De novo design and Rosetta-based assessment of high-affinity antibody variable regions (Fv) against the SARS-CoV-2 spike receptor binding domain (RBD)

Computational design and assessment of anti-SARS-CoV-2 RBD antibodies

Veda Sheersh Boorla[# a], Ratul Chowdhury[# a], Ranjani Ramasubramanian[#b], Brandon Ameglio [#c], Rahel Frick [d], Jeffrey J. Gray*[d] and Costas D. Maranas*[a]

[a]*Department of Chemical Engineering, The Pennsylvania State University, University Park. PA 16802*
[b]*Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA*
[c]*Program in Molecular Biophysics, Johns Hopkins University, Baltimore, MD, USA*
[d]*Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA*
[*]*Corresponding author* | Email: costas@psu.edu

[#] *Equal contribution*

Abstract

The continued emergence of new SARS-CoV-2 variants has accentuated the growing need for fast and reliable methods for the design of potentially neutralizing antibodies (Abs) to counter immune evasion by the virus. Here, we report on the *de novo* computational design of high affinity antibody variable regions (Fv) through the recombination of VDJ genes targeting the most solvent-exposed hACE2-binding residues of the SARS-CoV-2 spike receptor binding domain (RBD) protein using the software tool OptMAVEn-2.0.[2] Subsequently, we carried out computational affinity maturation of the designed variable regions through amino-acid substitutions for improved binding with the target epitope. Immunogenicity of designs was restricted by preferring designs that match sequences from a 9-mer library of "human antibodies" based on human string content score[74]. We generated 106 different designs and reported in detail on the top five that trade-off the greatest computational binding affinity for the RBD with human string content scores. We further describe computational evaluation of the top five designs produced by OptMAVEn-2.0[1] using a Rosetta-based approach. We used Rosetta *SnugDock*[2,3] for local docking of the designs to evaluate their potential to bind the spike RBD and performed "forward folding" with *DeepAb*[4] to assess their potential to fold into the designed structures. Ultimately, our results identified one designed antibody variable region, P1.D1, as a particularly promising candidate for experimental testing. This effort puts forth a computational workflow for the *de novo* design and evaluation of antibodies that can quickly be adapted to target spike epitopes of emerging SARS-CoV-2 variants or other antigenic targets.

Introduction

The COVID-19 pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) still poses a threat due to the continued emergence of variants.[5,6] Early on in the pandemic, convalescent plasma extracted from SARS-CoV-2-infected patients was used as a treatment.[7,8] Characterization of the plasma samples of convalescent individuals showed that the majority of the neutralizing antibodies (nAbs) target the receptor binding domain (RBD) of the spike protein indicating its high immunodominance.[9,10] Structural studies of nAbs in complex with the RBD indicated antigenic sites throughout the RBD. These varied antigenic sites suggest that nAbs can have different mechanisms of antiviral action ranging from completely blocking receptor binding to only disrupting some interactions with the receptor.[11,12] Newly emerging variants contain an increasing fraction of amino acid changes in the RBD that confer immune evasion rather than increased affinity with the receptor protein hACE2 (human receptor - angiotensin converting enzyme).[13,14] Furthermore, as newer immune evading variants emerge, existing antibodies for both therapeutic and serological testing appear to be less effective.[15,16] For example, amino acid changes at the positions Phe456 and Glu484 on the RBD (present in Beta, Gamma and Omicron variants) have been shown to prevent neutralization by monoclonal antibodies and convalescent plasma.[17,18] Moreover, a recent study showed that SARS-CoV-2 can completely escape from highly neutralizing COVID-19 convalescent plasma after several months of laboratory evolution.[19] This escape can occur due to mutations and the presence of glycans in the N-terminal domain and the RBD of the virus. Therefore, the continued emergence of immune evading variants motivates the development of faster and cheaper methods for targeted design of nAbs.

Monoclonal antibodies derived from both convalescent plasma and phage display libraries have shown excellent neutralizing properties against SARS-CoV-2 in both animal models[20,21] and live-virus or pseudo-virus assays.[22,23] Some potential antibody-based therapeutics have also shown effectiveness in randomized clinical trials.[24,25,26] Several hundreds of these antibodies have been isolated and characterized through detailed biochemical and structural studies (refer reviews[27,28] for a detailed discussion of existing antibodies). Several computational studies have focused on the redesign of existing antibodies against other coronaviruses towards better binding with SARS-CoV-2 RBD using machine-learning,[29,30] structure guided mutagenesis[31,32] and CDR grafting approaches.[33] Additionally, computationally designed peptide inhibitors have shown promise in preventing SARS-CoV-2 infection in mammalian cells.[34,35] However, studies that perform structure guided design of fully *de novo* high affinity antibodies against specific epitopes of SARS-CoV-2 spike protein are still lacking. Motivated by this shortcoming, here we explore the *de novo* design

of antibody variable regions targeting the most solvent-exposed residues of the spike protein that are also part of the residue contact map involved in hACE2 binding, and trade-off binding energy against human sequence content in the variable region. We focus on hACE2 binding because antibody designs which may interfere with hACE2 binding regions on the RBD could be neutralizing by disrupting RBD-hACE2 interaction and hence block SARS-CoV-2 cell entry[76]. Our goal is to exhaustively explore the sequence space of all possible variable region designs using the antibody design software OptMAVEn-2.0[1] and report a range of diverse solutions that can serve as potentially neutralizing antibodies (nAb). We find that many different combinations of VDJ genes followed by mutation can yield potentially high affinity variable regions (scored using the Rosetta energy function) against an epitope of the spike protein RBD. Pareto optimal designs with respect to binding affinity vs. human content were drawn and five affinity matured designs are detailed in the results section. By grafting the designed heavy (VH) and light (VL) chain variable regions onto a human framework region, high-affinity and potentially neutralizing full-length monoclonal antibodies can also be constructed. Additionally, we describe a Rosetta-based computational evaluation of our top five designs where we compared the designs generated with OptMAVEn-2.0[1] to structural models generated by DeepAb and Rosetta to help identify the most promising designs prior to experimental testing.

Results

We first performed solvent accessibility analysis using the STRIDE[36] program on the 21 hACE2-binding residues of the SARS-CoV-2 spike protein (S-protein) RBD to define our binding epitope. The top seven residues with the highest solvent accessibility scores (i.e., SAS) are (Arg346, Phe347, Ala348, Tyr351, Ala352, Asn354, and Arg355) comprising our binding epitope (see Figure 1). Furthermore, the epitope is accessible for binding to RBD in the open conformation of the full spike protein (See Supplementary Supplementary SI_1 Fig. S9).
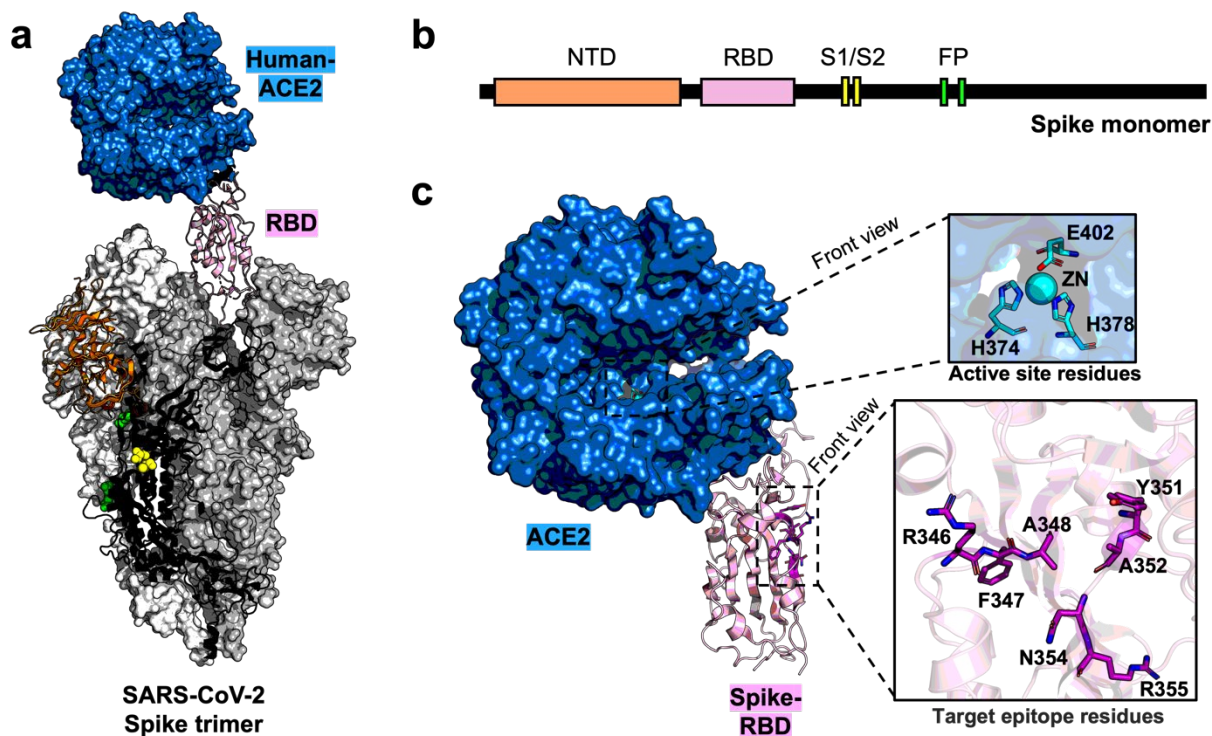
Figure 1. (a) The SARS-CoV-2 spike trimer structure in RBD-open conformation (PDB-id: 6VYB) superposed with human-ACE2-RBD complex (PDB-id: 6LZG). One of the trimer units is shown in cartoon representation color coded as per the schematic in **(b).** NTD, N-terminal domain; RBD, Receptor binding domain; S1/S2, ends of S1/S2 cleavage site; FP, ends of Fusion peptide. **(c)** Zoomed-in view of the human-ACE2-RBD complex from **(a).** ACE2's active-site residues and the target epitope residues on the RBD are shown as viewed from the front in inset boxes (top & bottom respectively). The numbering scheme for the ACE2 and RBD residues is the same as that in the crystal structure 6LZG (rcsb.org/structure/6LZG or 6VW1[37] and 6M0J[38]).

