Free Energy Perturbation Calculations of Mutation Effects on SARS-CoV-2 RBD:: ACE2 Binding Affinity

Alina P. Sergeeva [1], **Phinikoula S. Katsamba** [2], **Junzhuo Liao** [3], **Jared M. Sampson** [3,4], **Fabiana Bahna** [2], **Seetha Mannepalli** [2], **Nicholas C. Morano** [2], **Lawrence Shapiro** [2,5]∗, **Richard A. Friesner** [3]∗ and **Barry Honig** [1,2,5,6]∗

1 - *Department of Systems Biology,* Columbia University Medical Center, New York, NY 10032, USA
2 - *Zuckerman Mind Brain and Behavior Institute,* Columbia University, New York, NY 10027, USA
3 - *Department of Chemistry,* Columbia University, New York, NY 10027, USA
4 - *Schrödinger,* Inc., New York, NY 10036, USA
5 - *Department of Biochemistry and Molecular Biophysics,* Columbia University, New York, NY 10032, USA
6 - *Department of Medicine,* Columbia University, New York, NY 10032, USA

Correspondence to Lawrence Shapiro, Richard A. Friesner and Barry Honig: Columbia University, New York, NY, 10027, USA. *lss8@columbia.edu* (L. Shapiro), *raf8@columbia.edu* (R.A. Friesner), *bh6@columbia.edu* (B. Honig) *@_Alina_Sergeeva* 🐦 (A.P. Sergeeva)
https://doi.org/10.1016/j.jmb.2023.168187
Edited by Michael Sternberg

Abstract

The strength of binding between human angiotensin converting enzyme 2 (ACE2) and the receptor binding domain (RBD) of viral spike protein plays a role in the transmissibility of the SARS-CoV-2 virus. In this study we focus on a subset of RBD mutations that have been frequently observed in infected individuals and probe binding affinity changes to ACE2 using surface plasmon resonance (SPR) measurements and free energy perturbation (FEP) calculations. Our SPR results are largely in accord with previous studies but discrepancies do arise due to differences in experimental methods and to protocol differences even when a single method is used. Overall, we find that FEP performance is superior to that of other computational approaches examined as determined by agreement with experiment and, in particular, by its ability to identify stabilizing mutations. Moreover, the calculations successfully predict the observed cooperative stabilization of binding by the Q498R N501Y double mutant present in Omicron variants and offer a physical explanation for the underlying mechanism. Overall, our results suggest that despite the significant computational cost, FEP calculations may offer an effective strategy to understand the effects of interfacial mutations on protein–protein binding affinities and, hence, in a variety of practical applications such as the optimization of neutralizing antibodies.

Introduction

The ability to accurately predict binding affinity changes upon mutations of interfacial residues is a problem of significant importance, ranging from the general problem of understanding of interaction specificity and the design of therapeutics such as potent monoclonal antibodies that target antigens to revealing the mechanism of action of cancer driver mutations. Multiple approaches to the problem have been developed including machine learning,[1–5] statistical potentials[6] and various force field related scoring functions[7–11] embedded in programs such as

FoldX[11] and Rosetta.[7] Each approach is associated with its own set of issues, such as conformational changes upon mutation, that are nicely discussed in reviews of Bonvin and co-workers.[12] Moreover, some methods succeed on some test sets and fail on others, suggesting either over-training or simply that some protein–protein interfaces have different properties than others. Issues of experimental validation can also arise[13]; not all experimental methods are equally accurate and, as discussed below, the nuances of the experimental system can have significant effects on the outcome.

Detailed atomic-level simulations have not been extensively applied to the prediction of mutation effects, in part due to the computational requirements involved. Free-energy perturbation (FEP) methods have the potential to impact the field as physics-based force fields are, in principle, agnostic to the system being studied. Most current applications have involved the optimization of ligand–protein interaction in the context of small molecule drug design (reviewed in [14]) but recent publications have begun to explore the use of FEP methods to the study of protein–protein interactions (PPIs); specifically, to the effects of interfacial mutations on protein–protein binding free energies.[8,9,15–19] This is an inherently complex problem since, as opposed to relatively rigid ligand binding pockets, protein–protein interfaces are often quite large and less constrained so that they can more easily undergo conformational changes as a result of a mutation. Moreover, FEP calculations involve a complex computational infrastructure and are extremely time consuming. However, fast graphical processing units (GPUs) make such calculations feasible and a number of recent publications, involving different software packages, suggest that the methodology has reached the point that good correlation with experiment is to be expected.[8,9,15–19] Clearly, if FEP methods are capable of providing meaningful results, then in many applications, the computational cost will be worthwhile.

Here we explore the ability of FEP calculations to reproduce the effects of mutations on the binding of the receptor binding domain (RBD) of the SARS-CoV-2 spike protein with the human angiotensin converting enzyme 2 (ACE2) using the FEP+ implementation (see Methods). Given that the pathogen entry into the host cell is mediated by RBD::ACE2 binding, the problem has attracted considerable interest and multiple experimental[20–46] and computations studies[33,47–63] have been reported. We chose to study a set of 23 frequently observed RBD mutations (Table S1) located in the RBD::ACE2 interface (Figure 1) of Alpha, Beta, Gamma, Delta, or Omicron SARS-CoV-2 variants (Table S2). Surface Plasmon Resonance (SPR) experiments were carried out for each and compared to previous experimental work.[20–46] A number of computational methods were applied to predict binding free energy changes upon mutations, $\Delta\Delta G$. We found FEP to be the best performer and, moreover, FEP trajectory analysis facilitates the characterization of the biophysical effects that underlie mutational effects. We show that FEP successfully recapitulates the stabilizing epistatic effect of the Q498R N501Y mutant present in every Omicron variant. The ability to anticipate non-additive effects of multiple mutations is likely to be an important element of future efforts in protein interface design.

Results

SPR measurements of binding affinity changes

The second column in Table 1 lists experimental changes in binding affinity ($\Delta\Delta G$) of the ACE2::RBD complex when RBD is mutated. Of 23 single point RBD mutations probed, only four (N501Y, Y453F, S477N and N501T, Table 1, see Figure S1 for corresponding fitted SPR data) were identified as stabilizing with $\Delta\Delta G$ values ≤ -0.4 kcal/mol which is our measure of experimental accuracy (see Methods). The third column lists $\Delta\Delta G$ values from the deep mutational scanning study of Starr et al.[20] Although the methods are quite different and our SPR results are obtained
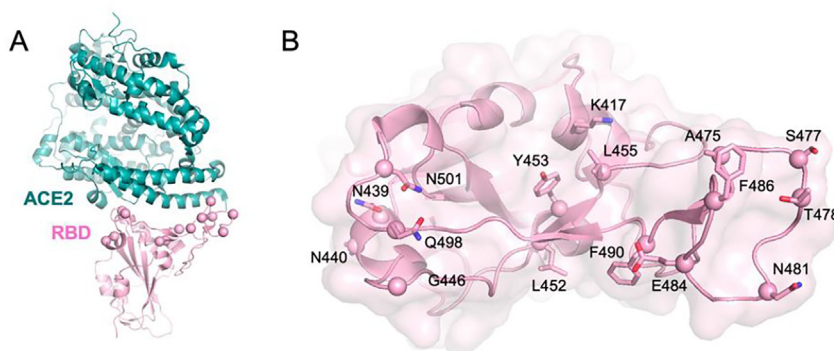

Figure 1. The ACE2/RBD system. (A) Ribbon representation of the ACE2/RBD complex. (B) Side chains of interfacial RBD residues in contact with ACE2 probed in this study are shown in stick representation with Cα atoms shown as spheres.

Table 1 Experimental ACE2/RBD binding affinity changes for RBD mutants.

Mutation	ΔΔG (SPR, this study)	ΔΔG (Yeast Display, Starr et al.)[20]	ΔΔG (other studies) SPR[21–33] BLI[34–46] Yeast Display[64] (range)
N501Y	−0.8	−0.3	(−1.7 . . . −0.5)
Y453F	−0.7	−0.3	(−1.2 . . . −0.8)
S477N	−0.5	−0.1	(−0.6 . . . −0.1)
N501T	−0.5	−0.1	(−0.9 . . . 0.0)
N439K	−0.1	0.0	(−0.4 . . . 0.0)
N440K	0.0	−0.1	
F490S	0.0	0.0	
L452M	0.0	−0.1	
L452R	0.0	0.0	−0.2
E484Q	0.1	0.0	
T478K	0.1	0.0	0.2
N481K	0.1	0.0	
E484K	0.1	0.1	(−0.6 . . . 0.3)
Q498R	0.2	0.1	(−0.5 . . . 0.5)
S477I	0.2	0.1	
G446V	0.2	0.4	0.5
T478R	0.2	0.1	
S477R	0.3	0.0	−0.5
A475V	0.3	0.2	0.2
L455F	0.4	0.3	
K417T	0.4	0.4	(0.1 . . . 0.7)
F486L	0.6	0.6	0.9
K417N	0.6	0.6	(−0.5 . . . 1.1)

Experimental binding affinity difference values were calculated based on binding affinity (K_D) measurements for wild-type (WT) and single mutant (MT) proteins using the following formula:
ΔΔG = RT ln (K_D(MT)/K_D(WT)), in units of kcal/mol.

with monomeric ACE2 while Starr et al. used dimeric ACE2, the results are in good agreement (high Pearson correlation coefficient (PCC) = 0.9 and low root mean square error (RMSE) = 0.2 kcal/mol). Given that our calculations are carried out on a structure containing monomeric ACE2 and the likelihood that the SPR results are more accurate than the high-throughput yeast display values, we use the SPR values to compare to computational predictions.

