA	100.00% (41/41)	35	329-363	Non-Toxin
12	S protein	NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIY QAGSTPCNGVEGFNCYFPLQSYGFQPTN	100.00% (41/41)	62	440-501	Non-Toxin

643
644 **Characterisation of potential epitopes**

645 *Epitope conservation analysis.* Sequence conservation analysis of the screened
646 CTL, HTL and B cell epitopes have shown highly conserved nature of the
647 shortlisted epitopes. Both the CTL epitopes and the HTL epitope were found to
648 be significantly conserved with their 100% amino acid sequence amongst the
649 NCBI retrieved protein sequences of SARS-CoV-2 (CTL epitopes 94.44% to
650 100% conserved and HTL epitopes were 97.22% to 100% conserved (Table 1, 2
651 4, Supplementary table S8, S9, S10 & S11).

652

653 *Epitope toxicity prediction.* Toxicity analysis of all the screened CTL, HTL and B
654 Cell epitopes was also performed. The ToxinPred study of all the shortlisted
655 epitopes shows that they all are non-toxic in nature (Table 1, 2 4, Supplementary
656 table S8, S9, S10 & S11).

657

658 *Overlapping residue analysis.* Amino acid sequence overlap analysis amongst
659 the shortlisted CTL, HTL and B cell epitopes from eleven SARS-CoV-2 proteins
660 was performed by the Multiple Sequence Alignment (MSA) analysis tool Clustal
661 Omega. The analysis has shown that several epitopes of CTL, HTL and B cell
662 were having amino acid sequence overlap. The CTL, HTL and B cell epitopes
663 having two or more than two amino acid residues overlap are shown in Fig.2.

664

665

666

Overlapping T & B cell epitopes across proteome of SARS-CoV-2

ORF1ab

```
                                                                                  --------YILFTRFFYV----
                                                                                  WFLAYILFTRFFYVL---
                                                                                  -----LAYILFTRFFYVLGL-
                                                                                  ---FLAYILFTRFFYVLG--
                                                                                  -------AYILFTRFFYVLGLA
                                                                                                          -----YIFFASFYYV
                                                                                                          RMYIFFASFY--
          -IILASFSASTSAFVE                  ----SPFVMMSAPPAQYEL
          AIILASFSASTSAFV-                  QESPFVMMSAPPAQY--
                                            -ESPFVMMSAPPAQYE-
474-489   AIILASFSASTSAFVE        1800-1816 QESPFVMMSAPPAQYEL        2327-2344 WFLAYILFTRFFYVLGLA        2382-2393 RMYIFFASFYYV

                                            RAMPNMLRIMASLVL----
                                            -------PNMLRIMASLVLARK-
                                            ---------NMLRIMASLVLARKH
          ----------CTDDNALAY---------      ----MPNMLRIMASLVLAR--
          ----------CTDDNALAYY-------      --AMPNMLRIMASLVLA---
          GTTQTACTDDNALAYYNTTK
4157-4176 GTTQTACTDDNALAYYNTTK      5016-5034 RAMPNMLRIMASLVLARKH
```

Surface protein

```
          --------LLIVNNATNVVIKVC
          ------SLLIVNNATNVVIKV-
          ----QSLLIVNNATNVVIK--
          --TQSLLIVNNATNVVI---
          KTQSLLIVNNATNVV----
113-131   KTQSLLIVNNATNVVIKVC

                                            ---------------WTAGAAAYYV
                                            TPGDSSSGWTA-------
                                            TPGDSSSGWTAGAAAYYV      250-267

                                                                    FPNITNLCPF-----------------------
                                                                    FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA
                                                                    FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA
                                                                    329-363

          ----------------NYNYLYRLFR----------------------------------------
          --------------------NYLYRLFR----------------------------------------
          NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN
440-501   NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN
```

ORF3a

```
          --------------------FVRATATIPIQASLP
          QGEIKDATPSDF--------------
17-42     QGEIKDATPSDFVRATATIPIQASLP

                                            IPYNSVTSSI
                                            -PYNSVT---
                                            158-167 IPYNSVTSSI
```

Envelope protein

```
          -LLFLAFVVF------
          -LLFLAFVVFLLVTLA
          VLLFLAFVVFLLVTL-
17-32     VLLFLAFVVFLLVTLA
```

Membrane Protein

```
          -------------YFIASFRLFAR
          --LMWLSYFIASFRLFA-
          GLMWLSYFIASFRLF--
89-105    GLMWLSYFIASFRLFAR

                                            ATSRTLSYYK----------
                                            ----------LSYYKLGASQRVAGD-
                                            ------------------KLGASQRVAGDS
                                            171-191 ATSRTLSYYKLGASQRVAGDS
```

ORF6

```
          AEILLIIMRTF--------------
          -------------RTFKVSIW----------
          ------------------KVSIWNLDYIINLII
          ----------------FKVSIWNLDYIINLI-
          --------------TFKVSIWNLDYIINL--
12-37     AEILLIIMRTFKVSIWNLDYIINLII
```

ORF7a

```
          -ILFLALITLATCELY-------
          IILFLALITLATCEL--------
          --------------LYHYQECVR
3-25      IILFLALITLATCELYHYQECVR

                                            -------------RARSVSPK----------
                                            ----------QLRARSVSPK----------
                                            VKHVYQLRARSVSPKLFIRQEEVQEL
                                            71-96 VKHVYQLRARSVSPKLFIRQEEVQEL
```

ORF7b

```
          ------FLAFLLFLV-----
          ----CFLAFLLFLVLIMLI
          --LCFLAFLLFLVLIML-
          YLCFLAFLLFLVLIM--
10-26     YLCFLAFLLFLVLIMLI
```

ORF8

```
          ----------------------------------HFYSKWYIR--------
          -------------QPYVVDDPCPIHFYS-------------
          ----------HQPYVVDDPCPIHFY--------------
          -----CTQHQPYVVDDPCPI-----------------
          QSCTQHQPYVVDDPCPIHFYSKW-----------
          ------------------------------------------------RVGARKSAP
23-56     QSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAP
```

Nucleocapsid Protein

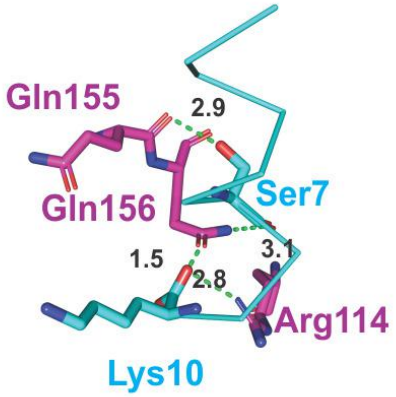
```
          ------------------------------------------------------------KPRQKRTAT--
          RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA
226-267   RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA

                                            -----AQFAPSASAFFGMSR
                                            ----IAQFAPSASAFFGMS-
                                            PQIAQFAPSASAFFG---
                                            302-319 PQIAQFAPSASAFFGMSR
```

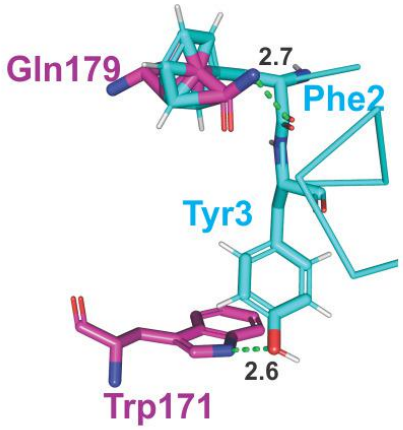
ORF10

```
          ---GYINVFAFPF------
          MGYINVFAF--------
          -----YINVFAFPFTIYSLL
1-17      MGYINVFAFPFTIYSLL
```

690 **Figure 3. Overlapping SARS-CoV-2 CTL, HTL and B cell epitopes**. Multiple
691 sequence alignment performed by Clustal Omega at EBI to identify the
692 consensus overlapping regions of CTL (red), HTL (blue) and B cell epitopes
693 (green) amongst shortlisted epitopes. Epitopes with overlapping region amongst
694 all the three types of epitopes (CTL, HTL and B Cell epitopes) were chosen for
695 further studies (encircled).
696
697 *Epitope selected for molecular interaction study with HLA allele and TAP*
698 *transporter.* The epitopes showing overlap amongst all the three type of epitopes
699 i.e CTL, HTL and B cell epitopes have been encircled in Fig.2 and are chosen for
700 further study for their interaction with HLA allele and TAP (Transporter
701 Associated with Antigen Processing) transporter.
702
703 **Molecular interaction analysis of selected epitopes with HLA allele and TAP**
704 **transporter.**
705 *Molecular interaction analysis of chosen CTL and HTL epitopes with HLA alleles.*
706 The molecular docking study of chosen CTL and HTL epitopes with their
707 respective HLA class I and II allele binders was performed by PatchDock tool.
708 The study revealed a significant molecular interaction between all the chosen
709 epitopes and their HLA allele binders showing multiple hydrogen bond formations
710 (Fig.3). Furthermore, the B-factor analysis of all the epitope-HLA allele
711 complexes has also shown the epitope ligand to have stable (blue) binding

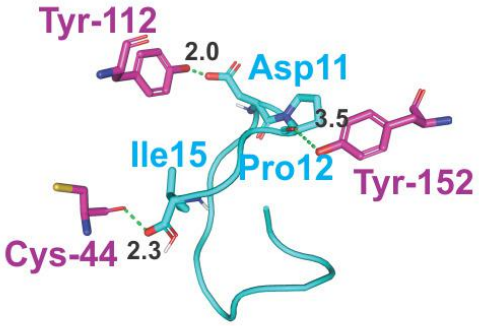
712 conformation in complex with the HLA allele molecule (VIBGYOR colour
713 presentation, blue being very stable) Fig.4.
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734

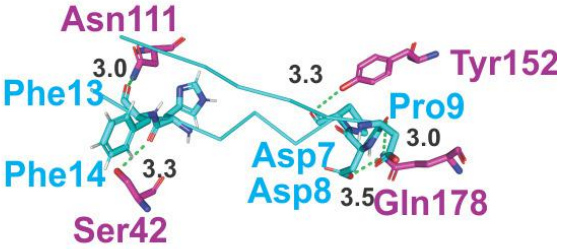




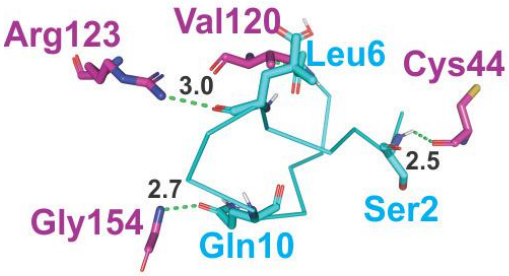
1. ATSRTLSYYK-HLA-A*11:01 2. HFYSKWYIR-HLA-A*31:01

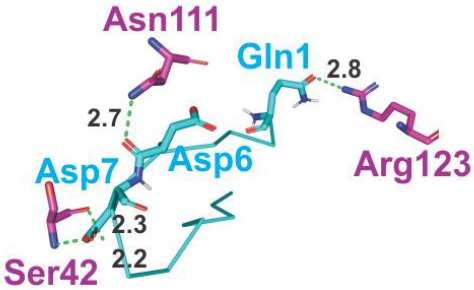
CTL epitopes - HLA Class I allele complexes





3. CTQHQPYVVDDPCPI-HLA-DRB3*01:01 4. HQPYVVDDPCPIHFY-HLA-DRB3*01:01

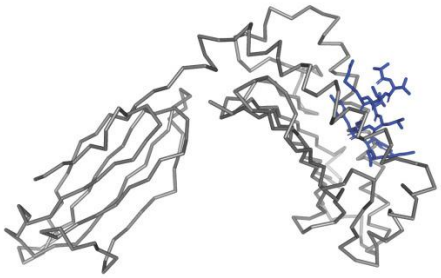




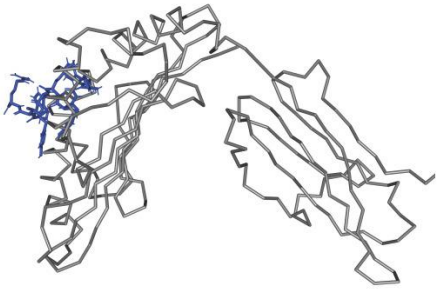
5. LSYYKLGASQRVAGD-HLA-DRB1*09:01 6. QPYVVDDPCPIHFYS-HLA-DRB3*01:01

HTL epitopes - HLA Class II allele complexes

758 **Figure 4. Molecular Docking analysis of SARS-CoV-2 CTL epitopes and**
759 **HLA alleles.** Molecular docking of chosen CTL and HTL epitopes (cyan sticks)
760 with binding amino acid residues of their respective HLA class I and class II allele
761 binders (magenta sticks). The study shows the docked complexes to form stable
762 complex with multiple hydrogen bonds (green dots) formation. Images were
763 generated by the PyMOL Molecular Graphics System, Version 2.0 Schrödinger,
764 LLC.
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780

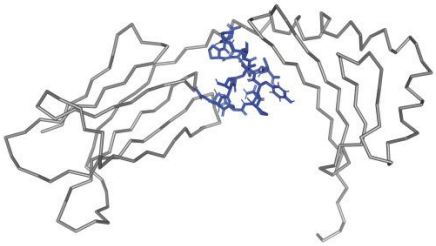


1. ATSRTLSYYK-HLA-A*11:01

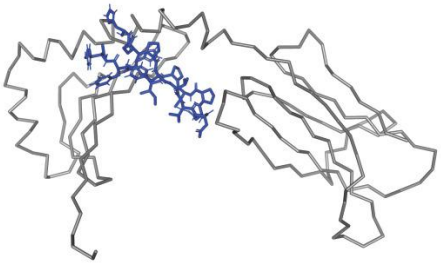


2. HFYSKWYIR-HLA-A*31:01

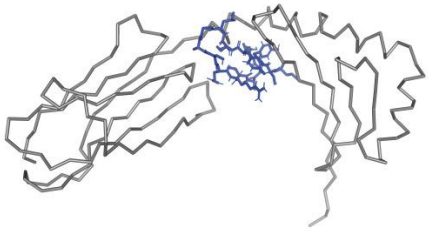
CTL epitopes - HLA Class I allele complexes



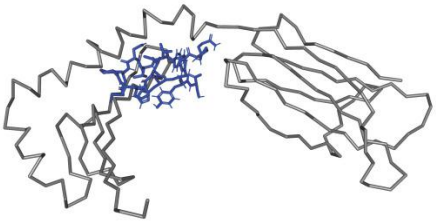
3. CTQHQPYVVDDPCPI-HLA-DRB3*01:01



4. HQPYVVDDPCPIHFY-HLA-DRB3*01:01



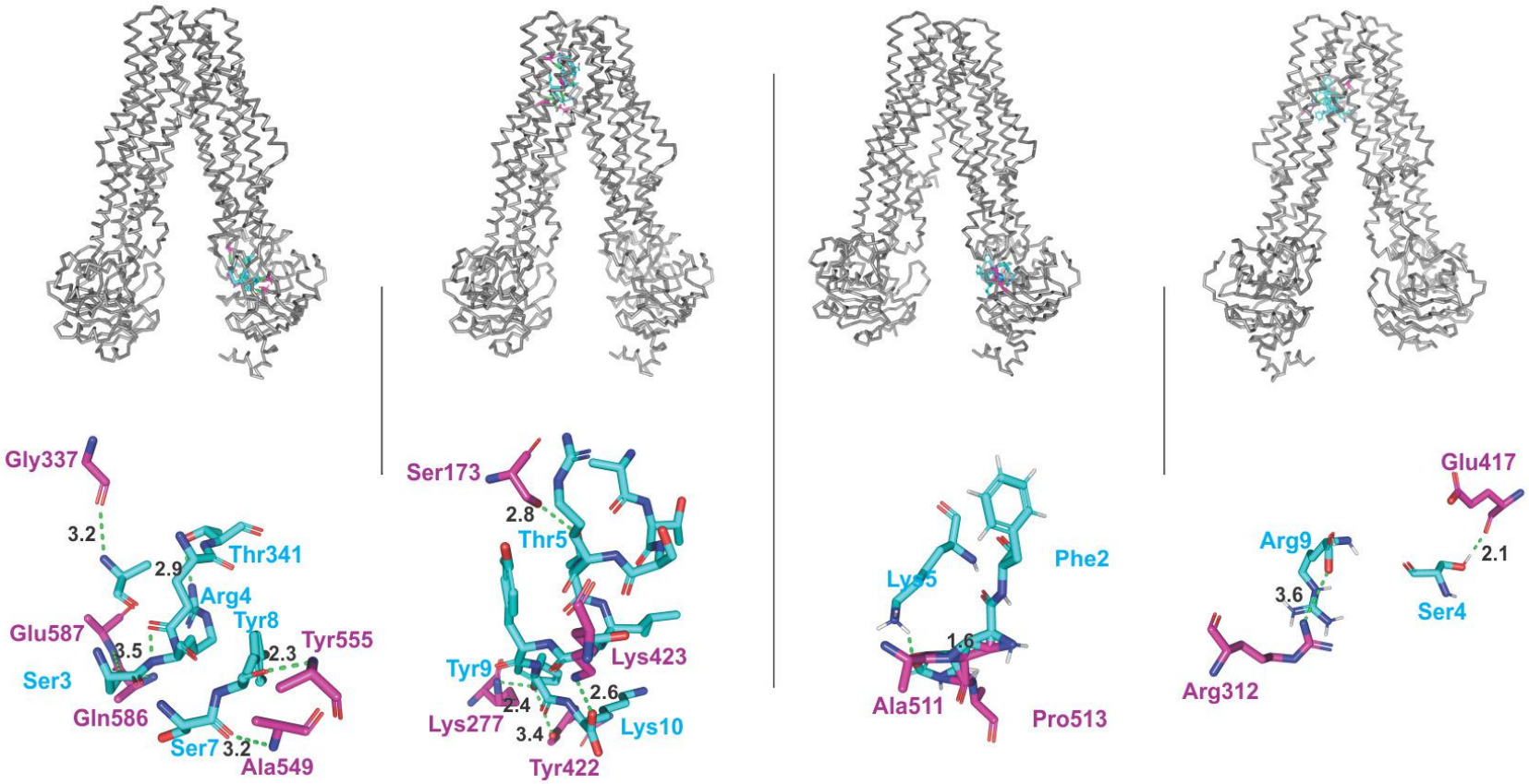
5. LSYYKLGASQRVAGD-HLA-DRB1*09:01



6. QPYVVDDPCPIHFYS-HLA-DRB3*01:01

HTL epitopes - HLA Class II allele complexes

804 **Figure 5. B-Factor of CTL and HTL epitope in complex with HLA class I and II**
805 **allele.** CTL and HTL Epitopes are shown in sticks and HLA Class I and II alleles are
806 shown in ribbon. The regions of the epitope in complex are shown in rainbow
807 (VIBGYOR), the regions in blue being very stable and the region towards red being
808 relatively unstable. In the complexes shown above, most of the regions of epitopes are in
809 blue indicating the complexes to be highly stable.
810
811 *Molecular interaction analysis of selected CTL epitopes with TAP transporter.*
812 The molecular docking interaction analysis of the chosen CTL epitopes with the
813 TAP transporter cavity has shown a significantly strong molecular interaction with
814 several hydrogen bonds formation at different sites of the TAP transporter cavity.
815 Two sites of interaction were of particular interest, one closer to the cytoplasmic
816 end and another closer to the ER lumen (Fig.5). This study confirms the
817 feasibility of transportation of chosen CTL epitopes from the cytoplasm to the ER
818 lumen which is an essential event for the representation of epitope by the HLA
819 allele molecules on the surface of antigen presenting cells.
820
821
822
823
824
825
826



Epitope ATSRTLSYYK through TAP cavity **Epitope HFYSKWYIR through TAP cavity**

827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849

850 **Figure 6. Molecular docking analysis of CTL epitopes within the TAP**
851 **transporter cavity.** Molecular interaction of CTL epitopes (cyan sticks) within the
852 TAP cavity (gray ribbon/sticks) is shown. Detailed interaction between the
853 residues of epitopes and TAP transporter residues have been shown with
854 hydrogen formation shown with green dots.
855
856 **Characterisation and molecular interaction analysis of designed Multi-**
857 **Epitope Vaccines with immunological receptor**
858 ***Characterisation of designed Multi-Epitope Vaccines***
859 *Physicochemical property analysis of designed MEVs.* ProtParam analysis for
860 both the CTL and HTL MEVs was performed to analyse their physiochemical
861 properties. The empirical physiochemical properties of the CTL and HTL MEVs
862 are given in the table 5. The aliphatic index and grand average of hydropathicity
863 (GRAVY) of both the MEVs indicate globular and hydrophilic nature of the both
864 the MEVs. The instability index score of both the MEVs indicate stable nature of
865 the protein molecules (Table 5).
866
867
868
869
870
871

872 **Table 5. Physicochemical property analysis based on amino acid sequence**
873 **of designed CTL and HTL multi-epitope vaccine.**

MEVs	Amino acid length	Molecular Weight	Theoretical pI	Expected half-life	Aliphatic index	Grand average of hydropathicity (GRAVY)	Instability index score
CTL MEV	704 aa	72.62 kDa	9.70	E.Coli: 10 Hr Yeast: 20 Hr Mammalian cell: 30 Hr	61.09	-0.090	44.31
HTL MEV	989 aa	101.31 kDa	9.39	Coli: 10 Hr Yeast: 20 Hr Mammalian cell: 30 Hr	96.35	0.539	43.31

874

875

876 *Interferon-gamma inducing epitope prediction.* Interferon-gamma (IFN-γ)
877 inducing epitopes are involved in both the adaptive as well as in the innate
878 immune response. The IFN-γ inducing 15 mer peptide epitopes were screened
879 from the amino acid sequence of CTL and HTL MEVs by utilizing the IFNepitope
880 server. A total of 20 CTL MEV and 20 HTL MEV INF-γ inducing POSITIVE
881 epitopes with a score of 1 or more than 1 were shortlisted (Supplementary table
882 S2).

883

884

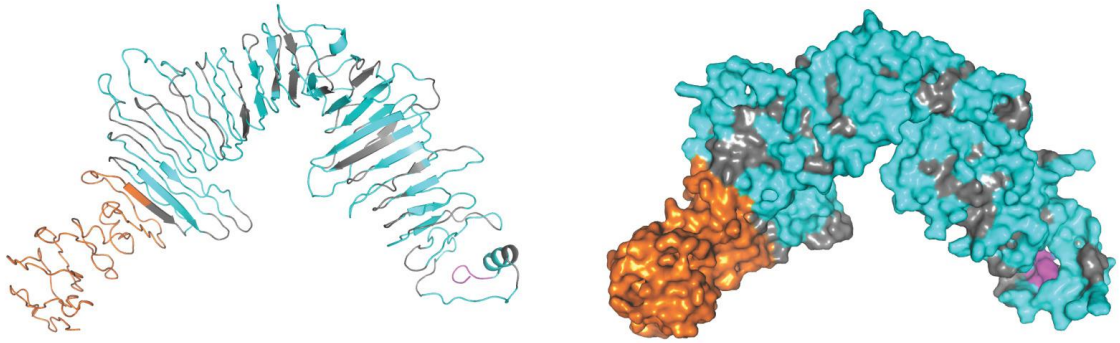
885

886

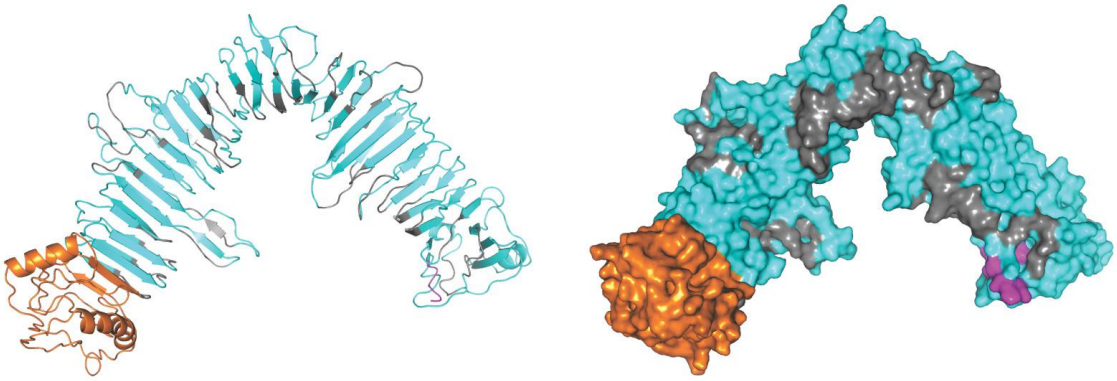
887

888

889

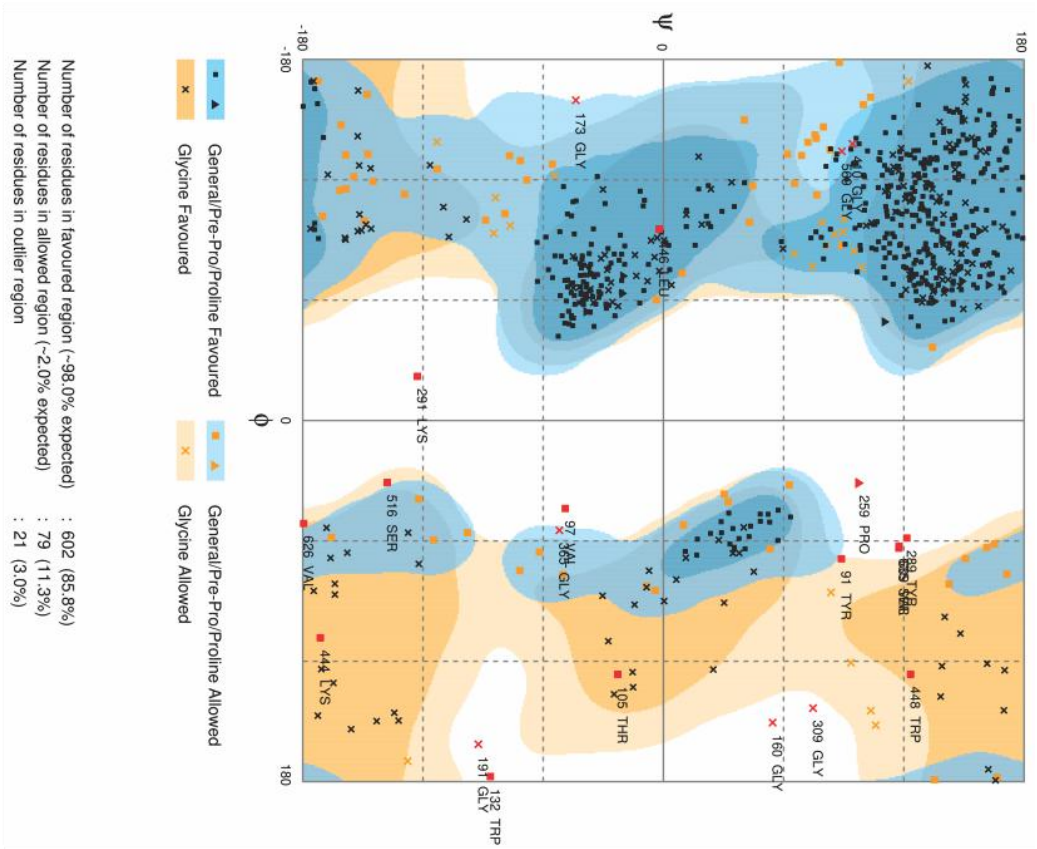


A. CTL Multi-epitope vaccine model

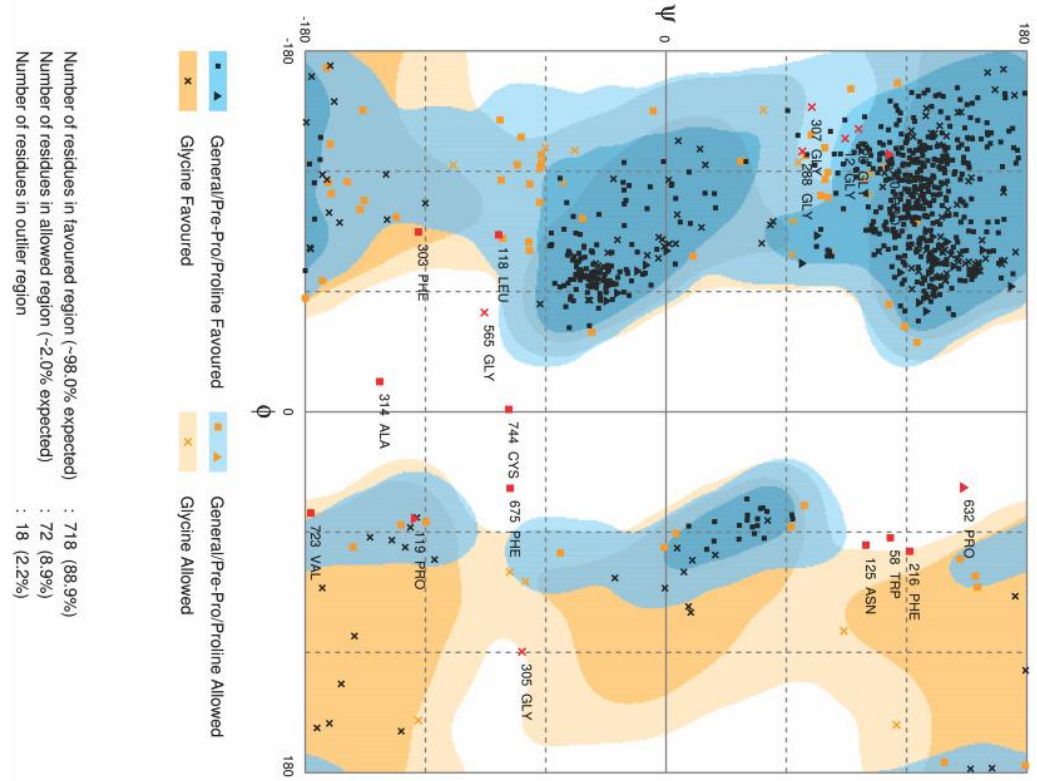


B. HTL Multi-epitope vaccine model

Figure 7. Tertiary structure modelling of CTL and HTL Multi-Epitope Vaccines. Tertiary structural models of CTL and HTL MEVs have been shown. The epitopes are shown in cyan. The adjuvant (Ov-ASP-1) is shown in orange. The linkers are shown in gray. Both the cartoon and surface presentation of both the MEVs are shown.



