Design Strategy

Our goal was to design a binding peptide for ACE2 interface of the SARS-COV-2 spike protein RBD. Given the limited computation time available, we decided to focus on improving the single helix binding domain of ACE2, with the knowledge that an experimentally feasible mini-protein binder would likely require additional domains to be stable. Our design strategy, in summary, was as follows: 1) generate new backbones with Remodel, 2) optimize backbone design with FastDesign and Docking movers, and 3) explore further optimization of best scoring decoy with NCAA substitution.

Generating new backbones with Remodel

Our hypothesis was that by introducing loop secondary structure ("breaking the helix", so to speak) in the ACE2 binding domain, we would make it more flexible and thus better able to fit into the spike protein RBD interface. We used as the starting helix residues 23-46 of chain A of PDB 6M0J, as was done in Cao 2020. We generated 22 blueprints, specifying LLE at each possible segment of 3 residues. The other residues were set to helical redesign. 5 designs were generated from each blueprint, using the native helix in complex with the RBD as the PDB template in order to constrain the design to a structure that has shape complementarity with the original RBD interface. 12 designs across the blueprints that had the lowest total energy score were selected for optimization.

Optimization with FastDesign and Docking movers

Each of the 12 Remodel designs underwent further optimization with a round of the FastDesign and DockingProtocol Movers. The peptide interface was permitted to redesign during FastDesign and Docking; the interface of the RBD was restricted to repacking during both steps; and the non-interface residues were restricted to repacking during FastDesign and prevented from repacking during Docking. ShapeComplementarity, Ddg, and BuriedUnsatHBonds filters were applied only for evaluation (confidence set to 0). Inspection of the interaction domain in PyMol indicated that 8 A was a reasonable definition of interface residues, and a short design/docking run with interface residue definition set to 10 A did not produce noticeably different results. In addition to the 12 Remodel-designed backbones, we ran FastDesign/Docking on the native helix as a control in order to evaluate the efficacy of our Remodel method for generating better binders. The three filters were also evaluated for the original, native helix and PDB 7JZU (the mini-protein binder with the lowest dissociation constant from Cao 2020, LCB1, in complex with RBD).

Initially, we attempted three rounds of FastDesign/Docking, but this proved too time-consuming; comparing the results of one round versus three rounds from one remodel design did not reveal any noticeable differences in energy scores. Since there would be another round of optimization following this, we also sacrificed the number of trajectories explored for the sake of time; we only explored 20 trajectories per design, which yielded about 200-300 decoys per design. Docking was conducted at both centroid and full-atom levels.

In selecting a decoy for the next step, we considered interface energy score, total complex energy score, binding affinity (Ddg), number of buried unsatisfied hydrogen bonds, and shape complementarity. There was generally good correlation between the first three, such that the lowest interface score decoy for each backbone design usually also had the lowest or close to the lowest total score and binding affinity. We chose the lowest-scoring decoy for Remodel design 16_5 for further optimization, since it had the lowest interface score overall, close to lowest total score, and highest binding affinity (lowest Ddg).

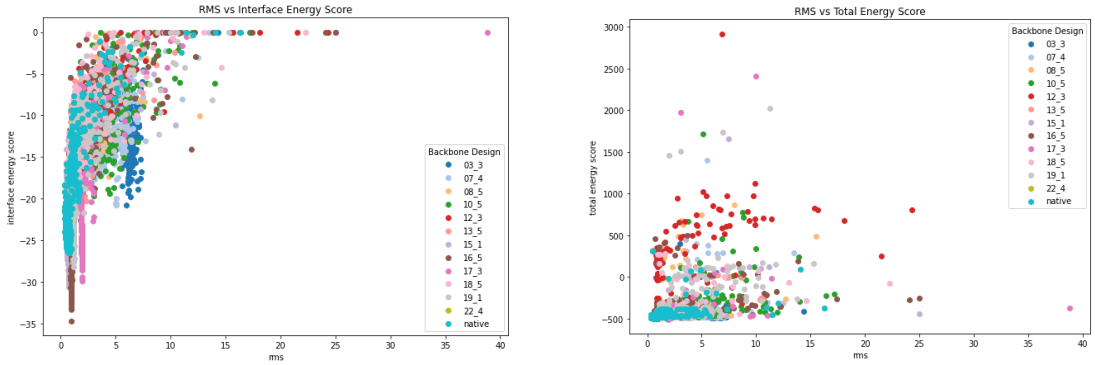
Optimization with NCAA substitution

The next step in our design process was to incorporate non-canonical amino acids into the structure to provide flexibility and possibly allow us to find a peptide with a lower interface score. The first step of this process was to randomly incorporate NCAAs into different positions along the helix and find the helix with lowest energy based on these iterations. The starting helix was design 16_5 for the reasons mentioned previously.

