1   **Targeting "Immunogenic Hotspots" in Dengue and Zika Virus:**

2   **A Novel Approach to a Common Vaccine**

3

4   Dhrubajyoti Mahata [1, 2,¶], Debangshu Mukherjee[1,¶], Vanshika Malviya[3], Gayatri Mukherjee [1 *]

5

6   [1] School of Medical Science and Technology, Indian Institute of Technology Kharagpur, India

7   [2]School of Bioscience, Indian Institute of Technology Kharagpur, India

8   [3] Department of Biotechnology, Indian Institute of Technology Kharagpur, India

9

10   *To whom correspondence should be addressed:

11

12   Gayatri Mukherjee,

13   School of Medical Science and Technology, Indian Institute of Technology Kharagpur, West Bengal -

14   721302, India Phone: 032222284902,

15   Email: gayatri.mukherjee@smst.iitkgp.ac.in

16

17   ¶These authors have contributed equally to the work and share first authorship.

18

19

20

21

Abstract

Diseases caused by Dengue (DENV) and Zika (ZIKV) viruses cause significant mortality and illness globally. Due to the high sequence similarity of the viral proteins and the purported cross-reactive immune responses against the viruses, we envisioned a common multi-epitope vaccine (MEV) against both viruses by adopting a novel approach of identifying "immunogenic hotspots". These stretches of the structural and non-structural proteins are enriched with MHC class I and class II supertype-restricted T cell epitopes, and B cell epitopes, in addition to being highly conserved between different DENV serotypes and ZIKV. Such an approach ensures inclusion of multiple overlapping T and B cell epitopes common to both viruses, and also warrants high population coverage. Importantly, epitopes known to cause antibody-dependent-enhancement of infection have been excluded. These immunogenic hotspots have then been stitched together with linkers in-silico along with an adjuvant, CTxB to develop the MEV candidate. Four structural models of the MEV were selected on the basis of conformational preservation of CTxB, and their biophysical parameters, which also conserved the immunogenicity of the multiple epitopes. Importantly, each of the MEV candidates were found to interact with TLR4-MD2 complex by molecular docking studies, indicative of their ability to induce TLR-mediated immune responses.

Keyword: Dengue virus, Zika virus, Mutli-epitope rich immunogenic Hotspots, common multi-epitope vaccine

Introduction

Vector-borne viral diseases, like Dengue fever and Zika viral fever cause severe illness, significant mortality and resultant socio-economic burden worldwide, thus proving to be a major public health concern of the modern era. The Dengue virus (DENV) has four serotypes (DENV1-4), which are responsible for dengue fever, dengue hemorrhagic fever, and dengue shock syndrome in humans[1]. Globally, it is the fastest-growing mosquito-borne disease, causing nearly 400 million infections every year with a mortality rate of 20% when untreated (https://www.who.int/en/news-room/fact-sheets/detail/dengue-and-severe-dengue). On the other hand, the Zika virus (ZIKV) causes the zika viral fever which, in addition to causing headache, muscle and joint pain, rash, edema of extremities, retro-orbital pain and conjunctival hyperemia [2],often lead to congenital abnormalities in unborn fetus [3]. Due to such long-lasting damage by ZIKV, WHO declared this virus a public health emergency of international concern in 2016.

Being members of the genus Flavivirus, DENV and ZIKV share some common genetic as well as phenotypic features. Both are borne by different species of the *Aedes* mosquito, and have approximately 11 kb+ ssRNA genome that encodes three structural and seven non-structural proteins. The three structural proteins are capsid (C), envelope (E) and the membrane glycoprotein (M); while the seven non-structural proteins are called NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5[4,5]. Importantly, the 4 serotypes of DENV share a high genetic similarity with ZIKV, especially in the E protein and the NS3 and NS5 proteins[6].

The B cell [7,8] and T cell [9,10,11]mediated immune response against different DENV serotypes show significant cross reactivity not only among themselves but also with ZIKV and vice versa. Such cross-reactive immune response often leads to Antibody Dependent Enhancement (ADE) of infection and disease virulence in successive natural infections with different serotypes of DENV following the first infection. Additionally, ADE has also been reported in ZIKV infections following prior DENV infections and vice versa [12,13,14]. This could be due to binding of the non-neutralizing or weakly neutralizing, but cross-reactive antibodies from the first infection, to the heterotypic virus

3

71   particles, which could lead to increased viral uptake by monocytes and macrophages via their Fcγ

72   receptors, thus resulting in increased viral load [15,16].Reports have shown that Dengvaxia, the only

73   licensed vaccine against DENV currently available, poses considerable risk of ADE of infection upon

74   encounter with dengue serotypes in vaccinated individuals who were seronegative at the time of

75   vaccination [17,18,19]. Therefore, it is imperative to design and develop a new generation vaccine

76   which will be able to protect against all DENV serotypes and ZIKV, while circumventing the

77   deleterious effects of ADE.

78   In the current study, we first identified the viral proteins from all four DENV serotypes and

79   the Asian strain of ZIKV that have significant antigenic probability and phylogenetic relatedness.

80   These proteins were then probed for T cell epitopes restricted to MHC class I and class II supertypes,

81   instead of individual MHC alleles, thereby ensuring a high population coverage, to account for the

82   high degree of polymorphisms in the HLA alleles. Next, we proceeded to map the T cell epitopes on

83   the selected proteins, and identified amino acid stretches of these proteins that- a) contained multiple

84   T cell epitopes restricted to multiple HLA supertypes; b) was present on all four DENV serotypes as

85   well as ZIKV; and c) was highly conserved among the selected DENV strains and ZIKV. These

86   amino acid stretches were henceforth referred to as "immunogenic hotspots". The hotspots were

87   further probed for the presence of linear and conformational B cell epitopes that were completely

88   conserved in all five viral strains. However, we excluded the immunogenic hotspots that contained B

89   cell epitopes previously reported to be responsible for ADE among DENV serotypes and/or ZIKV, to

90   minimize the deleterious side effects of the vaccine candidate. With this strategy, we identified four

91   completely conserved, multi-epitope rich immunogenic hotspots, which were then joined together

92   with linkers in silico along with CTxB, a well-established adjuvant used for multi-epitope vaccine

93   (MEV) design. Three-dimensional structures of the MEV were predicted with different order of the

94   immunogenic hotspots, using trRosetta server(https://yanglab.nankai.edu.cn/trRosetta/). Models with

95   the highest structural preservation of the CTxB adjuvant were further selected, followed by

96   calculation of the free energy of folding of each model. Thereby four MEV models have been

97   predicted, which were also found to interact with TLR4-MD2 complex by molecular docking studies.

98 To the best of our knowledge, this work represents the first attempt to utilize conserved immunogenic

99 hotspots in the viral protein sequences, to engineer a common MEV candidate with the potential to

100 protect against DENV serotypes and ZIKV, with high population coverage.

101

Methods

Sequence retrieval

104 The protein sequences for all the four serotypes of DENV (DENV1, DENV2, DENV3 and

105 DENV4) together with the Asian lineage of ZIKV were retrieved in FASTA format from ViPR (Virus

106 Pathogen Resource Database, https://www.viprbrc.org/brc/home.spg?decorator=vipr). These include

107 the structural proteins, viz. Membrane Glycoprotein (M), Membrane Glycoprotein Precursor (PreM),

108 Anchored Capsid Protein (AncC), Capsid Protein (C) and Envelope Protein (E); and seven non-

109 structural proteins for all the above-mentioned strains. The GenBank accession numbers are

110 KJ755855 (DENV1), KY427085 (DENV2), KU216209 (DENV3), KX845005 (DENV4) and

111 LC21972 (ZIKV).

Determination of antigenicity

113 Next, the antigenicity of all the proteins was evaluated using the VaxiJen 2.0 server

114 (http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html). Unlike most other prediction methods

115 which are based on alignment of protein sequences to known antigenic proteins, VaxiJen is based on

116 an alignment independent prediction method which takes into consideration the individual properties

117 of the amino acids in the protein sequence[20]. The prediction method takes into account protein

118 classifications, structural and functional similarities between peptides, principal properties of the

119 amino acid itself and those in its neighborhood, to give a more specific and accurate antigenic

120 probability. The protein sequences were fed into the server in FASTA format and the threshold value

121 of antigenicity was set at the recommended cutoff of 0.4, having highest accuracy for viral proteins.

122   The proteins with the antigenicity score more than the set cutoff were considered as most probable

123   antigens.

Multiple sequence alignment and phylogenetic analysis

125   The protein sequence of all the 13 proteins common for all 4 DENV serotypes and ZIKV was

126   subjected to multiple sequence alignment (MSA) to find out the conserved sequences. Clustal Omega

127   server (https://www.ebi.ac.uk/Tools/msa/clustalo/)was used for the alignments; this server utilizes

128   pairwise alignment methods followed by sequence clustering to get distance scores. Following this, a

129   guide tree is constructed and further alignments are done by progressive alignment packages to give

130   the final output and a specific phylogenetic tree of the proteins that were analyzed[21]. Boxshade

131   server (https://embnet.vital-it.ch/software/BOX_form.html) was used for the representation of the

132   MSA data. The amino acid at a particular position was considered conserved if present in the protein

133   sequences of four or more viral strains under consideration. The phylogenetic trees were used to

134   analyze the phylogenetic hierarchy between the proteins from the viral strains.