A. CTL Multi-epitope vaccine RAMPAGE analysis



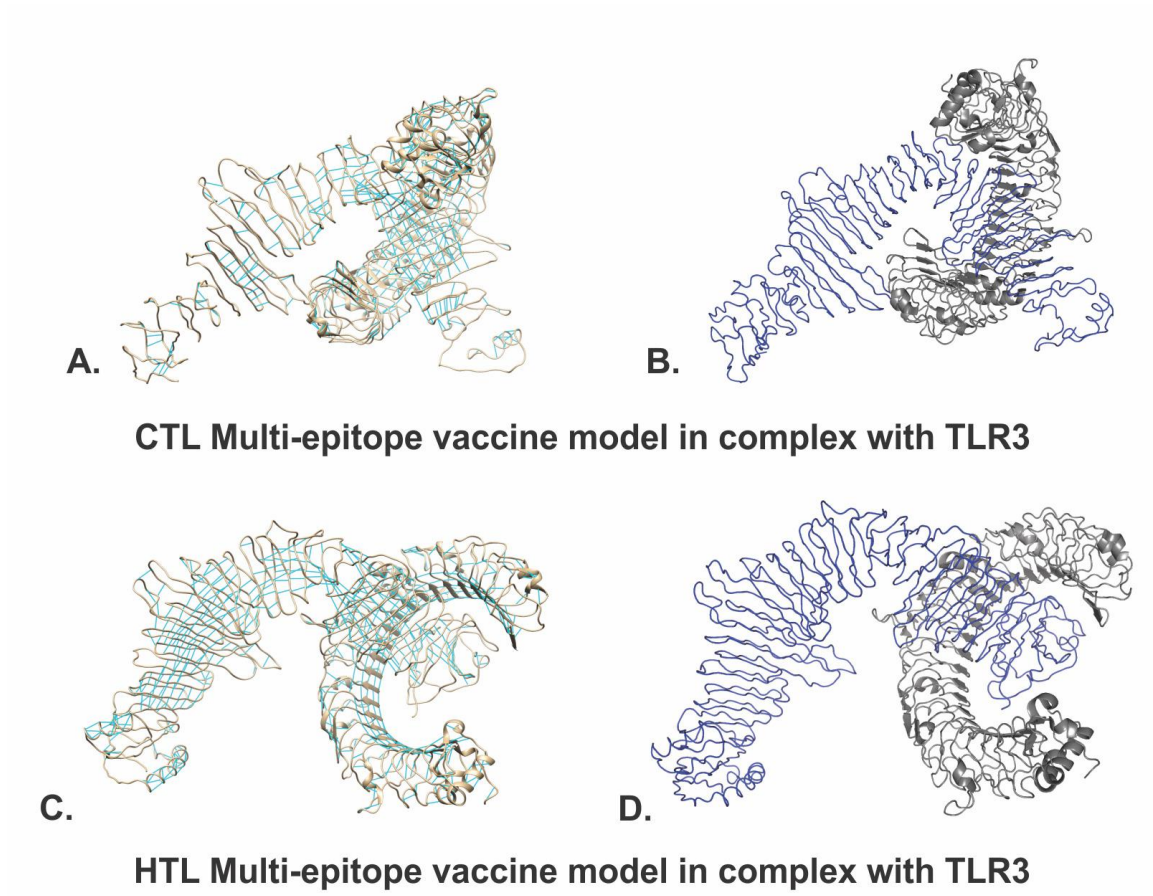
B. HTL Multi-epitope vaccine model RAMPAGE analysis

936 **Figure 8. RAMPAGE analysis of CTL and HTL MEVs.** The RAMPAGE analysis
937 of both the CTL and HTL MEVs have been done and shown here.
938
939 *MEVs allergenicity and antigenicity prediction.* Both the CTL and HTL MEVs
940 were found to be NON-ALLERGEN by the AlgPred analysis (scoring -
941 0.95185601 and -1.1293352 respectively; threshold being -0.4). The CTL and
942 HTL MEVs were also analyzed by VaxiJen to be probable ANTIGENS (prediction
943 score 0.4485 and 0.4215 respectively; default threshold being 0.4). Hence with
944 the mentioned analysis tools both the CTL and HTL MEVs are predicted to be
945 non-allergic as well as potentially antigenic in nature.
946
947 *Tertiary structure modelling and refinement of MEVs.* 3D homology models were
948 generated for both the CTL and HTL MEVs by utilizing the I-TASSER modeling
949 tool (Fig.7). The models were generated for CTL Multi-epitope vaccine (PDB hit
950 5n8pA, Norm. Z-score of 1.49, Cov of 0.92, TM-score of 0.916 and RMSD of
951 1.04 Å) and HTL Multi-epitope vaccine (PDB hit 5n8pA, Norm. Z-score of 1.52,
952 Cov of 0.97, TM-score of 0.916 and RMSD of 1.04 Å).
953 Both the generated CTL and HTL 3D models were further refined by
954 ModRefiner to fix any gaps and then followed by GalaxyRefine refinement. The
955 refinement by ModRefiner showed the TM-score of 0.9189 and 0.9498 for the
956 CTL and HTL models respectively, hence being close to 1, the initial and the
957 refined models were structurally similar. After refinement, the RMSD for CTL and
958 HTL models with respect to the initial model was 3.367Å and 2.318Å respectively.

959 Further, both the CTL and HTL MEVs models were refined by GalaxyRefine and
960 model 1 was chosen on the basis of best scorings parameters. The CTL MEV
961 model refinement output model (Rama favoured was 83.6%, GDT-HA was
962 0.9371, RMSD was 0.459, MolProbity was 2.539, Clash score was 23.2, and
963 Poor rotamers was 1.8) and the HTL MEV model refinement output model (Rama
964 favoured was 87.7%, GDT-HA was 0.9552, RMSD was 0.402, MolProbity was
965 2.537, Clash score was 27.9, and Poor rotamers was 1.6) show a well refined
966 and acceptable models for both of the MEVs. After refinement, all the mentioned
967 parameter have found to be improved significantly in comparison to the initial
968 CTL and HTL MEV models (Supplementary table S3).
969
970 *Validation of CTL and HTL MEVs refined models.* Both the CTL and HTL model
971 were analyzed by the RAMPAGE analysis tool after refinement. The refined CTL
972 MEV model was found to have 85.8% residues in favored region, 11.3% residues
973 in allowed region, and only 3.0% residues in the outlier region; while the refined
974 HTL MEV model was found to have 88.9% of residues in favored region, 8.9%
975 residues in allowed region, and only 2.2% residues in the outlier region (Fig.8)
976
977 *Linear and Discontinuous B-cell epitope prediction from MEVs.* The linear and
978 discontinuous B-cell epitope prediction was performed to enlist potential linear
979 and discontinuous epitopes from the refined 3D models of CTL and HTL MEVs
980 utilizing the ElliPro tool available on IEDB server. The screening revealed that the
981 CTL MEV carries 17 linear and 2 potential discontinuous B cell epitopes and that

982 of HTL MEV carries 17 linear and 4 potential discontinuous epitopes. The high
983 range of the PI (Protrusion Index) score of the linear and discontinuous epitopes
984 from CTL and HTL MEV s show high potential of the epitopes to cause humoral
985 immune response (PI score: CTL MEV linear & discontinuous B cell epitopes -
986 0.511 to 0.828 & 0.664 to 0.767 respectively; HTL MEV linear and discontinuous
987 B cell epitopes - 0.518 to 0.831 & 0.53 to 0.776 respectively) (Supplementary
988 table S4,S5, S6, S7).
989
990 ***Molecular interaction analysis of MEVs with immunological receptor.***
991 *Molecular docking and dynamics simulation study of MEVs and TLR-3.* Both the
992 refined models of CTL and HTL MEVs were further studied for their molecular
993 interaction with the ectodomain (ECD) of human TLR-3. Therefore, molecular
994 docking of CTL and HTL MEVs model with the TLR-3 crystal structure model
995 (PDB ID: 2A0Z) was performed utilizing the PatchDock tool. Generated docking
996 conformation with highest scores of 20776 and 20350 for CTL and HTL MEVs
997 respectively were chosen for further study. The highest docking score indicates
998 the best geometric shape complementarity fitting conformation of MEVs and the
999 TLR-3 receptor as predicted by the PatchDock tool. Both the CTL and HTL MEVs
1000 were fitting into the ectodomain region of TLR-3 after docking (Fig.9A, 9C). The
1001 CTL and HTL MEVs have shown to form multiple hydrogen bonds within the
1002 ectodomain cavity region of TLR-3.
1003 The B-factor analysis of MEVs-TLR3 complexes was also performed. The
1004 B-factor indicates the displacement of the atomic positions from an average

(mean) value i.e. the more flexible an atom is the larger the displacement from the mean position will be (mean-squares displacement) (Fig.9B, 9D). The B-factor analysis of the CTL and HTL MEVs bound to the TLR3 receptor shows that most of the regions of MEVs bound to TLR3 are stable nature in nature. The B-Factor analysis has been represented by the VIBGYOR color presentation with blue represents low B-factor and red represents high B-factor. Hence, the results suggest a stable complex formation tendency for both the CTL and HTL MEVs with the ectodomain of the human TLR-3 receptor (Fig.9B, 9D).



1028 **Figure 9. Molecular Docking study of CTL and HTL MEVs with TLR-3**
1029 **receptor. (A), (C)** The docking complex of CTL-TLR and HTL-TLR3 have been
1030 shown. The TLR3 is shown in cartoon, and the MEVs are shown in ribbon. The
1031 Hydrogen bond formation is shown by cyan lines. **(B), (D)** B-factor presentation
1032 for the docked MEVs to the TLR3 receptor. The presentation is in VIBGYOR
1033 color, with blue showing low B-factor and red show high B-factor. Here most of
1034 the MEV regions are in blue showing low B-factor and hence indicate stable
1035 complex formation with TLR3 receptor. Images were generated by PyMol and
1036 UCSF Chimera (Pettersen et al., 2004).
1037

1038 ***In-silico* analysis of MEVs for cloning and expression potency**

1039 *Analysis of cDNA of both the MEVs for cloning and expression in mammalian*
1040 *host cell line.* Complementary DNA optimized for CTL and HTL expression in the
1041 mammalian host cell line (Human) was generated by utilizing the Java Codon
1042 Adaptation Tool. Further, the generated optimized cDNA's for both the MEVs
1043 were analyzed by utilizing the GenScript Rare Codon Analysis Tool. The analysis
1044 revealed that the codon optimized cDNA of both the CTL and HTL MEVs bear all
1045 the crucial and favorable compositions for high level expression in mammalian
1046 cell line (CTL-MEV: GC content 70.40%, CAI (Codon Adaptation Index) score
1047 1.00 and 0% tandem rare codons; HTL MEV: GC content 69.26%, CAI score
1048 1.00 and 0% tandem rare codons). Ideally, the GC content of a cDNA should be
1049 30% to 70%, CAI score that indicates the possibility of cDNA expression in a
1050 chosen expression system should be between 0.8-1.0, and the tandem rare

1051 codon frequency that indicates the presence of low-frequency codons in cDNA
1052 should be <30%. The tandem rare codons may hinder the proper expression of
1053 the cDNA or even interrupt the translational machinery of the chosen expression
1054 system. Hence as per the GenScript Rare Codon Analysis, the cDNA of both the
1055 MEVs satisfy all the mentioned parameters and are predicted to have high
1056 expression in the mammalian host cell line (Human).
1057

1058 **CONCLUSION**

1059 In the present study, we have designed and proposed two multi-epitope
1060 vaccines derived from multiple CTL and HTL epitopes against SARS-CoV-2
1061 (COVID19). The chosen CTL and HTL epitopes show significant sequence
1062 overlap with screened linear B cell epitopes. The shortlisted CTL and HTL
1063 epitopes were utilized to design CTl and HTL multi-epitope vaccine. Both the
1064 generated CTL and HTL multi-epitope vaccine tertiary models have shown to
1065 carry potential linear and discontinuous B cell epitopes as well as potential INF-γ
1066 epitopes. Hence the designed MEVs are predicted to have the potential to elicit
1067 humoral as well as cellular immune responses. Since Onchocerca volvulus
1068 activation-associated secreted protein-1 (Ov-ASP-1) binds to the APCs and
1069 trigger pro-inflammatory cytokine production via Toll-like receptor 3 (TLR3), the
1070 truncated (residues 10-153) Ov-ASP-1 has been utilized as an adjuvant at N
1071 terminal of both the CTL and HTL MEVs models. Chosen overlapping clustering
1072 epitopes were validated for their molecular interaction with their respective HLA
1073 allele binders by molecular docking studies. Molecular interaction of the chosen

1074 CTL epitopes within the TAP transporter cavity was also analyzed. Analysis of
1075 the average world population coverage by both the shortlisted CTL and HTL
1076 epitopes combined revealed coverage of 96.10% world population. The
1077 molecular interaction analysis of both the CTL and HTL MEVs with the
1078 immunoreceptor TLR3 have shown very convincing structural fitting of the MEVs
1079 into the ectodomain of TLR3 receptor cavity. The cDNA for both the MEVs were
1080 generated considering codon-biasing for expression in the mammalian host cell
1081 line (Human). Both the cDNA has been optimized in respect of their GC content
1082 and zero tandem rare codons for the cDNA to have high expression possibility in
1083 the mammalian host cell line (Human). Hence for further studies, both the design
1084 of CTL and HTL MEVs could be cloned, expressed and tested for *in-vivo*
1085 validations and animal trials as potential vaccine candidates against SARS-CoV-
1086 2 infection.

1087

1088 **Supplementary table S1. Homology modeling for HLA alleles.** Tertiary
1089 structures of HLA alleles were modeled by homology modeling using SwissModel
1090 server. Templates were chosen with highest sequence identity. Generated
1091 models with acceptable QMEAN value were chosen for further studies.
1092
1093 **Supplementary table S2. INF-γ epitopes from CTL and HTL MEVs.** INF-γ
1094 inducing (POSITIVE) epitopes from CTL and HTL MEVs were screened by using
1095 "Motif and SVM hybrid" (MERCI & SVM) approaches.
1096
1097 **Supplementary table S3. Refinement models of CTL and HTL MEVs.** Both
1098 the CTL and HTL MEVs models were refined by GalaxyWEB server. After
1099 refinement in particular Rama favored residues increased significantly.
1100
1101 **Supplementary table S4. B Cell linear epitopes of CTL MEVs.** Linear B Cell
1102 epitopes predicted by ElliPro (IEDB) from CTL MEVs.
1103
1104 **Supplementary table S5. B Cell discontinuous epitopes of CTL MEVs.**
1105 Discontinuous B Cell epitopes predicted by ElliPro (IEDB) from CTL MEVs.

Supplementary table S6. B Cell linear epitopes of HTL MEVs. Linear B Cell epitopes predicted by ElliPro (IEDB) from HTL MEVs.

Supplementary table S7. B Cell discontinuous epitopes of HTL MEVs. Discontinuous B Cell epitopes predicted by ElliPro (IEDB) from HTL MEVs.

Supplementary table S8. Shortlisted high scoring CTL epitopes (MHC-I Binding Predictions). Selected high scoring CTL epitopes and their respective HLA alleles binders predicted by "MHC-I Binding Predictions" IEDB tool. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy and high immunogenicity.

Supplementary table S9. Shortlisted high scoring CTL epitopes (MHC-I Processing Predictions). Selected high scoring CTL epitopes and their respective HLA alleles binders predicted by "MHC-I Processing Predictions" IEDB tool. The screening gives detailed and combined scoring "Total score" for Proteasomal cleavage/TAP transport/MHC class I combined. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy with high immunogenicity.

Supplementary table S10. Shortlisted high scoring HTL epitopes (MHC-II Binding Predictions). Selected high "Percentile rank" HTL epitopes with their respective HLA class II alleles binders predicted by the "**MHC-II Binding Predictions**" tool of IEDB are listed. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy.

Supplementary table S11. Shortlisted B Cell epitopes (BepiPred Linear B Cell Prediction). B cell linear epitopes with length of 4 to 20 amino acids, predicted by the "BepiPred Linear B Cell Prediction" IEDB tool, from eleven SARS-CoV-2 proteins, are listed here. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant amino acid sequence conservancy.

AUTHOR CONTRIBUTION:

Protocol design: S.S., M.K.; Methodology performed by S.S., S.V., M.K., R.K.; Global Economic risk analysis: R.K.B.; Data analysis, scientific writing and revising the article: S.S., S.V., M.K., R.K., R.K.B., A.K.S., H.J.S., K.P., M.Kolbe.

1146 **ADDITIONAL INFORMATION:**

1147 Authors declare to have no competing interests.

1148 **REFERENCES:**

1149 Abele, R. and Tampé, R., 2004. The ABCs of immunology: structure and function
1150 of TAP, the transporter associated with antigen processing. Physiology,
1151 19(4), pp.216-224.
1152 Antoniou, A.N., Powis, S.J. and Elliott, T., 2003. Assembly and export of MHC
1153 class I peptide ligands. Current opinion in immunology, 15(1), pp.75-81.
1154 Arnold K, Bordoli L, Kopp J, and Schwede T (2006). The SWISS-MODEL
1155 Workspace: A web-based environment for protein structure homology
1156 modelling. Bioinformatics.,22,195-201.
1157 Bell, J.K., Botos, I., Hall, P.R., Askins, J., Shiloach, J., Segal, D.M. and Davies,
1158 D.R., 2005. The molecular structure of the Toll-like receptor 3 ligand-
1159 binding domain. Proceedings of the National Academy of Sciences of the
1160 United States of America, 102(31), pp.10976-10980.
1161 Benkert, P., Tosatto, S.C. and Schomburg, D., 2008. QMEAN: A comprehensive
1162 scoring function for model quality assessment. Proteins: Structure,
1163 Function, and Bioinformatics, 71(1), pp.261-277.
1164 Bui H. H, Sidney J, Dinh K, Southwood S, Newman M. J, Sette A. 2006.
1165 Predicting population coverage of T-cell epitope-based diagnostics and
1166 vaccines. BMC Bioinformatics 17:153.
1167 Bui HH, Sidney J, Li W, Fusseder N, Sette A. 2007. Development of an epitope
1168 conservancy analysis tool to facilitate the design of epitope-based
1169 diagnostics and vaccines. BMC Bioinformatics 8:361.
1170 Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A,
1171 Kesmir C, Peters B. 2013. Properties of MHC class I presented peptides
1172 that enhance immunogenicity. PloS Comp. Biol. 8(1):361.
1173 Chen, X., Zaro, J.L. and Shen, W.C., 2013. Fusion protein linkers: property,
1174 design and functionality. Advanced drug delivery reviews, 65(10),
1175 pp.1357-1369.
1176 Delneste, Y., Beauvillain, C. and Jeannin, P., 2007. Innate immunity: structure
1177 and function of TLRs. Medecine sciences: M/S, 23(1), pp.67-73.
1178 Dhanda, S. K., Vir, P. & Raghava, G. P. Designing of interferon-gamma inducing
1179 MHC class-II binders. Biol. Direct. 8, 30 (2013).
1180 Dong Xu and Yang Zhang. Improving the Physical Realism and Structural
1181 Accuracy of Protein Models by a Two-step Atomic-level Energy
1182 Minimization. Biophysical Journal, vol 101, 2525-2534 (2011).

1183 Duhovny D, Nussinov R, Wolfson HJ. Efficient Unbound Docking of Rigid
1184 Molecules. In Gusfield et al., Ed. Proceedings of the 2'nd Workshop on
1185 Algorithms in Bioinformatics(WABI) Rome, Italy, Lecture Notes in
1186 Computer Science 2452, pp. 185-200, Springer Verlag, 2002
1187 Farina, C., Krumbholz, M., Giese, T., Hartmann, G., Aloisi, F. and Meinl, E., 2005.
1188 Preferential expression and function of Toll-like receptor 3 in human
1189 astrocytes. Journal of neuroimmunology, 159(1-2), pp.12-19.
1190 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M.R., Appel,
1191 R.D. and Bairoch, A., 2005. Protein identification and analysis tools on the
1192 ExPASy server (pp. 571-607). Humana Press.
1193 Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R.H., Peters, B. and Sette, A.,
1194 2020. A Sequence Homology and Bioinformatic Approach Can Predict
1195 Candidate Targets for Immune Responses to SARS-CoV-2. Cell Host &
1196 Microbe.
1197 Guo, J., Yang, Y., Xiao, W., Sun, W., Yu, H., Du, L., Lustigman, S., Jiang, S.,
1198 Kou, Z. and Zhou, Y., 2015. A truncated fragment of Ov-ASP-1 consisting
1199 of the core pathogenesis-related-1 (PR-1) domain maintains adjuvanticity
1200 as the full-length protein. Vaccine, 33(16), pp.1974-1980.
1201 Gupta, S., Kapoor, P., Chaudhary, K., Gautam, A., Kumar, R., Raghava, G.P.
1202 and Open Source Drug Discovery Consortium, 2013. In silico approach for
1203 predicting toxicity of peptides and proteins. PLoS One, 8(9), p.e73957.
1204 Hajighahramani, N., Nezafat, N., Eslami, M., Negahdaripour, M., Rahmatabadi,
1205 S.S. and Ghasemi, Y., 2017. Immunoinformatics analysis and in silico
1206 designing of a novel multi-epitope peptide vaccine against Staphylococcus
1207 aureus. Infection, Genetics and Evolution, 48, pp.83-94.
1208 He, Y., Barker, S.J., MacDonald, A.J., Yu, Y., Cao, L., Li, J., Parhar, R., Heck, S.,
1209 Hartmann, S., Golenbock, D.T. and Jiang, S., 2009. Recombinant Ov-
1210 ASP-1, a Th1-biased protein adjuvant derived from the helminth
1211 Onchocerca volvulus, can directly bind and activate antigen-presenting
1212 cells. The Journal of Immunology, 182(7), pp.4005-4016.
1213 Hoof, I., Peters, B., Sidney, J., Pedersen, L.E., Sette, A., Lund, O., Buus, S. and
1214 Nielsen, M., 2009. NetMHCpan, a method for MHC class I binding
1215 prediction beyond humans. Immunogenetics, 61(1), p.1.
1216 Hu W, Li F, Yang X, Li Z, Xia H, Li G, Wang Y, Zhang Z. A flexible peptide linker
1217 enhances the immunoreactivity of two copies HBsAg preS1 (21-47) fusion
1218 protein. J Biotechnol. 2004;107:83–90.
1219 Irini A Doytchinova and Darren R Flower. VaxiJen: a server for prediction of
1220 protective antigens, tumour antigens and subunit vaccines. BMC
1221 Bioinformatics. 2007 8:4.