PyRosetta was used to run a Monte Carlo simulation to find the optimal sequence for the helix while randomly mutating residues to NCAAs. The NCAA rotamer library files were downloaded via a study that had previously incorporated NCAAs into Rosetta. A Rosetta MonteCarlo object was used for saving and recording the lowest pose during the search. The mover MutateResidue was implemented to mutate a residue at a random location along chain A (the helix). 10,000 iterations were applied, updating the pose after every 100 iterations to the lowest-energy configuration determined by the MonteCarlo object. After 10,000 iterations, a FastRelax was applied and the pdb was outputted. The script was run 7 times and conformers were generated that contained different combinations of residues of chain A mutated to NCAAs. Looking at the sequences in PyMol (see mutants.pse), four of the sequences converge to replacing NCAAs in only a few key positions, at residues 4, 7, 9, 19, 11, 21, and 24.

After NCAA substitution was performed with the Monte Carlo Pyrosetta script, the structures were used for further optimization using FastDesign/Docking, specifically using the mutant_helix_design_docking .xml script. This script differs from the remodeled_helix_docking_global_short.xml only in the DockingProtocol construction, where it is specified that 'docking_local_refine=true'. This is due to the fact that centroid-level parameter files are not included in Rosetta for NCAAs, only full-atom .params files exist.

Results



Figures 1 (left) and 2 (right). 1) Docking funnel plot depicting interface score versus RMS for the lowest interface score decoys of each backbone designed through Remodel, as well as for the native ACE2 helix backbone. **2)** Docking funnel plot depicting total score versus RMS for the lowest interface score decoys of each backbone designed through Remodel, as well as for the native ACE2 helix backbone.

Remodel backbone generation

The top 12 designs all had total energy scores (in complex with RBD) of around -420 (compared to -312 for the native complex after relaxing). The 12 designs were each from a different blueprint. All had very different amino acid sequences from the native helix. Not all successfully "broke" the helix with loops; some still appeared to be a continuous helix but had a coil segment instead of true helical secondary structure.

Design 16_5 appears to shift the helix closer to the RBD interface (see 16_5.pse). Design 03_3 appears to have identified an entirely different interaction site (see 03_3.pse). Other designs effectively shortened the helix length in proximity to the RBD interface by introducing a large kink, reminiscent of the shortened helix segments in LCB1 (as compared to the native ACE2 helix).

backbone	I_sc	score	Irms	rms	buns_filter	ddg_filter	sc_filter
03_3	-20.804	-460.461	2.542	5.953	19.0	-7.322	-0.002
07_4	-20.744	-476.925	3.153	5.041	20.0	-7.024	-0.007
08_5	-25.006	-396.323	0.476	0.865	23.0	-22.296	-0.003
10_5	-19.642	-465.830	0.375	2.992	21.0	-19.276	-0.005
12_3	-27.636	260.804	1.522	0.961	22.0	-21.733	-0.002
13_5	-29.403	-480.356	1.457	0.851	17.0	-17.232	-0.008
15_1	-30.570	-490.420	0.587	0.725	21.0	-24.295	-0.001
16_5	-34.629	-499.144	0.365	0.958	21.0	-29.067	-0.003
17_3	-29.831	-503.914	0.267	1.928	22.0	-27.377	-0.003
18_5	-20.619	-461.707	1.042	0.769	22.0	-14.909	-0.003
19_1	-28.202	-443.251	0.620	1.144	23.0	-20.455	-0.006
22_4	-26.508	-467.154	0.571	0.740	19.0	-23.378	-0.005
native (redesigned)	-25.136	-440.940	1.367	0.435	24.0	-11.011	-0.004
native (original)	N/A	-401.111	N/A	0.0	20.0	-13.018	-1.0
LCB1 (original)	N/A	-730.711	N/A	24.325	17.0	-63.931	0.004

Table 1. Decoys with the lowest interface score (I_sc, calculated by the DockingProtocol Mover), as well as filter values for the native ACE2 helix and LCB1 from Cao 2020.

FastDesign/Docking optimization

For each design, the decoys with the lowest interface score (I_sc, calculated by the DockingProtocol Mover) are listed in Table 1, along with comparisons to the filter values for the native helix and LCB1.

