


 Cite this: *RSC Adv.*, 2021, **11**, 14737

# Single point mutations can potentially enhance infectivity of SARS-CoV-2 revealed by *in silico* affinity maturation and SPR assay†

 Ting Xue,<sup>‡a</sup> Weikun Wu,<sup>‡b</sup> Ning Guo,<sup>b</sup> Chengyong Wu,<sup>a</sup> Jian Huang,<sup>b</sup> Lipeng Lai,<sup>b</sup> Hong Liu,<sup>a</sup> Yalun Li,<sup>a</sup> Tianyuan Wang<sup>\*b</sup> and Yuxi Wang<sup>ID\*<sup>a</sup></sup>

The RBD (receptor binding domain) of the SARS-CoV-2 virus S (spike) protein mediates viral cell attachment and serves as a promising target for therapeutics development. Mutations on the S-RBD may alter its affinity to the cell receptor and affect the potency of vaccines and antibodies. Here we used an *in silico* approach to predict how mutations on RBD affect its binding affinity to hACE2 (human angiotensin-converting enzyme2). The effect of all single point mutations on the interface was predicted. SPR assay results show that 6 out of 9 selected mutations can strengthen binding affinity. Our prediction has reasonable agreement with the previous deep mutational scan results and recently reported mutants. Our work demonstrated the *in silico* method as a powerful tool to forecast more powerful virus mutants, which will significantly benefit the development of broadly neutralizing vaccine and antibody.

Received 18th January 2021

Accepted 11th April 2021

DOI: 10.1039/d1ra00426c

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## 1. Introduction

Since it first broke out in late 2019, COVID-19 has soon spread out worldwide and has been defined as a world pandemic by WHO. The disease is caused by a novel coronavirus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2),<sup>1</sup> which is a beta-coronavirus closely related to the known SARS-CoV. Many efforts have been distributed in developing the prevention/treatment to the disease,<sup>2</sup> including small molecule drugs,<sup>3</sup> vaccine,<sup>4,5</sup> neutralizing antibodies and<sup>6,7</sup> other engineered proteins. However, none of the methods has been comprehensively tested or publicly applied so far.

The adhesion of the virus to the target cell and the following membrane fusion process are essential steps in virus infection thus those two processes are promising targets for drug development. As is typical for coronavirus, the homo-trimeric spike glycoprotein (S protein, comprising S1 and S2 subunit in each monomer) on the envelope of SARS-CoV-2 is responsible for the cell adhesion process.<sup>8</sup> SARS-CoV-2 uses hACE2 as the receptor for host cell entry, and<sup>8</sup> the dissociation constant  $K_D$  of S protein RBD binding to hACE2 was determined to be in the scale of nM. X-ray crystal structures of RBD was resolved in complex with the

hACE2 receptor,<sup>9</sup> revealing essential interactions on the binding surface.

As cell entry is the very first step of virus infection, blocking the binding of S protein to hACE2 can potentially prevent virus transmission. Many developing therapeutics are based on this concept.<sup>5</sup> Monoclonal neutralizing antibodies separated from convalescent patient showed complete competition with RBD and can reduce virus titers in infected lungs in a mouse model.<sup>10</sup> Recombinant vaccine that comprises residues from the S protein RBD was shown to induce functional antibody response in immunized animals.<sup>6,7</sup> What's more, *de novo* protein inhibitors with  $K_D$  of picomolar-level has been designed to neutralize the virus by targeting S protein RBD.

While therapeutics relying on the RBD binding surface require the surface to be consistent enough, SARS-CoV-2 is an RNA virus that is known to have high mutation rate. According to reported data,<sup>11</sup> 196 missense mutations in the gene encoding the RBD domain have been discovered. Although most mutants on the RBD domains were determined to be less infectious,<sup>12</sup> some variants indeed became more resistant to neutralizing antibodies. We are interested in finding RBD mutants with higher affinity, which may escape future treatments. Thus inspiring us to make precautionary therapeutics in the near future.

<sup>13</sup>*In silico* affinity maturation technology is widely used in protein engineering and antibody discovery. Rosetta Flex ddG method is a ddG estimation method developed within the Rosetta macromolecule modeling suite. The ddG represents the difference in protein–protein interaction strength upon mutation. The protocol generates an ensemble of models to include conformational plasticity around a specific/given mutation site

<sup>a</sup>Targeted Tracer Research and Development Laboratory, Precision Medicine Research Center, Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu 610041, P. R. China. E-mail: yuxiwang@scu.edu.cn.com

<sup>b</sup>XtalPi AI Research Center, 7F, Tower B, Dongsheng Building, No. 8, Zhongguancun East Road, Haidian District, Beijing 100083, P. R. China. E-mail: tianyuan.wang@xtalpi.com

† Electronic supplementary information (ESI) available: ESI 1 and ESI file 2. See DOI: 10.1039/d1ra00426c

‡ Co-first author.



and then calculates the average ddG over the ensemble. This method has been shown to outperform previous methods,<sup>14</sup> and considerable improvement on predicting binding-stabilizing mutations was observed.

Since the Rosetta Flex ddG protocol was published, some researcher had applied the algorithm to help their project.

<sup>20</sup>Sophia *et al.* used this protocol to identify hot spot interactions in the  $\alpha$ IIB and  $\beta$  3 stalks that regulate  $\alpha$ IIB  $\beta$  3 function, and found that the correlation between prediction and the extent of  $\alpha$ IIB  $\beta$  3 activation reached 0.59. The Pearson correlation coefficient matched well with the original benchmark. Moreover,<sup>22</sup> Huy *et al.* used powerful technology (MRBLE-pep) with Rosetta Flex ddG protocol to correctly predict that I3 and I5 mutations decrease affinity of CN-binding peptides.

