Predicting Virus Fitness: Towards a structure-based computational model

Shivani Thakur[1], Kasper Planeta Kepp[2], Rukmankesh Mehra[1,3*]

[1] *Department of Chemistry, Indian Institute of Technology Bhilai, Sejbahar, Raipur-492015, Chhattisgarh, India. E-mail:* * rukmankesh@iitbhilai.ac.in.

[2] *DTU Chemistry, Technical University of Denmark, Building 206, 2800, Kongens Lyngby, Denmark.*

[3] *Department of Bioscience and Biomedical Engineering, Indian Institute of Technology Bhilai, Sejbahar, Raipur-492015, Chhattisgarh, India.*

* Correspondence: rukmankesh@iitbhilai.ac.in

ORCID

Rukmankesh Mehra: 0000-0001-6010-1514

Kasper P. Kepp: 0000-0002-6754-7348

Shivani Thakur: 0000-0002-7869-6648

Abstract

Predicting the impact of new emerging virus mutations is of major interest in surveillance and for understanding the evolutionary forces of the pathogen. The SARS-CoV-2 surface spike-protein (S-protein) binds to human ACE2 receptors as a critical step in host cell infection. At the same time, S-protein binding to human antibodies neutralizes the virus and prevents interaction with ACE2. Here we combine these two binding properties in a simple virus fitness model, using structure-based computation of all possible mutation effects averaged over 10 ACE2 complexes and 10 antibody complexes of the S-protein (~3,80,000 computed mutations), and validated the approach against diverse experimental binding/escape data of ACE2 and antibodies. The ACE2-antibody selectivity change caused by mutation (i.e., the differential change in binding to ACE2 vs. immunity-inducing antibodies) is proposed to be a key metric of fitness model, enabling systematic error cancelation when evaluated. In this model, new mutations become fixated if they increase the selective binding to ACE2 relative to circulating antibodies, assuming that both are present in the host in a competitive binding situation. We use this model to categorize viral mutations that may best reach ACE2 before being captured by antibodies. Our model may aid the understanding of variant-specific vaccines and molecular mechanisms of viral evolution in the context of a human host.

Keywords: SARS-CoV-2; spike protein; antibody; ACE2; mutations; computation; fitness

Introduction

The pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1–3] has been characterized by the emergence of variants involving mutations in the surface spike glycoprotein (S-protein) that is involved in human host cell entry and is important for immune recognition.[4–6] The direct health implications of these variants have put the field of protein evolution into a central role, with a public understanding of the importance of mutations leading to more infectious or antibody- or vaccine-resistant variants.[6–11] Accordingly, there is substantial interest in using protein evolution and protein modelling methods to rationalize and predict these events.

The S-protein enters human host cells by fusion with the cell-surface receptor angiotensin-converting enzyme 2 (ACE2), and the high binding affinity of the S-protein variants towards ACE2 is a prerequisite for infection.[12–14] However at the same time, the S-protein is presented to the human immune system leading to development of immunity in the population after infection or after vaccination, since this presentation is the design rationale behind many vaccines.[15,16] These two binding processes can be viewed as opposed, with one favoring infection and the other limiting it, and thus, the fitness of the virus is linked to both these binding events, as explored further in this study.

Evolution occurs by changing mutations in a folded protein structure,[17–21] and thus it is important to provide the structural context to the mutations in order to predict and rationalize their impacts. Fortunately, the few years before the pandemic witnessed major technical breakthroughs in the field of cryo-electron microscopy applied to solving the 3-dimensional structures of macromolecules.[22–25] This development made it possible during the pandemic for research groups to publish hundreds of structures of the S-protein solved for all relevant conformational states either by itself or in complex with ACE2 or a large variety of antibodies, typically at resolutions of 2-4 Å that establish well the polypeptide backbone structures.[26] This means that structure-based evolution models of the SARS-CoV-2 S-protein are in principle feasible.

The present work develops a new method by using ensembles of these experimental structures and computational models estimating binding affinities to evaluate mutation effects in a model of virus fitness that includes both ACE2 and antibody binding as a competitive binding phenomenon. We propose that the host-virion interaction can be viewed as a situation where the virion seeks to enter the cell via binding its S-protein to ACE2 before the S-protein

is bound by circulating antibodies. In this model, the virus fitness becomes directly proportional to the binding affinity towards ACE2 minus the binding affinity towards a representative cocktail of antibodies, and the mutation effect on fitness (the selection coefficient) becomes proportional to the change in this difference relative to the wild-type or reference strain.

One of the advantages of this "selectivity model" is that it compares differences of differences (differences in binding differences of mutant and wild type to ACE2 and antibodies). This protocol removes systematic errors that otherwise always exist in the experimental assays and especially in the computer models, from bias towards some amino acid mutation types[27,28] and reliance on a folded wild-type structure for estimating the mutation effects.[27–35] Our results correlated well with diverse experimental binding/escape datasets for ACE2 and antibodies. Overall, our work demonstrates a simple, intuitively appealing, and computable model to estimate the fitness function of the virus.

Methods

Strategy and hypothesis

ACE2 and antibodies (AB) both bind to the S-protein. ACE2 binding induces conformational changes in the S-protein, mainly in the receptor binding domain (RBD). AB binding may also be associated with conformational changes and the binding sites of different ABs may vary, although many bind to the RBD.

We hypothesize that the virus particle, in order to be successful, needs to bind with its S-protein to ACE2 *before* being bound to an antibody, and we assume that the ACE2 receptors and antibodies are available in a competitive binding equilibrium (the relative amount of them may differ, but this is supposed to be the same condition for all variants when evaluating the fitness effects, and if so, does not affect the model). Under these conditions, the selectivity (the differential binding affinity to ACE2 vs. the ensemble of circulating antibodies) defines the fraction of virions that infects cells before they are neutralized by antibodies of the immune system. We also propose that this selectivity defines a fitness function of the virus as it accounts both for optimized binding to ACE2 and antigenic drift (selection pressure to weaken binding to antibodies). Combining the two binding affinities directly seems important because many mutations may strengthen binding to many proteins including both ACE2 and antibodies, but it is the *differential* binding effect that we envision is important.

Since cryo-EM structures and computer models are available to compute the binding affinity change caused by mutation, we can compute these for an ensemble of structures to get more robust results given the structural heterogeneity, for both ACE2 and an ensemble of antibodies, and subtract these binding effects of mutation, to yield a fitness function F:

$$F = \Delta G_{bind,ACE2} - \Delta G_{bind,AB} \quad (1)$$

Where $\Delta G_{bind,AB}$ is the binding free energy of the S-protein of the variant to the weighted ensemble of circulating antibodies, and $\Delta G_{bind,ACE2}$ is the binding affinity to ACE2 for the same variant. The selection coefficients then become proportional to the change in fitness:

$$\Delta F = F_{(MUT)} - F_{(WT)} = \Delta G_{bind(MUT,ACE2)} - \Delta G_{bind(MUT,AB)} - \Delta G_{bind(WT,ACE2)} + \Delta G_{bind(WT,AB)} \quad (2)$$

Or written in a shorter form,

$$\Delta F = \Delta\Delta G_{bind(ACE2)} - \Delta\Delta G_{bind(AB)} \quad (3)$$

A major advantage of this model is not only that it combines ACE2 and antibody binding directly in a similar treatment, but also that this reduces systematic errors on the final metric *s* (the selectivity coefficient, which is the difference in fitness ΔF of the mutant and wild type or reference strain). Since any denominator of the selection coefficient is a constant for the purpose of the present work (using the same reference strain, the Wuhan variant), we have set it to 1 in the following to reduce the computation complexity, and thus only consider ΔF. The values thus need to be scaled with a constant if estimating true selection coefficients, which is beyond the scope of the present work.

Protein structural datasets

To develop the model described above, we need to use an ensemble of structures, since using one structure for each of the two cases is insufficient due to the variation in ABs requiring an "ensemble" or cocktail of these to be studied, and also, even for the same chemical system, due to coordinate heterogeneity for published structures necessitating an ensemble treatment.[36] We thus report below average binding effects calculated from ensembles of ten ACE2 and ten AB structures to reduce the impact of these variations.

Specifically, we used two structural datasets: 10 AB- (**Figure S1** and **Table S1**) and 10 ACE2- S-protein (**Table S2)** complexes were selected from Protein Data Bank (PDB). The criteria for the selection include: **(i)** RBD should preferably be in down conformation in AB complexes whereas up in ACE2. **(ii)** For the S-protein-AB complexes, AB should be bound to RBD. **(iii)** The S-protein should be represented by a nearly full-length structure (N > 900, N: number of residues in each chain, **Table S3**). **(iv)** Structures should be of good resolution. **(v)** In addition, all ABs are human monoclonal ABs obtained from SAR-CoV (S309) and SARS-CoV2 infected patients.[37-38] These criteria ensured consistency in the two datasets (10 ACE2 and 10 AB), binding to the RBD and having similar-length structures to reduce noise and maintain a good level of uniformity for comparing S-protein-ACE2 and -AB complexes with RBD binding regions (**Table S4b**).

Figure S1 shows how the antibodies bind to the RBD (**Supplementary Section 1**). These structures include 6WPS by Pinto et al.[37], 6XEY by Liu et al.[39], 7JV6 by Piccoli et al.[40], 7K8S and 7K90 by Barnes et al.[41], 7ND7 by Dejnirattisai et al.[42], 7LS9 by Cerutti et al.[43], 7K43 by Tortorici et al.[44], 7M6E by Scheid et al.[45] and 7L56 by Rapp et al.[38] In six structures (**Figure S1a-S1f),** one antibody unit (one heavy and one light chain) is bound to each RBD, whereas in

the remaining four structures (**Figure S1g-S1j**), two antibody units interact with each RBD (**Table S1**). We included these two subsets to ensure that we cover all possible AB binding regions on RBD and broad S-protein conformations, since the ensemble of these conformations might form an active AB interacting state. Residue-specific details regarding the binding interface between antibodies and S-protein are shown in **Tables S5-S6**.

The PDB codes of ten S-protein-ACE2 complexes include 7A98 by Benton et al.[46], 7CT5 by Guo et al.[47], 7DX8 and 7DX9[48] by Yan et al.[48], 7KJ3 and 7KJ4 by Xiao et al.[49], 7KMS, and 7KMZ, 7KNH, and 7KNI by Zhou et al.[50] Out of these, six structures have all RBD in the up-conformation (7A98, 7CT5, 7DX9, 7KJ4, 7KMS, 7KNI) and four with two up and one down conformation (7DX8, 7KJ3, 7KMZ, 7KNH). Five structures are three-ACE2 bound, and the other five are two-ACE2 bound S-protein complexes (**Table S2**). By considering multiple structures of these types, we account for such variability in our estimates. In summary, the main results reported below are, unless otherwise noted, based on computed numbers averaged over all the 10 structures of the AB or ACE2 S-protein complexes, respectively.

Comparative estimates of binding affinity changes in AB vs. ACE2 datasets

In total, we computed the change in binding affinity upon mutation ($\Delta\Delta G_{bind}$ in kcal/mol) in 20 structures (ten S-protein-AB and ten S-protein-ACE2 complexes, three chains in each structure) using the BeAtMuSiC[51] program, which has yielded promising accuracy in independent benchmarks.[52,53] Mutations were introduced simultaneously in all identical chains, and then used to compute the $\Delta\Delta G_{bind} = \Delta G_{bind\ (MUT)} - \Delta G_{bind\ (WT)}$, i.e., the change in binding free energy due to mutation. The negative value indicates an increase in binding affinity and *vice-versa*. Each structure had approximately ~1000 residues, giving a total of ~19,000 mutations (19 mutations × 1000) per structure, leading to a total of ~3,80,000 computational mutation calculations in this work.

Investigating groups of mutations

Computational mutation models suffer from some limitations partly due to the biased training datasets, the algorithm's accuracy, the ignorance of structural heterogeneity, and the assumption that mutant properties can be derived from a static wild-type structure.[27,28,31,54] We have previously shown that the average mutation effect of an ensemble of structures of a group

of mutations compared to a control group of mutations is necessary to reduce systematic errors in computational models, much in the same way as in randomized controlled trials but applied in the context of computational (bio)chemistry.[55,56] In this way we can determine if a group of mutations in e.g., a variant, have properties significantly different from those of a reference strain, with much higher precision and accuracy than standard absolute estimates.