1222 Ko, J., Park, H., Heo, L., and Seok, C., GalaxyWEB server for protein structure
1223 prediction and refinement, Nucleic Acids Res. 40 (W1), W294-W297
1224 (2012).
1225 Kringelum, J.V., Lundegaard, C., Lund, O. and Nielsen, M., 2012. Reliable B cell
1226 epitope predictions: impacts of method development and improved
1227 benchmarking. PLoS computational biology, 8(12).
1228 Larsen JE, Lund O, Nielsen M. 2006. Improved method for predicting linear B-
1229 cell epitopes. Immunome Res 2:2.
1230 MacDonald, A.J., Cao, L., He, Y., Zhao, Q., Jiang, S. and Lustigman, S., 2005.
1231 rOv-ASP-1, a recombinant secreted protein of the helminth Onchocerca
1232 volvulus, is a potent adjuvant for inducing antibodies to ovalbumin, HIV-1
1233 polypeptide and SARS-CoV peptide antigens. Vaccine, 23(26), pp.3446-
1234 3452.
1235 Morla, S., Makhija, A. & Kumar, S. Synonymous codon usage pattern in
1236 glycoprotein gene of rabies virus. Gene. 584, 1–6 (2016).
1237 Nagpal, G., Gupta, S., Chaudhary, K., Dhanda, S.K., Prakash, S. and Raghava,
1238 G.P., 2015. VaccineDA: Prediction, design and genome-wide screening of
1239 oligodeoxynucleotide-based vaccine adjuvants. Scientific reports, 5,
1240 p.12478.
1241 NCBI protein sequence database: SARS-CoV-2 (Severe acute respiratory
1242 syndrome coronavirus 2) Sequences
1243 (https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/)
1244 Nielsen, M., Lundegaard, C. and Lund, O., 2007. Prediction of MHC class II
1245 binding affinity using SMM-align, a novel stabilization matrix alignment
1246 method. BMC bioinformatics, 8(1), p.238.
1247 Oldham, M.L., Grigorieff, N. and Chen, J., 2016. Structure of the Transporter
1248 associated with antigen processing trapped by herpes simplex virus. eLife,
1249 5, p.e21829.
1250 Peters B, Bulik S, Tampe R, Van Endert PM, Holzhutter HG. 2003. Identifying
1251 MHC class I epitopes by predicting the TAP transport efficiency of epitope
1252 precursors. J Immunol171:1741-1749.
1253 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M.,
1254 Meng, E.C. and Ferrin, T.E., 2004. UCSF Chimera—a visualization
1255 system for exploratory research and analysis. Journal of computational
1256 chemistry, 25(13), pp.1605-1612.
1257 Ponomarenko JV, Bui H, Li W, Fusseder N, Bourne PE, Sette A, Peters B. 2008.
1258 ElliPro: a new structure-based tool for the prediction of antibody epitopes.
1259 BMC Bioinformatics 9:514.
1260 PyMOL: The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

Ramakrishnan, C. and Ramachandran, G.N., 1965. Stereochemical criteria for polypeptide and protein chain conformations: II. Allowed conformations for a pair of peptide units. Biophysical journal, 5(6), pp.909-933.

Roy, A., Kucukural, A. and Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature protocols, 5(4), p.725.

S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson (2002) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450.

Saha, S. & Raghava, G. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic. Acids. Res. 34, W202–W209 (2006).

Schneidman-Duhovny D, Inbar Y, Nussinov R, Wolfson HJ. PatchDock and SymmDock: servers for rigid and symmetric docking. Nucl. Acids. Res. 33: W363-367, 2005.

Shen, Y., Maupetit, J., Derreumaux, P. and Tufféry, P., 2014. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. Journal of chemical theory and computation, 10(10), pp.4745-4758.

Shin, W.H., Lee, G.R., Heo, L., Lee, H. and Seok, C., 2014. Prediction of protein structure and interaction by GALAXY protein modeling programs. Bio Design, 2(1), pp.1-11.

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A. and Peters, B., 2008. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. Immunome research, 4(1), p.2.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J. and Thompson, J.D., and Higgins D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7(1), p.539.

Srivastava, S., Kamthania, M., Kumar Pandey, R., Kumar Saxena, A., Saxena, V., Kumar Singh, S., Kumar Sharma, R. and Sharma, N., 2019. Design of novel multi-epitope vaccines against severe acute respiratory syndrome validated through multistage molecular interaction and dynamics. Journal of Biomolecular Structure and Dynamics, 37(16), pp.4345-4360.

Srivastava, S., Kamthania, M., Singh, S., Saxena, A.K. and Sharma, N., 2018. Structural basis of development of multi-epitope vaccine against middle east respiratory syndrome using in silico approach. Infection and drug resistance, 11, p.2377.

1300 Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U.,
1301 Braxenthaler, M., Gallazzi, F., Protti, M.P., Sinigaglia, F. and Hammer, J.,
1302 1999. Generation of tissue-specific and promiscuous HLA ligand
1303 databases using DNA microarrays and virtual HLA class II matrices.
1304 Nature biotechnology, 17(6), p.555.
1305 Tenzer S, Peters B, Bulik S, Schoor O, Lemmel C, Schatz MM, Kloetzel PM,
1306 Rammensee HG, Schild H, Holzhutter HG. 2005. Modeling the MHC class
1307 I pathway by combining predictions of proteasomal cleavage, TAP
1308 transport and MHC class I binding. Cell Mol Life Sci 62:1025-1037.
1309 Totura, A.L., Whitmore, A., Agnihothram, S., Schäfer, A., Katze, M.G., Heise,
1310 M.T. and Baric, R.S., 2015. Toll-like receptor 3 signaling via TRIF
1311 contributes to a protective innate immune response to severe acute
1312 respiratory syndrome coronavirus infection. MBio, 6(3), pp.e00638-15.
1313 United nations conference on trade and development report 9th March 2020
1314 (https://unctad.org/en/pages/PressRelease.aspx?OriginalVersionID=548)
1315 Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M. and Peters, B.,
1316 2010. Peptide binding predictions for HLA DR, DP and DQ molecules.
1317 BMC bioinformatics, 11(1), p.568.
1318 Wang, Z. and Xu, J., 2013. Predicting protein contact map using evolutionary and
1319 physical constraints by integer programming. Bioinformatics, 29(13),
1320 pp.i266-i273.
1321 World Health Organization report, Coronavirus disease 2019 (COVID-19)
1322 Situation Report – 61, 21st March 2020
1323 (https://www.who.int/docs/default-source/coronaviruse/situation-
1324 reports/20200321-sitrep-61-covid-19.pdf?sfvrsn=6aa18912_2)
1325 Wu, X., Wu, S., Li, D., Zhang, J., Hou, L., Ma, J., Liu, W., Ren, D., Zhu, Y. and
1326 He, F., 2010. Computational identification of rare codons of Escherichia
1327 coli based on codon pairs preference. Bmc Bioinformatics, 11(1), p.61.
1328

SUPPLEMENTARY:

Structural basis to design multi-epitope vaccines against Novel Coronavirus 19 (COVID19) infection, the ongoing pandemic emergency: an in silico approach

Sukrit Srivastava*[a,b], Sonia Verma[c], Mohit Kamthania[d], Rupinder Kaur[e], Ruchi Kiran Badyal[f], Ajay Kumar Saxena[b], Ho-Joon Shin[g], Kailash Pandey[c], Michael Kolbe[h,i]

a Department of Biotechnology, Institute of Bio-Medical Education and Research, Mangalayatan University, 202146 Aligarh, India.
b Molecular Medicine Lab., School of Life Science, Jawaharlal Nehru University, 110067 New Delhi, India.
c National Institute of Malaria Research-Indian Council of Medical Research, New Delhi, India
d Department of Biotechnology, Institute of Applied Medicines and Research, 201206 Ghaziabad, India.
e Department of Chemistry, Guru Nanak Dev University, Amritsar, India
f Department of economics, Mangalayata University, India
g TIDCL, Ajou University, South Korea
h Department for Structural Infection Biology, Centre for Structural Systems Biology (CSSB) & Helmholtz-Centre for Infection Research, Notkestraße 85, 22607 Hamburg, Germany.
i Faculty of Mathematics, Informatics and Natural Sciences, University of Hamburg, Rothenbaumchaussee 19, 20148 Hamburg, Germany.

*Correspondence should be addressed to S.S.
(sukrit.srivastav@mangalayatan.edu.in; srivastav.sukrit@gmail.com)

Supplementary table S1. Homology modeling for HLA alleles. Tertiary structures of HLA alleles were modeled by homology modeling using SwissModel server. Templates were chosen with highest sequence identity. Generated models with acceptable QMEAN value were chosen for further studies.

S.No	HLA Class I Allele	Template used for modeling	% sequence identity	QMEAN
1	A*31:01	6o9b.1.A	94.29%	0.73
2	A*11:01	6o9b.1.A	97.50%	1.14
S.No	HLA Class II Allele	Template used for modeling	% sequence identity	QMEAN
3	DRB3*01:01	4is6.1.B	90.10%	(-)1.84
4	DRB1*09:01	1bx2.1.B	87.43%	(-)0.46

Supplementary table S2. INF-γ epitopes from CTL and HTL MEVs. INF-γ inducing (POSITIVE) epitopes from CTL and HTL MEVs were screened by using "Motif and SVM hybrid" (MERCI & SVM) approaches.

CLT Epitopes also predicted to be IFN-gamma epitopes					
S.No	Start-END	Sequence	Method	Result	Score
1	67-82	RWANQCIFGHSPRQQ	MERCI	POSITIVE	2
2	68-83	WANQCIFGHSPRQQR	MERCI	POSITIVE	2
3	69-84	ANQCIFGHSPRQQRE	MERCI	POSITIVE	2
4	70-85	NQCIFGHSPRQQREG	MERCI	POSITIVE	2
5	71-86	QCIFGHSPRQQREGV	MERCI	POSITIVE	2
6	83-98	EGVGENVYAYWSSVS	MERCI	POSITIVE	1
7	84-99	GVGENVYAYWSSVSV	MERCI	POSITIVE	2
8	120-135	LYENNPSNNMTWKVA	MERCI	POSITIVE	1
9	121-136	YENNPSNNMTWKVAG	MERCI	POSITIVE	1
10	122-137	ENNPSNNMTWKVAGQ	MERCI	POSITIVE	2
11	413-428	CTDDNALAYYGGGGS	MERCI	POSITIVE	1
12	428-443	TDDNALAYYGGGGSM	MERCI	POSITIVE	1
13	429-444	DDNALAYYGGGGSMY	MERCI	POSITIVE	1
14	439-454	GGGGSSIINNTVYTK	MERCI	POSITIVE	1
15	440-455	GGGSSIINNTVYTKG	MERCI	POSITIVE	2
16	441-456	GGSSIINNTVYTKGG	MERCI	POSITIVE	3
17	442-457	GSSIINNTVYTKGGG	MERCI	POSITIVE	3
18	443-458	SSIINNTVYTKGGGG	MERCI	POSITIVE	3
19	444-459	SIINNTVYTKGGGGS	MERCI	POSITIVE	3
20	445-460	IINNTVYTKGGGGSL	MERCI	POSITIVE	3
HLT Epitopes also predicted to be IFN-gamma epitopes					
S.No	Start-END	Sequence	Method	Result	Score
1	67-82	RWANQCIFGHSPRQQ	MERCI	POSITIVE	2
2	68-83	WANQCIFGHSPRQQR	MERCI	POSITIVE	2
3	69-84	ANQCIFGHSPRQQRE	MERCI	POSITIVE	2
4	70-85	NQCIFGHSPRQQREG	MERCI	POSITIVE	2
5	71-86	QCIFGHSPRQQREGV	MERCI	POSITIVE	2
6	83-98	EGVGENVYAYWSSVS	MERCI	POSITIVE	1
7	84-99	GVGENVYAYWSSVSV	MERCI	POSITIVE	2
8	120-135	LYENNPSNNMTWKVA	MERCI	POSITIVE	1
9	121-136	YENNPSNNMTWKVAG	MERCI	POSITIVE	1
10	122-137	ENNPSNNMTWKVAGQ	MERCI	POSITIVE	2
11	585-600	GGGGSFAWWTAFVTN	MERCI	POSITIVE	1
12	586-601	GGGSFAWWTAFVTNV	MERCI	POSITIVE	1
13	587-602	GGSFAWWTAFVTNVN	MERCI	POSITIVE	1
14	588-603	GSFAWWTAFVTNVNA	MERCI	POSITIVE	1
15	589-604	SFAWWTAFVTNVNAS	MERCI	POSITIVE	1
16	560-575	FAWWTAFVTNVNASS	MERCI	POSITIVE	1
17	561-576	AWWTAFVTNVNASSG	MERCI	POSITIVE	1
18	806-821	GGGSCTQHQPYVVDD	MERCI	POSITIVE	1
19	807-822	GGSCTQHQPYVVDDP	MERCI	POSITIVE	1
20	808-823	GSCTQHQPYVVDDPC	MERCI	POSITIVE	1

Supplementary table S3. Refinement models of CTL and HTL MEVs. Both the CTL and HTL MEVs models were refined by GalaxyWEB server. After refinement in particular Rama favored residues increased significantly.

Galaxy Refinement for CTL MEV						
Model	GDT-HA	RMSD	Mol Probity	Clash score	Poor rotamers	Rama favored
Initial	1.0000	0.000	3.749	100.5	4.5	75.8
MODEL 1	0.9371	0.459	2.539	23.2	0.4	83.6
Galaxy Refinement for HTL MEV						
Initial	1.0000	0.000	3.554	131.2	2.2	81.3
MODEL 1	0.9552	0.402	2.537	27.9	0.7	87.7

Supplementary table S4. B Cell linear epitopes of CTL MEVs. Linear B Cell epitopes predicted by ElliPro (IEDB) from CTL MEVs.

CTL Multi-Epitope vaccine model linear epitopes					
S.No.	Start	End	Peptide	Number of residues	Score
1	1	97	IVVAVTGYNCPGGKLTALERKKIVGQNNKYRSDLINGKLKNRNGTY MPRGKNMLELTWDCKLESSAQRWANQCIFGHSPRQQREGVGENVYAYWSSV	97	0.828
2	604	704	LRARSVSPKGGGGSFLAFLLFLVGGGGSHFYSKWYIRGGGGSWTAG AAAYYVGGGGSFPNITNLCPFGGGGSNYNYLYRLFRGGGGSNYLYRLFRHHHHHH	101	0.812
3	400	415	LNGSCGSVGGGGSCTD	16	0.693
4	265	277	SSEMVMCGGSLYG	13	0.678
5	422	438	GGGGSCTDDNALAYYGG	17	0.662
6	584	596	FGGGGSRARSVSP	13	0.619
7	219	231	GGGSKPRQKRTAT	13	0.618
8	490	502	WSMATYYGGGGSY	13	0.618
9	294	304	GGGSISNSWLM	11	0.607
10	362	372	FYGGGGSFLFV	11	0.599
11	380	393	GGGSRYFRLTLGVY	14	0.599
12	320	327	KGGGGSQE	8	0.55
13	445	456	GLPWNVVRGGGG	12	0.548
14	99	109	VEGLKKTAGTD	11	0.547
15	342	349	FNVPMEKG	8	0.534
16	521	524	ASFY	4	0.528
17	464	477	VYTKGGGGSLPVNV	14	0.511

Supplementary table S5. B Cell discontinuous epitopes of CTL MEVs. Discontinuous B Cell epitopes predicted by ElliPro (IEDB) from CTL MEVs.

CTL Multi-Epitope vaccine model discontinuous epitopes			
S.No.	Residues	Number of residues	Score
1	I1, V2, V3, A4, V5, T6, G7, Y8, N9, C10, P11, G12, G13, K14, L15, T16, A17, L18, E19, R20, K21, K22, I23, V24, G25, Q26, N27, N28, K29, Y30, R31, S32, D33, L34, I35, N36, G37, K38, L39, K40, N41, R42, N43, G44, T45, Y46, M47, P48, R49, G50, K51, N52, M53, L54, E55, L56, T57, W58, D59, C60, K61, L62, E63, S64, S65, A66, Q67, R68, W69, A70, N71, Q72, C73, I74, F75, G76, H77, S78, P79, R80, Q81, Q82, R83, E84, G85, V86, G87, E88, N89, V90, Y91, A92, Y93, W94, S95, S96, V97, S98, V99, E100, G101, L102, K103, K104, T105, G107, T108, D109, K112, W114, W115, N125, P126, S127, N128, N129, K133, V134, A135, G136	120	0.767
2	G160, A187, R188, G189, G219, G220, G221, S222, K223, P224, R225, Q226, R228, T229, A230, T231, G246, G247, G248, G249, S250, G264, S265, S266, E267, M268, V269, M270, C271, G272, G273, S274, L275, Y276, G277, G294, G295, G296, S297, I298, S299, N300, S301, W302, L303, M304, K320, G321, G322, G323, G324, S325, Q326, E327, I328, T341, F342, N343, V344, P345, M346, E347, K348, G349, Y363, G364, G365, G366, G367, S368, F369, L370, F371, V372, A374, G380, G381, G382, S383, R384, Y385, F386, R387, L388, T389, L390, G391, V392, L400, N401, G402, S403, C404, G405, S406, V407, G408, G409, G410, G411, S412, C413, T414, D415, G422, G423, G424, G425, S426, C427, T428, D429, D430, N431, A432, L433, A434, Y435, Y436, G445, L446, P447, W448, N449, V450, V451, R452, G453, G454, G455, G456, G469, G470, G471, S472, L473, P474, V475, N476, V477, W490, M492, A493, T494, Y495, Y496, G497, G498, G499, G500, S501, Y502, A521, S522, F523, Y524, I546, P547, Y548, N549, S550, K564, V565, S566, I567, W568, F584, G585, G586, G587, G588, S589, R590, A591, R592, S593, V594, S595, P596, G598, G599, G600, G601, Q603, L604, R605, A606, R607, S608, V609, S610, P611, K612, G613, G614, G615, G616, S617, F618, L619, A620, F621, L622, L623, F624, L625, V626, G627, G628, G629, G630, S631, H632, F633, Y634, S635, K636, W637, Y638, I639, R640, G641, G642, G643, G644, S645, W646, T647, A648, G649, A650, A651, A652, Y653, Y654, V655, G656, G657, G658, G659, S660, F661, P662, N663, I664, T665, N666, L667, C668, P669, F670, G671, G672, G673, G674, S675, N676, Y677, N678, Y679, L680, Y681, R682, L683, F684, G686, G687, G688, G689, S690, N691, Y692, L693, Y694, R695, R698, H699, H700, H701, H702, H703, H704	282	0.664

Supplementary table S6. B Cell linear epitopes of HTL MEVs. Linear B Cell epitopes predicted by ElliPro (IEDB) from HTL MEVs.

HTL Multi-Epitope vaccine model linear epitopes					
S.No.	Start	End	Peptide	Number of residues	Score
1	691	810	INVFAFPFTIYSLLGGGGSKTQSLLIVNNATNVVGGGGSLLIVNNATNVVIK VCGGGGSQSLLIVNNATNVVIKGGGGSSLLIVNNATNVVIKVGGGGSTQ SLLIVNNATNVVIHHHHHH	120	0.831
2	1	138	IVVAVTGYNCPGGKLTALERKKIVGQNNKYRSDLINGKLKNRNGTYMPR GKNMLELTWDCKLESSAQRWANQCIFGHSPRQQREGVGENVYAYWSS VSVEGLKKTAGTDAGKSWWSKLPKLYENNPSNNMTWKVAGQG	138	0.777
3	452	468	VSIWNLDYIINLIGGGG	17	0.713
4	323	333	FVGGGGSESPF	11	0.69
5	474	489	WNLDYIINLIIGGGGS	16	0.685
6	434	444	TVYSHLLLVAA	11	0.679
7	543	551	LYGGGGSCF	9	0.651
8	416	424	TIPIQASLP	9	0.63
9	278	287	SAFFGMSGGG	10	0.608
10	351	360	ILASFSASTS	10	0.597
11	496	508	WNLDYIINLGGGG	13	0.586
12	399	404	PAQYEL	6	0.572
13	378	383	SAPPAQ	6	0.565
14	673	684	FAFPFTIYSLLL	12	0.556
15	515	527	ALITLATCELGGG	13	0.544
16	570	573	LCFL	4	0.54
17	655	660	DDPCPI	6	0.518

Supplementary table S7. B Cell discontinuous epitopes of HTL MEVs. Discontinuous B Cell epitopes predicted by ElliPro (IEDB) from HTL MEVs.

HTL Multi-Epitope vaccine model discontinuous epitopes			
S. No.	Residues	Number of residues	Score
1	C639, P640, I641, D655, D656, P657, C658, P659, I660, H661, F662, F673, A674, F675, P676, F677, T678, I679, Y680, S681, L682, L683, L684, G687, G688, I691, N692, V693, F694, A695, F696, P697, F698, T699, I700, Y701, S702, L703, L704, G705, G706, G707, G708, S709, K710, T711, Q712, S713, L714, L715, I716, V717, N718, N719, A720, T721, N722, V723, V724, G725, G726, G727, G728, S729, L730, L731, I732, V733, N734, N735, A736, T737, N738, V739, V740, I741, K742, V743, C744, G745, G746, G747, G748, S749, Q750, S751, L752, L753, I754, V755, N756, N757, A758, T759, N760, V761, V762, I763, K764, G765, G766, G767, G768, S769, S770, L771, L772, I773, V774, N775, N776, A777, T778, N779, V780, V781, I782, K783, V784, G785, G786, G787, G788, S789, T790, Q791, S792, L793, L794, I795, V796, N797, N798, A799, T800, N801, V802, V803, I804, H805, H806, H807, H808, H809, H810	145	0.776
2	I1, V2, V3, A4, V5, T6, G7, Y8, N9, C10, P11, G12, G13, K14, L15, T16, A17, L18, E19, R20, K21, K22, I23, V24, G25, Q26, N27, N28, K29, Y30, R31, S32, D33, L34, I35, N36, G37, K38, L39, K40, N41, R42, N43, G44, T45, Y46, M47, P48, R49, G50, K51, N52, M53, L54, E55, L56, T57, W58, D59, C60, K61, L62, E63, S64, S65, A66, Q67, R68, W69, A70, N71, Q72, C73, I74, F75, G76, H77, S78, P79, R80, Q81, Q82, R83, E84, G85, V86, G87, E88, N89, V90, Y91, A92, Y93, W94, S95, S96, V97, S98, V99, E100, G101, L102, K103, K104, T105, A106, G107, T108, D109, A110, G111, K112, S113, W114, W115, S116, K117, L118, P119, K120, L121, Y122, E123, N124, N125, P126, S127, N128, N129, M130, T131, W132, K133, V134, A135, G136, Q137, G138, E145, A146, A147, A148, L150, F152, A154, V156, L163, A164, G165, G168, S169	151	0.748
3	S278, A279, F280, F281, G282, M283, S284, G285, G286, G287, F303, G304, G305, A322, F323, V324, G325, G326, G327, G328, S329, E330, S331, P332, F333, I350, I351, L352, A353, S354, F355, S356, A357, S358, T359, S360, S378, A379, P380, P381, A382, Q383, P399, A400, Q401, Y402, E403, L404, G405, G406, G407, T416, I417, P418, I419, Q420, A421, S422, L423, P424, T434, V435, Y436, S437, H438, L439, L440, L441, V442, A444, G446, V452, S453, I454, W455, N456, L457, D458, Y459, I460, I461, N462, L463, I464, G465, G466, G467, G468, W474, N475, L476, D477, Y478, I479, I480, N481, L482, I483, I484, G485, G486, G487, G488, S489, I495, W496, N497, L498, D499, Y500, I501, I502, N503, L504, G505, G506, G507, G508, A515, L516, I517, T518, A520, T521, C522, E523, L538, A539, L543, Y544, G545, G546, G547, G548, S549, C550, F551, L570, C571, F572, L573, F596, L597, L598, F599, P615, Y616, V617, V618, D619	150	0.609
4	L524, G525, G526, G527, G528	5	0.53

Supplementary table S8. Shortlisted high scoring CTL epitopes (MHC-I Binding Predictions). Selected high scoring CTL epitopes and their respective HLA alleles binders predicted by "MHC-I Binding Predictions" IEDB tool. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy and high immunogenicity.

SARS-CoV-2 CTL epitopes (MHC-I Binding Predictions)											
S.No.	Proteins	Peptide	Allele	Start	End	Length	Immunogenicity	Conservancy	Toxicity	Method used	Percentile Rank
1	E Protein	FLAFVVFLL	A*02:01	20	28	9	0.30188	100.00% (35/35)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
2	E Protein	FLAFVVFLL	A*02:03	20	28	9	0.30188	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.25
3	E Protein	FLAFVVFLLV	A*02:01	20	29	10	0.30526	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.15
4	E Protein	FLAFVVFLLV	A*02:03	20	29	10	0.30526	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.23
5	E Protein	LAFVVFLLV	B*51:01	21	29	9	0.2141	100.00% (35/35)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
6	E Protein	LVKPSFYVY	B*15:01	51	59	9	-0.11106	100.00% (35/35)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
7	E Protein	RVKNLNSSR	A*31:01	61	69	9	-0.32968	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.16
8	E Protein	SLVKPSFYVY	B*15:01	50	59	10	-0.2443	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.28
9	E Protein	SVLLFLAFV	A*02:06	16	24	9	0.19022	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.33
10	E Protein	TLAILTALR	A*68:01	30	38	9	0.1989	97.14% (34/35)	Non-Toxin	Consensus (ann/smm)	0.21
11	E Protein	VLLFLAFVV	A*02:01	17	25	9	0.26315	100.00% (35/35)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.3
12	E Protein	VSLVKPSFY	A*30:02	49	57	9	-0.25372	100.00% (35/35)	Non-Toxin	Consensus (ann/smm)	0.33
13	M Protein	FIASFRLFAR	A*68:01	26	35	10	0.12185	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.12
14	M Protein	LIFLWLLWPV	A*02:06	51	60	10	0.40176	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.08
15	M Protein	LSYFIASFR	A*68:01	23	31	9	0.21181	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.11
16	M Protein	MWLSYFIASF	A*23:01	21	30	10	0.00197	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.11
17	M Protein	MWLSYFIASFR	A*68:01	21	31	11	0.03554	100.00% (41/41)	Non-Toxin	ann	0.09
18	M Protein	MWLSYFIASFR	A*33:01	21	31	11	0.03554	100.00% (41/41)	Non-Toxin	ann	0.1
19	M Protein	QWNLVIGFLF	A*23:01	19	28	10	0.28076	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.12
20	M Protein	RFLYIIKLIF	A*23:01	44	53	10	0.11728	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.12
21	M Protein	RTRSMWSF	A*32:01	35	42	8	-0.25178	100.00% (41/41)	Non-Toxin	ann	0.1
22	N Protein	AQFAPSASAF	B*15:01	25	34	10	-0.17446	100.00% (40/40)	Non-Toxin	Consensus (ann/smm)	0.16
23	N Protein	AQFAPSASAFF	B*15:01	25	35	11	-0.11074	100.00% (40/40)	Non-Toxin	ann	0.12
24	N Protein	FTALTQHGK	A*68:01	53	61	9	-0.0226	100.00% (40/40)	Non-Toxin	Consensus (ann/smm)	0.18
25	N Protein	KMKDLSPR	A*31:01	30	37	8	-0.22357	100.00% (40/40)	Non-Toxin	ann	0.15
26	N Protein	KTFPPTEPK	A*11:01	11	19	9	0.1306	100.00% (40/40)	Non-Toxin	Consensus (ann/smm)	0.11
27	N Protein	KTFPPTEPK	A*30:01	11	19	9	0.1306	100.00% (40/40)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
28	N Protein	RQKRTATKAY	B*15:01	49	58	10	-0.06462	100.00% (40/40)	Non-Toxin	Consensus (ann/smm)	0.2
29	N Protein	SSPDDQIGYY	A*01:01	8	17	10	0.07924	100.00% (40/40)	Non-Toxin	Consensus (ann/smm)	0.2
30	ORF-1ab	ASFYYVWKSY	A*30:02	8	17	10	0.00073	100.00% (36/36)	Non-Toxin	Consensus (ann/smm)	0.07
31	ORF-1ab	CLAYYFMRFRR	A*33:01	49	59	11	0.12614	100.00% (36/36)	Non-Toxin	ann	0.07
32	ORF-1ab	ETISLAGSY	A*26:01	33	41	9	-0.1653	100.00% (36/36)	Non-Toxin	Consensus (ann/smm)	0.1
33	ORF-1ab	IPLMYKGLPW	B*53:01	55	64	10	-0.37784	94.44% (34/36)	Non-Toxin	Consensus (ann/smm)	0.07
34	ORF-1ab	IPLMYKGLPW	B*53:01	55	64	10	-0.37784	94.44% (34/36)	Non-Toxin	Consensus (ann/smm)	0.07
35	ORF-1ab	KMNYQVNGY	A*30:02	36	44	9	-0.06542	97.22% (35/36)	Non-Toxin	Consensus (ann/smm)	0.07
36	ORF-1ab	KMNYQVNGY	A*30:02	36	44	9	-0.06542	97.22% (35/36)	Non-Toxin	Consensus (ann/smm)	0.07
37	ORF-1ab	STNVTIATY	A*30:02	33	41	9	0.25822	100.00% (36/36)	Non-Toxin	Consensus (ann/smm)	0.09

