1 **De novo design of picomolar SARS-CoV-2 miniprotein inhibitors**
2
3 Longxing Cao[1,2], Inna Goreshnik[1,2], Brian Coventry[1,2,3], James Brett Case[4], Lauren Miller[1,2], Lisa
4 Kozodoy[1,2], Rita E. Chen[4,5], Lauren Carter[1,2], Lexi Walls[1],Young-Jun Park[1], Lance Stewart[1,2],
5 Michael Diamond[4,6], David Veesler[1,], David Baker[1,2,7*]
6
7 1. Department of Biochemistry, University of Washington, Seattle, WA 98195
8 2. Institute for Protein Design, University of Washington, Seattle, WA 98195
9 3. Molecular Engineering Graduate Program, University of Washington, Seattle, WA 98195
10 4. Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110,
11 USA
12 5. Department of Pathology and Immunology, Washington University School of Medicine, St.
13 Louis, MO 63110, USA
14 6. The Andrew M. and Jane M. Bursky Center for Human Immunology & Immunotherapy
15 Programs, Washington University School of Medicine, St. Louis, MO 63110, USA
16 7. Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195
17 *Corresponding author. E-mail: dabaker@uw.edu
18
19
20 **Abstract**
21
22 We used two approaches to design proteins with shape and chemical complementarity to the
23 receptor binding domain (RBD) of SARS-CoV-2 Spike protein near the binding site for the
24 human ACE2 receptor. Scaffolds were built around an ACE2 helix that interacts with the RBD,
25 or de novo designed scaffolds were docked against the RBD to identify new binding modes. In
26 both cases, designed sequences were optimized first in silico and then experimentally for target
27 binding, folding and stability. Nine designs bound the RBD with affinities ranging from 100pM to
28 10nM, and blocked bona fide SARS-CoV-2 infection of Vero E6 cells with IC_{50} values ranging
29 from 35 pM to 35 nM; the most potent of these — 56 and 64 residue hyperstable proteins made
30 using the second approach — are roughly six times more potent on a per mass basis (IC_{50} ~
31 0.23 ng/ml) than the best monoclonal antibodies reported thus far. Cryo-electron microscopy
32 structures of the SARS-CoV-2 spike ectodomain trimer in complex with the two most potent
33 minibinders show that the structures of the designs and their binding interactions with the RBD
34 are nearly identical to the computational models, and that all three RBDs in a single Spike
35 protein can be engaged simultaneously. These hyperstable minibinders provide promising
36 starting points for new SARS-CoV-2 therapeutics, and illustrate the power of computational
37 protein design for rapidly generating potential therapeutic candidates against pandemic threats.
38
39 **Main text**
40
41 SARS-CoV-2 infection generally begins in the nasal cavity, with virus replicating there for
42 several days before spreading to the lower respiratory tract (*1*). Delivery of a high concentration
43 of a viral inhibitor into the nose and into the respiratory system generally might therefore provide
44 prophylactic protection and/or therapeutic benefit for treatment of early infection, and could be
45 particularly useful for healthcare workers and others coming into frequent contact with infected
46 individuals. A number of monoclonal antibodies are in development as systemic treatments for
47 COVID-19 (*2-6*), but these proteins are not ideal for intranasal delivery as antibodies are large
48 and often not extremely stable molecules and the density of binding sites is low (two per 150
49 KDa. antibody); antibody-dependent disease enhancement (*7-9*) is also a potential issue. High-
50 affinity Spike protein binders that block the interaction with the human cellular receptor
51 angiotensin-converting enzyme 2 (ACE2) (*10*) with enhanced stability and smaller sizes to

52 maximize the density of inhibitory domains could have advantages over antibodies for direct
53 delivery into the respiratory system through intranasal administration, nebulization or dry powder
54 aerosol. We found previously that intranasal delivery of small proteins designed to bind tightly
55 to the influenza hemagglutinin can provide both prophylactic and therapeutic protection in
56 rodent models of lethal influenza infection (*11*).
57
58 We set out to design high-affinity protein minibinders to the COVID-19 Spike RBD that compete
59 with ACE2 binding. We explored two strategies: first we incorporated the alpha-helix from ACE2
60 which makes the majority of the interactions with the RBD into small designed proteins that
61 make additional interactions with the RBD to attain higher affinity (**Fig 1A**). Second, we
62 designed binders completely from scratch without relying on known RBD-binding interactions
63 (**Fig 1B**). An advantage of the second approach is that the range of possibilities for design is
64 much larger, and so potentially a greater diversity of high-affinity binding modes can be
65 identified. For the first approach, we used the Rosetta blueprint builder to generate miniproteins
66 which incorporate the ACE2 helix (human ACE2 residues 23 to 46). For the second approach,
67 we used RIF docking (*12*) and design using large miniprotein libraries (*11*) to generate binders
68 to distinct regions of the RBD surface surrounding the ACE2 binding site (**Fig 1 and Fig S1**).
69
70 Large pools of designed minibinders (see Methods) made using the first and second
71 approaches, were encoded in long oligonucleotides and screened for binding to fluorescently
72 tagged RBD displayed on the surface of yeast cells. Deep sequencing identified three ACE2
73 helix scaffolded designs (*Approach 1*), and 105 de novo interface designs (*Approach 2*) that
74 were enriched following fluorescence activated cell sorting (FACS) for RBD binding. All three
75 ACE2-scaffolded designs and twelve of the *de novo* designs were expressed in *E. coli* and
76 purified. One of the ACE2-scaffolded designs expressed solubly and bound RBD with an affinity
77 of ~2 uM in biolayer interferometry (BLI) experiments (**Fig S2**). Eleven of the twelve *de novo*
78 designs were soluble and bound RBD with affinities ranging from 100nM to 2uM (**Fig S3**).
79 Affinity maturation of the ACE2-scaffolded design by PCR mutagenesis led to a variant, AHB1,
80 which bound RBD with an affinity of ~1 nM (**Fig 3A**) and blocked binding of ACE2 to the RBD
81 (**Fig S4**), consistent with the design model.
82
83 For 50 minibinders made using Approach 2, we generated site saturation mutagenesis libraries
84 (SSMs) in which every residue in each design was substituted with each of the 20 amino acids
85 one at a time. Deep sequencing before and after FACS sorting for RBD binding revealed that
86 residues at the binding interface and protein core were largely conserved in 40 out of the 50
87 cases (**Fig 2 and Fig S5**). For most of these minibinders, a small number of substitutions were
88 enriched in the FACS sorting; combinatorial libraries incorporating these substitutions were
89 constructed for the eight highest affinity designs and again screened for binding to the RBD at
90 concentrations down to 20pM. Each library converged on a small number of closely related
91 sequences; one of these was selected for each design (LCB1-LCB8) and found to bind the RBD
92 with high affinity on the yeast surface in a manner competed by ACE2 (**Fig 3 and Fig S6**).
93
94 LCB1–LCB8 were expressed, purified from *E. coli,* and binding to the RBD assessed by BLI.
95 For six of the designs, the K_D values ranged from 1–20 nM (**Fig 3 and Fig S6**), and for two
96 (LCB1 and LCB3), the K_D values were below 1 nM, which is too strong to measure reliably with
97 this technique (**Fig 3**). On the surface of yeast cells, LCB1 and LCB3 showed binding signals at
98 5 pM of RBD following protease (trypsin and chymotrypsin) treatment (**Fig S7**). Circular
99 dichroism spectra of the purified minibinders were consistent with the design models, and the
100 melting temperatures for most were greater than 90C (Figure 3 and FigS6). The designs
101 retained full binding activity after 14 days at room temperature (**Fig S8**).
102

