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

Veda Sheersh Boorla[#], Ratul Chowdhury[#], and Costas D. Maranas[*]

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

[*]*Corresponding author* | Email: costas@psu.edu

[#] *Equal contribution*

Abstract

The emergence of SARS-CoV-2 is responsible for the pandemic of respiratory disease known as COVID-19, which emerged in the city of Wuhan, Hubei province, China in late 2019. Both vaccines and targeted therapeutics for treatment of this disease are currently lacking. Viral entry requires binding of the viral spike receptor binding domain (RBD) with the human angiotensin converting enzyme (ACE2). In an earlier paper[1], we report on the specific residue interactions underpinning this event. Here we report on the *de novo* computational design of high affinity antibody variable regions through the recombination of VDJ genes targeting the most solvent-exposed ACE2-binding residues of the SARS-CoV-2 spike protein using the software tool OptMAVEn-2.0[2]. Subsequently, we carry out computational affinity maturation of the designed prototype variable regions through point mutations for improved binding with the target epitope. Immunogenicity was restricted by preferring designs that match sequences from a 9-mer library of "human string content" (HSC)[3]. We generated 60 different variable region designs and report in detail on the top five that trade-off the greatest affinity for the spike epitope (quantified using the Rosetta binding energies) with low immunogenicity scores. By grafting these designed variable regions with frameworks, high-affinity monoclonal antibodies can be constructed. Having a potent antibody that can recognize the viral spike protein with high affinity would be enabling for both the design of sensitive SARS-CoV-2 detection devices and for their deployment as neutralizing antibodies.

Main

Over the last few weeks, several studies on using human or humanized antibodies targeted at the SARS-CoV-2 spike protein have been reported[4,5,6,7]. In addition, multiple efforts by laboratories and companies (Cellex, GeneTex etc.) for the development of antibody-based tests for SARS-CoV-2 detection are ongoing[8]. At the same time, significant progress towards the isolation and design of vaccines (mRNA-1273 vaccine © 2020 Moderna, Inc) and neutralizing antibodies[9] has been made. A computational study

identified the structural basis for multi-epitope vaccines[10,11] whereas in another study, the glycosylation patterns of the spike SARS-CoV-2 protein were computationally deduced[12]. In one study[13], fully human single domain anti-SARS-CoV-2 antibodies with sub-nanomolar affinities were identified from a phage-displayed single-domain antibody library by grafting naïve CDRs into framework regions of an identified human germline IGHV allele using SARS-CoV-2 RBD and S1 protein as antigens. In another study[14], a human antibody 47D11 was identified to have cross neutralizing effect on SARS-CoV-2 by screening through a library of SARS-Cov-1 antibodies. In two other studies, potent neutralizing antibodies were isolated from the sera of convalescent COVID-19 patients[15,16]. To the best of our knowledge, none of these neutralizing antibody sequences are publicly available. In another very encouraging study[17], human antibody CR3022 (which is neutralizing against SARS-CoV-1[18]) has been shown to bind to SARS-CoV-2 RBD in a cryptic epitope but without a neutralizing effect for SARS-CoV-2 *in vitro*. Moreover, we did not find any studies that performed guided design of high affinity antibodies against specific epitopes of SARS-CoV-2 proteins such as targeting the spike protein and subsequently prevent its binding with human ACE2.

Motivated by these shortcomings, here we explore the *de novo* design of antibody variable regions targeting most solvent-exposed residues of the spike protein that are involved in ACE2 binding, and trade-off binding energy against human sequence content in the variable region. The goal was to exhaustively explore the sequence space of all possible variable region designs and report a range of diverse solutions that can serve as neutralizing antibodies (nAb). We find that many different combinations of VDJ genes followed by mutation can yield potentially high affinity nAbs (scored using the Rosetta binding energy function) against the epitope of the spike protein involved with ACE2 binding. Pareto optimal designs with respect to binding affinity vs. human content were drawn and five choicest designs have been discussed in the results section.

We first performed solvent accessibility analysis using the STRIDE[19] program on the 21 ACE2-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).