38	ORF-1ab	YAYLRKHFSM	B*08:01	28	37	10	-0.13937	97.22% (35/36)	Non-Toxin	Consensus (ann/smm)	0.07
39	ORF3a	FMRIFTIGTV	A*02:03	4	13	10	0.47908	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.17
40	ORF3a	FTIGTVTLK	A*68:01	8	16	9	0.18024	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.16
41	ORF3a	HVTFFIYNK	A*68:01	17	25	9	0.36278	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.16
42	ORF3a	IIMRLWLCWK	A*03:01	53	62	10	0.27346	96.43% (27/28)	Non-Toxin	Consensus (ann/smm)	0.13
43	ORF3a	VEHVTFFIY	B*44:03	15	23	9	0.3766	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.18
44	ORF3a	YFLQSINFVR	A*33:01	43	52	10	-0.03483	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.2
45	ORF6	AEILLIIM	B*40:01	12	19	8	0.25884	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.41
46	ORF6	FQVTIAEIL	A*02:06	7	15	9	0.38115	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.52
47	ORF6	ILLIIMRTFK	A*03:01	14	23	10	0.2388	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.17
48	ORF6	IMRTFKVSI	A*32:01	18	26	9	-0.09496	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.44
49	ORF6	KVSIWNLDY	A*30:02	23	31	9	0.29343	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.41
50	ORF6	LLIIMRTFK	A*03:01	15	23	9	0.156	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.36
51	ORF6	LSKSLTENKY	A*30:02	40	49	10	-0.24668	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.48
52	ORF6	SIWNLDYII	A*32:01	25	33	9	0.15011	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
53	ORF6	VTIAEILLI	B*58:01	9	17	9	0.28951	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.34
54	ORF7a	CVRGTTVLLK	A*03:01	23	32	10	0.14952	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.17
55	ORF7a	EVQELYSPI	A*26:01	22	30	9	-0.09723	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.21
56	ORF7a	EVQELYSPIF	A*26:01	22	31	10	-0.03858	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.22
57	ORF7a	FALTCFSTQF	B*53:01	54	63	10	-0.09369	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.22
58	ORF7a	FLIVAAIVF	B*15:01	31	39	9	0.29611	100.00% (28/28)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
59	ORF7a	FLIVAAIVFI	A*02:01	31	40	10	0.38946	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.21
60	ORF7a	ITLCFTLKRK	A*03:01	40	49	10	-0.06825	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.17
61	ORF7a	QELYSPIFLI	B*44:02	24	33	10	0.03838	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.22
62	ORF7a	QELYSPIFLI	B*44:03	24	33	10	0.03838	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.25
63	ORF7a	RSVSPKLFI	B*58:01	10	18	9	-0.30783	100.00% (28/28)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
64	ORF7b	DFYLCFLAF	A*23:01	8	16	9	0.05884	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.29
65	ORF7b	FLAFLLFLVL	A*02:01	13	22	10	0.23386	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.26
66	ORF7b	FLLFLVLIML	A*02:01	16	25	10	0.14288	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.3
67	ORF7b	FYLCFLAFLL	A*23:01	9	18	10	0.1745	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.13
68	ORF7b	FYLCFLAFLL	A*24:02	9	18	10	0.1745	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.28
69	ORF7b	IELSLIDFY	B*44:03	2	10	9	0.03153	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.19
70	ORF7b	LVLIMLIIFW	B*53:01	20	29	10	0.25452	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.32
71	ORF7b	MIELSLIDFY	A*01:01	1	10	10	0.06184	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.35
72	ORF7b	MLIIFWFSL	A*02:01	24	32	9	0.50177	100.00% (2/2)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
73	ORF7b	SLIDFYLCFL	A*02:01	5	14	10	0.18838	100.00% (2/2)	Non-Toxin	Consensus (ann/smm)	0.14
74	ORF7b	YLCFLAFLL	A*02:01	10	18	9	0.21865	100.00% (2/2)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.3
75	ORF8	CPIHFYSKW	B*53:01	37	45	9	-0.07935	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.2
76	ORF8	CPIHFYSKWY	B*53:01	37	46	10	-0.02216	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.2
77	ORF8	CSFYEDFLEY	A*01:01	32	41	10	0.31272	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.24
78	ORF8	DFLEYHDVR	A*33:01	37	45	9	0.16684	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.13
79	ORF8	FTINCQEPK	A*68:01	16	24	9	-0.04683	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.18
80	ORF8	GARKSAPLI	A*30:01	50	58	9	-0.34031	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.3
81	ORF8	GIITTVAAF	B*15:01	8	16	9	0.2148	100.00% (34/34)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.3
82	ORF8	GSLVVRCSFY	A*30:02	26	35	10	0.00657	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.24

83	ORF8	IQYIDIGNY	A*30:02	1	9	9	0.30442	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.12
84	ORF8	LGIITTVAAF	B*15:01	7	16	10	0.34746	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.31
85	ORF8	LQSCTQHQPY	B*15:01	22	31	10	-0.25674	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.17
86	ORF8	NYTVSCLPF	A*23:01	8	16	9	-0.17355	58.82% (20/34)	Non-Toxin	Consensus (ann/smm)	0.23
87	ORF8	QSCTQHQPY	A*01:01	23	31	9	-0.16503	100.00% (34/34)	Non-Toxin	Consensus (ann/smm)	0.28
88	ORF10	FAFPFTIYSL	A*02:06	7	16	10	0.20414	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.48
89	ORF10	FPFTIYSL	B*51:01	9	16	8	0.06356	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.39
90	ORF10	FPFTIYSLL	B*51:01	9	17	9	0.05708	100.00% (28/28)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.3
91	ORF10	FPFTIYSLLL	B*53:01	9	18	10	0.03149	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.13
92	ORF10	FPFTIYSLLL	B*51:01	9	18	10	0.03149	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.14
93	ORF10	FTIYSLLLCR	A*68:01	11	20	10	-0.18372	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.12
94	ORF10	FTIYSLLLCR	A*33:01	11	20	10	-0.18372	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.35
95	ORF10	MGYINVFAFPF	A*23:01	1	11	11	0.40977	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.41
96	ORF10	TIYSLLLCR	A*68:01	12	20	9	-0.22977	100.00% (28/28)	Non-Toxin	Consensus (ann/smm)	0.47
97	ORF10	YINVFAFPF	B*53:01	3	11	9	0.28259	100.00% (28/28)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.35
98	ORF10	YINVFAFPF	A*32:01	3	11	9	0.28259	100.00% (28/28)	Non-Toxin	Consensus (ann/comblib_sidney2008/smm)	0.4
99	S Protein	FVFLVLLPLV	A*02:06	2	11	10	0.02996	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.1
100	S Protein	GNYNYLYRLFR	A*33:01	27	37	11	0.08205	100.00% (41/41)	Non-Toxin	ann	0.11
101	S Protein	KSFTVEKGIY	A*30:02	24	33	10	0.11812	100.00% (41/41)	Non-Toxin	Consensus (ann/smm)	0.11

Supplementary table S9. Shortlisted high scoring CTL epitopes (MHC-I Processing Predictions). Selected high scoring CTL epitopes and their respective HLA alleles binders predicted by "MHC-I Processing Predictions" IEDB tool. The screening gives detailed and combined scoring "Total score" for Proteasomal cleavage/TAP transport/MHC class I combined. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy with high immunogenicity.

SARS-CoV-2 CTL epitopes (MHC-I Processing Predictions)															
S.No.	Proteins	Allele	Start	End	Peptide Length	Peptide	Immunogenicity	Conservancy	Toxicity	Proteasome Score	TAP Score	MHC Score	Processing Score	Total Score	MHC IC50[nM]
1	E Protein	B*15:01	50	59	10	SLVKPSFYVY	-0.2443	100.00% (35/35)	Non-Toxin	1.51	1.36	-1.39	2.86	1.47	24.6
2	E Protein	B*15:01	51	59	9	LVKPSFYVY	-0.11106	100.00% (35/35)	Non-Toxin	1.51	1.35	-1.42	2.86	1.44	26.3
3	E Protein	A*30:02	51	59	9	LVKPSFYVY	-0.11106	100.00% (35/35)	Non-Toxin	1.51	1.35	-1.61	2.86	1.25	40.4
4	E Protein	A*02:01	20	28	9	FLAFVVFLL	0.30188	100.00% (35/35)	Non-Toxin	1.45	0.41	-0.81	1.86	1.05	6.5
5	E Protein	B*15:01	34	42	9	LTALRLCAY	0.01886	97.14% (34/35)	Non-Toxin	1.42	1.27	-1.69	2.69	1	49.5
6	M Protein	A*30:02	31	39	9	ATSRTLSYY	-0.11604	100.00% (41/41)	Non-Toxin	1.26	1.34	-1.12	2.6	1.48	13.3
7	M Protein	A*11:01	31	39	9	ATSRTLSYY	-0.11604	100.00% (41/41)	Non-Toxin	1.26	1.34	-1.52	2.6	1.08	32.9
8	M Protein	B*08:01	33	42	10	FARTRSMWSF	-0.12986	100.00% (41/41)	Non-Toxin	1.41	1.12	-1.34	2.53	1.19	22.1
9	M Protein	A*02:01	53	62	10	FLWLLWPVTL	0.31272	100.00% (41/41)	Non-Toxin	1.85	0.46	-1.16	2.31	1.16	14.3
10	M Protein	A*68:01	23	31	9	LSYFIASFR	0.21181	100.00% (41/41)	Non-Toxin	0.84	0.72	-0.46	1.56	1.1	2.9
11	M Protein	A*23:01	23	33	11	LSYFIASFRLF	0.2706	100.00% (41/41)	Non-Toxin	1.25	1.18	-1.36	2.43	1.07	23
12	M Protein	A*23:01	21	30	10	MWLSYFIASF	0.00197	100.00% (41/41)	Non-Toxin	1.38	1.26	-0.91	2.63	1.73	8.1
13	M Protein	A*24:02	21	30	10	MWLSYFIASF	0.00197	100.00% (41/41)	Non-Toxin	1.38	1.26	-1.27	2.63	1.36	18.7
14	M Protein	B*15:01	21	30	10	MWLSYFIASF	0.00197	100.00% (41/41)	Non-Toxin	1.38	1.26	-1.33	2.63	1.3	21.6

15	M Protein	A*23:01	44	53	10	RFLYIIKLIF	0.11728	100.00% (41/41)	Non-Toxin	1.25	1.35	-1.33	2.6	1.27	21.3
16	M Protein	A*30:01	35	42	8	RTRSMWSF	-0.25178	100.00% (41/41)	Non-Toxin	1.41	1.28	-1.51	2.69	1.18	32.7
17	M Protein	A*30:02	51	59	9	SGFAAYSRY	0.00261	100.00% (41/41)	Non-Toxin	1.53	1.17	-1.4	2.69	1.3	25
18	M Protein	B*15:01	44	53	10	SQRVAGDSGF	0.0305	100.00% (41/41)	Non-Toxin	1.33	1.24	-1.35	2.58	1.23	22.4
19	M Protein	A*23:01	24	32	9	SYFIASFRL	0.18333	100.00% (41/41)	Non-Toxin	1.45	0.62	-0.99	2.08	1.09	9.8
20	M Protein	B*35:01	30	38	9	VATSRTLSY	-0.17295	100.00% (41/41)	Non-Toxin	1.34	1.31	-1.38	2.65	1.27	23.9
21	M Protein	A*30:02	56	64	9	YSRYRIGNY	0.21358	100.00% (41/41)	Non-Toxin	1.36	1.37	-1.61	2.73	1.12	40.6
22	N Protein	B*15:01	25	34	10	AQFAPSASAF	-0.17446	100.00% (40/40)	Non-Toxin	1.23	1.25	-0.7	2.48	1.78	5
23	N Protein	B*15:01	25	35	11	AQFAPSASAFF	-0.11074	100.00% (40/40)	Non-Toxin	1.3	1.25	-1.46	2.55	1.08	29
24	N Protein	A*30:02	32	41	10	KDLSPRWYFY	0.14332	100.00% (40/40)	Non-Toxin	1.58	1.21	-1.68	2.79	1.1	48.3
25	N Protein	A*30:02	30	39	10	KMKDLSPRWY	-0.05692	100.00% (40/40)	Non-Toxin	1.31	1.35	-1.56	2.66	1.1	36.2
26	N Protein	A*68:02	48	56	9	NTASWFTAL	0.22775	100.00% (40/40)	Non-Toxin	1.42	0.49	-0.79	1.91	1.12	6.2
27	N Protein	B*15:01	49	58	10	RQKRTATKAY	-0.06462	100.00% (40/40)	Non-Toxin	1.51	1.34	-1.55	2.85	1.3	35.5
28	N Protein	A*01:01	8	17	10	SSPDDQIGYY	0.07924	100.00% (40/40)	Non-Toxin	1.24	1.36	-1.16	2.6	1.45	14.3
29	ORF1ab	A*01:01	32	41	10	ACTDDNALAY	0.10055	100.00% (36/36)	Non-Toxin	1.49	1.33	-0.82	2.82	2	6.6
30	ORF1ab	A*01:01	33	41	9	CTDDNALAY	0.07355	100.00% (36/36)	Non-Toxin	1.49	1.23	-0.45	2.72	2.27	2.8
31	ORF1ab	A*26:01	33	41	9	ETISLAGSY	-0.1653	100.00% (36/36)	Non-Toxin	1.4	1.21	-0.53	2.61	2.08	3.4
32	ORF1ab	B*35:01	2	10	9	FAVDAAKAY	-0.04849	100.00% (36/36)	Non-Toxin	1.73	1.35	-0.46	3.08	2.62	2.9
33	ORF1ab	B*15:01	53	61	9	ILMTARTVY	0.12576	100.00% (36/36)	Non-Toxin	1.65	1.34	-1.1	2.99	1.89	12.5
34	ORF1ab	B*15:01	52	61	10	LILMTARTVY	0.0012	100.00% (36/36)	Non-Toxin	1.65	1.36	-1.23	3.01	1.78	16.9
35	ORF1ab	B*15:01	13	22	10	LMSNLGMPSY	-0.30933	100.00% (36/36)	Non-Toxin	1.55	1.34	-1	2.9	1.89	10.1
36	ORF1ab	B*35:01	1	9	9	LPSLATVAY	0.06748	100.00% (36/36)	Non-Toxin	1.47	1.15	-0.34	2.62	2.28	2.2
37	ORF1ab	A*32:01	2	10	9	RMYIFFASF	0.29328	100.00% (36/36)	Non-Toxin	1.32	1.36	-0.82	2.68	1.86	6.6
38	ORF1ab	A*30:02	2	11	10	RMYIFFASFY	0.32633	100.00% (36/36)	Non-Toxin	1.28	1.53	-0.97	2.81	1.83	9.4
39	ORF1ab	A*03:01	2	11	10	RMYIFFASFY	0.32633	100.00% (36/36)	Non-Toxin	1.28	1.53	-1	2.81	1.8	10.1
40	ORF1ab	B*44:03	26	34	9	SEFSSLPSY	-0.40603	100.00% (36/36)	Non-Toxin	1.49	1.35	-1.03	2.84	1.82	10.6
41	ORF1ab	B*15:01	18	26	9	VMYMGTLSY	-0.21438	100.00% (36/36)	Non-Toxin	1.3	1.5	-0.8	2.79	1.99	6.3
42	ORF1ab	A*03:01	18	26	9	VMYMGTLSY	-0.21438	100.00% (36/36)	Non-Toxin	1.3	1.5	-0.98	2.79	1.82	9.5
43	ORF3a	B*44:03	29	39	11	AGLEAPFLYLY	0.21841	100.00% (28/28)	Non-Toxin	1.51	1.29	-1.63	2.8	1.17	42.9
44	ORF3a	A*23:01	41	50	10	LVYFLQSINF	-0.05419	100.00% (28/28)	Non-Toxin	1.3	1.31	-1.39	2.61	1.23	24.3
45	ORF6	A*30:02	23	31	9	KVSIWNLDY	0.29343	100.00% (34/34)	Non-Toxin	1.21	1.33	-1.46	2.54	1.08	28.8
46	ORF7a	B*35:01	67	75	9	CPDGVKHVY	-0.07008	100.00% (28/28)	Non-Toxin	1.43	1.11	-1.34	2.54	1.2	22
47	ORF7b	A*02:01	26	34	9	IIFWFSLEL	0.2683	100.00% (2/2)	Non-Toxin	1.65	0.55	-1.2	2.21	1.01	15.7
48	ORF7b	A*02:01	5	14	10	SLIDFYLCFL	0.18838	100.00% (2/2)	Non-Toxin	1.7	0.5	-0.74	2.2	1.46	5.5
49	ORF7b	A*02:03	5	14	10	SLIDFYLCFL	0.18838	100.00% (2/2)	Non-Toxin	1.7	0.5	-0.79	2.2	1.41	6.2
50	ORF7b	A*02:06	5	14	10	SLIDFYLCFL	0.18838	100.00% (2/2)	Non-Toxin	1.7	0.5	-0.93	2.2	1.27	8.6
51	ORF8	B*15:01	8	16	9	GIITTVAAF	0.2148	100.00% (34/34)	Non-Toxin	1.27	1.13	-1.26	2.4	1.14	18.1
52	ORF8	A*33:01	40	48	9	HFYSKWYIR	-0.09452	100.00% (34/34)	Non-Toxin	1.08	0.76	-0.68	1.84	1.16	4.8
53	ORF8	A*31:01	40	48	9	HFYSKWYIR	-0.09452	100.00% (34/34)	Non-Toxin	1.08	0.76	-0.79	1.84	1.05	6.2
54	ORF8	B*15:01	7	16	10	LGIITTVAAF	0.34746	100.00% (34/34)	Non-Toxin	1.27	0.97	-1.1	2.24	1.14	12.7
55	ORF10	A*68:01	5	14	10	NVFAFPFTIY	0.40129	100.00% (28/28)	Non-Toxin	1.47	1.4	-1.84	2.86	1.02	69.7
56	S Protein	B*35:01	11	19	9	CVADYSVLY	-0.09595	100.00% (41/41)	Non-Toxin	1.51	1.38	-1.39	2.89	1.5	24.4
57	S Protein	B*35:01	58	66	9	FAMQMAYRF	-0.28061	100.00% (41/41)	Non-Toxin	1.45	1.05	-0.8	2.5	1.7	6.3
58	S Protein	B*58:01	44	53	10	KRSFIEDLLF	0.29624	100.00% (41/41)	Non-Toxin	1.23	1.26	-1.11	2.49	1.37	12.9
59	S Protein	A*23:01	21	30	10	KWPWYIWLGF	0.56424	100.00% (41/41)	Non-Toxin	1.26	1.25	-1.04	2.51	1.47	11
60	S Protein	B*35:01	54	64	11	LQIPFAMQMAY	-0.22124	100.00% (41/41)	Non-Toxin	1.42	1.4	-1.25	2.82	1.57	17.8

61	S Protein	A*01:01	25	33	9	LTDEMIAQY	0.02757	100.00% (41/41)	Non-Toxin	1.21	1.21	-0.72	2.42	1.71	5.2
62	S Protein	B*35:01	55	64	10	QIPFAMQMAY	-0.25308	100.00% (41/41)	Non-Toxin	1.42	1.35	-1.16	2.77	1.61	14.5
63	S Protein	A*30:02	7	15	9	RISNCVADY	-0.02787	100.00% (41/41)	Non-Toxin	1.16	1.47	-1.26	2.63	1.37	18.2
64	S Protein	B*58:01	45	53	9	RSFIEDLLF	0.27446	100.00% (41/41)	Non-Toxin	1.23	1.32	-0.72	2.54	1.82	5.3
65	S Protein	B*15:01	4	13	10	RVYSTGSNVF	-0.23394	100.00% (41/41)	Non-Toxin	1.51	1.32	-1.02	2.83	1.81	10.5
66	S Protein	B*35:01	22	30	9	SANNCTFEY	0.13273	100.00% (41/41)	Non-Toxin	1.18	1.3	-1.11	2.48	1.37	12.8
67	S Protein	A*30:02	28	37	10	YTNSFTRGVY	0.08467	95.12% (39/41)	Non-Toxin	1.36	1.28	-1.26	2.64	1.38	18.2

Supplementary table S10. Shortlisted high scoring HTL epitopes (MHC-II Binding Predictions). Selected high "Percentile rank" HTL epitopes with their respective HLA class II alleles binders predicted by the "**MHC-II Binding Predictions**" tool of IEDB are listed. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant conservancy.

SARS-CoV-2 HTL epitopes (MHC-II Binding Predictions)									
S.No	Proteins	Allele	Start	End	Peptide	Conservancy	Toxicity	Method used	Percentile Rank
1	E Protein	DPA1*03:01/DPB1*04:02	20	34	FLAFVVFLLVTLAIL	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
2	E Protein	DPA1*03:01/DPB1*04:02	19	33	LFLAFVVFLLVTLAI	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
3	E Protein	DPA1*01:03/DPB1*02:01	19	33	LFLAFVVFLLVTLAI	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
4	E Protein	DPA1*01/DPB1*04:01	19	33	LFLAFVVFLLVTLAI	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm)	0.08
5	E Protein	DPA1*02:01/DPB1*01:01	19	33	LFLAFVVFLLVTLAI	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
6	E Protein	DPA1*01:03/DPB1*02:01	15	29	NSVLLFLAFVVFLLV	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
7	E Protein	DPA1*03:01/DPB1*04:02	15	29	NSVLLFLAFVVFLLV	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
8	E Protein	DPA1*01/DPB1*04:01	15	29	NSVLLFLAFVVFLLV	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm)	0.07
9	E Protein	DPA1*02:01/DPB1*01:01	15	29	NSVLLFLAFVVFLLV	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
10	E Protein	DPA1*02:01/DPB1*01:01	14	28	VNSVLLFLAFVVFLL	100.00% (35/35)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
11	M Protein	DQA1*01:01/DQB1*05:01	49	63	IKLIFLWLLWPVTLA	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
12	M Protein	DQA1*01:01/DQB1*05:01	50	64	KLIFLWLLWPVTLAC	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.11
13	M Protein	DRB1*09:01	34	48	RTLSYYKLGASQRVA	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
14	M Protein	DRB1*01:01	34	48	RTLSYYKLGASQRVA	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.67
15	M Protein	DRB1*09:01	33	47	SRTLSYYKLGASQRV	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
16	M Protein	DRB1*09:01	37	51	SYYKLGASQRVAGDS	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.48
17	M Protein	DRB1*09:01	35	49	TLSYYKLGASQRVAG	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
18	M Protein	DRB1*01:01	35	49	TLSYYKLGASQRVAG	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.67
19	M Protein	DPA1*01:03/DPB1*02:01	18	32	VGLMWLSYFIASFRL	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
20	N Protein	DQA1*01:02/DQB1*06:02	12	26	ANNAAIVLQLPQGTT	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.2
21	N Protein	DRB1*11:01	12	26	DQIGYYRRATRRIRG	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.42
22	N Protein	DRB5*01:01	12	26	DQIGYYRRATRRIRG	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.58
23	N Protein	DQA1*01:02/DQB1*06:02	7	21	GTRNPANNAAIVLQL	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
24	N Protein	DRB1*07:01	48	62	GTWLTYTGAIKLDDK	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.58
25	N Protein	DRB1*11:01	15	29	GYYRRATRRIRGGDG	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.42
26	N Protein	DRB1*11:01	14	28	IGYYRRATRRIRGGD	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.42
27	N Protein	DQA1*01:02/DQB1*06:02	13	27	NNAAIVLQLPQGTTL	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.42
28	N Protein	DQA1*01:02/DQB1*06:02	10	24	NPANNAAIVLQLPQG	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03

29	N Protein	DQA1*01:02/DQB1*06:02	11	25	PANNAAIVLQLPQGT	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
30	N Protein	DRB1*09:01	23	37	QIAQFAPSASAFFGM	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.01
31	N Protein	DRB1*11:01	13	27	QIGYYRRATRRIRGG	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.39
32	N Protein	DRB5*01:01	13	27	QIGYYRRATRRIRGG	100.00% (40/40)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.58
33	N Protein	DQA1*01:02/DQB1*06:02	9	23	RNPANNAAIVLQLPQ	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
34	N Protein	DRB1*07:01	47	61	SGTWLTYTGAIKLDD	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.58
35	N Protein	DRB1*07:01	45	59	TPSGTWLTYTGAIKL	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.58
36	N Protein	DQA1*01:02/DQB1*06:02	8	22	TRNPANNAAIVLQLP	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
37	N Protein	DRB1*07:01	49	63	TWLTYTGAIKLDDKD	100.00% (40/40)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.58
38	ORF1ab	DPA1*01:03/DPB1*02:01	34	48	AAIMQLFFSYFAVHF	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
39	ORF1ab	DQA1*01:02/DQB1*06:02	45	59	AFASEAARVVRSIFS	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
40	ORF1ab	DPA1*01:03/DPB1*02:01	35	49	AIMQLFFSYFAVHFI	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
41	ORF1ab	DRB3*02:02	24	38	ANYIFWRNTNPIQLS	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.05
42	ORF1ab	DQA1*05:01/DQB1*03:01	35	49	ASIVAGGIVAIVVTC	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
43	ORF1ab	DRB3*02:02	56	70	AWWTAFVTNVNASSS	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.01
44	ORF1ab	DQA1*05:01/DQB1*03:01	32	46	DISASIVAGGIVAIV	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
45	ORF1ab	DPA1*03:01/DPB1*04:02	54	68	EETKFLTENLLLYID	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
46	ORF1ab	DQA1*05:01/DQB1*02:01	12	26	EIDFLELAMDEFIER	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
47	ORF1ab	DPA1*03:01/DPB1*04:02	55	69	ETKFLTENLLLYIDI	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
48	ORF1ab	DQA1*01:01/DQB1*05:01	48	62	FISNSWLMWLIINLV	94.44% (34/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
49	ORF1ab	DRB1*07:01	19	33	FSAVGNICYTPSKLI	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
50	ORF1ab	DRB1*07:01	52	66	FTPLVPFWITIAYII	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
51	ORF1ab	DQA1*01:01/DQB1*05:01	47	61	HFISNSWLMWLIINL	94.44% (34/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
52	ORF1ab	DQA1*05:01/DQB1*02:01	13	27	IDFLELAMDEFIERY	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
53	ORF1ab	DQA1*05:01/DQB1*03:01	33	47	ISASIVAGGIVAIVV	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
54	ORF1ab	DQA1*01:01/DQB1*05:01	49	63	ISNSWLMWLIINLVQ	94.44% (34/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
55	ORF1ab	DPA1*03:01/DPB1*04:02	55	69	KLINIIIWFLLLSVC	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
56	ORF1ab	DPA1*02:01/DPB1*05:01	45	59	KVTLVFLFVAAIFYL	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
57	ORF1ab	DPA1*01:03/DPB1*02:01	33	47	LAAIMQLFFSYFAVH	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
58	ORF1ab	DPA1*03:01/DPB1*04:02	53	67	LEETKFLTENLLLYI	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
59	ORF1ab	DPA1*03:01/DPB1*04:02	56	70	LINIIIWFLLLSVCL	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
60	ORF1ab	DRB1*07:01	55	69	LVPFWITIAYIICIS	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
61	ORF1ab	DQA1*05:01/DQB1*02:01	11	25	MEIDFLELAMDEFIE	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
62	ORF1ab	DPA1*02:01/DPB1*14:01	34	48	MNLKYAISAKNRART	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.08
63	ORF1ab	DPA1*01:03/DPB1*02:01	37	51	MQLFFSYFAVHFISN	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
64	ORF1ab	DPA1*01:03/DPB1*02:01	3	17	MYIFFASFYYVWKSY	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
65	ORF1ab	DRB1*11:01	25	39	NEFYAYLRKHFSMMI	97.22% (35/36)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.02
66	ORF1ab	DRB3*02:02	25	39	NYIFWRNTNPIQLSS	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.04
67	ORF1ab	DRB1*07:01	54	68	PLVPFWITIAYIICI	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.02
68	ORF1ab	DQA1*05:01/DQB1*02:01	10	24	QMEIDFLELAMDEFI	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
69	ORF1ab	DPA1*02:01/DPB1*14:01	33	47	QMNLKYAISAKNRAR	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.07
70	ORF1ab	DRB1*01:01	49	63	QQESPFVMMSAPPAQ	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
71	ORF1ab	DQA1*01:01/DQB1*05:01	26	40	QSTQWSLFFFLYENA	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
72	ORF1ab	DPA1*01:03/DPB1*02:01	29	43	QWSLFFFLYENAFLP	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
73	ORF1ab	DPA1*01:03/DPB1*02:01	2	16	RMYIFFASFYYVWKS	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
74	ORF1ab	DQA1*05:01/DQB1*03:01	34	48	SASIVAGGIVAIVVT	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
75	ORF1ab	DRB1*07:01	20	34	SAVGNICYTPSKLIE	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
76	ORF1ab	DQA1*05:01/DQB1*03:01	36	50	SIVAGGIVAIVVTCL	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03