103 We characterized the structures of LCB1 and LCB3 in complex with the SARS-CoV-2 spike
104 ectodomain trimer at 2.7 Å and 3.1 Å resolution, respectively, revealing stoichiometric binding of
105 the minibinders to each RBD within the spike trimer (**Fig 4 A-B, E-F and Fig S9, S11**). Although
106 the spike predominantly harbored two open RBDs for both complexes, we identified a subset of
107 particles with three RBDs open for the LCB3 complex (**Fig 4 A-B, E-F and Fig S9, S11**). We
108 subsequently improved the resolvability of the RBD/LCB1 and RBD/LCB3 densities using
109 focused classification and local refinement yielding maps at 3.1 and 3.5 Å resolution enabling
110 visualization of the interactions formed by each minibinder with the RBD (**Fig 4 C, G and Fig**
111 **S9-12)**.

113 LCB1 and LCB3 dock with opposite orientations in the crevice formed by the RBD receptor-
114 binding motif through extensive electrostatic interactions and shape complementarity mediated
115 by two out of the three minibinder α-helices (**Fig 4 C-D, G-H**). Similar to ACE2, the LCB1 and
116 LCB3 binding sites are buried in the closed S conformational state and require opening of at
117 least two RBDs to allow simultaneous recognition of the three binding sites (**Fig 4 A-B, E-F**).
118 Both LCB1 and LCB3 form numerous hydrogen bonds and salt bridges with the receptor-
119 binding motif resulting in the burial of a surface area of ~1,000Å^2 and ~800Å^2, respectively
120 (**Fig S10 and Fig S12**). The extensive networks of polar and hydrophobic interactions explain
121 the subnanomolar affinities of these inhibitors. As designed, the binding sites for LCB1 and
122 LCB3 overlap with that of ACE2, and hence should compete for binding to the RBD and inhibit
123 viral attachment at the host cell surface.

125 Superimposition of the designed LCB1/RBD or LCB3/RBD models to the corresponding cryoEM
126 structures, using the RBD as reference, show that the binding poses closely match the design
127 with backbone Cα rmsd of 1.27 Å and 1.9Å for LCB1 and LCB3, respectively (Fig. 4D, H) and
128 most of the polar interactions in the design models match nicely with the CryoEM structure (**Fig**
129 **S10, S12**). These data show that the computational design method can have quite high
130 accuracy. The structure comparisons in Fig. 4D and 4H are to the original design models; the
131 substitutions that increased binding affinity are quite subtle and have very little effect on
132 backbone geometry.

134 We investigated the capacity of AHB1 and LCB1-5 to prevent infection of human cells by *bona*
135 *fide* SARS-CoV-2. Varying concentrations of minibinders were incubated with 100 focus-forming
136 units (FFU) of SARS-CoV-2 and then added to Vero E6 monolayers. AHB1 strongly neutralized
137 SARS-CoV-2 (IC_{50} of 35nM), whereas a control influenza minibinder showed no neutralization
138 activity (**Fig 5A**). Next, we tested the Approach 2 designed minibinders LCB1-5. We observed
139 even more potent neutralization of SARS-CoV-2 by LCB1 and LCB3 with IC_{50} values of 34.46pM
140 and 48.1pM, respectively (**Fig 5B).** On a molar basis, these values are approximately 6-fold
141 lower than the most potent anti-SARS-CoV-2 monoclonal antibody described to date (*13*); on a
142 mass basis, because of their very small size, the designs are more potent than any of these
143 antibodies.

145 The minibinders designed in this work have potential advantages over antibodies as potential
146 therapeutics. Together, they span a range of binding modes, and in combination viral mutational
147 escape would be quite unlikely (**Fig S1**). The retention of activity after extended time at elevated
148 temperatures suggests they might not require a cold chain. The designs are 20-fold smaller than
149 a full antibody molecule, and hence in an equal mass have 20-fold more potential neutralizing
150 sites, increasing the potential efficacy of a locally administered drug. The cost of goods and the
151 ability to scale to very high production should be lower for the much simpler miniproteins, which
152 unlike antibodies, do not require expression in mammalian cells for proper folding. The small
153 size and high stability should also make them amenable to formulation in a gel for nasal

154 application, and to direct delivery into the respiratory system by nebulization or as a dry powder.
155 We will be exploring alternative routes of delivery in the months ahead as we seek to translate
156 the high potency neutralizing proteins into SARS-Cov2 therapeutics and prophylactics.
157 Immunogenicity is a potential problem with any foreign molecule, but for previously
158 characterized small de novo designed proteins little or no immune response has been observed
159 (*11, 14*), perhaps because the high solubility and stability together with the small size makes
160 presentation on dendritic cells less likely.
161
162 Timing is critical in a pandemic outbreak: potent therapeutics are needed in as short a time as
163 possible. We began to design minibinders in January 2020 based on a Rosetta model of the
164 SARS-CoV-2 Spike structure and switched to the crystal structures once they became available.
165 By the end of May 2020, we had identified very potent neutralizers of infectious virus; during
166 this same time, a number of neutralizing monoclonal antibodies were identified. We believe that
167 with continued development, the computational design approach can become much faster. First,
168 as structure prediction methods continue to increase in accuracy, target models suitable for
169 design could be generated within a day of determining the genome sequence of a new
170 pathogen. Second, with continued improvement in computational design methods, it should be
171 possible to streamline the workflow described here, which required screening of large sets of
172 computational designs, followed by experimental optimization, to identify very high affinity
173 binders. The very close agreement of the cryoEM structures of LCB1 and LCB2 with the
174 computational design models suggest that the main challenges to overcome are not in the de
175 novo design of proteins with shape and chemical complementarity to the target surface, but in
176 recognizing the best candidates and identifying a small number of affinity increasing
177 substitutions. The large amount of data collected in protein interface design experiments such
178 as those described here should inform the improvement of the detailed atomic models at the
179 core of Rosetta design calculations, as well as complementary machine learning approaches,
180 to enable recognition and improved sequence design of the best candidates; this would enable
181 the rapid in silico design of pM inhibitors like LCB1 and LCB3. With continued methods
182 development, we believe that it will become possible to generate ultra-high-affinity, pathogen
183 neutralizing designs within weeks of obtaining genome sequence. Preparing against unknown
184 future pandemics is difficult, and such a capability could be an important component of a
185 general response strategy.
186
187 **Acknowledgements**
188
202
203
204