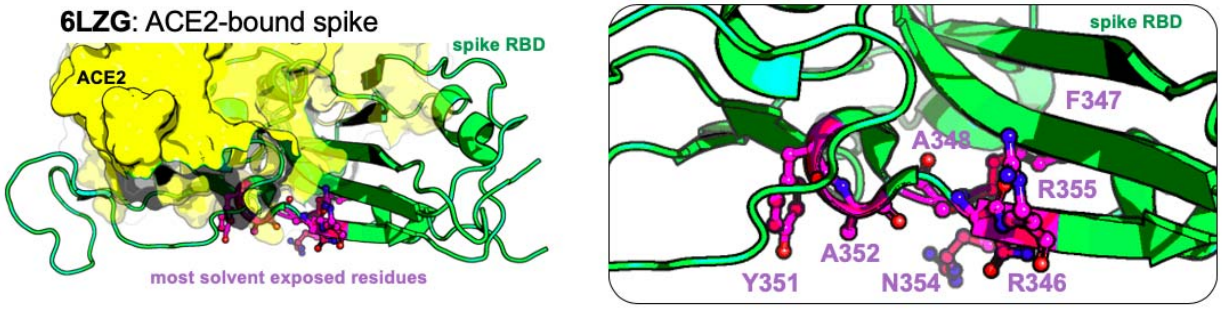


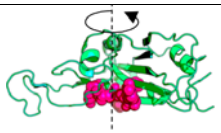
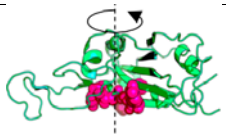
Figure 1. The ACE2-spike complex is shown along with the most solvent accessible residues at the binding interface highlighted in purple. These seven residues form the epitope for the variable region of the nAb. The numbering scheme for the S-protein residues matches the one used in PDB with accession id 6LZG (rcsb.org/structure/6LZG or 6VW1[7], and 6M0J[6]).

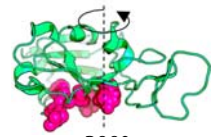
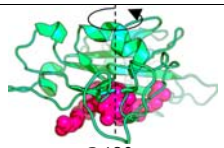
We next used the previously developed OptMAVEn-2.0[2] software to computationally identify the combination of VDJ genes forming the variable region that best binds the desired epitope. OptMAVEn[20] has been used before successfully to design five high affinity CDRs against a FLAG tetrapeptide[21] , three 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[22] and two thermostable, high affinity variable heavy chain domains (V_HH) against α-synuclein peptide responsible for Parkinson's disease[23]. 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[24]. MAPs (see Supp. Info. 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)[25] 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[26]. 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 on the basis of their Rosetta binding energies with the spike epitope and classified into 27 clusters (using *k-means*[27]) 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[2]). 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 design P5 uses a lambda (L) light chain instead of a kappa (K). Figure 1a 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* gene ids for the top five prototype variable region designs and corresponding Rosetta binding energies[28,29].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						**Antigen pose (rotation of epitope about vertical axis)**	**Rosetta binding energy (kcal/mol)**
Prototype design	**HV***	**HCDR3**	**HJ***	**L/KV***	**L/KCDR3**	**L/KJ***		
P1	82	315	5	61(K)	4(K)	3(K)	 0°	-41.19
P2	52	94	1	61(K)	17(K)	3(K)	 0°	-37.86

P3	105	12	5	6(K)	23(K)	4(K)	 300°	-34.93
P4	79	204	1	2(K)	1(K)	4(K)	 240°	-35.67
P5	108	212	1	37 (L)	5 (L)	5 (L)	360°	-44.31

Inspection of the interaction of design P1 with the spike epitope reveals strong electrostatic contacts between the S-protein residues Tyr351, Asn354, and Arg355 (see Figure 1c) all of which have been deemed important for ACE2 binding[1]. The strongest contacts with the three epitope residues are established by five antibody residues spanning both the heavy and light chains (shown in yellow in Figure 2b). Spike Tyr351 interacts with Ser64 in the HV* domain, Asn354 interacts with Glu38 and Tyr114 in HV* and KV* domains respectively, while spike Arg355 interacts with Asn37 and Asp110 of HV* and HCDR3 domains, respectively, in the stable spike-antibody complex (see Figure 2c).