77	ORF1ab	DPA1*03:01/DPB1*04:02	54	68	SKLINIIIWFLLLSV	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
78	ORF1ab	DQA1*01:01/DQB1*05:01	50	64	SNSWLMWLIINLVQM	94.44% (34/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
79	ORF1ab	DQA1*05:01/DQB1*02:01	9	23	SQMEIDFLELAMDEF	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
80	ORF1ab	DPA1*01:03/DPB1*02:01	27	41	STQWSLFFFLYENAF	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
81	ORF1ab	DQA1*01:01/DQB1*05:01	27	41	STQWSLFFFLYENAF	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
82	ORF1ab	DPA1*03:01/DPB1*04:02	56	70	TKFLTENLLLYIDIN	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
83	ORF1ab	DPA1*03:01/DPB1*04:02	52	66	TLEETKFLTENLLLY	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
84	ORF1ab	DRB1*07:01	53	67	TPLVPFWITIAYIIC	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
85	ORF1ab	DPA1*01:03/DPB1*02:01	28	42	TQWSLFFFLYENAFL	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.04
86	ORF1ab	DQA1*01:01/DQB1*05:01	28	42	TQWSLFFFLYENAFL	97.22% (35/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
87	ORF1ab	DRB1*11:01	24	38	VNEFYAYLRKHFSMM	97.22% (35/36)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.05
88	ORF1ab	DRB1*01:01	48	62	VQQESPFVMMSAPPA	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.1
89	ORF1ab	DQA1*01:02/DQB1*06:02	44	58	YAFASEAARVVRSIF	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
90	ORF1ab	DPA1*01:03/DPB1*02:01	4	18	YIFFASFYYVWKSYV	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
91	ORF1ab	DRB3*02:02	26	40	YIFWRNTNPIQLSSY	97.22% (35/36)	Non-Toxin	NetMHCIIpan	0.05
92	ORF1ab	DPA1*01:03/DPB1*02:01	21	35	YILFTRFFYVLGLAA	100.00% (36/36)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
93	ORF3a	DPA1*01/DPB1*04:01	33	47	APFLYLYALVYFLQS	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.12
94	ORF3a	DPA1*02:01/DPB1*14:01	27	41	DFVRATATIPIQASL	100.00% (28/28)	Non-Toxin	NetMHCIIpan	0.12
95	ORF3a	DPA1*01:03/DPB1*02:01	12	26	DTGVEHVTFFIYNKI	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.1
96	ORF3a	DPA1*02:01/DPB1*05:01	12	26	DTGVEHVTFFIYNKI	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.6
97	ORF3a	DRB1*04:05	20	34	FFIYNKIVDEPEEHV	100.00% (28/28)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.94
98	ORF3a	DPA1*01/DPB1*04:01	35	49	FLYLYALVYFLQSIN	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.12
99	ORF3a	DPA1*01:03/DPB1*02:01	14	28	GVEHVTFFIYNKIVD	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.2
100	ORF3a	DRB1*04:05	17	31	HVTFFIYNKIVDEPE	100.00% (28/28)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.84
101	ORF3a	DPA1*01/DPB1*04:01	34	48	PFLYLYALVYFLQSI	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.12
102	ORF3a	DRB1*04:05	19	33	TFFIYNKIVDEPEEH	100.00% (28/28)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.81
103	ORF3a	DPA1*01:03/DPB1*02:01	13	27	TGVEHVTFFIYNKIV	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.93
104	ORF3a	DPA1*02:01/DPB1*05:01	13	27	TGVEHVTFFIYNKIV	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.4
105	ORF3a	DRB1*04:05	18	32	VTFFIYNKIVDEPEE	100.00% (28/28)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.81
106	ORF3a	DPA1*01/DPB1*04:01	37	51	YLYALVYFLQSINFV	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.14
107	ORF6	DRB1*15:01	13	27	EILLIIMRTFKVSIW	100.00% (34/34)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.1
108	ORF6	DRB1*15:01	14	28	ILLIIMRTFKVSIWN	100.00% (34/34)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.1
109	ORF6	DQA1*01:01/DQB1*05:01	24	38	VSIWNLDYIINLIIK	100.00% (34/34)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
110	ORF7a	DRB1*01:01	2	16	KIILFLALITLATCE	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.16
111	ORF7b	DRB4*01:01	15	29	AFLLFLVLIMLIIFW	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.14
112	ORF7b	DPA1*03:01/DPB1*04:02	8	22	DFYLCFLAFLLFLVL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
113	ORF7b	DPA1*01:03/DPB1*02:01	8	22	DFYLCFLAFLLFLVL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
114	ORF7b	DPA1*01/DPB1*04:01	8	22	DFYLCFLAFLLFLVL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm)	0.09
115	ORF7b	DPA1*03:01/DPB1*04:02	13	27	FLAFLLFLVLIMLII	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
116	ORF7b	DRB4*01:01	13	27	FLAFLLFLVLIMLII	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.14
117	ORF7b	DRB4*01:01	16	30	FLLFLVLIMLIIFWF	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.19
118	ORF7b	DPA1*03:01/DPB1*04:02	9	23	FYLCFLAFLLFLVLI	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
119	ORF7b	DPA1*01:03/DPB1*02:01	9	23	FYLCFLAFLLFLVLI	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
120	ORF7b	DPA1*01/DPB1*04:01	9	23	FYLCFLAFLLFLVLI	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm)	0.09
121	ORF7b	DPA1*01:03/DPB1*02:01	7	21	IDFYLCFLAFLLFLV	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.06
122	ORF7b	DPA1*03:01/DPB1*04:02	7	21	IDFYLCFLAFLLFLV	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.16
123	ORF7b	DPA1*03:01/DPB1*04:02	14	28	LAFLLFLVLIMLIIF	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
124	ORF7b	DRB4*01:01	14	28	LAFLLFLVLIMLIIF	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.12

125	ORF7b	DPA1*01:03/DPB1*02:01	6	20	LIDFYLCFLAFLLFL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.05
126	ORF7b	DPA1*01:03/DPB1*02:01	4	18	LSLIDFYLCFLAFLL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.15
127	ORF7b	DQA1*01:01/DQB1*05:01	20	34	LVLIMLIIFWFSLEL	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.14
128	ORF7b	DPA1*01:03/DPB1*02:01	5	19	SLIDFYLCFLAFLLF	100.00% (2/2)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.07
129	ORF8	DQA1*01:01/DQB1*05:01	98	112	LVVRCSFYEDFLEYH	100.00% (34/34)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.45
130	ORF8	DRB3*01:01	27	41	QHQPYVVDDPCPIHF	100.00% (34/34)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
131	ORF8	DRB3*01:01	26	40	TQHQPYVVDDPCPIH	100.00% (34/34)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.08
132	ORF10	DPA1*01/DPB1*04:01	7	21	FAFPFTIYSLLLCRM	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.56
133	ORF10	DPA1*02:01/DPB1*01:01	7	21	FAFPFTIYSLLLCRM	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.1
134	ORF10	DPA1*03:01/DPB1*04:02	7	21	FAFPFTIYSLLLCRM	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.1
135	ORF10	DPA1*01:03/DPB1*02:01	4	18	INVFAFPFTIYSLLL	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.29
136	ORF10	DPA1*01/DPB1*04:01	4	18	INVFAFPFTIYSLLL	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.46
137	ORF10	DPA1*02:01/DPB1*01:01	4	18	INVFAFPFTIYSLLL	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.72
138	ORF10	DPA1*01:03/DPB1*02:01	5	19	NVFAFPFTIYSLLLC	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.4
139	ORF10	DPA1*01/DPB1*04:01	5	19	NVFAFPFTIYSLLLC	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.56
140	ORF10	DPA1*02:01/DPB1*01:01	5	19	NVFAFPFTIYSLLLC	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.69
141	ORF10	DPA1*01/DPB1*04:01	6	20	VFAFPFTIYSLLLCR	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm)	0.56
142	ORF10	DPA1*02:01/DPB1*01:01	6	20	VFAFPFTIYSLLLCR	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.66
143	ORF10	DPA1*01:03/DPB1*02:01	6	20	VFAFPFTIYSLLLCR	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.71
144	ORF10	DPA1*03:01/DPB1*04:02	6	20	VFAFPFTIYSLLLCR	100.00% (28/28)	Non-Toxin	Consensus (comb.lib./smm/nn)	1.1
145	S protein	DRB3*01:01	47	61	ADSFVIRGDEVRQIA	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.49
146	S protein	DRB1*07:01	13	27	AIPTNFTISVTTEIL	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.4
147	S protein	DRB3*01:01	48	62	DSFVIRGDEVRQIAP	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.51
148	S protein	DRB3*02:02	42	56	EGVFVSNGTHWFVTQ	100.00% (41/41)	Non-Toxin	NetMHCIIpan	0.21
149	S protein	DRB5*01:01	52	66	FVFKNIDGYFKIYSK	100.00% (41/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.17
150	S protein	DRB1*11:01	27	41	GNYNYLYRLFRKSNL	100.00% (41/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.22
151	S protein	DRB1*07:01	12	26	IAIPTNFTISVTTEI	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.47
152	S protein	DRB5*01:01	23	37	INITRFQTLLALHRS	97.56% (40/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.32
153	S protein	DRB5*01:01	25	39	ITRFQTLLALHRSYL	97.56% (40/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.26
154	S protein	DPA1*02:01/DPB1*14:01	25	39	ITRFQTLLALHRSYL	97.56% (40/41)	Non-Toxin	NetMHCIIpan	0.43
155	S protein	DRB1*13:02	48	62	LIVNNATNVVIKVCE	100.00% (41/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.03
156	S protein	DRB1*01:01	23	37	LSFELLHAPATVCGP	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
157	S protein	DRB5*01:01	24	38	NITRFQTLLALHRSY	97.56% (40/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.32
158	S protein	DPA1*02:01/DPB1*14:01	24	38	NITRFQTLLALHRSY	97.56% (40/41)	Non-Toxin	NetMHCIIpan	0.45
159	S protein	DRB1*07:01	15	29	PTNFTISVTTEILPV	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.51
160	S protein	DPA1*02:01/DPB1*14:01	31	45	QLIRAAEIRASANLA	100.00% (41/41)	Non-Toxin	NetMHCIIpan	0.31
161	S protein	DPA1*02:01/DPB1*14:01	30	44	QQLIRAAEIRASANL	100.00% (41/41)	Non-Toxin	NetMHCIIpan	0.2
162	S protein	DRB5*01:01	50	64	REFVFKNIDGYFKIY	100.00% (41/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.17
163	S protein	DRB3*02:02	41	55	REGVFVSNGTHWFVT	100.00% (41/41)	Non-Toxin	NetMHCIIpan	0.2
164	S protein	DRB1*01:01	24	38	SFELLHAPATVCGPK	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
165	S protein	DRB3*01:01	49	63	SFVIRGDEVRQIAPG	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.51
166	S protein	DRB1*13:02	42	56	SKTQSLLIVNNATNV	100.00% (41/41)	Non-Toxin	Consensus (smm/nn/sturniolo)	0.03
167	S protein	DRB1*01:01	22	36	VLSFELLHAPATVCG	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
168	S protein	DRB1*01:01	21	35	VVLSFELLHAPATVC	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.03
169	S protein	DRB1*01:01	20	34	VVVLSFELLHAPATV	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.09
170	S protein	DRB3*01:01	46	60	YADSFVIRGDEVRQI	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.49
171	S protein	DQA1*05:01/DQB1*03:01	25	39	YIWLGFIAGLIAIVM	100.00% (41/41)	Non-Toxin	Consensus (comb.lib./smm/nn)	0.51

Supplementary table S11. Shortlisted B Cell epitopes (BepiPred Linear B Cell Prediction). B cell linear epitopes with length of 4 to 20 amino acids, predicted by the "BepiPred Linear B Cell Prediction" IEDB tool, from eleven SARS-CoV-2 proteins, are listed here. *In-silico* analysis have shown all the selected epitopes to be non-toxic (Non-Toxin) as well as they show significant amino acid sequence conservancy.

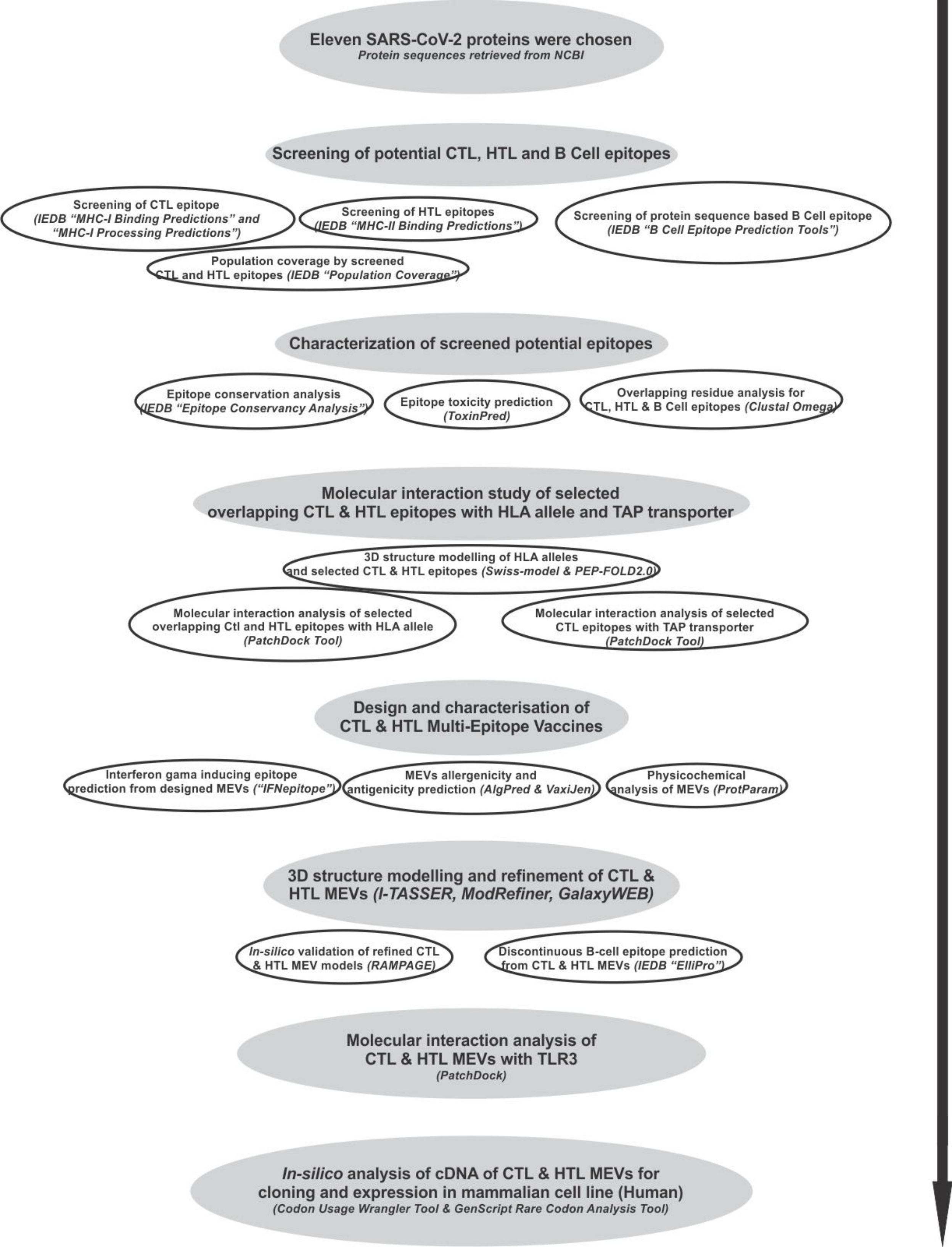
SARS-CoV-2 B Cell epitopes (Bepipred Linear Epitope Prediction)							
S.No.	Proteins	Start	End	Peptide	Conservancy	Toxicity	Length
1	E protein	6	9	SEET	100.00% (35/35)	Non-Toxin	4
2	E protein	57	71	YVYSRVKNLNSSRVP	100.00% (35/35)	Non-Toxin	15
3	M protein	132	137	PLLESE	100.00% (41/41)	Non-Toxin	6
4	M protein	180	191	KLGASQRVAGDS	100.00% (41/41)	Non-Toxin	12
5	M protein	5	20	NGTITVEELKKLLEQW	100.00% (41/41)	Non-Toxin	16
6	M protein	199	218	YRIGNYKLNTDHSSSSDNIA	100.00% (41/41)	Non-Toxin	20
7	N protein	343	348	DPNFKD	97.50% (39/40)	Non-Toxin	6
8	N protein	119	127	AGLPYGANK	100.00% (40/40)	Non-Toxin	9
9	N protein	4	15	NGPQNQRNAPRI	100.00% (40/40)	Non-Toxin	12
10	N protein	404	416	SKQLQQSMSSADS	100.00% (40/40)	Non-Toxin	13
11	ORF1ab	95	98	IQYG	100.00% (36/36)	Non-Toxin	4
12	ORF1ab	1064	1067	YLKH	100.00% (36/36)	Non-Toxin	4
13	ORF1ab	1318	1321	RKVP	100.00% (36/36)	Non-Toxin	4
14	ORF1ab	1831	1834	YQCG	100.00% (36/36)	Non-Toxin	4
15	ORF1ab	2112	2115	ELSR	100.00% (36/36)	Non-Toxin	4
16	ORF1ab	2457	2460	TFCA	100.00% (36/36)	Non-Toxin	4
17	ORF1ab	2763	2766	GKIV	100.00% (36/36)	Non-Toxin	4
18	ORF1ab	4528	4531	EGNC	97.22% (35/36)	Non-Toxin	4
19	ORF1ab	5221	5224	LPYP	97.22% (35/36)	Non-Toxin	4
20	ORF1ab	5337	5340	SLRC	97.22% (35/36)	Non-Toxin	4
21	ORF1ab	5516	5519	KVQI	97.22% (35/36)	Non-Toxin	4
22	ORF1ab	6255	6258	HDIG	97.22% (35/36)	Non-Toxin	4
23	ORF1ab	6974	6977	WNAD	97.22% (35/36)	Non-Toxin	4
24	ORF1ab	113	117	EIPVA	97.22% (35/36)	Non-Toxin	5
25	ORF1ab	1124	1128	KGEDI	100.00% (36/36)	Non-Toxin	5
26	ORF1ab	1181	1185	NLYDK	100.00% (36/36)	Non-Toxin	5
27	ORF1ab	1689	1693	KFNPP	100.00% (36/36)	Non-Toxin	5
28	ORF1ab	1756	1760	GQQQT	100.00% (36/36)	Non-Toxin	5
29	ORF1ab	1936	1940	KFADD	100.00% (36/36)	Non-Toxin	5
30	ORF1ab	2128	2132	AAVNS	100.00% (36/36)	Non-Toxin	5
31	ORF1ab	2547	2551	SKCEE	100.00% (36/36)	Non-Toxin	5
32	ORF1ab	2903	2907	IEYTD	100.00% (36/36)	Non-Toxin	5
33	ORF1ab	3035	3039	IQPIG	100.00% (36/36)	Non-Toxin	5
34	ORF1ab	3243	3247	NSGSD	100.00% (36/36)	Non-Toxin	5
35	ORF1ab	3553	3557	EFTPF	100.00% (36/36)	Non-Toxin	5
36	ORF1ab	3899	3903	LLAKD	100.00% (36/36)	Non-Toxin	5
37	ORF1ab	3932	3936	EEMLD	100.00% (36/36)	Non-Toxin	5

38	ORF1ab	4017	4021	RSEDK	100.00% (36/36)	Non-Toxin	5
39	ORF1ab	5288	5292	TGHML	97.22% (35/36)	Non-Toxin	5
40	ORF1ab	5877	5881	AHSCN	97.22% (35/36)	Non-Toxin	5
41	ORF1ab	6736	6740	QTGSS	97.22% (35/36)	Non-Toxin	5
42	ORF1ab	18	23	LPVLQV	100.00% (36/36)	Non-Toxin	6
43	ORF1ab	31	36	FGDSVE	97.22% (35/36)	Non-Toxin	6
44	ORF1ab	76	81	ARTAPH	100.00% (36/36)	Non-Toxin	6
45	ORF1ab	208	213	AGKASC	100.00% (36/36)	Non-Toxin	6
46	ORF1ab	220	225	DFIDTK	100.00% (36/36)	Non-Toxin	6
47	ORF1ab	1813	1818	QYELKH	100.00% (36/36)	Non-Toxin	6
48	ORF1ab	1960	1965	FPDLNG	100.00% (36/36)	Non-Toxin	6
49	ORF1ab	2297	2302	SLDTYP	100.00% (36/36)	Non-Toxin	6
50	ORF1ab	2312	2317	SSFKWD	100.00% (36/36)	Non-Toxin	6
51	ORF1ab	2414	2419	KRNRAT	100.00% (36/36)	Non-Toxin	6
52	ORF1ab	2442	2447	KGFCKL	100.00% (36/36)	Non-Toxin	6
53	ORF1ab	2701	2706	AQVAKS	100.00% (36/36)	Non-Toxin	6
54	ORF1ab	3881	3886	VESSSK	100.00% (36/36)	Non-Toxin	6
55	ORF1ab	5905	5910	LYDKLQ	97.22% (35/36)	Non-Toxin	6
56	ORF1ab	6496	6501	ENKTTL	97.22% (35/36)	Non-Toxin	6
57	ORF1ab	6873	6878	DKGVAP	97.22% (35/36)	Non-Toxin	6
58	ORF1ab	1092	1098	ATNGPLK	100.00% (36/36)	Non-Toxin	7
59	ORF1ab	2678	2684	NKVENMT	100.00% (36/36)	Non-Toxin	7
60	ORF1ab	2747	2753	TRQVVNV	100.00% (36/36)	Non-Toxin	7
61	ORF1ab	2861	2867	EVGFVVP	100.00% (36/36)	Non-Toxin	7
62	ORF1ab	2924	2930	DASGKPV	100.00% (36/36)	Non-Toxin	7
63	ORF1ab	3009	3015	DYYRSLP	100.00% (36/36)	Non-Toxin	7
64	ORF1ab	3169	3175	GVSFSTF	100.00% (36/36)	Non-Toxin	7
65	ORF1ab	3269	3275	MAFPSGK	100.00% (36/36)	Non-Toxin	7
66	ORF1ab	3825	3831	SQGLLPP	100.00% (36/36)	Non-Toxin	7
67	ORF1ab	3858	3864	VQSKMSD	100.00% (36/36)	Non-Toxin	7
68	ORF1ab	3924	3930	AVDINKL	100.00% (36/36)	Non-Toxin	7
69	ORF1ab	3992	3998	DRDAAMQ	100.00% (36/36)	Non-Toxin	7
70	ORF1ab	4196	4202	FPKSDGT	100.00% (36/36)	Non-Toxin	7
71	ORF1ab	4336	4342	HPNPKGF	100.00% (36/36)	Non-Toxin	7
72	ORF1ab	5104	5110	GNKIADK	97.22% (35/36)	Non-Toxin	7
73	ORF1ab	5317	5323	YTPHTVL	97.22% (35/36)	Non-Toxin	7
74	ORF1ab	5524	5530	FEKGDYG	97.22% (35/36)	Non-Toxin	7
75	ORF1ab	5857	5863	VDSSQGS	97.22% (35/36)	Non-Toxin	7
76	ORF1ab	6078	6084	MYKGLPW	94.44% (34/36)	Non-Toxin	7
77	ORF1ab	6223	6229	IIGDELK	97.22% (35/36)	Non-Toxin	7
78	ORF1ab	6801	6807	QAWQPGV	97.22% (35/36)	Non-Toxin	7
79	ORF1ab	7012	7018	KPREQID	94.44% (34/36)	Non-Toxin	7
80	ORF1ab	5	12	VPGFNEKT	100.00% (36/36)	Non-Toxin	8
81	ORF1ab	175	182	RELNGGAY	100.00% (36/36)	Non-Toxin	8
82	ORF1ab	813	820	FTLKGGAP	100.00% (36/36)	Non-Toxin	8
83	ORF1ab	1720	1727	KTVGELGD	100.00% (36/36)	Non-Toxin	8