We next used OptMAVEn-2.0 to computationally identify the combination of VDJ genes forming the variable region that best binds the desired epitope. Previously, OptMAVEn[39] has been used successfully to design five high affinity CDRs against a FLAG tetrapeptide,[40] three thermally and conformationally stable antibody variable regions (sharing less than 75% sequence similarity to any naturally occurring antibody sequence) against a dodecapeptide mimic of carbohydrates[41] and two thermostable, high affinity variable heavy chain domains (V_HH) against α-synuclein peptide responsible for Parkinson's disease.[42] All these designs were experimentally constructed and nanomolar affinities for their respective target antigens was demonstrated.
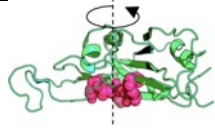
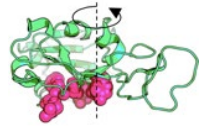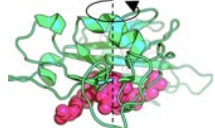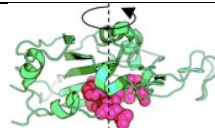Through a combination of rotations and translations, OptMAVEn-2.0 identified 3,234 unique antigen poses that presented the epitope to the antibody differently. The combinatorial space of different VDJ genes that upon recombination form the variable region of the prototype antibody was informed by the MAPs database of antibody parts.[43] MAPs (see Supplementary SI_1 S1 for link to full database) contains 929 modular

antibody (i.e., variable-V*, complementarity determining -CDR3, and joining-J*) parts from 1,168 human, humanized, chimeric, and mouse antibody structures (last updated in 2013). MAPs follows the antibody parts residue numbering convention as listed in the International iMmunoGeneTics (IMGT)[44] database. IMGT catalogs antibody parts as variable (V), diversity (D) and joining (J) structure libraries. MAPs stores complete CDR3 parts, C-terminus-shortened V parts (i.e. V* parts) and N-terminus-shortened J parts (J* parts). Note that CDR3 includes the entire D gene and also up to the C-terminus of the V gene and up to the N-terminus of the J gene. In the remainder of the manuscript, the list of parts used to design the variable region are referred to as CDR3, V* and J* parts.

For each one of the 3,234 spike poses, OptMAVEn-2.0 identified a variable region combination composed of end-to-end joined V*, CDR3, and J* region parts that minimized the Rosetta binding energy between the variable region and spike epitope formed by the seven residues. As part of OptMAVEn-2.0, the combinatorial optimization problem was posed and solved as a mixed-integer linear programming (MILP) problem using the cplex solver[45]. The solution of this problem identifies, for each one of the spike poses, the complete design of the variable region using parts denoted as HV*, HCDR3, HJ* for the heavy chain H and L/KV*, L/KCDR3 and L/KJ* for the light chain-L/K. Only 173 antigen-presenting poses out of 3,234 explored yielded non-clashing antigen-antibody complexes. These 173 poses were ranked based on their Rosetta binding energies with the spike epitope and classified into 27 clusters (using *k-means*[46]) in a 19-dimensional space defined by quantitative descriptors of sequence similarity, three-dimensional spatial pose, and the angle at which they bind to the target epitope (see details in original paper[1]). The top five prototype designs with the highest Rosetta binding energies were present in four clusters and spanned a highly diverse set of choices of MAPs parts (see Table 1) with minimal conservation of the same part among the five prototype designs. The number entries in Table 1 correspond to the id of the gene in the MAPs database (which are identical to the ids used in IMGT). Note that the design P5 uses a lambda (L) light chain instead of a kappa (K). Figure 2a plots the pairwise sequence similarity scores of the five antibody variable domains that were used in the top five designs. As expected, the top five prototype designs P1, P2, P3, P4, and P5 are the most dissimilar in their respective CDR3 domains in both light L, heavy H and HV* domain (but not LV*). They are the most similar in the choice of parts for the J* domains (see Figure 2a) reflecting the lack of diversity among possible choices for the J* domains in the MAPs database.

Table 1. V*, CDR3, J* part numbers chosen from the MAPs[43] database for the top five prototype variable region designs and their corresponding Rosetta binding energies.[47,48] Antigen poses are described with the angle that the vertical axis through the epitope (shown in pink) centroid and the Cβ carbon of the residue with greatest z-axis coordinate forms.

Modular Antibody Parts number chosen in each design		

Prototype design	HV*	HCDR3	HJ*	L/KV*	L/KCDR3	L/KJ*	Antigen pose (rotation of epitope about vertical axis)	Rosetta binding energy (kcal/mol)
P1	82	315	5	61(K)	4(K)	3(K)	0°	-36.77
P2	52	94	1	61(K)	17(K)	3(K)	0°	-27.57
P3	105	12	5	6(K)	23(K)	4(K)	300°	-29.20
P4	79	204	1	2(K)	1(K)	4(K)	240°	-19.78
P5	108	212	1	37 (L)	5 (L)	5 (L)	360°	-38.62

Inspection of the non-covalent interactions formed by the design P1 with the spike epitope revealed (Fig. 3a) both electrostatic (22 hydrogen bonds) and hydrophobic contacts at the interface (Supplementary SI_1 S3a). Lys356 and Asn354, Trp353, Tyr351, Arg454, Arg457 of RBD form two important clusters of interactions with light chain and heavy chain residues of P1 respectively (see Fig 3b, 3c). Notably, RBD residues Trp353 and Asn354 have been deemed important for hACE2 binding.[49] Similarly, the designs P2 through P5 capture the RBD by multiple hydrogen bonds, hydrophobic interactions, and salt-bridges (Supplementary SI_1 S3(b-e)).
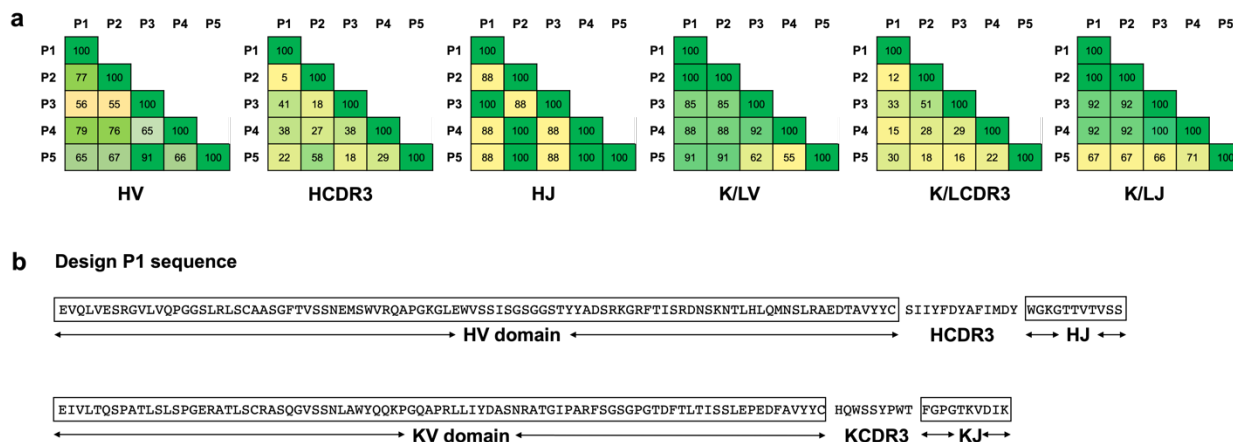
Figure 2. (a) Pairwise sequence similarity percentages between the members of the six parts that were used to construct the top five prototype variable regions with the lowest Rosetta binding energies with the viral spike epitope. **(b)** The amino acid sequence of prototype design P1 with the different domain parts of heavy-chain (HV, HCDR2, HJ) and light-chain (KV, KCDR3, KJ) labelled below.

We next applied Rosetta-based *in silico* affinity maturation (see Methods) for each one of the top five prototype designs shown in Table 1 to further improve non-covalent binding interactions between the antibody variable domains and the SARS-CoV-2 spike RBD. This computationally mimics the process of somatic hypermutation leading to eventual affinity maturation of antibodies in B cells. This procedure identified a total of 124 unique variable designs by introducing mutations in the five prototypes (see Figure 4a). We retained 106 designs of these which achieved both an improvement in the Rosetta binding energy over their respective prototype sequences and further stabilization (i.e., lower total Rosetta energy) of the antigen-antibody complex (see upper right quadrant of Figure 4a). On average, upon affinity maturation, the binding energy was improved by ~14 kcal/mol and the total energy of the complex was improved by ~37 kcal/mol. Supplementary file SI_1 S2 lists first the starting prototype design (i.e., P1, P2, P3, P4 or P5) followed by the 106 affinity matured designs (labeled as P1.D1, P1.D2, etc). On average, there were 4.5 mutations (Supplementary SI_1 S4) between computational affinity matured and prototype variable region designs.