To this end, the S-protein mutation dataset for each S-protein-AB and -ACE2 complexes was divided into nine subsets (**Table S4**): **(*a*)** "Full length" – all possible mutations in S-protein chain with N residues, i.e., N × 19 mutations; **(*b*)** "Interface" – including mutations in the binding interface residues of the S-protein (listed in **Table S5**). BeAtMuSiC identifies interface residues based on the condition that the difference between a residue's solvent accessibility in the complex and apo-protein is at least 5%[51]; **(*c*)** "All natural mutations studied" – mutations observed in the wild in different variants; **(*d*)** "Saturation mutagenesis at the natural mutation site" – all possible mutations at the natural mutation sites; **(*e*)** "Interface natural mutations" – those natural mutations found in the S-protein's interface residues; **(*f*)** "Gamma mutations" – mutations found in the gamma (P.1) variant; **(*g*)** "Delta mutations" – mutations found in the delta (B.1.617.2) variant; **(*h*)** "Omicron mutations" – mutations found in omicron (B.1.1.529) variant; and **(*i*)** "RBD escape mutations" – selected RBD mutations showing strong escape from some antibodies (**Table S7**). The datasets are described in additional detail in **Supplementary Section 2**.

Statistics

Statistical analyses were performed in the following ways: **(i)** Linear relation of average $\Delta\Delta G_{bind}$ per residue between AB- ($\Delta\Delta G_{bind(10\ AB)}$) and ACE2-bound ($\Delta\Delta G_{bind(10\ ACE2)}$) S-protein structures. The comparison was made for all possible S-protein mutations in the positions common to both sets of structures (10 AB and 10 ACE2 structures; 20197 data points). The correlation was also checked for the RBD mutations (3819 points), all natural mutations studied (70 points), and RBD natural mutations (25 points) (**Figure 1**). **(ii)** Linear regression of average $\Delta\Delta G_{bind(10\ ACE2)}$ versus $\Delta\Delta G_{bind(each\ AB)}$ (**Figure S2**), and average $\Delta\Delta G_{bind(10\ AB)}$ versus $\Delta\Delta G_{bind(each\ ACE2)}$ (**Figure S3**). For linear regressions, p-values were calculated. These analyses showed the pattern of $\Delta\Delta G_{bind}$ and how results depended on the choice of structures.

In addition, we determined significant differences in the structure-averaged binding effects of the mutation groups to AB and ACE2 structures, applied to comparisons of the nine

mutation groups defined above (**Table S8**). A Student's t-test comparison for the same mean binding effect was evaluated for the various combinations of mutation categories (36 combinations) for the AB-/ACE2-bound structures (**Tables S9-S10**).

Calculating the change in fitness function

As described above, we hypothesized that natural selection of SARS-CoV-2 variants balances antigenic drift with enhanced or maintained binding to ACE2. To model this, we computed the difference in change in affinity of ACE2-bound and AB-bound complexes due to specific mutations, i.e., the differential impact on affinity towards the two targets caused by a single-point substitution. In this way, the selection coefficient (without normalization to the wild-type effect) is proportional to the change in fitness (ΔF) for each mutation, as described by equations (1-3) defined above. We used three approaches to calculate the change in fitness (ΔF) for all natural mutations studied: (i) ΔF = Average $\Delta\Delta G_{bind(MUT,\ 10\ ACE2)} - \Delta\Delta G_{bind(MUT,\ each\ AB)}$, (ii) $\Delta F = \Delta\Delta G_{bind(MUT,\ each\ ACE2)}$ – Average $\Delta\Delta G_{bind(MUT,\ 10\ AB)}$, (iii) Average ΔF = Average $\Delta\Delta G_{bind(MUT,\ 10\ ACE2)}$ – Average $\Delta\Delta G_{bind(MUT,\ 10\ AB)}$.

Validation of the approach using experimental data

To validate our findings, we compared experimental data in three ways: Firstly, experimental ACE2 binding data from Starr et al. [57] (**dataset 0**) as shown in **Table S11** was correlated with computed average $\Delta\Delta G_{bind}$ values of ACE2-bound (3283 data points) complexes. The bin method was used to see the change in correlation for a range of experimental binding, as reported earlier.[58,59] The observed ACE2 binding data was only available for RBD mutations, and in the data, negative values indicate a decrease in ACE2 binding and *vice-versa*.

In the second validation approach, we compared experimental AB escape fraction data with the computed average $\Delta\Delta G_{bind}$ values of mutations. We took experimental data of SARS-CoV2 neutralizing ABs from two data sets (**datasets 1 and 2; Table S11**). **Dataset 1** of 1924 data points was the average mutation escape of ten ABs, of which two were also present in our studied structures (C002 in 7K8S and C144 in 7K90 structure). **Dataset 2** comprised 2346 data points, which was also the average mutation escape of ten ABs and involve ABs different from those present in our studied AB structures. In addition, we also correlated the computed binding

of individual antibodies C002 (1950 data points) and C144 (1949 data points) with the experimental mutation escape according to **dataset 1**.

Lastly, the experimental differences in ACE2 binding (**dataset 0**) and AB escape (**datasets 1** and **2**) were correlated with computed binding differences (change in fitness, ΔF, equation 3) for RBD mutations**.** Experimental average AB escape values (**datasets 1** and **2; Table S11**) were subtracted from experimental ACE2 binding values (**dataset 0**), and further, these experimental difference (ACE2 – AB) for each mutation was correlated with the computed binding difference (computed change in fitness = Average $\Delta\Delta G_{bind(MUT,\ 10\ ACE2)}$ - Average $\Delta\Delta G_{bind(MUT,\ 10\ AB)}$). The data points involved in two correlations are 1847 (**dataset 1**) and 2245 (**dataset 2**) respectively.

For all comparisons, mutation escape fractions were converted to -log values. The experimental data (**datasets 0, 1** and **2**) were highly squeezed and therefore, the bin method was used with 0.25 bins to see the change in correlation for a range of experimental binding/escape, as reported earlier.[58,59]

Estimating epistatic effects

Epistatic effects were estimated using FoldX[60], by calculating the change in folding stability upon mutation (ΔΔG) after introducing all mutations simultaneously for a given variant, i.e., for the combined effect of mutations allowing them to interact, for gamma 1, gamma 2, delta, and omicron, and additionally selected AB escape mutations (**Table S12)**. The average ΔΔG values were calculated using the ΔΔG values of eight structures of the apo S-protein (average $\Delta\Delta G_{(apo)}$), AB-bound protein (average $\Delta\Delta G_{(AB)}$), and the ACE2-bound S-protein (average $\Delta\Delta G_{(ACE2)}$) for each S-protein variant. The apo S-protein dataset include 6VXX[61], 6X6P[62], 6X79[63], 6XF5[64], 6Z97[65], 7CAB[66], 7DDD[67] and 7DF3[68]. The S-protein-AB (holo, AB) structures include 6WPS[37], 6XEY[39], 7JV6[40], 7K8S[41], 7K43[44], 7K90[41], 7M6E[45] and 7ND7[42]. Whereas the S-protein-ACE2 (holo, ACE2) dataset comprises 7DX8[48], 7DX9[48], 7KJ3[49], 7KJ4[49], 7KMS[50], 7KMZ[50], 7KNH[50] and 7KNI[50] complexes.

Results and discussion

Computed affinities in S-protein-antibodies versus -ACE2 structures

Figure 1 summarizes the impact of all possible mutations averaged over the ten used structures on S-protein binding to ACE2 and the ten studied antibodies. We have previously shown that computed ACE2 binding affinity changes correlate surprisingly well with the experimental data from Bloom et al. once the data sets were curated.[58] **Figure 1** shows that the correlation coefficients illustrate well the expectation that in general, mutations that tend to cause decreased S-protein affinity to one type of proteins (ACE2) also tend to reduce binding to the other type of proteins (antibodies), i.e., the mutation effects have some generic protein-protein interaction effects, related to e.g., hydrophobic packing. Results in **Figures S2-S3** also confirm this, which shows good correlation for the comparison of each AB structure with the average of 10 ACE2 (**Figure S2)**, and *vice-versa* (**Figure S3**).

From an evolution point of view and related to the hypothesis of this work, outliers deviating from the regression line in **Figure 1** are of interest as having a potential fitness effect in a competitive binding model, as the binding is not similar for the two protein targets. To understand these outliers better, we separately looked at all mutations together (**Figure 1a**), all mutations in the RBD (**Figure 1b**), all studied naturally occurring mutations (**Figure 1c**) and natural mutations in the RBD (**Figure 1d**). From **Figure 1c**, we found Y145D, W152C, E484A, F490S, and Y505H as especially showing differential binding (>1 kcal/mol) between AB and ACE2 structures. Interestingly, most of these (E484A, F490S, and Y505H) are located on the RBD (**Figure 1d**). Natural mutations Y145D, W152C and Y505H are estimated to have higher binding affinities for antibodies, while E484A and F490S have higher computed binding affinities for ACE2. The F490S mutant is reported to escape several monoclonal antibodies.[69]

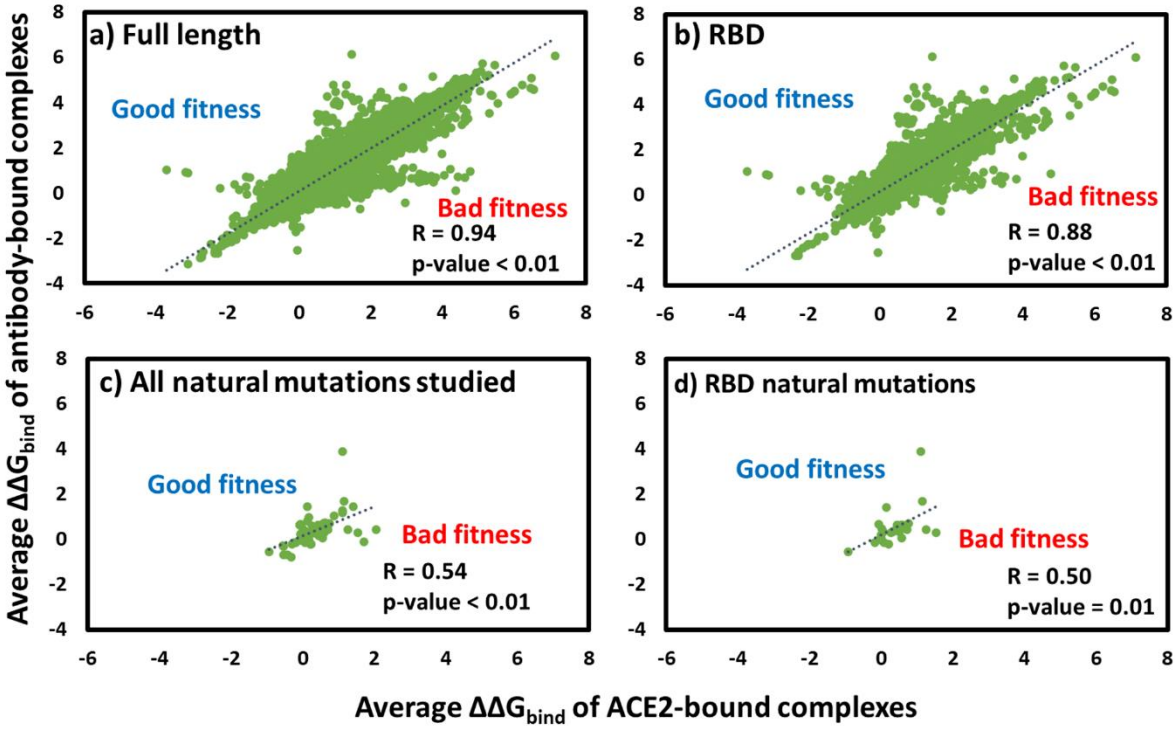

Figure 1. Linear relation between $\Delta\Delta G_{bind}$ values of S-protein mutations bound to AB and ACE2. (a) Full length mutations. **(b)** Mutation in RBD residues. **(c)** All natural mutations studied. **(d)** RBD natural mutations.