Docking funnel plots for each Remodel-designed backbone, as well as for the native ACE2 helix, are shown in Figures 1 and 2. While the plot of the total energy scores versus RMS (Fig. 2) does not display the characteristic funnel shape, interface energy score versus RMS (Fig. 1) does. Design 16_5 has the deepest funnel, and several designs have deeper funnels than the native ACE2 helix backbone. As previously mentioned, the lowest interface-score decoy for design 16_5 (decoy 16_5_10_10) had the lowest interface score among all decoys, close to the lowest total score, and the highest binding affinity (lowest Ddg). While the number of unsatisfied hydrogen bonds is somewhat high, it's only one more than the original, native helix-RBD complex. Shape complementarity was negative, but much higher than that of the native. The sequence of decoy 16_5_10_10 was not modified through FastDesign. We ran decoy 16_5_10_10 through trRosetta to predict if the designed sequence will fold into our desired structure. Results can be seen in Figure 3 and the Pymol session is entitled comparison.pse. It seems as though the decoy 16_5_10_10 will fold into a straight alpha-helix, but the desired design has a bend in the structure hypothesized to bring it closer to the RBD.

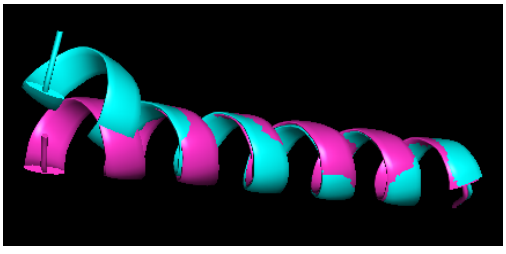


Figure 3. Superimposed designed structure 16_5 (cyan) on actual generated structure from trRosetta.

Optimization with NCAA Substitution

The Monte Carlo NCAA substitution using the Pyrosetta script generated structures with a total score of around -640 for each structure. The sequences for each simulation can be found in Table 1. Running the Docking xml script produced interface scores ranging from -43 to -47 and total scores from -620 to -650. Design 7 seemed to perform the best and had the least amount of buns (14) as seen in Table 2. Confirmation via trRosetta was unsuccessful as the program was not able to recognize the NCAA substitutions. For further confirmation we went with decoy 16_5.

Design	Sequence	Total Score	I_sc	BUNS Filter
1	B97-E-C41-A91-AE-C27-Q-A12-R-B74-AQTELKKGE-C27-EK-A82	-637	-47.234	19
2	EEEEAEKQ-A12-R-B74-AQTELKKGE-C27-EK-A83	-637	-46.486	20
3	EEEEAEKQFR-A94-AQTELKKGE-A83-EK-B28	-620	-43.229	15
4	EEE-A91-AE-C27-Q-A12-R-B74-AQTELKKGE-A07-EK	-633	-43.360	18
5	EEE-A91-AE-C27-Q-A12-R-B74-AQTELKKGE-B74-EK-B28	-638	-43.310	17
6	EEE-A91-AE-C27-Q-A12-R-B74-AQTELKKGE-A07-EK-A06	-636	-43.899	21
7	EEE-B74-AE-C27-Q-A12-R-B74-AQTELKKGE-B74-EK-B28	-642	-45.884	14

Table 2. NCAA substitution summary of metrics.

Comparison to existing binders

As seen in Table 1, the binding affinity of mini-protein binder LCB1 from Cao 2020 is much greater than our highest binding affinity design. This is likely mostly driven by the presence of two helix segments that interact with the RBD interface instead of just the one in our case. LCB1 is also likely much more stable as a result of stabilizing interactions between the three helices in the mini-protein bundle. We tried to predict interface hotspots with the KFC Server. It did predict more hotspots for LCB1-RBD than 16_5_10_10-RBD, but the accuracy of the prediction is doubtful; the server only correctly predicted one of the hotspots characterized for the native 6M0J structure.

References

1. Cao L, Goreshnik I, Coventry B, Case JB, Miller L, Kozodoy L, Chen RE, Carter L, Walls AC, Park YJ, Strauch EM, Stewart L, Diamond MS, Veesler D, Baker D. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science. 2020 Oct 23;370(6515):426-431. doi: 10.1126/science.abd9909. Epub 2020 Sep 9. PMID: 32907861; PMCID: PMC7857403.
2. Renfrew, P Douglas et al. “Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design.” PloS one vol. 7,3 (2012): e32637. doi:10.1371/journal.pone.0032637