1

Epistasis at the SARS-CoV-2 RBD Interface and the Propitiously Boring Implications for Vaccine Escape

4

Nash D. Rochman[1*], Guilhem Faure[2], Yuri I. Wolf[1], Peter L. Freddolino[3,4], Feng Zhang[2,5,6,7,8*], and Eugene V. Koonin[1*]

[1]National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894
[2]Broad Institute of MIT and Harvard, Cambridge, MA 02142; [3]Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, MI, USA; [4]Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA; [5]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; [6]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139; [7]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; and [8]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

*For correspondence: nash.rochman@nih.gov, zhang@broadinstitute.org, koonin@ncbi.nlm.nih.gov

17

Keywords: SARS-CoV-2, Delta Variant, Gamma Variant, Escape Mutants, Epistasis, Protein Structure Modelling, Rosetta

20

21

Abstract

At the time of this writing, August 2021, potential emergence of vaccine escape variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a grave global concern. The interface between the receptor-binding domain (RBD) of SARS-CoV-2 spike (S) protein and the host receptor (ACE2) overlap with the binding site of principal neutralizing antibodies (NAb), limiting the repertoire of viable mutations. Nonetheless, variants with multiple mutations in the RBD have rose to dominance. Non-additive, epistatic relationships among RBD mutations are apparent, and assessing the impact of such epistasis on the mutational landscape is crucial. Epistasis can substantially increase the risk of vaccine escape and cannot be completely characterized through the study of the wild type (WT) alone. We employed protein structure modeling using Rosetta to compare the effects of all single mutants at the RBD-NAb and RBD-ACE2 interfaces for the WT, Gamma (417T, 484K, 501Y), and Delta variants (452R, 478K). Overall, epistasis at the RBD surface appears to be limited and the effects of most multiple mutations are additive. Epistasis at the Delta variant interface weakly stabilizes NAb interaction relative to ACE2, whereas in the Gamma variant, epistasis more substantially destabilizes NAb interaction. These results suggest that the repertoire of potential escape mutations for the Delta variant is not substantially different from that of the WT, whereas Gamma poses a moderately greater risk for enhanced vaccine escape. Thus, the modest ensemble of mutations relative to the WT shown to reduce vaccine efficacy might constitute the majority of all possible escape mutations.

Significance

Potential emergence of vaccine escape variants of SARS-CoV-2 is arguably the most pressing problem during the COVID-19 pandemic as vaccines are distributed worldwide. We employed a computational approach to assess the risk of antibody escape resulting from mutations in the receptor-binding domain of the spike protein of the wild type SARS-CoV-2 virus as well as the Gamma and Delta variants. The results indicate that emergence of escape mutants is somewhat less likely for the Delta variant than for the wild type and moderately more likely for the Gamma variant. We conclude that the small set of escape-

52   enhancing mutations already identified for the wild type is likely to include the majority of all

53   possible mutations with this effect, a welcome finding.

54

55

56

57

Introduction

When severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first emerged as a global public health concern early in 2020, there has been considerable debate regarding whether the low mutation rate of the virus and the relatively inflexible receptor-binding domain of the antigenic spike (S) protein would admit robust host adaptation(1, 2). By 2021, it became clear that SARS-CoV-2 has access to a broad mutational repertoire enabling extensive diversification(3) and that without vaccination, SARS-CoV-2 would likely result in substantial global disease burden for a protracted period(4, 5). The development of multiple, effective vaccines against SARS-CoV-2(6) make it possible to dramatically reduce this burden. However, at the time of writing, in August, 2021, the majority of the global population remains unvaccinated and the potential emergence of vaccine escape variants(7) is of major international concern https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html.

The interface between the receptor-binding domain (RBD) of the S protein and the host receptor (ACE2) largely overlaps with the binding sites for the most potent neutralizing antibodies (NAb)(8, 9), limiting the scope of viable mutations. Nevertheless, multiple variants containing single mutations in the RBD that, to different extents, reduce NAb binding have begun to circulate(8-10). Moreover, variants with multiple mutations in the RBD have risen to dominance outcompeting the wild type (WT, identical to Wuhan-Hu-1) and single mutants (see below). This dynamic of SARS-CoV-2 variants could result from non-additive, epistatic, interactions among the mutated sites(10, 11) or simply from additive effects of multiple mutations(11). The effects of all single mutations in the RBD relative to the WT have been studied(8, 12). Epistasis among RBD mutations has the potential to substantially increase the risk of escape variant emergence and cannot be characterized through the study of the WT alone.

Using the Rosetta software suite https://rosettacommons.org(13), we set out to study and compare the effects of all single non-synonymous mutants at the RBD-NAb and RBD-ACE2 interfaces for the WT, Gamma variant (417T, 484K, 501Y), and Delta variant (452R, 478K). The Gamma and Delta variants were considered to be of the

89 greatest relevance at the time of writing because both, particularly Delta, have become

90 dominant in different regions of the world, and both were still on the ascent(14, 15). We

91 establish the distribution of RBD mutations on the plane bounded by the costs of ACE2

92 and NAb binding and classify the direction and magnitude of epistatic interactions

93 between variant mutations and the broader mutational repertoire. The results reveal

94 only weak epistasis, which is more pronounced for the Gamma than for the Delta

95 variant, and accordingly, a limited potential for the emergence of escape variants.

96

97

98

99 **Results**

100

101 **Rationale**

102 Epistatic interactions among mutations in the RBD are of interest and concern because

103 they might substantially increase the risk of vaccine escape. Mutations in the RBD

104 subtly change the shapes of the interfaces between RBD and ACE2, and between RBD

105 and NAb (Figure 1A). While the structure of the WT RBD-ACE2 interface is highly

106 similar to that of the RBD-NAb interface (see below), a single mutation in the RBD can

107 result in distinct shape changes in both interfaces. These changes can be depicted by

108 the position of each mutant on the plane bounded by the receptor binding cost and the

109 antibody cost (Figure 1B). The cost is the increase (positive cost) or decrease (negative

110 cost) in the ACE2 or Nab binding affinity relative to the WT. The four quadrants of this

111 plane represent four broad categories of mutations. Mutations in the top, right quadrant

112 are strongly destabilizing relative to both ACE2 and the NAb. The bottom, right quadrant

113 contains mutants that strongly destabilize the interaction with ACE2 but not with NAb.

114 Most mutants in these quadrants are not evolutionarily viable. Mutants in the bottom, left

115 quadrant stabilize or only weakly destabilize both interfaces. These mutations may or

116 may not provide a selective advantage to the virus depending on the fraction of the host

117 population that has been vaccinated or has recovered from prior infection. The top, left

118 quadrant contains mutations most likely to admit vaccine escape, those which strongly

119 destabilize the interaction with NAb but not with ACE2.

120

121

Single-mutant vaccine escape candidates for the wild type RBD

Starting with the two crystal structures of interest, RBD in complex with ACE2

https://www.rcsb.org/structure/6M0J(16) and RBD in complex with the NAb CV30,

 https://www.rcsb.org/structure/6XE1, we generated a representative ensemble of 50

native conformations per complex following standard Rosetta protocols (see Methods

and Discussion for details). Although NAb that bind epitopes, which do not overlap with

the RBD have been identified(8), at the time of writing, the antibodies most critical for

assessing the risk of vaccine escape appear to reside within the RBD(9) and are well

represented by CV30. Regions important for antibody binding are known to overlap

broadly among human coronaviruses(17). We then identified the RBD residues at the

interface for each conformation and, in all conformations, introduced all single amino

acid substitutions at these sites. For the WT and Delta variant, 52 residues (Table 1)

were identified at the interface of at least one conformation for either complex. For the

Gamma variant, 4 additional residues were identified. Sites 480 and 488 are connected

by a disulfide bond and were found to be unsuitable for substitution.