Prediction of MHC class I and II restricted T cell epitopes

136   Following the VaxiJen analysis, MSA and phylogenetic analysis; three proteins, (E, PreM and

137   NS5) were considered the most antigenic. The MHC class I and II epitopes were predicted for these

138   proteins using the IEDB server(IEDB.org: Free epitope database and prediction resource). The

139   extremely polymorphic MHC molecules called HLA (human leukocyte antigens) in humans, can be

140   clustered on the basis of their main anchor specificity into 9 groups (in case of MHC class I) and 7

141   groups (with respect to MHC class II) called supertypes[22,23]. Between the supertypes, global

142   population coverage is found to be nearly 100%. Therefore, mapping supertype based epitopes gives

143   us the maximum opportunity to mine out conserved immunodominant regions that can be useful in

144   designing a common vaccine. The latest IEDB recommended 2.22 prediction method was used for

145   epitope prediction for the nine MHC class I allele supertypes, viz. HLA- A01, A02, A03, A24, B07,

146   B27, B44, B58, and B60. For MHC class II, IEDB recommended prediction method was used to

147 predict epitopes for a predefined set of MHC class II alleles that gives maximum population

148 coverage[24]. The epitope length was set as 9mer and 15mer for MHC class I and class II molecules

149 respectively. The output result was in the form of percentile rank and the lower the rank of an epitope

150 the stronger the binding of that particular epitope. A cutoff of 2.0 percentile rank was used to get the

151 best predicted epitopes which was then used to compute the immunodominant regions of the proteins.

Prediction of linear and conformational B cell epitopes

153 Linear B cell epitopes were predicted using the ABCpred server

154 (https://webs.iiitd.edu.in/raghava/abcpred/). The FASTA sequence of each protein was submitted to

155 get the predicted epitopes. The threshold score was set at 0.51 (between 0.1 to 1) and length of epitope

156 was set as 16mer because the server shows 65.93% accuracy with equal sensitivity and specificity at

157 this default set up[25]. The highly specific predicted epitopes with a score >0.75 were selected for

158 further analysis.

159 For prediction of conformational B cell epitopes, Discotope 2.0 server

160 (http://tools.iedb.org/discotope/) was used. The three-dimensional structures (resolution <5 Å) of the

161 envelope protein, PreM and NS5 protein of different DENV serotypes and ZIKV were retrieved from

162 Protein Data Bank (https://www.rcsb.org/) and fed to the Discotope server. The threshold value is set

163 at the recommended -3.7 cutoff, which corresponds to 75% specificity[26].

Identification of epitope rich "Immunogenic Hotspots"

165 The predicted CD4 and CD8 T cell specific epitopes were further mapped on to the selected

166 proteins. Upon doing so, it was found that certain stretches of the amino acid sequences have given a

167 higher number of overlapping epitopes restricted to multiple MHC alleles. Among these, few amino

168 acid stretches were found in all four serotypes of DENV and ZIKV. These common immunodominant

169 regions were referred to as "immunogenic hotspot". Further, the epitopes obtained from these

170 immunogenic hotspots were mapped back to the consensus sequences generated from multiple

171   sequence alignment to identify the conserved epitope-rich stretches for each selected protein across all

172   viral strains.

173   Similarly, the predicted linear B cell epitopes were also analyzed to identify epitope-rich

174   immunodominant regions. These immunodominant regions were then further mapped back to the

175   conserved stretches of T cell epitopes. The conserved amino acid stretches that included common T

176   cell and B cell epitopes in all the DENV serotypes as well as ZIKV were identified. These stretches of

177   amino acids are henceforth referred to as immunogenic hotspots, which have the potential to elicit

178   both T cell and B cell mediated immune response, thereby making them lucrative candidates for MEV

179   design.

Multi-epitopevaccine design

181   The conserved immunogenic hotspots were further used for the designing of multi epitope

182   peptide vaccine (MEV). We have selected four immunogenic hotspot sequences derived from E and

183   NS5 protein. These four immunogenic stretches were linked together with GPGPGPG linkerswhich

184   minimizes conformational changes in the tertiary structure and provides flexibility to the peptide

185   chains. Peptide adjuvant CTxB (GenBank Id WP_000593519.1), the B subunit of cholera toxin was

186   included in the N terminus of the MEV using rigid EAAAK linkers. The 103 amino acid long

187   functional domain of the CTxB interacts with the GM1 lipid present on the plasma membrane thereby

188   facilitating the entry of the MEV candidate into antigen presenting cells (APC)[27,28]. Further, this N

189   terminal CTxB adjuvant can also activate the TLR mediated immune response by interacting with

190   TLR4[29], thus making it a suitable peptide adjuvant. To predict the most stable structure of the

191   MEV, the four immunogenic stretches were marked as A, B, C & D and joined one after another in all

192   possible combinations keeping the adjuvant fixed at N terminus which gives a total of 24 models,

193   each 198 amino acids long. Furthermore, the tertiary structures of the MEV candidates were predicted

194   using the trRosetta server (https://yanglab.nankai.edu.cn/trRosetta/). trRosetta builds the protein

195   structure based on direct energy minimizations with a restrained Rosetta. The restraints include inter-

196   residue distance and orientation distributions, predicted by a deep residual neural network[30]. Each

197   combination gave one predicted tertiary structure, which was named as model 1 to model 24. They

198   were further aligned with native structure of CTxB (PDB ID 5LZG) using Pymol software (The

199   PyMOL Molecular Graphics System, Version 1.1 Schrödinger, LLC.), and RMSD values were

200   calculated. Models which have aligned more than 80% of the CTxB's Cα atoms with its native

201   structure and also have RMSD values <1 Å were further selected and checked for stability by

202   calculating Gibbs free energy for folding. The free energy for folding of each model was calculated

203   using the FoldX tool (http://foldxsuite.crg.eu/command/Stability). These parameters are reflective of

204   the overall stability of the predicted protein. Each of the four finally selected MEV models were

205   further docked with TLR4 (PDB ID: 3FXI) using the ZDOCK server version 3.0.2

206   (https://zdock.umassmed.edu/) [31]. Such interaction would be indicative of uptake of the MEV by

207   immune cells leading to protective responses. The results of docking were visualized using PyMOL

208   software.

209

Results and Discussion

Selection of proteins for epitope prediction

212   DENV and ZIKV belong to the genus Flavivirus and have common structural proteins

213   consisting of conserved regions. In case of DENV and ZIKV infection, it is observed that neutralizing

214   antibodies are generated against E, M and its precursor PreM, C and anchored capsid protein

215   (AncC)[32,33,34]. Additionally, T cell mediated immune response in DENV and ZIKV infections has

216   been found to be mostly generated against the non-structural proteins[9,35]. In this study, instead of

217   selecting highly immunogenic individual T and B cell epitopes as vaccine candidates, we aimed to

218   identify T and B cell epitope-rich stretches of both structural and non-structural proteins that are

219   highly conserved among the four DENV serotypes and ZIKV, and formulate a multi-epitope vaccine

220   candidate with these conserved "immunogenic hotspots" of the virus proteome. To select the

221   structural proteins, we used the VaxiJen server which identifies the most probable antigens in an

9

222   alignment independent manner and considers the properties of the amino acid composition of the

223   protein, protein classification, structural and functional similarities of the peptides. VaxiJen

224   recommends a cut-off of 0.4 in order to select the most probable viral antigens. To satisfy our

225   approach of developing a common vaccine against all four DENV serotypes and ZIKV, we needed to

226   consider the antigenicity of the viral proteins in each individual serotype. Accordingly, we first

227   applied the cutoff to each structural protein (E, M, PreM, C and AncC) of all the DENV serotypes and

228   ZIKV individually, in addition to calculating the average antigenic probability, the cut-off for which

229   was set at 0.6. Only the E (Average antigenicity score =0.6091) and PreM (Average antigenicity score

230   =0.684) satisfied both criteria and therefore were selected for further studies (**Table 1**). This is

231   consistent with the observation that envelope protein induces a strong antibody response against

232   ZIKV[34] and all serotypes of DENV[33]. The cleavage of the PreM into functional M protein is not

233   entirely efficient and as a result, 30-40% of mature virions have PreM protein on their surface[36].

234   This leads to a considerable antibody response being generated against the PreM protein[37]and thus

235   justifies the selection of the PreM protein at this stage.

236

237   **Table 1**: Antigenic probability of the structural proteins as determined by the VaxiJen server.