84	ORF1ab	2603	2610	NVPMEKLK	100.00% (36/36)	Non-Toxin	8
85	ORF1ab	4255	4262	GNATEVPA	100.00% (36/36)	Non-Toxin	8
86	ORF1ab	4286	4293	SGGQPITN	100.00% (36/36)	Non-Toxin	8
87	ORF1ab	6756	6763	EIIKSQDL	97.22% (35/36)	Non-Toxin	8
88	ORF1ab	860	868	YTVELGTEV	100.00% (36/36)	Non-Toxin	9
89	ORF1ab	1301	1309	IPTKKAGGT	100.00% (36/36)	Non-Toxin	9
90	ORF1ab	1333	1341	GLNGYTVEE	100.00% (36/36)	Non-Toxin	9
91	ORF1ab	1423	1431	YFYTSKTTV	100.00% (36/36)	Non-Toxin	9
92	ORF1ab	1586	1594	MSMTYGQQF	100.00% (36/36)	Non-Toxin	9
93	ORF1ab	2621	2629	AKNVSLDNV	100.00% (36/36)	Non-Toxin	9
94	ORF1ab	2814	2822	GGVTRDIAS	100.00% (36/36)	Non-Toxin	9
95	ORF1ab	4751	4759	VNLHSSRLS	97.22% (35/36)	Non-Toxin	9
96	ORF1ab	5125	5133	RNRDVDTDF	97.22% (35/36)	Non-Toxin	9
97	ORF1ab	5261	5269	LTKHPNQEY	97.22% (35/36)	Non-Toxin	9
98	ORF1ab	5465	5473	TEETFKLSY	97.22% (35/36)	Non-Toxin	9
99	ORF1ab	5967	5975	IPGIPKDMT	97.22% (35/36)	Non-Toxin	9
100	ORF1ab	6351	6359	FHTPAFDKS	97.22% (35/36)	Non-Toxin	9
101	ORF1ab	6464	6472	KGHFDGQQG	97.22% (35/36)	Non-Toxin	9
102	ORF1ab	6636	6644	GVVQQLPET	97.22% (35/36)	Non-Toxin	9
103	ORF1ab	6674	6682	IERYKLEGY	94.44% (34/36)	Non-Toxin	9
104	ORF1ab	56	65	VEKGVLPQLE	100.00% (36/36)	Non-Toxin	10
105	ORF1ab	245	254	SEKSYELQTP	100.00% (36/36)	Non-Toxin	10
106	ORF1ab	373	382	CHNSEVGPEH	100.00% (36/36)	Non-Toxin	10
107	ORF1ab	720	729	CVKSREETGL	100.00% (36/36)	Non-Toxin	10
108	ORF1ab	1510	1519	SYKDWSYSGQ	100.00% (36/36)	Non-Toxin	10
109	ORF1ab	1851	1860	ALLTKSSEYK	100.00% (36/36)	Non-Toxin	10
110	ORF1ab	1943	1952	QLTGYKKPAS	100.00% (36/36)	Non-Toxin	10
111	ORF1ab	2076	2085	ILKPANNSLK	100.00% (36/36)	Non-Toxin	10
112	ORF1ab	3500	3509	YNYEPLTQDH	100.00% (36/36)	Non-Toxin	10
113	ORF1ab	3965	3974	EQAVANGDSE	100.00% (36/36)	Non-Toxin	10
114	ORF1ab	4312	4321	PEANMDQESF	100.00% (36/36)	Non-Toxin	10
115	ORF1ab	4413	4422	PCGTGTSTDV	97.22% (35/36)	Non-Toxin	10
116	ORF1ab	4448	4457	FQEKDEDDNL	97.22% (35/36)	Non-Toxin	10
117	ORF1ab	4468	4477	TFSNYQHEET	97.22% (35/36)	Non-Toxin	10
118	ORF1ab	4493	4502	FFKFRIDGDM	97.22% (35/36)	Non-Toxin	10
119	ORF1ab	4936	4945	LKYAISAKNR	97.22% (35/36)	Non-Toxin	10
120	ORF1ab	5162	5171	YASQGLVASI	97.22% (35/36)	Non-Toxin	10
121	ORF1ab	5192	5201	WTETDLTKGP	97.22% (35/36)	Non-Toxin	10
122	ORF1ab	5990	5999	QVNGYPNMFI	97.22% (35/36)	Non-Toxin	10
123	ORF1ab	6540	6549	YKRDAPAHIS	97.22% (35/36)	Non-Toxin	10
124	ORF1ab	6578	6587	RVDGQVDLFR	97.22% (35/36)	Non-Toxin	10
125	ORF1ab	1652	1662	HTKKWKYPQVN	100.00% (36/36)	Non-Toxin	11
126	ORF1ab	2478	2488	KRPINPTDQSS	100.00% (36/36)	Non-Toxin	11
127	ORF1ab	4708	4718	LFSTVFPPTSF	97.22% (35/36)	Non-Toxin	11
128	ORF1ab	5493	5503	VGKPRPPLNRN	97.22% (35/36)	Non-Toxin	11
129	ORF1ab	6329	6339	RVLSNLNLPGC	97.22% (35/36)	Non-Toxin	11

130	ORF1ab	6558	6568	DIAKKPTETIC	97.22% (35/36)	Non-Toxin	11
131	ORF1ab	387	398	YHNESGLKTILR	100.00% (36/36)	Non-Toxin	12
132	ORF1ab	743	754	EGETLPTEVLTE	100.00% (36/36)	Non-Toxin	12
133	ORF1ab	1973	1984	KHYTPSFKKGAK	100.00% (36/36)	Non-Toxin	12
134	ORF1ab	2638	2649	RQGFVDSDVETK	100.00% (36/36)	Non-Toxin	12
135	ORF1ab	3195	3206	SDVLLPLTQYNR	100.00% (36/36)	Non-Toxin	12
136	ORF1ab	4138	4149	KLQNNELSPVAL	100.00% (36/36)	Non-Toxin	12
137	ORF1ab	4772	4783	MHAASGNLLLDK	97.22% (35/36)	Non-Toxin	12
138	ORF1ab	5068	5079	KPGGTSSGDATT	97.22% (35/36)	Non-Toxin	12
139	ORF1ab	5300	5311	TNDNTSRYWEPE	97.22% (35/36)	Non-Toxin	12
140	ORF1ab	5371	5382	PYVCNAPGCDVT	97.22% (35/36)	Non-Toxin	12
141	ORF1ab	6018	6029	GCHATREAVGTN	97.22% (35/36)	Non-Toxin	12
142	ORF1ab	6599	6610	SVKGLQPSVGPK	97.22% (35/36)	Non-Toxin	12
143	ORF1ab	6898	6909	LNDFVSDADSTL	97.22% (35/36)	Non-Toxin	12
144	ORF1ab	1354	1366	YILPSIISNEKQE	100.00% (36/36)	Non-Toxin	13
145	ORF1ab	1456	1468	GLNLEEAARYMRS	100.00% (36/36)	Non-Toxin	13
146	ORF1ab	1553	1565	FDNLKTLLSLREV	100.00% (36/36)	Non-Toxin	13
147	ORF1ab	1777	1789	EQFKKGVQIPCTC	100.00% (36/36)	Non-Toxin	13
148	ORF1ab	3310	3322	EDMLNPNYEDLLI	100.00% (36/36)	Non-Toxin	13
149	ORF1ab	5919	5931	RNVATLQAENVTG	97.22% (35/36)	Non-Toxin	13
150	ORF1ab	6932	6944	PKTKNVTKENDSK	97.22% (35/36)	Non-Toxin	13
151	ORF1ab	3211	3224	YNKYKYFSGAMDTT	100.00% (36/36)	Non-Toxin	14
152	ORF1ab	6002	6015	EEAIRHVRAWIGFD	97.22% (35/36)	Non-Toxin	14
153	ORF1ab	326	340	CGETSWQTGDFVKAT	100.00% (36/36)	Non-Toxin	15
154	ORF1ab	763	777	LQPLEQPTSEAVEAP	100.00% (36/36)	Non-Toxin	15
155	ORF1ab	886	900	VSELLTPLGIDLDEW	100.00% (36/36)	Non-Toxin	15
156	ORF1ab	2797	2811	SKHTDFSSEIIGYKA	100.00% (36/36)	Non-Toxin	15
157	ORF1ab	5730	5744	PAPRTLLTKGTLEPE	97.22% (35/36)	Non-Toxin	15
158	ORF1ab	6512	6526	KRNIKPVPEVKILNN	97.22% (35/36)	Non-Toxin	15
159	ORF1ab	6709	6723	RFKESPFELEDFIPM	97.22% (35/36)	Non-Toxin	15
160	ORF1ab	345	361	GTENLTKEGATTCGYLP	100.00% (36/36)	Non-Toxin	17
161	ORF1ab	2833	2849	DFDTWFSQRGGSYTNDK	100.00% (36/36)	Non-Toxin	17
162	ORF1ab	3356	3372	TANPKTPKYKFVRIQPG	100.00% (36/36)	Non-Toxin	17
163	ORF1ab	4546	4562	DDYFNKKDWYDFVENPD	97.22% (35/36)	Non-Toxin	17
164	ORF1ab	454	471	LQKEKVNINIVGDFKLNE	100.00% (36/36)	Non-Toxin	18
165	ORF1ab	2135	2152	WDTIANYAKPFLNKVVST	100.00% (36/36)	Non-Toxin	18
166	ORF1ab	1482	1500	AVTAYNGYLTSSSKTPEEH	100.00% (36/36)	Non-Toxin	19
167	ORF1ab	4380	4398	GCSCDQLREPMLQSADAQS	100.00% (36/36)	Non-Toxin	19
168	ORF1ab	6053	6071	PNNTDFSRVSAKPPPGDQF	97.22% (35/36)	Non-Toxin	19
169	ORF1ab	6817	6835	RMLLEKCDLQNYGDSATLP	97.22% (35/36)	Non-Toxin	19
170	ORF1ab	1398	1417	IVSTIQRKYKGIKIQEGVVD	100.00% (36/36)	Non-Toxin	20
171	ORF1ab	4157	4176	GTTQTACTDDNALAYYNTTK	100.00% (36/36)	Non-Toxin	20
172	ORF1ab	7035	7054	IQLSSYSLFDMSKFPLKLRG	97.22% (35/36)	Non-Toxin	20
173	ORF3a	135	140	SKNPLL	92.86% (26/28)	Non-Toxin	6
174	ORF3a	159	164	PYNSVT	100.00% (28/28)	Non-Toxin	6
175	ORF3a	216	225	STQLSTDTGV	100.00% (28/28)	Non-Toxin	10

176	ORF3a	61	71	KIITLKKRWQL	100.00% (28/28)	Non-Toxin	11
177	ORF3a	17	28	QGEIKDATPSDF	100.00% (28/28)	Non-Toxin	12
178	ORF6	44	57	LTENKYSQLDEEQP	100.00% (34/34)	Non-Toxin	14
179	ORF7a	17	25	LYHYQECVR	100.00% (28/28)	Non-Toxin	9
180	ORF7b	33	39	ELQDHNE	100.00% (2/2)	Non-Toxin	7
181	ORF8	91	95	QEPKL	100.00% (34/34)	Non-Toxin	5
182	ORF8	106	111	EDFLEY	100.00% (34/34)	Non-Toxin	6
183	ORF8	48	56	RVGARKSAP	100.00% (34/34)	Non-Toxin	9
184	ORF8	63	78	DEAGSKSPIQYIDIGN	100.00% (34/34)	Non-Toxin	16
185	S protein	293	296	LDPL	100.00% (41/41)	Non-Toxin	4
186	S protein	580	583	QTLE	100.00% (41/41)	Non-Toxin	4
187	S protein	1203	1206	LGKY	100.00% (41/41)	Non-Toxin	4
188	S protein	602	606	TNTSN	100.00% (41/41)	Non-Toxin	5
189	S protein	988	992	EAEVQ	100.00% (41/41)	Non-Toxin	5
190	S protein	773	779	EQDKNTQ	100.00% (41/41)	Non-Toxin	7
191	S protein	555	562	SNKKFLPF	100.00% (41/41)	Non-Toxin	8
192	S protein	807	814	PDPSKPSK	100.00% (41/41)	Non-Toxin	8
193	S protein	1035	1043	GQSKRVDFC	100.00% (41/41)	Non-Toxin	9
194	S protein	250	260	TPGDSSSGWTA	100.00% (41/41)	Non-Toxin	11
195	S protein	634	644	RVYSTGSNVFQ	100.00% (41/41)	Non-Toxin	11
196	S protein	656	666	VNNSYECDIPI	100.00% (41/41)	Non-Toxin	11
197	S protein	1107	1118	RNFYEPQIITTD	100.00% (41/41)	Non-Toxin	12
198	S protein	177	189	MDLEGKQGNFKNL	100.00% (41/41)	Non-Toxin	13
199	S protein	786	800	KQIYKTPPIKDFGGF	100.00% (41/41)	Non-Toxin	15
200	S protein	828	842	LADAGFIKQYGDCLG	100.00% (41/41)	Non-Toxin	15
201	S protein	206	221	KHTPINLVRDLPQGFS	100.00% (41/41)	Non-Toxin	16
202	S protein	695	710	YTMSLGAENSVAYSNN	100.00% (41/41)	Non-Toxin	16
203	S protein	138	154	DPFLGVYYHKNNKSWME	100.00% (41/41)	Non-Toxin	17
204	S protein	616	632	NCTEVPVAIHADQLTPT	100.00% (41/41)	Non-Toxin	17
205	S protein	304	322	KSFTVEKGIYQTSNFRVQP	100.00% (41/41)	Non-Toxin	19
206	S protein	672	690	ASYQTQTNSPRRARSVASQ	100.00% (41/41)	Non-Toxin	19


Eleven SARS-CoV-2 proteins were chosen
Protein sequences retrieved from NCBI
Screening of potential CTL, HTL and B Cell epitopes
Screening of CTL epitope (IEDB "MHC-I Binding Predictions" and "MHC-I Processing Predictions")
Screening of HTL epitopes (IEDB "MHC-II Binding Predictions")
Screening of protein sequence based B Cell epitope (IEDB "B Cell Epitope Prediction Tools")
Population coverage by screened CTL and HTL epitopes (IEDB "Population Coverage")
Characterization of screened potential epitopes
Epitope conservation analysis (IEDB "Epitope Conservancy Analysis")
Epitope toxicity prediction (ToxinPred)
Overlapping residue analysis for CTL, HTL & B Cell epitopes (Clustal Omega)
Molecular interaction study of selected overlapping CTL & HTL epitopes with HLA allele and TAP transporter
3D structure modelling of HLA alleles and selected CTL & HTL epitopes (Swiss-model & PEP-FOLD2.0)
Molecular interaction analysis of selected overlapping Ctl and HTL epitopes with HLA allele (PatchDock Tool)
Molecular interaction analysis of selected CTL epitopes with TAP transporter (PatchDock Tool)
Design and characterisation of CTL & HTL Multi-Epitope Vaccines
Interferon gama inducing epitope prediction from designed MEVs ("IFNepitope")
MEVs allergenicity and antigenicity prediction (AlgPred & VaxiJen)
Physicochemical analysis of MEVs (ProtParam)
3D structure modelling and refinement of CTL & HTL MEVs (I-TASSER, ModRefiner, GalaxyWEB)
In-silico validation of refined CTL & HTL MEV models (RAMPAGE)
Discontinuous B-cell epitope prediction from CTL & HTL MEVs (IEDB "ElliPro")
Molecular interaction analysis of CTL & HTL MEVs with TLR3
(PatchDock)
In-silico analysis of cDNA of CTL & HTL MEVs for cloning and expression in mammalian cell line (Human)
(Codon Usage Wrangler Tool & GenScript Rare Codon Analysis Tool)

A. CTL epitopes

B. HTL epitopes

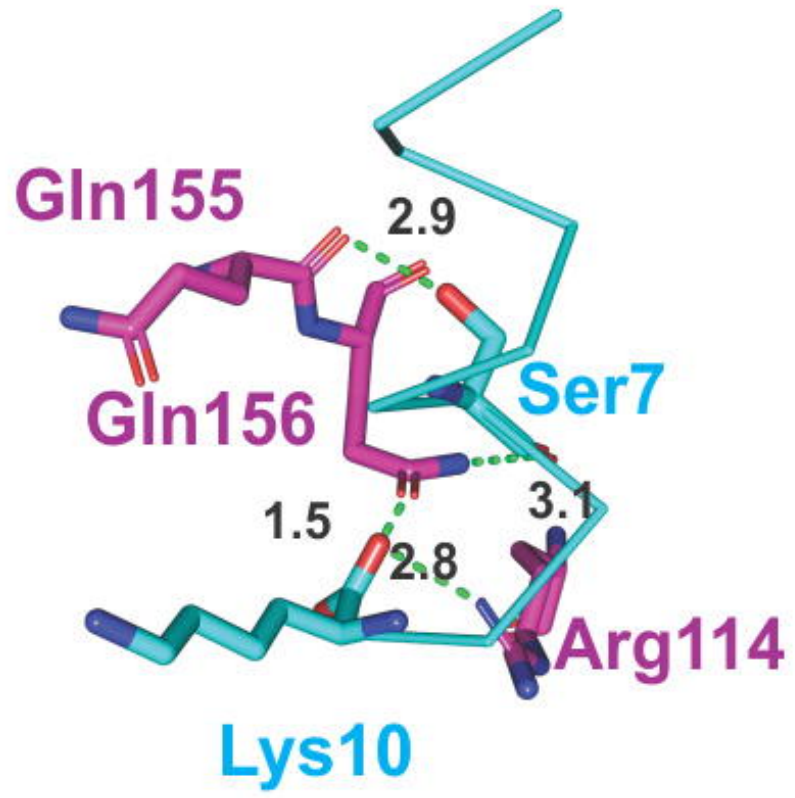
Ov-ASP-1 (truncated (10-153))

IVVAVTGYNCPGGKLTALERKKIVGQNNKYRSDLINGKL
KNRNGTYMPRGKNMLELTWDCKLESSAQRWANQCIFG
HSPRQQREGVGENVYAYWSSVSVEGLKKTAGTDAGKS
WWSKLPKLYENNPSNNMTWKVAGQGVLHFTQ

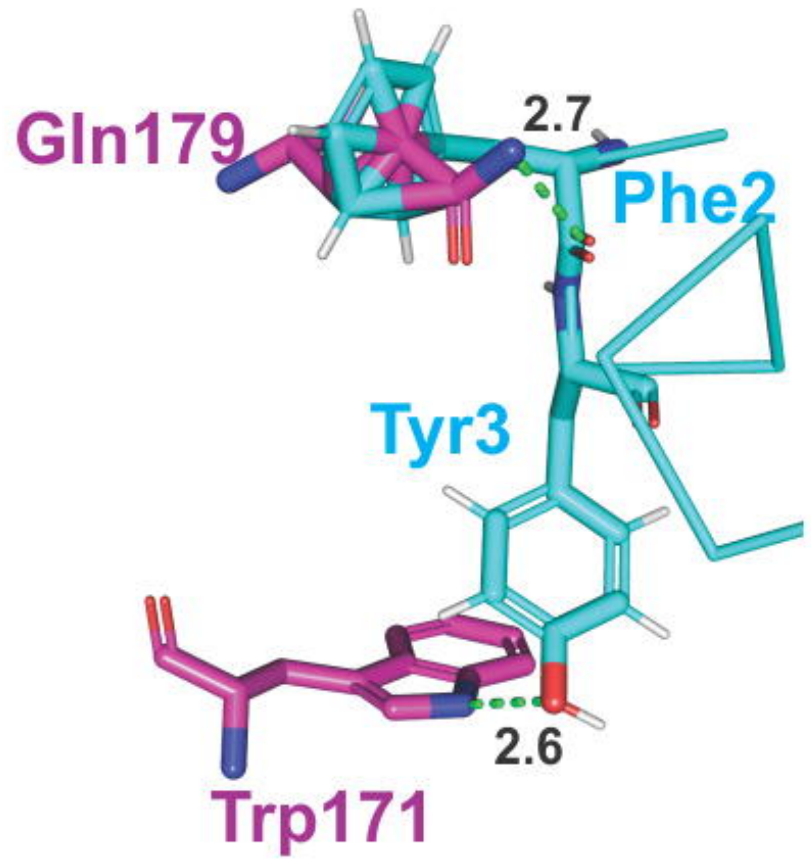
	A. CTL epitopes	B. HTL epitopes
ORF1ab	EAAAK	EAAAK
	SEMVMCGGSLY	WFLAYILFTRFFYVL
	GGGGS	GGGGS
	FYWFFSNYLKR	FLAYILFTRFFYVLG
	GGGGS	GGGGS
	ISNSWLMW	LAYILFTRFFYVLGL
	GGGGS	GGGGS
	ETISLAGSYK	AYILFTRFFYVLGLA
	GGGGS	GGGGS
	QEILGTVSW	RAMPNMLRIMASLVL
	GGGGS	GGGGS
	STFNVPMEK	AMPNMLRIMASLVLA
	GGGGS	GGGGS
	RMYIFFASFY	MPNMLRIMASLVLAR
	GGGGS	GGGGS
	FLFVAAIFYL	PNMLRIMASLVLARK
	GGGGS	GGGGS
	RYFRLTLGVY	QESPFVMMSAPPAQY
	GGGGS	GGGGS
	FLNGSCGSV	NMLRIMASLVLARKH
	GGGGS	GGGGS
	CTDDNALAY	ESPFVMMSAPPAQYE
	GGGGS	GGGGS
	CTDDNALAYY	SPFVMMSAPPAQYEL
	GGGGS	GGGGS
	MYKGLPWNVVR	AIILASFSASTSAFV
	GGGGS	GGGGS
	SIINNTVYTK	IILASFSASTSAFVE
	GGGGS	GGGGS
	LPVNVAFELW	FAWWTAFVTNVNASS
	GGGGS	GGGGS
	DEWSMATYY	KTQSLLIVNNATNVV
	GGGGS	GGGGS
	YILFTRFFYV	TQSLLIVNNATNVVI
	GGGGS	GGGGS
	YIFFASFYYV	QSLLIVNNATNVVIK
	GGGGS	GGGGS
Surface protein	WTAGAAAYYV	SLLIVNNATNVVIKV
	GGGGS	GGGGS
	FPNITNLCPF	LLIVNNATNVVIKVC
	GGGGS	GGGGS
	NYNYLYRLFR	LLFVTVYSHLLLVAA
	GGGGS	GGGGS
	NYLYRLFR	FVRATATIPIQASLP
	GGGGS	GGGGS
ORF3a	YLYALVYFL	VLLFLAFVVFLLVTL
	GGGGS	GGGGS
	IPYNSVTSSI	LLFLAFVVFLLVTLA
	GGGGS	GGGGS
Envelope protein	LLFLAFVVF	GLMWLSYFIASFRLF
	GGGGS	GGGGS
	LTALRLCAY	LMWLSYFIASFRLFA
	GGGGS	GGGGS
	YFIASFRLFAR	LSYYKLGASQRVAGD
	GGGGS	GGGGS
Membrane protein	ATSRTLSYYK	TFKVSIWNLDYIINL
	GGGGS	GGGGS
	RTFKVSIW	FKVSIWNLDYIINLI
	GGGGS	GGGGS
ORF6	AEILLIIMRTF	KVSIWNLDYIINLII
	GGGGS	GGGGS
	RARSVSPK	IILFLALITLATCEL
	GGGGS	GGGGS
ORF7a	QLRARSVSPK	ILFLALITLATCELY
	GGGGS	GGGGS
	FLAFLLFLV	YLCFLAFLLFLVLIM
	GGGGS	GGGGS
ORF7b	HFYSKWYIR	LCFLAFLLFLVLIML
	GGGGS	GGGGS
	MEVTPSGTW	CFLAFLLFLVLIMLI
	GGGGS	GGGGS
	KPRQKRTAT	CTQHQPYVVDDPCPI
	GGGGS	GGGGS
ORF8	MGYINVFAF	HQPYVVDDPCPIHFY
	GGGGS	GGGGS
	GYINVFAFPF	QPYVVDDPCPIHFYS
	HHHHHH	GGGGS
Nucleocapsid phosphoprotein		PQIAQFAPSASAFFG
		GGGGS
		IAQFAPSASAFFGMS
		GGGGS
		AQFAPSASAFFGMSR
		GGGGS
ORF10		YINVFAFPFTIYSLL
		HHHHHH

Overlapping T & B cell epitopes across proteome of SARS-CoV-2

ORF1ab	-IILASFSASTSAFVE AIILASFSASTSAFV- 474-489 AIILASFSASTSAFVE ----SPFVMMSAPPAQYEL QESPFVMMSAPPAQY-- -ESPFVMMSAPPAQYE- 1800-1816 QESPFVMMSAPPAQYEL --------YILFTRFFYV---- WFLAYILFTRFFYVL--- -----LAYILFTRFFYVLGL- ---FLAYILFTRFFYVLG-- -------AYILFTRFFYVLGLA 2327-2344 WFLAYILFTRFFYVLGLA -----YIFFASFYYV RMYIFFASFY-- 2382-2393 RMYIFFASFYYV ----------CTDDNALAY--------- ----------CTDDNALAYY------- GTTQTACTDDNALAYYNTTK 4157-4176 GTTQTACTDDNALAYYNTTK RAMPNMLRIMASLVL---- -------PNMLRIMASLVLARK- ---------NMLRIMASLVLARKH ----MPNMLRIMASLVLAR-- --AMPNMLRIMASLVLA--- 5016-5034 RAMPNMLRIMASLVLARKH
Surface protein	--------LLIVNNATNVVIKVC ------SLLIVNNATNVVIKV- ----QSLLIVNNATNVVIK-- --TQSLLIVNNATNVVI--- KTQSLLIVNNATNVV---- 113-131 KTQSLLIVNNATNVVIKVC ----------WTAGAAAYYV TPGDSSSGWTA------- 250-267 TPGDSSSGWTAGAAAYYV FPNITNLCPF---------- FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA 329-363 FPNITNLCPFGEVFNATRFASVYAWNRKRISNCVA ----------NYNYLYRLFR---------- ----------NYLYRLFR---------- NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN 440-501 NLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIYQAGSTPCNGVEGFNCYFPLQSYGFQPTN
ORF3a	----------FVRATATIPIQASLP QGEIKDATPSDF---------- 17-42 QGEIKDATPSDFVRATATIPIQASLP IPYNSVTSSI -PYNSVT--- 158-167 IPYNSVTSSI
Envelope protein	-LLFLAFVVF------ -LLFLAFVVFLLVTLA VLLFLAFVVFLLVTL- 17-32 VLLFLAFVVFLLVTLA
Membrane Protein	----------YFIASFRLFAR --LMWLSYFIASFRLFA- GLMWLSYFIASFRLF-- 89-105 GLMWLSYFIASFRLFAR ATSRTLSYYK---------- ----------LSYYKLGASQRVAGD- ----------KLGASQRVAGDS 171-191 ATSRTLSYYKLGASQRVAGDS
ORF6	AEILLIIMRTF---------- ----------RTFKVSIW---------- ----------KVSIWNLDYIINLI ----------FKVSIWNLDYIINLI- ----------TFKVSIWNLDYIINL-- 12-37 AEILLIIMRTFKVSIWNLDYIINLI
ORF7a	-ILFLALITLATCELY------- IILFLALITLATCEL-------- ----------LYHYQECVR 3-25 IILFLALITLATCELYHYQECVR ----------RARSVSPK---------- ----------QLRARSVSPK---------- VKHVYQLRARSVSPKLFIRQEEVQEL 71-96 VKHVYQLRARSVSPKLFIRQEEVQEL
ORF7b	------FLAFLLFLV----- ----CFLAFLLFLVLIMLI --LCFLAFLLFLVLIML- YLCFLAFLLFLVLIM-- 10-26 YLCFLAFLLFLVLIMLI
ORF8	----------HFYSKWYIR-------- ----------QPYVVDDPCPIHFYS---------- ----------HQPYVVDDPCPIHFY---------- -----CTQHQPYVVDDPCPI---------- QSCTQHQPYVVDDPCPIHFYSKW---------- ----------RVGARKSAP 23-56 QSCTQHQPYVVDDPCPIHFYSKWYIRVGARKSAP
Nucleocapsid Protein	----------KPRQKRTAT-- RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA 226-267 RLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQKRTATKA -----AQFAPSASAFFGMSR ----IAQFAPSASAFFGMS- PQIAQFAPSASAFFG--- 302-319 PQIAQFAPSASAFFGMSR
ORF10	---GYINVFAFPF------ MGYINVFAF-------- -----YINVFAFPFTIYSLL 1-17 MGYINVFAFPFTIYSLL

