

1 ***De novo* design of high-affinity antibody variable regions (Fv) against the SARS-CoV-2**  
2 **spike protein**

3  
4 Veda Sheersh Boorla<sup>#</sup>, Ratul Chowdhury<sup>#</sup>, and Costas D. Maranas<sup>\*</sup>

5 Department of Chemical Engineering, The Pennsylvania State University, University Park. PA 16802

6  
7 <sup>\*</sup>*Corresponding author* | Email: [costas@psu.edu](mailto:costas@psu.edu)

8 <sup>#</sup> *Equal contribution*

9  
10  
11 **Abstract**

12 The emergence of SARS-CoV-2 is responsible for the pandemic of respiratory disease known as COVID-  
13 19, which emerged in the city of Wuhan, Hubei province, China in late 2019. Both vaccines and targeted  
14 therapeutics for treatment of this disease are currently lacking. Viral entry requires binding of the viral spike  
15 receptor binding domain (RBD) with the human angiotensin converting enzyme (hACE2). In an earlier  
16 paper<sup>1</sup>, we report on the specific residue interactions underpinning this event. Here we report on the *de novo*  
17 computational design of high affinity antibody variable regions through the recombination of VDJ genes  
18 targeting the most solvent-exposed hACE2-binding residues of the SARS-CoV-2 spike protein using the  
19 software tool OptMAVE<sup>n</sup>-2.0<sup>2</sup>. Subsequently, we carry out computational affinity maturation of the  
20 designed prototype variable regions through point mutations for improved binding with the target epitope.  
21 Immunogenicity was restricted by preferring designs that match sequences from a 9-mer library of “human  
22 antibodies” based on H-score (human string content, HSC)<sup>3</sup>. We generated 106 different designs and report  
23 in detail on the top five that trade-off the greatest affinity for the spike RBD epitope (quantified using the  
24 Rosetta binding energies) with low H-scores. By grafting the designed Heavy (VH) and Light (VL) chain  
25 variable regions onto a human framework (Fc), high-affinity and potentially neutralizing full-length  
26 monoclonal antibodies (mAb) can be constructed. Having a potent antibody that can recognize the viral  
27 spike protein with high affinity would be enabling for both the design of sensitive SARS-CoV-2 detection  
28 devices and for their deployment as therapeutic antibodies.

29  
30 **Main**

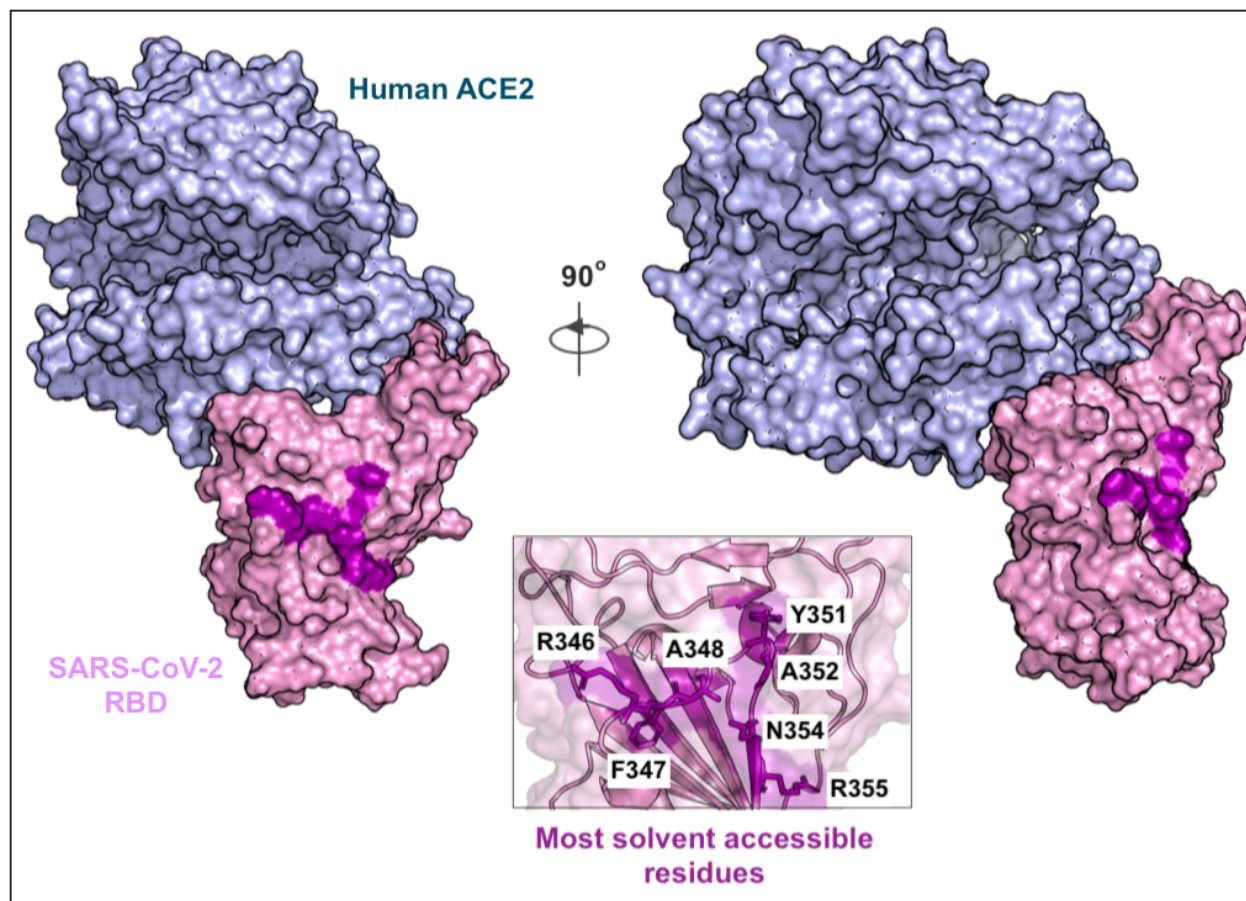
31 Over the last few weeks, several studies on using human or humanized antibodies targeted at the SARS-  
32 CoV-2 spike protein have been reported<sup>4,5,6,7</sup>. In addition, multiple efforts by laboratories and companies  
33 (Cellex, GeneTex etc.) for the development of antibody-based tests for SARS-CoV-2 detection are  
34 ongoing<sup>8</sup>. At the same time, significant progress towards the isolation and design of vaccines (mRNA-1273