Natural mutations in S-protein in antibody and ACE2 complexes

To clarify the fitness effects of **Figure 1** on a residue basis, the $\Delta\Delta G_{bind}$ values for the 70 studied natural S-protein mutations binding to ACE2 and the studied antibodies are summarized in the heatmap in **Figure 2** (Note: Out of 79 natural mutations in the dataset, only 70 were present in the experimental S-protein structures and thus computable). The values for each of the 10 structures and the average are shown in the heatmap. The green color represents increased affinity, whereas red shows decreased affinity. Neutral effects are represented in yellow. White spaces indicate missing mutation sites in the respective structures. The heatmap shows differential average affinity changes, i.e., effect on AB/ACE2 selectivity. For mutations with only a few coordinates available (mostly white patches in **Figure 2**), no significant conclusion could be drawn due to variation in the $\Delta\Delta G_{bind}$ values, which shows the importance of accounting for structural heterogeneity by averaging, as discussed before.[36]

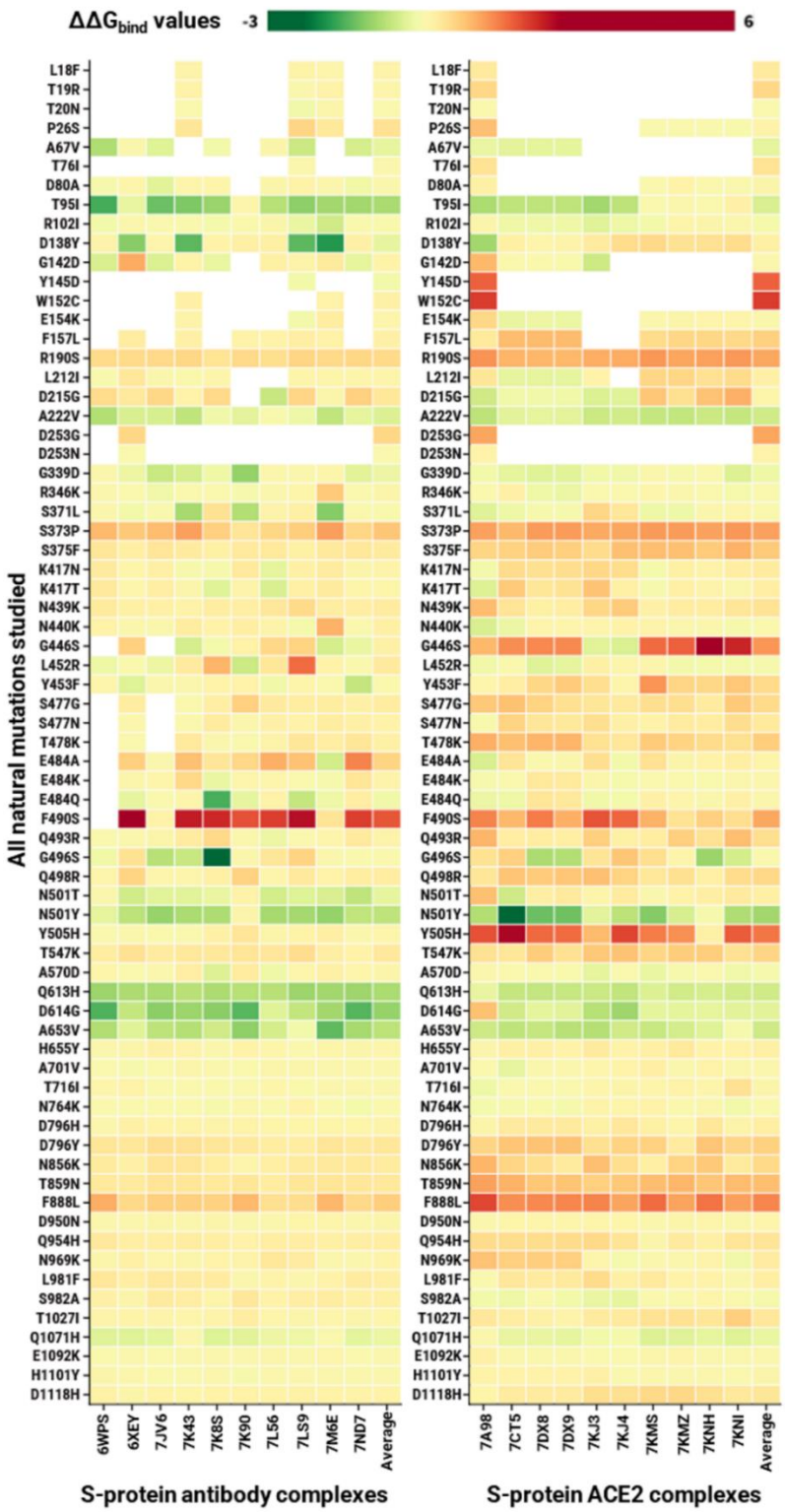

Figure 2. Heatmap of changes in binding affinity estimated for the studied natural mutations. The mutations with comparatively increased binding affinity are shown in green, whereas mutations that reduce binding affinity relative to the Wuhan reference strain are shown in red. Mutations in sites that were not available in the experimental structures are shown in white.

Interestingly, D614G showed increased affinity towards both antibodies and ACE2, but more towards the antibodies. This would be consistent with a relative loss of fitness in a population where these antibodies circulated, i.e., the antibodies were made to bind the S-protein with this very early fixated mutation present. The literature suggests that the D614G variant increases infectivity but reduces the affinity for ACE2 because of a faster dissociation rate.[70] Increased selectivity for the antibody ensemble was observed for T95I, D138Y, R190S, S373P, S375F, K417N, K417T, N439K, G446S, Y453F, S477G, S477N, T478K, E484Q, Q493R, Q498R, Y505H, T547K, Q613H, D614G, A653V, D796H, D796Y, N856K, T859N, F888L, Q954H, N969K, L981F, T1027I, and D1118H. Approximately half of these are located on the RBD. These observations are consistent with the studied set of human monoclonal antibodies having been developed to target the circulating natural mutations.

In contrast, mutations such as D215G, A222V, E484A, F490S, N501Y, A570D and S982A showed increased selectivity for ACE2. Comparatively, a very small number of mutations exhibit increased affinity for ACE2, and these might exhibit the antibody escape tendency. The N501Y mutant is reported to display 5-10 times higher binding to ACE2 than wild-type.[57,71–73] These mutations might be suspected to induce a future selective advantage by having more antigenic drift vs. the studied antibodies.

$\Delta\Delta G_{bind}$ variations due to S-protein structure and types of mutations

As mentioned in the Methods section, the mutation dataset was divided into nine groups (**Table S4**). **Figures S4-S5** show the detailed comparison of these groups for each of the 20 used experimental structures. In general, natural mutations exhibited significantly less reduced affinity compared to the control group of random mutations and even all possible hypothetical mutations located at the binding interface in both AB and ACE2 sets. When averaging the 10 structures and using group comparison (**Figure 3** and **Figure S6)**, a similar trend was observed for the nine mutation groups. However, the effect of interface and delta mutations was highly variable in the AB dataset. Interface residues showed the lowest average $\Delta\Delta G_{bind}$ compared to any other mutation group. Of the nine groups, delta mutations exhibited the highest average affinity change, followed by gamma, whereas omicron showed similarly decreased affinity in both datasets. For the antibody escape mutations, we observed a similar lowered binding, consistent with substantial antibody escape.

Figure 3 overall indicates that natural mutations had higher affinity impacts compared to random mutations in the full S-protein, including interface sites separately. Also, the natural mutations in the binding interface showed decreased affinity compared to full group of natural mutations, including gamma, delta, and omicron groups. We note that the standard deviations (SD) for the mutation groups are higher for the AB dataset than for ACE2, consistent with the larger expected structural epitope variance of the antibodies, whereas ACE2 represents one type of protein receptor only. This heterogeneity also explains why results using just one antibody as representative of antigenic drift is not meaningful.

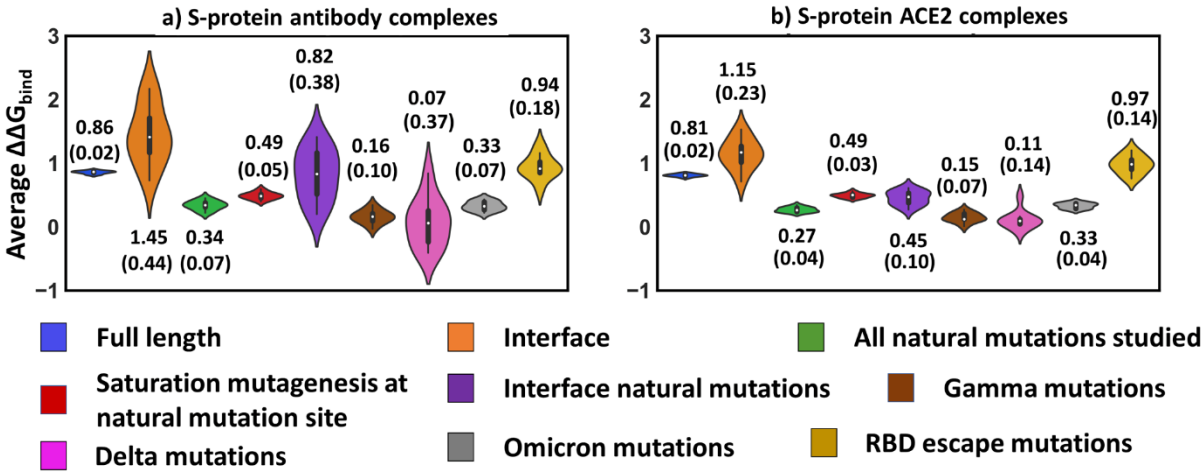

Figure 3. Average change in S-protein's binding affinity ($\Delta\Delta G_{bind}$) to antibodies and ACE2 for the studied mutation categories, averaged over the 10 structures. The average and standard deviation values (in brackets) are shown for each mutation group. A positive value indicates a decreased binding and *vice-versa*.

Figure 4 reveals differences between the behavior of the mutation groups for the two datasets: (i) random mutations in full length S-protein show significantly different binding affinity changes (blue), i.e., lower binding for AB structures; (ii) the interface region exhibits comparatively more variable and low average affinity in AB structures (orange), though the difference is not significant; (iii) the natural mutations show significantly lower antibody binding (green), which may relate to their AB escape tendency; (iv) similar prominent effects were observed for the interface natural mutations, i.e. lower AB binding. All these observations suggest weaker binding of the S-protein to antibody structures for natural and random

mutations. However, no significant differences were observed for the individual comparisons between gamma, delta, omicron, and escape mutations. This might be due to the smaller dataset for comparison because a significant difference was noted for the combined natural mutation datasets.

Overall, the interface natural mutations had the largest significant change in selectivity (binding energy difference between the two protein targets), 0.82 vs. 0.45 kcal/mol, as indicated in **Figure 4e**, where these positive values indicate reduced binding to both with respect to wild-type, but more reduced binding to AB than ACE2. This indicates that the interface natural mutations have comparatively more fitness towards ACE2. The group of all possible and all natural mutations also showed a similar significant effect of comparatively reduced antibody binding suggesting viral fitness for ACE2.