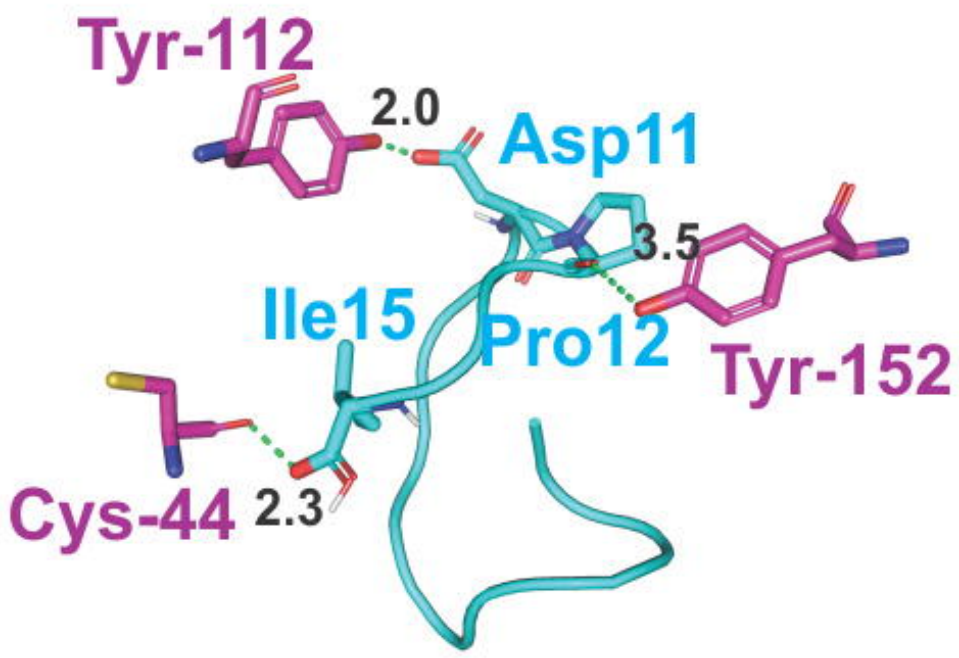


1. ATSRTLSYYK-HLA-A*11:01

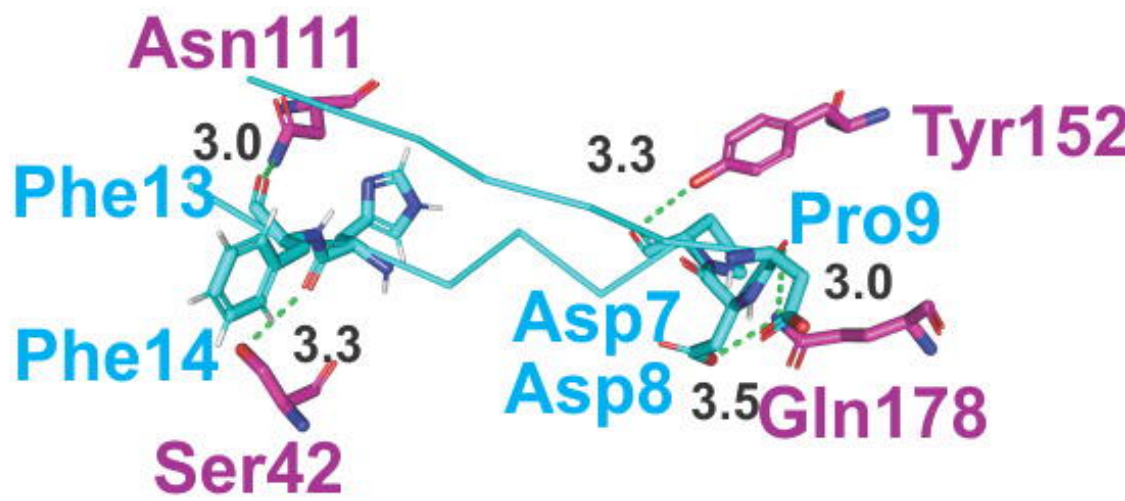


2. HFYSKWYIR-HLA-A*31:01

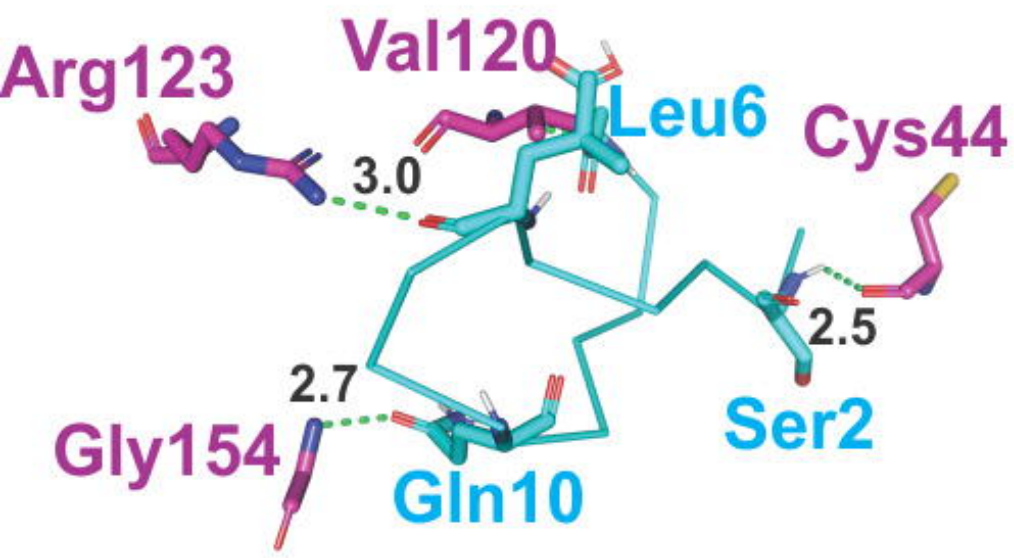
CTL epitopes - HLA Class I allele complexes



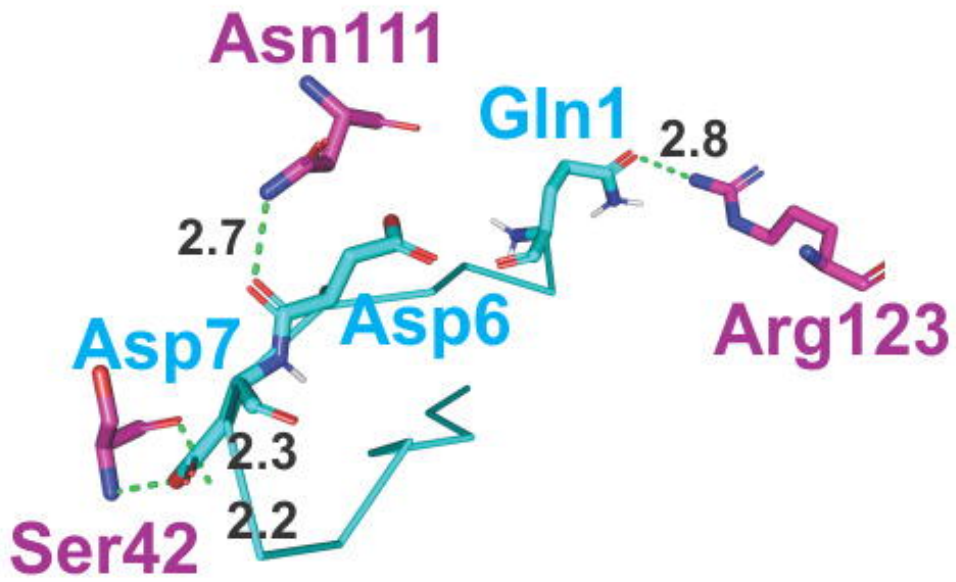
3. CTQHQPYVVDDPCPI-HLA-DRB3*01:01



4. HQPYVVDDPCPIHFY-HLA-DRB3*01:01

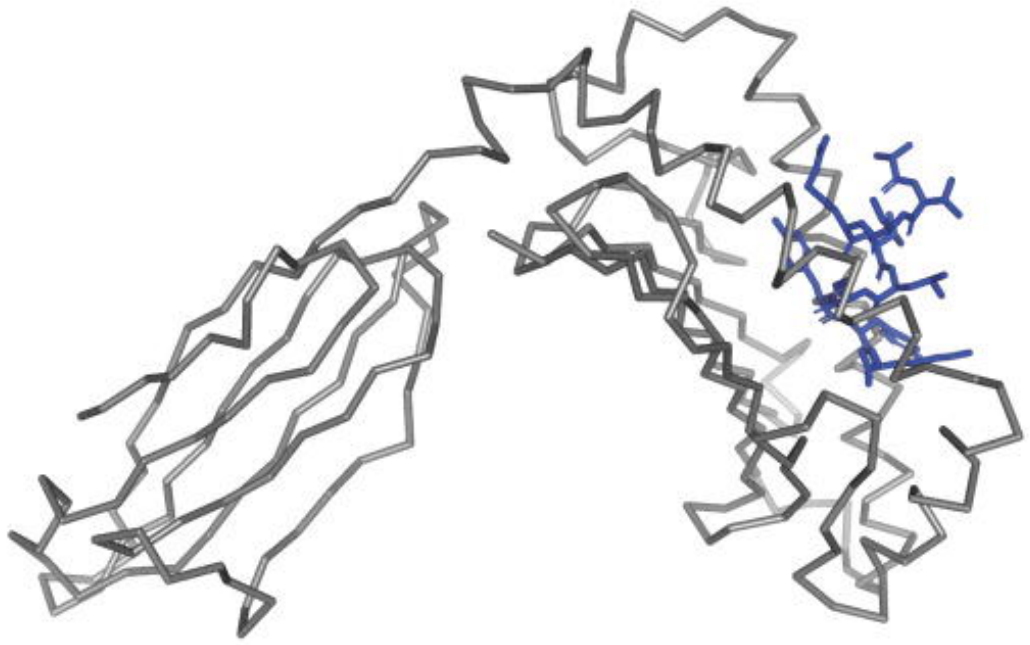


5. LSYYKLGASQRVAGD-HLA-DRB1*09:01

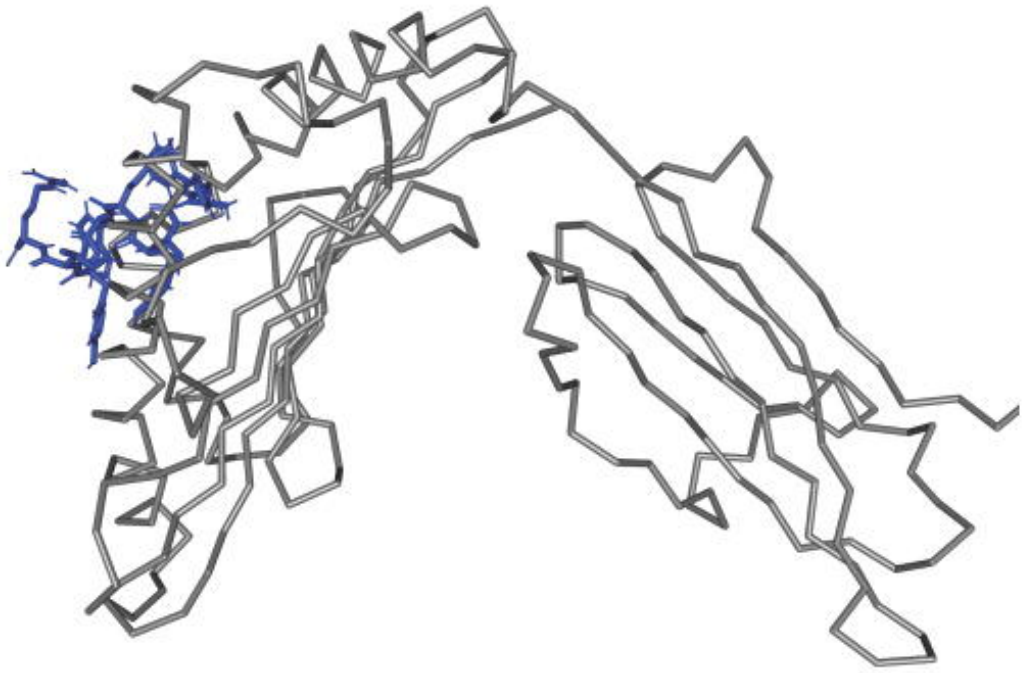


6. QPYVVDDPCPIHFYS-HLA-DRB3*01:01

HTL epitopes - HLA Class II allele complexes

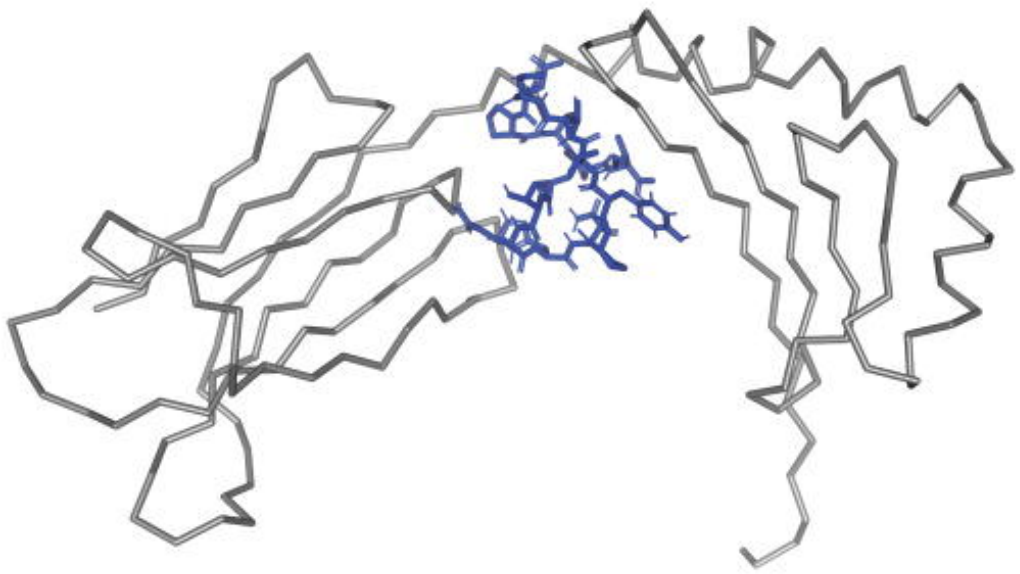


1. ATSRTLSYYK-HLA-A*11:01

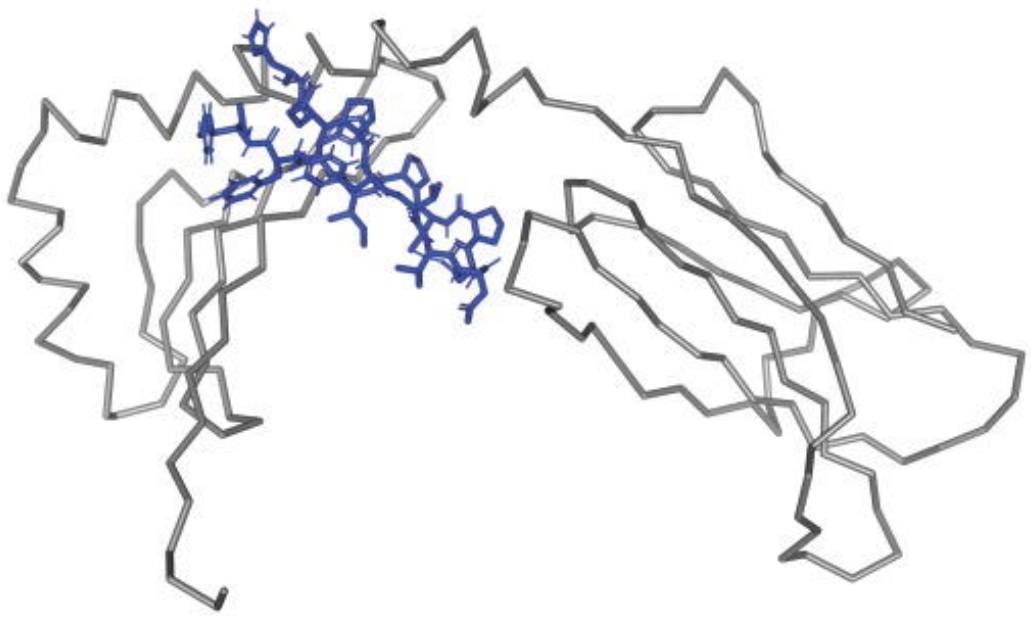


2. HFYSKWYIR-HLA-A*31:01

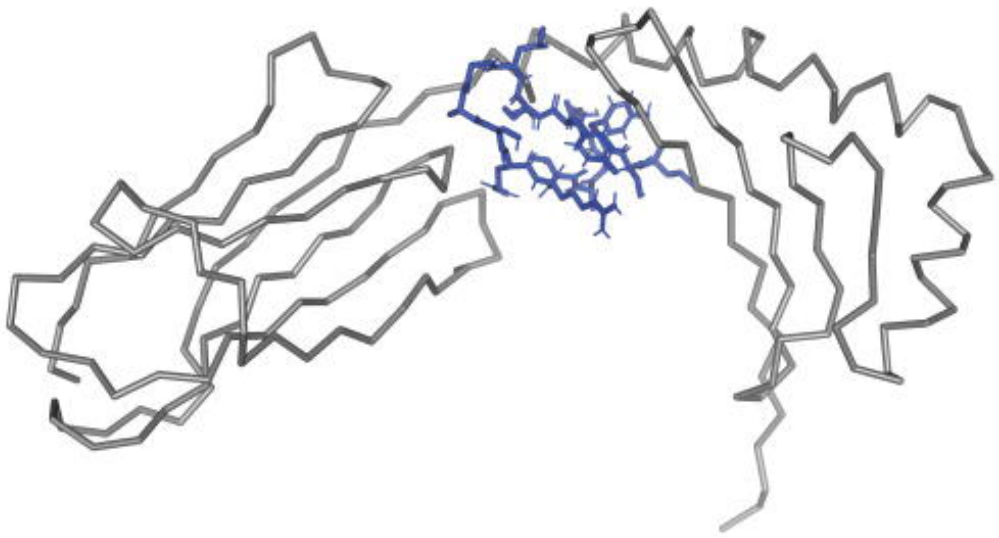
CTL epitopes - HLA Class I allele complexes



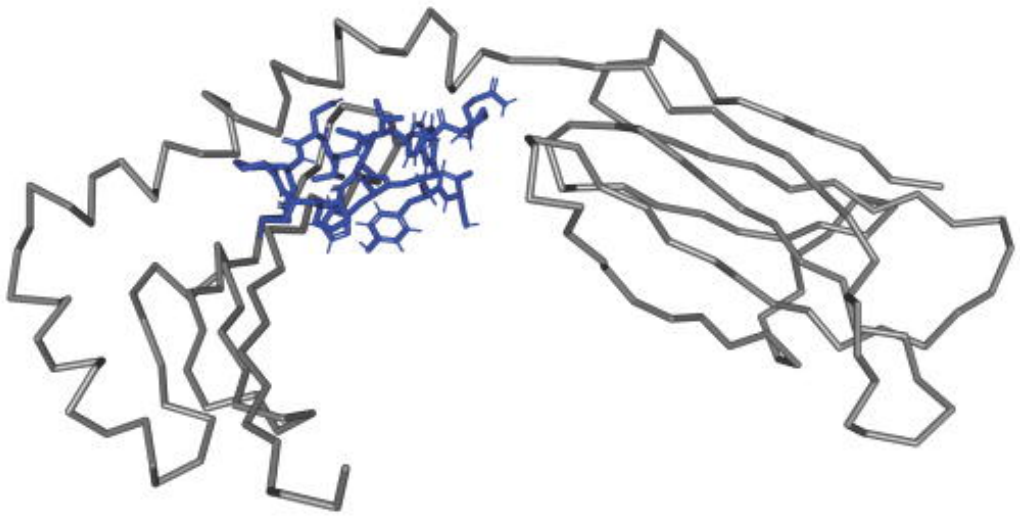
3. CTQHQPYVVDDPCPI-HLA-DRB3*01:01



4. HQPYVVDDPCPIHFY-HLA-DRB3*01:01

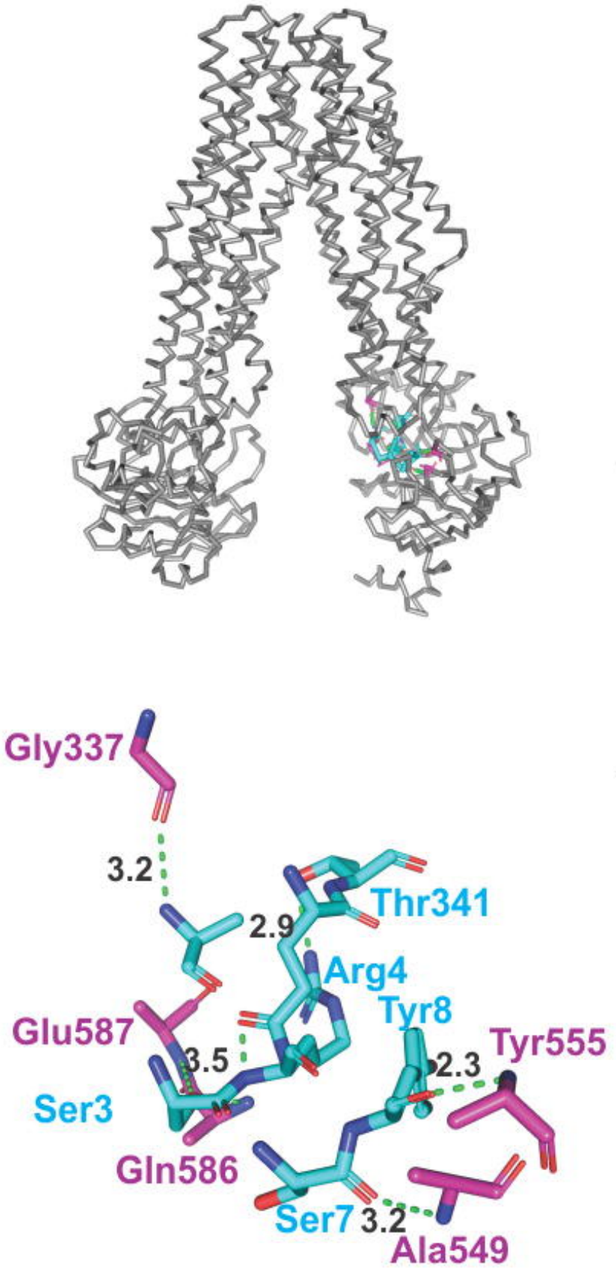


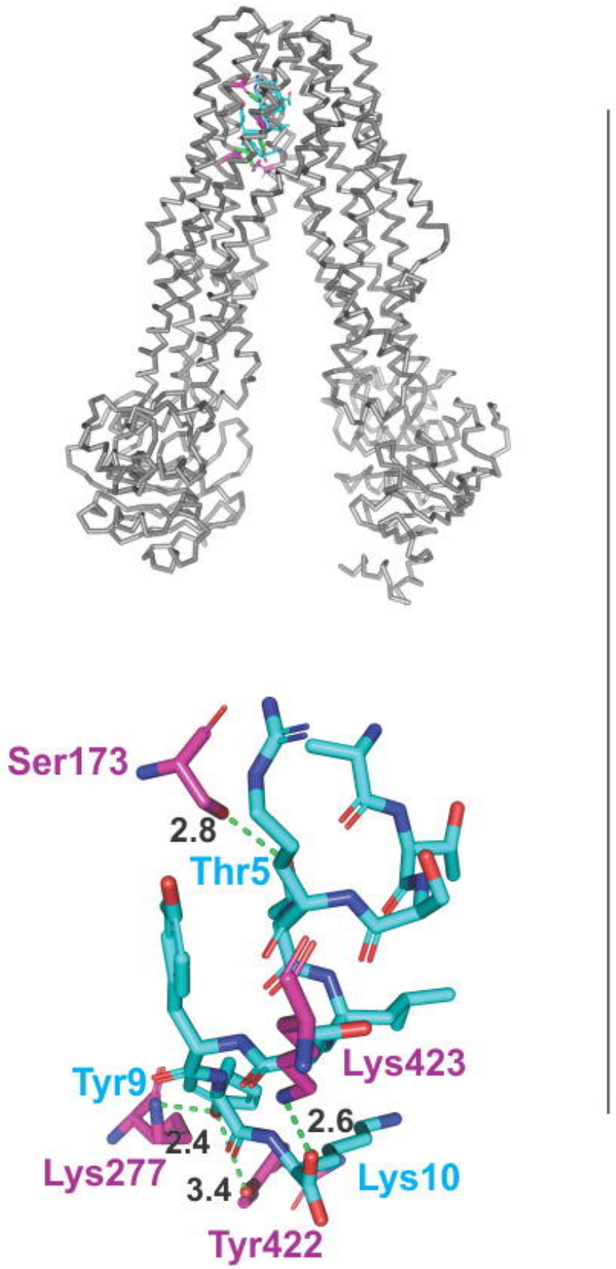
5. LSYYKLGASQRVAGD-HLA-DRB1*09:01



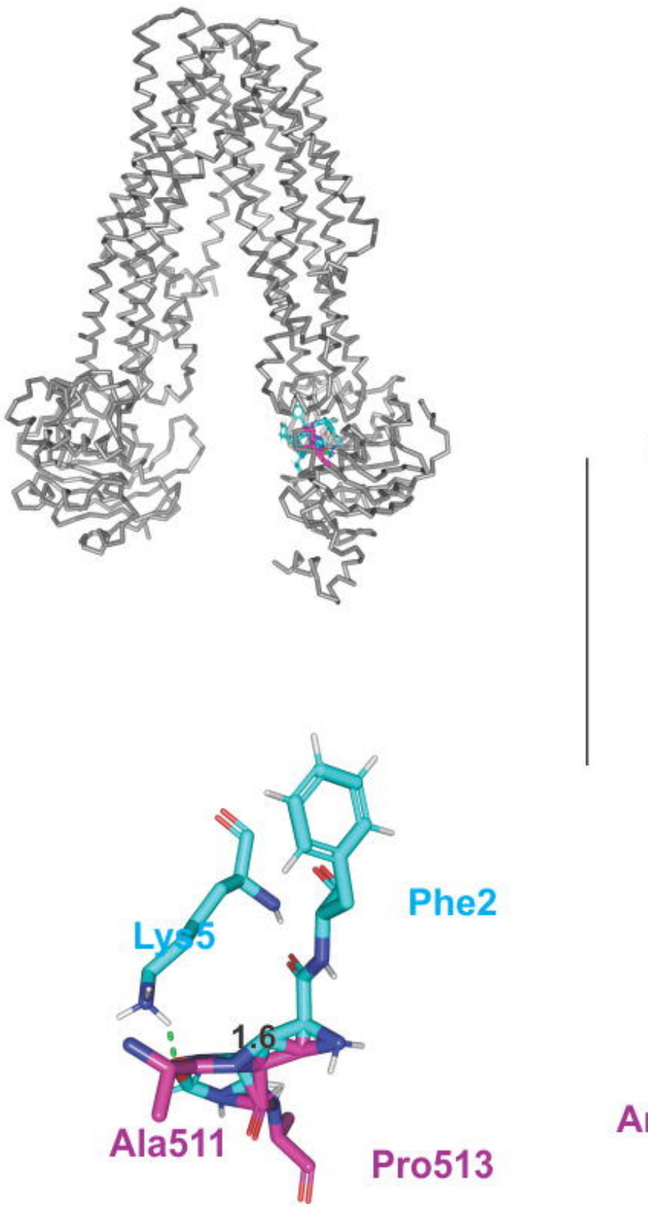
6. QPYVVDDPCPIHFYS-HLA-DRB3*01:01

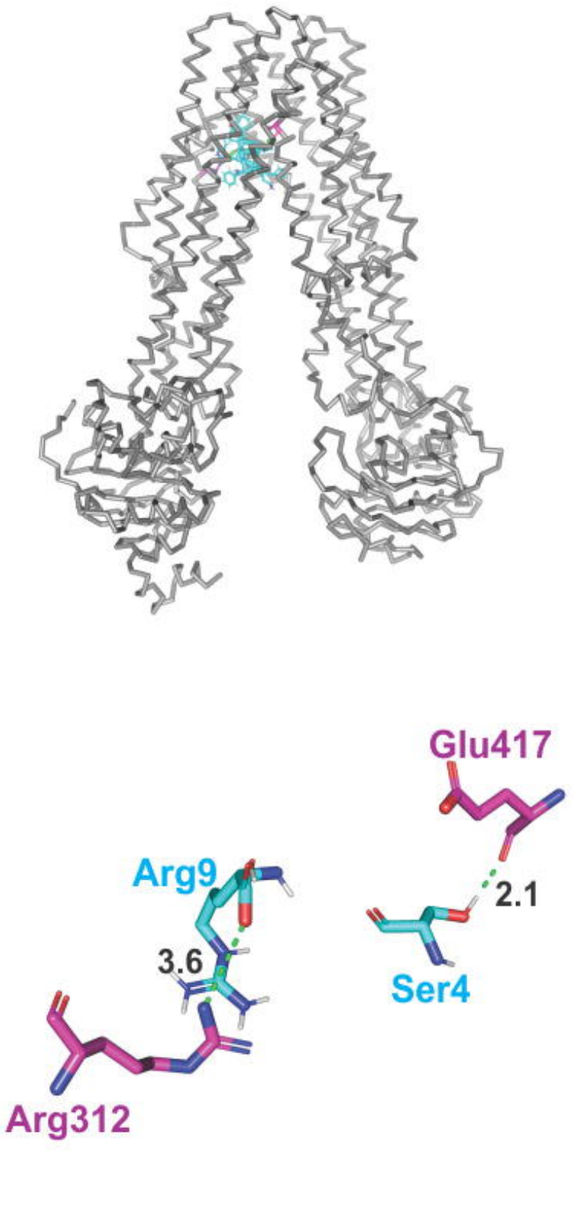
HTL epitopes - HLA Class II allele complexes