We then assessed the departure of the 106 affinity-matured variable regions from fully human antibody sequences using H-Score.[50] H-score is defined as the sum of the minimum edit distance between all possible 9-mer fragments in the designed variable region from a library of all 9-mer sequences (prepared in *T.Li et. al.*[39]) observed in human antibodies.[39] The value of H-score is scaled to 100 and normalized by the length of the sequence. Between two sequences, the sequence with greater human antibody sequence content will have a higher value of H-score and vice-versa. Figure 4b illustrates the trade-off between the Rosetta

binding energy vs. H-score for these affinity-matured variable region designs. For comparison, we calculated the H-score for the human antibodies CR3022,[51] 80R,[52] S230,[53] and M396[54] which are known to be neutralizing against SARS-CoV-1. These antibodies (including only Fv regions) have an average H-score of 62.98 (stdev: 4.9) which are in the same range as our most human designs (Fig 4b and 4c). Note that one could use a 9-mer sequence library prepared using antibody sequences relevant to a particular application (here, anti-RBD antibodies), and hence tune the H-score calculation to obtain perfect scores (close to 100) for known anti-RBD human antibodies.
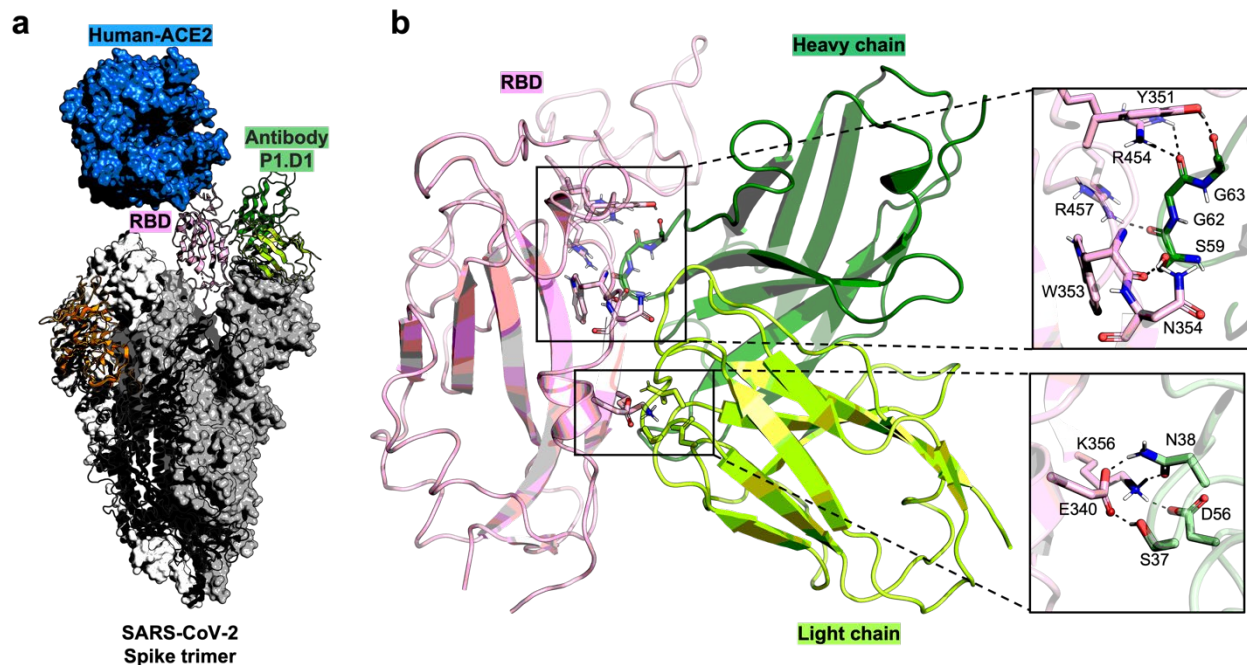
Figure 3. **(a)** The SARS-CoV-2 spike trimer structure in RBD-open conformation superposed with human-ACE2-RBD complex and the antibody design P1.D1; RBD and NTD are shown in orange and pink cartoons, antibody is shown in green cartoon and ACE2 is shown in blue surface. **(b)** Structural view of the interface between the RBD (shown in pink) and antibody design P1 with amino-acid residues forming representative non-covalent interactions shown as stick representations. Both Light (shown in pale green) and Heavy chains (shown in green) of the antibody contribute to the interface with the RBD. The insets show zoomed-in views of the interactions formed at the Heavy-chain (top) and Light-chain interfaces, respectively.

We selected five designs that were on the Pareto optimum curve shown in Figure 4b. The Pareto optimum curve is defined as the collection of designs for which no other design exists that can simultaneously improve upon both criteria (i.e., Rosetta binding energy and H-score). Designs P1.D1 and P1.D2 shown in blue (in Fig. 4b) have the lowest Rosetta binding energies whereas P3.D1 and P3.D3 shown in yellow correspond to the ones with the highest H-scores. Design P4.D1 is an intermediate design that balances both binding energy and H-score. The lowest binding energy designs (P1.D1, P1.D2), irrespective of H-score, would be relevant in ELISA-based in vitro detection assays whereas the lowest H-score designs (P3.D1, P3.D3) may offer the highest potential as therapeutic antibodies. In addition, we calculated the Rosetta binding energy between the human CR302255 (anti-SARS-CoV-1 antibody) and the SARS-CoV-2 spike protein RBD using complex structure (PDB-id:6W41) to be -56.4 kcal/mol which is very close to
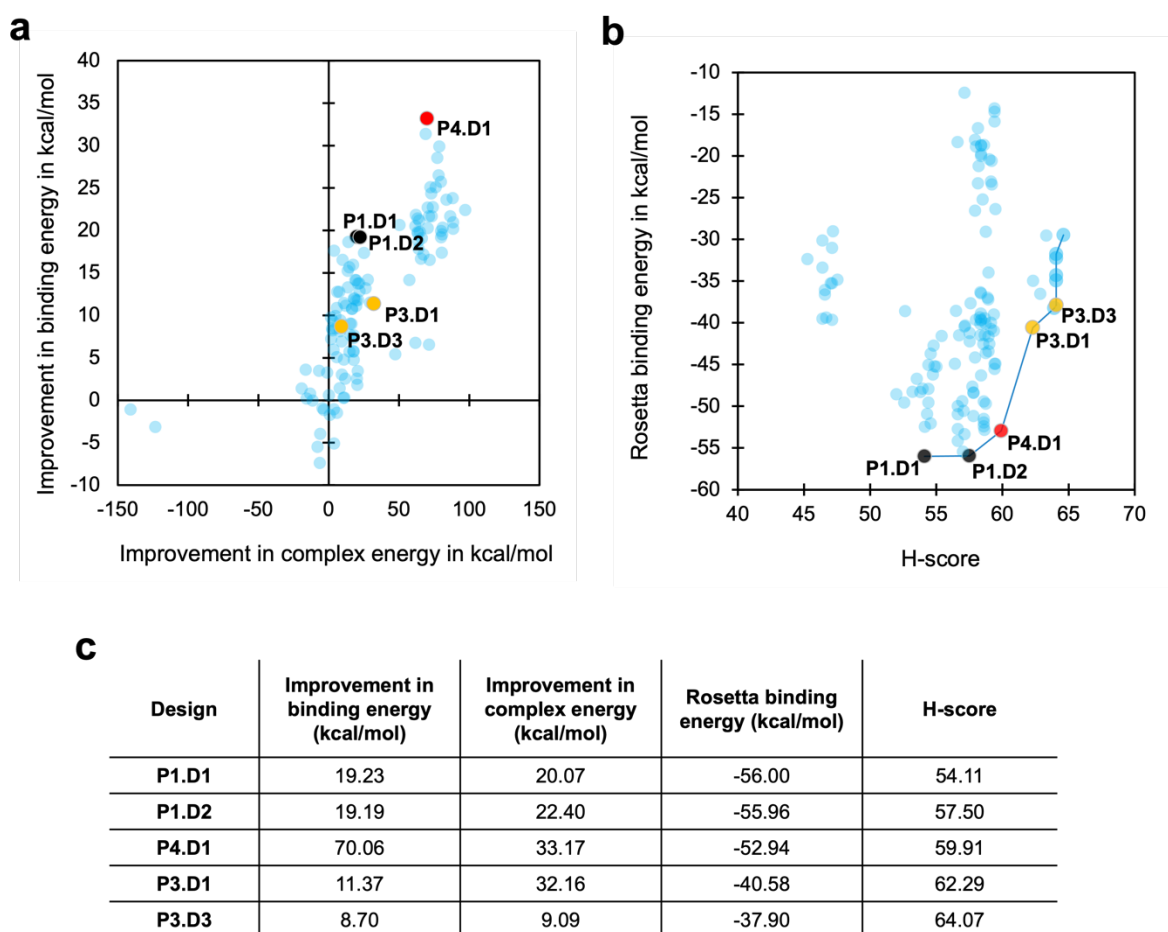

Design	Improvement in binding energy (kcal/mol)	Improvement in complex energy (kcal/mol)	Rosetta binding energy (kcal/mol)	H-score
P1.D1	19.23	20.07	-56.00	54.11
P1.D2	19.19	22.40	-55.96	57.50
P4.D1	70.06	33.17	-52.94	59.91
P3.D1	11.37	32.16	-40.58	62.29
P3.D3	8.70	9.09	-37.90	64.07

Figure 4. (a) Scatter plot of improvement in Rosetta binding energy vs. complex energy for the 124 affinity matured designs. **(b)** Scatter plot of the Rosetta binding energy vs. H-score for the 106 affinity-matured sequences that fall in top-right quadrant of plot a. The blue line connects the designs on the Pareto optimum curve that trade-off binding energy with H-score. The three best-binding affinity matured designs (emerging from P1) – P1.D1 and P1.D2 are highlighted in black whereas the two most human designs (emerging from P3) – P3.D1 and P3.D3 are highlighted in yellow. An intermediate design P4.D1 (emerging from P4) is highlighted in red.

the Rosetta binding energy of designs P1.D1 and P1.D2. However, P1.D1 and P1.D2 bind a different epitope on the spike RBD than the one that CR3022 targets (see Supplementary SI_1 Fig S8). Supplementary excel file SI_2 contains the list of Rosetta energies and H-scores for all 106 affinity matured designs.