Proteins	Antigenic Probability					Average Probability
	DENV1	DENV2	DENV3	DENV4	ZIKV	
Anchored Capsid Protein (AncC)	0.4003	0.3877	0.4335	0.2336	0.4677	0.3846
Capsid Protein (C)	0.3778	0.3693	0.3641	0.1636	0.3856	0.332
Membrane Glycoprotein Precursor (PreM)	0.6391	0.5796	0.6585	0.6273	0.541	0.6091
Membrane Glycoprotein (M)	0.6168	0.502	0.5769	0.5101	0.38	0.5172

Envelope Protein (E)	0.7495	0.6695	0.7367	0.6442	0.6201	0.684

238

239   Like the structural proteins, the non-structural proteins were also investigated on the VaxiJen

240   server to determine their antigenicity. The results showed that all the non-structural proteins qualified

241   the 0.4 cutoff and could be considered probable antigens (**S1Table**). Proteins with high sequence

242   homology and phylogenetic relatedness have greater probability of generating similar or identical

243   epitopes across the species under consideration. In flavivirus, it has been shown that phylogenetic

244   relatedness plays a role in antigenicity, wherein individual distinct clades generate identical epitopes

245   and therefore such relatedness must be considered while predicting the common epitopes[38].

246   Therefore, to select the best candidates from the non-structural proteins of the DENV serotypes and

247   ZIKV, we probed these proteins for sequence homology determination and phylogenetic relatedness.

248   Our studies as well as already published data [6]show that NS3 and NS5 have the highest sequence

249   homology(greater than 60% homology) between ZIKV and all the DENV serotypes. Therefore, we

250   selected these two proteins for further analysis of phylogenetic relatedness. The NS5 protein of ZIKV

251   and DENV4 share three common ancestors and the pair shares two common ancestors with DENV2,

252   indicative of high degree of relatedness between these three viral species. However, NS5 of DENV1

253   and DENV3 are phylogenetically separated from these three (**Fig 1(A)**). In comparison, NS3 of ZIKV

254   and DENV4 share two common ancestors, while that of DENV1 and DENV3 share a separate pair of

255   common ancestors, reflective of the greater phylogenetic difference between these two clades.

256   Additionally, further divergence is seen with the NS3 protein of DENV2 from the rest of the members

257   (**Fig 1(B)**). Thus, the data demonstrates that NS5 is more closely related across all the viral species

258   than NS3 and hence it was selected as a candidate for further vaccine design.

259

260   **Fig 1:Phylogenetic tree(A)** NS5 and **(B)** NS3 protein as computed by CLUSTALW server

261

11

262 # Identification of MHC class I restricted epitopes and

263 # 'Immunogenic Hotspots' from the selected proteins

264   Based on the antigenicity score and phylogenetic relationship, E, PreM, and NS5 were

265 selected for prediction of MHC class I restricted human CD8 T cell specific epitopes from the IEDB

266 server. The peptide binding cleft of MHC class I can accommodate 8 to 10mer peptides and therefore

267 9mer peptides were predicted for the nine MHC class I allele supertypes giving maximum population

268 coverage. After analyzing the identified epitopes, it was found that there are distinct amino acid

269 stretches that contain a high number of overlapping epitopes restricted to the different HLA

270 supertypes. These epitope rich regions of the proteins are termed as "immunogenic hotspots". It was

271 seen that in case of PreM, there was one amino acid stretch that was dominant across all viral species,

272 and contained multiple MHC class I supertype restricted epitopes (**Table 2**). However, upon further

273 analysis, this immunogenic stretch did not yield a conserved peptide stretch that can be used as a

274 hotspot peptide in our MEV candidate.

275

276 **Table 2**: Immunogenic hotspots enriched with multiple MHC class I restricted epitopes from PreM, E

277 and NS5 proteins across all 5 viral species

Virus	Protein	Position	Number of predicted epitopes restricted to MHC class I supertypes
DENV1	PreM	103-165	16
DENV2		103-165	27
DENV3		109-165	18
DENV4		103-165	21
ZIKV		116-165	17

12

DENV1	E	200-226	12
		291-310	6
DENV2		196-240	16
		286-310	10
DENV3		194-244	17
		289-308	6
DENV4		195-231	14
		291-306	6
ZIKV		195-236	10
		291-306	10
DENV1	NS5	443-514	18
		749-769	10
		362-386	9
DENV2		447-525	17
		750-769	8
		372-388	5
DENV3		438-514	22
		750-769	9
		364-401	6
DENV4		450-521	21
		751-778	11
		365-388	5
ZIKV		468-498	11
		752-778	10
		373-384	5

278

13

279 In contrast, the E protein yielded 2 such dominant immunogenic hotspot stretches **(Table 2)**.

280 Interestingly, E protein, better known for eliciting humoral immune response, gave a 15mer conserved

281 sequence **WLVHKQWFLDLPLPW** within the 200-226 hotspot stretch **(Fig 2(A))**. This conserved

282 sequence has a few amino acid substitutions where L205M, K208R and L212F in DENV3, K210R in

283 DENV2and Q216E, L219H and L221I in ZIKV. Since, the majority of these substitutions are with

284 biochemically similar amino acids and are not in the anchor regions of the epitopes generated from

285 the stretch, this particular hotspot was selected for our MEV candidate. The second hotspot of the E

286 protein did not yield any conserved sequence and therefore was not considered further.

287

288 **Fig 2: Conserved Immunogenic hotspots identified from the E and NS5 proteins of all DENV**

289 **serotypes and ZIKV.(A)** 15mer conserved hotspot from E protein having single position amino acid

290 differences highlighted. **(B)** 12mer sequence from NS5 protein having 100% identity and overlapping

291 epitopes. **(C)** 15mer sequence from NS5 protein having 100% identity and overlapping epitopes. **(D)**

292 Another 15mer hotspot from NS5 protein having single position amino acid differences highlighted.

293 **(E)** 12mer conserved hotspot from NS5 protein having single position amino acid differences

294 highlighted.

295

296 NS5, which is known to elicit strong T cell responses, gave three immunogenic hotspot

297 stretches **(Table 2)** that were found in the similar regions across all viral species. Upon investigation

298 of the these stretches, the 450-500 amino acid stretch of NS5 yielded a 12mer **CVYNMMGKREKK**

299 and 15mer **AKGSRAIWYMWLGAR** sequence that was fully conserved containing overlapping

300 epitopes**(Fig 2(B), 2(C))**. Another 15mer sequence **FLEFEALGFLNEDHW** was obtained from the

301 same hotspot, but with two highly-related substitutions (F482Y in DENV3 and L491M in DENV1)

302 **(Fig 2(D))**. A fourth 12mer conserved sequence **GWSLRETACLGK** was mined out from the

303 hotspot stretch 750-769 **(Fig 2(E))**. This sequence had a substitution of two amino acids (L750I and

304 G757A) specifically in ZIKV which are biochemically similar as well. Moreover, these substitutions

14

305 are not in the anchor region of the epitopes (important for the formation of stable pMHC complexes

306 and antigen presentation by APC) that are generated from these hotspots. Thus, these four

307 "Immunogenic Hotspots" were selected as potential candidates for generation of multi-epitope

308 vaccine.

309 Identification of MHC class II restricted epitopes and 310 'Immunogenic Hotspot' from the selected proteins

311   The MHC class II peptide binding cleft, unlike its class I counterpart, is more open allowing a

312 peptide with higher length and various side chains to interact with the T cell receptor. An optimum of

313 13 to 17mer peptides is presented by the MHC class II molecules. Therefore, we predicted 15mer

314 epitopes for the MHC class II HLA molecules reference set which has the highest population

315 coverage available on the IEDB server. The predicted epitopes were analyzed to map the

316 immunogenic hotspots. PreM and E proteins each had one hotspot stretch (**Table 3)** that was common

317 for all the viral species. However, upon analysis by multiple sequence alignment, these stretches did

318 not give any conserved peptides for MHC II binding that could be further considered for MEV

319 candidate designing. It is to be noted, that although there was partial overlap of the E protein hotspot

320 with the conserved MHC class I peptide (**WLVHKQWFLDLPLPW**), the overlapping sequence was

321 shorter than the minimum peptide length necessary for MHC class II binding.