In spite of the fact that Rosetta Flex ddG protocol was developed to predict the mutations in protein–protein interface, Matteo *et al.* found that combining Rosetta Flex ddG protocol with Molecular Dynamics simulation was able to quantitatively predict changes of ligand binding affinity upon protein mutations.<sup>21</sup> The data is a close match to the experimentally determined  $\Delta\Delta G$  values, with a root-meansquare error of 1.2 kcal mol<sup>-1</sup>.

In this article, Rosetta Flex ddG protocol was used to predict the binding affinity change of point mutations on the RBD binding surface. Candidate mutants with large negative predicted ddG score were selected for further experimental validation. 6 of the 9 recommended mutants showed improved affinity to hACE2 in SPR affinity assay.

## 2. Methods

### 2.1. Rosetta Flex ddG calculation

The S protein RBD-hACE2 complex structure was downloaded from PDB database (PDB ID 6M0J).<sup>9</sup> The structure was relaxed using Rosetta FastRelax Mover. Resfiles describing saturate point mutation were generated for each residue on the S protein within 8 Å of S protein and hACE2 interface in the relaxed structure. The Flex ddG protocol defined in previous literature<sup>14</sup> was refactored for in-house high-performance computing platform, and implemented using pyRosetta API (Fig. S1†). For each mutant defined by a resfile, backrub sampling was applied around the mutation site. The structure was then allowed to repack and relax globally with both the WT and the mutant. The binding energy dG<sub>cross</sub> was calculated using the InterfaceAnalyzerMover, and the dG<sub>cross</sub> difference between the mutant and the WT model was taken as ddG. 48 independent ddG calculations were performed for each mutant. The mutants were sorted according to their average ddG score. After subsequent manual examination of top scored structures, 9 structures were selected for further SPR wet-lab experimental validation.

### 2.2. SPR assay

The affinity between SARS-CoV-2 Spike Protein (RBD, His tag) and hACE2 was measured using a Reichert4SPR system (Reichert Technologies, Depew, NY, USA) in single-cycle mode.

SARS-CoV-2 Spike Protein (RBD, His tag) and its mutants were immobilized to an mSAM sensor chip (planar polyethylene glycol/carboxyl sensor chip P/N 13206061) at approximately 500 response units. The experiment data were obtained at 25 °C with running buffer PBST (8 mM Na<sub>2</sub>HPO<sub>4</sub>, 136 mM NaCl, 2 mM KH<sub>2</sub>PO<sub>4</sub>, 2.6 mM KCl, and 0.05% (v/v) Tween 20, pH 7.4). Gradient concentrations of hACE2 (from 100 nM to 6.25 nM with 2-fold dilution) were then flowed over the chip surface. The resulting data were fit to a 1 : 1 binding model using Scrubber and Clamp software. Kinetic rate constant  $k_a$  and  $k_d$  were calculated from the above analysis and the apparent  $K_D$  was calculated as  $k_d/k_a$ .

### 2.3. Plasmid construction, protein expression and purification

The receptor binding domain (RBD, corresponding to Spike 319-514AA) of spike protein of SARS-COV-2 have been previously codon optimized and synthesized into pTT5 vector (EcoRI + BamHI) through PCR based overlapping oligonucleotides assembly. Kozak sequence was inserted before start codon to enhance translational efficiency. To obtain the secreted RBD protein, a mouse V-set immunoregulatory receptor signal peptide was fused to the N-terminus of the protein (1-32, MGVPAPVEASSPRWGTLLLAIFLAASRGLVAA). Ten mutations were introduced using primers with the desired mutation in a PCR protocol which amplifies the entire plasmid template (the above clone as the template). Restriction enzyme digestion and Sanger sequencing confirmed that all 10 clones were correct.

The mutated RBD proteins were prepared using the Expi293 Expression System (code: A14635). The specific operation was performed following the kit instruction. Expi293F cells were maintained in serum-free Expi293 expression medium, and the expression plasmid transfected into Expi293 cells by using an ExpiFectamine 293 Transfection Kit (all from Gibco). Five days after transfection the medium was collected, and the protein was purified by Ni-NTA (QIAGEN) column chromatography (nonspecifically binding contaminants were washed on Ni-NTA column using PBS, pH 7.4 containing 20 mM imidazole and eluted with PBS, pH 7.4 supplemented with 300 mM imidazole.). The eluted fractions were pooled and dialyzed against PBS (pH7.4) to remove the imidazole. Purified proteins were checked by SDS-PAGE and protein concentrations were quantified using Nano-Drop.

## 3. Results and discussion

### 3.1. Mutations with increased binding affinity are accurately predicted through Rosetta Flex ddG and validated by SPR assay

Rosetta Flex ddG calculation was carried out for 25 residues on the RBD binding surface. For each residue, the wild type (WT) amino acid was mutated to the other 19 natural amino acids individually and Flex ddG calculation was applied. 48 models for each mutant were generated and the ddG score were averaged over 48 models. In all 475 possible single point mutants (25 times 19), 114 mutants have negative ddG score, which



**Table 1** Results of the binding affinity predicted by Rosetta Flex ddG and measured by SPR assay

Mutant	Flex ddG- $\Delta\Delta G_{\text{binding}}^a$	SPR- $K_D^b$	SPR- $\Delta\Delta G_{\text{binding}}^c$
WT	0.00	21.08 $\pm$ 3.01	0.00
Q498W	-3.66 $\pm$ 1.80	7.10	-2.70
Q498R	-2.04 $\pm$ 1.34	11.60	-1.48
T500W	-1.90 $\pm$ 0.56	21.80	0.08
S477H	-1.39 $\pm$ 1.16	13.90	-1.03
Y505W	-1.23 $\pm$ 0.41	16.10	-0.67
T500R	-1.21 $\pm$ 1.38	12.20	-1.36
N501V	-1.02 $\pm$ 1.09	158.50	5.00
Y489W	-1.01 $\pm$ 0.50	38.90	1.52
Q493M	-0.82 $\pm$ 1.39	6.90	-2.77

<sup>a</sup>  $\Delta\Delta G_{\text{binding}}$  calculated by Rosetta Flex ddG.  $dG_{\text{cross}}$  is used as ddG score, which means the binding energy of the interface calculated with cross-interface energy terms. The unit is REU. Uncertainties were estimated from the standard deviations of 48 replicated trajectories.