Table 1 also lists ΔΔG values obtained previously with SPR[21–33] and other experimental methods: bio-layer interferometry (BLI)[34–46] and yeast display.[64] Direct comparisons are difficult since, for example, different constructs were used in different experiments and different proteins were used as analytes in some SPR experiments (see Methods for details). The differences in constructs result from the choice of monomeric vs. multimeric forms of interacting proteins as well as the selection of protein domain boundaries. Nevertheless, overall, there is good agreement among most experimental results with the outliers attributable to the factors mentioned here. Moreover, there is good consensus regarding the identity of the most stabilizing mutations. For example, our results for N501Y (−0.8 kcal/mol), Y453F (−0.7 kcal/mol), S477N (−0.5 kcal/mol) and N501T (−0.5 kcal/mol) are in

good agreement with previously published values[20–32,34,36–38,40–46,65] regardless of constructs/experimental setup differences.

Computational prediction of the effect of mutations on binding affinity

Table 2 presents FEP results for the 23 experimental ΔΔG values obtained from SPR measurements listed in Table 1. Correlation plots for the data in Table 2 are given in Figure S2. The overall performance of FEP is in line with previous work[8,9,14,66–68]; a PCC of 0.6 and an RMSE of 0.8 kcal/mol (Table 2). The FEP calculations clearly predict that all four stabilizing mutations, N501Y, N501T, S477N, and Y453F have ΔΔG values < 0 although the prediction for S477N is a weak one. Of note, previous FEP calculations on N501Y[53,55,56] have yielded results very similar to ours attesting to the robustness of the method. The most significant failure of FEP is its prediction that A475V is stabilizing when the SPR results indicate that it is weakly destabilizing. A likely explanation for this result is that A475 is located close to the N-terminal residue of ACE2 (Q18) for which no coordinates were assigned in the crystal structure.[69] Another problematic result, a largely overestimated ΔΔG for Q498R, will be discussed below.

Table 2 **Calculated ACE2/RBD binding affinity changes for RBD mutants.** Pearson correlation coefficient (PCC), and root mean square error (RMSE) are calculated for every method tested based on comparison to SPR results in Table 1. Stabilizing mutations with $\Delta\Delta G \leq -0.4$ kcal/mol in green, destabilizing mutations with $\Delta\Delta G \geq 0.4$ kcal/mol in red. Experimental binding affinity difference values were calculated based on SPR binding affinity (K_D) measurements for wild-type (WT) and single mutant (MT) proteins using the following formula: $\Delta\Delta G = RT \ln (K_D(MT)/K_D(WT))$, in units of kcal/mol. Correlation plots for all theoretical methods are provided in Figure S2. Calculations were performed on a crystal structure of ACE2/RBD (PDBID 6M0J). Protein specific residue numbering of all the mutants as in Uniprot ID P0DTC2.

Mutation	ΔΔG experiment SPR	ΔΔG FEP 100ns	ΔΔG Mutabind2	ΔΔG mCSM-PPI2	ΔΔG SAAMBE-3D	ΔΔG BeAtMusic	ΔΔG FoldX	ΔΔG Rosetta flex ddG	ΔΔG MM/PB-SA	ΔΔG MM/GB-SA
N501Y	-0.8	-1.2	0.7	0.5	0.0	0.1	6.0	0.4	5.8	2.6
Y453F	-0.7	-0.6	-0.2	0.1	0.1	0.3	-0.4	-0.2	1.3	2.2
S477N	-0.5	-0.1	-0.1	-0.1	0.2	0.1	0.0	0.1	5.4	9.3
N501T	-0.5	-1.8	-0.6	-0.9	-0.1	0.4	-0.9	-0.4	3.1	1.8
N439K	-0.1	0.6	-0.1	0.5	0.6	0.3	0.0	0.0	-8.8	-0.8
N440K	0.0	-0.4	0.1	-0.1	0.5	0.0	-0.1	0.0	-1.9	-2.9
F490S	0.0	-0.1	0.6	0.2	0.8	0.7	0.0	0.0	-1.0	2.2
L452M	0.0	0.2	0.0	0.5	-0.3	0.2	0.0	0.0	6.3	2.1
L452R	0.0	-0.3	-0.8	0.0	0.1	0.2	-0.3	0.0	-6.8	-3.8
E484Q	0.1	0.3	0.1	0.3	0.5	0.2	-0.1	0.0	7.7	11.3
T478K	0.1	0.0	0.2	0.0	0.1	0.2	0.0	0.1	0.8	4.9
N481K	0.1	-0.1	0.0	0.0	0.6	0.2	0.0	0.0	-13.2	-7.4
E484K	0.1	0.2	0.2	0.3	0.3	0.1	-0.2	-0.1	3.1	7.2
Q498R	0.2	2.7	0.4	1.4	1.1	0.8	-0.5	-0.6	-1.0	-6.4
S477I	0.2	0.0	0.1	0.0	0.0	0.2	0.1	0.0	-1.3	6.1
G446V	0.2	0.4	-0.7	0.1	-0.3	1.3	0.1	0.0	6.0	6.6
T478R	0.2	-0.1	0.1	0.0	0.3	0.2	0.0	0.0	1.6	3.4
S477R	0.3	-0.2	0.2	-0.1	0.0	0.1	0.0	-0.2	0.6	3.1
A475V	0.3	-1.1	0.9	0.0	0.1	0.1	1.0	0.2	10.0	12.2
L455F	0.4	1.7	1.5	-0.7	0.2	-0.1	5.0	-0.6	6.5	5.1
K417T	0.4	0.9	0.2	0.4	0.5	0.5	0.8	0.7	11.4	14.1
F486L	0.6	1.3	0.3	0.9	0.5	0.7	1.1	1.8	5.9	7.8
K417N	0.6	0.9	0.6	0.5	0.4	0.5	0.8	0.9	12.2	13.1
PCC		0.6	0.3	0.2	0.3	0.2	-0.1	0.4	0.2	0.3
RMSE		0.8	0.5	0.5	0.5	0.5	1.7	0.5	6.5	6.9

The FEP results in Table 2 were obtained from 100 ns simulations. Extensive prior work[8,9] has demonstrated that shorter simulations are often inadequate for some mutations, typically those involving residues with a low degree of solvent exposure (i.e. partially or fully buried at the protein–protein interface). The results of 10 ns FEP simulations are shown in the Supplementary Material (Table S3) for comparison. Of note, the N501Y mutation, which is predicted to be destabilizing at 10 ns is correctly predicted to be stabilizing at 100 ns.

Evaluation of theoretical methods on the ACE2::RBD dataset (Table 2) shows that machine learning (ML) algorithms (Mutabind2,[5] mCSM-PPI2[3] and SAAMBE-3D[1]) uniformly have a weak correlation with experiment (PCC < 0.4). As we have pointed out previously[60] ML methods tend to overpredict destabilizing mutations presumably due to the preponderance of destabilizing mutations in training sets. A related factor likely accounts for relatively low RMSEs of ML methods since most mutations in training sets have only small effects on binding affinities. BeAtMuSiC evaluates mutation effects using a statistical potential[6] while FoldX uses an empirical physics-based force field[10, 11]. Both methods assume a rigid backbone although FoldX allows for side chain rearrangement upon mutation. Neither method produces a meaningful

correlation with experiment. Of note, poor performance of FoldX can be attributed to outliers (Figure S2) whereby mutations require backbone rearrangement. BeAtMuSiC identifies no stabilizing mutations while FoldX correctly identifies N501T and Y453F. Other than FEP, Rosetta flex ddG[7] is the only method that allows for backbone flexibility but its PCC is still significantly smaller than that of FEP and, as can be seen in Table 2, Rosetta flex ddG predicts two of the four stabilizing mutations (although the prediction for Y453F is a weak one). Nevertheless, even its partial success highlights the need to account for the ability of proteins to relax in response to interfacial mutations. Finally, we also tested Molecular Mechanics/Poisson–Boltzmann Surface Area (MM/PB-SA) and Molecular Mechanics/Generalized Born Surface Area (MM/GB-SA) methods, which utilize molecular mechanics energies along with a continuum representation of the solvent. As can be seen in Table 2, these methods performed poorly (see Methods for the protocol details).

The ACE2::RBD dataset has limitations for statistical analysis given the narrow affinity range of the experimental values (from -0.8 to $+0.6$ kcal/mol, Table 2) and the preponderance of neutral (15 out of 23) over stabilizing and destabilizing mutations (4 each). For a more robust assessment, we have compared

performance of various theoretical methods on a larger and more balanced dataset. Specifically, we have expanded the ACE2::RBD dataset with mutations from another system (DIP::Dpr). DIPs and Dprs are families of neuronal adhesion proteins whose members bind to one another with distinct subgroup-specific affinities.[70] We have previously studied a number of DIP::Dpr mutants using SPR[71–73] to establish specificity deteminants in this family. Here, we chose 18 single point mutations in Dpr6::DIP-α and Dpr10::DIPα complexes to combine with 23 ACE2::RBD mutations for a total of 41 data points. The combined ACE2::RBD DIP:: Dpr dataset has a 2.5 times wider range of $\Delta\Delta G$ values (−1.3 to +2.2 kcal/mol) than just ACE2::RBD alone and a more balanced distribution of stabilizing, neutral, and destabilizing mutations (Table S3).

Table S3 contains the results of different computational methods applied to the combined data set. We use multiple statistical tests to measure performance. Based on the same measures used in Table 2 (PCC and RMSE), all methods exhibit similar performance but with somewhat improved PCCs (except MM/GB-SA). FEP gives the best results with a PCC of 0.8 and RMSE of 1.0 kcal/mol. The next best performing method was Mutabind2 (PCC of 0.6 and RMSE of 0.8). Rosetta flex ddG and FoldX have a PCC of 0.5 and 0.4, respectively, with FoldX having a large RMSE error (1.8 kcal/mol) mostly due to high energetic penalties for mutations that introduce clashes at the interface which are difficult to correct when the backbone fixed. The continuum methods displayed poor performance (PCC < 0.3) and large errors (RMSE $\sim$ 6–7).