322

323 **Table 3**: Immunogenic hotspots enriched with multiple MHC class II restricted epitopes from PreM,

324 E and NS5 proteins across all 5 viral species

Virus	Protein	Position	Number of predicted epitopes restricted to MHC class II supertypes
DENV1	PreM	130-166	19
DENV2		123-165	58

DENV3		129-166	32
DENV4		121-166	67
ZIKV		123-168	63
DENV1	E	188-224	21
DENV2		189-224	23
DENV3		186-222	20
DENV4		190-224	28
ZIKV		192-227	28
DENV1	NS5	221-254	23
		374-418	23
		445-495	42
		558-586	24
		754-783	12
DENV2		200-252	60
		398-416	9
		446-497	47
		558-585	27
		752-784	17
DENV3		159-254	58
		397-434	26
		445-496	51
		575-581	26
		752-784	17
DENV4		222-241	14
		376-420	34
		446-498	48
		559-589	25

		753-786	19
ZIKV		224-242	9
		395-422	27
		447-499	48
		545-584	12
		756-783	7

325

326   The NS5 protein gave five immunogenic hotspots, enriched with MHC class II restricted

327   peptides, with three of them coinciding with the hotspot stretches enriched with MHC class I

328   restricted peptides as mentioned in the previous section. Additionally, two hotspot stretches within the

329   150-250 and 550-590 range having overlapping epitopes was mined out **(Table 3)**.Further

330   investigation revealed a conserved sequence **AKGSRAIWYMWLGARFLEFEALGFLNEDHW**

331   having overlapping epitopes **(Fig 3).** This 30mer sequence essentially is a combination of the two

332   15mer hotspots **(Fig 2(C), 2(D))** that were obtained in the case of MHC class I epitopes having

333   identical amino acid replacements. This makes this 30mer hotspot stretch a prime candidate in MEV

334   vaccine design capable of generating a strong $CD4^+$ and $CD8^+$ T cell response and thereby result in

335   generation of effective memory T cells.

336

337   **Fig 3: 30mer conserved sequence obtained from NS5 protein with single amino acid differences**

338   **highlighted.**

339

340   # Prediction of linear B cell epitopes and identification of

341   # 'Immunogenic Hotspot' from selected proteins

342   Like T cell epitopes, B cells epitopes were also identified from the candidate proteins. Online

343   epitope prediction tool ABCpred was used for linear B cell epitope prediction of the selected proteins.

344   The 16mer epitopes with threshold value more than 0.75 were selected and further analyzed to

345   identify the immunogenic hotspots.

346   DENV E and M proteins elicit an initial B cell response that generates neutralizing antibodies

347   against these two proteins. These antibodies against PreM and E proteins have been reported to

348   facilitate secondary infection by a different serotype and increase the viral load via ADE[32,37]. The

349   E protein contains a 15mer immunodominant region at position 97-111 called **fusion peptide loop**

350   (FPL), and antibodies to this highly conserved region frequently show broad flavivirus cross-

351   reactivity resulting in ADE[39,40].Though this region has generated linear epitopes for all DENV

352   serotypes and ZIKV in our studies, but considering its contribution in ADE, it is excluded as an

353   immunogenic hotspot. Furthermore, although an immunodominant region was observed in the E

354   protein, no conserved stretches of considerable length were observed (**S2Table**). Our observation of

355   the conformational B cell epitopes generated from the FPL is described in later sections.

356   Similar analysis of the PreM protein demonstrated that each serotype of DENV and ZIKV had

357   immunogenic stretches in their sequences that contain multiple linear 16meric B cell epitopes

358   (**S2Table**). However, none of these stretches showed significant overlap. Interestingly, experimental

359   data has reported the generation of at least three antibodies from the PreM protein; one of which is

360   against an identical epitope in ZIKV and DENV serotypes, while the other two are specific to either

361   DENV or ZIKV respectively[41]. The epitope common to both DENV and ZIKV encompasses amino

362   acids derived from position 56-72 of the PreM protein. However, the sequence is not 100% conserved

363   among the viral species, with the conserved sequences being confined to positions 61-70

364   (**EPdDvDCWCN**) only, with similar amino acid substitutions in position 63 and 65 (denoted with

365   small letter). As discussed earlier about the effect of ADE specifically for these flavivirus infections,

366   the lack of complete conserved sequence of the epitope may lead to greater probability of such

367   incidence;therefore we refrained from utilizing this epitope for the MEV design.

368 NS5 protein gave maximum number of immunodominant regions for both T cell and linear B

369 cell epitopes (**S2Table**), many of them are in a similar position of the protein sequence. Regions like

370 440-500, 514-570, and 720-760 have generated B cell epitopes as well as part of both CD4 and CD8

371 T cell epitopes. The 30mer conserved immunogenic stretch

372 (**AKGSRAIWYMWLGARFLEFEALGFLNEDHW**) (**Fig 3**) previously described also contains

373 multiple linear B cell epitopes. This indicates that this region of NS5 is important for both cell

374 mediated and humoral immune response.

Conformational B cell epitopes

376 High resolution (<5Å) and full length three dimensional structures are imperative for structure

377 based computational work including prediction of conformational B cell epitopes. Despite extensive

378 structural analysis of the DENV and ZIKV proteins, there is a lack of fully solved 3-dimensional

379 structures of these proteins in protein data bank (PDB). Most of the available data are either for E

380 protein or the RNA dependent RNA polymerase domain of NS5; both of which have been used for the

381 prediction of conformational B cell epitopes conserved among all serotypes in this study. 3-D

382 structures of either the PreM or M proteins are not available in PDB for all serotypes of the viruses

383 and therefore, these two proteins were excluded from the study. Taking into account the above

384 considerations, conformational B cell epitopes of E and NS5 proteins of DENV serotypes and ZIKV

385 were identified with the help of Discotope 2.0 server using the available 3D structures from PDB.

386 Only a few regions of the E protein were found to give consistent conformational epitopes.

387 Prominent among them, position 100-102 (GWG) of E protein of ZIKV (PDB ID: 5IRE) and all

388 DENV serotypes except DENV1 (PDB ID: 3G7T) was predicted as epitope. These amino acids are

389 part of the FPL of flavivirus[39,40] which is involved in cross reactivity/ADE among flavivirus

390 members. We further investigated the number of contact residues, which is the number of Cα atoms in

391 the protein within a distance of 10 Å of the residue's Cα atom, for each epitope. Interestingly, the

392 contact residues of the amino acids constituting the GWG epitope was different for each DENV

393 serotype as well as ZIKV (**S3Table**). Importantly, a similar analysis of the other residues of the FPL,

19

394   wherever applicable, also demonstrated considerable diversity in the number of contact residues. This

395   could indicate a conformational difference in this region among the viral serotypes of DENV and

396   ZIKV, which may lead to weak interaction of antibodies raised against the FPL of one serotype with

397   the FPL of another,resulting in ADE. Another sequence, 341-346 was predicted as epitopes in all

398   serotypes of DENV, albeit with variable sequence, but not in ZIKV.

399   Next, we analyzed the conformational epitopes derived from the NS5 protein of ZIKV,

400   DENV2 and DENV3. It is to be noted here that all of the reported structures of the NS5 protein of

401   DENV3 are shorter than actual protein size (899 amino acid residues long), with many of them

402   missing internal sequences. Moreover, due to a complete lack of NS5 structures of DENV1 and

403   DENV4, we could only include the DENV2 and DENV3 NS5 protein structures in our search for B

404   cell conformational epitopes. The conformational epitopes obtained, barring a few, have low sequence

405   homology (**S4Table**). Among them, sequence 14-38 of ZIKV NS5 was found to be overlapping with

406   a common region of predicted conformational epitopes of NS5 protein of DENV 2 and DENV 3

407   where 23E, 24F, and 27-31 (YKKSG) are conserved. Sequence 107-109 (GPG) was also conserved

408   and predicted as an epitope for both ZIKV and the DENV serotypes. Region 303-318 of ZIKV NS5

409   also have common embedded conserved epitopes with DENV2 and DENV3, such as 303-305 (TWA),

410   310Y, 311E. Region 346-369 of ZIKV has a 9mer stretch (VFKEKVDTR) which is also a part of B

411   cell epitope in all DENV serotype under investigation. Similarly, 719-725 and 817-834 has two 4-mer

412   conserved sequences: 721-724 (KDGR) and 828-831(DKTP). However, like the E protein, we further

413   observed the contact residues for each epitope predicted from NS5 (**S4Table**) and likewise found that

414   in majority of the cases, the number of contact residues for each position are different for DENV

415   serotypes and ZIKV. This indicates a difference in conformation of NS5 protein among the serotypes

416   of DENV and ZIKV.

417   Thus, for both the proteins, conformational B cell epitopes, even when conserved, indicated

418   structural dissimilarities, which could translate into variable affinity of cross-reactive antibodies.

419   Since our aim was to design a single MEV which would be able to induce T and B cell responses

420   without causing ADE, it seemed prudent not to include such conformational epitopes in the candidate.

20

421

Multi-epitope vaccine design

423   From the above observations and data, it was clear that the MEV candidate must be designed

424   using the conserved immunogenic hotspots derived from E and NS5 protein of DENV and ZIKV that

425   includes MHC class I and MHC class II restricted T cell epitopes and linear B cell epitopes(**Table 4**).

426   To increase the immunogenicity and the probability of uptake by cells, the adjuvant CTxB was

427   included in the N terminus of the MEV. The epitopes were connected with GPGPGPG linker and the

428   adjuvant was connected to the epitopes by EAAAK linker.