35 vaccine © 2020 Moderna, Inc) and neutralizing antibodies<sup>9</sup> has been made. A computational study  
36 identified the structural basis for multi-epitope vaccines<sup>10,11</sup> whereas in another study, the glycosylation  
37 patterns of the spike SARS-CoV-2 protein were computationally deduced<sup>12</sup>. In addition, fully human single  
38 domain anti-SARS-CoV-2 antibodies with sub-nanomolar affinities were identified from a phage-displayed  
39 single-domain antibody library. Naïve CDRs were grafted to framework regions of an identified human  
40 germline IGHV allele using SARS-CoV-2 RBD and S1 protein as antigens<sup>13</sup>. In another study<sup>14</sup>, a human  
41 antibody 47D11 was identified to have cross neutralizing effect on SARS-CoV-2 by screening a library of  
42 SARS-CoV-1 antibodies. In two other studies, potent neutralizing antibodies were isolated from the sera of  
43 convalescent COVID-19 patients<sup>15,16</sup>. To the best of our knowledge, none of these neutralizing antibody  
44 sequences are publicly available. In a follow up effort<sup>17</sup>, human antibody CR3022 (which is neutralizing  
45 against SARS-CoV-1<sup>18</sup>) has been shown to bind to SARS-CoV-2 RBD in a cryptic epitope but without a  
46 neutralizing effect for SARS-CoV-2 *in vitro*. Desautels et al<sup>19</sup>, performed a machine learning based *in silico*  
47 mutagenesis of SARS-CoV-1 neutralizing antibodies to bind to SARS-CoV-2 spike protein. Walter et al<sup>20</sup>  
48 generated a number of synthetic nanobodies by *in vitro* screening large combinatorial libraries for binding  
49 to SARS-CoV-2 spike RBD<sup>20</sup>. However, studies that perform structure guided design of high affinity  
50 antibodies against specific epitopes of SARS-CoV-2 spike protein that may interfere with hACE2 binding  
51 are still lacking.

52  
53 Motivated by these shortcomings, here we explore the *de novo* design of antibody variable regions targeting  
54 the most solvent-exposed residues of the spike protein that are also part of the residue contact map involved  
55 in hACE2 binding, and trade-off binding energy against human sequence content in the variable region.  
56 The goal here is to exhaustively explore the sequence space of all possible variable region designs and  
57 report a range of diverse solutions that can serve as potentially neutralizing antibodies (nAb). We find that  
58 many different combinations of VDJ genes followed by mutation can yield potentially high affinity variable  
59 regions (scored using the Rosetta binding energy function) against an epitope of the spike protein RBD.  
60 Pareto optimal designs with respect to binding affinity vs. human content were drawn and five affinity  
61 matured designs are detailed in the results section.

62  
63 We first performed solvent accessibility analysis using the STRIDE<sup>21</sup> program on the 21 hACE2-binding  
64 residues of the SARS-CoV-2 spike protein (S-protein) RBD to define our binding epitope. The top seven  
65 residues with the highest solvent accessibility scores (i.e., SAS) are (Arg346, Phe347, Ala348, Tyr351,  
66 Ala352, Asn354, and Arg355) comprising our binding epitope (see Figure 1). Furthermore, the epitope is  
67 accessible for binding to RBD in the open confirmation of the full spike protein (See Supp. Fig. S8).

68



69  
70 **Figure 1.** The SARS-CoV-2 spike protein RBD in complex with Human ACE2 protein (PDB-id: 6LZG) is shown  
71 along with the most solvent accessible residues at the binding interface highlighted in purple. A zoomed view of these  
72 seven epitope residues is shown in the inset box. The numbering scheme for the S-protein residues is same as in PDB  
73 accession id 6LZG ([rcsb.org/structure/6LZG](https://rcsb.org/structure/6LZG) or 6VW1<sup>7</sup> and 6M0J<sup>6</sup>).  
74

75 We next used the previously developed OptMAVEN-2.0<sup>2</sup> software to computationally identify the  
76 combination of VDJ genes forming the variable region that best binds the desired epitope. OptMAVEN<sup>22</sup>  
77 has been used before successfully to design five high affinity CDRs against a FLAG tetrapeptide<sup>23</sup>, three  
78 thermally and conformationally stable antibody variable regions (sharing less than 75% sequence similarity  
79 to any naturally occurring antibody sequence) against a dodecapeptide mimic of carbohydrates<sup>24</sup> and two  
80 thermostable, high affinity variable heavy chain domains (V<sub>H</sub>H) against  $\alpha$ -synuclein peptide responsible  
81 for Parkinson's disease<sup>25</sup>. All these designs were experimentally constructed and nanomolar affinities for  
82 their respective target antigens was demonstrated.

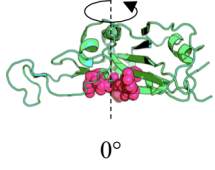
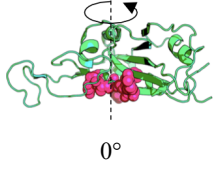
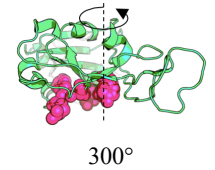
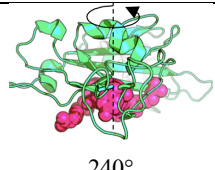
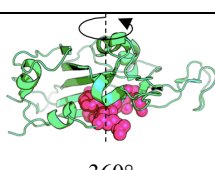
83  
84 Through a combination of rotations and translations, OptMAVEN-2.0 identified 3,234 unique antigen poses  
85 that presented the epitope to the antibody differently. The combinatorial space of different VDJ genes that  
86 upon recombination form the variable region of the prototype antibody was informed by the MAPs database