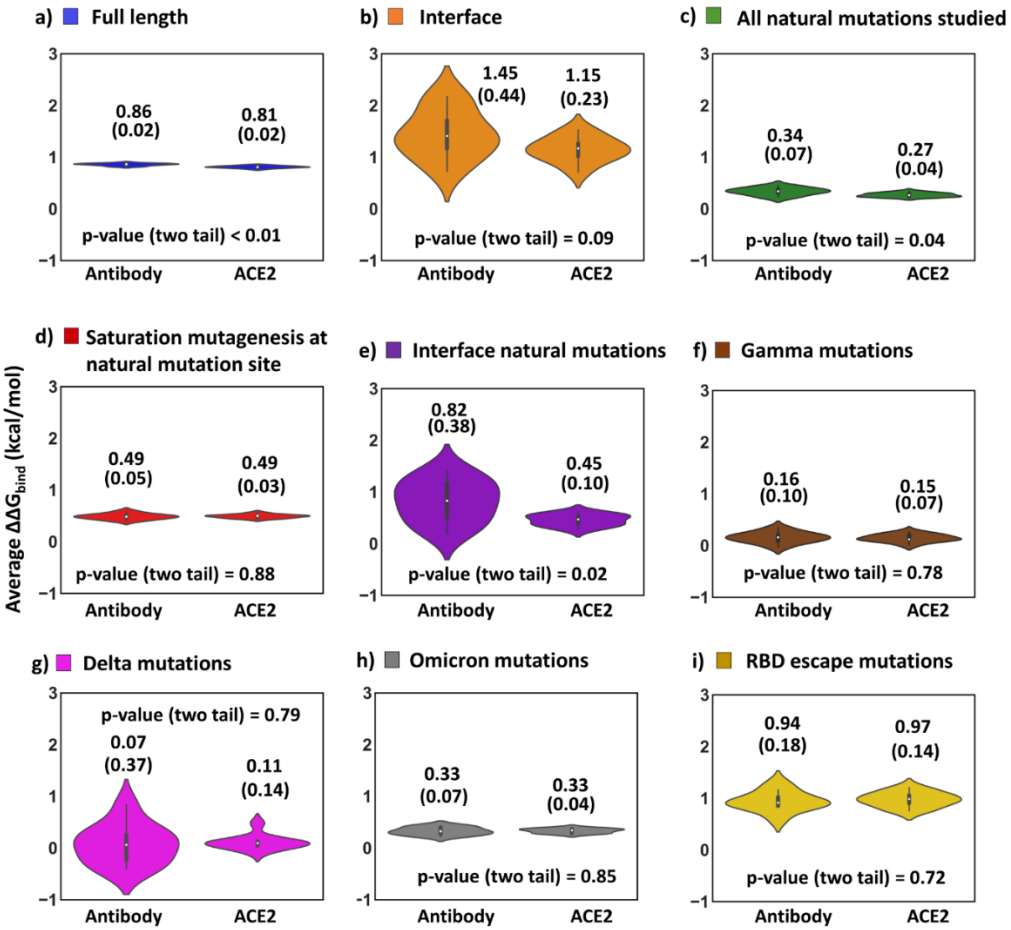

Figure 4. Comparison of binding affinity changes for S-protein binding to antibody or ACE2. Average $\Delta\Delta G_{bind}$ value, standard deviation (in bracket), and p-value (two tail) in each subplot. A p-value smaller than 0.05 was used to indicate high probability (95% confidence level) that the mean mutation effects differ for the two types of protein targets.

We further studied the conformational dependence of the binding affinity changes by dividing structures into subsets based on the number of AB/ACE2 binding partners and RBD conformation as detailed in **Supplementary section 3**. A very similar effect of each of these structural subsets was found, which is consistent with the combined effect of the ten AB/ACE2 structures, indicating that the results are not influenced by the change in the number of binding partners or different RBD conformations.

Estimating fitness effects by a simple model

The calculated change in fitness values (ΔF, equation 3) were plotted in a heatmap (**Figure 5** and **S9**) for all natural mutations studied for each ACE2 and antibody-bound complex. ΔF was calculated in three ways: **Figure 5a** displays ΔF as the difference between average $\Delta\Delta G_{bind}$ for 10 structures in ACE2 set and $\Delta\Delta G_{bind}$ for each structure in the AB set; similarly, **Figure 5b** shows ΔF as the difference between $\Delta\Delta G_{bind}$ for each structure in ACE2 set and the average $\Delta\Delta G_{bind}$ for 10 structures in AB set (X-axis shows the PDB id; the average column in the average value for the 10 structures); whereas **Figure 5c** shows average ΔF as the difference in the average $\Delta\Delta G_{bind}$ values of 10 structures between each ACE2 and AB sets.

Clearly, these individual mutations and average values in each **Figure 5a, 5b** and **5c** differ by the choice of structural set used for computation. Therefore, relying on one structure is not sufficient for drawing any conclusions. The consensus obtained in **Figure 5c** is of interest because it considers variable antibody epitope and structural conformations. Here, a more negative (green) value indicates more ACE2 binding or more antibody escape, whereas a more positive (red) value shows that the mutation increases binding to antibodies relative to ACE2. Residues P26S, G142D, E154K, D215G, N440K, S373P, L452R, E484A, E484K, F490S, N501Y, A570D, L981F, and S982A have more affinity for ACE2 receptor (cutoff < -0.25). These variations might lead to increased viral fitness and cause antibody escape. T19R, T76I, T95I, D138Y, Y145D, W152C, G446S, Y453F, T478K, N501T, Y505H, and D614G reveal a less binding affinity for ACE2 but more binding for antibodies (cutoff > 0.25). This mutation set reveals higher antibody binding that may neutralize the viral particles.

This simple fitness estimate, the difference between change in binding of ACE2 and antibody, can be used to anticipate important mutations leading to increased or decreased viral fitness. This initial computational effort to model fitness is further validated by comparison with experimental data below.

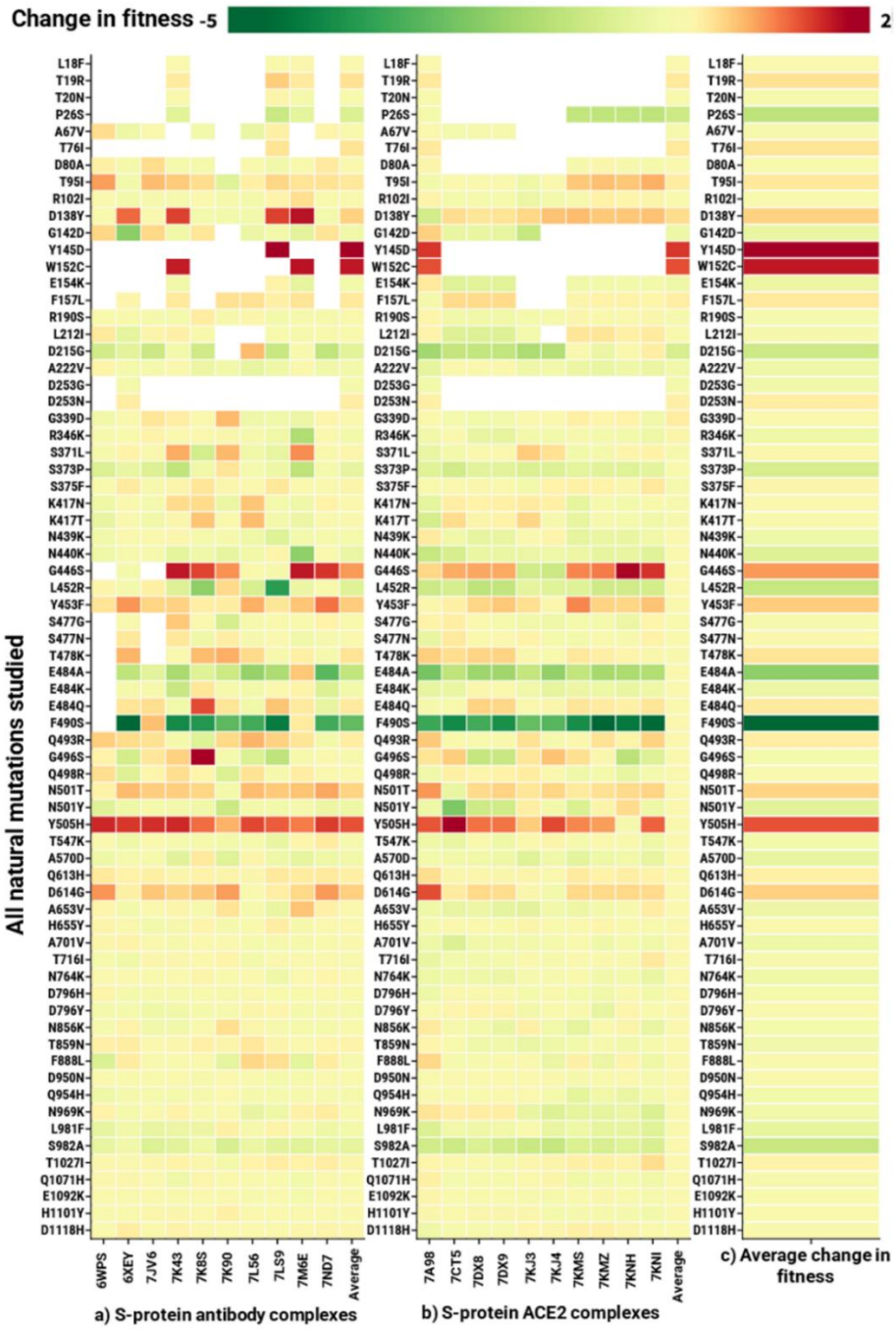

Figure 5. **Heatmap showing the estimated change in fitness using a simple two-property fitness model. (a)** ΔF = Average $\Delta\Delta G_{bind(MUT,\ 10\ ACE2)} - \Delta\Delta G_{bind(MUT,\ each\ AB)}$. The x-axis shows the PDB ID of the antibody set, with values being averages of 10 structures. **(b)** $\Delta F = \Delta\Delta G_{bind(MUT,\ each\ ACE2)}$ – Average $\Delta\Delta G_{bind(MUT,\ 10\ AB)}$. The x-axis shows the PDB ID of the ACE2 set, with values being averages of 10 structures. **(c)** Average ΔF = Average $\Delta\Delta G_{bind(MUT,\ 10\ ACE2)}$ – Average $\Delta\Delta G_{bind(MUT,\ 10\ AB)}$.

Validation of the approach

To validate our computational strategy, we compared our results with the three experimental datasets (0-2) defined in **Table S11**. The experimental ACE2 binding data[57] (**dataset 0**) was compared with the computed $\Delta\Delta G_{bind}$ for ACE2 (**Figure S10**). The average $\Delta\Delta G_{bind}$ for mutations in ACE2 complexes showed a highly significant correlation with R = 0.37 ($p < 0.01$), and when using binned method with 0.25 binned data, the correlation increased to R = 0.84 ($p < 0.01$) (**Figure S10**), as reported earlier.[58] This indicating the experimental ACE2 binding data are in parts defined by our computations. It is important to note here that three experimental datasets are highly squeezed with very less difference between several data points. However, comparing while taking average of small intervals in the data defined as bins may be more suitable for any comparison, as reported earlier.[59] Therefore, we used the binned strategy (0.25 bins) for all further comparisons.

Figure S11 shows the comparison of the experimental ACE2 binding (**dataset 0**) versus -log (antibody escape fractions) (**datasets 1** and **2**). Clearly, the direct correlations (R = 0.84 for dataset 0 vs. 1, and R = 0.56 for dataset 0 vs. 2 using 0.25 bins) confirm our argument about the generic effects of mutations in protein-protein interactions, i.e., mutation that tend to decrease S-protein affinity for one type of protein (ACE2) also reduces binding for the other type of protein (say antibodies).

The second approach involved comparison of two sets of experimental AB escape fraction data (**datasets 1** and **2**; **Table S11**) with our computed AB binding effects. Each of the datasets 1 and 2 was comprised of the average effect of ten antibodies. The average $\Delta\Delta G_{bind}$ for each mutation was correlated with experimental values for AB average mutation escape.[74,75] Computed change in binding correlates significantly with mutation escape data with the same and meaningful direction (**Figure S12a;** dataset 1, R = 0.25 and $p < 0.01$; dataset 2, R = 0.21 and $p < 0.01$). Interestingly, the correlation increased to R = 0.85 ($p < 0.01$) and 0.60 ($p = 0.06$) for datasets 1 and 2 using 0.25 bins (**Figure S12a**). We also specifically checked the correlation of the AB-bound S-proteins present in our dataset with the experimental AB escape data (**Figure S12b**). We observed a good correlation for experimental mutation escape of C002 antibody with calculated $\Delta\Delta G_{bind}$ for the C002 PDB structure (PDB code 7K8S; R= 0.66, p-value = 0.02 for 0.25 bins) with similar direction of regression line as with datasets 1 and 2. Similar trend was observed for C144 antibody correlation (PDB code 7K90; R = 0.65, p-value

= 0.01 for 0.25 bins) (**Figure S12b**). This suggests that our computations partially explain the mutation escape data.