429

430

431   **Table 4**: "Immunogenic Hotspots" selected for designing of multi-epitope vaccine from the E and

432   NS5 protein of DENV serotypes and ZIKV

Sequence of conserved Immunogenic Hotspots	Protein	Specificity	Denoted as
AKGSRAIWYMWLGARFLEFEALGFLNEDHW	NS5	CD4, CD8 & B cell	A
CVYNMMGKREKK	NS5	CD8 T cell	B
GWSLRETACLGK	NS5	CD8 T cell	C
WLVHKQWFLDLPLPW	E	CD8 T cell	D

433

434   The four immunogenic hotspots were marked alphabetically A, B, C, & D (**Table 4**) and

435   arranged in all possible combinations keeping the CTxB adjuvant at N terminus. The 24 MEV

436   combinations thus generated, were then used for tertiary structure prediction using the trRosetta

437   server. Each combination gave one best predicted structure in PDB format which were named model

438   1 to 24 and was further evaluated to determine the best candidate(s). The tertiary structures of the

439   terminal adjuvant CTxB is important for the MEV's internalization into the host APC and thus must

440 remain in its native three-dimensional form. Therefore, the 24 models were aligned with the reported

441 structure of CTxB (PDB ID 5LZG) and the RMSD values were calculated. Models with more than

442 80% (greater than 82) of the Cα atom aligned with the structure of CTxB and having RMSD less than

443 1 Å were selected to get the highest possible conserved structures of CTxB. Five models (**8, 9, 13, 19**

444 **and 21**) out of 24 that qualified the criteria (**S5 Table**) were further checked for their stability by

445 calculating the Gibbs free energy for folding. This was done using the FoldX server which is used for

446 measurement of protein stability from its 3D structure, by calculating the change in relative free

447 energy for folding of the protein from its unfolded state[42,43]. Except for model 13, the other four

448 models have shown <100 Kcal/mole for folding, indicating a highly stable structure (**S6 Table**). **Fig 4**

449 shows the predicted tertiary structures of the four models (**8, 9, 19 and 21**) as visualized in Pymol.

450 Molecular docking of the individual MEV candidates with TLR4 showed thateach candidate MEV

451 interacts with different regions of the TLR4 and its associated complex MD2 (**Fig 5**). This is

452 indicative of the potential of the selected MEV candidates to activate TLR mediated immune

453 response.These four vaccine candidates were further probed for T and B cell epitope generation using

454 the respective epitope-prediction tools discussed above and it was found that the highly immunogenic

455 T and B cell epitopes have been preserved in these candidates (**S7 Table**).

456 **Fig 4: Predicted tertiary structure of the designed multi-epitope vaccine candidates (8, 9, 19 and**

457 **21)** (colored by rainbow from N to C terminus) (IH- Immunogenic Hotspot)

458 **Fig 5: Molecular Docking of MEV candidates with TLR4/MD2 complex.(A)** Model 8 **(B)** Model 9

459 **(C)** Model 19 **(D)** Model 21. TLR4 is in blue, MD2 is in green and MEV is in red.

460

Conclusion

462 Conventional MEVs are designed to incorporate multiple MHC class I and MHC class II

463 restricted immunodominant T cell epitopes with the potential to activate multiple T cell clones, which

464 can differentiate into cytotoxic and helper T cells respectively. In addition, such MEVs also include B

465 cell epitopes that can activate B cells leading to their differentiation into antibody-producing plasma

466 cells, which can neutralize the pathogens. One of the most important aspects of MEVs is the exclusion

467 of potentially detrimental parts of antigenic determinants that can otherwise initiate unwanted and

468 harmful immune-responses, like ADE. Especially in the case of flavivirus like DENV and ZIKV,

469 which have a high degree of sequence similarity, ADE is a matter of serious concern, with numerous

470 reports of increased virulence of subsequent heterotypic infections. Therefore, any vaccine made

471 against these viruses, must be specifically checked to exclude any potential ADE, while ensuring the

472 extension of the protection to all serotypes concerned.

473   Given the close phylogenetic relationship between DENV serotypes and ZIKV, we ventured

474 to design a common vaccine which will not merely incorporate multiple single T and B cell epitopes,

475 but rather include immunogenic stretches or hotspots, rich in multiple HLA supertype-restricted T cell

476 epitopes and B cell epitopes that are common to the pathogens, as well as completely conserved

477 among species. Only four such hotspots could fulfill the stringent criteria, which also specifically

478 excludedthose stretches that contain known ADE-associated epitopes.Notably, in the predicted MEV

479 candidates, only linear B cell epitopes could be included as the predicted conformational B cell

480 epitopes, were not found to be completely conserved sequences among the considered viruses. More

481 importantly, the contact residues of these predicted conformational epitopes varied in number among

482 the serotypes, which could purportedly lead to differences in antibody affinities against these epitopes.

483 Further investigation needs to be done to ascertain the contribution of this observation to ADE.

484   Furthermore, the designed MEV described here, also incorporates an adjuvant, the CTxB,

485 which can be takenup by cells via the GM1-gangliosides, and has been found to be recognized by

486 TLR4[29], thereby having the ability to initiate TLR mediated signaling pathways that can activate

487 the protective immune response. Lastly, the immunogenic hotspots described in this study include T

488 cell epitopes that are restricted to MHC class I supertypes and MHC class II alleles with maximum

489 population coverage, thus ensuring its efficacy among the global population. In sum, the MEV

490 candidates described here can be considered as a stepping stone towards a common vaccine against

491 DENV and ZIKV.

Acknowledgment

Author contributions

Conceptualization: Gayatri Mukherjee

Data Curation: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Formal Analysis: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Funding Acquisition: Gayatri Mukherjee

Investigation: Dhrubajyoti Mahata, Debangshu Mukherjee, Vanshika Malviya, Gayatri Mukherjee

Methodology: Dhrubajyoti Mahata, Debangshu Mukherjee, Vanshika Malviya, Gayatri Mukherjee

Project Administration: Gayatri Mukherjee

Resources: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Software: Dhrubajyoti Mahata, Debangshu Mukherjee

Supervisor: Gayatri Mukherjee

Validation: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Visualization: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Writing- Original Draft Preparation: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

Writing- Review and Editing: Dhrubajyoti Mahata, Debangshu Mukherjee, Gayatri Mukherjee

24

References

1. World Health Organization. WHO report on global surveillance of epidemic-prone infectious diseases—dengue and dengue haemorrhagicfever. (2001). [Accessed January 5, 2020] http://www.who.int/csr/resources/publications/dengue/CSR_ISR_2000_1/en/

2. Ioos S, Mallet HP, Leparc Goffart I, Gauthier V, Cardoso T, Herida M. Current Zika virus epidemiology and recent epidemics. Med Mal Infect. 2014; 44(7):302-7. doi: 10.1016/j.medmal.2014.04.008. PMID: 25001879.

3. Rather IA, Lone JB, Bajpai VK, Park YH. Zika Virus Infection during Pregnancy and Congenital Abnormalities. Front Microbiol. 2017; 8:581. doi: 10.3389/fmicb.2017.00581. PMID: 28421065.

4. Rice CM, Lenches EM, Eddy SR, Shin SJ, Sheets RL, Strauss JH. Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. Science. 1985; 229(4715):726-33. doi: 10.1126/science.4023707. PMID: 4023707.

5. Wong SS, Poon RW, Wong SC. Zika virus infection-the next wave after dengue? J Formos Med Assoc. 2016; 115(4):226-42. doi: 10.1016/j.jfma.2016.02.002. PMID: 26965962

6. Subramaniam KS, Lant S, Goodwin L, Grifoni A, Weiskopf D, Turtle L. Two Is Better Than One: Evidence for T-Cell Cross-Protection Between Dengue and Zika and Implications on Vaccine Design. Front Immunol. 2020; 11:517. doi: 10.3389/fimmu.2020.00517. PMID: 32269575.

7. Kam YW, Lee CY, Teo TH, Howland SW, Amrun SN, Lum FM, et al. Cross-reactive dengue human monoclonal antibody prevents severe pathologies and death from Zika virus infections. JCI insight. 2017; 2(8):e92428. doi: 10.1172/jci.insight.92428. PMID: 28422757.

8. Stettler K, Beltramello M, Espinosa DA, Graham V, Cassotta A, Bianchi S, et al. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. Science. 2016; 353(6301):823-6. doi: 10.1126/science.aaf8505. PMID: 27417494.

538   9.   Wen J, Shresta S. T Cell Immunity to Zika and Dengue Viral Infections. J Interferon
539   Cytokine Res. 2017; 37(11):475-479. doi: 10.1089/jir.2017.0106. PMID: 29135369.

540   10. Regla-Nava JA, Elong Ngono A, Viramontes KM, Huynh AT, Wang YT, Nguyen AT, et al.
541   Cross-reactive Dengue virus-specific CD8+ T cells protect against Zika virus during
542   pregnancy. Nat Commun. 2018; 9(1):3042. doi: 10.1038/s41467-018-05458-0. PMID:
543   30072692.