<sup>b</sup>  $K_D$  (dissociation constant) assayed by SPR. The unit is nM. For WT, uncertainties were estimated from the standard deviations of 3 replicated experiments. For other mutants, the experiments were carried out only once. <sup>c</sup>  $\Delta\Delta G_{\text{binding}}$  calculated from  $K_D$  measured by SPR. The unit is kcal mol<sup>-1</sup>.

implies potential mutations with improved affinity. The lowest predicted ddG score was -3.66 REU (Rosetta Energy Unit) from mutation Q498W. Mutants with large negative ddG values were examined manually and 9 mutants were selected for SPR (Surface Plasmon Resonance) assay. The selected mutants and their corresponding predicted ddG scores are shown in Table 1.

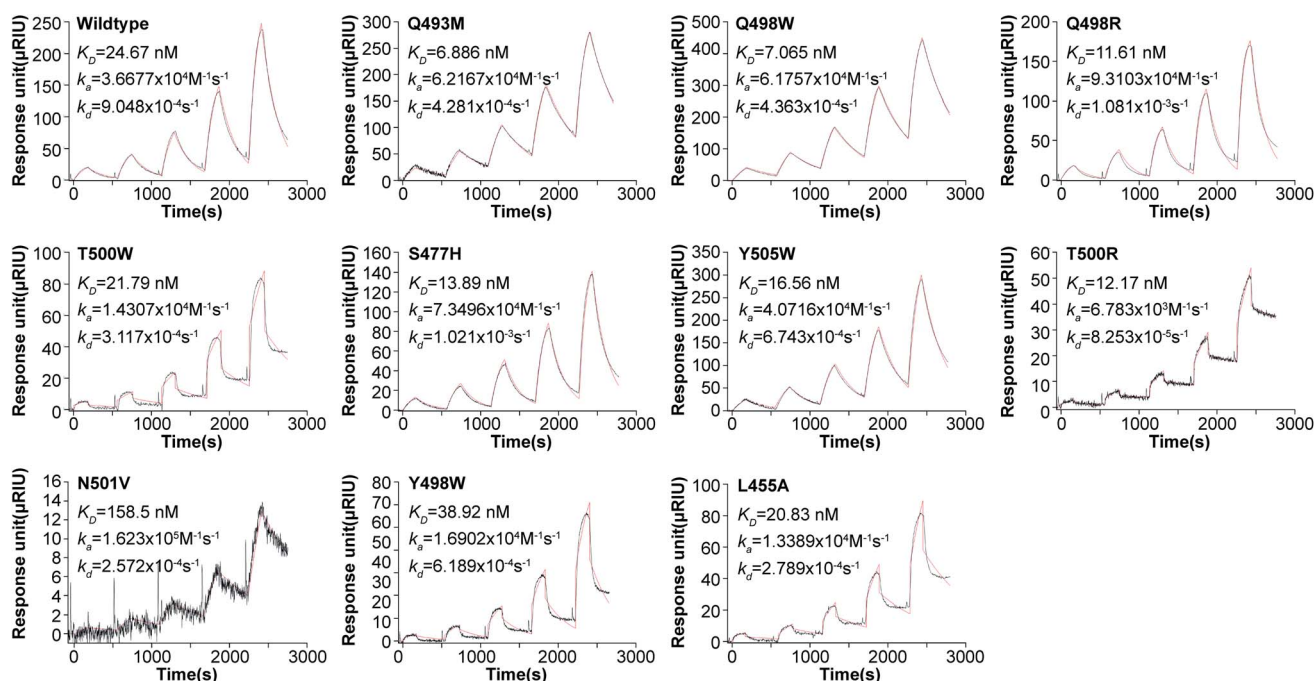
Next, we carried out SPR affinity measurement for WT RBD and 9 potentially stabilizing mutants to validate the prediction result. Fig. 1 shows the SPR sensorgram of WT and all mutants, the SPR sensorgram are shown.

The experimental  $K_D$  are reported in Table 1. 6 of the 9 selected mutants with predicted elevated binding affinity have lower  $K_D$  to hACE2 than WT RBD in SPR assay. The prediction accuracy is comparable to that reported in previous work.<sup>14</sup> Among the 9 selected mutants, the best binder is mutation Q493M with a  $K_D$  of 6.90 nM, showing a 3-fold improvement in affinity compared with that of WT. There is a modest Pearson correlation coefficient of 0.41 between predicted ddG and SPR measured  $K_D$  (Fig. 2).

Among the 5 RBD sites (477, 493, 498, 500 and 505) which give affinity-increased mutants,<sup>15</sup> 4 of them were identified as key residues interacting with hACE2 (493, 498, 500 and 505)<sup>9</sup> and 3 of them were SARS-CoV-2 unique mutants different from SARS-CoV-1 (477, 493 and 498) in previous research. The increased affinity coupled with those mutations shows that the key residues in natural binding pattern aren't necessarily optimized for binding. And even after evolution towards more potent binding, there is still great capacity for increased binding affinity.

### 3.2. Analysis of mutant structures indicates the stabilizing interaction mainly come from newly formed hydrogen bonds and hydrophobic interactions

To explain the affinity-increase effect in mutants, we analyzed the interface characters other than  $dG_{\text{cross}}$  reported by Rosetta InterfaceAnalyzeMover. We chose to analyze the  $hbond_{\text{int}}$  (count of hydrogen bonds crossing the interface) and the  $dSASA$  (buried solvent accessible area at the interface, Å<sup>2</sup>) as those terms are closely related to  $dG_{\text{cross}}$  value. In Table 2, each term was calculated by averaging over the 48 difference between the reported values of the WT model and the mutant model.