87 of antibody parts<sup>26</sup>. MAPs (see Supp. Info. S1 for link to full database) contains 929 modular antibody (i.e.,  
88 variable-V\*, complementarity determining -CDR3, and joining-J\*) parts from 1,168 human, humanized,  
89 chimeric, and mouse antibody structures (last updated in 2013). MAPs follows the antibody parts residue  
90 numbering convention as listed in the International iMmunoGeneTics (IMGT)<sup>27</sup> database. IMGT catalogs  
91 antibody parts as variable (V), diversity (D) and joining (J) structure libraries. MAPs stores complete CDR3  
92 parts, C-terminus-shortened V parts (i.e. V\* parts) and N-terminus-shortened J parts (J\* parts). Note that  
93 CDR3 includes the entire D gene and also up to the C-terminus of the V gene and up to the N-terminus of  
94 the J gene. In the remainder of the manuscript, the list of parts used to design the variable region are referred  
95 to as CDR3, V\* and J\* parts.

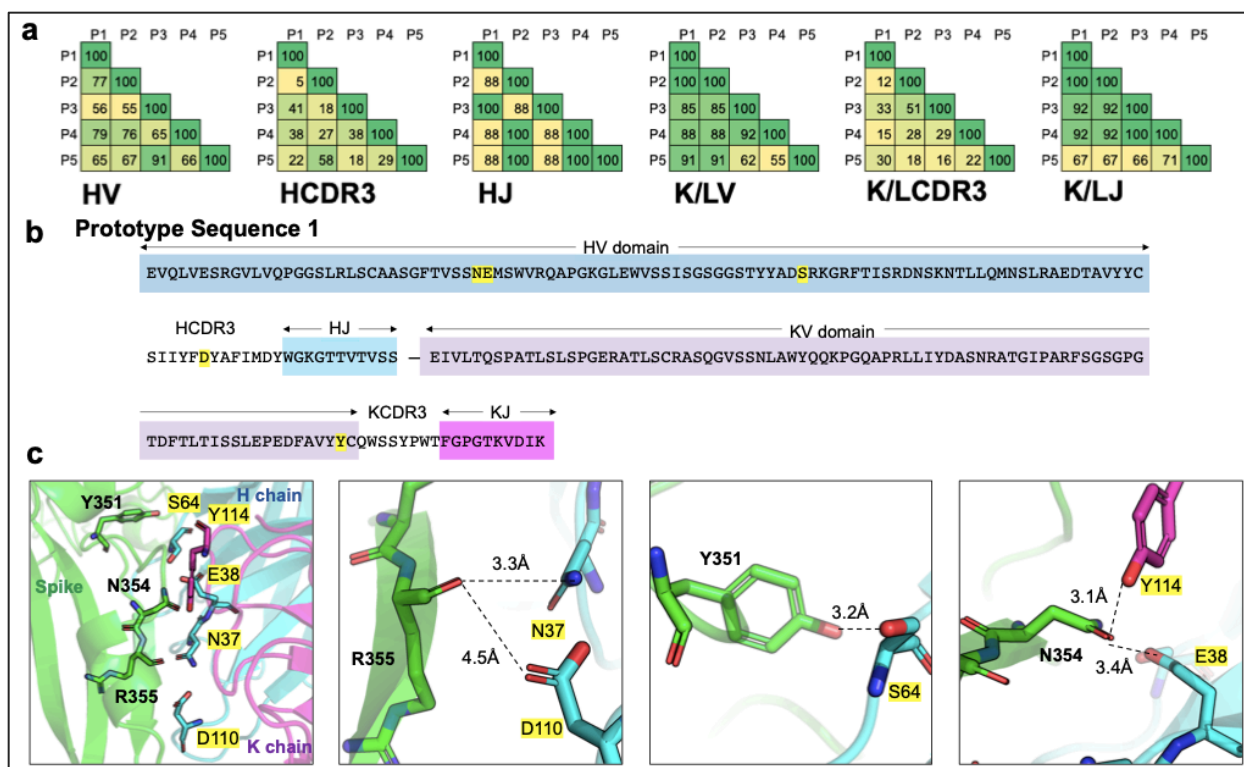
96  
97 For each one of the 3,234 spike poses, OptMAVEN-2.0 identified a variable region combination composed  
98 of end-to-end joined V\*, CDR3, and J\* region parts that minimized the Rosetta binding energy between  
99 the variable region and spike epitope formed by the seven residues. As part of OptMAVEN-2.0, the  
100 combinatorial optimization problem was posed and solved as a mixed-integer linear programming (MILP)  
101 problem using the cplex solver<sup>28</sup>. The solution of this problem identifies, for each one of the spike poses,  
102 the complete design of the variable region using parts denoted as HV\*, HCDR3, HJ\* for the heavy chain  
103 H and L/KV\*, L/KCDR3 and L/KJ\* for the light chain-L/K. Only 173 antigen-presenting poses out of  
104 3,234 explored, yielded non-clashing antigen-antibody complexes. These 173 poses were ranked on the  
105 basis of their Rosetta binding energies with the spike epitope and classified into 27 clusters (using *k-*  
106 *means*<sup>29</sup>) in a 19-dimensional space defined by quantitative descriptors of sequence similarity, three-  
107 dimensional spatial pose, and the angle at which they bind to the target epitope (see details in original  
108 paper<sup>2</sup>). The top five prototype designs with the highest Rosetta binding energies were present in four  
109 clusters and spanned a highly diverse set of choices of MAPs parts (see Table 1) with minimal conservation  
110 of the same part among the five prototype designs. The number entries in Table 1 correspond to the id of  
111 the gene in the MAPs database (which are identical to the ids used in IMGT). Note that design P5 uses a  
112 lambda (L) light chain instead of a kappa (K). Figure 1a plots the pairwise sequence similarity scores of the  
113 five antibody variable domains that were used in the top five designs. As expected, the top five prototype  
114 designs P1, P2, P3, P4, and P5 are the most dissimilar in their respective CDR3 domains in both light L,  
115 heavy H and HV\* domain (but not LV\*). They are the most similar in the choice of parts for the J\* domains  
116 (see Figure 2a) reflecting the lack of diversity among possible choices for the J\* domains in the MAPs  
117 database.