205
Material and Methods

206

207

Designing miniprotein binders based on the ACE2 helix

208

209

The crystal structure (PDB: 6M0J) of the RBD bound to ACE2 (*4*) was refined with the Rosetta
FastRelax protocol with coordinate constraints, and the main binding helix of ACE2 (residue 23 -
46) was extracted and used as the starting point for miniprotein generation. Miniprotein binder
generation began by defining a variety of 3-helical bundles using the RosettaRemodel blueprint
(*15*)} format with the requirement that the ACE2 helix was incorporated into one of the helices.
1,974 blueprints that differed in the lengths of the helices and the loop types were generated.
The blueprints were used to generate backbones using the Rosetta Monte Carlo-based
fragment assembly protocol (*16*), while keeping the ACE2 helix fixed. Specially, the RBD
structure was loaded into a hashing grid for fast clash checking to guarantee the newly
generated backbones are not clashing with the RBD. The generated miniprotein was subjected
to a 3-stage sequence design protocol. Firstly, only interface residues on the miniproteins were
designed using the PackRotamersMover to quickly select designs that could make extra
contacts with RBD basides of the ACE2 helix. The whole miniprotein except the key interacting
residues on the ACE2 helix were then designed using the FastDesign protocol, activating
between side-chain rotamer optimization and gradient-descent-based energy minimization.
Lastly, the interface residues were further optimized with another round of FastDesign to
improve shape and chemically complementary. The monomeric metrics and the interface
metrics were calculated using Rosetta. Top 5,000 designs were selected based on the protein
folding metrics (score_pre_res, buried non-polar area, local structure and sequence agreement)
and interface metrics (ddg, shape complementarity, interface buried solvent accessible area).
Both the quality of the designed monomeric structure and the accuracy of designed interface
interactions are important for successfully making miniprotein binders, which is hard to achieve
through one Monte Carlo fragment assembly trajectory. Thus, we performed a second-round
design based on the above generated design models. In detail, 273 helix hairpins that make
good contacts with the RBD are selected based on the above-mentioned interface metrics.
Starting from these helix hairpins, the third helix was built using the same fragment assembly
protocol. Sequence design was performed on the generated backbone, except the first stage
sequence design was omitted. The outputs were filtered by protein folding metrics and interface
binding metrics, and 18,000 designs were selected.

239

De novo binder design

240

241

The crystal structure (PDB: 6M0J) of the RBD bound to ACE2 (*4*) was refined with the Rosetta
FastRelax protocol with coordinate constraints, and the RBD subunit was extracted for
subsequent steps. Initial docking conformations were generated by RifDock (*12*). Briefly, billions
of individual disembodied amino acids were docked against the ACE2 binding region on RBD
using RifDock. The ones that passed a specific energy cutoff value (-1.5 Rosettta energy unit)
were stored and the corresponding inverse rotamers were generated. The de novo scaffold
library (*11*) of 19,000 mini-proteins (in length 56 - 65 residues) were docked into the field of the
inverse rotamers to produce initial docked conformations. These docked conformations were

250 further optimized using the FastDesign protocol to generate shape and chemically
251 complementary interfaces. Computational metrics of the final design models were calculated
252 using Rosetta, which includes ddg, shape complementary and interface buried solvent
253 accessible surface area. Top 100,000 designs based on the above metrics were selected and
254 an Agilent DNA library were ordered for DNA synthesis.

DNA library preparation

256 All protein sequences were padded to a uniform length (88 aa for approach 1 and 65 for
257 approach 2) by adding a (GGGS)n linker at the C terminal of the designs, to avoid the biased
258 amplification of short DNA fragments during PCR reactions. The protein sequences were
259 reversed translated and optimized using DNAworks2.0 (*17*) with the S. cerevisiae codon
260 frequency table. Homologous to the pETCON plasmid Oligo libraries encoding the ACE2 helix
261 scaffolded designs were ordered from Twist Bioscience. Oligo pool encoding the *de novo*
262 designs and the point mutant library were ordered from Agilent Technologies. The error-prone
263 library of the initial hit of the ACE2 helix scaffolded design was constructed by error-prone PCR
264 with a GeneMorph II Random Mutagenesis Kit (Agilent Technologies) with manufacturer's
265 instructions. 3 to 4 mutations were found on each sequence as verified by colony sequencing.
266 Combinatorial libraries were ordered as IDT (Integrated DNA Technologies) ultramers with the
267 final DNA diversity ranging from 1e6 to 1e7.

268 All libraries were amplified using Kapa HiFi Polymerase (Kapa Biosystems) with a qPCR
269 machine (BioRAD CFX96). In detail, the libraries were firstly amplified in a 25 ul reaction, and
270 PCR reaction was terminated when the reaction reached half the maximum yield to avoid over
271 amplification. The PCR product was loaded to a DNA agarose gel. The band with the expected
272 size was cut out and DNA fragments were extracted using QIAquick kits (Qiagen Inc). Then, the
273 DNA product was re-amplified as before to generate enough DNA for yeast transformation. The
274 final PCR product was cleaned up with a QIAquick Clean up kit (Qiagen Inc). For the yeast
275 transformation, 2-3 µg of digested modified pETcon vector (pETcon3) and 6 µg of insert were
276 transformed into EBY100 yeast strain using the protocol as described in (*11*).

277 DNA libraries for deep sequencing were prepared using the same PCR protocol, except the first
278 step started from yeast plasmid prepared from 5×10^7 to 1×10^8 cells by Zymoprep (Zymo
279 Research). Illumina adapters and 6-bp pool-specific barcodes were added in the second qPCR
280 step. Gel extraction was used to get the final DNA product for sequencing. All libraries include
281 the native library and different sorting pools were sequenced using Illumina NextSeq/MiSeq
282 sequencing.

Yeast surface display

284 *S. cerevisiae* EBY100 strain cultures were grown in C-Trp-Ura media and induced in SGCAA
285 media following the protocol in (*11*) . Cells were washed with PBSF (PBS with 1% BSA) and
286 labelled with biotinylated RBD using two labeling methods, with-avidity and without-avidity
287 labeling. For the with-avidity method, the cells were incubated with biotinylated RBD, together
288 with anti-c-Myc fluorescein isothiocyanate (FITC, Miltenyi Biotech) and streptavidin–
289 phycoerythrin (SAPE, ThermoFisher). The concentration of SAPE in the with-avidity method
290 was used at ¼ concentration of the biotinylated RBD. The with-avidity method was used in the
291 first few rounds of screening of the original design to fish out weak binder candidates. For the
292 without-avidity method, the cells were firstly incubated with biotinylated RBD, washed,
293 secondarily labelled with SAPE and FITC. For the ACE2 helix scaffolded designs, three rounds

294 of with-avidity sorts were applied at 1 uM concentration of RBD. For the original library of de
295 novo designs, the library was sorted twice using the with-avidity method at 1 uM RBD, followed
296 by several without-avidity sort in the third round of sorting with RBD concentrations at 1uM,
297 100nM and 10 nM. The SSM library was screened using the without-avidity method for four
298 rounds, with RBD concentrations at 1uM, 100nm, 10nM and 1nM. The error-prone library and
299 the combinatorial libraries were sorted to convergence by decreasing the target concentration
300 with each subsequent sort and collecting only the top 0.1% of the binding population. The final
301 sorting pool of the error-prone library was sequenced using Illumina MiSeq kit to identify all
302 beneficial mutations. The final sorting pools of the combinatorial libraries were plated on C-trp-
303 ura plates and the sequences of individual clones were determined by Sanger sequencing.