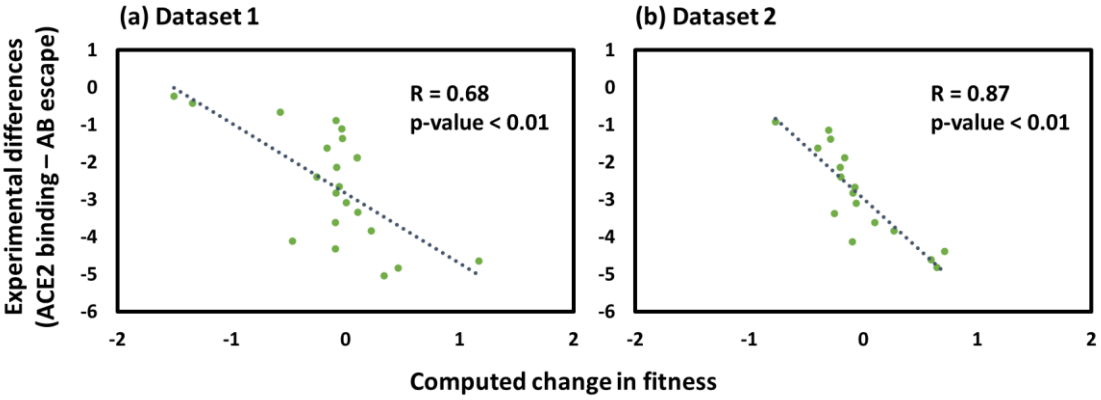

Figure 6. Scatter plots for the comparison of computed change in fitness with experimental differences (ACE2 binding – AB escape). Experimental data were binned into 0.25 bins and correlated with the computed data. **(a)** Dataset 1. **(b)** Dataset 2.

Most importantly, in the third approach, the experimental differences (ACE2 – AB), using **Table S11** data, were compared with computed change in fitness for the RBD mutations (**Figures 6** and **S13**). Interestingly, for both comparisons (ACE2 – AB datasets 1 and 2), experimental differences correlated significantly with computed differences, which was a fair comparison because the computed fitness is the difference between change in S-protein binding to ACE2 and AB (i.e. ACE2 – AB). The experimental values are the difference between ACE2 binding and -log (AB escape fraction). The binned data showed good correlations with R values of 0.68 (p-value <0.01) and 0.87 (p-value <0.01) for datasets 1 and 2 respectively (**Figure 6**). The directions of correlations were meaningful and in accordance with the experimental datasets, as the computed fitness value decreased (more ACE2 binding), the experimental difference increased (more ACE2 binding), thus validating our approach.

Overall, our computed binding data was able to partially reproduce the effect shown by diverse experimental studies on ACE2 and antibodies (**Figures S10-S13**), indicating the potential of our approach in computing binding effects, and therefore the fitness.

Considering intraprotein epistasis

In order to understand the effect (including non-additive effects, i.e., "intraprotein epistasis") of multiple mutations simultaneously present in the S-protein, we compared ΔΔG (kcal/mol) calculated using FoldX[60] for different variants with all their mutations present simultaneously, as seen in the variants (**Figure 7a**). We observed that for gamma variants (G1 and G2), omicron and RBD escape show resemblance in stability, with apo S-protein being the most stable and ACE2 complexes being the least stable. However, for the delta variant, the ACE2 complexes were more stabilized than the AB complexes with the least stability change for apo S-protein. In the same analysis, we also compared the stability of different variants for apo and holo structures (**Figure S14**). We found that the delta variant has high stability in AB- and ACE2-bound complexes, whereas RBD escape mutation decreases the stability of the protein in apo and holo forms. In addition, omicron mutations are decreasing stability of ACE2 complexes largely compared to AB complexes.

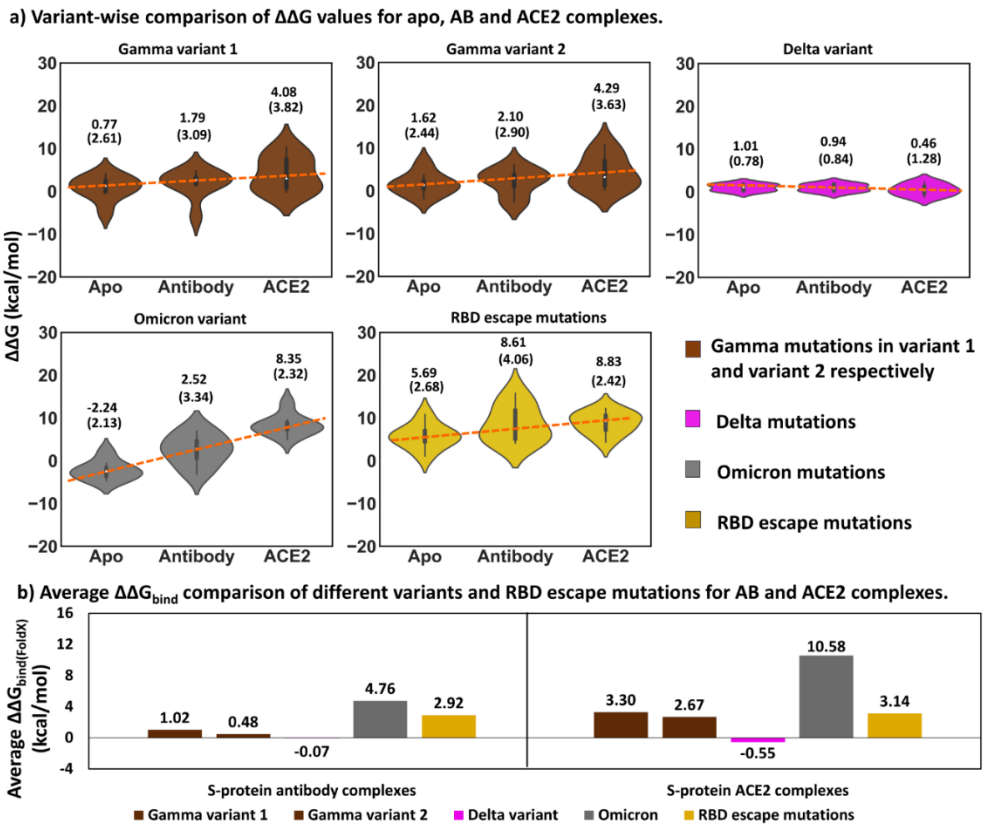

Figure 7. Summarizing epistasis in relation to ACE2 and AB binding. (a) ΔΔG values of apo, AB- and ACE2-bound complexes were compared for different variants using epistasis effect. **(b)** Comparing average change in binding free energy of different variants upon epistasis for AB-bound and ACE2-bound complexes. Positive ΔΔG show destabilizing and negative stabilizing effect of mutation. Similarly, positive $\Delta\Delta G_{bind}$ represents decreased binding.

When comparing the change in binding free energy ($\Delta\Delta G_{bind(FoldX)}$ = average $\Delta\Delta G_{(AB/ACE2)}$ – average $\Delta\Delta G_{(apo)}$) value for AB vs. ACE2 caused by mutation, a similar pattern was seen for the variants and RBD escape mutations (**Figure 7b**). A negative $\Delta\Delta G_{bind(FoldX)}$ value indicates increased binding affinity and *vice-versa*. The average values suggest that delta mutations stabilize both AB and ACE2 complexes, whereas omicron mutations tended to do the opposite, but the net effect is again the most important from a fitness perspective, as discussed above.

Limitations and strengths of the method

The fitness function defined here is a proof of the concept and not exact. It shows the ability of our approach to model fitness exceptionally well, realising the fact that some variables contributing to fitness were kept constant; still the simple two-property model can reproduce the effect, which is the novelty of our work. The model did not include the roles of innate and adaptive immunity in infection clearance, the evolved viral mechanisms for reduced detection,[76,77] and the dynamics of an immune response.[78] The advanced model might be expected to include such immune factors. However, our simplified model could partly reproduce the effects of several experimental datasets. It is an initial step towards using protein structure-based tools for making fitness functions at the virus organism level.

The model is based on the ten antibodies dataset binding to the RBD region similar to the ten ACE2 dataset binding. Such approach ensured consistency to reduce noise and maintain uniformity in computations. We do not deny the possibility of including more antibodies binding to multiple regions of S-protein, which is not considered here, since this may lead to heterogeneity of the approach. Some antibodies may not compete with ACE2 for binding to the S-protein yet still neutralize the virus.[41,79,80] Most of the experimental AB escape fraction data used for validating our approach are different from the set of computed AB data used in the present study, still the comparisons showed good correlations with consistent and meaningful direction of regression line, indicating the strength of our approach.

The experimental binding assay data and the experimental structures do not represent accurately the in vivo S-protein structure, which is heavily glycosylated.[81] Even beyond this glycan effect, the cryo-EM structures may not reflect the physiologically most representative conformation state due to the exposure of the samples to the cryo-temperature.[23,24,82] The rapid cooling may freeze some conformational dynamics,[83–91] and some of the conformations in the

S-protein may be pH- or temperature dependent in ways that will add complexity to real-world applications of the results.[92] Also, the experimental assays additionally are done on single mutants and thus do not reflect the same effects in variants where some introduced mutation may correlate (same-gene epistasis).

The use of differences in binding affinity caused by mutation, and differences of these between ACE2 and antibodies reduce systematic errors considerably, and the use of structure-averaged mean effects reduce structural noise and increase precision, which partly explain the robustness of our results. However, we note that systematic effects due to glycosylation could change conclusions reported both here and in the experimental assays, so the work here should be seen as a starting point for establishing fitness estimates using molecular modeling, not as final best estimates.

Conclusions

We have described a simple model and a computational protocol to estimate the fitness of arising mutations in the SARS-CoV-2 virus, based on the structure-based computational estimates of S-protein binding to a cocktail of important antibodies and human ACE2, emphasizing that fitness relates to a competitive binding situation, where the virion seeks to bind ACE2 before it is bound to circulating antibodies. The model uses many protein structures to compute robust estimates of these effects and takes advantage of a large degree of systematic error cancellation by considering differences between differences in estimated binding affinities (changes in ACE2–AB binding selectivity). Thus, while the model and protocol still have some limitations, it represents the best model so far in our estimate on a path towards predicting and rationalizing the impact of arising mutations in SARS-CoV-2 at the molecular level, work that is still in progress and requires much expansion before being applicable to e.g., other viruses as well.

Supplementary information

Details about the work are provided in the supplementary file. It contains details about the antibody and ACE2 datasets used in the study, group of mutations, conformational dependence of the binding affinity changes, experimental ACE2/antibody binding/escape datasets for

validation, detailed statistical tests (violin plots, correlation plots and t-test results), and epistatic analyses.

Acknowledgments

RM acknowledges SERB-SRG, Govt of India for funding via grant SRG/2022/000304-C. RM, ST acknowledge IIT Bhilai for supporting this work via Research Initiation Grant (RIG), number 2005900. ST acknowledges the Ministry of Education, Government of India, for PMRF fellowship grant number 5001800.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

RM and KPK conceptualized the work. RM supervised the work. ST performed the work. ST, RM and KPK wrote and refined the manuscript.

References

[1] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-Z.Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E.C. Holmes, and Y.-Z.Y.-L. Zhang, "A new coronavirus associated with human respiratory disease in China," Nature **579**(7798), 265–269 (2020).

[2] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G.F. Gao, and W. Tan, "A Novel Coronavirus from Patients with Pneumonia in China, 2019," N. Engl. J. Med. **382**(8), 727–733 (2020).