137

All 19 substitutions in each of the remaining 54 sites were investigated, with the

exception of WT reversion for the variants. Altogether, this analysis produced 307,300

structures, which necessitated the development of a computationally efficient protocol.

To meet this need, mutants were introduced into each conformation without repacking

of adjacent sidechains or backbone minimization. This minimalist approach yielded

favorable comparisons to available experimental data (see below). However, generally,

substitutions might introduce steric or charge clashes within the conformations, in which

the mutations were introduced (without repacking and minimization). Inference of the

relative change in binding affinity for the ACE2 and NAb complexes is limited for such

mutations. However, we observed a favorable comparison to experimental data in this

respect as well, whereby few experimentally predicted escape mutations (relative to the

WT) fall into this, inference-limited, category (see below).

150

151　Structure stability was estimated by the total score, S, in arbitrary units produced by the

152　empirically-driven Rosetta Energy Function 2015(18) (labelled REU for "Rosetta Energy

153　Units", https://new.rosettacommons.org/docs/latest/rosetta_basics/Units-in-Rosetta).

154　The total score was calculated for each of the 50 conformations of the NAb and ACE2

155　complexes generated, and the mean value was assessed with, S_M^C, and without, S_{WT}^C,

156　the mutation. The receptor cost and antibody cost were estimated as $[S_M^C - S_{WT}^C]_{ACE2}$

157　and $[S_M^C - S_{WT}^C]_{NAb}$, respectively. The interface free energy ($\varDelta G$) was also more directly

158　approximated by the difference between the total score of the unbound state and the

159　complex, $S^C - S^U$. The effect of the mutation on this value ($\varDelta\varDelta G$) was reported for both

160　complexes, $(S_M^C - S_M^U) - (S_{WT}^C - S_{WT}^U)$. Figure 1C shows the distribution of RBD

161　mutations on the plane bounded by the ACE2 and NAb binding costs and putative NAb

162　escape candidates, for which $[S_M^C - S_{WT}^C]_{NAb} - [S_M^C - S_{WT}^C]_{ACE2} > 1$ or $\varDelta\varDelta G_{NAb}\text{-}\varDelta\varDelta G_{ACE2}>1$

163　and $[S_M^C - S_{WT}^C]_{ACE2} < 13$. The threshold value of 13 was selected to remove from

164　consideration mutations that likely produce steric or charge clashes in the structure; few

165　experimentally validated escape candidates were observed above this value (see

166　below).

167

168　Mutants showed strong clustering along the diagonal (identity line: receptor cost is

169　equal to antibody cost), indicating that most mutations similarly affected the WT RBD-

170　ACE2 and RBD-NAb complexes. Mutations in the top, left quadrant of the plane, which

171　corresponds to strong destabilization of the interaction with NAb but not with ACE2, are

172　the strongest candidates for vaccine escape, followed by those in the bottom, left

173　quadrant, which includes weakly destabilizing mutations. The selective advantage (or

174　lack thereof) of mutations in this quadrant depends on the fraction of the host population

175　that has been vaccinated or has recovered from prior infection (see also Discussion). In

176　a fully vaccinated population, mutations that substantially reduce infectivity through the

177　destabilization of receptor binding could still provide a selective advantage. In particular,

178　multiple mutations in 6 sites (417, 477, 484, 491, 493, 499) were found to substantially

179　destabilize the RBD-NAb complex relative to the RBD-ACE2 complex (Figure S1).

180　Additionally, we identified site 453 to harbor mutations that simultaneously stabilize the

181　RBD-ACE2 complex and destabilize the RBD-NAb complex. These observations are

182   broadly consistent with the results of deep mutational scanning(8, 9, 19-22). Therefore,

183   we conservatively considered all mutations, for which the antibody cost exceeded the

184   receptor cost, to be viable escape candidates.

185

186   **Single-mutant vaccine escape candidates for the Gamma and Delta variant RBDs**

187   **and predicted epistatic interactions**

188   Having charted the WT RBD landscape, we sought to identify the most prominent

189   combinations of RBD mutations circulating over the course of the pandemic. As of June,

190   16th, 2021, there were 53 countries, from which more than 1,000 SARS-CoV-2 isolates

191   were contributed to the GISAID(23) database (see Data Availability). For each of these

192   locations, we randomly selected 1,000 isolates and reported the frequency of each

193   combination of RBD mutations among the 53,000 selected isolates over time. Figure 2A

194   displays this region-normalized global prevalence of the 10 most common combinations

195   of RBD mutations. The 6 RBD single-mutants (501Y/Alpha Variant, 477N, 439K, 484K,

196   478K, and 459F) began emerging between July and November 2020. All these single

197   mutants were eventually displaced by 4 RBD multi-mutants (452R|478K/Delta Variant,

198   417T|484K|501Y/Gamma Variant, 417N|484K|501Y/Beta Variant, and

199   346K|484K|501Y), which began emerging in November, 2020, with the exception of

200   Beta, which according to our analysis, first appeared in July. By March, 2021, the WT

201   had become less prevalent than the Alpha, Gamma, and Delta variants. We pursued

202   further analysis for the Gamma and Delta variant RBDs given their high and rising

203   global prevalence. The complexes were prepared starting from the WT crystal

204   structures and treated identically to the WT. However, we also demonstrate that

205   comparable results can be obtained starting directly with the native conformations

206   approximated for the WT, which significantly reduces computational burden (see

207   Discussion and Extended Methods in the SI Appendix).

208

209   The rapid emergence and subsequent displacement of RBD single mutants might in

210   part result from epistasis among the RBD variant residues or from purely additive

211   interactions. The most prominent trend was the displacement of the single mutant, 501Y

212   (Alpha variant) by the Beta and Gamma variants (both also containing 501Y). Residue

8

213   501Y has been shown to substantially increase the binding affinity with ACE2 which,

214   however, is reduced with the addition of mutation 417N in the Beta variant(11). In

215   contrast, 417N severely reduces the neutralizing activity of a variety of NAb(24). These

216   observations imply that mutations in site 417 provide a selective advantage through

217   destabilization of the NAb complex, but given the large overlap between the RBD-NAb

218   and RBD-ACE2 interfaces, maintenance of sufficient infectivity requires a compensatory

219   mutation, such as 501Y, that stabilizes the RBD-ACE2 complex.

220

221   Examination of the interface footprints, defined as the ensemble of sites predicted to lie

222   at the interface of at least one of the 50 conformations for each complex, for the 6

223   complexes of interest demonstrates that RBD makes a greater number of contacts with

224   NAb than with ACE2 within the same range of sites, 403-506 (Figure 2B). The footprints

225   of the WT and Delta variant interfaces in both the RBD-ACE2 and the RBD-NAb

226   complexes are identical. The WT/Delta RBD-ACE2 footprint consists of 37 sites

227   whereas the WT/Delta RBD-NAb footprint consists of 51 sites (Table 1). The Gamma

228   RBD-ACE2 footprint consists of 41 sites including all those in the WT/Delta footprint,

229   with the single notable exception of site 484, and five additional sites. The Gamma

230   RBD-NAb footprint consists of 53 sites including all those in the WT/Delta interface and,

231   in addition, sites 408 and 480. Sites in the RBD-NAb footprint that are not shared by the

232   RBD-ACE2 footprint might provide routes for the emergence of vaccine escape

233   variants. However, because the RBD-ACE2 footprint is smaller than the RBD-NAb

234   interface, the former is more sensitive to perturbation than the latter, for example, from

235   mutations in site 417, which is part of the footprint of all 6 complexes. The Gamma