Figure 5 shows the sequence alignment of the five selected affinity-matured sequences (i.e., P1.D1, P1.D2, P4.D2, P3.D1, and P3.D3). A total of 156 out of 226 aligned positions are conserved among all designs. Table 2 lists the most important (strongest) contacts with the spike protein as informed by an *in-silico* alanine scanning (Supplementary SI_1 Table S5) on the spike-binding residues of the variable region designs. In essence, the alanine scanning analysis identifies the loss in binding energy that is incurred upon mutating each residue (one at a time) to alanine.
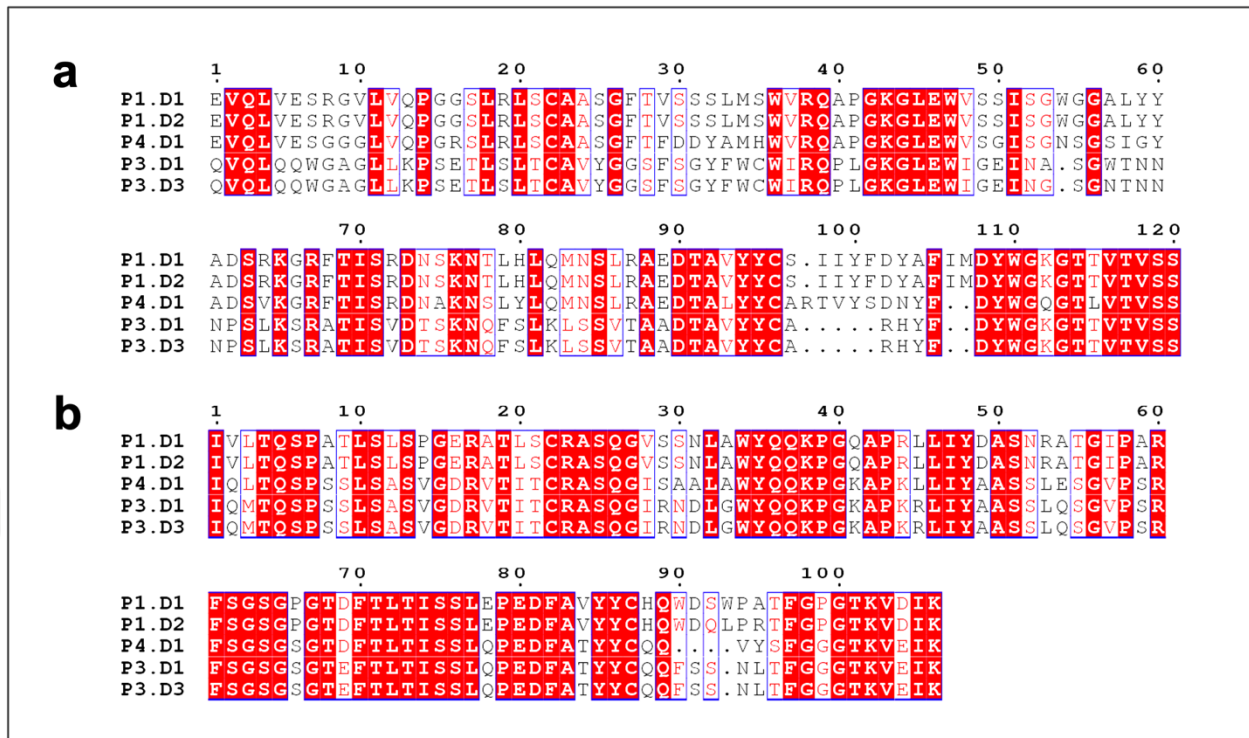
Figure 5. Sequence alignment of (a) Heavy chain sequences and (b) light chain sequences of the five pareto optimal affinity matured sequences with the positions conserved amino acids across designs highlighted in red with white font and those with conserved residue type are highlighted in white with red font.

Table 2. List of important contacts between the RBD epitope and each of the selected affinity matured designs. For each contact, the corresponding amino acid residues from the antibody and RBD are listed along with the loss in binding energy upon mutating the antibody amino acid to alanine.

Matured antibody Design id	Interface residue from antibody	Interacting spike residue(s)	Loss in binding energy upon mutation to alanine (kcal/mol)	Corresponding variable region
P1.D1	G62	A352	9.69	HV*
	G63	Y351	2.13	HV*
	L65	Y489	1.26	HV*
	Y66	S349	0.53	HV*
	I56	C488	0.94	HV*
P1.D2	F109	K356	2.77	HCDR3
	T65	Y351	1.42	HV*
	Y66	A348,S349	1.06	HV*
	I56	C488	0.94	HV*
	S57	A352	0.71	HV*
P4.D1	A56	L452	0.573	KV*
	D35	R357	0.08	HV*
	G28	T478,G482	0.01	KV*
	T85	N481	0.00	KV*
	L67	V445	0.00	KV*
P3.D1	W64	A352	2.30	HV*
	N57	N354	0.65	HV*
	F107	T345	0.57	KCDR3
	S29	E340	0.14	HV*
	S108	R346	0.12	KCDR3
P3.D3	N57	N354	0.99	HV*
	F107	T345	0.33	KCDR3
	D38	R346	0.18	KV*
	S29	E340	0.22	HV*
	Q106	R346	0.03	KCDR3

Antibodies that strongly bind to the RBD but do not inhibit hACE2 binding have been shown to be neutralizing for SARS-CoV-2 (47D11[56]) and for SARS-CoV-1 (CR3022 in combination with CR3014).[51] The mechanisms of neutralization of such antibodies are not completely known.[56, 75] It is possible that upon binding, these antibodies perturb the interaction network of the RBD with hACE2 thereby rendering RBD-hACE2 binding less effective. Using the cryo-EM structure of 47D11 (PDB-id: 7AKD[75]) in complex with RBD, we found that the Rosetta binding energy is -52.56 kcal/mol which is close to the energy of our best binding designs (Fig. 3c).In addition, Daniel et al[57] showed that nanobody VHH-72 raised against SARS-CoV-1 had a neutralizing effect despite binding to an epitope that does not overlap with the hACE2 residue binding domain. By fusing VHH-72 with a human IgG1 they demonstrated SARS-CoV-2 neutralizing activity. They hypothesized that binding of the nanobody with the trimeric spike protein may disrupt conformational dynamics and consequently prevent binding to hACE2.

In comparison, our design P1.D1 forms strong contacts (see Table 2) with many residues of the RBD which in turn also indirectly interact with hACE2 (see Fig. 6). For example, residues L455 and T470 of the RBD are in contact with both hACE2 contacting RBD residues Y449, F490 and P1.D1 contacting RBD residues