118  
119  
120

121 **Table 1.** V\*, CDR3, J\* gene ids for the top five prototype variable region designs and corresponding Rosetta  
 122 binding energies<sup>30,31</sup>. Antigen poses are described with the angle that the vertical axis through the epitope (shown in  
 123 pink) centroid and the C $\beta$  carbon of the residue with greatest z-axis coordinate forms.

Prototype design	Modular Antibody Parts number chosen in each design						Antigen pose (rotation of epitope about vertical axis)	Rosetta binding energy (kcal/mol)
	HV*	HCDR3	HJ*	L/KV*	L/KCDR3	L/KJ*		
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

124  
 125 Inspection of the interaction of design P1 with the spike epitope reveals strong electrostatic contacts  
 126 between the S-protein residues Tyr351, Asn354, and Arg355 (see Figure 1c) all of which have been deemed  
 127 important for hACE2 binding<sup>1</sup>. The strongest contacts with the three epitope residues are established by  
 128 five antibody residues spanning both the heavy and light chains (shown in yellow in Figure 2b). Spike  
 129 Tyr351 interacts with Ser64 in the HV\* domain, Asn354 interacts with Glu38 and Tyr114 in HV\* and KV\*  
 130 domains respectively, while spike Arg355 interacts with Asn37 and Asp110 of HV\* and HCDR3 domains,  
 131 respectively, in the stable spike-antibody complex (see Figure 2c).



132

133 **Figure 2. (a)** Pairwise sequence similarity percentages between the members of the six parts that were used to  
 134 construct the top five prototype variable regions with the lowest Rosetta binding energies with the viral spike epitope.  
 135 **(b)** The amino acid sequence of prototype design P1 with the different domain parts highlighted in different colors.  
 136 Spike epitope binding residues are highlighted in yellow. **(c)** Structural view of the strongest epitope-prototype  
 137 variable region interactions for P1. They imply strong electrostatic capture of three epitope residues by five variable  
 138 region residues spanning both heavy (H) and light (K) chains.

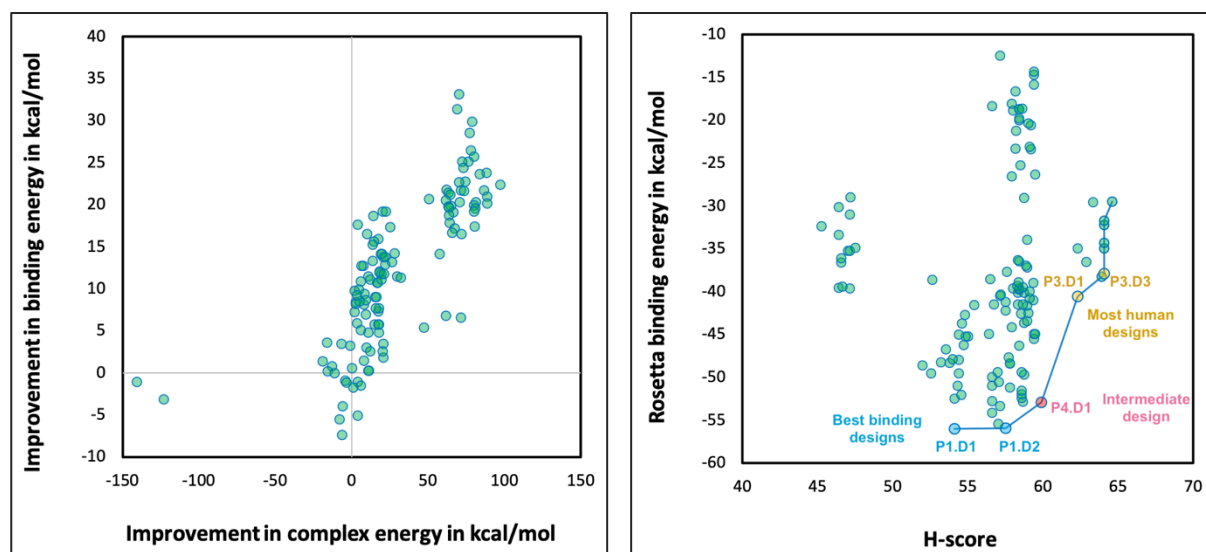
139

140 We next applied Rosetta-based *in silico* affinity maturation (see Methods) for each one of the top five  
 141 prototype designs shown in Table 1 to further enhance the non-covalent binding between the antibody  
 142 variable domains and the SARS-CoV-2 spike RBD. This computationally mimics the process of somatic  
 143 hypermutation leading to eventual affinity maturation of antibodies in B cells. This procedure identified a  
 144 total of 124 unique variable designs by introducing mutations in the five prototypes (see Figure 3a). We  
 145 retained 106 designs which achieved both an improvement in the Rosetta binding energy over their  
 146 respective prototype sequences and also further stabilization (i.e., lower overall Rosetta energy) of the  
 147 spike-antibody complex (see upper right quadrant of Figure 3a). On average, upon affinity maturation, the  
 148 binding energy was improved by ~14 kcal/mol and the overall energy was improved by ~37 kcal/mol.  
 149 Supplementary S2 lists first the starting prototype design (i.e., P1, P2, P3, P4 or P5) followed by the 106  
 150 affinity matured designs (labeled as P1.D1, P1.D2, etc). On average, there were 4.5 mutations (Supp. info.  
 151 S3) between computational affinity matured and prototype variable region designs.