236   variant has a larger footprint with ACE2 compared to the other variants, potentially

237   reducing this sensitivity. Notably, however, site 484 is absent from the Gamma RBD-

238   ACE2 footprint (although Gamma variant contains mutation 484K), but remains in the

239   RBD-NAb footprint.

240

241   When a second mutation, M_j, is introduced in addition to a prior mutation (Figure 2C),

242   M_i, the resulting conformational change can be additive so that the effect of the two

243   mutations is the sum of the effects of the two individual mutations. In this case, the

9

244   position of the double-mutant $M_{i,j}$ on the plane defined by the receptor cost and antibody

245   cost relative to the single mutant, M_i, will be the same as that of the single-mutant, M_j,

246   relative to the WT. If the conformational change is non-additive, representing an

247   epistatic relationship, the resulting trends can be classified by their impact on potential

248   vaccine escape. Such trends could be escape-neutral when the ensemble of candidate

249   vaccine escape mutations differs from that for the WT, but the number of such

250   candidates is the same; escape-minimizing when the antibody cost is on average

251   reduced relative to the receptor cost across all mutations for the mutant vs the WT; or

252   escape-exacerbating where the antibody cost is on average increased.

253

254   Consistent with the differences in the footprints, we found the Gamma variant RBD

255   conformation in complex with ACE2 to be moderately different from that of the WT and

256   Delta variants, which could not be differentiated from one another. While both the

257   Gamma and Delta variant RBD conformations in complex with the NAb were found to

258   be significantly different from that of the WT, the magnitude of this difference was

259   modest and smaller in magnitude than the variability among the RBD-NAb

260   conformations (Figure 3A). Despite these differences, the effects of mutations in the

261   RBD were found to be principally additive, that is, there seems to be little epistasis

262   (Figure 3B, see Extended Methods in the SI Appendix).

263

264   The landscape of mutants predicted to enhance vaccine escape for the Delta variant

265   was almost identical to that of the WT but differed significantly from the Gamma variant

266   landscape. Mutations in the Gamma variant tend to have a lower receptor cost relative

267   to the WT, possibly, due to the larger RBD-ACE2 footprint (see above), resulting in an

268   increased number of escape candidates. These trends are summarized in Figure 3C,

269   which tabulates all non-shared candidates. There are 15(13) escape candidates in the

270   WT that were not predicted to enhance escape for Delta and 6(2) candidates in Delta

271   but not WT (values in parentheses are mutations with $[S_M^C - S_{WT}^C]_{ACE2} < 13$, regardless

272   of whether or not the mutation is a candidate, included to mitigate potential artifacts

273   caused by steric or charge clashes). In contrast, in the case of Gamma, there were

274   32(28) candidates identified in WT but not Gamma, and 86(66) candidates identified in

275 Gamma but not the WT. Thus, we identified 9(11) fewer candidates for Delta compared

276 to the WT but 54(38) additional candidates for Gamma.

277

278 Consistent with the more dramatic conformational change observed in the RBD-ACE2

279 complex relative to the RBD-NAb complex, the non-additive effects observed in the

280 Gamma variant appear to predominantly result from the decreased sensitivity of the

281 RBD-ACE2 interface to mutation. This conclusion is compatible with the available

282 experimental results. Figure 4A shows the distribution of the receptor cost,

283 $[S_M^C - S_{WT}^C]_{ACE2}$, for two categories of mutations for each of the three receptor

284 complexes: those at the interface that have been experimentally demonstrated to

285 reduce neutralizing activity of antibodies COV2-2050 and COV2-2479 in the WT(8), and

286 all others (included in the same experimental study) at the interface. As discussed

287 above, the upper bound for the receptor cost, $[S_M^C - S_{WT}^C]_{ACE2}$, is lower for mutations

288 predicted to reduce NAb activity than for other mutations. However, only the Gamma

289 candidate ensemble exhibits a reduced median receptor cost (see bootstrap analysis in

290 Figure 4A). In other words, in the Gamma variant, mutations that are predicted to

291 reduce NAb activity are also less likely than other mutations to reduce the receptor

292 binding affinity relative to the WT and Delta variant.

293

294 Figure 4B summarizes the magnitude of the increase of the risk of vaccine-escape for

295 each mutation at the RBD interface (given $[S_M^C - S_{WT}^C]_{ACE2} < 13$). Epistasis increasing

296 the risk of vaccine-escape is apparent in three regions of the RBD interface: site 417,

297 site 477, and site 494 together with the surrounding neighborhood. The trend in site 417

298 was observed only in Gamma, which already contains mutation 417T, showing that

299 further changes to this site could result in enhanced vaccine escape. However, the

300 epidemiological implications of this finding are limited considering that mutations in site

301 417 are likely to pose a risk of vaccine escape in most variants. The enhanced escape

302 associated with mutations in site 477 for both variants relative to the WT, together with

303 the early spread of 477N, suggest that this site could play an important role in future

304 host adaptation. Most prominently, mutations in site 494 and the surrounding

305 neighborhood are likely to enhance vaccine escape in Gamma (and to a much lesser

11

306   degree in Delta). Indeed, 494P has both been found in circulation and experimentally

307   demonstrated to reduce antibody neutralization capacity of convalescent sera(25).

308

309   In addition to these apparent differences among the ensembles of candidate vaccine-

310   escape mutations, we observed sites that harbored no candidates but nevertheless

311   displayed signatures of increased risk of vaccine escape for Gamma. The two most

312   notable trends were observed in sites 408 and 504 (Figure 4C). Axes represent

313   antibody cost vs receptor cost as in Figures 1C and 3B. All but one substitution in site

314   408 enhance vaccine-escape for Gamma, but strikingly, all have the opposite effect in

315   Delta. Similarly, all substitutions at site 504 substantially enhance vaccine-escape in

316   Gamma, but exert a modest opposite effect in Delta. However, these mutations are not

317   considered candidates in our analysis because, even in the case of Gamma, they

318   destabilize the ACE2 interaction to a greater extent than the interaction with NAb.

319   Additionally, $[S_M^C - S_{WT}^C]_{ACE2} \gg 13$ for substitutions at site 504, which limits confidence in

320   the assessment of trends at this site.

321

Discussion

323   Here we report the results of a computational study predicting the effects of all single

324   mutants at the RBD-NAb and RBD-ACE2 interfaces for the WT, Gamma variant (417T,

325   484K, 501Y), and Delta variant (452R, 478K) of SARS-CoV-2 on receptor and antibody

326   binding. For the WT, we found multiple mutations in 6 sites (417, 477, 484, 491, 493,

327   499) that are predicted to significantly destabilize the RBD-NAb complex relative to the

328   RBD-ACE2 complex and appear to pose a risk of vaccine escape, which is broadly

329   consistent with the results of deep mutational scanning(8, 9, 19-22). Overall, most

330   mutations at the interface were found to similarly effect the WT and both variants

331   indicating limited epistasis at the interface. Non-additive, epistatic interactions predicted

332   to increase the risk of vaccine-escape were apparent, however, at site 477 and site 494

333   as well as in the surrounding neighborhood. This trend is particularly prominent in the

334   Gamma variant so that, across all sites at the interface, we predicted 22% more escape

335   candidate mutations for the Gamma variant than for the WT. In contrast, there is little

336   apparent epistasis in the Delta variant, and across all sites at the interface, we predicted

337 4% fewer candidate mutations compared to WT. This information potentially could be

338 leveraged when making decisions as to which variants warrant tailored vaccine

339 development. In addition to the observed differences in infectivity and/or vaccine

340 resistance, variants, such as Gamma, that apparently have an easier access to vaccine

341 escape than other variants seem to merit closer surveillance.