Table S3 also reports PCC values for the 18 observations in the DIP::Dpr data set alone. As is evident from the Table, FEP is by far the best performer with a PCC of 0.9, well above that of other methods. Of note, all methods perform better on the DIP::Dpr than on the ACE2::RBD data set. We attribute this observation to the fact that the DIP:Dpr interface is not prone to backbone rearrangement upon mutation.

We categorized mutations as stabilizing ($\Delta\Delta G \leq −0.4$ kcal/mol, green), neutral ($−0.4 < \Delta\Delta G < 0.4$ kcal/mol), or destabilizing ($\Delta\Delta G \geq 0.4$ kcal/mol, red). The cutoff of ±0.4 kcal/mol was chosen based on expected reproducibility error in SPR measurements (see Methods). We evaluated the ability of various theoretical methods to correctly classify mutations into each of these three categories - essentially yes or no regarding placement in a particular category as well as their ability to correctly classify the effects of each mutant into one of three categories (using Matthews correlation coefficient, MCC). As shown in Table S3, FoldX is best at classification into three categories (MCC (triple class) = 0.7; 32 out of 41 mutations correctly classified) and destabilizing mutations (MCC (dest) = 0.9). FEP is

the best method for prediction of stabilizing mutations (MCC (stab) = 0.5) and second best for three category classification (MCC (triple class) = 0.6; 31 out of 41 mutations correctly classified). The probability of achieving this result by random chance is approximately $10^{−8}$.

To assess the statistical significance of our FEP predictions, we calculated a p-value for the classification of stabilizing mutations, which was approximately $3.0 \times 10^{−11}$ (Table S3) under the assumption that the reproducibility error of the calculations was close to zero. To test this assumption, we performed ten independent repeats of the 100 ns FEP simulations for a select set of mutants. The standard deviation of $\Delta\Delta G$ values and the standard error of the mean (SEM) are reported in Table S4. On average, the $\Delta\Delta G$ values obtained starting with different random velocities deviate by approximately 0.2 kcal/mol (SEM $\sim$ 0.06). These small fluctuations would only affect the classification of mutations whose $\Delta\Delta G$ values are very close to the cutoff. In fact, categorizing stabilizing mutations is not affected by using a mean value of multiple independent runs (Table S4).

Physical insights from trajectory analysis

The N501Y mutation is responsible for a high infectivity and transmissibility of the Alpha variant of SARS-CoV-2[74] and has the largest stabilizing effect (Table 1). The N501T mutation found in the SARS-CoV-2 variants transmitted from mink to humans[75,76] occurs at the same position. Analysis of FEP trajectories reveals that the stabilization effect associated with N501Y and N501T mutations is due to substitution of the asparagine with less polar side chains of tyrosine and threonine. N501 has only one of its polar groups satisfied in the wild-type (WT) structure while, throughout the course of the relevant trajectories, the hydroxyl groups of both tyrosine and threonine participate in hydrogen bonds (see dashed lines, Figure 2 (A)). In addition to enhanced stability due to the absence of unsatisfied hydrogen bonds, both mutants undergo stabilizing interactions with Y41 of ACE2; the aromatic ring of Y501 participates in π-π stacking interactions (see purple lines in Figure 2(A)) while the methyl group of T501 forms a hydrophobic contact with the aromatic ring of Y41 (see gray shading in Figure 2(A)). Of note, previous studies have identified the role of π-π stacking as a source of stabilization of the N501Y mutant[47,53,56,58] but the role of the unsatisfied hydrogen bond in the WT protein has not been emphasized.

Analysis of trajectories associated with the Y453F mutation shows that in the WT protein, a hydrogen bond between the hydroxyl group of Y453 and the Nε atom of H34 is present in $\sim$25% of the WT trajectory (Figure 2(B)) while, in most cases, these two residues form hydrogen bonds with trapped

solvent molecules. In the Phe mutant, there is no need to satisfy the buried hydroxyl of the tyrosine while the Nε of H34 is satisfied by structured waters or backbone atoms (see dashed lines showing hydrogen bonds in Figure 2(B)). Thus, the enhanced stability of the mutant is likely due to the greater hydrophobicity of a Phe relative to a Tyr. Of note, our trajectory analysis is in agreement with a previous study comparing crystallographic structures of the WT and Y453F mutant.[32]

The simulations correctly predict that the S477N mutation is stabilizing but only weakly so. This residue faces the N-terminal of ACE2, and, specifically, the Q18 residue for which coordinates were missing in the crystal structure and, hence, modelled as an acetyl group cap in our simulations (Figure 2(C)). Thus, the calculations may suffer from conformational uncertainties in this region. The trajectories reveal that the longer Asn side chain makes more contacts with the ACE2 N-terminal than does the WT Ser. We show only one snapshot from wild-type and mutant trajectories in Figure 2(C), but in reality, due to the flexibility of the ACE2 N-terminal, no specific contact is retained throughout the simulations.

The largest error in the FEP calculations is for the Q498R mutant whose destabilizing effect is over-predicted by about 2.5 kcal/mol. Analysis of the 100 ns Q498R trajectory reveals that the Arg side chain samples different conformations with the most prevalent state (~53%) involving an unfavorable polar-hydrophobic contact between N501 and the aliphatic chain of the Arg, leaving both polar groups of the Asn unsatisfied (Figure 3 (A)). This is likely a contributor to the destabilizing ΔΔG value. As we have noted previously,[9] a computationally unfavorable outlier of this magnitude is typically attributable to the failure of the molecular dynamics trajectories (100 ns in this study) to achieve a converged reorganization of the protein structure. To test this possibility, we carried out a 300 ns simulation which appears to reach convergence at ~200 ns (Figure S3) where the destabilizing effect of Q498R is calculated to be 1.9 kcal/mol; reduced from 3.6 kcal/mol at 10 ns and 2.7 kcal/mol at 100 ns. It may well be the case that the system has converged to a metastable state at 300 ns and that there are lower energy states not sampled in the course of the simulation.

Epistatic effect of the Q498R N501Y double mutant in the omicron variant

Zahradnik et al.[64] demonstrated that the Q498R N501Y double mutant is more stabilizing than the additive effect of two single point mutations as estimated by their yeast display assay (a "cooperativity energy" of ~−1.7 kcal/mol). Our SPR results on the single Q498R and N501Y mutants predict, if their effects were additive, that the double mutant would be stabilizing by −0.6 kcal/mol (−0.8 + 0.2 kcal/mol for the single mutations, respectively) while the experimental value for the double mutant is −1.2 kcal/mol yielding a cooperativity energy of −0.6 kcal/mol (see Figure 3(B) for the experimental ΔΔG values and Figure S1 for corresponding fitted data and dissociation constants). While the experimental SPR and yeast display ΔΔG values differ, both methods indicate that a substantial epistatic effect is playing a role, perhaps contributing to the greater infectivity of the Omicron variants where these mutations are present.

FEP calculations on the double mutant predict a ΔΔG of −1.4 kcal/mol whereas the predicted additive effect of the two single mutants (Table 2) is +1.5 kcal/mol – corresponding to a cooperativity energy of −2.9 kcal/mol (Figure 3(B)). Thus, the 100 ns calculations successfully predict the existence of cooperativity. A detailed examination of the structure of the double mutant provides a compelling physical interpretation as to why it behaves so differently from the single Q498R mutation. In the double mutant, Arg 498 forms a favorable pairing with the side chain of Tyr 501. The aliphatic portion of the Arg side chain packs against the Tyr aromatic ring creating an enhanced hydrophobic contact in the double mutant compared to the N501Y single mutant (grey shading, Figure 3(A)), while one hydrogen of the Arg guanidinium head group forms a hydrogen bond with the oxygen of the Tyr hydroxyl group (dashed lines, Figure 3(A)). The geometries of the two residues are such that this pairing can be carried out without introducing any conformational strain, either in the backbone or in the side chains themselves. Thus, in the presence of N501Y, the mutation from Gln to Arg actually enhances binding rather than diminishing it.

As a control, we considered a second double mutant (L452R T478K) where no cooperativity is observed experimentally (Figure 3(C)). The FEP calculations at 100 ns predict values of both single and double mutant within 0.4 kcal/mol from the experiment and a cooperativity energy of −0.1 kcal/mol (same as SPR) (Figure 3(C)). Thus, FEP accurately predicts the absence of cooperativity in the double mutant belonging to the Delta SARS-CoV-2 variant (Table S2).