152

153 We next assessed the departure of the 106 designed variable regions from fully-human antibody sequences  
154 using H-Score<sup>3</sup>. H-score is defined as the sum of the minimum edit distance between all possible 9-mer  
155 fragments in the designed variable region from a library of all 9-mer sequences observed in human  
156 antibodies<sup>22</sup>. The value of H-score is scaled to 100 and normalized by the length of the sequence. An  
157 antibody sequence with all 9-mers exactly matching 9-mers of human antibodies will have a perfect H-  
158 score of 100. Figure 3b illustrates the trade-off between the Rosetta binding energy vs. H-score for these  
159 affinity matured variable region designs. For comparison, we calculated the H-score for the human  
160 antibodies CR3022<sup>32</sup>, 80R<sup>33</sup>, S230<sup>34</sup> and M396<sup>35</sup> which are known to be neutralizing against SARS-CoV-  
161 1. These antibodies (including only Fv regions) have an average H-score of 62.98 (stdev: 4.9) which are in  
162 the same range as our most human designs (e.g., P3.D1 and P3.D3).

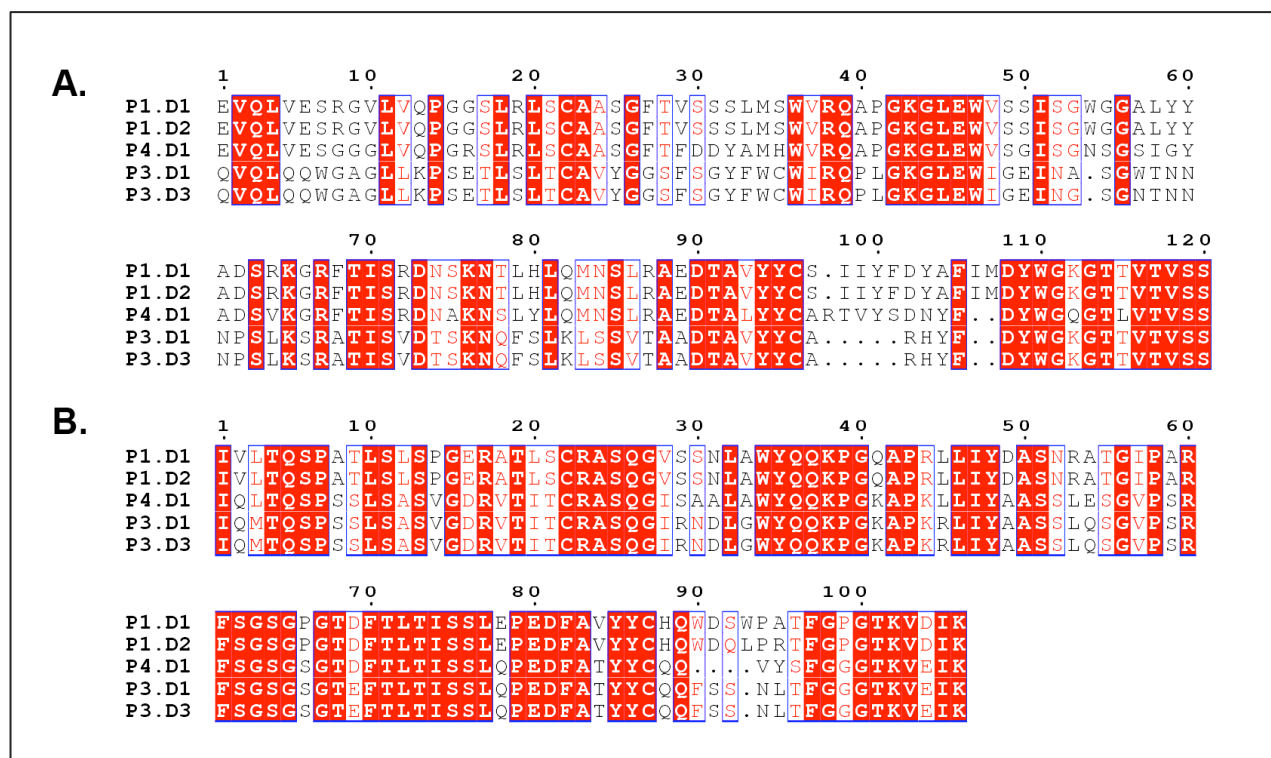
163  
164 We selected five designs that were on the Pareto optimum curve shown in Figure 3b. The Pareto optimum  
165 curve is defined as the collection of designs for which no other design exists that can simultaneously  
166 improve upon both criteria (i.e., Rosetta binding energy and H-score). Designs P1.D1 and P1.D2 shown in  
167 blue (in Fig. 3b) have the lowest Rosetta binding energies whereas P3.D1 and P3.D3 shown in yellow  
168 correspond to the ones with the highest H-scores. Design P4.D1 is an intermediate design that balances  
169 both binding energy and H-score. The lowest binding energy designs (P1.D1, P1.D2), irrespective of H-  
170 score, would be relevant in ELISA-based *in vitro* detection assays whereas the lowest H-score designs



**Figure 3.** (a) The 106 out of 124 Rosetta-affinity matured designs that improve upon both energy criteria fall in the top right quadrant. (b) Plot of the Rosetta binding energy vs. H-score for the 106 retained affinity-matured sequences. The blue line connects the designs on the Pareto optimum curve between these two design objectives. The three best-binding affinity matured designs (emerging from P1) – P1.D1 and P1.D2 are shown in blue whereas the two most human designs (emerging from P3) – P3.D1 and P3.D3 are shown in yellow. An intermediate design P4.D1 (emerging from P4) is shown in red.

171 (P3.D1, P3.D3) may offer the highest potential as therapeutic antibodies. In addition, we calculated the  
 172 Rosetta binding energy between the human CR3022<sup>17</sup> (anti-SARS-CoV-1 antibody) and the SARS-CoV-2  
 173 spike protein RBD using complex structure (PDB-id:6W41) to be -56.4 kcal/mol which is very close to the  
 174 Rosetta binding energy of designs P1.D1 and P1.D2. However, P1.D1 and P1.D2 bind a different epitope  
 175 on the spike RBD than the one that CR3022 targets (see Supp. Fig S7).