342

343 Epistasis is a major if not the principal driver of protein evolution(26). Compensatory

344 mutations are particularly strong epistatic interactions that can result in

345 chemotherapeutic(27) or antimicrobial(28) drug resistance and are commonly observed

346 throughout species evolution(29). In a completely susceptible population, mutation

347 501Y, which likely substantially increases infectivity(11), is expected to evolve under

348 positive selection. As a population gains immunity, through prior exposure and/or

349 vaccination, selective pressures rapidly change to promote the emergence of resistant

350 variants(7). Under these conditions, 501Y and other mutations, which increase

351 infectivity, might primarily play the role of compensators for mutations destabilizing NAb

352 interactions, such as 417T. As global vaccinations rise, it can be expected that more

353 mutations emerge that destabilize the interactions of the RBD with both NAb and ACE2,

354 thus resulting in (partial) escape variants that, however, also have reduced infectivity.

355 However, variants such as Gamma that carry both an antibody destabilizing mutation

356 and a compensatory mutation have the potential to undercut this trend.

357

358 We conducted structural modelling for the WT, Gamma, and Delta variants to

359 approximate an ensemble of native conformations using a fairly expensive

360 computational approach. While massively parallelizable, and in principle requiring less

361 than 72 hours to compute, this method has practical limitations. To address this issue,

362 we computed alternative ensembles of both variant conformations beginning with the

363 WT ensemble (a computationally cheap approach) and demonstrated satisfactory

364 agreement with the results obtained with the original protocol (see Extended Methods in

365 the SI Appendix). With this modified approach, given the WT conformations reported in

366 this work, the vaccine-escape landscape can be cheaply and rapidly established for any

367 emergent variant of concern.

368

369 **Limitations**

370 The work presented here is strictly computational, and although we demonstrate

371 agreement with experimental results where possible, many features not captured by the

372 models presented (involving protein expression, docking, and other factors) could

373 modulate antigen-receptor or antigen-antibody binding. Furthermore, although we

374 explore many conformations for both the RBD-ACE2 and RBD-NAb interfaces, we start

375 from a single crystal structure for each. We believe the conformational ensembles

376 selected to represent each complex are diverse enough to accurately reflect the relative

377 destabilization of the NAb and ACE2 complexes across the spectrum of RBD interface

378 mutations, which is of primary concern. However, this conformational diversity makes it

379 difficult to demonstrate stabilizing interactions, which are typically much weaker than

380 destabilizing ones(12). While multiple low-energy conformations were resolved for each

381 variant and the WT, the average behavior of the conformational ensemble selected to

382 represent the Delta variant relative to the WT was found to be weakly destabilizing for

383 NAb and neutral for ACE2, whereas in the case of Gamma, it was weakly destabilizing

384 for both complexes. This is unlikely to accurately reflect the relative binding affinities

385 between these variants and the WT given the enhanced infectivity of both variants,

386 particularly Delta(14). However, it is important to recognize that the relationship

387 between the measures of interface stability we report and viral life history traits

388 (infectivity, immune activity, etc.) is complex. Although we believe we proposed sensible

389 thresholds for determining which structures can be analyzed with high confidence and

390 the biological implications of the relative destabilization of the NAb vs ACE2, the effects

391 studied in this work do not represent the diversity of possible host adaptation. It is

392 incompletely understood at the time of writing why the Delta variant appears to replicate

393 faster than the WT(14) and substitutions outside the Spike protein may play key roles in

394 immune modulation(30, 31). A targeted exploration of the lowest-energy conformations

395 achievable for each variant might yield better agreement with the known properties of

396 these variants. However, this would likely come at the cost of generalizability and

397 decrease the power with which our approach is able to predict relative destabilization of

398 interface mutations between NAb and ACE2 complexes.

14

399 **Conclusions**

400 We employed a computational approach to study the effects of all single mutations at

401 the RBD-NAb and RBD-ACE2 interfaces for the WT, Gamma variant (417T, 484K,

402 501Y), and Delta variant (452R, 478K) of SARS-CoV-2. Overall, little epistasis at the

403 RBD interface was detected, with additive effects on the binding affinities observed for

404 most pairs of mutations. In the Delta variant, the detected non-additive trends weakly

405 stabilize the interaction of the RBD with the NAb relative to the interaction with ACE2,

406 whereas in the Gamma variant, epistasis is predicted to more substantially destabilize

407 interaction with the NAb relative to ACE2. These results suggest that the mutational

408 repertoire of the Delta variant is slightly less prone to the emergence of new escape

409 mutations than that of the WT, whereas the Gamma variant poses a moderately greater

410 risk of enhanced vaccine escape. The modest ensemble of mutations relative to the WT

411 that are currently known to reduce vaccine efficacy is likely to comprise the majority of

412 all possible escape mutations for future variants.

413

414 **Brief Methods**

415 The RBD-ACE2 interface was modelled starting with the crystal structure of the RBD in

416 complex with ACE2 https://www.rcsb.org/structure/6M0J(16). Residues outside the RBD

417 that might affect binding (e.g. 614(32)) were not considered. The RBD-NAb interface

418 was modelled starting with the crystal structure of the respective complex:

419 https://www.rcsb.org/structure/6XE1(33, 34). Epitopes outside the RBD that might be

420 epidemiologically relevant (35, 36) were not considered. The Rosetta(13, 37) software

421 suite was used to approximate native conformations(38) relative to the crystal structure

422 through a process of backbone minimization and side chain repacking. The total score,

423 S, in arbitrary units(18) (labelled REU for "Rosetta Energy Units",

424 https://new.rosettacommons.org/docs/latest/rosetta_basics/Units-in-Rosetta), as well as

425 ΔG separated, which is the difference in the total score between the bound (complex)

426 and unbound state, S^C-S^U, derived from separating the binding partners were used to

427 assess interface stability. A lower total score, S^C, and lower ΔG separated, S^C-S^U<0

428 indicate relative stability. We first searched for an "energy funnel"(39) among the

429 approximated native conformations. However, a funnel was not identified, so we

430 proceeded to select an ensemble of conformations (50 per interface) based on a total

431 score and ΔG separated ranking protocol to represent each interface rather than the

432 single lowest-energy conformation, which may be an entropically disfavored state(40).

433 We then predicted 52 residues to lie at the interface of at least one structure for

434 WT/Delta and 56 for Gamma. We introduced each of the19 possible mutations at each

15

435 interface site for these 100 structures for the WT, Delta, and Gamma variants, with the
436 exception of the WT reversion for the variants. Mutations were introduced without
437 repacking and minimization for a maximally computationally efficient approach.
438 Alternative approaches have been reviewed (41-44), but the minimal approach pursued
439 maximizes the breadth of candidate mutations considered and is able to recapitulate
440 experimental results. As an alternative to constructing the variant ensembles starting
441 with the crystal structure, we considered starting with the conformational ensemble
442 constructed for the WT and applying a reduced protocol of iterative minimization and
443 repacking for conformational optimization. This approach produced results compatible
444 with those obtained with the original protocol and can be utilized for rapid evaluation of
445 emerging variants of concern (see Extended Methods in the SI Appendix).

446

Data Availability

448 The data has been deposited through Zenodo(45)
449 (https://doi.org/10.5281/zenodo.5297699) including GISAID acknowledgements.
450 Previously published data were used for this work: GISAID(23). Data is additionally
451 made available through FTP: https://ftp.ncbi.nih.gov/pub/wolf/_suppl/SARSstruct21/

Author contributions

453 NDR, and GF collected data; NDR, YIW, GF, PF, FZ, and EVK analyzed data; NDR and
454 EVK wrote the manuscript that was edited and approved by all authors.