Discussion

We have carried out experimental and computational studies of a series of RBD mutants located in the RBD::ACE2 interface. The choice was based in part on their frequency in infective SARS-CoV-2 variants and in part because they have been extensively studied.[20–63] A central goal has been to assess the ability of the free energy perturbation methodology to predict the effect of mutations at the protein–protein interface on binding affinities but this in turn required an assessment of experimental accuracy. To this end we carried out
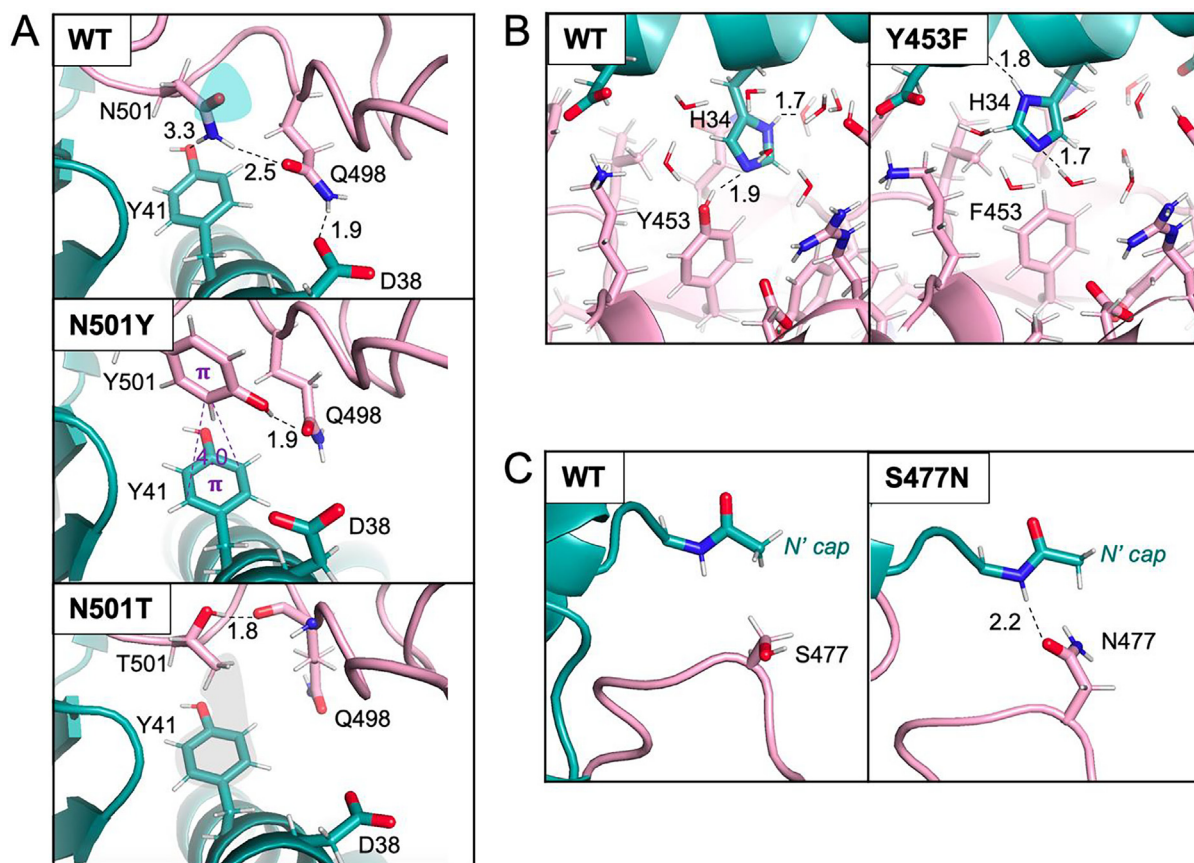
Figure 2. Structural origins of stabilizing ACE2/RBD interactions. Closeups of the ACE2(cyan)/RBD(pink) showing key interactions in wild-type (WT) and mutant proteins. Hydrophobic contacts are in grey, π-π interactions are in purple, hydrogen bonds are shown as black dashed lines, unsatisfied polar group of N501 is in cyan, favorable hydrophobic contact between T501 and Y41 in grey, and all distances are in Å.

SPR experiments on each mutant and compared our findings to those previously reported. We also constructed a balanced dataset of mutations with sufficient energy spread and compared performance of easily accessible computational methods to those obtained from FEP.

As shown in Table 1 and discussed above, there is excellent agreement between our SPR results and previous yeast display results from the Bloom group.[20] Moreover, both sets of results are in the range reported in other studies. In particular, all methods agree on the identification of stabilizing mutations although the experimental values vary by as much as 1 kcal/mol, likely due to some of the issues discussed in Methods. As pointed out above, FEP yields the best correlation with our SPR results and, crucially, is the most effective in identifying stabilizing mutations (Table 2, Table S3).

The mutations we used in our study are not a part of common ML training databases such as SKEMPI[77]; hence, our dataset for methods evaluation can be considered as "blind". A modest level of performance for ML methods (Table 2, Table S3) is in line with a study by Bonvin and co-workers where various ML methods were tested on a blind dataset of 487 mutations in 56 complexes.[78] FoldX and Rosetta flex ddG categorize neutral and destabilizing mutations well and are at least partially successful in identifying stabilizing mutations (Table S3), though their overall performance (PCC of 0.4–0.5) is inferior to FEP (PCC of 0.8).

In addition to the methods we were able to test on our data set, many other studies on RBD mutants have been reported. Table S5 lists published results on RBD::ACE2 on different sets of mutants than studied here but it is of interest to compare cases where there is overlap with previous work. Previous FEP studies[53,55,56] also carried out long simulations and all identify N501Y and N501T as stabilizing. TopNetTree,[49] a topology-based ML method, does not perform well on our data set. The MM/GBSA study from the Bahar group[58] successfully predicts that N501Y is stabilizing but was not run on other mutants in our data set. We could not reproduce this result in our MM/GB-SA calculations likely due to protocol differences and use of different crystal structures as input. Finally, Maranas and co-workers[47] used physical interactions extracted from MM/GBSA simulations to train a neural network to predict ΔΔG values (NN_MM-
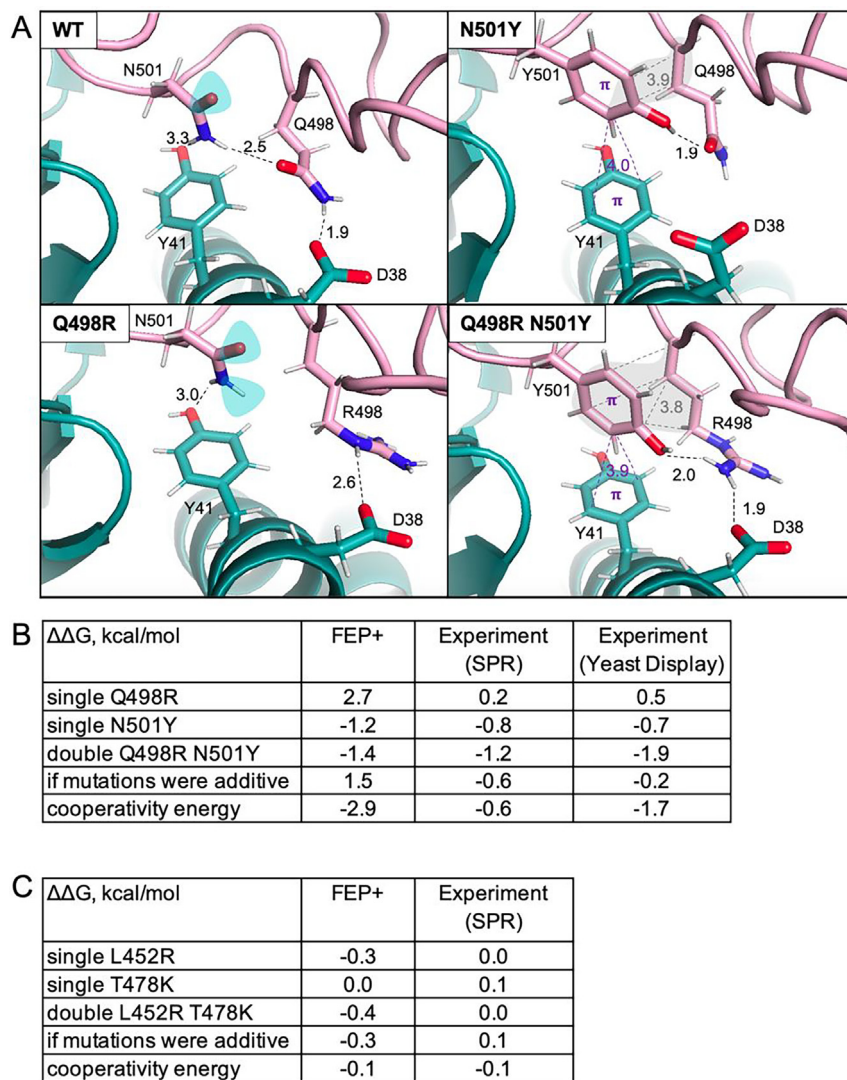
Figure 3. Epistatic effect of the double Q498R N501Y mutant. (A) A closeup of the complex between ACE2 (in cyan) and RBD (in pink) showing chemical interactions involving RBD residues 501 and 498. Hydrophobic contacts are in grey, π-π interactions are in purple, hydrogen bonds are shown as black dashed lines, unsatisfied polar groups are in cyan, and all distances are measured in Å. (B) Cooperativity of the Q498R N501Y double mutant probed by FEP and SPR computed as a difference between a hypothetical ΔΔG (if the double mutant had an additive effect of two single point mutations) and the actual ΔΔG of the double mutant. Yeast display assay values are from Zahradnik et al.[64] (C) Absence of cooperativity probed by FEP and SPR for the L452R T478K double mutant. Experimental result for the Delta RBD variant is from Liu et al.[92] (see Methods for details), ΔΔG of single mutants are from the current study (Table 2). The FEP+ results are from 100 ns trajectories.

GBSA method). Since many of the mutants in our set were used in the NN_MM-GBSA training (Table S5) it is difficult to make comparisons with our own results.

Despite the success of the FEP approach revealed in this work, significant challenges remain. In previous work on a variety of systems, we have demonstrated that the FEP results correlate well with experiment (PCC values on the order of 0.6–0.8) and display RMS errors in the range expected for the OPLS4[79] and related molecular mechanics force fields (0.5–1 kcal/-mol).[8,9,14,66–68] Precision beyond the above cited statistics is very difficult to obtain, indeed, experimental reproducibility errors are typically on the order of 0.4 kcal/mol, even for the high quality SPR results that we report here (see Methods).