176  
 177 Figure 4 shows the sequence alignment of these five selected affinity matured sequences (i.e., P1.D1,  
 178 P1.D2, P4.D2, P3.D1, and P3.D3). Shown in red boxes are the conserved positions and in red font the  
 179 positions with different but of similar type residues. A total of 156 out of 226 aligned positions are  
 180 conserved among all designs. Table 2 lists the most important (strongest) contacts with the spike protein as  
 181 informed by an *in silico* alanine scanning (Supp. info. S4) on the spike-binding residues of the variable  
 182 region designs. In essence, the alanine scanning analysis identifies the loss in binding energy that is incurred  
 183 upon mutating each residue (one at a time) to alanine.



184 **Figure 4.** Sequence alignment of Heavy chain sequences (panel A) and light chain sequences (panel B) of the five  
 185 pareto optimal affinity matured sequences. Conserved positions are outlined in blue boxes. Positions with same  
 186 residues across designs are highlighted in red and positions with different amino acid but similar residue type are  
 187 highlighted in white with red text.

188  
 189 **Table 2.** List of important contacts between the spike protein epitope residues and residues of each of the selected  
 190 affinity matured designs. For each contact, the loss in binding energy upon mutation of antibody residue from the  
 191 interface to alanine is tabulated. The corresponding interacting spike residue is also shown.  
 192



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

193

194

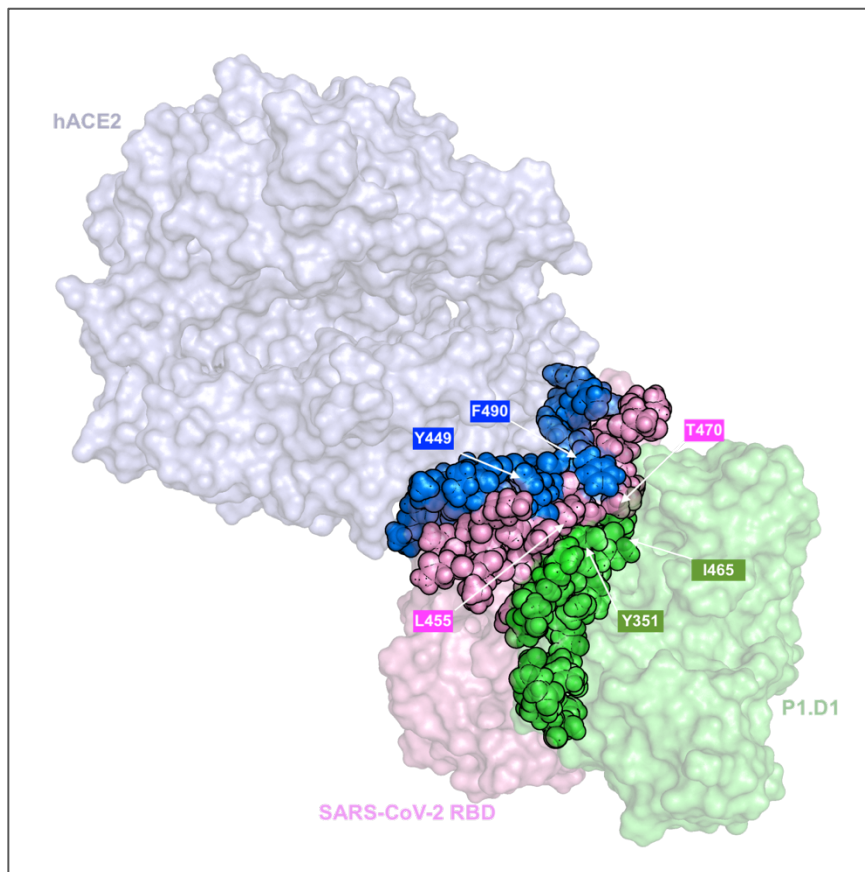
195 Antibodies that strongly bind to the RBD but do not inhibit hACE2 binding have been shown to be  
 196 neutralizing for SARS-CoV-2 (47D11<sup>14</sup>) and for SARS-CoV-1 (CR3022 in combination with CR3014<sup>32</sup>).

197 The mechanisms of neutralization of such antibodies are not completely known<sup>14</sup>. It is possible that upon  
 198 binding, these antibodies perturb the interaction network of the RBD with hACE2 thereby rendering RBD-  
 199 hACE2 binding less effective. In addition, Daniel et al<sup>36</sup> showed that nanobody VHH-72 raised against  
 200 SARS-CoV-1 had a neutralizing effect despite binding to an epitope that does not overlap with the hACE2  
 201 residue binding domain. By fusing VHH-72 with a human IgG1 they demonstrated SARS-CoV-2  
 202 neutralizing activity. They hypothesized that binding of the nanobody with the trimeric spike protein may  
 203 disrupt conformational dynamics and consequently prevent binding to hACE2.

204

205 In comparison, our design P1.D1 forms strong contacts (see Table 2) with many residues of the RBD which  
 206 in turn also indirectly interact with hACE2 (see Figure 5). For example, residues L455 and T470 of the  
 207 RBD are in contact with both hACE2 contacting RBD residues Y449, F490 and P1.D1 contacting RBD

208 residues Y351, I465. By perturbing the inter-residue interaction network of RBD-hACE2 it is plausible that  
209 a neutralizing effect can be achieved.



**Figure 5.** Space filling plot of the P1.D1-RBD complex superimposed on the hACE2-RBD complex. In blue are shown the RBD residues in contact with hACE2 and in green the RBD residues in contact with the P1.D1 designed antibody. In pink are shown all the RBD residues that are in contact with both hACE2 contacting RBD residues and P1.D1 contacting RBD residues. A few residues that are a part of these inter-residue contacts are labelled.