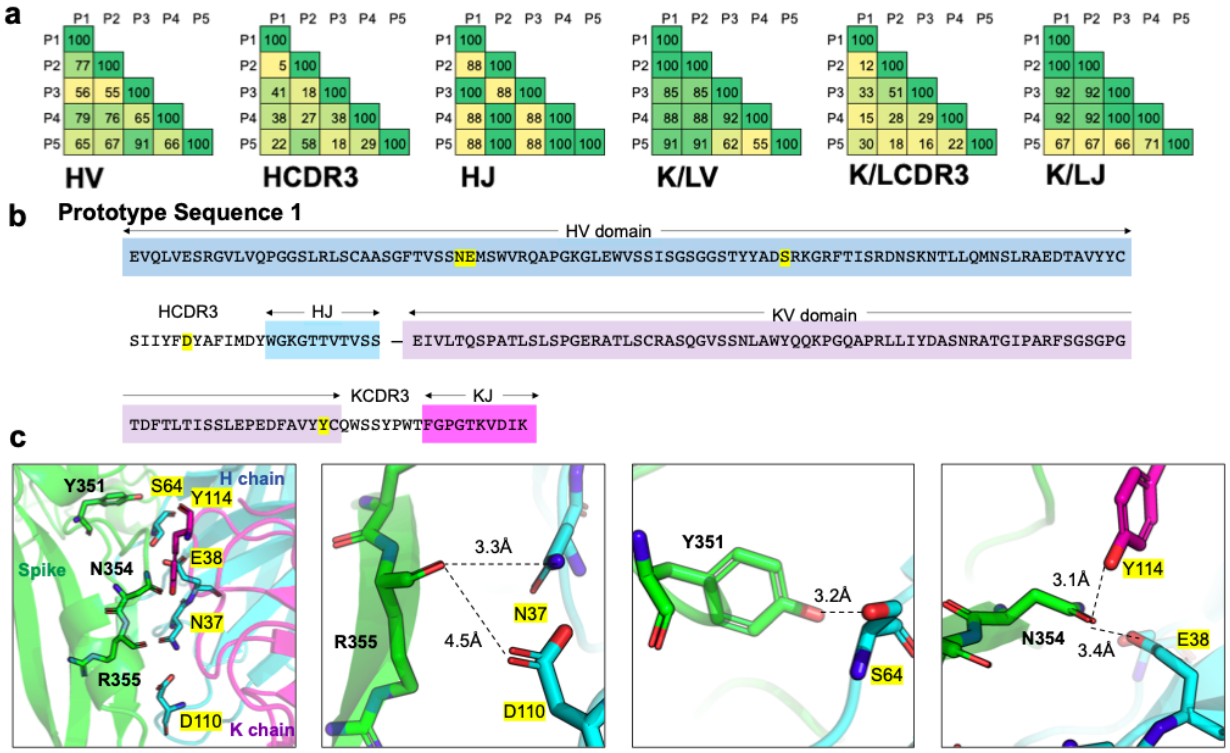


Figure 2. (a) Pairwise sequence similarity percentages between the members of the six parts that were used to construct the top five prototype variable regions with the lowest Rosetta binding energies with the viral spike epitope. **(b)** The amino acid sequence of prototype design P1 with the different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

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 enhance the non-covalent binding 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 3a). We retained the top 60 designs which achieved both an improvement in the Rosetta binding energy over their respective prototype sequences and also further stabilization (i.e., lower overall Rosetta energy) of the spike-antibody complex (see upper right quadrant of Figure 3a). On average, upon affinity maturation, the Rosetta binding energy was improved by ~7.35 kcal/mol and the Rosetta overall energy was improved by ~140 kcal/mol. Supplementary S2 lists the starting prototype design (i.e., P1, P2, P3, P4 or P5) and below it these 60 affinity matured designs (labeled as P5.D1, P5.D2, etc). On average, there were 4.5 mutations (Supp. info. S3) between computational affinity matured and prototype variable region designs. Unsurprisingly, P1 (with 14 affinity matured positions) gave rise to the most retained affinity matured sequences (i.e. 30 sequences) whereas P4 (with eight affinity matured positions) the fewest (i.e., two sequences).