As highlighted in this work, the major challenge in using FEP to predict mutation effects on protein–protein binding affinities, as opposed to small molecule binding, is the possibility of significant conformational changes induced by a mutation, for example if a buried charge is created by mutating a buried hydrophobic residue at the

interface to one with a net charge. The key issue is whether the conformational changes required to make accurate predictions are accessible on the timescale of the FEP simulation. When the conformational changes consist primarily of side chain rearrangements and relatively minor backbone motion, FEP will typically deliver reliable results (as in N501Y). When there are significant backbone conformational changes, the accuracy of the FEP results will depend upon whether the barrier to conformational change can be surmounted on the time scale of the simulation.

In terms of overall accuracy, the results obtained here are consistent, in terms of both RMSD errors and correlation coefficients, with those we have reported previously in studying HIV derived gp120-antibody binding and also more diverse sets of protein–protein complexes via FEP simulations[8,9]. The most significant errors may result from structural uncertainties (as in A475V) or from the prediction of overly unfavorable free energy changes (as in Q498R) which often result from electrostatic or steric clashes or from the burial of a charge in a hydrophobic pocket.[9] However, mutations that are predicted to be highly destabilizing would be of little or no consequence in the context of binding optimization project since such mutations would be rejected in an initial screen even if converged results were obtained.

The accuracy of binding free energy differences determined by FEP or other structure-based methods is crucially dependent on the quality and completeness of the complex structures used in the simulations. While FEP performs well with experimentally determined structures (as shown in Table S3), achieving comparable success with modeled complexes obtained by docking or homology models of poor quality is much harder. Our study reveals that using experimental structures helps with general convergence, as multiple independent FEP runs result in SEM < 0.1 kcal/mol (as shown in Table S4). However, when FEP simulations were performed on modeled protein-peptide complexes, a much larger deviation (SEM $\sim$ 1.5 kcal/mol) was observed.[79]

Although FEP is a successful method for predicting the effect of point mutations on binding free energy differences, it is not suitable for direct evaluation of deletions or insertions effects using alchemical transformations. A combination of MD and advanced sampling techniques to guarantee the adequate exploration of each relevant degree of freedom can be applied to obtain binding free energies within experimental accuracy in the process of progressive separation of proteins via the potential of mean force (PMF).[80,81] The PMF approaches are not limited to single point mutations, but their success requires sufficiently accurate approximation of the native wild-type and mutant states.

Our results suggest FEP might be used in a strategy to optimize the binding of two proteins, for example in the optimization of antibodies. FEP calculations appear to have sufficient rank ordering capability to enable prioritization of specific single residue mutations. The initial step in a design strategy would be an exhaustive screening of single mutations at the various positions across the protein–protein interface (perhaps a few hundred to a few thousand calculations, which would require relatively modest expenditure, given the steadily decreasing cost of GPU-based computation). A key advantage of FEP is that analysis of trajectories can reveal insights as to the response of the interface to different perturbations, enhanced when needed by loop modeling procedures that more efficiently sample conformational space. These insights could then be translated into the investigation of a selected subset of double and triple mutations, with the goal of achieving favorable nonadditive effects. The understanding we gained about the Q498R N501Y double mutant could not have been accomplished by any of the other methods tested here.

It is important to clarify that, despite identifying cooperativity of the N501Y Q498R double mutant, the FEP results have not reached the point where the calculated values have experimental accuracy. It is then interesting to consider how a researcher interested in designing a stabilizing mutation would respond to calculated single mutant values of −1.2 kcal/mol for N501Y, +2.7 kcal/mol for Q498R and −1.4 kcal/mol for the double mutant. We would argue that the large calculated cooperativity energy of −2.9 kcal/mol, and the physical basis of this effect revealed by the simulations, would provide a strong hint that the double mutant is worth testing experimentally. Moreover, the experience we gained in this study would make that decision more likely. In conclusion, a careful exploration of a particular system of biological importance as enabled by FEP simulations would then appear to offer a way forward in many practical applications.

Materials and Methods

ACE2::RBD dataset

We focused on missense RBD mutations at the interface with ACE2 that occurred most frequently in the US at the beginning of 2021 or were a part of known variants of concern. Among single point mutations with a frequency above 100 as of Jan 4, 2021, only seven mutations (S477N, N439K, N501Y, L452R, Y453F, S477R and S477I) were both missense and interfacial. To expand the dataset, we added missense interfacial mutations with a lower frequency (in the range of 10–100) if mutations were stabilizing or destabilizing in the

study of Bloom and co-workers[20] while nearly neutral positions were ignored. Stabilizing mutations were of interest as potentially increasing infectivity of the virus, while destabilizing mutations were a necessary addition to create a balanced dataset for a proper testing of ΔΔG predictors. The K417T and Q498R mutations were added due to occurrence in the SARS-CoV-2 variants of concern and not based on frequency counts. The K417T mutation was added due to its emergence in Brazil (as a part of the Gamma variant) at the time. The Q498R mutant was studied in the context of double mutant effects alongside N501Y due to the co-occurrence of these two mutations in Omicron variant (Table S2) and a recent study on *in vitro* evolution suggesting cooperativity between the two mutations.[64] Although the cumulative frequency (i.e. a number of times a given mutation has been found in the sequenced SARS-CoV-2 genomes) has changed during 2021, the majority of the most frequently observed mutations in January 2021 were still among the most recurrent mutations at the end of the year, with our data set including 16 out of 20 interfacial missense RBD mutations that had the highest frequency (>1000) on December 31, 2021 (Table S2).

Frequencies of SARS-CoV-2 mutations were obtained from the Mutation Tracker resource (https://users.math.msu.edu/users/weig/SARS-CoV-2_Mutation_Tracker.html)[82] that relies on data from the GISAID database of coronavirus genomes (https://www.gisaid.org/).

DIP::Dpr dataset

We included single point mutations of DIP-α, Dpr6 or Dpr10 that have been previously tested by SPR.[71–73]

ΔΔG calculations

All calculations presented in Table 2 and Table S3 were performed using crystal structures of highest available resolution: the SARS-CoV-2 RBD complex with ACE2 (2.45 Å, PDBID: 6M0J),[69] the DIP-α::Dpr6 complex (2.3 Å, PDBID: 5EO9)[70] or the DIP-α::Dpr10 complex (1.8 Å, PDBID: 6NRQ).[83]

Mutabind2 predictions were run on the https://lilab.jysw.suda.edu.cn/research/mutabind2/ webserver. mCSM-PPI2 predictions were submitted as a query to the following webserver: https://biosig.unimelb.edu.au/mcsm_ppi2/.

Standalone version (https://compbio.clemson.edu/saambe_webserver/standaloneCode.zip) was used for binding affinity calculations of SAAMBE-3D. FoldX calculations were performed as described in Sergeeva et al.[72] Rosetta flex ddG calculations were run using a standalone version (https://github.com/Kortemme-Lab/flex_ddG_tutorial) with the following parameters (considered to give optimized performance of this method[7]: nstruct = 35, max_minimization_iter = 5000, abs_s

core_convergence_thresh = 1.0, number_back rub_trials = 35000, backrub_trajectory_stride = 35000.

The MM/GB-SA and MM/PB-SA methods rely on the end-point approximation of the wild-type (WT) and mutant (MT) states to calculate binding free energies using the following formula: $\Delta\Delta G = \Delta G_{binding}(MT) - \Delta G_{binding}(WT)$, where $\Delta G_{binding}(P_1::P_2) = G(P_1::P_2) - G(P_1) - G(P_2)$. To generate the mutated proteins, we used the "Mutate Residue" function followed by a local minimization of the mutated residue in Maestro, using the crystal structure of the WT as input. We employed the AMBER software package and pmemd.cuda[84,85] for minimization. The $P_1::P_2$ complex is minimized in explicit OPC water, followed by removal of water molecules and energy calculations in implicit solvent. The isolated P_1 and P_2 proteins are assumed to have the same conformation as in the complex. The energy calculations were carried out using MMPBSA.py.[86] The mbondi2 radii set was used in both GB and PB calculations, and the OBC2 model was used for GB calculations.

We used Schrödinger software (2021-2 release) and default FEP+ protocols (implementing guidelines for protein–protein interactions published earlier[8,9] for our predictions of binding affinity changes in the ACE2/RBD complex upon RBD mutations. The release incorporates a recently developed OPLS4 force field,[79] replica exchange with solute tempering (REST) enhanced sampling methodology for mutated residues, and an improved grand canonical Monte Carlo (GCMC) protocol for sampling solvent molecules around mutated residues.[87] FEP+ is a fully physics-based model that uses explicitly represented water. During FEP+ simulations, an alchemical transformation of a wild type amino acid residue into a mutant residue is conducted, which is implemented by running a series of separate molecular dynamics simulations ("lambda windows") with varied energy weighting. The differences between adjacent lambda windows are first calculated using a perturbative expansion and then summed up to estimate the total free energy change between a wild-type and mutant states. To enhance the convergence of the free energy calculations, our default protocols use 12 lambda windows for charge conserving mutations and 24 lambda windows for the charge changing mutations. All mutations were run for 10 ns and 100 ns (see Table S3). Calculations of double mutant effects were performed by simultaneous alchemical transformation of the two mutated residues. This procedure minimizes errors associated with a more common FEP protocol where mutations are introduced sequentially. For example, the ΔΔG (Q498R N501Y) is predicted to be −1.4 kcal/mol using the simultaneous protocol and −0.6 kcal/mol using the sequential protocol. Of note, the results from the simultaneous protocol are in better agreement with the experimental value of −1.2 kcal/mol.

The FEP+ methodology automates generation of the mutated end-point, while the user provides a prepared wild-type complex as input. This process involves sampling and ranking the rotamers of the mutated residue, followed by minimization of the mutated residue. Additionally, the protein–protein complex is equilibrated before the FEP/REST2 production run. This protocol is typically effective in accurately identifying the mutated end-point throughout the FEP/REST2 trajectory, except in cases where a significant conformational change is required upon mutation.