Acknowledgements

461

References

1. Zhan SH, Deverman BE, & Chan YA (2020) SARS-CoV-2 is well adapted for humans. What does this mean for re-emergence? *BioRxiv*.

2. van Dorp L*, et al.* (2020) No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nature communications* 11(1):1-8.

3. Rochman ND*, et al.* (2021) Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proceedings of the National Academy of Sciences* 118(29).

4. Rochman ND, Wolf YI, & Koonin EV (2020) Evolution of Human Respiratory Virus Epidemics. *medRxiv*.

5. Saad-Roy CM*, et al.* (2020) Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. *Science* 370(6518):811-818.

6. Amanat F & Krammer F (2020) SARS-CoV-2 vaccines: status report. *Immunity* 52(4):583-589.

7. Rochman ND, Wolf YI, & Koonin EV (2020) Substantial Impact of Post Vaccination Contacts on Cumulative Infections during Viral Epidemics. *medRxiv*.

8. Greaney AJ*, et al.* (2021) Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell host & microbe* 29(1):44-57. e49.

9. Greaney AJ*, et al.* (2021) Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. *Nature Communications* 12(1):1-14.

10. Nelson G*, et al.* (2021) Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y. V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *BioRxiv*.

11. Laffeber C, de Koning K, Kanaar R, & Lebbink JH (2021) Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants. *Journal of Molecular Biology* 433(15):167058.

12. Starr TN*, et al.* (2020) Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182(5):1295-1310. e1220.

13. Leaver-Fay A*, et al.* (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods in enzymology* 487:545-574.

14. Li B*, et al.* (2021) Viral infection and transmission in a large well-traced outbreak caused by the Delta SARS-CoV-2 variant. *MedRxiv*.

15. Duong D (2021) Alpha, Beta, Delta, Gamma: What's important to know about SARS-CoV-2 variants of concern? (Can Med Assoc).

16. Lan J*, et al.* (2020) Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581(7807):215-220.

17. Jiang S, Hillyer C, & Du L (2020) Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. *Trends in immunology* 41(5):355-359.

18. Alford RF*, et al.* (2017) The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation* 13(6):3031-3048.

19. Starr TN*, et al.* (2021) Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* 371(6531):850-854.

20. Starr TN, Greaney AJ, Dingens AS, & Bloom JD (2021) Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Reports Medicine* 2(4):100255.

21. Wang Z*, et al.* (2021) mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* 592(7855):616-622.

508 22. Greaney AJ, *et al.* (2021) Comprehensive mapping of mutations in the SARS-CoV-2 receptor-
509   binding domain that affect recognition by polyclonal human plasma antibodies. *Cell host &*
510   *microbe* 29(3):463-476. e466.
511 23. Shu Y & McCauley J (2017) GISAID: Global initiative on sharing all influenza data–from vision to
512   reality. *Eurosurveillance* 22(13):30494.
513 24. Zhang Q*, et al.* (2021) Potent and protective IGHV3-53/3-66 public antibodies and their shared
514   escape mutant on the spike of SARS-CoV-2. *Nature communications* 12(1):1-12.
515 25. Alenquer M*, et al.* (2021) Amino acids 484 and 494 of SARS-CoV-2 spike are hotspots of immune
516   evasion affecting antibody but not ACE2 binding. *bioRxiv*.
517 26. Breen MS, Kemena C, Vlasov PK, Notredame C, & Kondrashov FA (2012) Epistasis as the primary
518   factor in molecular evolution. *Nature* 490(7421):535-538.
519 27. Sakai W*, et al.* (2008) Secondary mutations as a mechanism of cisplatin resistance in BRCA2-
520   mutated cancers. *Nature* 451(7182):1116-1120.
521 28. Levin BR, Perrot V, & Walker N (2000) Compensatory mutations, antibiotic resistance and the
522   population genetics of adaptive evolution in bacteria. *Genetics* 154(3):985-997.
523 29. Rochman ND, Wolf YI, & Koonin EV (2020) Deep phylogeny of cancer drivers and compensatory
524   mutations. *Communications biology* 3(1):1-11.
525 30. Zhang Y*, et al.* (2020) The ORF8 protein of SARS-CoV-2 mediates immune evasion through
526   potently downregulating MHC-I. *BioRxiv*.
527 31. Zinzula L (2021) Lost in deletion: The enigmatic ORF8 protein of SARS-CoV-2. *Biochemical and*
528   *biophysical research communications* 538:116-124.
529 32. Zhang J*, et al.* (2021) Structural impact on SARS-CoV-2 spike protein by D614G substitution.
530   *Science* 372(6541):525-530.
531 33. Hurlburt NK*, et al.* (2020) Structural basis for potent neutralization of SARS-CoV-2 and role of
532   antibody affinity maturation. *Nature communications* 11(1):1-7.
533 34. Seydoux E*, et al.* (2020) Analysis of a SARS-CoV-2-infected individual reveals development of
534   potent neutralizing antibodies with limited somatic mutation. *Immunity* 53(1):98-105. e105.
535 35. Voss WN*, et al.* (2021) Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-
536   RBD spike epitopes. *Science* 372(6546):1108-1112.
537 36. Garushyants SK, Rogozin IB, & Koonin EV (2021) Insertions in SARS-CoV-2 genome caused by
538   template switch and duplications give rise to new variants of potential concern. *bioRxiv*.
539 37. Fleishman SJ*, et al.* (2011) RosettaScripts: a scripting language interface to the Rosetta
540   macromolecular modeling suite. *PloS one* 6(6):e20161.
541 38. Tyka MD*, et al.* (2011) Alternate states of proteins revealed by detailed energy landscape
542   mapping. *Journal of molecular biology* 405(2):607-618.
543 39. Tovchigrechko A & Vakser IA (2001) How common is the funnel-like energy landscape in protein-
544   protein interactions? *Protein science* 10(8):1572-1583.
545 40. Grünberg R, Nilges M, & Leckner J (2006) Flexibility and conformational entropy in protein-
546   protein binding. *Structure* 14(4):683-693.
547 41. Barlow KA*, et al.* (2018) Flex ddG: Rosetta ensemble-based estimation of changes in protein–
548   protein binding affinity upon mutation. *The Journal of Physical Chemistry B* 122(21):5389-5399.
549 42. Xue T*, et al.* (2021) Single point mutations can potentially enhance infectivity of SARS-CoV-2
550   revealed by in silico affinity maturation and SPR assay. *RSC Advances* 11(24):14737-14745.
551 43. Riahi S*, et al.* (2021) Application of an integrated computational antibody engineering platform
552   to design SARS-CoV-2 neutralizers. *bioRxiv*.
553 44. Desautels T, Zemla A, Lau E, Franco M, & Faissol D (2020) Rapid in silico design of antibodies
554   targeting SARS-CoV-2 using machine learning and supercomputing. *BioRxiv*.

18

555   45.   Rochman N, Faure, Guilhem, Wolf, Yuri, Freddolino, Peter, Zhang, Feng, Koonin, Eugene (2021)
556   Epistasis at the SARS-CoV-2 RBD Interface and the Propitiously Boring Implications for Vaccine
557   Escape [Data set]. (Zenodo).

558

Figure legends

559

Figure 1. Landscape of Vaccine Escape Mutants for the WT RBD

560

A. Cartoon depicting unique conformational changes to the RBD (blue) in complex with ACE2 (orange) and the NAb (green) associated with the same mutation. **B.** Cartoon depicting the landscape of vaccine escape mutations (the plane of receptor cost vs antibody cost). **C.** Landscape of vaccine escape mutations for the WT RBD. Circles with a black outline are NAb escape candidates. Color indicates propensity for escape as measured by $\Delta\Delta G$.