304 For protease treatment, the cells were firstly washed with TBS (20 mM Tris 100 mM NaCl pH
305 8.0). Protolysis was initiated by adding 250 μL of room temperature TBS with 1 uM of trypsin
306 and 0.5 uM of chymotrypsin, followed by vortexing and incubating the reaction at room
307 temperature for 5 min. The reaction was quenched by adding 1 mL of chilled PBSF, and cells
308 were immediately washed 4x in chilled PBSF. The cells were then labeled with RBD using the
309 above mentioned protocol.

310 **Protein expression**

311 Genes encoding the designed protein sequences were synthesized and cloned into modified
312 pET-29b(+) *E. coli* plasmid expression vectors (GenScript, N-terminal 8× His-tagged followed by
313 a TEV cleavage site). For all the designed proteins, the sequence of the N-terminal tag used is
314 MSHHHHHHHHSENLYFQSGGG (unless otherwise noted), which is followed immediately by
315 the sequence of the designed protein. Plasmids were then transformed into chemically
316 competent E. coli Lemo21 cells (NEB). Protein expression was performed using the Studier
317 autoinduction media supplemented with antibiotic, and grown overnight. The cells were
318 harvested by spinning at 4,000xg for 10 min and then resuspended in lysis buffer (300 mM
319 NaCl, 30 mM Tris-HCL, pH 8.0, with 0.25% CHAPS for cell assay samples) with DNAse and
320 protease inhibitor tablets. The cells were lysed with a QSONICA SONICATORS sonicator for 4
321 minutes total (2 minutes on time, 10 sec on-10 sec off) with an amplitude of 80%. Then the
322 soluble fraction was clarified by centrifugation at 20,000g for 30 min. The soluble fraction was
323 purified by Immobilized Metal Affinity Chromatography (Qiagen) followed by FPLC size-
324 exclusion chromatography (Superdex 75 10/300 GL, GE Healthcare). All protein samples were
325 characterized with SDS-PAGE with the purity higher than 95%. Protein concentrations were
326 determined by absorbance at 280 nm measured using a NanoDrop spectrophotometer (Thermo
327 Scientific) using predicted extinction coefficients.

328 **Circular dichroism**

329 Far-ultraviolet CD measurements were carried out with an JASCO-1500 equipped with a
330 temperature-controlled multi-cell holder. Wavelength scans were measured from 260 to 190 nm
331 at 25, 95°C and again at 25°C after fast refolding (~5 min). Temperature melts monitored
332 dichroism signal at 222 nm in steps of 2°C/minute with 30s of equilibration time. Wavelength
333 scans and temperature melts were performed using 0.3 mg/ml protein in PBS buffer (20mM
334 NaPO4, 150mM NaCl, pH 7.4) with a 1 mm path-length cuvette.

335 Melting temperatures were determined fitting the data with a sigmoid curve equation. 7 out of
336 the 9 designs retained more than half of the mean residue ellipticity values, which indicated the
337 Tm values are greater than 95°C. Tm values of AHB1 and LCB2 were determined as the
338 inflection point of the fitted function.

Biolayer interferometry

Biolayer interferometry binding data were collected in an Octet RED96 (ForteBio) and processed using the instrument's integrated software. For minibinder binding assays, biotinylated RBD (Acro Biosystems, for de novo miniprotein binders) or Fc tagged RBD (kindly provided by Erik Procko, for the ACE2 helix scaffolded binders) was loaded onto streptavidin-coated biosensors (SA ForteBio) or proteinA biosensors (ProteinA ForteBio) at 20 nM in binding buffer (10 mM HEPES (pH 7.4), 150 mM NaCl, 3 mM EDTA, 0.05% surfactant P20, 0.5% non-fat dry milk) for 360 s. Analyte proteins were diluted from concentrated stocks into binding buffer. After baseline measurement in the binding buffer alone, the binding kinetics were monitored by dipping the biosensors in wells containing the target protein at the indicated concentration (association step) and then dipping the sensors back into baseline/buffer (dissociation). For the ACE2 competition assay, ACE2-Fc protein (house made) was loaded onto proteinA biosensors (SA ForteBio) at 20 nM in binding buffer for 360 s, and the RBD protein (house made) was used as the ligand with a concentration of 100 nM. AHB1 protein was diluted into the ligand well at the indicated concentrations. Same Octet protocol as above mentioned was applied for acquiring the association and dissociation spectrums.

SARS-CoV-2 neutralization assay

SARS-CoV-2 strain 2019 n-CoV/USA_WA1/2020 was obtained from the Centers for Disease Control and Prevention (gift of Natalie Thornburg). Virus stocks were produced in Vero CCL81 cells (ATCC) and titrated by focus-forming assay on Vero E6 cells. Serial dilutions of minibinders were incubated with 10^2 focus-forming units (FFU) of SARS-CoV-2 for 1 h at 37°C. RBD binder-virus complexes were added to Vero E6 cell monolayers in 96-well plates and incubated for 1 h at 37°C. Next, cells were overlaid with 1% (w/v) methylcellulose in MEM supplemented to contain 2% FBS. Plates were harvested 30 h later by removing overlays and fixed with 4% PFA in PBS for 20 min at room temperature. Plates were washed and sequentially incubated with 1 µg/mL of CR3022 (*5*) anti-S antibody and HRP-conjugated goat anti-human IgG in PBS supplemented to contain 0.1% saponin and 0.1% BSA. SARS-CoV-2-infected cell foci were visualized using TrueBlue peroxidase substrate (KPL) and quantitated on an ImmunoSpot microanalyzer (Cellular Technologies). Data was processed using Prism software (GraphPad Prism 8.0).