Y351, I465. By perturbing the inter-residue interaction network of RBD-hACE2 a neutralizing effect can be achieved.
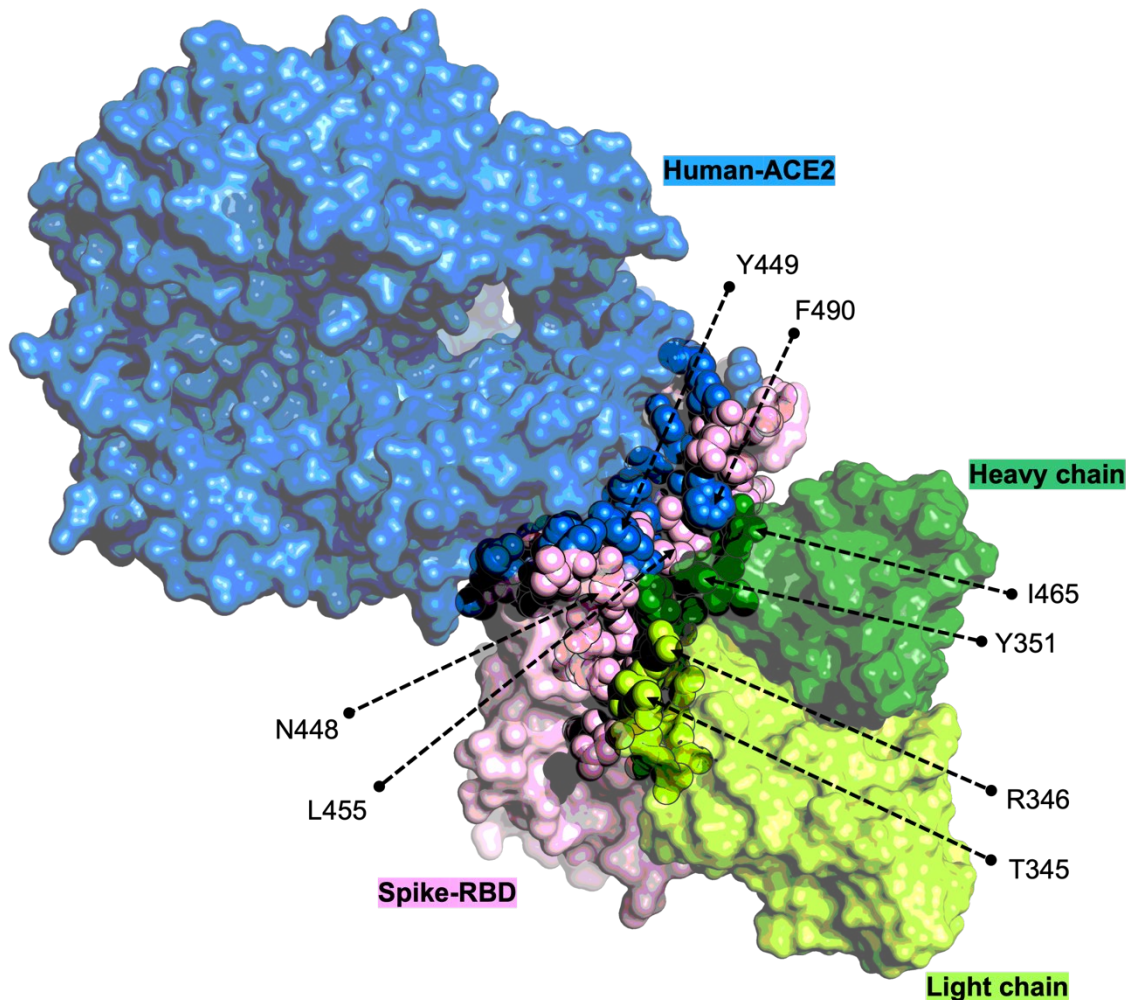

Figure 6. Surface representation view of the P1.D1-RBD complex superimposed on the ACE2-RBD complex with the contacting residues at interfaces shown as opaque spheres while the rest of the structure is shown as transparent surface. RBD residues in contact with ACE2 residues are shown in blue spheres and those in contact with the P1.D1's heavy chain and light chain are shown in dark green spheres and bright green spheres, respectively. RBD residues in contact with both ACE2 and P1.D1 are shown in pink spheres. RBD residues Y449, F490 are at the ACE2-RBD interface; I465, Y351 are at the RBD-heavy-chain interface; R346, T345 are at the RBD-light-chain interface and N448, L455 are in between the ACE2-RBD and RBD-heavy/light-chain interfaces.
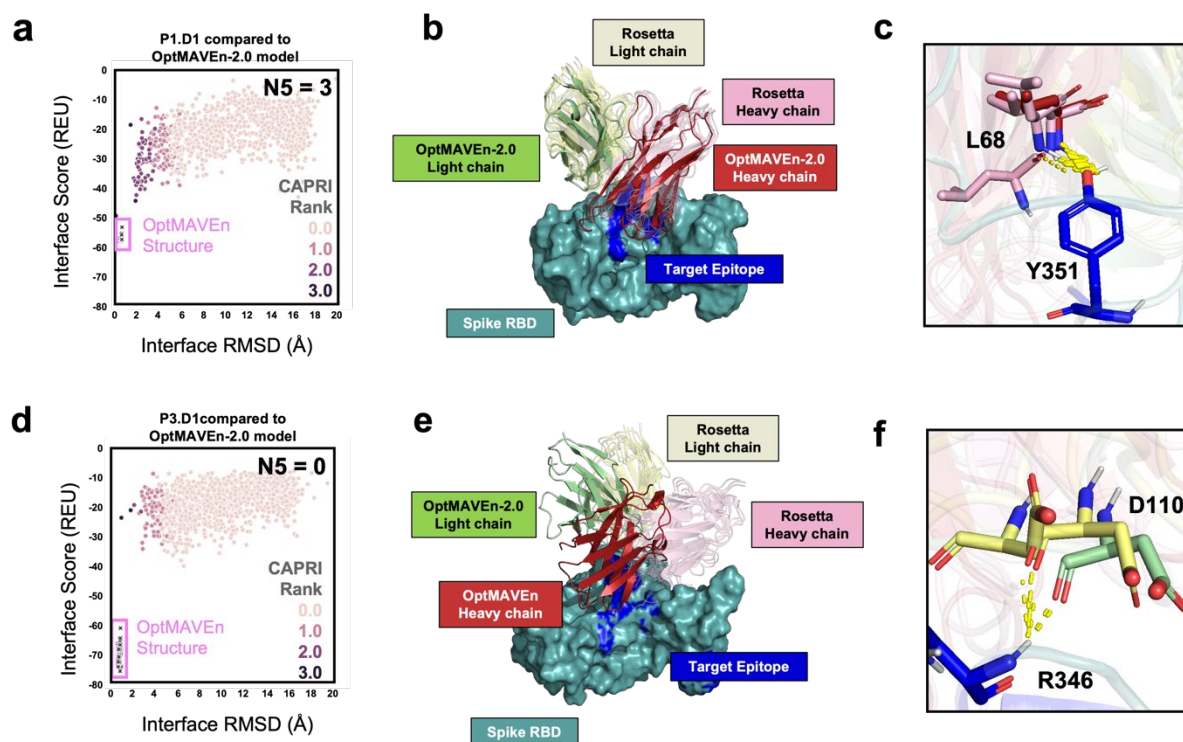
Figure 7. Computational evaluation of best binding designs using Rosetta docking studies. Funnel plots obtained for the docking results where interface RMSD is calculated with respect to the input complex, color coded by CAPRI criteria[59,60] for (a) Design P1.D1 and (d) Design P3.D1. Cartoon representations of the OptMAVEn-2.0 designed input structure (light and heavy chains shown in green and red respectively) and the lowest scoring decoys which agree the most in orientation (light and heavy chains shown in yellow and pink respectively) with respect to the RBD (shown in teal) for (b) Design P1.D1 which includes seven of the top ten lowest scoring decoys and (e) Design P3.D1 which includes four of the top ten lowest scoring decoys. (c,f) Preserved polar interactions at the interface between important residues (see Table 2) found in the top ten lowest scoring decoys for Design P1.D1.

Next, we used Rosetta's *SnugDock*[2,3] application to computationally predict if the designed binding complexes can be recovered through Rosetta's Monte Carlo approach to local docking. Predicted binding energy was evaluated by plotting the Rosetta interface score—an estimate for binding energy—against the interface RMSD when calculated from the designed OptMAVEn-2.0[1] structure. We considered an antibody to be a more likely binder if these plots show that models similar to the design (low-RMSD) have distinctly lower interface scores than the bulk of the models (Fig. 7a and 7d). To assess whether *SnugDock* was able to find energy minima, we calculated the interface RMSD from the input structure. We evaluated the top five lowest binding energy designs overall and the top two lowest binding energy designs for each prototype based on Rosetta binding energy. The top five designs produced the most promising energy funnels. These designs were all derived from the P1 prototype, of which P1.D1 was the top design overall and produced the deepest energy funnels (Fig. 7a and Supplementary SI_1 Fig. S10). To quantify the quality of the funnel, we use the N5 metric,[61] which is the number of top 5 lowest scoring decoys within 4.0Å from the designed structure. We consider an N5 of 3 or above a good funnel. For P1.D1 we observed an N5 of 3 (Fig. 7a). Structural visualization of the lowest scoring models and the designed structure further confirmed highly

similar binding modes for the design P1.D1. Seven of the top ten lowest scoring decoys appear to agree in orientation with each other and with the designed structure (see Fig. 7b).

By contrast, P3.D1 shows less promising energy funnels (Fig. 7c) with an N5 of 0, which is consistent with the fact that the initial OptMAVEn-2.0 P3 designs had a less favorable Rosetta binding energy than the top P1 designs (Table 1). The structures revealed vastly different binding modes between the low energy *SnugDock* models and the designed structure, with only partly overlapping footprints, and only moderate agreement among the low scoring models (Fig. 7d). Therefore, this design is unlikely to bind in the location designed for by OptMAVEn-2.0.