[3] M. Cevik, K. Kuppalli, J. Kindrachuk, and M. Peiris, "Virology, transmission, and pathogenesis of SARS-CoV-2," Bmj **371**, (2020).

[4] W.T. Harvey, A.M. Carabelli, B. Jackson, R.K. Gupta, E.C. Thomson, E.M. Harrison, C. Ludden, R. Reeve, A. Rambaut, and S.J. Peacock, "SARS-CoV-2 variants, spike mutations

and immune escape," Nat. Rev. Microbiol. **19**(7), 409–424 (2021).

[5] L. Guruprasad, "Human SARS CoV-2 spike protein mutations," Proteins Struct. Funct. Bioinforma. **89**(5), 569–576 (2021).

[6] L. van Dorp, C.J. Houldcroft, D. Richard, and F. Balloux, "COVID-19, the first pandemic in the post-genomic era," Curr. Opin. Virol., (2021).

[7] L. van Dorp, M. Acman, D. Richard, L.P. Shaw, C.E. Ford, L. Ormond, C.J. Owen, J. Pang, C.C.S.S. Tan, F.A.T.T. Boshier, A.T. Ortiz, and F. Balloux, "Emergence of genomic diversity and recurrent mutations in SARS-CoV-2," **83**, 104351 (2020).

[8] T.C. Williams, and W.A. Burgers, "SARS-CoV-2 evolution and vaccines: cause for concern?," Lancet Respir. Med. **9**(4), 333–335 (2021).

[9] K.R. McCarthy, L.J. Rennick, S. Nambulli, L.R. Robinson-McCarthy, W.G. Bain, G. Haidar, and W.P. Duprex, "Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape," Science (80-.). **371**(6534), 1139–1142 (2021).

[10] S.A. Kemp, D.A. Collier, R.P. Datir, I.A.T.M. Ferreira, S. Gayed, A.S. Jahun, M. Hosmillo, C. Rees-Spear, P. Mlcochova, I.U. Lumb, D.J. Roberts, A. Chandra, N. Temperton, S. Baker, G. Dougan, C. Hess, N. Kingston, P.J. Lehner, P.A. Lyons, N.J. Matheson, W.H. Owehand, C. Saunders, C. Summers, J.E.D. Thaventhiran, M. Toshner, M.P. Weekes, A. Bucke, J. Calder, L. Canna, J. Domingo, A. Elmer, S. Fuller, J. Harris, S. Hewitt, J. Kennet, S. Jose, J. Kourampa, A. Meadows, C. O'Brien, J. Price, C. Publico, R. Rastall, C. Ribeiro, J. Rowlands, V. Ruffolo, H. Tordesillas, B. Bullman, B.J. Dunmore, S. Fawke, S. Gräf, J. Hodgson, C. Huang, K. Hunter, E. Jones, E. Legchenko, C. Matara, J. Martin, F. Mescia, C. O'Donnell, L. Pointon, N. Pond, J. Shih, R. Sutcliffe, T. Tilly, C. Treacy, Z. Tong, J. Wood, M. Wylot, L. Bergamaschi, A. Betancourt, G. Bower, C. Cossetti, A. De Sa, M. Epping, S. Fawke, N. Gleadall, R. Grenfell, A. Hinch, O. Huhn, S. Jackson, I. Jarvis, D. Lewis, J. Marsden, F. Nice, G. Okecha, O. Omarjee, M. Perera, N. Richoz, V. Romashova, N.S. Yarkoni, R. Sharma, L. Stefanucci, J. Stephens, M. Strezlecki, L. Turner, E.M.D.D. De Bie, K. Bunclark, M. Josipovic, M. Mackay, S. Rossi, M. Selvan, S. Spencer, C. Yong, A. Ansaripour, A. Michael, L. Mwaura, C. Patterson, G. Polwarth, P. Polgarova, G. di Stefano, C. Fahey, R. Michel, S.-H. Bong, J.D. Coudert, E. Holmes, J. Allison, H. Butcher, D. Caputo, D. Clapham-Riley, E. Dewhurst, A. Furlong, B. Graves, J. Gray, T. Ivers, M. Kasanicki, E. Le Gresley, R. Linger, S. Meloy, F. Muldoon, N. Ovington, S. Papadia, I.

Phelan, H. Stark, K.E. Stirrups, P. Townsend, N. Walker, J. Webster, S.C. Robson, N.J. Loman, T.R. Connor, T. Golubchik, R.T. Martinez Nunez, C. Ludden, S. Corden, I. Johnston, D. Bonsall, C.C.P. Smith, A.R. Awan, G. Bucca, M. Estee Torok, K. Saeed, J.A. Prieto, D.K. Jackson, W.L. Hamilton, L.B. Snell, C. Moore, E.M. Harrison, S. Goncalves, D.J. Fairley, M.W. Loose, J. Watkins, R. Livett, S. Moses, R. Amato, S. Nicholls, M. Bull, D.L. Smith, J. Barrett, D.M. Aanensen, M.D. Curran, S. Parmar, D. Aggarwal, J.G. Shepherd, M.D. Parker, S. Glaysher, M. Bashton, A.P. Underwood, N. Pacchiarini, K.F. Loveson, A.M. Carabelli, K.E. Templeton, C.F. Langford, J. Sillitoe, T.I. de Silva, D. Wang, D. Kwiatkowski, A. Rambaut, J. O'Grady, S. Cottrell, M.T.G. Holden, E.C. Thomson, H. Osman, M. Andersson, A.J. Chauhan, M.O. Hassan-Ibrahim, M. Lawniczak, A. Alderton, M. Chand, C. Constantinidou, M. Unnikrishnan, A.C. Darby, J.A. Hiscox, S. Paterson, I. Martincorena, D.L. Robertson, E.M. Volz, A.J. Page, O.G. Pybus, A.R. Bassett, C. V Ariani, M.H. Spencer Chapman, K.K. Li, R.N. Shah, N.G. Jesudason, Y. Taha, M.P. McHugh, R. Dewar, A.S. Jahun, C. McMurray, S. Pandey, J.P. McKenna, A. Nelson, G.R. Young, C.M. McCann, S. Elliott, H. Lowe, B. Temperton, S. Roy, A. Price, S. Rey, M. Wyles, S. Rooke, S. Shaaban, M. de Cesare, L. Letchford, S. Silveira, E. Pelosi, E. Wilson-Davies, M. Hosmillo, The CITIID-NIHR BioResource COVID-19 Collaboration, The COVID-19 Genomics U K (COG-UK) Consortium, K. Sharrocks, E. Blane, Y. Modis, K. E. Leigh, J. A. G. Briggs, M. J. van Gils, K. G. C. Smith, J. R. Bradley, C.C.P. Smith, R. Doffinger, L. Ceron-Gutierrez, G. Barcenas-Morales, D. D. Pollock, R. A. Goldstein, A. Smielewska, J. P. Skittrall, T. Gouliouris, I. G. Goodfellow, E. Gkrania-Klotsas, C. J. R. Illingworth, L. E. McCoy, and R. K. Gupta, "SARS-CoV-2 evolution during treatment of chronic infection," Nature **592**(7853), 277–282 (2021).

[11] S.P. Otto, T. Day, J. Arino, C. Colijn, J. Dushoff, M. Li, S. Mechai, G. Van Domselaar, J. Wu, D.J.D. Earn, and N.H. Ogden, "The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic," Curr. Biol. **31**(14), R918–R929 (2021).

[12] M. Letko, A. Marzi, and V. Munster, "Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses," Nat. Microbiol. **5**(4), 562–569 (2020).

[13] A.R. Fehr, and S. Perlman, *Coronaviruses: An Overview of Their Replication and Pathogenesis* (New York, 2015).

[14] Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, and K.-Y. Yuen, "Structural and functional basis of SARS-CoV-2 entry by using human ACE2," Cell **181**(4), 894–904 (2020).

[15] C. Liu, Q. Zhou, Y. Li, L. V. Garner, S.P. Watkins, L.J. Carter, J. Smoot, A.C. Gregg, A.D. Daniels, S. Jervey, and D. Albaiu, "Research and Development on Therapeutic Agents and Vaccines for COVID-19 and Related Human Coronavirus Diseases," ACS Cent. Sci. **6**(3), 315–331 (2020).

[16] G. Forni, A. Mantovani, G. Forni, A. Mantovani, L. Moretta, R. Rappuoli, G. Rezza, A. Bagnasco, G. Barsacchi, G. Bussolati, M. Cacciari, P. Cappuccinelli, E. Cheli, R. Guarini, M.L. Bacci, M. Mancini, C. Marcuzzo, M.C. Morrone, G. Parisi, G. Pasquino, C. Patrono, A.Q. Curzio, G. Remuzzi, A. Roncaglia, S. Schiaffino, and P. Vineis, "COVID-19 vaccines: where we stand and challenges ahead," **28**(2), 626–639 (2021).

[17] A.E. Lobkovsky, Y.I. Wolf, and E. V Koonin, "Universal distribution of protein evolution rates as a consequence of protein folding physics.," Proc. Natl. Acad. Sci. U. S. A. **107**(7), 2983–8 (2010).

[18] P. Dasmeh, A.W.R. Serohijos, K.P. Kepp, and E.I. Shakhnovich, "Positively Selected Sites in Cetacean Myoglobins Contribute to Protein Stability," PLoS Comput. Biol. **9**(3), e1002929 (2013).

[19] D. a Liberles, S. a Teichmann, I. Bahar, U. Bastolla, J. Bloom, E. Bornberg-Bauer, L.J. Colwell, a P.J. de Koning, N. V Dokholyan, J. Echave, A. Elofsson, D.L. Gerloff, R. a Goldstein, J. a Grahnen, M.T. Holder, C. Lakner, N. Lartillot, S.C. Lovell, G. Naylor, T. Perica, D.D. Pollock, T. Pupko, L. Regan, A. Roger, N. Rubinstein, E. Shakhnovich, K. Sjölander, S. Sunyaev, A.I. Teufel, J.L. Thorne, J.W. Thornton, D.M. Weinreich, and S. Whelan, "The interface of protein structure, protein biophysics, and molecular evolution.," Protein Sci. **21**(6), 769–785 (2012).

[20] M. Bajaj, and T. Blundell, "Evolution and the tertiary structure of proteins," Annu. Rev. Biophys. Bioeng. **13**(1), 453–492 (1984).

[21] C.S. Wylie, and E.I. Shakhnovich, "A biophysical protein folding model accounts for most mutational fitness effects in viruses.," Proc. Natl. Acad. Sci. U. S. A. **108**(24), 9916–9921 (2011).

[22] T.L. Blundell, and A.K. Chaplin, “The resolution revolution in X-ray diffraction, Cryo-EM and other Technologies,” Prog. Biophys. Mol. Biol. **160**, 2–4 (2021).

[23] K. Murata, and M. Wolf, “Cryo-electron microscopy for structural analysis of dynamic biological macromolecules,” Biochim. Biophys. Acta - Gen. Subj. **1862**(2), 324–334 (2018).

[24] R. Fernandez-Leiro, and S.H.W. Scheres, “Unravelling biological macromolecules with cryo-electron microscopy,” Nature **537**(7620), 339–346 (2016).

[25] R. Danev, H. Yanagisawa, and M. Kikkawa, “Cryo-electron microscopy methodology: current aspects and future directions,” Trends Biochem. Sci. **44**(10), 837–848 (2019).

[26] R. Mehra, and K.P. Kepp, “Structure and Mutations of SARS-CoV-2 Spike Protein: A Focused Overview,” ACS Infect. Dis. **8**(1), 29–58 (2022).

[27] F. Pucci, K. V. Bernaerts, J.M. Kwasigroch, and M. Rooman, “Quantification of biases in predictions of protein stability changes upon mutations,” Bioinformatics **34**(21), 3659–3665 (2018).

[28] O. Caldararu, R. Mehra, T.L. Blundell, and K.P. Kepp, “Systematic Investigation of the Data Set Dependency of Protein Stability Predictors,” J. Chem. Inf. Model., (2020).