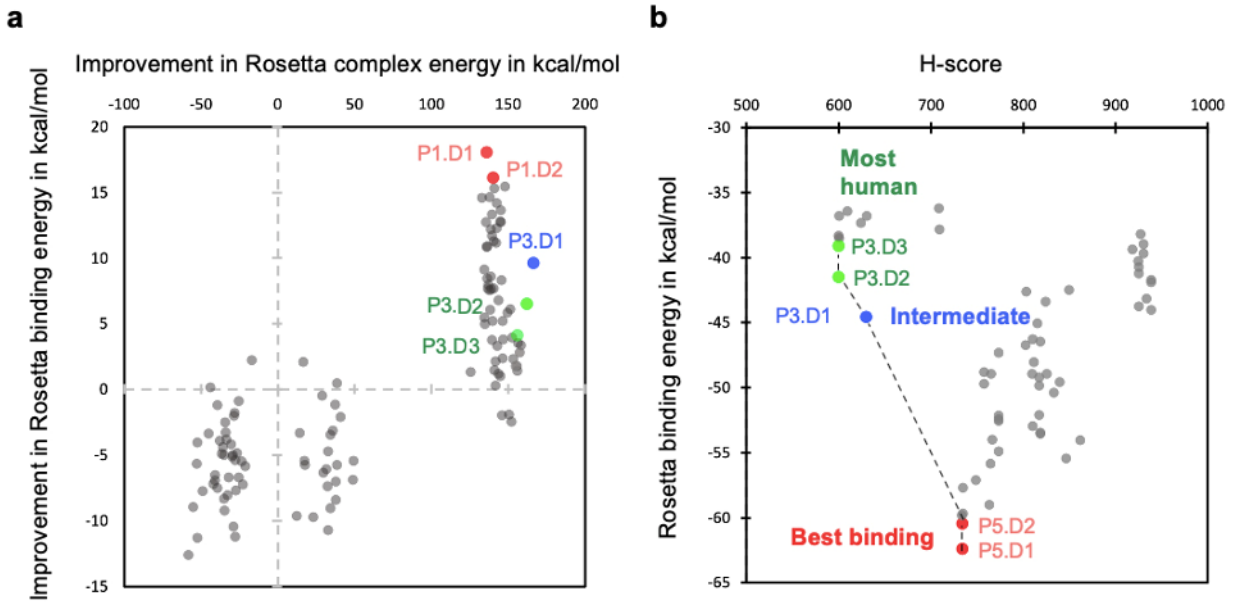


Figure 3. (a) The 60 out of 124 Rosetta-affinity matured designs that improve upon both energy criteria fall in the top right quadrant. **(b)** Plot of the Rosetta binding energy vs. H-score for the 60 retained affinity-matured sequences. The dotted line connects the designs on the Pareto optimum curve between these two design objectives. The two best-binding affinity matured designs (emerging from P5) – P5.D1 and P5.D2 are drawn in red whereas the two most human designs (emerging from P3) – P3.D2 and P3.D3 in green. Design P3.D1 with a balance between both criteria is shown in blue.

We then assessed the departure of the 60 designed variable regions from fully-human antibody sequences using H-Score[3]. 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 observed in human antibodies[20]. The smaller the value of H-score is, the closer the designed variable region to be fully human and thus less likely to trigger an immunogenic response. Figure 3b illustrates the trade-off between the Rosetta binding energy vs. H-score for the top 60 affinity matured variable region designs. Note that we calculated the H-score for the human CR3022[17] (anti-SARS-CoV-1 antibody) to be 667 which in the same range as our most human designs. We selected five designs that were on the Pareto optimum curve shown in Figure 3b. 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) at the same time. Designs P5.D1 and P5.D2 shown in red (in Fig. 3b) have the lowest Rosetta binding energies whereas P3.D2 and P3.D3 shown in green correspond to the ones with the lowest H-scores. Design P3.D1 balances Rosetta binding energy with low H-score. The lowest Rosetta energy designs (P5.D1, P5.D2), irrespective of H-score, would be relevant in ELISA-based *in vitro* detection assays whereas the lowest H-score designs (P3.D2, P3.D3) may offer the highest potential as neutralizing antibodies.