Analysis of polar interactions for P1.D1 showed that all the top ten lowest scoring *SnugDock* structures include at least one hydrogen bond between the desired RBD epitope and antibody, while eight of these structures include at least three hydrogen bonds. Although lower than the number of hydrogen bonds found in the MD simulations involving P1.D1, this is still a promising number of bonds since it exceeds the average number of hydrogen bonds between hACE2 and the RBD.[37] A hydrogen bond involving Tyr351 sidechain of the desired RBD epitope and Leu68 backbone of the antibody is conserved between the OptMAVEn-2.0 structure and six of the top ten lowest scoring *SnugDock* structures (Fig. 7c, Supplementary SI_1 Fig. S11). One additional *SnugDock* structure engages Tyr351 with Gly65 instead (Supplementary SI_1 Fig. S11-F). Similarly, a hydrogen bond between the Arg346 sidechain of the epitope and the Asp110 backbone of the antibody is conserved in seven out of the top ten lowest scoring *SnugDock* structures (Fig. 7f, Supplementary SI_1 Fig. S11). Two additional *SnugDock* structures predict an H-bond between Arg346 and the Phe111 backbone or the Gln27 sidechain instead (Supplementary SI_1 Fig. S11-E and S11-I). Additionally, many of the *SnugDock* lowest scoring structures include hydrogen bonds not found in the OptMAVEn-2.0 structure, some of which involve antibody sidechains. The notable ones are a hydrogen bond involving the sidechain of antibody light chain Trp135 and the sidechain of epitope Asn354 (Supplementary SI_1 Fig. S11-B), and a bond involving the sidechain of antibody heavy chain Tyr69 with the sidechain of epitope Ala352 (Supplementary SI_1 Fig. S11-A, D, H, I).
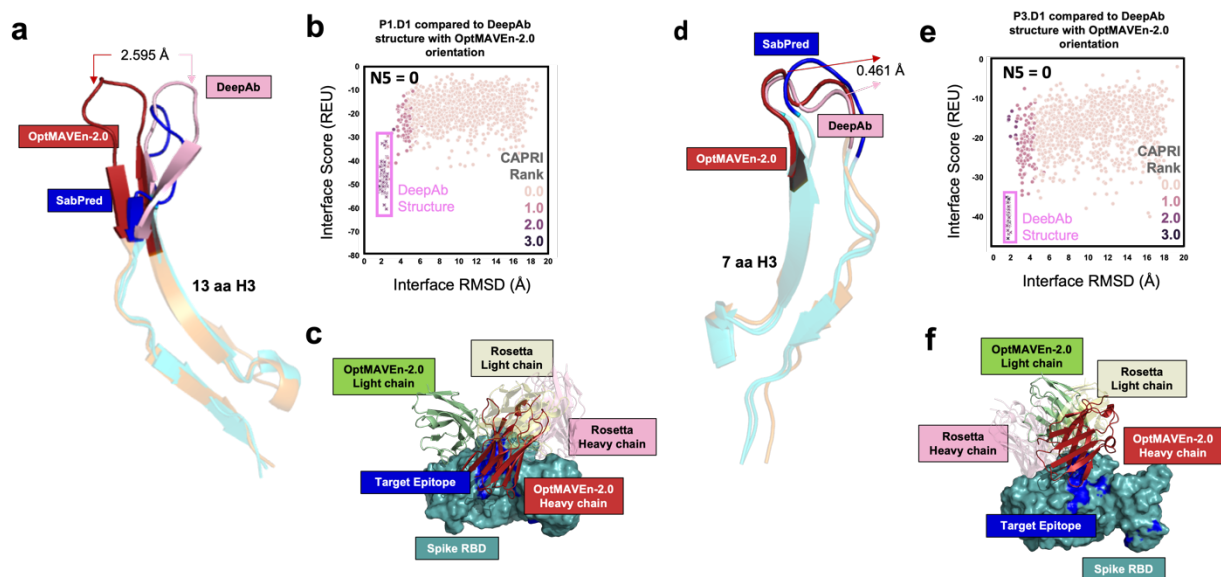
Figure 8. Computational evaluation of best binding designs by prediction of Fv structures. (a) P1.D1 H3 loop structures generated by OptMAVEn-2.0 (red), DeepAb (pink) and SAbPred (blue). (b) Plot of interface score versus interface RMSD (with respect to OptMAVEn-2.0 structure) from docking the DeepAb Fv model with the RBD. (c) P1.D1-RBD complex generated with OptMAVEn-2.0 along with the top ten lowest scoring decoys (yellow light chain, pink heavy chain) obtained from docking model (antibody chains of the docked decoys and OptMAVEn-2.0 complex are shown in transparent and solid cartoon representations respectively whereas the RBD is shown in surface representation). (d) Similar to a, featuring design P3.D1 (e) Similar to b, featuring design P3.D1 (f) Similar to c, featuring design P3.D1.

We also carried out an all-atom Molecular Dynamics (MD) simulation of the best binding design P1.D1 in complex with the RBD of the SARS-CoV-2 spike protein to assess the stability of the complex. Preliminary results for a 50ns trajectory counted an average of ~4 hydrogen bonds (sdt. dev: 1.5) present at the antibody-antigen interface (SI Fig. S6). This is quite encouraging, as in an earlier study,[58] MD simulation of the hACE2 receptor in complex with the spike protein RBD reported an average of only 2.7 hydrogen bonds at the interface. This implies that this design has the potential to competitively bind the RBD of the SARS-CoV-2 spike protein potentially sequestering it from hACE2. This is also corroborated by the Rosetta binding energy value of around -48.3 kcal/mol[49] calculated for the spike protein RBD with hACE2 (from PDB-id: 6lzg) which is weaker by over 7 kcal/mol compared to designs P1.D1 and P1.D2. Finally, it is important to stress that our designs rely on the accuracy of the Rosetta energy function to recapitulate experimental affinities and that carrying out experimental binding assays are needed to confirm or refute these findings. However, since protein folding requires MD simulations of micro-second timescales, they are unsuitable for validating if the antibody sequences fold as predicted by OptMAVEn in a high-throughput manner.

We hence used *DeepAb*[4], a deep learning method for predicting antibody structures from sequence[62] and *SAbPred*[63], which uses the *ABodyBuilder* algorithm[64,63] to generate structural models of the designed sequences. The models produced by these methods were aligned to the H3 loop of the OptMAVEn design using PyMOL (Fig. 8). For the P1.D1 design, there was a poor alignment amongst all three structures and the H3 loop RMSD between the OptMAVEn structure and the *DeepAb* structure was 2.6 Å. In contrast, the results for the P3.D1 design showed an excellent alignment between the *DeepAb* and OptMAVEn models, with an H3 RMSD of 0.5 Å, and a decent agreement with the *SAbPred model*. Thus, the designed sequence for P3.D1 is likely to fold into the intended structure, whereas the P1 designs may fold into alternative conformations, particularly in the CDR H3 loop.

Additionally, we re-docked the antibodies from DeepAb models to the RBD using *SnugDock* and generated funnel plots[61]. None of the DeepAb docking models resulted in converging energy funnel plots (Fig. 8b, 8d and Supplementary SI_1 Fig. S13). Furthermore, all the DeepAb docking models had higher interface binding energies compared to the corresponding OptMAVEn-2.0 structures (Supplementary SI_1 Fig. S14). Visual inspection of the low scoring models in PyMOL showed a variety of binding sites and orientations for the low energy models, suggesting that no clear energy minimum was found (see Fig. 8c and 8f).

Discussion

In summary, the goal of this computational analysis was to assess the range of possible antibody designs that can affect binding with the viral spike protein by interacting with residues involved in hACE2 binding. We reported on *de novo* prototype variable regions targeting the most solvent accessible seven-residue epitope in the spike RBD and their (computationally) affinity matured sequences with the lowest Rosetta binding energies. Designs were rank ordered not only in terms of their Rosetta binding energy but also their humanness score metric (H-score). We reported complete amino acid sequences for the 106 affinity matured designs as well as the five prototype sequences and V*, CDR3, and J* parts used. Importantly, we would like to note that high affinities of designed antibodies, as modeled using the Rosetta binding energy function, need not necessarily translate to therapeutic effectiveness. *De novo* antibody design methods require considerable screening of variants to isolate a high-affinity binder. Therefore, in addition to the OptMAVEn-2.0 protocol, we described a workflow for computational quality control and validation of designed sequences to be used before experimental testing. This workflow includes "forward folding" of the designed antibody sequences and docking of the antibodies to their antigen. These steps seek an energetic minimum of the antibody structure and the antibody-antigen complex. Designed structures and binding sites that agree with the predicted energetic minima are more promising candidates for experimental

testing. Other successful computational protein design studies targeting antigen epitopes have included forward folding and docking tests to prioritize candidates for testing.[73] This approach allows us to optimize the antibody design process by efficiently identifying promising antibody designs *in silico*. However, the evaluation method depends on the accuracy of the Rosetta energy function[47] but this limitation in accuracy is compensated for by the speed and low cost allowing for the evaluation of many potential binders simultaneously. Further, we compared the RBD residues that form non-covalent interactions with the antibody chains for five of our designs P1.D1, P2.D1, P3.D1, P4.D1 and P5.D1 with those of a set of 70 non-redundant anti-RBD antibodies taken from the CoV3D[78] database. Of these, we found only four antibodies with PDB-ids 6XKP, 7K8T, 7JX3 and 7CDJ that had around 35% common contacting RBD residues compared to our designs (Table 3, Appendix R). All other antibodies had lesser overlap of contacting residues indicating the novelty of our designs.

By analyzing our designed SARS-CoV2-RBD binders, we found that P1.D1 design binding sites were most efficiently recovered during local docking. More specifically, the P1.D1 design yielded the highest number of low energy structures which were within 1 Å of the original design. Similarly, the P1.D739 and P1.D544 models both yielded low energy docking structures within 1 Å of their respective designs. This suggests that the P1.D1, P1.D3, and P1.D4 designs are most likely to have an energetic minimum at their spike protein binding site and are the most promising for testing *in vitro*.