**Fig. 1** Surface plasmon resonance sensorgram showing the binding kinetics for human hACE2 and immobilized SARS-CoV-2 S protein RBD. Data are shown as black lines, and the best fit of the data to a 1 : 1 binding model is shown in red lines.



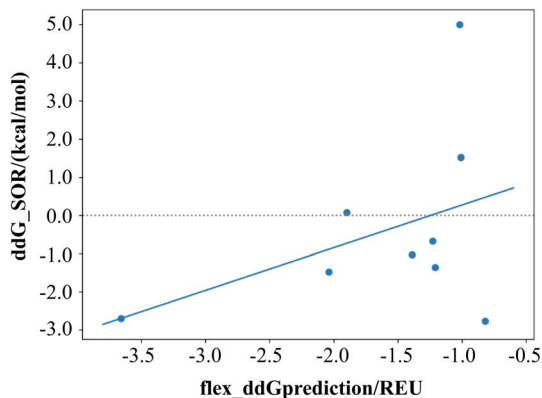


Fig. 2 ddG calculated from SPR result versus flex ddG predicted ddG value. The Pearson correlation coefficient of the regression is 0.41.

In the crystal structure, the WT Q498 forms a new hydrogen bond with nearby hACE2 42N (Fig. 3a). According to the best Q498W mutant model, the 498W residue forms a hydrogen bond with hACE2 38D (Fig. 3b). The Q498W mutation also increases hydrophobic contact with nearby residues, as indicated by the increment of hydrophobic SASA surface. For the Q498R mutant, in most of the mutant models, 498R forms hydrogen bond with hACE2 42N (Fig. 3c), and in some of the models also with hACE2 38D (Fig. 3d). This phenomenon could be explained by the fact that the longer hydrophobic part of arginine compared with glutamine could afford increased hydrophobic interaction. On average, there are 1.65 additional interface hydrogen bonds formed after mutation which may contribute to the increasing affinity. S477H mutation forms a new hydrogen bond with the terminal NH<sub>2</sub> group of hACE2 19S (Fig. 3e) according to the mutant model. Calculation result shows that additional 0.77 hydrogen bond formed in average along with increased hydrophobic interaction. For the Y505W mutant, the model shows hydrogen bond has no contribution. In the original crystal structure of WT, 505Y doesn't form any hydrogen bond with other residues. The mutation Y505W is likely increasing the affinity by evolving to bear better hydrophobic contact (Fig. 3f), as shown in the increase of hydrophobic interactions. For the T500R mutant, the dominant factor

is newly formed hydrogen bond interaction which is similar to Q498R mutant. The WT 500T forms a hydrogen bond with 41Y (Fig. 3g). In the best mutant model, T500R forms hydrogen bond with hACE2 329E and 330N (Fig. 3h), while in some other model T500R only forms hydrogen bonds to 330N (Fig. 3i). The Q493M mutant is the only affinity-increase mutant with significant decrease in hydrogen bond count. 493Q in WT originally forms a hydrogen bond with hACE2 35E (Fig. 3j). The van de waals contact increases significantly when mutating to a hydrophobic residue like methionine (Fig. 3k), indicated by the large dSASA\_hphobic in Table 2.

According to above analysis, the enhanced interaction mainly come from additional hydrogen bonds and hydrophobic interactions. The Q493M mutant also reveal that the hydrophobic term itself can be sufficient for leaning the equilibrium to more potent binding.

### 3.3. The false positive predictions are possibly due to inaccurate estimation of hydrogen bond interaction

3 of the 9 mutants (T500W, N501V and Y489W) are predicted to stabilize the interaction, however according to the SPR experiment result they are neutral or destabilizing. Next, we try to analyze the Flex ddG energy terms and structure model to elucidate the reason of the incorrect prediction. For the mutant T500W, the increase of SASA\_hphobic is the dominant factor to binding. However, the mutation to W destroy the hydrogen bond between 500T and hACE2 41Y as shown in T500R (Fig. 3g and 4a). Although in most of the WT model this hydrogen bonding pattern was well preserved, the hydrogen bond count didn't show this adverse factor as significant, with only a decrease of 0.08. This probably leads to wrong estimation of binding affinity. For the mutant N501V, in the crystal structure of N501 WT, 501N adopted the rotamer with the NH<sub>2</sub> of the amide pointing at hACE2 41Y, which forms a hydrogen bond (Fig. 4b). However, in most WT structure model generated after backrub (a method to perturbate protein backbone to generate an ensemble of conformations), the oxygen atom of the amide points to 41Y instead with an O–O distance of 3.6 Å (Fig. 4c). Therefore, the ddG calculation did not consider the destruction of the hydrogen bond, which is verified in the zero hbond count