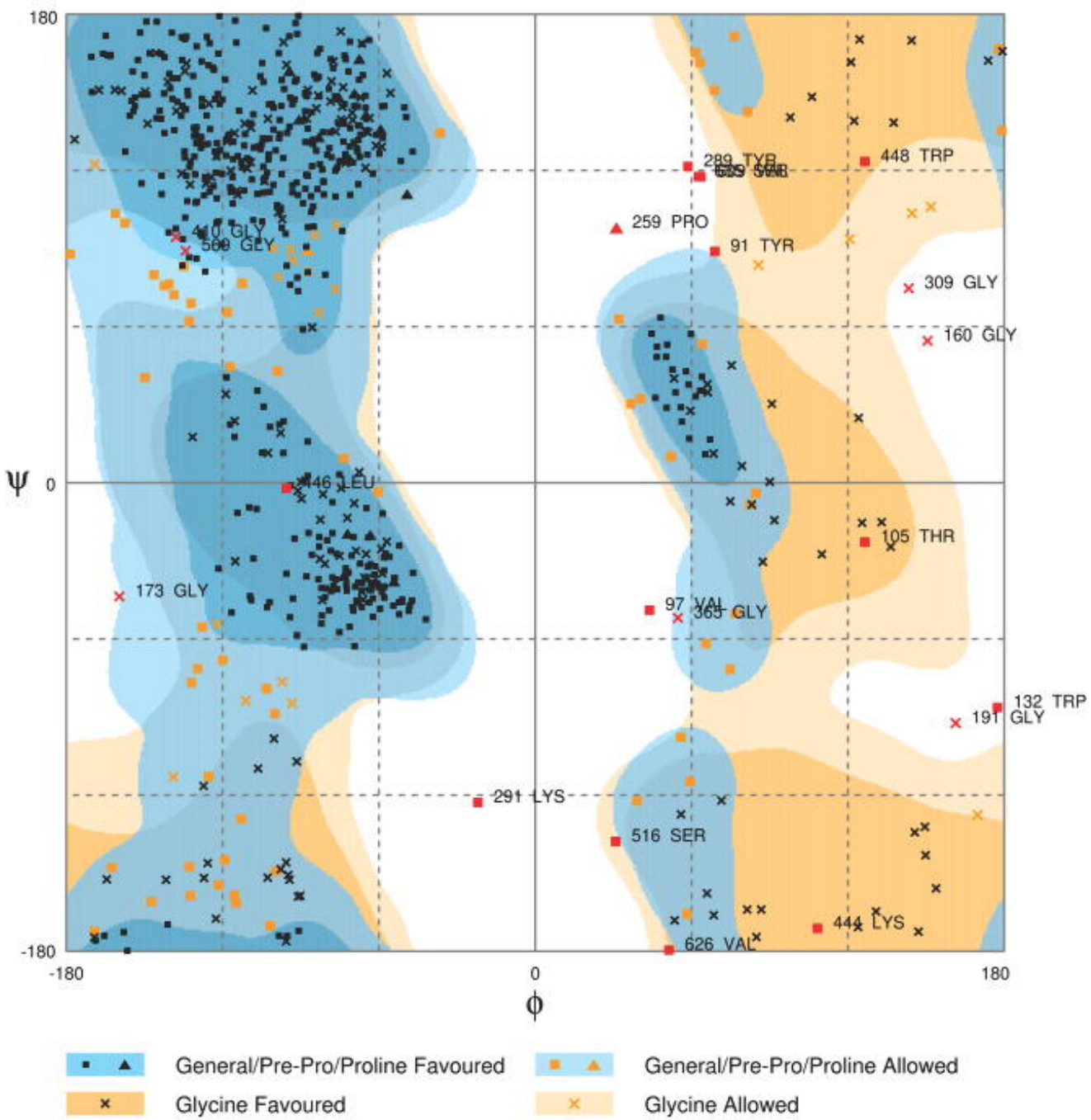


Epitope ATSRTLSYYK through TAP cavity



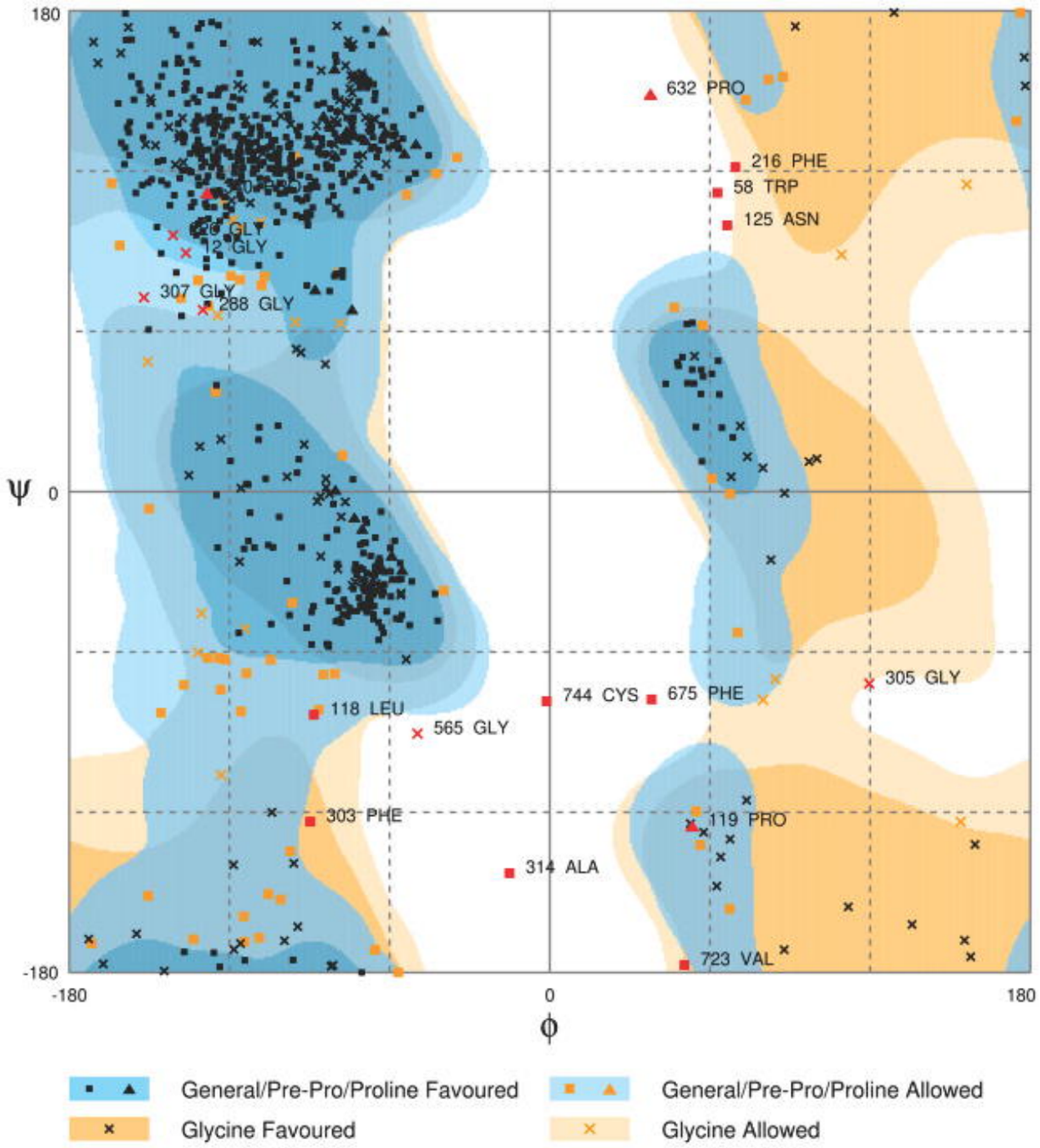


Epitope HFYSKWYIR through TAP cavity



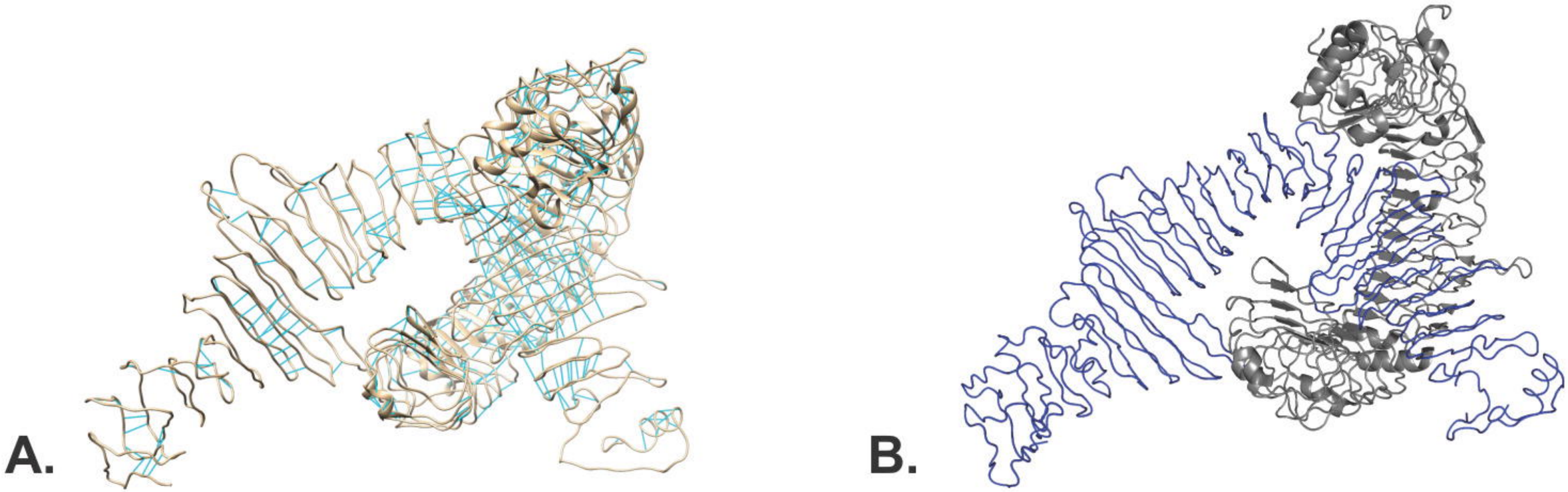
Number of residues in favoured region (~98.0% expected) : 602 (85.8%)
Number of residues in allowed region (~2.0% expected) : 79 (11.3%)
Number of residues in outlier region : 21 (3.0%)

A. CTL Multi-epitope vaccine RAMPAGE analysis

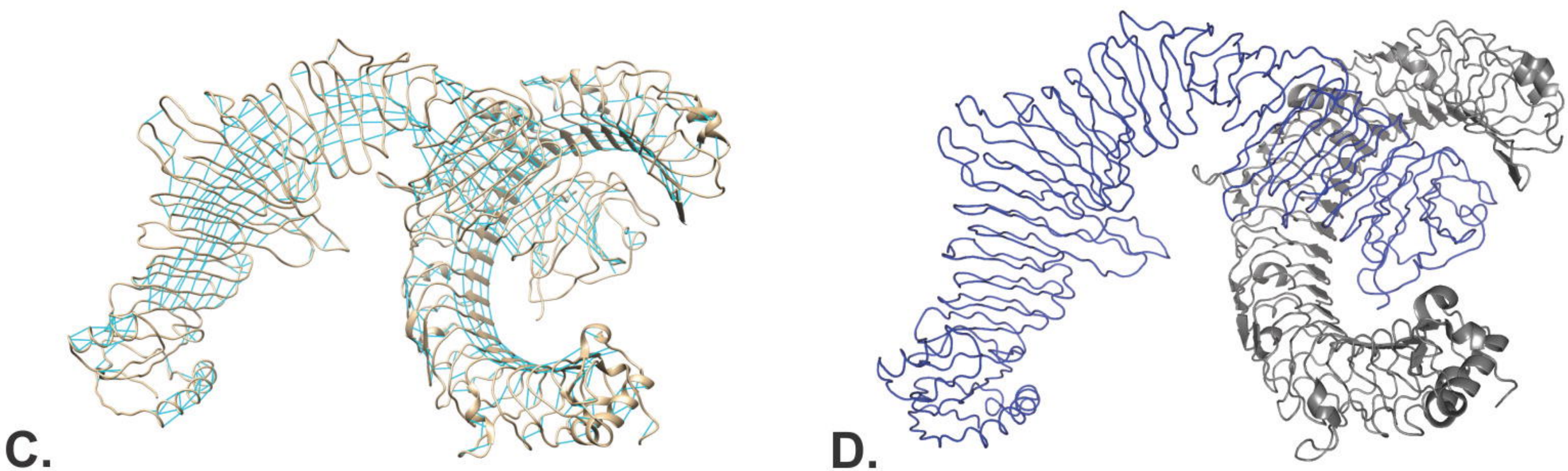


Number of residues in favoured region (~98.0% expected) : 718 (88.9%)
Number of residues in allowed region (~2.0% expected) : 72 (8.9%)
Number of residues in outlier region : 18 (2.2%)

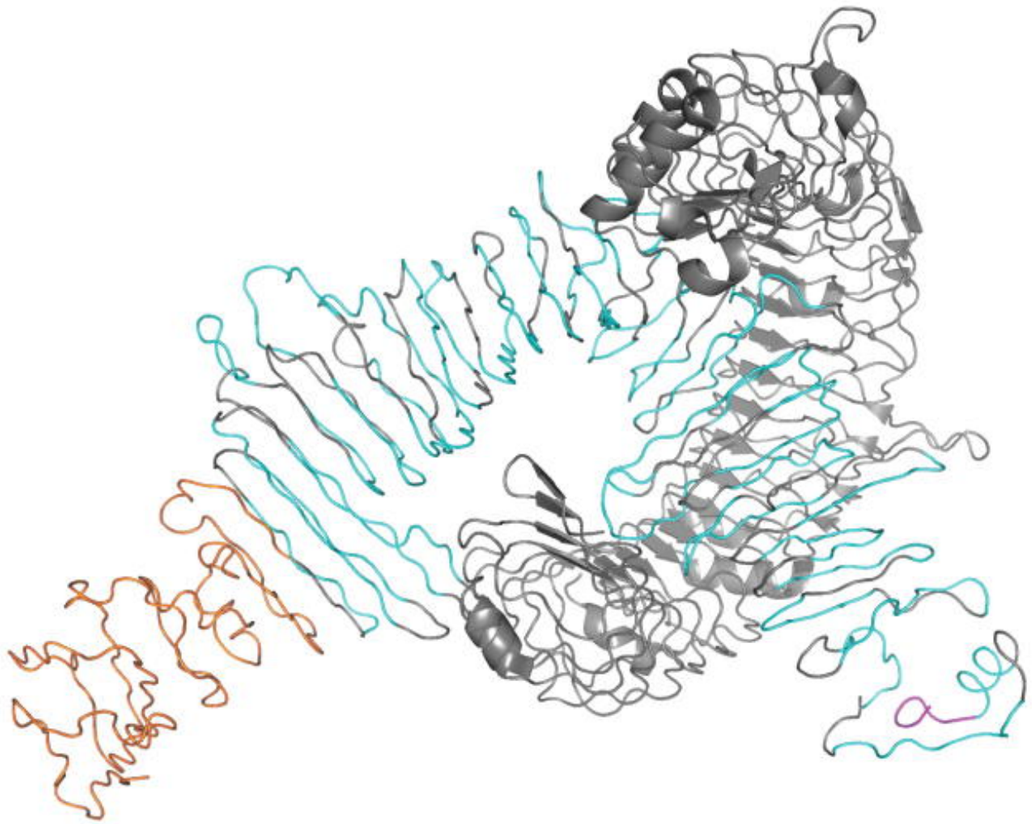
B. HTL Multi-epitope vaccine model RAMPAGE analysis



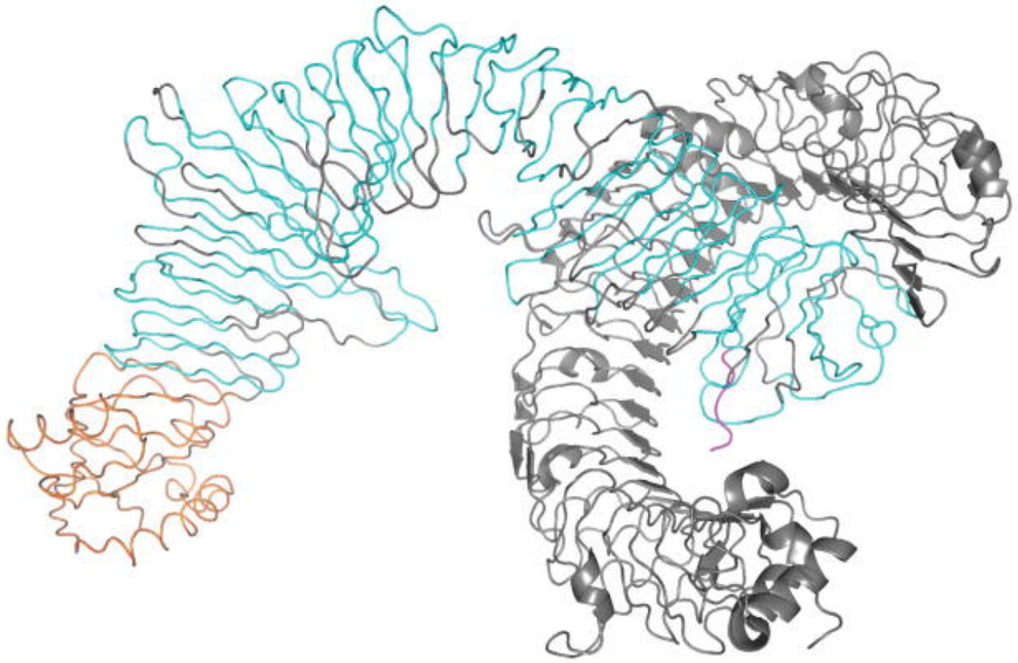
CTL Multi-epitope vaccine model in complex with TLR3



HTL Multi-epitope vaccine model in complex with TLR3

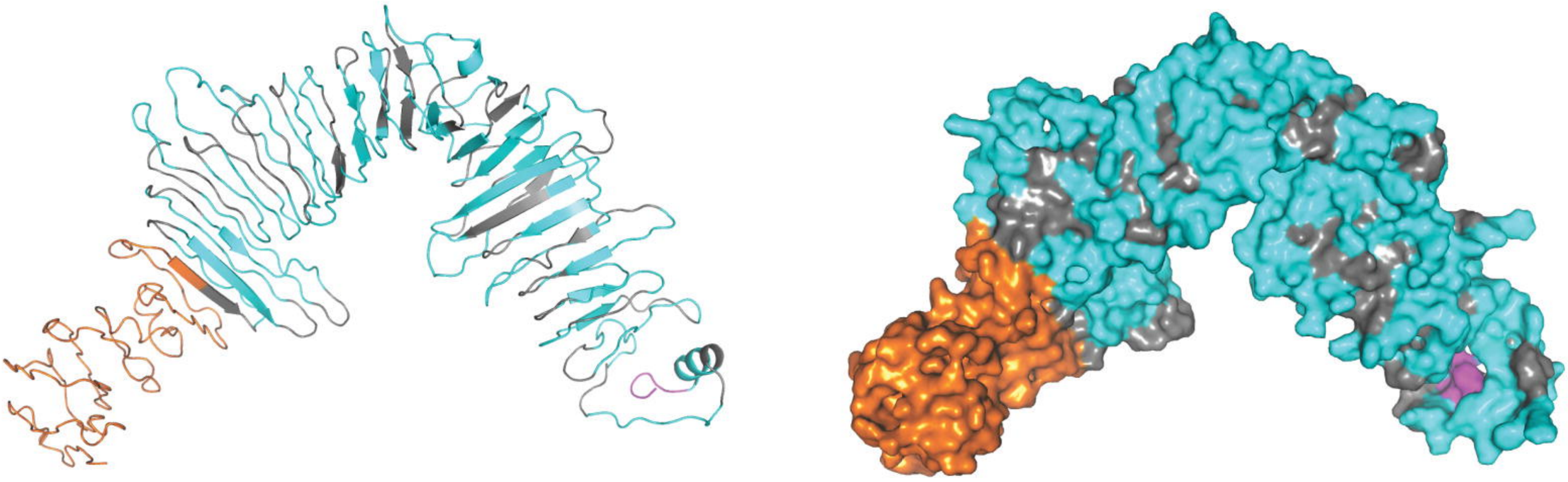


CTL Multi-epitope COVID19 vaccine - Toll Like Receptor 3 complex

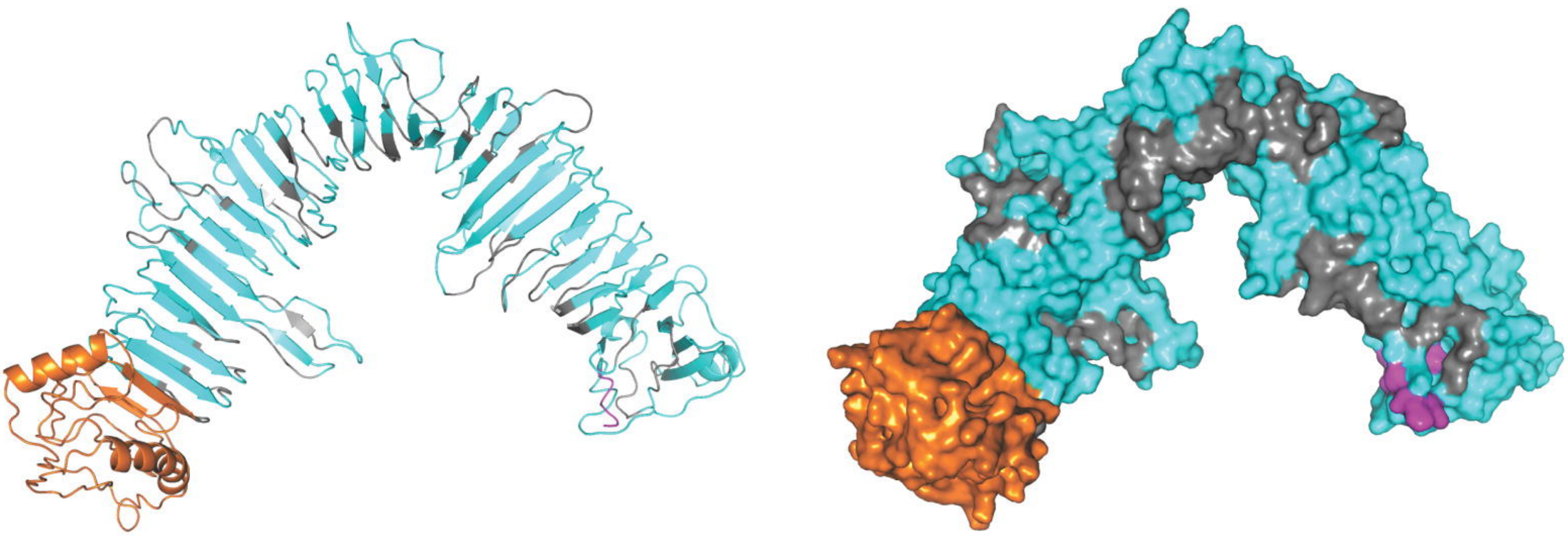


CTL Multi-epitope COVID19 vaccine - Toll Like Receptor 3 complex

Graphical abstract: The designed CTL (Cytotoxic T lymphocyte) and HTL (Helper T lymphocyte) multi-epitope vaccine (MEV) against COVID19 infection. Both the CTL and HTL MEVs models show a very stable and very fitting conformational complex formation tendency with the Toll like receptor 3. CTL and HTL MEVs: ribbon; Toll like receptor 3: gray cartoon; Adjuvant [truncated (residues 10-153) Onchocerca volvulus activation-associated secreted protein-1]: orange ribbon; Epitopes: cyan ribbons regions; 6xHis Tag: magenta ribbon region



A. CTL Multi-epitope vaccine model



B. HTL Multi-epitope vaccine model