544   11. Lim MQ, Kumaran EAP, Tan HC, Lye DC, Leo YS, Ooi EE, et al. Cross-Reactivity and
545   Anti-viral Function of Dengue Capsid and NS3-Specific Memory T Cells Toward Zika Virus.
546   Front Immunol. 2018; 9:2225. doi: 10.3389/fimmu.2018.02225. PMID: 30327651.

547   12. Dejnirattisai W, Supasa P, Wongwiwat W, Rouvinski A, Barba-Spaeth G, Duangchinda T, et
548   al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection
549   with zika virus. Nat Immunol. 2016; 17(9):1102-8. doi: 10.1038/ni.3515. Epub 2016 Jun 23.
550   PMID: 27339099.

551   13. Paul LM, Carlin ER, Jenkins MM, Tan AL, Barcellona CM, Nicholson CO, et al. Dengue
552   virus antibodies enhance Zika virus infection. Clin Transl Immunology. 2016; 5(12):e117.
553   doi: 10.1038/cti.2016.72. PMID: 28090318.

554   14. Kawiecki AB, Christofferson RC. Zika Virus-Induced Antibody Response Enhances Dengue
555   Virus Serotype 2 Replication In Vitro. J Infect Dis. 2016; 214(9):1357-1360. doi:
556   10.1093/infdis/jiw377. PMID: 27521359.

557   15. Halstead SB, O'Rourke EJ. Dengue viruses and mononuclear phagocytes. I. Infection
558   enhancement by non-neutralizing antibody. J Exp Med. 1977; 146(1):201-17. doi:
559   10.1084/jem.146.1.201. PMID: 406347.

560   16. Dejnirattisai W, Jumnainsong A, Onsirisakul N, Fitton P, Vasanawathana S, Limpitikul W, et
561   al. Cross-reacting antibodies enhance dengue virus infection in humans. Science. 2010;
562   328(5979):745-8. doi: 10.1126/science.1185181. PMID: 20448183.

563   17. Sun J, Du S, Zheng Z, Cheng G, Jin X. Defeat Dengue and Zika Viruses With a One-Two

564   Punch of Vaccine and Vector Blockade. Front Microbiol. 2020; 11:362. doi:

565   10.3389/fmicb.2020.00362. PMID: 32265852.

566   18. Halstead SB. Dengvaxia sensitizes seronegatives to vaccine enhanced disease regardless of

567   age. Vaccine. 2017; 35(47):6355-6358. doi: 10.1016/j.vaccine.2017.09.089. PMID:

568   29029938.

569   19. Shukla R, Ramasamy V, Shanmugam RK, Ahuja R, Khanna N. Antibody-Dependent

570   Enhancement: A Challenge for Developing a Safe Dengue Vaccine. Front Cell Infect

571   Microbiol. 2020; 10:572681. doi: 10.3389/fcimb.2020.572681. PMID: 33194810.

572   20. Doytchinova IA, Flower DR. VaxiJen: a server for prediction of protective antigens, tumour

573   antigens and subunit vaccines. BMC Bioinformatics. 2007; 8:4. doi: 10.1186/1471-2105-8-4.

574   PMID: 17207271.

575   21. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of

576   high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;

577   7:539. doi: 10.1038/msb.2011.75. PMID: 21988835.

578   22. Sidney J, Peters B, Frahm N, Brander C, Sette A. HLA class I supertypes: a revised and

579   updated classification. BMC Immunology. 2008; 9:1. doi: 10.1186/1471-2172-9-1. PMID:

580   18211710.

581   23. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of

582   class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a

583   surprising degree of repertoire sharing across supertypes. Immunogenetics. 2011; 63(6):325-

584   35. doi: 10.1007/s00251-011-0513-0. PMID: 21305276.

585   24. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, et al. Peptide binding predictions for

586   HLA DR, DP and DQ molecules. BMC Bioinformatics. 2010; 11:568. doi: 10.1186/1471-

587   2105-11-568. PMID: 21092157.

588   25. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent

589   neural network. Proteins. 2006; 65(1):40-8. doi: 10.1002/prot.21078. PMID: 16894596.

590 26. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions:

591   impacts of method development and improved benchmarking.PLoSComput Biol. 2012;

592   8(12):e1002829. doi: 10.1371/journal.pcbi.1002829. PMID: 23300419.

593 27. Jia S, Huang X, Li H, Zheng D, Wang L, Qiao X, et al. Immunogenicity evaluation of

594   recombinant Lactobacillus casei W56 expressing bovine viral diarrhea virus E2 protein in

595   conjunction with cholera toxin B subunit as an adjuvant. Microb Cell Fact. 2020; 19(1):186.

596   doi: 10.1186/s12934-020-01449-3. PMID: 33004035.

597 28. Davod J, Fatemeh DN, Honari H, Hosseini R. Constructing and transient expression of a gene

598   cassette containing edible vaccine elements and shigellosis, anthrax and cholera recombinant

599   antigens in tomato. Mol Biol Rep. 2018; 45(6):2237-2246. doi: 10.1007/s11033-018-4385-3.

600   PMID: 30244396.

601 29. Phongsisay V, Iizasa E, Hara H, Yoshida H. Evidence for TLR4 and FcRγ-CARD9 activation

602   by cholera toxin B subunit and its direct bindings to TREM2 and LMIR5 receptors. Mol

603   Immunol. 2015; 66(2):463-71. doi: 10.1016/j.molimm.2015.05.008. PMID: 26021803.

604 30. Yang J, Anishchenko I, Park H, Peng Z, Ovchinnikov S, Baker D. Improved protein structure

605   prediction using predicted interresidue orientations.Proc Natl Acad Sci U S A. 2020;

606   117(3):1496-1503. doi: 10.1073/pnas.1914677117. PMID: 31896580.

607 31. Pierce BG, Wiehe K, Hwang H, Kim BH, Vreven T, Weng Z. ZDOCK server: interactive

608   docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics.

609   2014;30(12):1771-3. doi: 10.1093/bioinformatics/btu097. PMID: 24532726.

610 32. Valdés K, Alvarez M, Pupo M, Vázquez S, Rodríguez R, Guzmán MG. Human Dengue

611   antibodies against structural and nonstructural proteins. Clin Diagn Lab Immunol. 2000;

612   7(5):856-7. doi: 10.1128/CDLI.7.5.856-857.2000. PMID: 10973471.

613 33. Lai CY, Tsai WY, Lin SR, Kao CL, Hu HP, King CC, et al. Antibodies to envelope

614   glycoprotein of dengue virus during the natural course of infection are predominantly cross-

615   reactive and recognize epitopes containing highly conserved residues at the fusion loop of

616   domain II. J Virol. 2008; 82(13):6631-43. doi: 10.1128/JVI.00316-08. PMID: 18448542.

28

34. Dai L, Song J, Lu X, Deng YQ, Musyoki AM, Cheng H, et al. Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective AntibodyCell Host Microbe. 2016; 19(5):696-704. doi: 10.1016/j.chom.2016.04.013. PMID: 27158114.

35. Rivino L, Kumaran EA, Jovanovic V, Nadua K, Teo EW, Pang SW, et al. Differential targeting of viral components by CD4+ versus CD8+ T lymphocytes in dengue virus infection. J Virol. 2013; 87(5):2693-706. doi: 10.1128/JVI.02675-12. PMID: 23255803.

36. Junjhon J, Edwards TJ, Utaipat U, Bowman VD, Holdaway HA, Zhang W, et al. Influence of pr-M cleavage on the heterogeneity of extracellular dengue virus particles. J Virol. 2010; 84(16):8353-8. doi: 10.1128/JVI.00696-10. PMID: 20519400.

37. Rodenhuis-Zybert IA, van der Schaar HM, da Silva Voorham JM, van der Ende-Metselaar H, Lei HY, Wilschut J, et al. Immature dengue virus: a veiled pathogen? PLoSPathog. 2010; 6(1):e1000718. doi: 10.1371/journal.ppat.1000718. PMID: 20062797.

38. Frank SA. Immunology and Evolution of Infectious Disease. Princeton (NJ): Princeton University Press; 2002. PMID: 20821852.

39. Allison SL, Schalich J, Stiasny K, Mandl CW, Heinz FX. Mutational evidence for an internal fusion peptide in flavivirus envelope protein E. J Virol. 2001; 75(9):4268-75. doi: 10.1128/JVI.75.9.4268-4275.2001. PMID: 11287576.

40. Stiasny K, Kiermayr S, Holzmann H, Heinz FX. Cryptic properties of a cluster of dominant flavivirus cross-reactive antigenic sites. J Virol. 2006; 80(19):9557-68. doi: 10.1128/JVI.00080-06. PMID: 16973559.

41. Amrun SN, Yee WX, Abu Bakar F, Lee B, Kam YW, Lum FM, Tan JJ, Lim VW, Watthanaworawit W, Ling C, Nosten F, Renia L, Leo YS, Ng LF. Erratum: Novel differential linear B-cell epitopes to identify Zika and dengue virus infections in patients. Clin Transl Immunology. 2020; 9(2):e01118. doi: 10.1002/cti2.1118. PMID: 32099653.

42. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. Nucleic Acids Res. 2005; 33(Web Server issue):W382-8. doi: 10.1093/nar/gki387. PMID: 15980494.

644   43. Buß O, Rudat J, Ochsenreither K. FoldX as Protein Engineering Tool: Better Than Random

645   Based   Approaches?   Comput   Struct   Biotechnol   J.   2018;   16:25-33.   doi:

646   10.1016/j.csbj.2018.01.002. PMID: 30275935.

647

648   # Supporting Information

649   **S1 Table:** Antigenic probabilities of the non-structural proteins as determined by the VaxiJen

650   server.

651   **S2 Table:** Linear B cell immunogenic hotspots identified from PreM, E and NS5 proteins of

652   all DENV serotypes and ZIKV. The stretches in bold depict common regions to MHC class I

653   and MHC class II restricted immunogenic hotspot.

654   **S3 Table:** Number of contact residues for thee predicted epitope (101-103) and the rest of the

655   fusion peptide loop region (97-111) of E protein, as predicted by Discotope 2.0 (PDB Id for

656   each protein is put in bracket).

657   **S4 Table:**Number of contact residues of the conserved epitopes from NS5 protein as

658   predicted by Discotope 2.0(PDB Id for each protein is put in bracket).

659   **S5 Table:**Alignment result between the known structure of CTxB and the predicted MEV

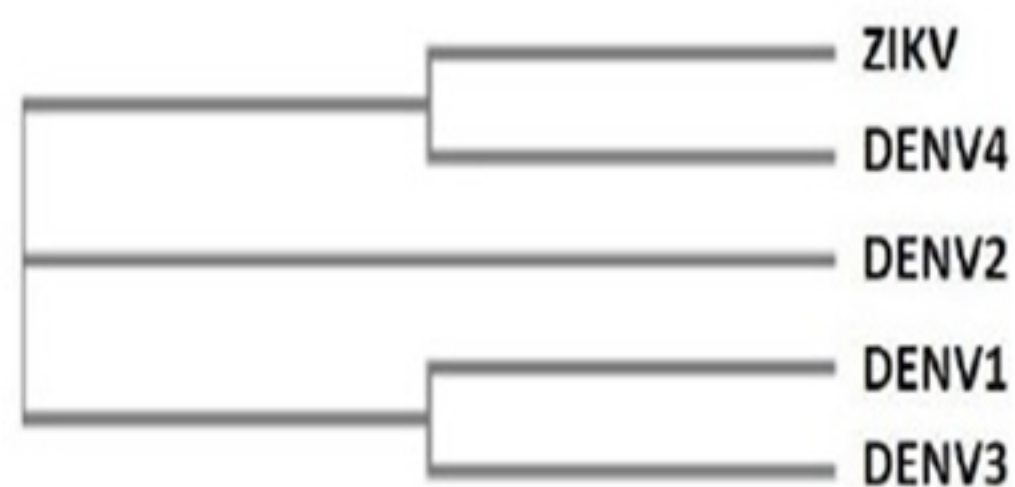660   models. The top five candidates are in bold.

661   **S6 Table:** Gibbs free energy for folding of the selected MEV models.

662   **S7 Table:**Conserved Epitopes for DENV serotypes and ZIKV as obtained from the MEV

663   models.

ZIKV
DENV4
DENV2
DENV1
DENV3

NS5

(A)

ZIKV
DENV4
DENV2
DENV1
DENV3

NS3

(B)

Fig 1

Fig 1

A

ZIKV 211 WLVHKEWFHDIPLPW 225
DENV4 206 WLVHKQWFLDLPLPW 220
DENV2 206 WLVHRQWFLDLPLPW 220
DENV1 206 WLVHKQWFLDLPLPW 220
CONSENSUS WLVHKQWFLDLPLPW

B

ZIKV 451 CVYNMMGKREKK 462
DENV4 450 CVYNMMGKREKK 461
DENV2 450 CVYNMMGKREKK 461
DENV1 449 CVYNMMGKREKK 460
DENV3 449 CVYNMMGKREKK 460
CONSENSUS CVYNMMGKREKK

C

ZIKV 469 AKGSRAIWYMWLGAR 483
DENV4 468 AKGSRAIWYMWLGAR 482
DENV2 468 AKGSRAIWYMWLGAR 482
DENV1 467 AKGSRAIWYMWLGAR 481
DENV3 467 AKGSRAIWYMWLGAR 481
CONSENSUS AKGSRAIWYMWLGAR

D

ZIKV 484 FLEFEALGFLNEDHW 498
DENV4 483 FLEFEALGFLNEDHW 497
DENV2 483 FLEFEALGFLNEDHW 497
DENV1 482 FLEFEALGFMNEDHW 496
DENV3 482 YLEFEALGFLNEDHW 496
CONSENSUS FLEFEALGFLNEDHW

E

ZIKV 747 GWSIRETACLAK 758
DENV4 746 GWSLRETACLGK 757
DENV2 745 GWSLRETACLGK 756
DENV1 744 GWSLRETACLGK 755
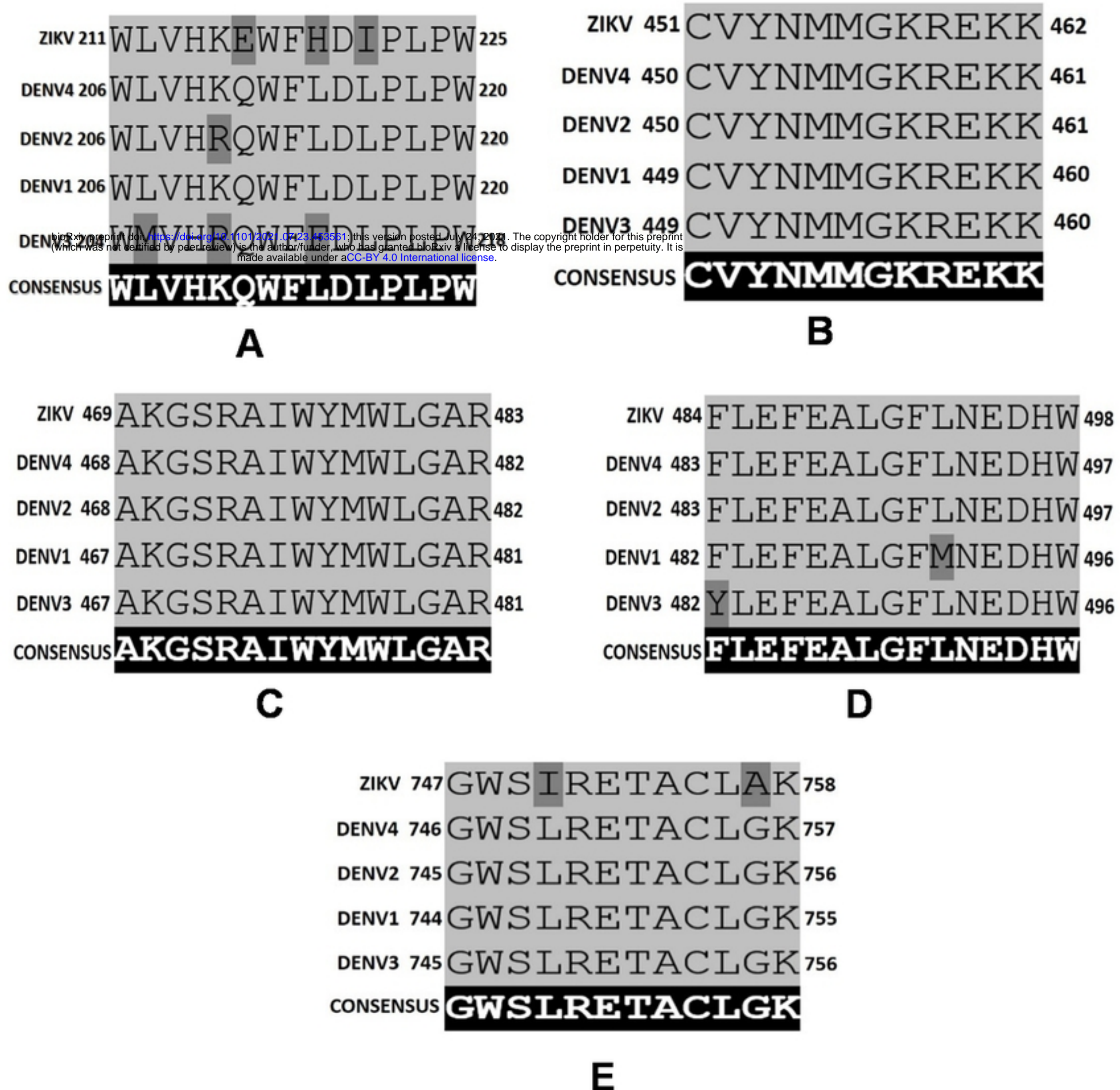DENV3 745 GWSLRETACLGK 756
CONSENSUS GWSLRETACLGK