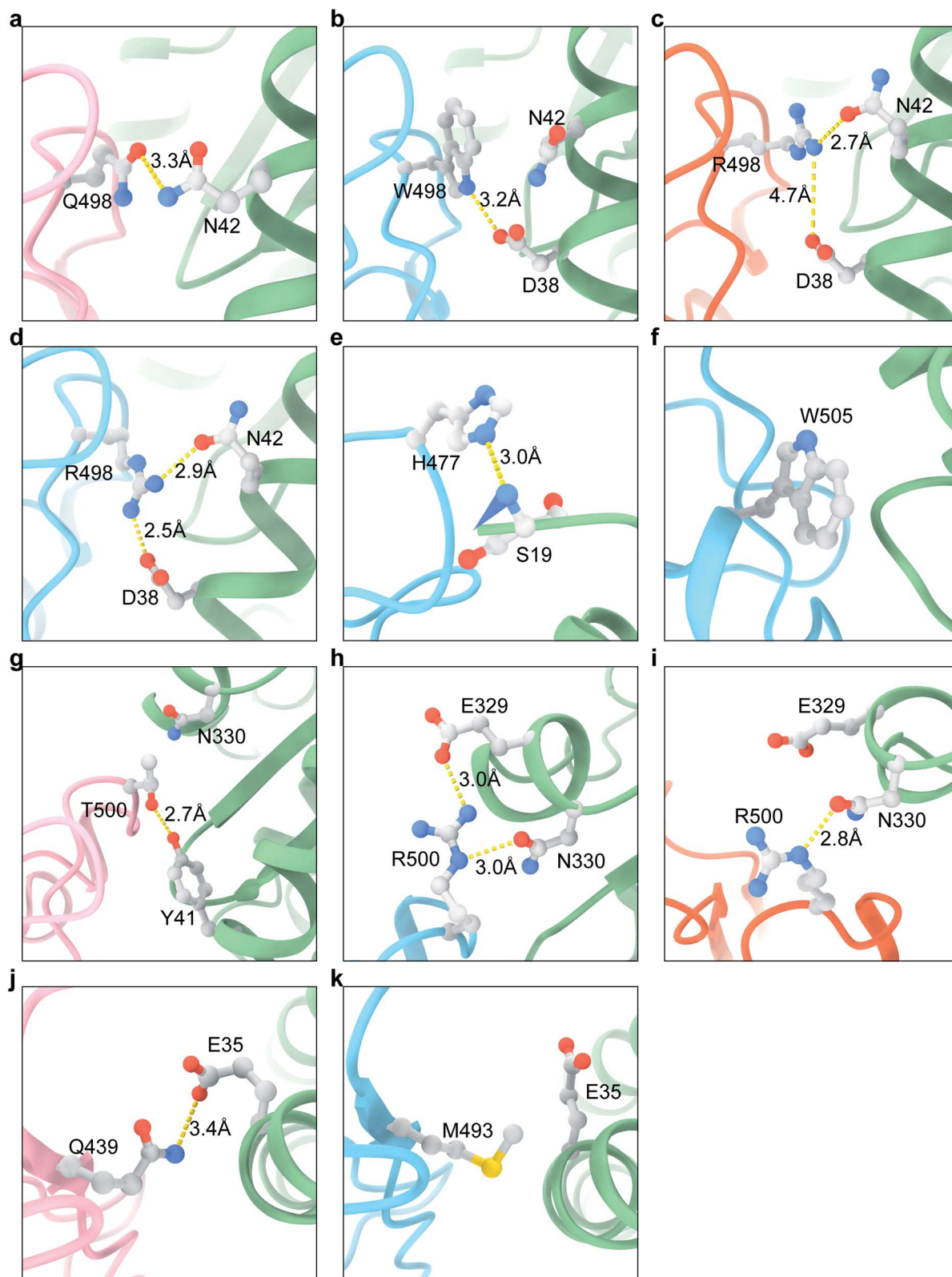
Table 2 Interface feature differences calculated from Rosetta InterfaceAnalyzer<sup>a</sup>

Mutant	mut_ddG	SPR result	hbonds_int	dSASA_int	dSASA_hphobic	dSASA_polar
Q498W	-3.66	Stabilize	0.31	39.61	58.39	-18.78
Q498R	-2.04	Stabilize	1.65	27.11	-8.08	35.20
S477H	-1.39	Stabilize	0.77	52.13	17.81	34.32
Y505W	-1.23	Stabilize	0.02	16.04	31.05	-15.01
T500R	-1.21	Stabilize	0.71	43.28	-5.05	48.33
Q493M	-0.82	Stabilize	-0.92	-4.67	71.23	-75.90
T500W	-1.90	Destabilize	-0.08	48.38	34.08	14.30
N501V	-1.02	Destabilize	0.00	0.05	12.30	-12.25
Y489W	-1.01	Destabilize	0.04	49.17	50.73	-1.56

<sup>a</sup> hbonds\_int: the change of hydrogen bond numbers at the interface. dSASA\_int: the change of solvent accessible area buried at the interface, in square Angstroms. dSASA\_hphobic: the change of the hydrophobic part of solvent accessible area buried at the interface, in square Angstroms. dSASA\_polar: the change of the polar part of solvent accessible area buried at the interface, in square Angstroms.







**Fig. 3** Predicted structure of mutants with enhanced affinity. The RBD of S protein and hACE2 are shown in cartoon representation. Residues of interest are shown in stick. Hydrogen bond is shown in yellow dash line. The distance between hydrogen bond donor and acceptor atoms is shown. (a) WT Q498 of RBD; (b) Q498W mutation of RBD; (c and d) Q498R mutation of RBD; (e) S477H mutation of RBD; (f) Y505W mutation of RBD; (g) WT T500 of RBD; (h and i) T500R mutation of RBD; (j) WT Q439 of RBD; (k) Q439M of RBD.