FEP+ requires GPU computing and the time required per single point mutation depends on the length of simulation (in nanoseconds), system size (in atoms), and type of mutation (with charge-changing mutations taking longer to compute compared to charge-neutral mutations as the number of lambda windows is twice as large). In the ACE2/RBD system ($\sim$100,000 atoms including explicit waters in the solvent box), the shortest 10 ns charge-neutral mutations take less than a day, while the longest 100 ns charge-changing simulations take $\sim$2 weeks when complex and solvent legs are run in parallel with each simulation leg using 4 GPUs.

Protein expression and purification

The SARS-CoV-2 RBD wild-type and its mutants (residues 331–528), were cloned into the pVRC-8400 mammalian expression plasmid, with a C-terminal 6XHis-tag and an intervening HRV-3C protease cleavage site. Plasmid constructs were transfected into HEK293 cells using polyethyleneimine (Polysciences). Cell growths were harvested four days after transfection, and the secreted proteins were purified from supernatant by nickel affinity chromatography using Ni-NTA IMAC Sepharose 6 Fast Flow resin (Cytiva) followed by size exclusion chromatography on a Superdex 200 column (Cytiva) in 10 mM Tris, 150 mM NaCl, pH 7.4.

A plasmid encoding ACE2 residues 1–615 (pcDNA3-sACE2(WT)-8his) was a gift from Erik Procko (Addgene plasmid #149,268; https://n2t.net/addgene:149268; RRID:Addgene_149268).[88] This plasmid was then mutated to encode ACE2 residues 1–620, followed by a C terminal HRV-3C protease cleavage site, and an 8X HIS tag. This construct was transfected into ExpiHEK293 cells using Expifectamine according to manufacturer's instructions. Supernatants were harvested seven days after transfection, and ACE2 was purified by nickel affinity chromatography using His60 Ni Superflow Resin (Takara) followed by size exclusion chromatography on a Superdex 200 column (Cytiva) in 10 mM Tris, 150 mM NaCl, pH 8.0. 400 ug of purified ACE2 was then digested overnight at 4 degrees with 20 units of HRV-3C protease (Thermo Scientific). Digested ACE2 was then incubated with His60 Ni Superflow Resin, which was then washed with 2 column volumes of 10 mM Tris, 150 mM NaCl, 5 mM imidazole, pH 8.0. The flow through and wash were determined by SDS-PAGE to contain cleaved ACE2, which was purified by size exclusion chromatography on a Superdex 200 column (Cytiva) in 10 mM Tris, 150 mM NaCl, pH 8.0.

Surface plasmon resonance

SPR binding assays for monomeric ACE2 binding to RBDs were performed using a Biacore T200 biosensor, equipped with a Series S CM5 chip, at 25 °C, in a running buffer of 10 mM HEPES pH 7.4, 150 mM NaCl, 0.1 mg/mL BSA and 0.01% (v/v) Tween-20 at 25 °C. Each RBD was captured through its C-terminal his-tag over an anti-his antibody surface, generated using the His-capture kit (Cytiva, MA) according to the instructions of the manufacturer.

During a binding cycle, each RBD was captured over individual flow cells at approximately 250 RU. An anti-his antibody surface was used as a reference flow cell to remove bulk shift changes from the binding signals. Monomeric ACE2 was prepared at six concentrations in running buffer using a three-fold dilution series, ranging from 1.1 to 270 nM. Samples were tested in order of increasing protein concentration, with each series tested in triplicate. Blank buffer cycles were performed by injecting running buffer instead of the analyte, after two ACE2 injections to remove systematic noise from the binding signal. The association and dissociation rates were each monitored for 180 s and 300 s respectively, at 50 μL/min. Bound RBD/Fab complexes were removed using a 10 s pulse of 15 mM H_3PO_4 at 100 μL/min, thus regenerating the anti-his surface for a new cycle of recapturing of each RBD, followed by a 60 s buffer wash at 100 μL/min. The data was processed and fit to 1:1 interaction model using the Scrubber 2.0 (BioLogic Software). The number in brackets for each kinetic parameter represents the error of the fit.

We have developed an SPR assay to measure the binding kinetics and affinities of interactions between ACE2 and wild-type or mutant RBD with the RBD captured to the chip to avoid compromised binding activity resulting from chemical immobilization or repeated surface harsh regeneration steps during the experiments. The SARS-CoV-2 RBD is a basic molecule with a pI $\sim$ 9, so capture to the chip surface will also minimize artifacts such as non-specific interactions between the positively charged RBDs and the negatively charged dextran layer of the sensor chips at physiological pH. Monomeric ACE2 was flowed over as analyte to avoid avidity effects resulting from using dimeric ACE2. Studies that have performed such experiments in both orientations (RBD tethered to the surface vs in solution as an analyte) showed that using RBD as

an analyte yielded affinities that were approximately three-fold stronger for the wild type RBD/ACE2 interaction, suggesting the presence of such non-specific interactions.[24] In our SPR experiments we have determined the K_D for wild type RBD binding to monomeric ACE2 is 162.9 nM (Figure S1), consistent with similar K_Ds reported from other groups that have used similar methodologies to perform their biosensor-based measurements.[21,89,90] Figure S1(A) shows the binding kinetics for interactions of mutant RBD proteins with ACE2 and Figure S1 (B) shows the kinetic parameters along with affinities calculated for each binding interaction, while Table 1 and Figure 3(B) list experimental changes in binding affinities ($\Delta\Delta G = RT\ln(K_{D(MT)}/K_{D(WT)})$) when RBD is mutated. Experimental reproducibility errors in our SPR data (association and dissociation rate constants) is expected to be ~15–20% according to previous estimates based on multiple independent measurements[72]; a 15–20% error corresponds to ~0.4 kcal/mol experimental error in the $\Delta\Delta G$(SPR) values reported in this study. Of 23 single point RBD mutations probed, four mutations were identified as stabilizing: N501Y, Y453F, S477N and N501T (Table 1).

Differences in experimental protocol affecting binding affinity changes

Previously reported experimental binding affinity changes upon RBD mutation of the ACE2/RBD complex span different choice of protein constructs and orientation of molecules used in the binding assays.

The differences in the constructs lie in the choice of monomeric vs. multimeric forms of interacting proteins and selection of protein domain boundaries. Some studies relied on monomeric ACE2 and RBD[21,22,24,25,27–30], whereas others used at least one of the molecule in a multimeric form (trimeric spike, dimeric ACE2 or monomeric ACE2 fused to a dimeric Fc tag).[20,25,26,31,32,64] RBD domain starts at residue 333 and ends at residue 526. It is common that constructs used in studies flank the RBD construct with a few residues before and after the domain boundary (e.g. 331–528 (this study), 331–531,[20] 328–531[21,22]) though some studies use constructs where such flanking regions are too long so they could result in non-specific binding (especially when containing unpaired cysteine residues, e.g. RBD 319–591.[25] Poor selection of protein domain boundaries can affect protein folding/integrity when a construct has incomplete domain sequence (e.g. RBD 343–532 construct is missing a β-strand).[23]

Orientation of molecules in the binding experiments (which molecule is tethered to the chip in SPR) can affect both absolute and relative binding affinities (discussed in SPR methods). Studies using a highly positively charged RBD

molecule as analyte and ACE2 immobilized on a chip[27–32] can be affected by non-specific binding of RBD to the chip. For example, using RBD as analyte in experiments measuring binding affinity of Alpha, Beta, Gamma, and Delta variants[91] results in stronger binding (by 0.4–0.6 kcal/mol) compared to a setup minimizing non-specific binding by immobilizing RBD on a chip.[92] We used the latter as a reference to assess performance of $\Delta\Delta G$ on predicting double mutant effects of the Delta variant (Figure 3 (C)).

CRediT authorship contribution statement

Alina P. Sergeeva: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data curation, Software, Visualization, Writing – original draft, Writing – review & editing. **Phinikoula S. Katsamba:** Validation, Formal analysis, Investigation, Visualization. **Junzhuo Liao:** Formal analysis, Investigation, Data curation, Software. **Jared M. Sampson:** Resources. **Fabiana Bahna:** Resources. **Seetha Mannepalli:** Resources. **Nicholas C. Morano:** Resources. **Lawrence Shapiro:** Conceptualization, Supervision, Funding acquisition. **Richard A. Friesner:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing. **Barry Honig:** Conceptualization, Supervision, Funding acquisition, Writing – review & editing.

DATA AVAILABILITY

Data will be made available on request.

DECLARATION OF COMPETING INTEREST

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: 'RAF is a founder of Schödinger and a consultant; JMS is an employee and BH is a consultant. The FEP calculations are based on Schrödinger-distributed software.'.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.jmb.2023.168187.

References

1. Pahari, S., Li, G., Murthy, A.K., Liang, S., Fragoza, R., Yu, H., et al., (2020). SAAMBE-3D: predicting effect of mutations on protein-protein interactions. *Int J Mol Sci* **21**, 2563.

2. Pires, D.E., Ascher, D.B., Blundell, T.L., (2014). mCSM: predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* **30**, 335–342.

3. Rodrigues, C.H.M., Myung, Y., Pires, D.E.V., Ascher, D.B., (2019). mCSM-PPI2: predicting the effects of mutations on protein-protein interactions. *Nucleic Acids Res* **47**, W338–W344.

4. Li, M., Simonetti, F.L., Goncearenco, A., Panchenko, A.R., (2016). MutaBind estimates and interprets the effects of sequence variants on protein-protein interactions. *Nucleic Acids Res* **44**, W494–W501.