Cryo Electron microscopy

To prepare SARS-CoV-2 S/LCB complexes, 10-fold molar excesses of LCB1 or LCB3 were incubated with SARS-CoV-2 HexaPro Spike protein (*18*) for 1 h on ice, and the complexes were purified on a Superose 6 Increase 10/300 GL column. 3 µL of 1 mg/ml purified SARS-CoV S/LCB complexes were loaded onto a freshly glow-discharged (30 s at 20 mA) 1.2/1.3 UltrAuFoil grid (300 mesh) prior to plunge freezing using a vitrobot Mark IV (ThermoFisher Scientific) using a blot force of 0 and 5.5 second blot time at 100% humidity and 25°C. Data were acquired using an FEI Titan Krios transmission electron microscope operated at 300 kV and equipped with a Gatan K2 Summit direct detector and Gatan Quantum GIF energy filter, operated in zero-loss mode with a slit width of 20 eV. Automated data collection was carried out using Leginon at a nominal magnification of 130,000x with a pixel size of 0.525 Å. The dose rate was adjusted to 8 counts/pixel/s, and each movie was acquired in super-resolution mode

385   fractionated in 50 frames of 200 ms. 2,500 micrographs were collected with a defocus range
386   between -0.5 and -2.5 μm. Movie frame alignment, estimation of the microscope contrast-
387   transfer function parameters, particle picking, and extraction were carried out using Warp.
388   Particle images were extracted with a box size of 800 binned to 400 yielding a pixel size of 1.05
389   Å. Reference-free 2D classification was performed using CryoSPARC to select well-defined
390   particle images. These selected particles were subjected to 3D classification in Relion. 3D
391   refinements were carried out using non-uniform refinement along with per-particle defocus
392   refinement in CryoSPARC. Selected particle images were subjected to the Bayesian polishing
393   procedure implemented in Relion3.0. After determining a refined 3D structure, the particles were
394   then subjected to focus 3D classification without refining angles and shifts using a soft mask on
395   Receptor binding motif (RBM) and LCB region with a tau value of 60. articles belonging to
396   classes with the best resolved LCB1/3 density were selected and subject to local refinement
397   with CryoSPARC, resulting in 3D reconstructions with improved quality for the minibinders.
398   Local resolution estimation, filtering, and sharpening were carried out using CryoSPARC.
399   Reported resolutions are based on the gold-standard Fourier shell correlation (FSC) of 0.143
400   criterion and Fourier shell correlation curves were corrected for the effects of soft masking by
401   high-resolution noise substitution.
402
403
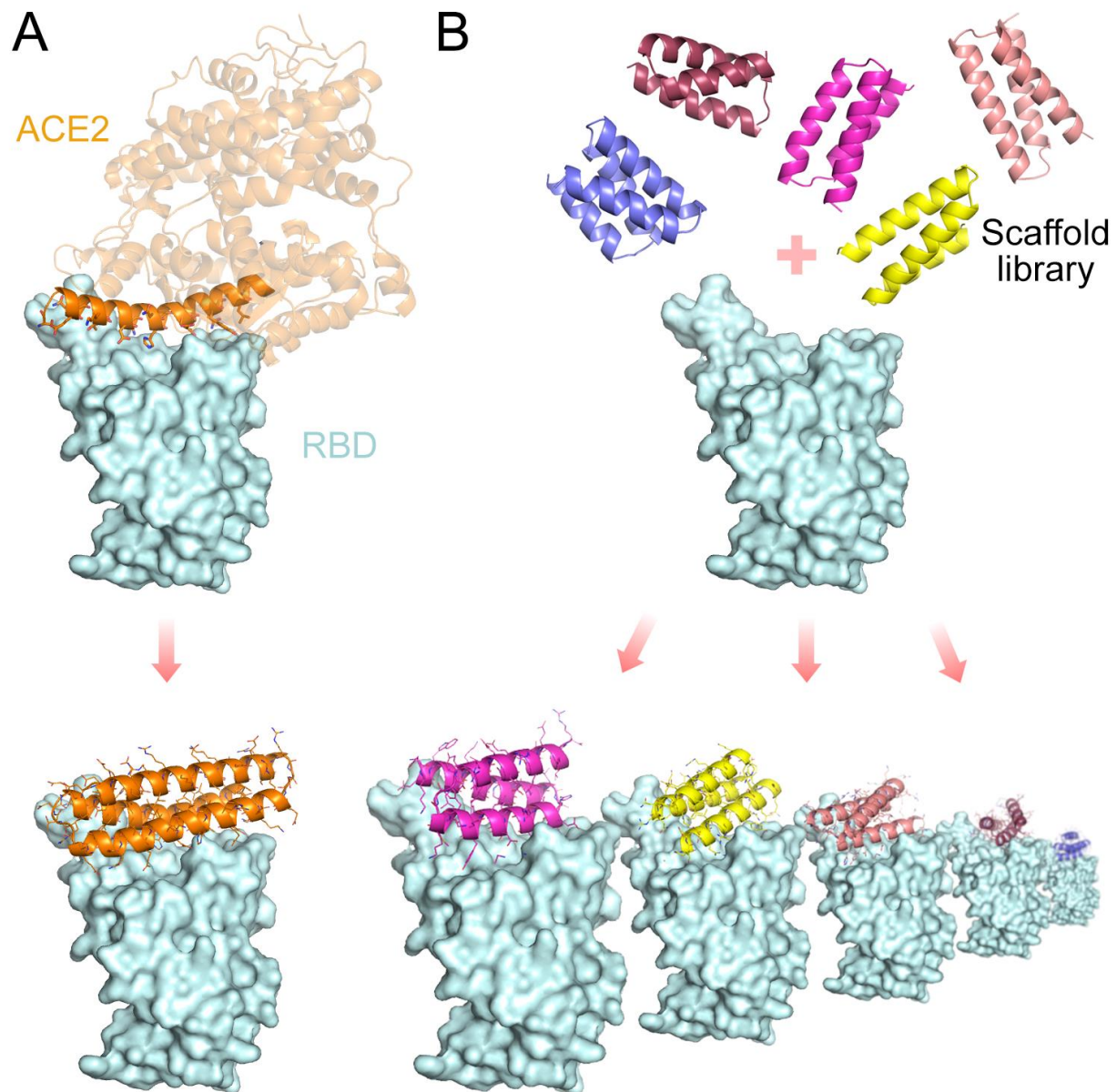404
405
406
407
408
409
410