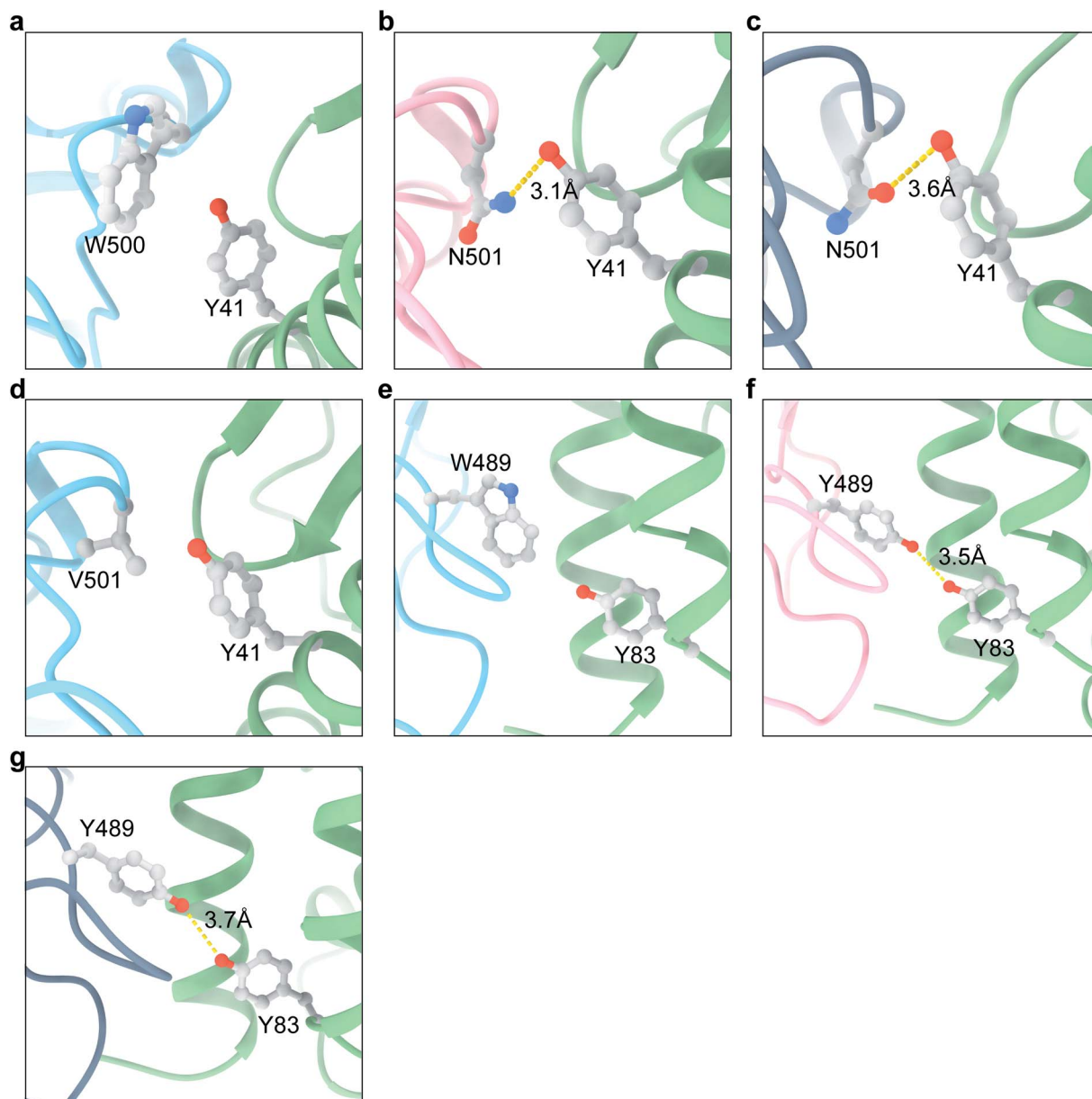


Fig. 4 Predicted structure of mutants with reduced affinity. The RBD of S protein and hACE2 are shown in cartoon representation. Residues of interest are shown in stick. Hydrogen bond is shown in yellow dash line. The distance between hydrogen bond donor and acceptor atoms is shown. (a) T500W of RBD; (b and c) WT N501 of RBD; (d) N501V mutation of RBD; (e) Y489W mutation of RBD; (f and g) WT Y489 of RBD.

difference. For the mutant Y489W (Fig. 4e), the situation is similar to N501V. In the crystal structure, the hydroxy oxygen of 489W is within 3.5 Å of hACE2 83Y (Fig. 4f). However, in most WT models generated by backrub, the distance is larger than 3.7 Å (Fig. 4g), resulting in loss of consideration for this hydrogen bond impact in the final ddG score.

By analyzing the models of false positive predictions, we illustrated that the main error may come from the broken of hydrogen bonds which are not considered in the ddG calculation. The missing hydrogen bond may be existing hydrogen bond in the scoring step and the hydrogen bond destroyed in the previous relaxing step. Visual examination of the model structure to see whether the original hydrogen bond is disrupted and comparing with the predicted

surface properties may help spot those inaccuracies and exclude some false positive instance before wet-lab experiments.

#### 3.4. The predicted binding affinity modestly agrees with deep mutational scanning (DMS) results

Starr *et al.* has systematically changed every amino acid in the S protein RBD and<sup>16</sup> determined the mutation effects on hACE2 binding using deep mutational scanning method. Although this high throughput method using flow cytometry has some contradict results compared with our SPR results (Fig. 5b), it is nevertheless a valuable data set to evaluate our Flex ddG prediction. Plotting  $\log(K_{D,mut}/K_{D,wt})$  measured by DMS *versus*



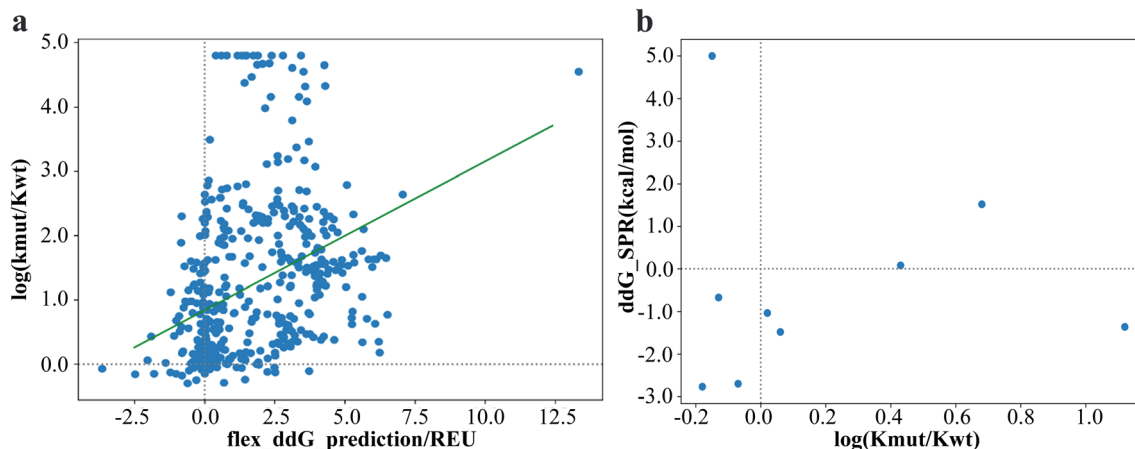


Fig. 5 (a) The ddG calculated from SPR versus  $\log(K_{d,mut}/K_{d,wt})$  from the DMS paper. The assays show contradict results on 4/9 of the mutants. (b) The  $\log(K_{d,mut}/K_{d,wt})$  from the DMS result versus our Flex ddG predicted ddG value.