To probe the structural predictions of the OptMAVEn-2.0 designs, we generated antibody Fv structural models with *DeepAb* and *ABodyBuilder*. Since antibody structure prediction is generally reliable for non-H3 CDR loops and framework regions[4] we focused on the CDR H3 loop conformations. The P3 models produced with *DeepAb* had a CDR H3 loop conformation similar to the ones originally predicted by OptMAVEn-2.0. We were not, however, able to recover CDR H3 conformations of any other OptMAVEn-2.0 designs within 1 Å when using *DeepAb* (Fig. 7d). These results suggest that the P3 models are most likely to fold into the intended structure among the prototype designs tested. Conversely, we were unable to recover H3 conformations of the P3 models or any of the other OptMAVEn-2.0 designs with *ABodyBuilder*. The difference in the CDR H3 loop structures may arise from the uncertainty of the DeepAb prediction.[4] Alternately, since OptMAVEn-2.0 incorporates the target antigen before relaxing whereas *DeepAb* and *SAbPred* are blind to the antigen, the conformation difference could simply reflect the conformational change upon binding. Finally, the OptMAVEn-2.0 designed structures may deviate because it is overweighting the interfacial interactions relative to the antibody structure and stability, as most CDR H3 loops change little upon binding.[77]

Among the top five prototype designs presented in this work, we have not identified a design that is predicted to both fold into the designed structure and bind at the designed location when stratifying our results with *ABodyBuilder*, *DeepAb*, and Rosetta *SnugDock*. In agreement with this observation, we were not able to recover the designed complex structure for any of the top five prototype designs tested. These results suggest that additional designs may need to be screened to identify successful high-affinity RBD binding antibodies. The fact that we were able to isolate designs predicted to bind as intended and other designs predicted to fold as intended, suggests that our workflow for antibody analysis may be an effective screening tool for in silico assessment of computational antibody designs. In summary, our computational pipeline for design and validation of *de novo* antibody variable regions provides a time-efficient approach to shortlist promising antibody designs against choice epitopes which can be further tested experimentally. Our work also contributes to the ongoing development and improvement of antibody design algorithms and computational validation approaches.[65]

Methods

Antibody design in OptMAVEn-2.0

The initial antibody variable domain sequences were predicted using *de novo* antibody design software tool, OptMAVEn-2.0.[1] Using an interatomic clash-cutoff of 1.25 Å, 173 antigen poses were sampled, and each of which yielded a successful (not necessarily unique) antibody design targeted at the seven most solvent accessible hACE2-binding residues of SARS-CoV-2 spike RBD.

Prior to identifying antibody sequences complementary to the epitopes, OptMAVEn-2.0 first minimizes the z-coordinate of the epitopes, with their collective centroid set at origin, to allow the *de novo* designed antibody regions (see Supplementary SI_1 S1 for link to the entire MAPs fragment library) to bind from the bottom. Next, an ensemble of starting antigen poses is generated by a combination of discrete rotations (about the z-axis) and translations (in x, y, and z) – each of which are subsequently passed into the antibody design step. We started out with 3234 such antigen poses for the SARS-CoV-2 spike protein with the epitopes occupying the most negative z-axis coordinates.

Affinity maturation design in Rosetta

The affinity maturation protocol consisted of an initial refinement of the complex by RosettaDock[66] followed by three iterations of backbone perturbation using RosettaBackrub,[67] interface design using RosettaDesign[68] and rotamer repacking of the complex using a Monte Carlo based simulated annealing

algorithm.[69,70] During the Rosetta affinity maturation, only amino acids in the variable region within 5 Å from any epitope residue are allowed to mutate. Each affinity matured designed complex was relaxed using FastRelax (with constraints) 10 times and energy minimized (using *Minimize*). For each of these relaxed poses, the binding energy (dG_separated) was calculated using the *InterfaceAnalyzer*[48] application. The entire protocol was implemented in RosettaScripts[71] using the REF2015 energy function[47] (see Supplementary SI_1 S7). This computational protocol was executed for 8,000 affinity matured sequence-design cycles. The top five variable region designs which show strong interaction energy scores with the viral spike and low immunogenicity (high H-scores) were further investigated to glean insight on the biophysics of interactions at the residue level.

Antibody modeling

Structural models of antibodies were generated using *DeepAb*[4] and *ABodyBuilder*.[64] *DeepAb* is a machine learning-based method that utilizes a deep residual convolutional network (AbResNet) to predict Fv structure.[4] The antibody models were generated using the heavy and light chain sequences from the top OptMAVEn[1] designs as input. *ABodyBuilder* was similarly used to generate template-based models through the SAbPred webserver.[63]

Antibody-antigen docking

Models of antibodies in complex with the SARS-CoV-2 RBD were generated using *SnugDock*[2,3] to predict the local energy landscape around the designed binding site. ref2015 was used as the default energy function for all Rossetta based applications. The OptMAVEn designs were minimized into the Rosetta energy function as described.[72] The lowest scoring *DeepAb* model for each prototype was similarly docked using SnugDock. Docking consisted of a spin around the Ab-Ag center-of-mass axis with uniform sampling of degrees 0-360. Random perturbations of translations and rotations were sampled and applied from Gaussian distributions centered at 3 Å and 8°, respectively. Docking included refining the CDR H2 and H3 loops through kinematic loop closure and refining the VL-VH orientation through VL-VH docking in addition to antibody-antigen docking. 1000 models were generated for each docking run, of which the top ten decoys were chosen based on low Rosetta energy. All structures are numbered using the Chothia numbering scheme.

Author contributions

RC designed and performed the OptMAVEn experiments. VSB designed and performed the affinity maturation experiments, MD simulations and analyses of designs. RR, BA, and RF performed the structure prediction and docking methods. VSB, RC, CDM, RR, BA, and RF wrote the manuscript.

Acknowledgements

Conflict of Interest

Dr. Jeffrey J. Gray is an unpaid board member of the Rosetta Commons. Under institutional participation agreements between the University of Washington, acting on behalf of the Rosetta Commons, Johns Hopkins University may be entitled to a portion of revenue received on licensing Rosetta software including applications mentioned in this manuscript. As a member of the Scientific Advisory Board, Dr. Gray has a financial interest in Cyrus Biotechnology. Cyrus Biotechnology distributes the Rosetta software, which may include methods mentioned in this manuscript.

References

1. Chowdhury, R., Allan, M. F. & Maranas, C. D. OptMAVEn-2.0: De novo Design of Variable Antibody Regions Against Targeted Antigen Epitopes. *Antibodies* **7**, 23 (2018).
2. Sircar, A. & Gray, J. J. SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models. *PLOS Comput. Biol.* **6**, e1000644 (2010).
3. Jeliazkov, J. R., Frick, R., Zhou, J. & Gray, J. J. Robustification of RosettaAntibody and Rosetta SnugDock. *PLoS One* **16**, e0234282 (2021).
4. Ruffolo J. A., Sulam J., Gray J. J., Antibody structure prediction using interpretable deep learning. Patterns (New York, N.Y.) 3, 100406 (2022).

5.   Tao, K. *et al.* The biological and clinical significance of emerging SARS-CoV-2 variants. *Nat. Rev. Genet.* **22**, 757–773 (2021).

6.   Abdool Karim, S. S. & de Oliveira, T. New SARS-CoV-2 Variants — Clinical, Public Health, and Vaccine Implications. *N. Engl. J. Med.* **384**, 1866–1868 (2021).

7.   Chen, L., Xiong, J., Bao, L. & Shi, Y. Convalescent plasma as a potential therapy for COVID-19. *Lancet Infect. Dis.* **20**, 398–400 (2020).

8.   Bloch, E. M. *et al.* Deployment of convalescent plasma for the prevention and treatment of COVID-19. *J. Clin. Invest.* **130**, 2757–2765 (2020).

9.   Piccoli, L. *et al.* Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **183**, 1024-1042.e21 (2020).

10. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437–442 (2020).

11. Huang, Y. *et al.* Neutralizing antibodies against SARS-CoV-2: current understanding, challenge and perspective. *Antib. Ther.* **3**, 285–299 (2020).

12. Alejandra, T. M. *et al.* Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science (80-.).* **370**, 950–957 (2020).

13. Matthew, M. *et al.* Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. *Science (80-.).* **0**, eabl8506 (2021).

14. Atlani-Duault, L., Lina, B., Chauvin, F., Delfraissy, J.-F. & Malvy, D. Immune evasion means we need a new COVID-19 social contract. *Lancet Public Heal.* **6**, e199–e200 (2021).

15. Wang, P. *et al.* Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *bioRxiv* 2021.01.25.428137 (2021) doi:10.1101/2021.01.25.428137.

16. Pereira, F. SARS-CoV-2 variants lacking a functional ORF8 may reduce accuracy of serological testing. *J Immunol M ethods* vol. 488 112906 (2021).

17. Starr, T. N. *et al.* Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).

18. Weisblum, Y. *et al.* Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, (2020).

19. Andreano, E. *et al.* SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proc. Natl. Acad. Sci.* **118**, e2103154118 (2021).

20. Rogers, T. F. *et al.* Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science (80-.).* **369**, 956 LP – 963 (2020).

21. Maisonnasse, P. *et al.* COVA1-18 neutralizing antibody protects against SARS-CoV-2 in three preclinical models. *Nat. Commun.* **12**, 6097 (2021).