411
412
Figure 1 | Overview of the computational design approaches. A. Design of helical proteins
incorporating ACE2 helix. B. Large scale de novo design of small helical scaffolds (top) followed
by RIF docking to identify shape and chemically complementary binding modes.
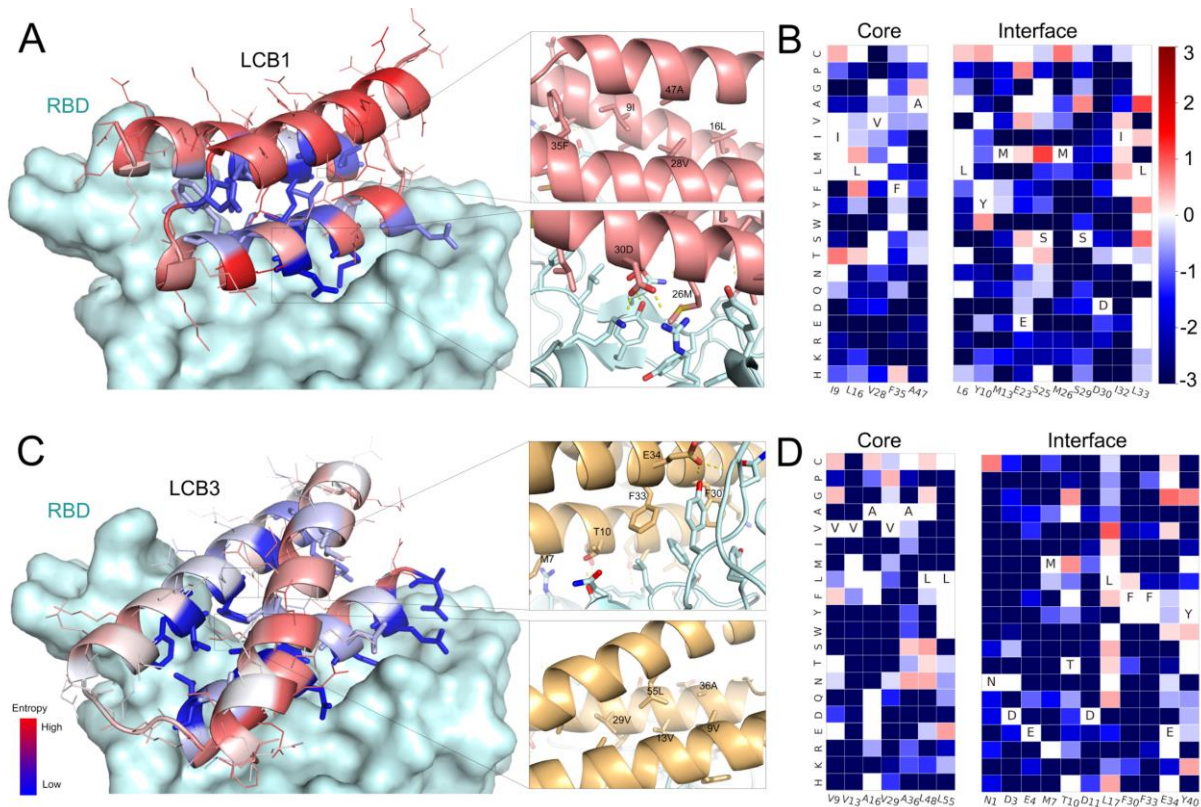
416
417
418
419
420

Figure 2 | High resolution sequence mapping of LCB1 and LCB3 prior to sequence optimization. (A) and (B) The designed binding proteins are colored by positional Shannon entropy from site saturation mutagenesis with blue indicating positions of low entropy (conserved) and red those of high entropy (not conserved). (C) and (D) Heat maps representing RBD-binding enrichment values for single mutations in the design model core (left) and the designed interface (right). Substitutions that are heavily depleted are shown in blue, and beneficial mutations in red. The depletion of most substitutions in both the binding site and the core suggest that the design models are largely correct, while the enriched substitutions suggest routes to improving affinity. Full SSM maps over all positions for all eight de novo designs are provided in Fig S5.

442

443

444

445

446

447

448

449

450

451
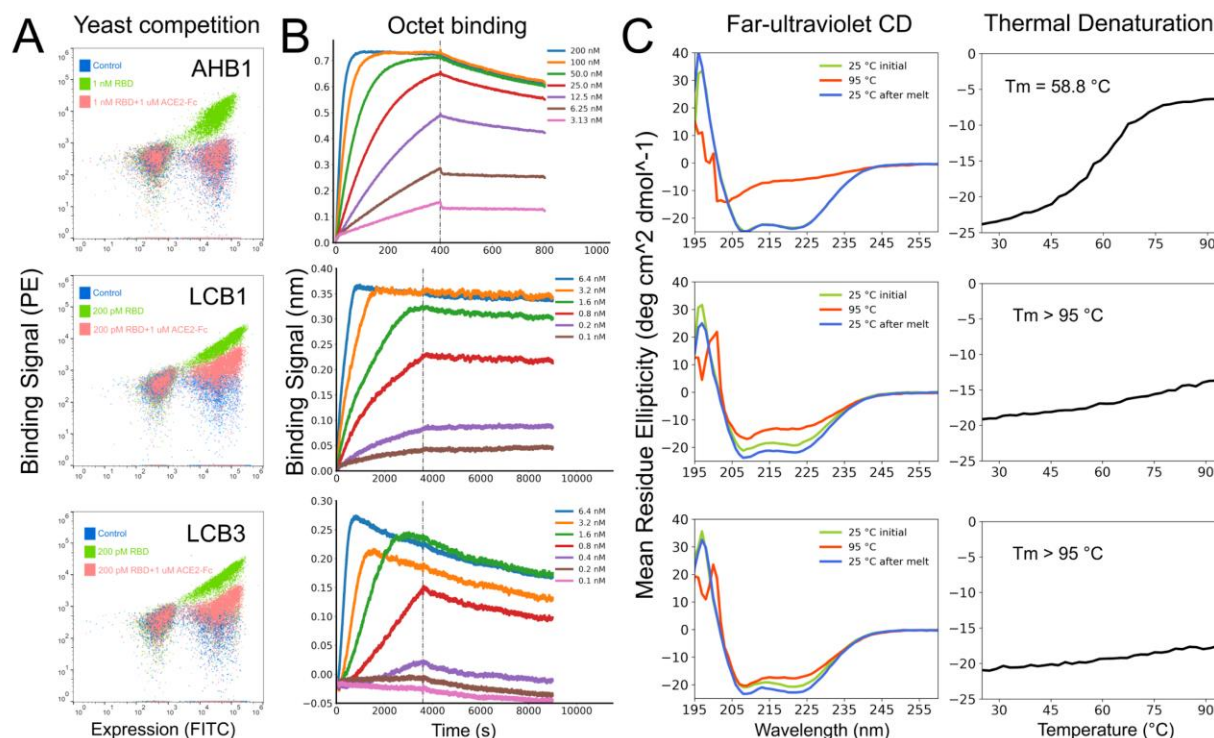
Figure 3 | The optimized designs bind with high affinity to the RBD, compete with ACE2, and are thermostable. A. ACE2 competes with the designs for binding to the RBD. Yeast cells displaying the indicated design were incubated with 1nM/200pM RBD in the presence or absence of 1uM ACE2, and RBD binding to cells (Y axis) was monitored by flow cytometry. B. Binding of purified miniproteins to the RBD monitored by biolayer interferometry. For LCB1 and LCB3 Kd's could not be accurately estimated due to lack of instrument sensitivity and long equilibration times below 200pM. C. Circular dichroism spectra at different temperatures, and D. CD signal at 222 nm wavelength as a function of temperature. The fully de novo designs LCB1 and LCB3 are more stable than the ACE2 scaffolded helix design AHB1.