567

Figure 2. Dominant Trends in Circulating RBD Mutations

A. Region-normalized global prevalence of the top 10 most common combinations of RBD mutations over time. Lines are solid up to peak prevalence and dashed afterwards. Shading indicates confidence intervals. **B.** Structural comparison of the complexes of ACE2 and NAb with RBD for WT, Delta, and Gamma variants. *Top*: WT footprints including the residues with an interaction within 4A of the partner. *Middle:* Visualization of the Delta variant interface. *Bottom:* Visualization of the Gamma variant interface. Mutations are labeled and represented as sticks. WT structures are superimposed for the RBD of each variant: WT, orange; variant, olive **C.** Cartoon illustrating additive and non-additive (epistatic) interactions between mutations. From left to right: additive, escape-neutral, escape-minimizing, and escape-exacerbating. *Mi, Mj, Mij* denote the effects of the single and double mutants.

580

Figure 3. Epistasis within the RBD.

A. Interface RMSD for NAb and ACE2 complexes relative to an arbitrary WT conformation. Asterisks denote p-values less than 0.02 for a Wilcoxon Rank Sum Test. **B.** *Top:* Landscape of vaccine escape mutations for the variant RBDs. Coloring, as in Figure 1C, indicates propensity for escape as measured by $\Delta\Delta G$. Circles with a black

586   outline denote NAb escape candidates. *Bottom:* Landscape of vaccine escape

587   mutations for the WT RBD. Black points are candidates for both WT and variant; gray

588   points are not candidates for either WT or variant; green points are only candidates for

589   WT; red points are only candidates for the variant. **C.** Tabulation of the total number of

590   escape candidates for the WT (241), Delta (232), and Gamma variants (295);

591   candidates present in WT but not variant; and size difference between ensembles, Delta

592   (-9) and Gamma (54). See main text for details.

593

594   **Figure 4. Landscape of Non-Additive Mutation Effects Supporting Enhanced**

595   **Vaccine Escape for the Gamma Variant**
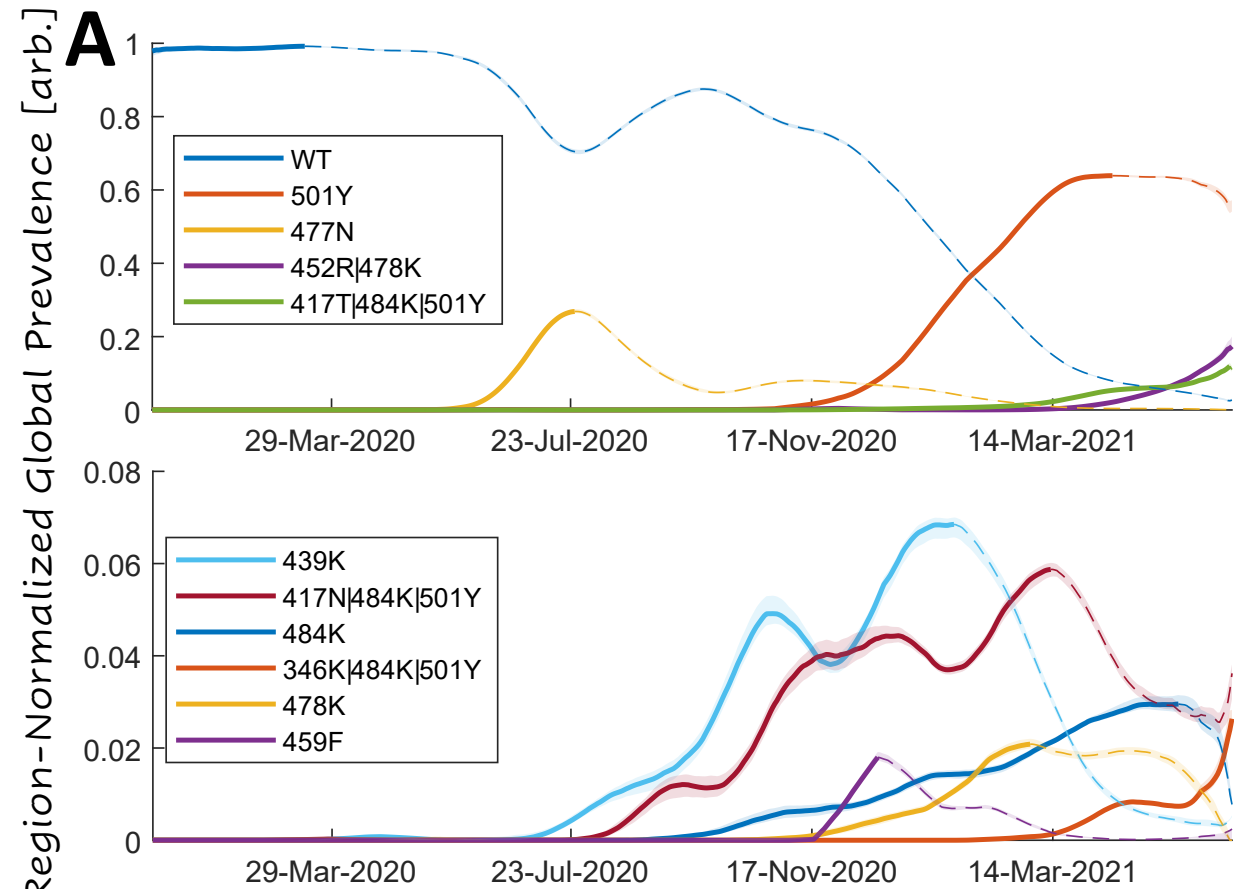
596   **A.** *Left:* Receptor cost for mutations at the interface that were experimentally

597   demonstrated to reduce NAb activity in the WT(8) (right) and other experimentally

598   studied mutations at the interface (left). Asterisks represent p-values less than 0.02 for a

599   Wilcoxon Rank Sum Test. *Right:* Distribution of median values over 1000x bootstrap. **B.**

600   Non-additive escape-exacerbating motifs in Delta (top) and Gamma (bottom) variants.

601   The size of each letter corresponds to the increased likelihood of vaccine escape for the

602   substitution in the variant relative to the WT. **C.** Movement within the plane of receptor

603   cost vs. antibody cost. Each arrow represents an amino acid substitution in site 408 or

604   504, and arrows point from the position on the plane corresponding to substitution in the