22. Shi, R. *et al.* A human neutralizing antibody targets the receptor binding site of SARS-CoV-2. *Nature* 1–5 (2020) doi:10.1038/s41586-020-2381-y.

23. Chen, X. *et al.* Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. *Cell. Mol. Immunol.* **17**, 647–649 (2020).

24. O'Brien, M. P. *et al.* Subcutaneous REGEN-COV Antibody Combination to Prevent Covid-19. *N. Engl. J. Med.* **385**, 1184–1195 (2021).

25. Cohen, M. S. *et al.* Effect of Bamlanivimab vs Placebo on Incidence of COVID-19 Among Residents and Staff of Skilled Nursing and Assisted Living Facilities: A Randomized Clinical Trial. *JAMA* **326**, 46–55 (2021).

26. Lundgren, J. D. *et al.* A Neutralizing Monoclonal Antibody for Hospitalized Patients with Covid-19. *N. Engl. J. Med.* **384**, 905–914 (2021).

27. Min, L. & Sun, Q. Antibodies and Vaccines Target RBD of SARS-CoV-2 . *Frontiers in Molecular Biosciences* vol. 8 247 (2021).

28. Jiang, S., Hillyer, C. & Du, L. Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. *Trends Immunol.* (2020) doi:10.1016/j.it.2020.03.007.

29. Desautels, T., Zemla, A., Lau, E., Franco, M. & Faissol, D. Rapid in silico design of antibodies targeting SARS-CoV-2 using machine learning and supercomputing. *bioRxiv* 2020.04.03.024885 (2020) doi:10.1101/2020.04.03.024885.

30. Magar, R., Yadav, P. & Barati Farimani, A. Potential neutralizing antibodies discovered for novel corona virus using machine learning. *Sci. Rep.* **11**, 5261 (2021).

31. Treewattanawong, W., Sitthiyotha, T. & Chunsrivirot, S. Computational redesign of Fab CC12.3 with substantially better predicted binding affinity to SARS-CoV-2 than human ACE2 receptor. *Sci. Rep.* **11**, 22202 (2021).

32. Riahi, S. *et al.* Application of an integrated computational antibody engineering platform to design SARS-CoV-2 neutralizers. *Antib. Ther.* **4**, 109–122 (2021).

33. Yang, J. *et al.* Computational design and modeling of nanobodies toward SARS-CoV-2 receptor binding domain. *Chem. Biol. Drug Des.* **98**, 1–18 (2021).

34. Cao, L. *et al.* De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science (80-.).* eabd9909 (2020) doi:10.1126/science.abd9909.

35. Valiente, P. A. *et al.* Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2. *J. Med. Chem.* **64**, 14955–14967 (2021).

36. Heinig, M. & Frishman, D. STRIDE: A web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500-2 (2004).

37. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 1–8 (2020) doi:10.1038/s41586-020-2179-y.

38. Lan, J. *et al.* Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *bioRxiv* 1–20 (2020) doi:10.1101/2020.02.19.956235.

39. Li, T., Pantazes, R. J. & Maranas, C. D. OptMAVEn – A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *PLoS One* **9**, e105954 (2014).

40. Entzminger, K. C. *et al.* De novo design of antibody complementarity determining regions binding a FLAG tetra-peptide. *Sci. Rep.* **7**, (2017).

41. Poosarla, V. G. *et al.* Computational de novo design of antibodies binding to a peptide with high affinity. *Biotechnol. Bioeng.* **114**, 1331–1342 (2017).

42. Tiller, K. E. *et al.* Facile affinity maturation of antibody variable domains using natural diversity mutagenesis. *Front. Immunol.* **8**, 986 (2017).

43. Pantazes, R. J. & Maranas, C. D. MAPs: A database of modular antibody parts for predicting tertiary structures and designing affinity matured antibodies. *BMC Bioinformatics* **14**, 168 (2013).

44. Lefranc, M.-P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **31**, 307–310 (2003).

45. IBM (2017) IBM ILOG CPLEX 12.7 User's Manual (IBM ILOG CPLEX Division, Incline Village, NV).1.

46. Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).

47. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

48. Benjamin Stranges, P. & Kuhlman, B. A comparison of successful and failed protein interface designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22**, 74–82 (2013).

49. Chowdhury, R. & Maranas, C. D. Biophysical characterization of the SARS-CoV2 spike protein binding with the ACE2 receptor explains increased COVID-19 pathogenesis. *bioRxiv* 2020.03.30.015891 (2020) doi:10.1101/2020.03.30.015891.

50. Lazar, G. A., Desjarlais, J. R., Jacinto, J., Karki, S. & Hammond, P. W. A molecular immunology approach to antibody humanization and functional optimization. *Mol. Immunol.* **44**, 1986–1998 (2007).

51. Ter Meulen, J. *et al.* Human monoclonal antibody combination against SARS coronavirus: Synergy and coverage of escape mutants. *PLoS Med.* **3**, 1071–1079 (2006).

52. Sui, J. *et al.* Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 2536–2541 (2004).

53. Walls, A. C. *et al.* Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **176**, 1026-1039.e15 (2019).

54. Zhu, Z. *et al.* Potent cross-reactive neutralization of SARS coronavirus isolates by human monoclonal antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12123–12128 (2007).

55. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science (80-.).* eabb7269 (2020) doi:10.1126/science.abb7269.

56. Wang, C. *et al.* A human monoclonal 1 antibody blocking SARS-CoV-2 infection. *bioRxiv* 2020.03.11.987958 (2020) doi:10.1101/2020.03.11.987958.

57. Wrapp, D. *et al.* Structural Basis for Potent Neutralization of Betacoronaviruses by Single-domain Camelid Antibodies. *Cell* 2020.03.26.010165 (2020) doi:10.1016/j.cell.2020.04.031.

58. Lupala, C. *et al.* Computational simulations reveal the binding dynamics between human ACE2 and the receptor binding domain of SARS-CoV-2 spike protein. *bioRxiv* 2020.03.24.005561 (2020) doi:10.1101/2020.03.24.005561.

59. Lensink, M. F. *et al.* Blind prediction of homo- and hetero-protein complexes: The CASP13-CAPRI experiment. *Proteins* **87**, 1200–1221 (2019).

60. Lensink, M. F., Nadzirin, N., Velankar, S. & Wodak, S. J. Modeling protein-protein, protein-peptide, and protein-oligosaccharide complexes: CAPRI 7th edition. *Proteins* **88**, 916–938 (2020).

61. Chaudhury, S. *et al.* Benchmarking and analysis of protein docking performance in Rosetta v3.2. *PLoS One* **6**, e22477 (2011).

62. Maguire, J. B., Grattarola, D., Mulligan, V. K., Klyshko, E. & Melo, H. XENet: Using a new graph convolution to accelerate the timeline for protein design on quantum computers. *PLOS Comput. Biol.* **17**, e1009037 (2021).

63. Dunbar, J. *et al.* SAbPred: a structure-based antibody prediction server. *Nucleic Acids Res.* **44**, W474–W478 (2016).

64. Leem, J., Dunbar, J., Georges, G., Shi, J. & Deane, C. M. ABodyBuilder: Automated antibody structure prediction with data-driven accuracy estimation. *MAbs* **8**, 1259–1268 (2016).

65. Schneider, C., Buchanan, A., Taddese, B. & Deane, C. M. DLAB: deep learning methods for structure-based virtual screening of antibodies. *Bioinformatics* btab660 (2021)

doi:10.1093/bioinformatics/btab660.

66. Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299 (2003).

67. Smith, C. A. & Kortemme, T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* **380**, 742–756 (2008).

68. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science (80-.).* **302**, 1364–1368 (2003).

69. Leaver-Fay, A., Kuhlman, B. & Snoeyink, J. An adaptive dynamic programming algorithm for the side chain placement problem. *Pac. Symp. Biocomput.* 16–27 (2005).

70. Leaver-Fay, A., Snoeyink, J. & Kuhlman, B. On-the-fly rotamer pair energy evaluation in protein design. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* vol. 4983 LNBI 343–354 (Springer, Berlin, Heidelberg, 2008).

71. Fleishman, S. J. *et al.* RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. *PLoS One* **6**, e20161 (2011).

72. Weitzner, B. D. *et al.* Modeling and docking of antibody structures with Rosetta. *Nat. Protoc.* **12**, 401–416 (2017).

73. Ovchinnikov, Sergey, and Po-Ssu Huang. Structure-based protein design with deep learning. *Curr O pin C hem B iol .* vol. 65 (2021).

74. Lazar, G.A. *et al.* A molecular immunology approach to antibody humanization and functional optimization. Mol. Immunol. 44, 1986–1998 (2007).

75. F. Juliette, *et al.*, Structural insights into the cross-neutralization of SARS-CoV and SARS-CoV-2 by the human monoclonal antibody 47D11. *Sci. Adv.* **7**, eabf5632 (2022).

76. Morales-Núñez J. J. *et al.*, Overview of Neutralizing Antibodies and Their Potential in COVID-19. *Vaccines* **9**, 1376 (2021).

77. Kuroda D., Gray J. J., Shape complementarity and hydrogen bond preferences in protein–protein interfaces: implications for antibody modeling and protein–protein docking. *Bioinformatics* **32**, 2451–2456 (2016).

78. Gowthaman R. *et al.,* CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Res*., 49(D1), D282–D287 (2021).