[29] O. Caldararu, T.L. Blundell, and K.P. Kepp, “A base measure of precision for protein stability predictors: structural sensitivity,” BMC Bioinformatics **22**(1), (2021).

[30] S. Iqbal, F. Li, T. Akutsu, D.B. Ascher, G.I. Webb, and J. Song, “Assessing the performance of computational predictors for estimating protein stability changes upon missense mutations,” Brief. Bioinform., bbab184 (2021).

[31] N.J. Christensen, and K.P. Kepp, “Accurate stabilities of laccase mutants predicted with a modified FoldX protocol,” J. Chem. Inf. Model. **52**(11), 3028–3042 (2012).

[32] K.P. Kepp, “Towards a ‘Golden Standard’ for computing globin stability: Stability and structure sensitivity of myoglobin mutants,” Biochim. Biophys. Acta - Proteins Proteomics **1854**(10), 1239–1248 (2015).

[33] S. Khan, and M. Vihinen, “Performance of protein stability predictors.,” **31**(6), 675–684 (2010).

[34] B.B. V Louis, and L.A. Abriata, “Reviewing Challenges of Predicting Protein Melting Temperature Change Upon Mutation Through the Full Analysis of a Highly Detailed Dataset

with High-Resolution Structures," Mol. Biotechnol. **in press**, (2021).

[35] F. Pucci, M. Schwersensky, and M. Rooman, "Artificial intelligence challenges for predicting the impact of mutations on protein stability," Curr. Opin. Struct. Biol. **72**, 161–168 (2022).

[36] R. Mehra, and K.P. Kepp, "Structural heterogeneity and precision of implications drawn from cryo-electron microscopy structures: SARS-CoV-2 spike-protein mutations as a test case," Eur. Biophys. J. **51**(7–8), 555–568 (2022).

[37] D. Pinto, Y.-J. Park, M. Beltramello, A.C. Walls, M.A. Tortorici, S. Bianchi, S. Jaconi, K. Culap, F. Zatta, A. De Marco, A. Peter, B. Guarino, R. Spreafico, E. Cameroni, J.B. Case, R.E. Chen, C. Havenar-Daughton, G. Snell, A. Telenti, H.W. Virgin, A. Lanzavecchia, M.S. Diamond, K. Fink, D. Veesler, and D. Corti, "Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody," Nature **583**(7815), 290–295 (2020).

[38] M. Rapp, Y. Guo, E.R. Reddem, J. Yu, L. Liu, P. Wang, G. Cerutti, P. Katsamba, J.S. Bimela, F.A. Bahna, S.M. Mannepalli, B. Zhang, P.D. Kwong, Y. Huang, D.D. Ho, L. Shapiro, and Z. Sheng, "Modular basis for potent SARS-CoV-2 neutralization by a prevalent VH1-2-derived antibody class," Cell Rep. **35**(1), (2021).

[39] L. Liu, P. Wang, M.S. Nair, J. Yu, M. Rapp, Q. Wang, Y. Luo, J.F.-W. Chan, V. Sahi, A. Figueroa, X. V Guo, G. Cerutti, J. Bimela, J. Gorman, T. Zhou, Z. Chen, K.-Y. Yuen, P.D. Kwong, J.G. Sodroski, M.T. Yin, Z. Sheng, Y. Huang, L. Shapiro, and D.D. Ho, "Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike," Nature **584**(7821), 450–456 (2020).

[40] L. Piccoli, Y.-J. Park, M.A. Tortorici, N. Czudnochowski, A.C. Walls, M. Beltramello, C. Silacci-Fregni, D. Pinto, L.E. Rosen, J.E. Bowen, O.J. Acton, S. Jaconi, B. Guarino, A. Minola, F. Zatta, N. Sprugasci, J. Bassi, A. Peter, A. De Marco, J.C. Nix, F. Mele, S. Jovic, B.F. Rodriguez, S. V Gupta, F. Jin, G. Piumatti, G. Lo Presti, A.F. Pellanda, M. Biggiogero, M. Tarkowski, M.S. Pizzuto, E. Cameroni, C. Havenar-Daughton, M. Smithey, D. Hong, V. Lepori, E. Albanese, A. Ceschi, E. Bernasconi, L. Elzi, P. Ferrari, C. Garzoni, A. Riva, G. Snell, F. Sallusto, K. Fink, H.W. Virgin, A. Lanzavecchia, D. Corti, and D. Veesler, "Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology," Cell **183**(4), 1024-1042.e21 (2020).

[41] C.O. Barnes, C.A. Jette, M.E. Abernathy, K.-M.A. Dam, S.R. Esswein, H.B. Gristick, A.G. Malyutin, N.G. Sharaf, K.E. Huey-Tubman, Y.E. Lee, D.F. Robbiani, M.C. Nussenzweig, A.P. West, and P.J. Bjorkman, "SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies," Nature **588**(7839), 682–687 (2020).

[42] W. Dejnirattisai, D. Zhou, H.M. Ginn, H.M.E. Duyvesteyn, P. Supasa, J.B. Case, Y. Zhao, T.S. Walter, A.J. Mentzer, C. Liu, B. Wang, G.C. Paesen, J. Slon-Campos, C. López-Camacho, N.M. Kafai, A.L. Bailey, R.E. Chen, B. Ying, C. Thompson, J. Bolton, A. Fyfe, S. Gupta, T.K. Tan, J. Gilbert-Jaramillo, W. James, M. Knight, M.W. Carroll, D. Skelly, C. Dold, Y. Peng, R. Levin, T. Dong, A.J. Pollard, J.C. Knight, P. Klenerman, N. Temperton, D.R. Hall, M.A. Williams, N.G. Paterson, F.K.R. Bertram, C.A. Siebert, D.K. Clare, A. Howe, J. Radecke, Y. Song, A.R. Townsend, K.-Y.A. Huang, E.E. Fry, J. Mongkolsapaya, M.S. Diamond, J. Ren, D.I. Stuart, and G.R. Screaton, "The antigenic anatomy of SARS-CoV-2 receptor binding domain," Cell **184**(8), 2183-2200.e22 (2021).

[43] G. Cerutti, M. Rapp, Y. Guo, F. Bahna, J. Bimela, E.R. Reddem, J. Yu, P. Wang, L. Liu, Y. Huang, D.D. Ho, P.D. Kwong, Z. Sheng, and L. Shapiro, "Structural basis for accommodation of emerging B.1.351 and B.1.1.7 variants by two potent SARS-CoV-2 neutralizing antibodies," Structure **29**(7), 655-663.e4 (2021).

[44] M.A. Tortorici, M. Beltramello, F.A. Lempp, D. Pinto, H. V. Dang, L.E. Rosen, M. McCallum, J. Bowen, A. Minola, S. Jaconi, F. Zatta, A. De Marco, B. Guarino, S. Bianchi, E.J. Lauron, H. Tucker, J. Zhou, A. Peter, C. Havenar-Daughton, J.A. Wojcechowskyj, J.B. Case, R.E. Chen, H. Kaiser, M. Montiel-Ruiz, M. Meury, N. Czudnochowski, R. Spreafico, J. Dillen, C. Ng, N. Sprugasci, K. Culap, F. Benigni, R. Abdelnabi, S.-Y.C.Y.C. Foo, M.A. Schmid, E. Cameroni, A. Riva, A. Gabrieli, M. Galli, M.S. Pizzuto, J. Neyts, M.S. Diamond, H.W. Virgin, G. Snell, D. Corti, K. Fink, and D. Veesler, "Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms," **370**(6519), 950–957 (2020).

[45] J.F. Scheid, C.O. Barnes, B. Eraslan, A. Hudak, J.R. Keeffe, L.A. Cosimi, E.M. Brown, F. Muecksch, Y. Weisblum, S. Zhang, T. Delorey, A.E. Woolley, F. Ghantous, S.-M. Park, D. Phillips, B. Tusi, K.E. Huey-Tubman, A.A. Cohen, P.N.P. Gnanapragasam, K. Rzasa, T. Hatziioanno, M.A. Durney, X. Gu, T. Tada, N.R. Landau, A.P. West Jr., O. Rozenblatt-Rosen, M.S. Seaman, L.R. Baden, D.B. Graham, J. Deguine, P.D. Bieniasz, A. Regev, D. Hung, P.J. Bjorkman, and R.J. Xavier, "B cell genomics behind cross-neutralization of

SARS-CoV-2 variants and SARS-CoV," Cell **184**(12), 3205-3221.e24 (2021).

[46] D.J. Benton, A.G. Wrobel, P. Xu, C. Roustan, S.R. Martin, P.B. Rosenthal, J.J. Skehel, and S.J. Gamblin, "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion," Nature **588**(7837), 327–330 (2020).

[47] L. Guo, W. Bi, X. Wang, W. Xu, R. Yan, Y. Zhang, K. Zhao, Y. Li, M. Zhang, X. Cai, S. Jiang, Y. Xie, Q. Zhou, L. Lu, and B. Dang, "Engineered trimeric ACE2 binds viral spike protein and locks it in 'Three-up' conformation to potently inhibit SARS-CoV-2 infection," Cell Res. **31**(1), 98–100 (2021).

[48] R. Yan, Y. Zhang, Y. Li, F. Ye, Y. Guo, L. Xia, X. Zhong, X. Chi, and Q. Zhou, "Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2," Cell Res., 1–3 (2021).

[49] T. Xiao, J. Lu, J. Zhang, R.I. Johnson, L.G.A. McKay, N. Storm, C.L. Lavine, H. Peng, Y. Cai, S. Rits-Volloch, S. Lu, B.D. Quinlan, M. Farzan, M.S. Seaman, A. Griffiths, and B. Chen, "A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent," Nat. Struct. Mol. Biol. **28**(2), 202–209 (2021).

[50] T. Zhou, Y. Tsybovsky, J. Gorman, M. Rapp, G. Cerutti, G.-Y. Chuang, P.S. Katsamba, J.M. Sampson, A. Schön, J. Bimela, J.C. Boyington, A. Nazzari, A.S. Olia, W. Shi, M. Sastry, T. Stephens, J. Stuckey, I.-T. Teng, P. Wang, S. Wang, B. Zhang, R.A. Friesner, D.D. Ho, J.R. Mascola, L. Shapiro, and P.D. Kwong, "Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains," Cell Host Microbe **28**(6), 867-879.e5 (2020).

[51] Y. Dehouck, J.M. Kwasigroch, M. Rooman, and D. Gilis, "BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations.," Nucleic Acids Res. **41**, W333–W339 (2013).

[52] A. Meseguer, L. Dominguez, P.M. Bota, J. Aguirre-Plans, J. Bonet, N. Fernandez-Fuentes, and B. Oliva, "Using collections of structural models to predict changes of binding affinity caused by mutations in protein–protein interactions," Protein Sci. **29**(10), 2112–2130 (2020).

[53] D.F.A.R. Dourado, and S.C. Flores, "A multiscale approach to predicting affinity changes in protein–protein interfaces," Proteins Struct. Funct. Bioinforma. **82**(10), 2681–2690 (2014).

[54] K.T. Bæk, and K.P. Kepp, "Data set and fitting dependencies when estimating protein

mutant stability: Toward simple, balanced, and interpretable models," J. Comput. Chem. **in press**(n/a), (n.d.).

[55] R. Mehra, and K.P.K.P. Kepp, "Computational analysis of Alzheimer-causing mutations in amyloid precursor protein and presenilin 1," Arch. Biochem. Biophys. **678**, 108168 (2019).

[56] B. Dehury, N. Tang, and K.P. Kepp, "Insights into membrane-bound presenilin 2 from all-atom molecular dynamics simulations," J. Biomol. Struct. Dyn., DOI: 10.1080/07391102.2019.1655481 (2019).

[57] T.N. Starr, A.J. Greaney, S.K. Hilton, D. Ellis, K.H.D. Crawford, A.S. Dingens, M.J. Navarro, J.E. Bowen, M.A. Tortorici, and A.C. Walls, "Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding," Cell **182**(5), 1295–1310 (2020).