5. Zhang, N., Chen, Y., Lu, H., Zhao, F., Alvarez, R.V., Goncearenco, A., et al., (2020). MutaBind2: predicting the impacts of single and multiple mutations on protein-protein interactions. *iScience* **23**, 100939

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

7. Barlow, K.A., S, O.C., Thompson, S., Suresh, P., Lucas, J. E., Heinonen, M., et al., (2018). Flex ddG: rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. *J Phys Chem B* **122**, 5389–5399.

8. Clark, A.J., Gindin, T., Zhang, B., Wang, L., Abel, R., Murret, C.S., et al., (2017). Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1. *J Mol Biol* **429**, 930–947.

9. Clark, A.J., Negron, C., Hauser, K., Sun, M., Wang, L., Abel, R., et al., (2019). Relative binding affinity prediction of charge-changing sequence mutations with FEP in protein-protein interfaces. *J Mol Biol* **431**, 1481–1493.

10. Guerois, R., Nielsen, J.E., Serrano, L., (2002). Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320**, 369–387.

11. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L., (2005). The FoldX web server: an online force field. *Nucleic Acids Res* **33**, W382–W388.

12. Geng, C., Xue, L.C., Roel-Touris, J., Bonvin, A.M.J.J., (2019). Finding the ΔΔG spot: Are predictors of binding affinity changes upon mutations in protein–protein interactions ready for it? *WIREs Comput Mol Sci* **9**, e1410.

13. Geng, C., Vangone, A., Bonvin, A.M.J.J., (2016). Exploring the interplay between experimental methods and the performance of predictors of binding affinity change upon mutations in protein complexes. *Protein Eng Design Select* **29**, 291–299.

14. Abel, R., Wang, L., Harder, E.D., Berne, B.J., Friesner, R. A., (2017). Advancing drug discovery through enhanced free energy calculations. *Acc Chem Res* **50**, 1625–1632.

15. Katz, D., DiMattia, M.A., Sindhikara, D., Li, H., Abraham, N., Leffler, A.E., (2021). Potency- and selectivity-enhancing mutations of conotoxins for nicotinic acetylcholine receptors can be predicted using accurate free-energy calculations. *Mar Drugs* **19**

16. Katz, D., Sindhikara, D., DiMattia, M., Leffler, A.E., (2021). Potency-enhancing mutations of gating modifier toxins for the voltage-gated sodium channel NaV1.7 can be predicted using accurate free-energy calculations. *Toxins (Basel)*, 13.

17. La Serra, M.A., Vidossich, P., Acquistapace, I., Ganesan, A.K., De Vivo, M., (2022). Alchemical free energy calculations to investigate protein-protein interactions: the case of the CDC42/PAK1 complex. *J Chem Inf Model* **62**, 3023–3033.

18. Nandigrami, P., Szczepaniak, F., Boughter, C.T., Dehez, F., Chipot, C., Roux, B., (2022). Computational assessment of protein-protein binding specificity within a family of synaptic surface receptors. *J Phys Chem B* **126**, 7510–7527.

19. Patel, D., Patel, J.S., Ytreberg, F.M., (2021). Implementing and assessing an alchemical method for calculating protein-protein binding free energy. *J Chem Theory Comput* **17**, 2457–2464.

20. Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., et al., (2020). Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20.

21. McCallum, M., Bassi, J., De Marco, A., Chen, A., Walls, A. C., Di Iulio, J., et al., (2021). SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. *Science* **373**, 648–654.

22. McCallum, M., Walls, A.C., Sprouse, K.R., Bowen, J.E., Rosen, L.E., Dang, H.V., et al., (2021). Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science* **374**, 1621–1626.

23. Deshpande, A., Harris, B.D., Martinez-Sobrido, L., Kobie, J.J., Walter, M.R., (2021). Epitope classification and RBD binding properties of neutralizing antibodies against SARS-CoV-2 variants of concern. *Front Immunol* **12**, 691715

24. Thomson, E.C., Rosen, L.E., Shepherd, J.G., Spreafico, R., da Silva, F.A., Wojcechowskyj, J.A., et al., (2021). Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* **184**, 1171–1187.e20.

25. Tian, F., Tong, B., Sun, L., Shi, S., Zheng, B., Wang, Z., et al., (2021). N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *Elife*, 10.

26. Gobeil, S.M., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Stalls, V., et al., (2021). Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* **373**

27. Laffeber, C., de Koning, K., Kanaar, R., Lebbink, J.H.G., (2021). Experimental evidence for enhanced receptor binding by rapidly spreading SARS-CoV-2 variants. *J Mol Biol* **433**, 167058

28. Liu, H., Wei, P., Zhang, Q., Chen, Z., Aviszus, K., Downing, W., et al., (2021). 501Y.V2 and 501Y.V3 variants of SARS-CoV-2 lose binding to bamlanivimab in vitro. *MAbs* **13**, 1919285.

29. Liu, H., Zhang, Q., Wei, P., Chen, Z., Aviszus, K., Yang, J., et al., (2021). The basis of a more contagious 501Y.V1 variant of SARS-CoV-2. *Cell Res* **31**, 720–722.

30. Barton, M.I., MacGowan, S.A., Kutuzov, M.A., Dushek, O., Barton, G.J., van der Merwe, P.A., (2021). Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *Elife*, 10.

31. Raghu, D., Hamill, P., Banaji, A., McLaren, A., Hsu, Y.T., (2021). Assessment of the binding interactions of SARS-CoV-2 spike glycoprotein variants. *J Pharm Anal*.

32. Han, P., Su, C., Zhang, Y., Bai, C., Zheng, A., Qiao, C., et al., (2021). Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nat Commun* **12**, 6103.

33. Xue, T., Wu, W., Guo, N., Wu, C., Huang, J., Lai, L., et al., (2021). Single point mutations can potentially enhance infectivity of SARS-CoV-2 revealed by in silico affinity maturation and SPR assay. *RSC Adv* **11**, 14737–14745.

34. Mannar, D., Saville, J.W., Zhu, X., Srivastava, S.S., Berezuk, A.M., Zhou, S., et al., (2021). Structural analysis of receptor binding domain mutations in SARS-CoV-2 variants of concern that modulate ACE2 and antibody binding. *Cell Rep* **37**, 110156

35. Saville, J.W., Mannar, D., Zhu, X., Srivastava, S.S., Berezuk, A.M., Demers, J.P., et al., (2022). Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. *Nat Commun* **13**, 742.

36. Zhu, X., Mannar, D., Srivastava, S.S., Berezuk, A.M., Demers, J.P., Saville, J.W., et al., (2021). Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. *PLoS Biol* **19**, e3001237.

37. Cai, Y., Zhang, J., Xiao, T., Lavine, C.L., Rawson, S., Peng, H., et al., (2021). Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* **373**, 642–648.

38. Zhang, J., Xiao, T., Cai, Y., Lavine, C.L., Peng, H., Zhu, H., et al., (2021). Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant. *Science* **374**, 1353–1360.

39. Yang, T.-J., Yu, P.-Y., Chang, Y.-C., Chang, N.-E., Tsai, Y.-X., Liang, K.-H., et al., (2021). Structure-activity relationships of B.1.617 and other SARS-CoV-2 spike variants. *bioRxiv* 2021.09.12.459978.

40. Bayarri-Olmos, R., Jarlhelt, I., Johnsen, L.B., Hansen, C. B., Helgstrand, C., Rose Bjelke, J., et al., (2021). Functional effects of receptor-binding domain mutations of SARS-CoV-2 B.1.351 and P.1 variants. *Front Immunol* **12**, 757197

41. Bayarri-Olmos, R., Johnsen, L.B., Idorn, M., Reinert, L.S., Rosbjerg, A., Vang, S., et al., (2021). The alpha/B.1.1.7 SARS-CoV-2 variant exhibits significantly higher affinity for ACE-2 and requires lower inoculation doses to cause disease in K18-hACE2 mice. *Elife*, 10.

42. Bayarri-Olmos, R., Rosbjerg, A., Johnsen, L.B., Helgstrand, C., Bak-Thomsen, T., Garred, P., et al., (2021). The SARS-CoV-2 Y453F mink variant displays a pronounced increase in ACE-2 affinity but does not challenge antibody neutralization. *J Biol Chem* **296**, 100536

43. Supasa, P., Zhou, D., Dejnirattisai, W., Liu, C., Mentzer, A. J., Ginn, H.M., et al., (2021). Reduced neutralization of SARS-CoV-2 B.1.1.7 variant by convalescent and vaccine sera. *Cell* **184** (2201–11), e7.

44. Yuan, M., Huang, D., Lee, C.D., Wu, N.C., Jackson, A.M., Zhu, X., et al., (2021). Structural and functional ramifications of antigenic drift in recent SARS-CoV-2 variants. *Science* **373**, 818–823.

45. Lopez, E., Haycroft, E.R., Adair, A., Mordant, F.L., O'Neill, M.T., Pymm, P., et al., (2021). Simultaneous evaluation of antibodies that inhibit SARS-CoV-2 variants via multiplex assay. *JCI Insight*, 6.

46. Collier, D.A., De Marco, A., Ferreira, I., Meng, B., Datir, R. P., Walls, A.C., et al., (2021). Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141.

47. Chen, C., Boorla, V.S., Banerjee, D., Chowdhury, R., Cavener, V.S., Nissly, R.H., et al., (2021). Computational prediction of the effect of amino acid changes on the binding affinity between SARS-CoV-2 spike RBD and human ACE2. *Proc Natl Acad Sci U S A* **118**

48. Liu, X., Luo, Y., Li, P., Song, S., Peng, J., (2021). Deep geometric representations for modeling effects of mutations on protein-protein binding affinity. *PLoS Comput Biol* **17**, e1009284.