605   WT to the position corresponding to the substitution in each variant. Color represents

606   the sign of $\Delta\Delta G_{NAb} - \Delta\Delta G_{ACE2}$ (blue, negative; red, positive).
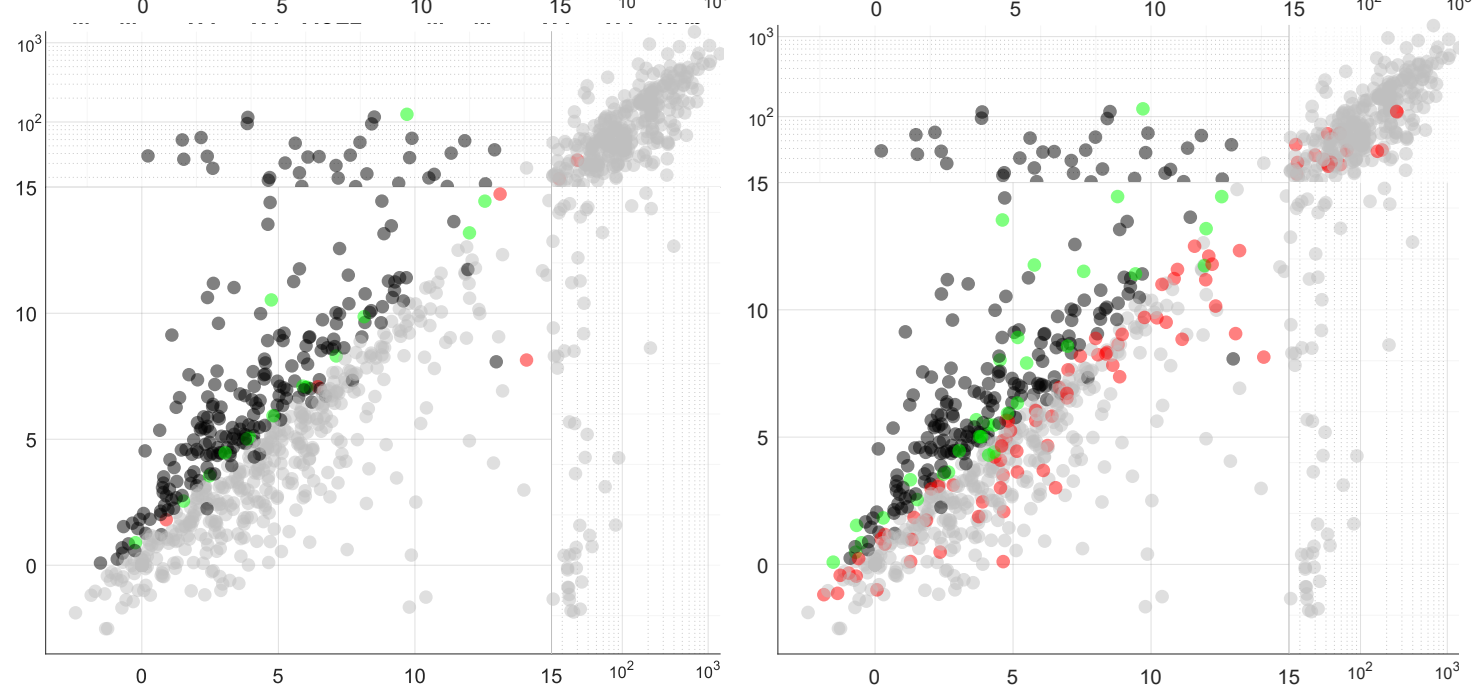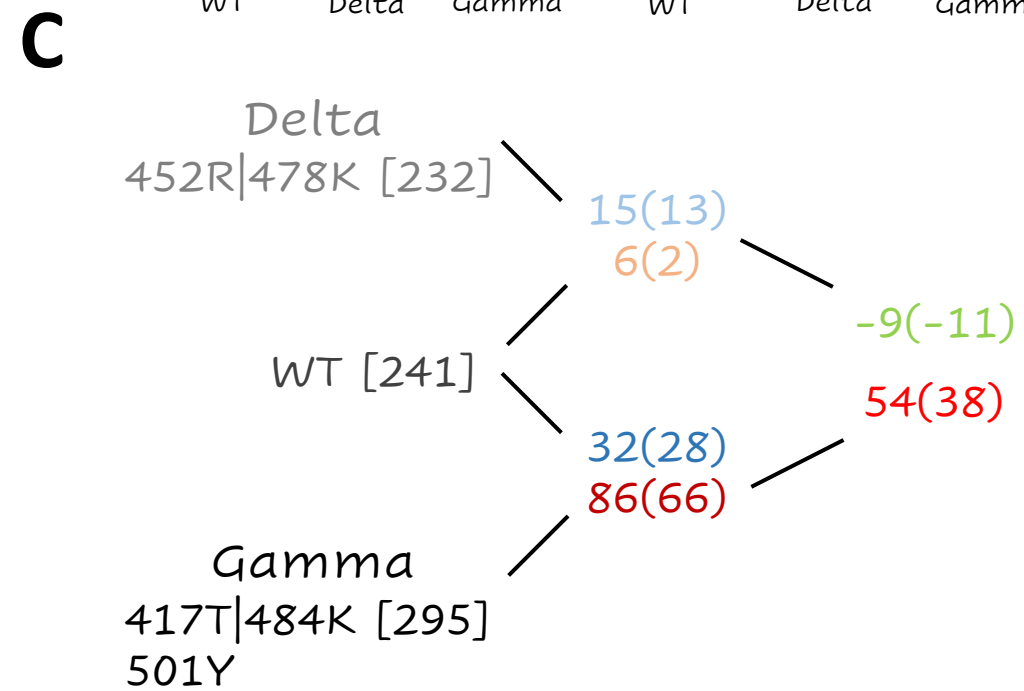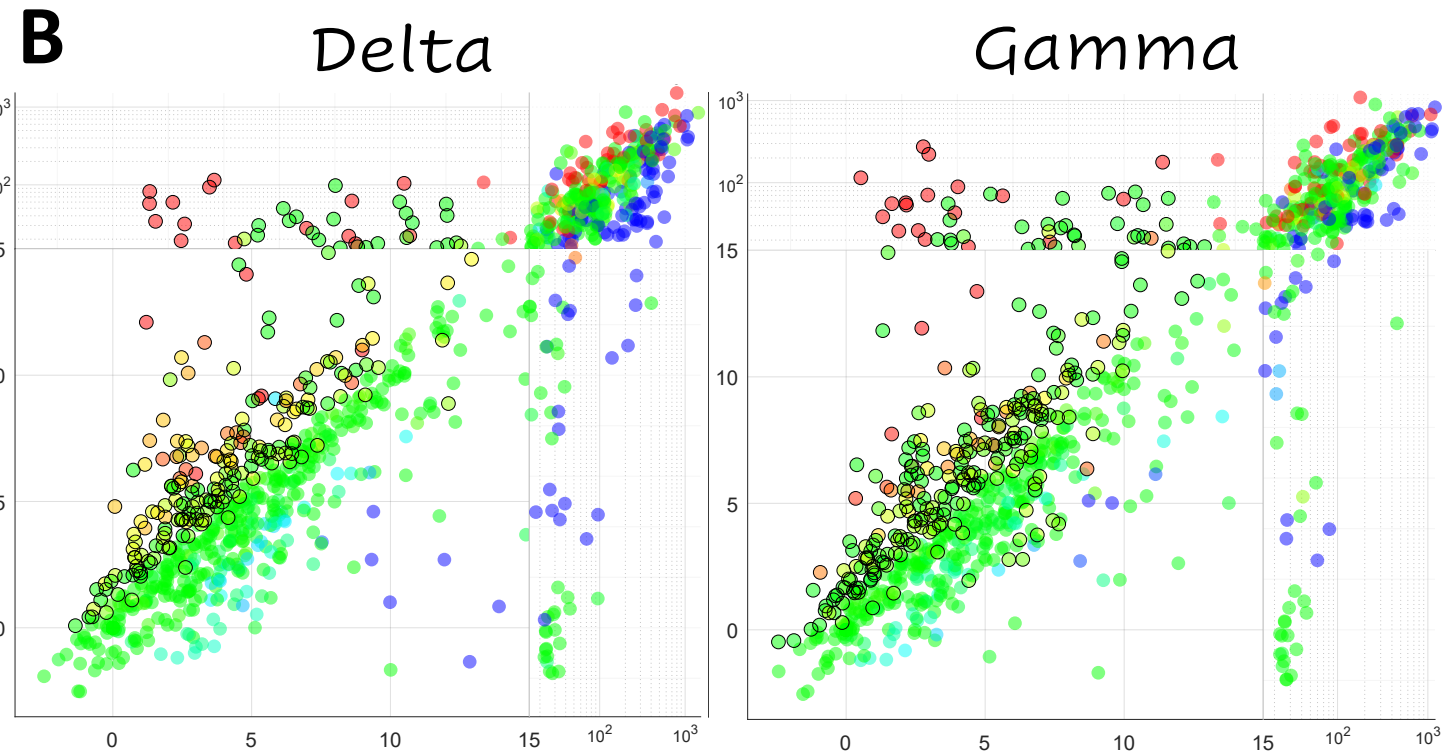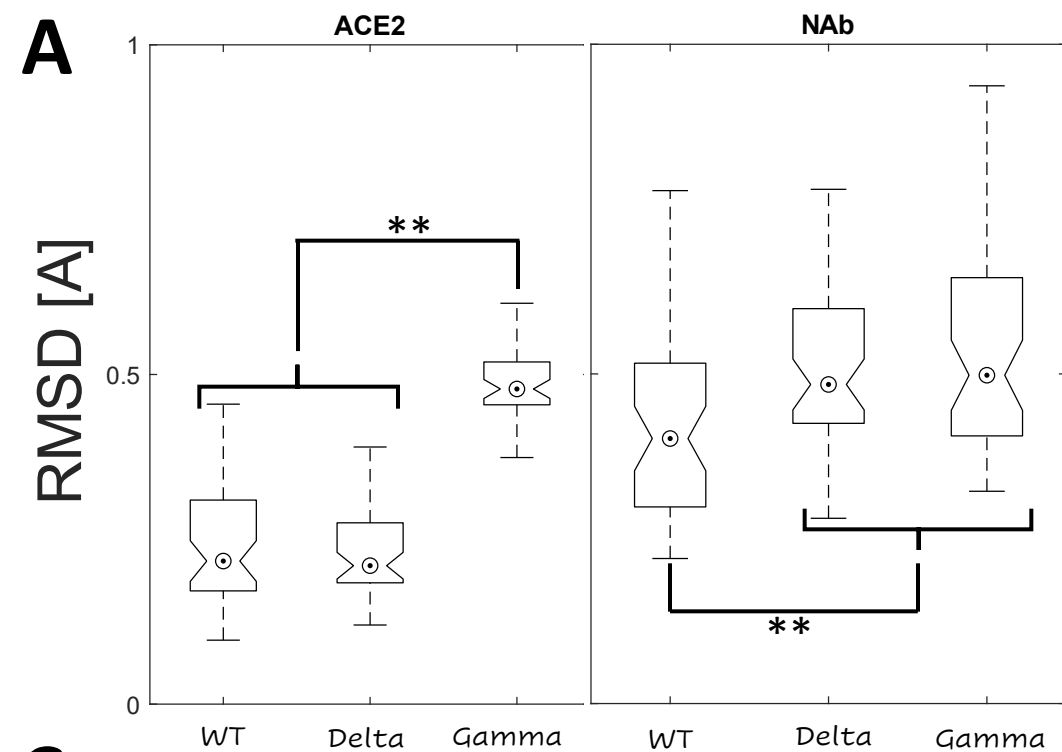
607