210  
211 We also carried out an all-atom Molecular Dynamics (MD) simulation of the best binding design P1.D1 in  
212 complex with the RBD of the SARS-CoV-2 spike protein to assess the stability of the complex. Preliminary  
213 results for a 50ns trajectory counted an average of ~4 hydrogen bonds (st. dev: 1.5) present at the antibody-  
214 antigen interface (Supp. Info. S5 for further details). This is quite encouraging, as in an earlier study<sup>37</sup>, MD  
215 simulation of the hACE2 receptor in complex with the spike protein RBD reported an average of only 2.7  
216 hydrogen bonds at the interface. This implies that this design has the potential to competitively bind the  
217 RBD of the SARS-CoV-2 spike protein potentially sequestering it from hACE2. This is also corroborated  
218 by the Rosetta binding energy value of around -48.3 kcal/mol<sup>1</sup> calculated for the spike protein RBD with  
219 hACE2 (from PDB-id: 6lzg) which is weaker by over 7 kcal/mol compared to designs P1.D1 and P1.D2.  
220 Finally, it is important to stress that our designs rely on the accuracy of the Rosetta energy function to

221 recapitulate experimental affinities and that carrying out experimental binding assays are needed to confirm  
222 or refute these findings.

223

## 224 **Summary**

225

226 In summary, the goal of this computational analysis was to assess the range of possible antibody designs  
227 that can affect binding with the viral spike protein by interacting with residues involved in hACE2 binding.

228 We reported on *de novo* prototype variable regions targeting the most solvent accessible seven-residue  
229 epitope in the spike and their (computationally) affinity matured sequences with the lowest Rosetta binding  
230 energies. Designs were rank ordered not only in terms of their Rosetta binding energy but also their  
231 humanness score metric H-score. We reported complete amino acid sequences for the 106 affinity matured  
232 designs as well as the five prototype sequences and V\*, CDR3, and J\* parts used. Importantly, we would  
233 like to note that high affinities of designed antibodies, as modeled using the Rosetta binding energy  
234 function, need not necessarily translate to therapeutic effectiveness. The exact mechanisms underlying the  
235 therapeutic action of monoclonal antibodies are quite complex and often only partially understood.  
236 Nevertheless, we hope that this analysis and data will contribute an important piece to help inform the  
237 discovery of high affinity mAb against SARS-CoV-2.

238

## 239 **Methods**

240

### 241 *Antibody design in OptMAVEN-2.0*

242

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

247

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

255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288

### *Affinity maturation design in Rosetta*

The affinity maturation protocol consisted of an initial refinement of the complex by RosettaDock<sup>38</sup> followed by three iterations of backbone perturbation using RosettaBackrub<sup>39</sup>, interface design using RosettaDesign<sup>40</sup> and rotamer repacking of the complex using a Monte Carlo based simulated annealing algorithm<sup>41,42</sup>. 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*<sup>31</sup> application. The entire protocol was implemented in RosettaScripts<sup>43</sup> using the REF2015 energy function<sup>30</sup> (see Supp. info. S6 for further details). 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.

### **Author contributions**

RC designed and performed the OptMAVEN experiments. VSB designed and performed the affinity maturation experiments, MD simulations and related analyses. VSB, RC and CDM wrote the manuscript.

### **Acknowledgements**

This activity was partially enabled by research conducted within the Center for Bioenergy Innovation of US Department of Energy (DE-SC0018420) and US National Science Foundation (NSF) grant CBET1703274.

All computations were performed using the High-Performance Computing facility at The Institute for Computational and Data Sciences Advanced Cyber Infrastructure (ICDS-ACI) at Pennsylvania State University.

## 289 References

290

- 291 1. Chowdhury, R. & Maranas, C. D. Biophysical characterization of the SARS-CoV2 spike protein  
292 binding with the ACE2 receptor explains increased COVID-19 pathogenesis. *bioRxiv*  
293 2020.03.30.015891 (2020) doi:10.1101/2020.03.30.015891.
- 294 2. Chowdhury, R., Allan, M. F. & Maranas, C. D. OptMAVEN-2.0: De novo Design of Variable  
295 Antibody Regions Against Targeted Antigen Epitopes. *Antibodies* **7**, 23 (2018).
- 296 3. Lazar, G. A., Desjarlais, J. R., Jacinto, J., Karki, S. & Hammond, P. W. A molecular immunology  
297 approach to antibody humanization and functional optimization. *Mol. Immunol.* **44**, 1986–1998  
298 (2007).
- 299 4. Yan, R. *et al.* Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2.  
300 *Science* **367**, 1444–1448 (2020).
- 301 5. Shang, J. *et al.* Structure of mouse coronavirus spike protein complexed with receptor reveals  
302 mechanism for viral entry. *PLoS Pathog.* **16**, e1008392 (2020).
- 303 6. Lan, J. *et al.* Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the  
304 ACE2 receptor. *bioRxiv* 1–20 (2020) doi:10.1101/2020.02.19.956235.
- 305 7. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 1–8 (2020)  
306 doi:10.1038/s41586-020-2179-y.
- 307 8. Petherick, A. Developing antibody tests for SARS-CoV-2. *Lancet* **395**, 1101–1102 (2020).
- 308 9. Jiang, S., Hillyer, C. & Du, L. Neutralizing Antibodies against SARS-CoV-2 and Other Human  
309 Coronaviruses. *Trends Immunol.* (2020) doi:10.1016/j.it.2020.03.007.
- 310 10. Saha, R. & Prasad, B. V. In silico approach for designing of a multi-epitope based vaccine against  
311 novel Coronavirus (SARS-COV-2). *bioRxiv* 2020.03.31.017459 (2020)  
312 doi:10.1101/2020.03.31.017459.
- 313 11. Srivastava, S. *et al.* Structural basis to design multi-epitope vaccines against Novel Coronavirus 19  
314 (COVID19) infection, the ongoing pandemic emergency: an in silico approach. *bioRxiv*  
315 2020.04.01.019299 (2020) doi:10.1101/2020.04.01.019299.
- 316 12. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O- glycosylation  
317 profile of the spike protein of novel coronavirus SARS-CoV-2. *bioRxiv* 2020.04.01.020966 (2020)  
318 doi:10.1101/2020.04.01.020966.
- 319 13. Wu, Y. *et al.* Fully human single-domain antibodies against SARS-CoV-2. *bioRxiv*  
320 2020.03.30.015990 (2020) doi:10.1101/2020.03.30.015990.
- 321 14. Wang, C. *et al.* A human monoclonal 1 antibody blocking SARS-CoV-2 infection. *bioRxiv*  
322 2020.03.11.987958 (2020) doi:10.1101/2020.03.11.987958.