Fig 2

Fig 2

ZIKV 469 AKGSRAIWYMWLGARFLEFEALGFLNEDHW 498
DENV4 468 AKGSRAIWYMWLGARFLEFEALGFLNEDHW 497
DENV2 468 AKGSRAIWYMWLGARFLEFEALGFLNEDHW 497
DENV1 467 AKGSRAIWYMWLGARFLEFEALGFMNEDHW 496
DENV3 467 AKGSRAIWYMWLGARYLEFEALGFLNEDHW 496
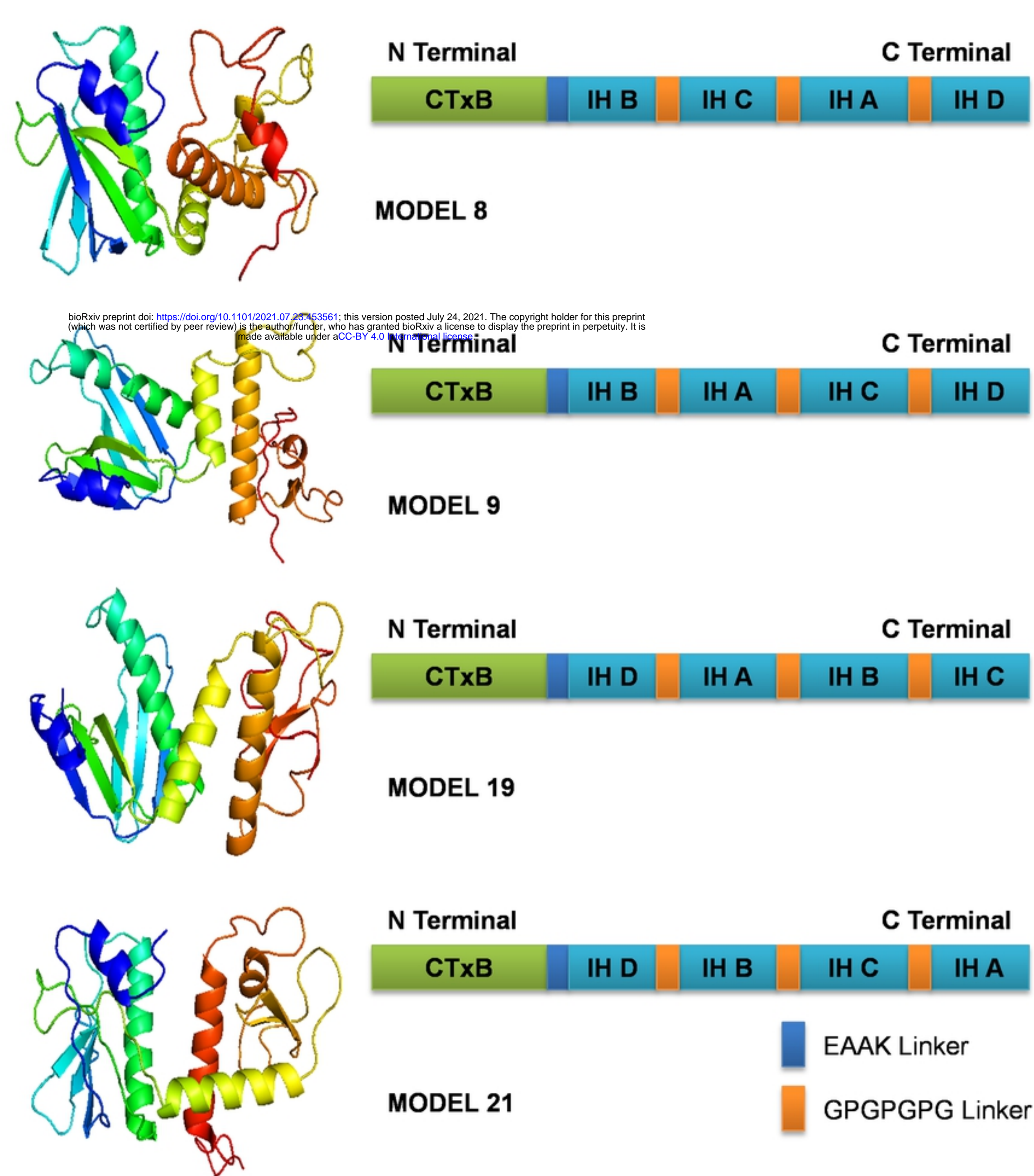CONSENSUS AKGSRAIWYMWLGARFLEFEALGFLNEDHW
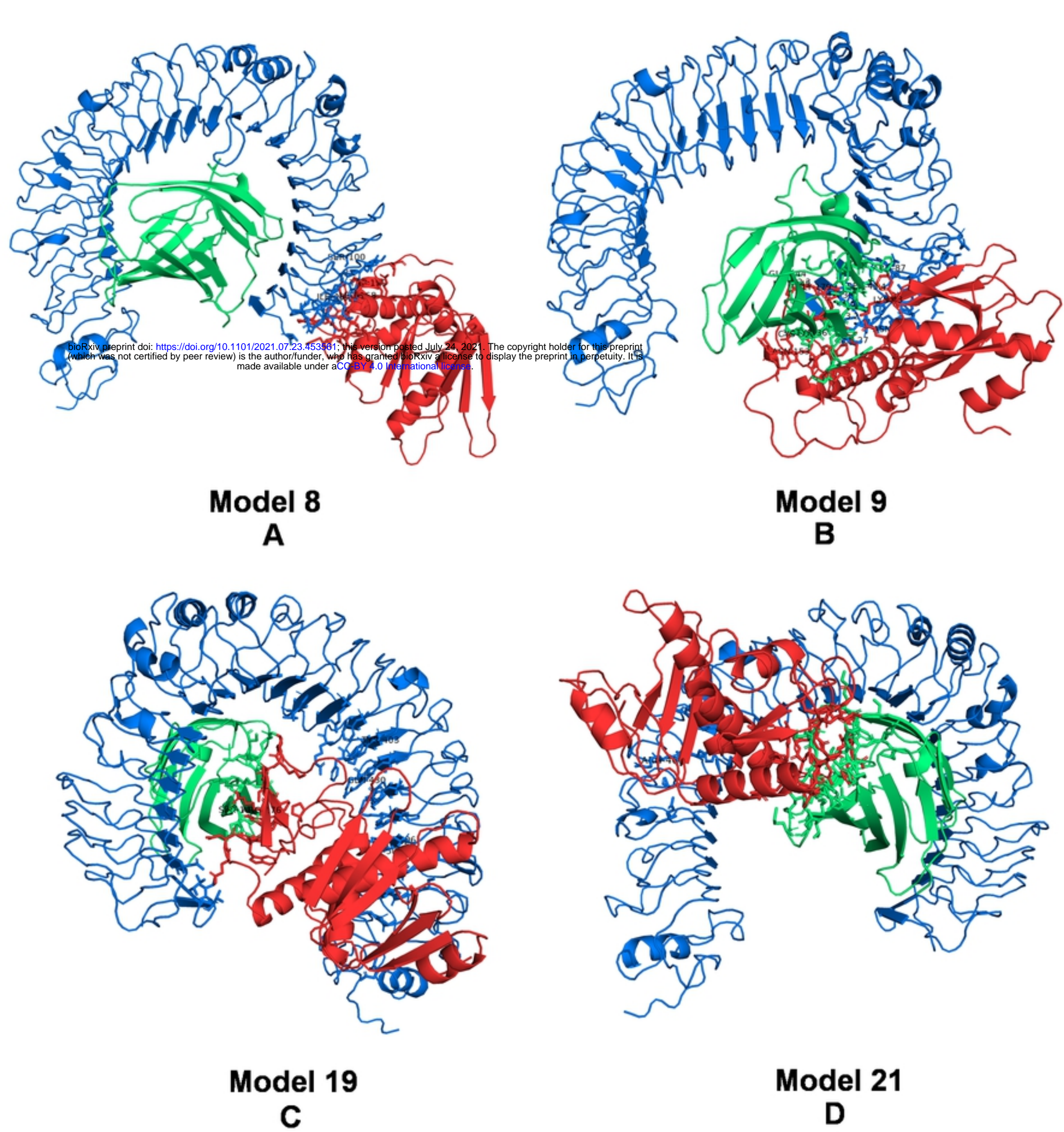
Fig 3

Fig 3

N Terminal   C Terminal

| CTxB | | IH B | IH C | IH A | IH D |

MODEL 8

N Terminal   C Terminal

| CTxB | | IH B | IH A | IH C | IH D |

MODEL 9

N Terminal   C Terminal

| CTxB | | IH D | IH A | IH B | IH C |

MODEL 19

N Terminal   C Terminal

| CTxB | | IH D | IH B | IH C | IH A |

MODEL 21

EAAK Linker

GPGPGPG Linker

Fig 4

Fig 4

Model 8

A

Model 9

B

Model 19

C

Model 21

D

Fig 5

Fig 5