49. Chen, J., Wang, R., Wang, M., Wei, G.W., (2020). Mutations Strengthened SARS-CoV-2 Infectivity. *J Mol Biol* **432**, 5212–5226.

50. Strokach, A., Lu, T.Y., Kim, P.M., (2021). ELASPIC2 (EL2): Combining contextualized language models and graph neural networks to predict effects of mutations. *J Mol Biol* **433**, 166810

51. Laurini, E., Marson, D., Aulic, S., Fermeglia, A., Pricl, S., (2021). Computational mutagenesis at the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 binding interface: comparison with experimental evidence. *ACS Nano* **15**, 6929–6948.

52. Laurini, E., Marson, D., Aulic, S., Fermeglia, M., Pricl, S., (2020). Computational alanine scanning and structural analysis of the SARS-CoV-2 spike protein/angiotensin-converting enzyme 2 complex. *ACS Nano* **14**, 11821–11830.

53. Fratev, F., (2021). N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human-derived antibody: a free energy of perturbation retrospective study. *J Chem Inf Model* **61**, 6079–6084.

54. Fratev, F., (2021). The High transmission of SARS-CoV-2 Omicron (B.1.1.529) variant is not only due to its hACE2 binding: a free energy of perturbation study. *bioRxiv* 2021.12.04.471246.

55. Pavlova, A., Zhang, Z., Acharya, A., Lynch, D.L., Pang, Y. T., Mou, Z., et al., (2021). Machine learning reveals the critical interactions for SARS-CoV-2 spike protein binding to ACE2. *J Phys Chem Lett* **12**, 5494–5502.

56. Luan, B., Wang, H., Huynh, T., (2021). Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the

human ACE2 receptor: insights from molecular dynamics simulations. *FEBS Lett* **595**, 1454–1461.

57. Khan, A., Zia, T., Suleman, M., Khan, T., Ali, S.S., Abbasi, A.A., et al., (2021). Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J Cell Physiol* **236**, 7045–7057.

58. Cheng, M.H., Krieger, J.M., Banerjee, A., Xiang, Y., Kaynak, B., Shi, Y., et al., (2022). Impact of new variants on SARS-CoV-2 infectivity and neutralization: A molecular assessment of the alterations in the spike-host protein interactions. *iScience* **25**, 103939

59. Buratto, D., Saxena, A., Ji, Q., Yang, G., Pantano, S., Zonta, F., (2021). Rapid assessment of binding affinity of SARS-COV-2 spike protein to the human angiotensin-converting enzyme 2 receptor and to neutralizing biomolecules based on computer simulations. *Front Immunol* **12**, 730099

60. Villoutreix, B.O., Calvez, V., Marcelin, A.G., Khatib, A.M., (2021). In silico investigation of the new UK (B.1.1.7) and South African (501Y.V2) SARS-CoV-2 variants with a focus at the ACE2-spike RBD interface. *Int J Mol Sci*, 22.

61. Wu, L., Zhou, L., Mo, M., Liu, T., Wu, C., Gong, C., et al., (2022). SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Signal Transduct Target Ther* **7**, 8.

62. Gan, H.H., Twaddle, A., Marchand, B., Gunsalus, K.C., (2021). Structural modeling of the SARS-CoV-2 spike/ human ACE2 complex interface can identify high-affinity variants associated with increased transmissibility. *J Mol Biol* **433**, 167051

63. Bozdaganyan, M.E., Shaitan, K.V., Kirpichnikov, M.P., Sokolova, O.S., Orekhov, P.S., (2022). Computational analysis of mutations in the receptor-binding domain of SARS-CoV-2 spike and their effects on antibody binding. *Viruses* **14**

64. Zahradnik, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., et al., (2021). SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nat Microbiol* **6**, 1188–1198.

65. Liu, H., Wei, P., Zhang, Q., Aviszus, K., Linderberger, J., Yang, J., et al., (2021). The Lambda variant of SARS-CoV-2 has a better chance than the Delta variant to escape vaccines. *bioRxiv*.

66. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J.Y., et al., (2016). OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J Chem Theory Comput* **12**, 281–296.

67. Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J.M., Lu, C., et al., (2019). OPLS3e: Extending force field coverage for drug-like small molecules. *J Chem Theory Comput* **15**, 1863–1874.

68. Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., et al., (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* **137**, 2695–2703.

69. Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., et al., (2020). Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220.

70. Carrillo, R.A., Ozkan, E., Menon, K.P., Nagarkar-Jaiswal, S., Lee, P.T., Jeon, M., et al., (2015). Control of synaptic connectivity by a network of drosophila IgSF cell surface proteins. *Cell* **163**, 1770–1782.

71. Cosmanescu, F., Katsamba, P.S., Sergeeva, A.P., Ahlsen, G., Patel, S.D., Brewer, J.J., et al., (2018). Neuron-subtype-specific expression, interaction affinities, and specificity determinants of DIP/Dpr cell recognition proteins. *Neuron*.

72. Sergeeva, A.P., Katsamba, P.S., Cosmanescu, F., Brewer, J.J., Ahlsen, G., Mannepalli, S., et al., (2020). DIP/Dpr interactions and the evolutionary design of specificity in protein families. *Nat Commun* **11**, 2125.

73. Xu, S., Sergeeva, A.P., Katsamba, P.S., Mannepalli, S., Bahna, F., Bimela, J., et al., (2022). Affinity requirements for control of synaptic targeting and neuronal cell survival by heterophilic IgSF cell adhesion molecules. *Cell Rep* **39**, 110618

74. Liu, Y., Liu, J., Plante, K.S., Plante, J.A., Xie, X., Zhang, X., et al., (2021). The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature*.

75. Burkholz, S., Pokhrel, S., Kraemer, B.R., Mochly-Rosen, D., Carback 3rd, R.T., Hodge, T., et al., (2021). Paired SARS-CoV-2 spike protein mutations observed during ongoing SARS-CoV-2 viral transfer from humans to minks and back to humans. *Infect Genet Evol* **93**, 104897

76. Munnink, B.B.O., Sikkema, R.S., Nieuwenhuijse, D.F., Molenaar, R.J., Munger, E., Molenkamp, R., et al., (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* **371**, 172–177.

77. Jankauskaite, J., Jimenez-Garcia, B., Dapkunas, J., Fernandez-Recio, J., Moal, I.H., (2018). SKEMPI 2.0: An updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*.

78. Geng, C., Vangone, A., Folkers, G.E., Xue, L.C., Bonvin, A., (2018). iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. *Proteins*.

79. Lu, C., Wu, C., Ghoreishi, D., Chen, W., Wang, L., Damm, W., et al., (2021). OPLS4: Improving force field accuracy on challenging regimes of chemical space. *J Chem Theory Comput* **17**, 4291–4300.

80. Blazhynska, M., Coderc, G., de Lacam, E., Chen, H., Roux, B., Chipot, C., (2022). Hazardous shortcuts in standard binding free energy calculations. *J Phys Chem Lett* **13**, 6250–6258.

81. Fu, H., Chen, H., Blazhynska, M., Coderc, G., de Lacam, E., Szczepaniak, F., Pavlova, A., et al., (2022). Accurate determination of protein:ligand standard binding free energies from molecular dynamics simulations. *Nat Protoc* **17**, 1114–1141.

82. Wang, R., Chen, J., Gao, K., Hozumi, Y., Yin, C., Wei, G. W., (2021). Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants. *Commun Biol* **4**, 228.

83. Cheng, S., Ashley, J., Kurleto, J.D., Lobb-Rabe, M., Park, Y.J., Carrillo, R.A., et al., (2019). Molecular basis of synaptic specificity by immunoglobulin superfamily receptors in Drosophila. *Elife* **8**

84. Gotz, A.W., Williamson, M.J., Xu, D., Poole, D., Le Grand, S., Walker, R.C., (2012). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J Chem Theory Comput* **8**, 1542–1555.

85. Salomon-Ferrer, R., Gotz, A.W., Poole, D., Le Grand, S., Walker, R.C., (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. *J Chem Theory Comput* **9**, 3878–3888.

86. Miller 3rd, B.R., McGee Jr., T.D., Swails, J.M., Homeyer, N., Gohlke, H., Roitberg, A.E., (2012). MMPBSA.py: An efficient program for end-state free energy calculations. *J Chem Theory Comput* **8**, 3314–3321.

87. Ross, G.A., Russell, E., Deng, Y., Lu, C., Harder, E.D., Abel, R., et al., (2020). Enhancing water sampling in free energy calculations with grand canonical Monte Carlo. *J Chem Theory Comput* **16**, 6061–6076.

88. Chan, K.K., Dorosky, D., Sharma, P., Abbasi, S.A., Dye, J. M., Kranz, D.M., et al., (2020). Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261–1265.

89. Meng, B., Abdullahi, A., Ferreira, I., Goonawardane, N., Saito, A., Kimura, I., et al., (2022). Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* **603**, 706–714.

90. Zhou, T., Teng, I.T., Olia, A.S., Cerutti, G., Gorman, J., Nazzari, A., et al., (2020). Structure-based design with tag-based purification and in-process biotinylation enable streamlined development of SARS-CoV-2 spike molecular probes. *Cell Rep* **33**, 108322

91. Lan, J., Chen, P., Liu, W., Ren, W., Zhang, L., Ding, Q., et al., (2022). Structural insights into the binding of SARS-CoV-2, SARS-CoV, and hCoV-NL63 spike receptor-binding domain to horse ACE2. *Structure* **30**, 1432–1442. e4.

92. Li, L., Liao, H., Meng, Y., Li, W., Han, P., Liu, K., et al., (2022). Structural basis of human ACE2 higher binding affinity to currently circulating Omicron SARS-CoV-2 sub-variants BA.2 and BA.1.1. *Cell* **185**, 2952–2960.e10.