461
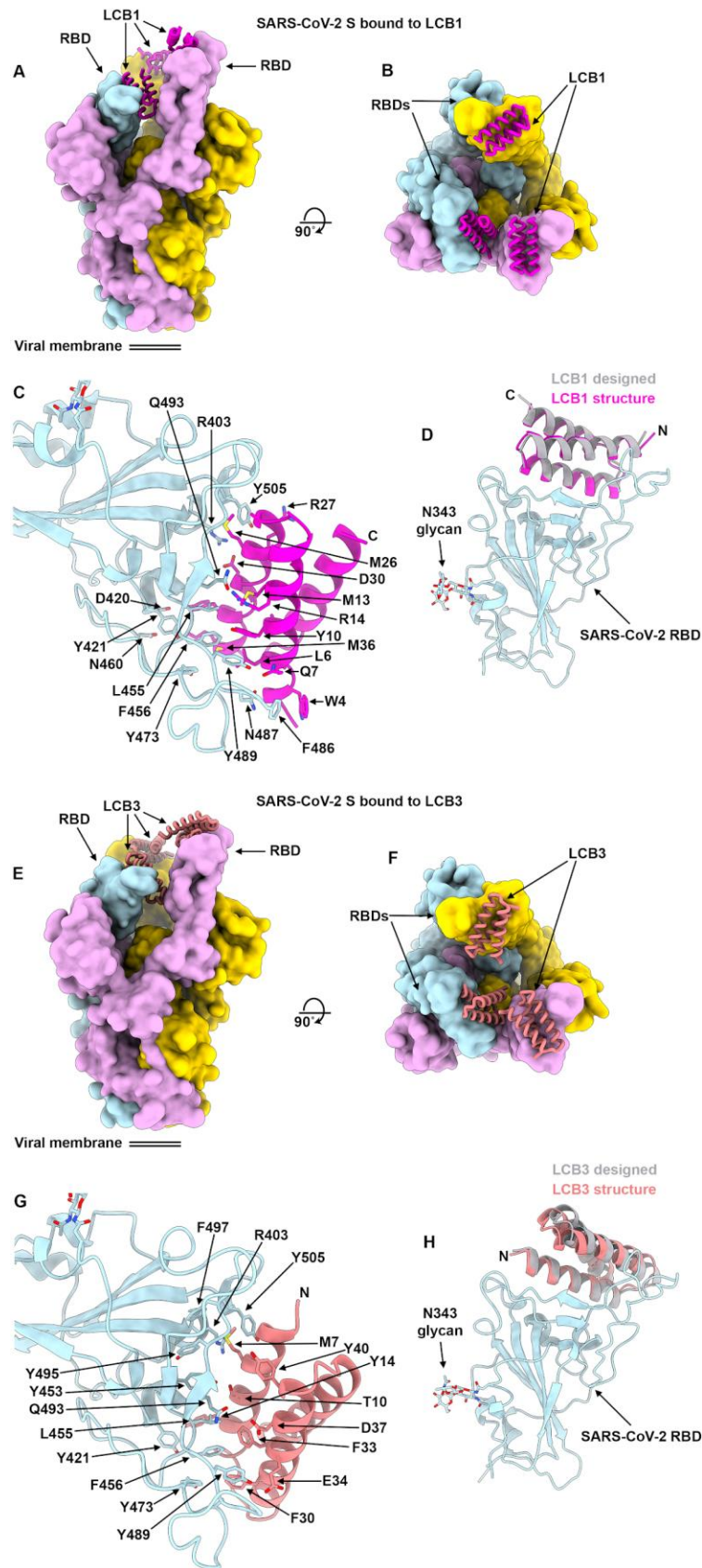
463 **Figure 4 | CryoEM characterization of the LCB1 and LCB3 minibinders in complex with**
464 **SARS-CoV-2 S. A-B,** Molecular surface representation of LCB1 bound to the SARS-CoV-2 S
465 ectodomain trimer viewed along two orthogonal orientations. **C,** Zoomed-in view of the
466 interactions between LCB1 (magenta) and the SARS-CoV-2 RBD (cyan) showing selected
467 interacting side chains. **D,** Superimposition of the designed model (grey) and refined cryoEM
468 structure (magenta) of LCB1 (using the map obtained through local refinement) bound to the
469 RBD (cyan). **E-F,** Molecular surface representation of LCB3 bound to the SARS-CoV-2 S
470 ectodomain trimer viewed along two orthogonal orientations. **G,** Zoomed-in view of the
471 interactions between LCB3 (salmon) and the SARS-CoV-2 RBD (cyan) showing selected
472 interacting side chains. **H,** Superimposition of the designed model (grey) and refined cryoEM
473 structure (salmon) of LCB3 (using the map obtained through local refinement) bound to the RBD
474 (cyan) In panels A-B and E-F, each S protomer is colored distinctly (cyan, plum and gold). For
475 panels D and H, the RBDs were superimposed to evaluate the binding pose deviations between
476 designed models and refined structure of each minibinder.
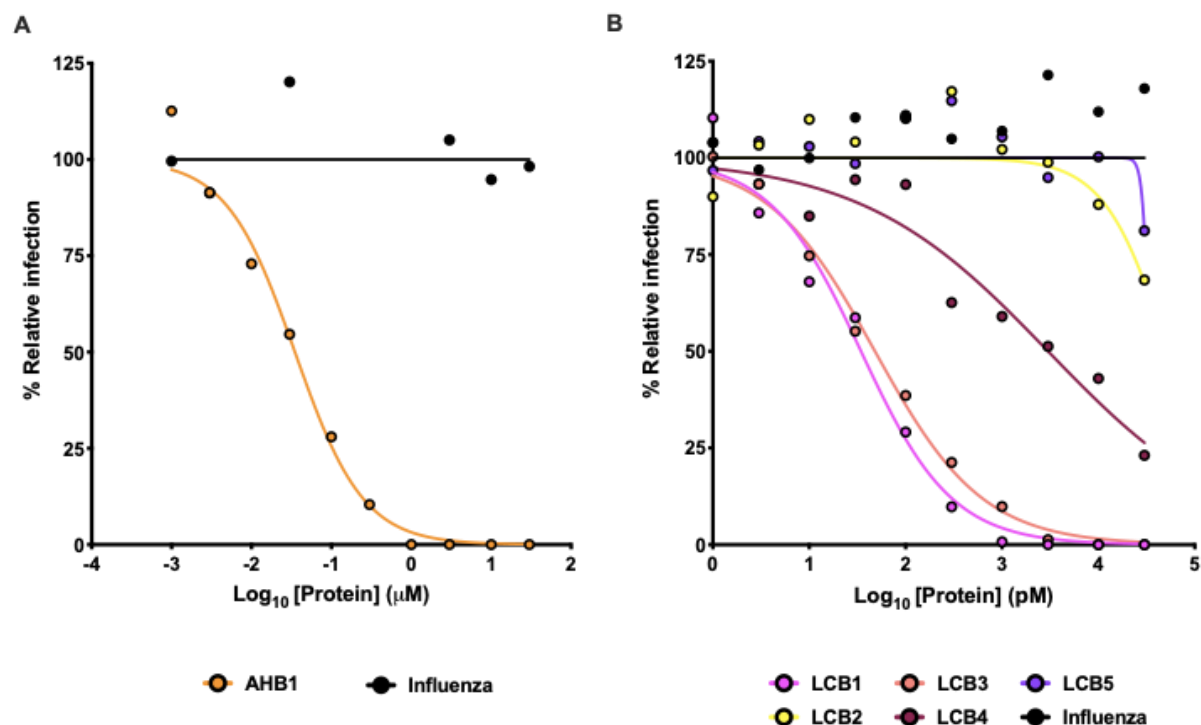477
478
479
480
481
482
483
484
485
486
487