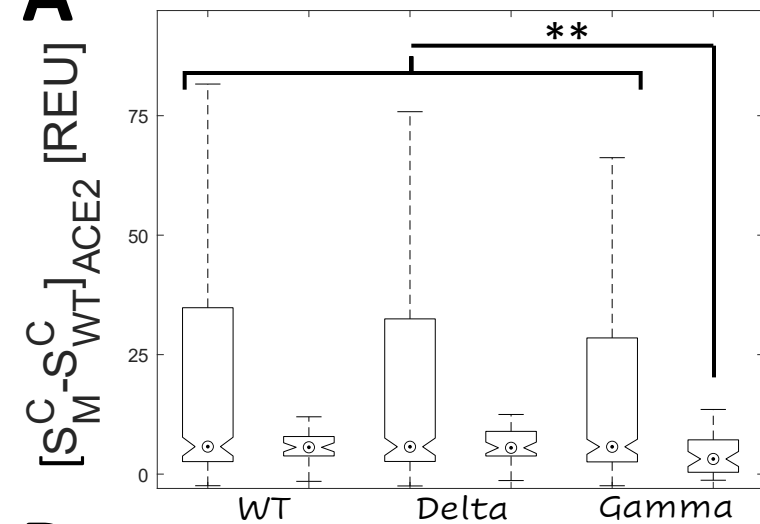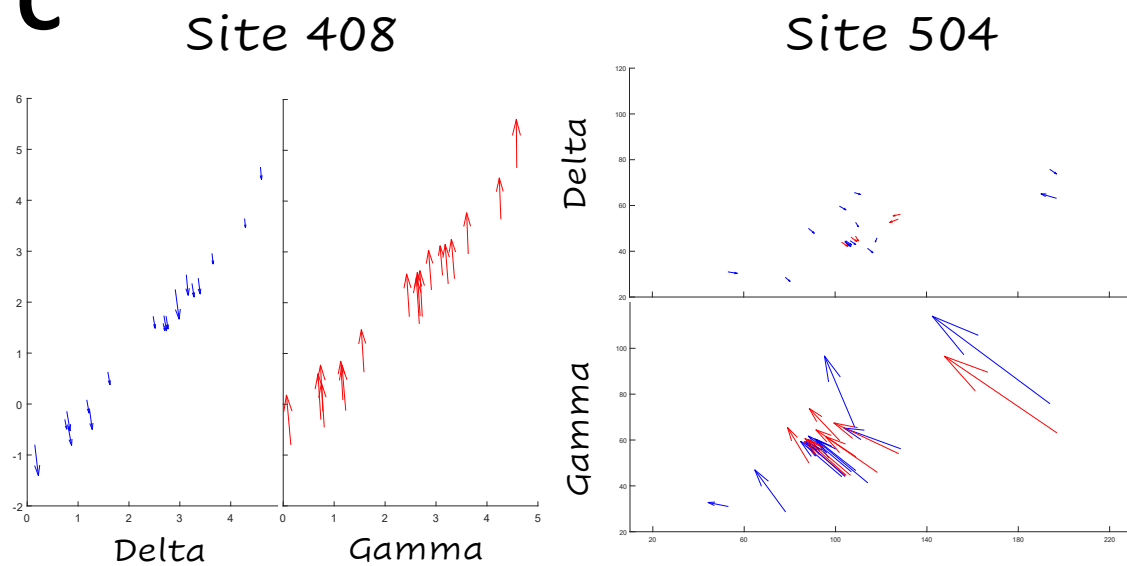608   **Table 1: Receptor-binding and antibody-binding interface footprints in the RBD**

WT RBD-ACE2 Footprint	403, 405, 417, 445-447, 449, 453, 455-456, 473-478, 484-491, 493-498, 500-506
Delta RBD-ACE2 Footprint	Same as WT
Gamma RBD-ACE2 Footprint	WT + 404, 406, 408, 439, 499 - 484
WT RBD-NAb Footprint	403-406, 409, 414-417, 419-421, 446-447, 449, 453, 455-461, 473-478, 484-498, 500-506
Delta RBD-NAb Footprint	Same as WT
Gamma RBD-NAb Footprint	WT + 408, 480

609

A WT → M_i

ACE2
RBD

RBD   NAb

B

Antibody Cost (vertical axis)
Receptor Cost (horizontal axis)

Escape | Not Viable

Neutral/Escape | Not Viable

C

$[(S_M^C-S_M^U)-(S_{WT}^C-S_{WT}^U)]_{NAb} - [(S_M^C-S_M^U)-(S_{WT}^C-S_{WT}^U)]_{ACE2}$ [REU]

y-axis: $[S_M^C-S_{WT}^C]_{NAb}$ [REU]

x-axis: $[S_M^C-S_{WT}^C]_{ACE2}$ [REU]

colorbar: [REU]