[58] S. Thakur, R.K. Verma, K.P. Kepp, and R. Mehra, "Modelling SARS-CoV-2 spike-protein mutation effects on ACE2 binding," J. Mol. Graph. Model. **119**, 108379 (2023).

[59] K.T. Bæk, R. Mehra, and K.P. Kepp, "Stability and expression of SARS-CoV-2 spike-protein mutations," (2022).

[60] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, "The FoldX web server: An online force field," Nucleic Acids Res. **33**, W382–W388 (2005).

[61] A.C. Walls, Y.-J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein," Cell **181**(2), 281–292 (2020).

[62] N.G. Herrera, N.C. Morano, A. Celikgil, G.I. Georgiev, R.J. Malonis, J.H. Lee, K. Tong, O. Vergnolle, A.B. Massimi, L.Y. Yen, A.J. Noble, M. Kopylov, J.B. Bonanno, S.C. Garrett-Thomson, D.B. Hayes, R.H. Bortz, A.S. Wirchnianski, C. Florez, E. Laudermilch, D. Haslwanter, J.M. Fels, M.E. Dieterle, R.K. Jangra, J. Barnhill, A. Mengotto, D. Kimmel, J.P. Daily, L. Pirofski, K. Chandran, M. Brenowitz, S.J. Garforth, E.T. Eng, J.R. Lai, and S.C. Almo, "Characterization of the SARS-CoV-2 S Protein: Biophysical, Biochemical, Structural, and Antigenic Analysis," ACS Omega **6**(1), 85–102 (2021).

[63] M. McCallum, A.C. Walls, J.E. Bowen, D. Corti, and D. Veesler, "Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation," Nat. Struct. Mol. Biol. **27**(10), 942–949 (2020).

[64] T. Zhou, I.T. Teng, A.S. Olia, G. Cerutti, J. Gorman, A. Nazzari, W. Shi, Y. Tsybovsky, L. Wang, S. Wang, B. Zhang, Y. Zhang, P.S. Katsamba, Y. Petrova, B.B. Banach, A.S. Fahad, L. Liu, S.N. Lopez Acevedo, B. Madan, M. Oliveira de Souza, X. Pan, P. Wang, J.R. Wolfe, M. Yin, D.D. Ho, E. Phung, A. DiPiazza, L.A. Chang, O.M. Abiona, K.S. Corbett, B.J. DeKosky, B.S. Graham, J.R. Mascola, J. Misasi, T. Ruckwardt, N.J. Sullivan, L. Shapiro, and P.D. Kwong, "Structure-Based Design with Tag-Based Purification and In-Process Biotinylation Enable Streamlined Development of SARS-CoV-2 Spike Molecular Probes," Cell Rep. **33**(4), (2020).

[65] J. Huo, Y. Zhao, J. Ren, D. Zhou, H.M.E. Duyvesteyn, H.M. Ginn, L. Carrique, T. Malinauskas, R.R. Ruza, P.N.M. Shah, T.K. Tan, P. Rijal, N. Coombes, K.R. Bewley, J.A. Tree, J. Radecke, N.G. Paterson, P. Supasa, J. Mongkolsapaya, G.R. Screaton, M. Carroll, A. Townsend, E.E. Fry, R.J. Owens, and D.I. Stuart, "Neutralization of SARS-CoV-2 by Destruction of the Prefusion Spike," Cell Host Microbe **28**(3), 445-454.e6 (2020).

[66] Z. Lv, Y.-Q. Deng, Q. Ye, L. Cao, C.-Y. Sun, C. Fan, W. Huang, S. Sun, Y. Sun, L. Zhu, Q. Chen, N. Wang, J. Nie, Z. Cui, D. Zhu, N. Shaw, X.-F. Li, Q. Li, L. Xie, Y. Wang, Z. Rao, C.-F. Qin, and X. Wang, "Structural basis for neutralization of SARS-CoV-2 and SARS-CoV by a potent therapeutic antibody," Science (80-.). **369**(6510), 1505–1509 (2020).

[67] C. Zhang, Y. Wang, Y. Zhu, C. Liu, C. Gu, S. Xu, Y. Wang, Y. Zhou, Y. Wang, and W. Han, "Development and structural basis of a two-MAb cocktail for treating SARS-CoV-2 infections," Nat. Commun. **12**(1), 1–16 (2021).

[68] C. Xu, Y.Y.Y. Wang, C. Liu, C. Zhang, W. Han, X. Hong, Y.Y.Y. Wang, Q. Hong, S. Wang, Q. Zhao, Y.Y.Y. Wang, Y. Yang, K. Chen, W. Zheng, L. Kong, F. Wang, Q. Zuo, Z. Huang, and Y. Cong, "Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM," **7**(1), (2021).

[69] P.-A. Koenig, H. Das, H. Liu, B.M. Kümmerer, F.N. Gohr, L.-M. Jenster, L.D.J. Schiffelers, Y.M. Tesfamariam, M. Uchima, J.D. Wuerth, K. Gatterdam, N. Ruetalo, M.H. Christensen, C.I. Fandrey, S. Normann, J.M.P. Tödtmann, S. Pritzl, L. Hanke, J. Boos, M. Yuan, X. Zhu, J.L. Schmid-Burgk, H. Kato, M. Schindler, I.A. Wilson, M. Geyer, K.U. Ludwig, B.M. Hällberg, N.C. Wu, and F.I. Schmidt, "Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape," Science (80-.). **371**(6530), eabe6230 (2021).

[70] L. Yurkovetskiy, X. Wang, K.E. Pascal, C. Tomkins-Tinch, T.P. Nyalile, Y. Wang, A. Baum, W.E. Diehl, A. Dauphin, C. Carbone, K. Veinotte, S.B. Egri, S.F. Schaffner, J.E. Lemieux, J.B. Munro, A. Rafique, A. Barve, P.C. Sabeti, C.A. Kyratsous, N. V Dudkina, K. Shen, and J. Luban, "Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant," Cell **183**(3), 739-751.e8 (2020).

[71] L. Makowski, W. Olson-Sidford, and J.W. Weisel, "Biological and Clinical Consequences of Integrin Binding via a Rogue RGD Motif in the SARS CoV-2 Spike Protein.," Viruses **13**(2), 146–146 (2021).

[72] F. Tian, B. Tong, L. Sun, S. Shi, B. Zheng, Z. Wang, X. Dong, and P. Zheng, "N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2," Elife **10**, (2021).

[73] C. Laffeber, K. de Koning, R. Kanaar, and J.H.G. Lebbink, "Experimental Evidence for Enhanced Receptor Binding by Rapidly Spreading SARS-CoV-2 Variants," J. Mol. Biol. **433**(15), (2021).

[74] A.J. Greaney, T.N. Starr, C.O. Barnes, Y. Weisblum, F. Schmidt, M. Caskey, C. Gaebler, A. Cho, M. Agudelo, S. Finkin, Z. Wang, D. Poston, F. Muecksch, T. Hatziioannou, P.D. Bieniasz, D.F. Robbiani, M.C. Nussenzweig, P.J. Bjorkman, and J.D. Bloom, "Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies," Nat. Commun. **12**(1), (2021).

[75] A.J. Greaney, T.N. Starr, P. Gilchuk, S.J. Zost, E. Binshtein, A.N. Loes, S.K. Hilton, J. Huddleston, R. Eguia, K.H.D. Crawford, A.S. Dingens, R.S. Nargi, R.E. Sutton, N. Suryadevara, P.W. Rothlauf, Z. Liu, S.P.J. Whelan, R.H. Carnahan, J.E. Crowe, and J.D. Bloom, "Complete Mapping of Mutations to the SARS-CoV-2 Spike Receptor-Binding Domain that Escape Antibody Recognition," Cell Host Microbe **29**(1), 44–57 (2021).

[76] V. Naranbhai, A. Nathan, C. Kaseke, C. Berrios, A. Khatri, S. Choi, M.A. Getz, R. Tano-Menka, O. Ofoman, A. Gayton, F. Senjobe, Z. Zhao, K.J. St Denis, E.C. Lam, M. Carrington, W.F. Garcia-Beltran, A.B. Balazs, B.D. Walker, A.J. Iafrate, and G.D. Gaiha, "T cell reactivity to the SARS-CoV-2 Omicron variant is preserved in most but not all individuals," Cell **185**(6), 1041-1051.e6 (2022).

[77] G. Breton, P. Mendoza, T. Hägglöf, T.Y. Oliveira, D. Schaefer-Babajew, C. Gaebler, M. Turroja, A. Hurley, M. Caskey, and M.C. Nussenzweig, "Persistent cellular immunity to

SARS-CoV-2 infection," J. Exp. Med. **218**(4), (2021).

[78] V.K. Shah, P. Firmal, A. Alam, D. Ganguly, and S. Chattopadhyay, "Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past," Front. Immunol. **11**, 553450 (2020).

[79] X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, and W. Chen, "A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2," Science (80-.). **369**(6504), 650–655 (2020).

[80] M.I.J. Raybould, A. Kovaltsuk, C. Marks, and C.M. Deane, "CoV-AbDab: the coronavirus antibody database," Bioinformatics **37**(5), 734–735 (2021).

[81] L. Casalino, Z. Gaieb, J.A. Goldsmith, C.K. Hjorth, A.C. Dommer, A.M. Harbison, C.A. Fogarty, E.P. Barros, B.C. Taylor, J.S. McLellan, E. Fadda, and R.E. Amaro, "Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein," ACS Cent. Sci. **6**(10), 1722–1734 (2020).

[82] A. Meents, S. Gutmann, A. Wagner, and C. Schulze-Briese, "Origin and temperature dependence of radiation damage in biological samples at cryogenic temperatures," Proc. Natl. Acad. Sci. **107**(3), 1094–1099 (2010).

[83] R. Mehra, B. Dehury, and K.P. Kepp, "Cryo-temperature effects on membrane protein structure and dynamics," Phys. Chem. Chem. Phys. **in press.**, DOI: 10.1039/C9CP06723J (2020).

[84] A.H. Linden, W.T. Franks, Ü. Akbey, S. Lange, B.J. Van Rossum, and H. Oschkinat, "Cryogenic temperature effects and resolution upon slow cooling of protein preparations in solid state NMR," J. Biomol. NMR **51(3)**(3), 283 (2011).

[85] R.F. Tilton Jr, J.C. Dewan, and G.A. Petsko, "Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K," Biochemistry **31**(9), 2469–2481 (1992).

[86] W. Doster, A. Bachleitner, R. Dunau, M. Hiebl, and E. Lüscher, "Thermal properties of water in myoglobin crystals and solutions at subzero temperatures," Biophys. J. **50(2)**(2), 213–219 (1986).

[87] K. V Dunlop, R.T. Irvin, and B. Hazes, "Pros and cons of cryocrystallography: should we

also collect a room-temperature data set?," Acta Crystallogr. Sect. D Biol. Crystallogr. **61**(1), 80–87 (2005).

[88] B. Dehury, N. Tang, R. Mehra, T.L. Blundell, and K.P. Kepp, "Side-by-side comparison of Notch- And C83 binding to γ-secretase in a complete membrane model at physiological temperature," RSC Adv. **10**(52), 31215–31232 (2020).

[89] R. Mehra, and K.P. Kepp, "Understanding familial Alzheimer's disease: The fit-stay-trim mechanism of γ-secretase," Wiley Interdiscip. Rev. Comput. Mol. Sci. **12**(1), (2022).

[90] R. Mehra, and K.P. Kepp, "Computational prediction and molecular mechanism of γ-secretase modulators," Eur. J. Pharm. Sci., 105626 (2020).

[91] R. Mehra, and K.P. Kepp, "Identification of Structural Calcium Binding Sites in Membrane-Bound Presenilin 1 and 2," J. Phys. Chem. B **124**(23), 4697–4711 (2020).

[92] R.J. Edwards, K. Mansouri, V. Stalls, K. Manne, B. Watts, R. Parks, K. Janowska, S.M.C. Gobeil, M. Kopp, and D. Li, "Cold sensitivity of the SARS-CoV-2 spike ectodomain," Nat. Struct. Mol. Biol. **28**(2), 128–131 (2021).