- 323 15. Poh, C. M. *et al.* Potent neutralizing antibodies in the sera of convalescent COVID-19 patients are  
324 directed against conserved linear epitopes on the SARS-CoV-2 spike protein. *bioRxiv*  
325 2020.03.30.015461 (2020) doi:10.1101/2020.03.30.015461.
- 326 16. Ju, B. *et al.* Potent human neutralizing antibodies elicited by SARS-CoV-2 infection. *bioRxiv*  
327 2020.03.21.990770 (2020) doi:10.1101/2020.03.21.990770.
- 328 17. Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor-binding domains of SARS-  
329 CoV-2 and SARS-CoV. *Science (80-. )*. eabb7269 (2020) doi:10.1126/science.abb7269.
- 330 18. ter Meulen, J. *et al.* Human Monoclonal Antibody Combination against SARS Coronavirus:  
331 Synergy and Coverage of Escape Mutants. *PLoS Med.* **3**, e237 (2006).
- 332 19. Desautels, T., Zemla, A., Lau, E., Franco, M. & Faissol, D. Rapid in silico design of antibodies  
333 targeting SARS-CoV-2 using machine learning and supercomputing. *bioRxiv* 2020.04.03.024885  
334 (2020) doi:10.1101/2020.04.03.024885.
- 335 20. Walter, J. D. *et al.* Synthetic nanobodies targeting the SARS-CoV-2 receptor-binding domain.  
336 *bioRxiv* 2020.04.16.045419 (2020) doi:10.1101/2020.04.16.045419.
- 337 21. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known  
338 atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500-2 (2004).
- 339 22. Li, T., Pantazes, R. J. & Maranas, C. D. OptMAVEEn – A New Framework for the de novo Design  
340 of Antibody Variable Region Models Targeting Specific Antigen Epitopes. *PLoS One* **9**, e105954  
341 (2014).
- 342 23. Entzminger, K. C. *et al.* De novo design of antibody complementarity determining regions binding  
343 a FLAG tetra-peptide. *Sci. Rep.* **7**, (2017).
- 344 24. Poosarla, V. G. *et al.* Computational de novo design of antibodies binding to a peptide with high  
345 affinity. *Biotechnol. Bioeng.* **114**, 1331–1342 (2017).
- 346 25. Tiller, K. E. *et al.* Facile affinity maturation of antibody variable domains using natural diversity  
347 mutagenesis. *Front. Immunol.* **8**, 986 (2017).
- 348 26. Pantazes, R. J. & Maranas, C. D. MAPs: A database of modular antibody parts for predicting  
349 tertiary structures and designing affinity matured antibodies. *BMC Bioinformatics* **14**, 168 (2013).
- 350 27. Lefranc, M.-P. IMGT, the international ImMunoGeneTics database. *Nucleic Acids Res.* **31**, 307–  
351 310 (2003).
- 352 28. IBM (2017) IBM ILOG CPLEX 12.7 User's Manual (IBM ILOG CPLEX Division, Incline  
353 Village, NV).1.
- 354 29. Lloyd, S. P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
- 355 30. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and  
356 Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

- 357 31. Benjamin Stranges, P. & Kuhlman, B. A comparison of successful and failed protein interface  
358 designs highlights the challenges of designing buried hydrogen bonds. *Protein Sci.* **22**, 74–82  
359 (2013).
- 360 32. Ter Meulen, J. *et al.* Human monoclonal antibody combination against SARS coronavirus:  
361 Synergy and coverage of escape mutants. *PLoS Med.* **3**, 1071–1079 (2006).
- 362 33. Sui, J. *et al.* Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a  
363 human mAb to S1 protein that blocks receptor association. *Proc. Natl. Acad. Sci. U. S. A.* **101**,  
364 2536–2541 (2004).
- 365 34. Walls, A. C. *et al.* Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus  
366 Fusion. *Cell* **176**, 1026-1039.e15 (2019).
- 367 35. Zhu, Z. *et al.* Potent cross-reactive neutralization of SARS coronavirus isolates by human  
368 monoclonal antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12123–12128 (2007).
- 369 36. Wrapp, D. *et al.* Structural Basis for Potent Neutralization of Betacoronaviruses by Single-domain  
370 Camelid Antibodies. *Cell* 2020.03.26.010165 (2020) doi:10.1016/j.cell.2020.04.031.
- 371 37. Lupala, C. *et al.* Computational simulations reveal the binding dynamics between human ACE2  
372 and the receptor binding domain of SARS-CoV-2 spike protein. *bioRxiv* 2020.03.24.005561  
373 (2020) doi:10.1101/2020.03.24.005561.
- 374 38. Gray, J. J. *et al.* Protein-protein docking with simultaneous optimization of rigid-body  
375 displacement and side-chain conformations. *J. Mol. Biol.* **331**, 281–299 (2003).
- 376 39. Smith, C. A. & Kortemme, T. Backrub-Like Backbone Simulation Recapitulates Natural Protein  
377 Conformational Variability and Improves Mutant Side-Chain Prediction. *J. Mol. Biol.* **380**, 742–  
378 756 (2008).
- 379 40. Kuhlman, B. *et al.* Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*  
380 (80- ). **302**, 1364–1368 (2003).
- 381 41. Leaver-Fay, A., Kuhlman, B. & Snoeyink, J. An adaptive dynamic programming algorithm for the  
382 side chain placement problem. *Pac. Symp. Biocomput.* 16–27 (2005).
- 383 42. Leaver-Fay, A., Snoeyink, J. & Kuhlman, B. On-the-fly rotamer pair energy evaluation in protein  
384 design. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*  
385 *Intelligence and Lecture Notes in Bioinformatics)* vol. 4983 LNBI 343–354 (Springer, Berlin,  
386 Heidelberg, 2008).
- 387 43. Fleishman, S. J. *et al.* RosettaScripts: A Scripting Language Interface to the Rosetta  
388 Macromolecular Modeling Suite. *PLoS One* **6**, e20161 (2011).
- 389