Figure 4 shows the sequence alignment of these five selected affinity matured sequences (i.e., P5.D1, P5.D2, P3.D1, P3.D2, and P3.D3). Shown in cyan are the positions with different residues and in yellow the positions with different but similar residues. A total of 137 out of 218 residues 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 (Supp. info. S4) 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.

```
1_P5.D1:      EVQLQESGPGLVKPSETLSLTCAVYGGSFSGYWWSWIRQPPGKGLEWIGQINHSGATMYN      60
2_P5.D2:      EVQLQESGPGLVKPSETLSLTCAVYGGSFSGYWWSWIRQPPGKGLEWIGMINHSGATMYN      60
3_P3.D1:      QVQLQQWGAGLLKPSETLSLTCAVYGGSFSGYFWCWIRQPLGKGLEWIGEINASGWTNNN      60
4_P3.D2:      QVQLQQWGAGLLKPSETLSLTCAVYGGSFSGYFWCWIRQPLGKGLEWIGEINHSGWTNNN      60
5_P3.D3:      QVQLQQWGAGLLKPSETLSLTCAVYGGSFSGYFWCWIRQPLGKGLEWIGEINHSGNTNNN      60

1_P5.D1:      PSLKSRITMSVDTSKNQFYLKLSSVTAADTAVYYCATLTGDLDAFDVWGQGTLVTVSSYE      120
2_P5.D2:      PSLKSRITMSVDTSKNQFYLKLSSVTAADTAVYYCATLTGDLDAFDVWGQGTLVTVSSYE      120
3_P3.D1:      PSLKSRATISVDTSKNQFSLKLSSVTAADTAVYYCAR-----HYFDYWGKGTTVTVSSIQ      115
4_P3.D2:      PSLKSRATISVDTSKNQFSLKLSSVTAADTAVYYCAR-----HYFDYWGKGTTVTVSSIQ      115
5_P3.D3:      PSLKSRATISVDTSKNQFSLKLSSVTAADTAVYYCAR-----HYFDYWGKGTTVTVSSIQ      115

1_P5.D1:      LTQ-PLSVSVALGQAARITCGGNNLGYKSVHWYQQKPGQAPVLVIYRDNNRPSGIPERFS      179
2_P5.D2:      LTQ-PLSVSVALGQAARITCGGNNLGYKSVHWYQQKPGQAPVLVIYRDNNRPSGIPERFS      179
3_P3.D1:      MTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKRLIYAASSLQSGVPSRFS      175
4_P3.D2:      MTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKRLIYAASSLQSGVPSRFS      175
5_P3.D3:      MTQSPSSLSASVGDRVTITCRASQGIRNDLGWYQQKPGKAPKRLIYAASSLQSGVPSRFS      175

1_P5.D1:      GSNSGNTATLTISRAQAGDEADYYCQSYDGSNVVFGSGTKVTVL  223
2_P5.D2:      GSNSGNTATLTISRAQAGDEADYYCQSYDGSNVVFGSGTKVTVL  223
3_P3.D1:      GSGSGTEFTLTISSLQPEDFATYYCQQFS-SNLTFGGGTKVEIK  218
4_P3.D2:      GSGSGTEFTLTISSLQPEDFATYYCQQFS-SNLTFGGGTKVEIK  218
5_P3.D3:      GSGSGTEFTLTISSLQPEDFATYYCQQFS-SNLTFGGGTKVEIK  218
```

Figure 4. Sequence alignment of top five pareto optimal affinity matured sequences. Variable positions are highlighted in cyan and variable though similar in residue type are highlighted in yellow.

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

Matured antibody Design id	Interface residue from antibody	Interacting spike residue	Change in binding energy upon mutation to alanine (kcal/mol)	Corresponding variable region
P5.D1	G109	R346	2.42	LCDR3
	M66	W353	1.14	HV*
	I56	C488	0.90	HV*
	W38	A352	0.89	HV*
	G63	F490	0.30	HV*
P5.D2	G109	R346	2.48	LCDR3
	M66	A348	1.01	HV*
	N57	W353	0.05	HV*
	T77	C488	0.12	HV*
	S59	R457	0.02	HV*
P3.D1	W64	A352	2.30	HV*
	N57	R355	0.65	HV*
	F107	T345	0.59	KCDR3
	S108	G446	0.44	KCDR3
	S29	E340	0.14	HV*
P3.D2	W64	A352, S349	2.52	HV*
	G36	E340	0.84	HV*
	F107	T345	0.77	KCDR3
	N57	N354	0.76	HV*
	F38	F347, A348	0.43	HV*
P3.D3	G36	C336, E340	1.65	HV*
	N57	N354	1.28	HV*
	F107	T345	0.69	KCDR3