Table 3 The classification performance of flex ddG prediction result using DMS dataset as ground truth

	DMS stabilizing	DMS neutral	DMS destabilizing	Precision
Predicted stabilizing	20	1	93	0.18
Predicted neutral	0	0	0	0
Predicted destabilizing	20	2	339	0.94
Recall	0.50	0	0.78	

predicted ddG gives Pearson correlation coefficient of 0.37 (Fig. 5a).

Using the result measured by DMS as the ground truth, the prediction performance of Rosetta Flex ddG method is calculated. Remarkably, the precision on predicting the stabilizing mutation is 0.18, and the recall is 0.50 given a ratio of 40/432 stabilized and destabilizing mutations, which shows the potential of the Flex ddG method to suggest stabilized mutations for enhanced binding affinity (Table 3).

## 4. Conclusions

In this article, we utilized the Rosetta Flex ddG protocol to predict mutations on the SARS-CoV-2 RBD binding surface that can strengthen its interaction to hACE2. SPR experiments confirmed 6 of the 9 predicted mutations had increased affinity, with the best mutation showed a 3-fold improvement of  $K_D$ .

<sup>17,18</sup>The SARS-CoV-2 mutants carrying the mutation Y453F were discovered in Denmark to circulate between human and minks and have brought up concerns recently. Our prediction ( $-0.30$  REU, in ESI file 2†) and<sup>16</sup> the deep mutational scan results both shows that the Y453F mutation will strengthen the interaction between RBD to hACE2, which is consistent with the maintained transmissibility and pathogenicity. Although more studies on the potential effect on treatment, diagnostic test and virus antigenicity are still ongoing, preliminary results showed that the cluster-5 strain which carries another 3 mutations outside the RBD domains was more difficult to be recognized by convalescent sera.<sup>19</sup> It is also a concern that spreading of the virus in minks may bring up more fatal variants.

The still-mutating SARS-CoV-2 alerts us of the need of an efficient mechanism in dealing with newly discovered pathogens. For such RNA virus under constant mutagenesis, it is necessary to quickly spot possible mutants with increased pathogenicity or infectivity. Given the relatively low cost and short period of computational methods, we can build a computational forecast system on virus pathogenicity in the early stage of a virus breakout. The Flex ddG method used in this work would be a constructive part in this system, predicting pathogenicity by evaluating mutant binding affinity to cell receptor. The mutations on the binding surface are especially essential in vaccine or antibody development and the predicted mutations can be considered in designing multi valent vaccine to prevent possible immune escape.

Furthermore, *in silico* affinity maturation itself is a promising technology in the development of macromolecule drugs. For instance, it can help reduce the scale of experimental screening in optimizing the affinity of peptide or protein drugs. Our work demonstrated the ability of the previously reported Flex ddG method to serve as an efficient affinity maturation method in predicting stabilizing mutants with no need for the time-consuming MD process. We anticipate that combined with machine learning and deep learning technology, Rosetta Flex ddG-based *in silico* affinity maturation can be improved to be faster and more accurate in the near future.

## Author contributions

Conceptualization, Wang T. and Wang Y.; methodology, Xue T. and Wu W.; software, Huang J.; writing—original draft preparation, Wang T.; writing—review and editing, Wang Y., Lai L.,



Liu H. and Wu C.; funding acquisition, Wang T. and Wang Y. All authors have read and agreed to the published version of the manuscript.

## Funding

Part of this work was supported by Alibaba Cloud Anti-COVID-19 Program. Part of this work was supported by Special Funds for COVID-19 Prevention and Control of West China Hospital of Sichuan University (HX-2019-nCoV-003), Sichuan Science and Technology Program (2019YFS0003), Technological Special Project for 'Significant New Drugs Development' (2018ZX09201018-021), China Postdoctoral Science Foundation (2018T110984, 2017M610607).

## Informed consent statement

Not applicable.

## Data availability statement

Relevant data are within the paper and its ESI files.† Other data and parameters generated or analyzed during the study are available from the corresponding author on reasonable request.

## Conflicts of interest

Weikun Wu, Ning Guo, Jian Huang, Lipeng Lai and Tianyuan Wang are employees of XtalPi – AI Research Center (XARC). The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## Acknowledgements