	N64	N354	0.38	HV*
	S29	E340	0.14	HV*

Finally, to gain further insight into the biophysics of interactions, an all-atom Molecular Dynamics (MD) simulation of the best binding design P5.D1 in complex with the RBD of the SARS-CoV-2 spike protein was performed. Preliminary results for a 60ns trajectory counted an average of 4.7 hydrogen bonds (st. dev: 1.8) present at the antibody-antigen interface (Supp. Info. S5 for further details). This is quite encouraging, as in an earlier study[30], MD simulation of the ACE2 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 thus sequestering it from ACE2. In addition, we calculated the Rosetta binding energy between the human CR3022[17] (anti-SARS-CoV-1 antibody) and the SARS-CoV-2 spike protein RBD using complex structure (PDB-id:6W41) to be -48.83 kcal/mol and both our best binding designs P5.D1 and P5.D2 have better binding energies (~12 kcal/mol more negative) than CR3022. It is important to stress that our designs rely on the accuracy of the Rosetta energy function to recapitulate experimental affinities and that running experimental binding assays are needed to confirm or refute these findings.

Summary

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 occupying the residues involved in ACE2 binding. We reported on *de novo* prototype variable regions targeting the seven-residue epitope in the spike 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 all 60 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. The exact mechanisms underlying the therapeutic action of monoclonal antibodies are quite complex and often only partially understood. Nevertheless, we hope that this analysis and data will contribute an important piece to help inform the discovery of effective neutralizing mAb against SARS-CoV-2. We remain poised to help with any follow up computational tasks.

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[2]. 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 ACE2-binding residues of SARS-CoV-2 spike RBD.

Prior to identifying antibody sequences complimentary 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 Supp. Info. S1 for link to 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[31] followed by three iterations of backbone perturbation using RosettaBackrub[32], interface design using RosettaDesign[33] and rotamer repacking of the complex using a Monte Carlo based simulated annealing algorithm[34,35]. During the Rosetta affinity maturation, only amino acids in the variable region within 5 Å from any epitope residue are allowed to mutate. At the end of these affinity maturation iterations, the entire spike-antibody complex energy was minimized and the Rosetta binding energy was calculated using the *InterfaceAnalyzer*[29] application. The entire protocol was implemented in RosettaScripts[36] using the REF2015 energy function[28] (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.

Acknowledgements

We acknowledge funding sources: US National Science Foundation (NSF) grant CBET1703274 and Center for Bioenergy Innovation of US Department of Energy (DE-SC0018420).

References

1. 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.
2. 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).
3. 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).
4. Yan, R. *et al.* Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
5. Shang, J. *et al.* Structure of mouse coronavirus spike protein complexed with receptor reveals mechanism for viral entry. *PLoS Pathog.* **16**, e1008392 (2020).
6. 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.
7. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 1–8 (2020) doi:10.1038/s41586-020-2179-y.
8. Petherick, A. Developing antibody tests for SARS-CoV-2. *Lancet* **395**, 1101–1102 (2020).
9. 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.
10. Saha, R. & Prasad, B. V. In silico approach for designing of a multi-epitope based vaccine against novel Coronavirus (SARS-COV-2). *bioRxiv* 2020.03.31.017459 (2020) doi:10.1101/2020.03.31.017459.
11. Srivastava, S. *et al.* Structural basis to design multi-epitope vaccines against Novel Coronavirus 19 (COVID19) infection, the ongoing pandemic emergency: an in silico approach. *bioRxiv* 2020.04.01.019299 (2020) doi:10.1101/2020.04.01.019299.
12. Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O- glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *bioRxiv* 2020.04.01.020966 (2020) doi:10.1101/2020.04.01.020966.
13. Wu, Y. *et al.* Fully human single-domain antibodies against SARS-CoV-2. *bioRxiv* 2020.03.30.015990 (2020) doi:10.1101/2020.03.30.015990.
14. 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.

15. Poh, C. M. *et al.* Potent neutralizing antibodies in the sera of convalescent COVID-19 patients are directed against conserved linear epitopes on the SARS-CoV-2 spike protein. *bioRxiv* 2020.03.30.015461 (2020) doi:10.1101/2020.03.30.015461.

16. Ju, B. *et al.* Potent human neutralizing antibodies elicited by SARS-CoV-2 infection. *bioRxiv* 2020.03.21.990770 (2020) doi:10.1101/2020.03.21.990770.

17. 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.

18. ter Meulen, J. *et al.* Human Monoclonal Antibody Combination against SARS Coronavirus: Synergy and Coverage of Escape Mutants. *PLoS Med.* **3**, e237 (2006).

19. 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).

20. 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).

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

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

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

24. 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).

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

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

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

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

29. 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).

30. 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.

31. 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).

32. 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).

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

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

35. 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).

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