We gratefully thank Jingjia Liu for assisting in preparation of 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.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes and Y.-Z. Zhang, *Nature*, 2020, **579**, 265.
- 2 V. Tiwari, J. C. Beer, N. V. Sankaranarayanan, M. Swanson-Mungerson and U. R. Desai, *Drug Discov. Today*, 2020, **25**, 1535–1544.
- 3 W. H. Chen, U. Strych, P. J. Hotez and M. E. Bottazzi, *Curr. Trop. Med. Rep.*, 2020, **7**, 61–64.
- 4 C. Wang, W. Li, D. Drabek, N. M. A. Okba, R. van Haperen, A. Osterhaus, F. J. M. van Kuppeveld, B. L. Haagmans, F. Grosveld and B. J. Bosch, *Nat. Commun.*, 2020, **11**, 2251.
- 5 Y. Wu, F. Wang, C. Shen, W. Peng, D. Li, C. Zhao, Z. Li, S. Li, Y. Bi, Y. Yang, Y. Gong, H. Xiao, Z. Fan, S. Tan, G. Wu, W. Tan, X. Lu, C. Fan, Q. Wang, Y. Liu, C. Zhang, J. Qi, G. F. Gao, F. Gao and L. Liu, *Science*, 2020, eabc2241, DOI: 10.1126/science.abc2241.
- 6 L. Cao, I. Goreshnik, B. Coventry, J. B. Case, L. Miller, L. Kozodoy, R. E. Chen, L. Carter, A. C. Walls, Y.-J. Park, E.-M. Strauch, L. Stewart, M. S. Diamond, D. Veessler and D. Baker, *Science*, 2020, **370**, 426–431.
- 7 T. W. Linsky, R. Vergara, N. Codina, J. W. Nelson, M. J. Walker, W. Su, C. O. Barnes, T.-Y. Hsiang, K. Esser-Nobis, K. Yu, Z. B. Reneer, Y. J. Hou, T. Priya, M. Mitsumoto, A. Pong, U. Y. Lau, M. L. Mason, J. Chen, A. Chen, T. Berrocal, H. Peng, N. S. Clairmont, J. Castellanos, Y.-R. Lin, A. Josephson-Day, R. S. Baric, D. H. Fuller, C. D. Walkey, T. M. Ross, R. Swanson, P. J. Bjorkman, M. Gale, L. M. Blancas-Mejia, H.-L. Yen and D.-A. Silva, *Science*, 2020, eabe0075, DOI: 10.1126/science.abe0075.
- 8 A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veessler, *Cell*, 2020, **181**, 281.
- 9 J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang and X. Wang, *Nature*, 2020, **581**, 215–220.
- 10 J. Yang, W. Wang, Z. Chen, S. Lu, F. Yang, Z. Bi, L. Bao, F. Mo, X. Li, Y. Huang, W. Hong, Y. Yang, Y. Zhao, F. Ye, S. Lin, W. Deng, H. Chen, H. Lei, Z. Zhang, M. Luo, H. Gao, Y. Zheng, Y. Gong, X. Jiang, Y. Xu, Q. Lv, D. Li, M. Wang, F. Li, S. Wang, G. Wang, P. Yu, Y. Qu, L. Yang, H. Deng, A. Tong, J. Li, Z. Wang, J. Yang, G. Shen, Z. Zhao, Y. Li, J. Luo, H. Liu, W. Yu, M. Yang, J. Xu, J. Wang, H. Li, H. Wang, D. Kuang, P. Lin, Z. Hu, W. Guo, W. Cheng, Y. He, X. Song, C. Chen, Z. Xue, S. Yao, L. Chen, X. Ma, S. Chen, M. Gou, W. Huang, Y. Wang, C. Fan, Z. Tian, M. Shi, F.-S. Wang, L. Dai, M. Wu, G. Li, G. Wang, Y. Peng, Z. Qian, C. Huang, J. Y.-N. Lau, Z. Yang, Y. Wei, X. Cen, X. Peng, C. Qin, K. Zhang, G. Lu and X. Wei, *Nature*, 2020, **586**, 572–577.
- 11 Z. Gong, J. W. Zhu, C. P. Li, S. Jiang, L. N. Ma, B. X. Tang, D. Zou, M. L. Chen, Y. B. Sun, S. H. Song, Z. Zhang, J. F. Xiao, Y. B. Xue, Y. M. Bao, Z. L. Du and W. M. Zhao, *J. Zool. Res.*, 2020, **41**, 705–708.
- 12 Q. Li, J. Wu, J. Nie, L. Zhang, H. Hao, S. Liu, C. Zhao, Q. Zhang, H. Liu, L. Nie, H. Qin, M. Wang, Q. Lu, X. Li, Q. Sun, J. Liu, L. Zhang, X. Li, W. Huang and Y. Wang, *Cell*, 2020, **182**, 1284.
- 13 J. Zhao, R. Nussinov, W. J. Wu and B. Ma, *Antibodies*, 2018, **7**, 22.
- 14 K. A. Barlow, O. C. S, S. Thompson, P. Suresh, J. E. Lucas, M. Heinonen and T. Kortemme, *J. Phys. Chem. B*, 2018, **122**, 5389–5399.
- 15 Q. Wang, Y. Zhang, L. Wu, S. Niu, C. Song, Z. Zhang, G. Lu, C. Qiao, Y. Hu, K. Y. Yuen, Q. Wang, H. Zhou, J. Yan and J. Qi, *Cell*, 2020, **181**, 894–904.
- 16 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, A. C. Walls, N. P. King, D. Veessler and J. D. Bloom, *Cell*, 2020, **182**, 1295–1310.
- 17 S. Mallapaty, *Nature*, 2020, **587**, 340–341.
- 18 B. B. Oude Munnink, R. S. Sikkema, D. F. Nieuwenhuijse, R. J. Molenaar, E. Munger, R. Molenkamp, A. van der Spek, P. Tolma, A. Rietveld, M. Brouwer, N. Bouwmeester-Vincken, F. Harders, R. Hakze-van der Honing,





- M. C. A. Wegdam-Blans, R. J. Bouwstra, C. GeurtsvanKessel, A. A. van der Eijk, F. C. Velkers, L. A. M. Smit, A. Stegeman, W. H. M. van der Poel and M. P. G. Koopmans, *Science*, 2021, **371**, 172–177.
- 19 Detection of new SARS-CoV-2 variants related to mink. Available online: <https://www.ecdc.europa.eu/sites/default/files/documents/RRA-SARS-CoV-2-in-mink-12-nov-2020.pdf> (accessed on Dec 24th, 2020) (2020).
- 20 S. K. Tan, K. P. Fong, N. F. Polizzi, *et al.*, Modulating integrin  $\alpha$ Iib $\beta$ 3 activity through mutagenesis of allosterically regulated intersubunit contacts, *Biochemistry*, 2019, **58**(30), 3251–3259.
- 21 M. Aldeghi, V. Gapsys and B. L. de Groot, Accurate estimation of ligand binding affinity changes upon protein mutation, *ACS Cent. Sci.*, 2018, **4**(12), 1708–1718.
- 22 H. Q. Nguyen, J. Roy, B. Harink, *et al.*, Quantitative mapping of protein-peptide affinity landscapes using spectrally encoded beads, *eLife*, 2019, **8**, e40499.