Figure 5 | Neutralization of live virus by designed miniprotein inhibitors. Neutralization activity of (A) AHB1 or (B) LCB1-5 were measured by FRNT. Indicated concentrations of minibinders were incubated with 100 FFU of authentic SARS-CoV-2 and subsequently transferred onto Vero E6 monolayers. AHB1, LCB1, and LCB3 potently neutralize SARS-CoV-2, with EC_{50} values < 50nM (AHB1) or < 50pM (LCB1 and LCB3). Data are representative of two independent experiments, each performed in technical duplicate.
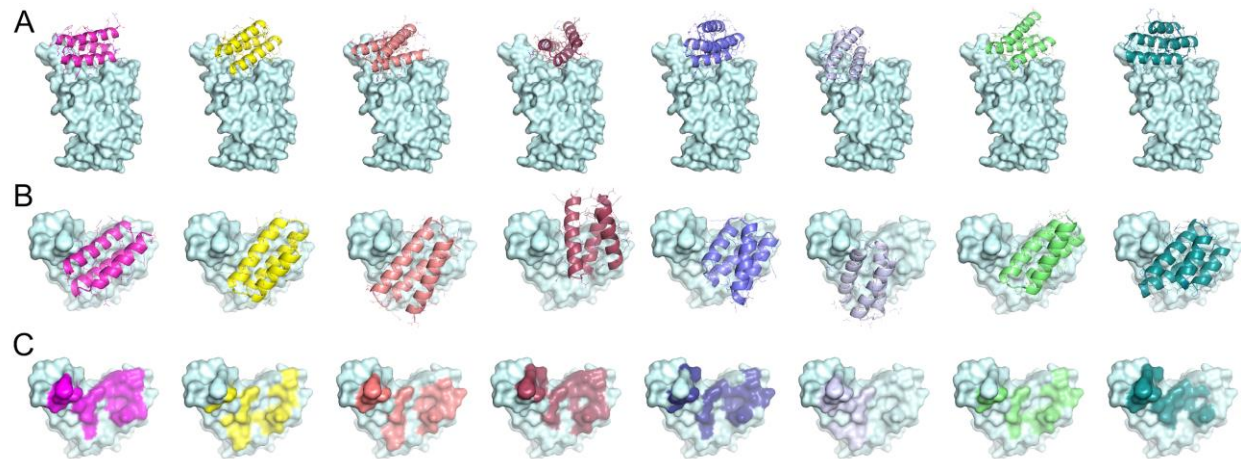
514

515
516

517 **Figure S1 |** Gallery of *de novo designed* miniprotein binders that bind to SARS-CoV-2 RBD with
518 different binding configurations. Side views of *de novo* miniprotein binder-RBD complexes
519 (LCB1-8) are shown in (A) and top views are shown in (B). (C) Residues on RBD that are within
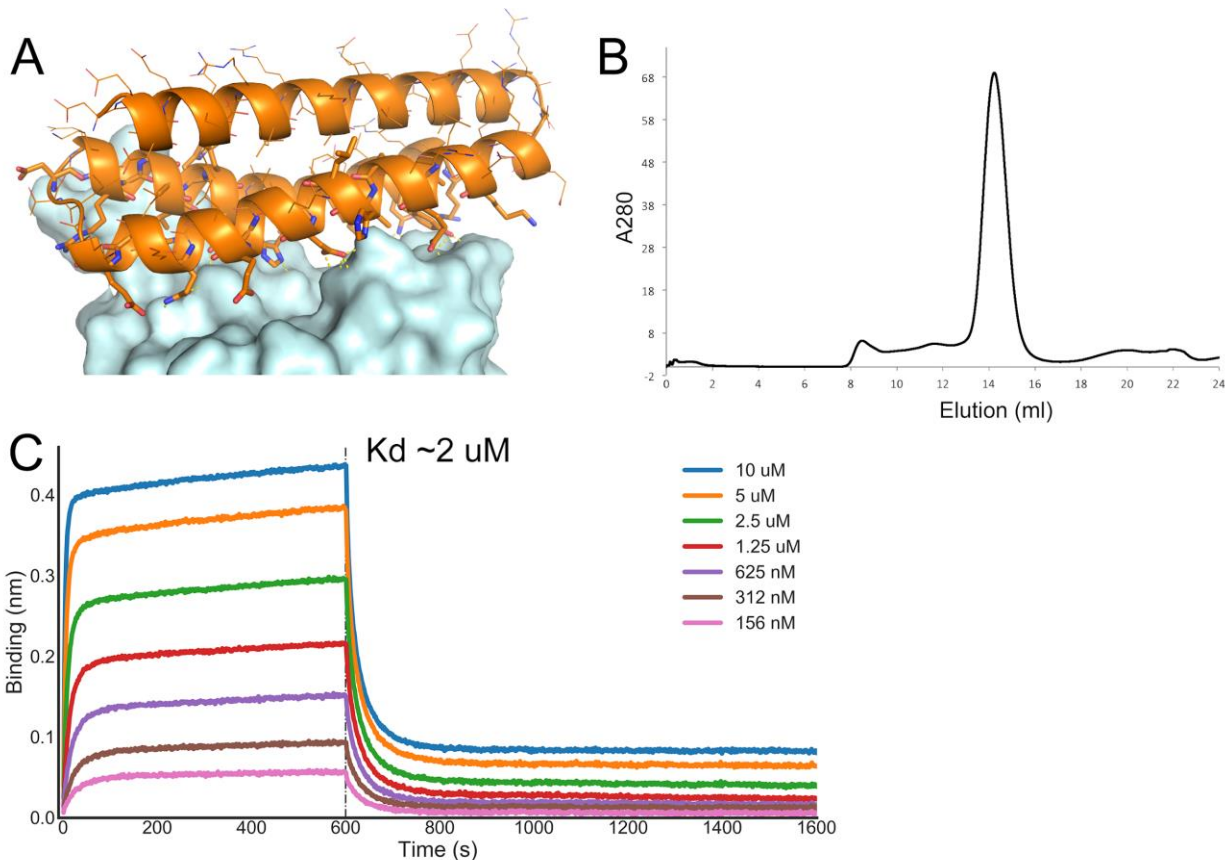520 8Å Cβ distance are highlighted.

521

522

523

524

525

526

527

528

529 **Figure S2 |** Expression and characterization of ACE2 scaffolded design. (A) Design model of
530 the identified ACE2 scaffolded design in complex with RBD. (B) Size Exclusion chromatography
531 profile of the E. coli expressed protein. (C) Binding of ACE2 scaffolded design to RBD in the
532 biolayer interferometry experiments. Fc-tagged RBD protein was loaded onto Protein A
533 biosensors, and allowed to equilibrate before setting the baseline to zero at t=0. The BLI tips
534 were then placed into different concentrations of AHB1 as indicated for 600 seconds. The tips
535 were then placed into buffer solution and the dissociation was monitored for an additional 1,000
536 seconds.

537

538

539

540

541

542

543
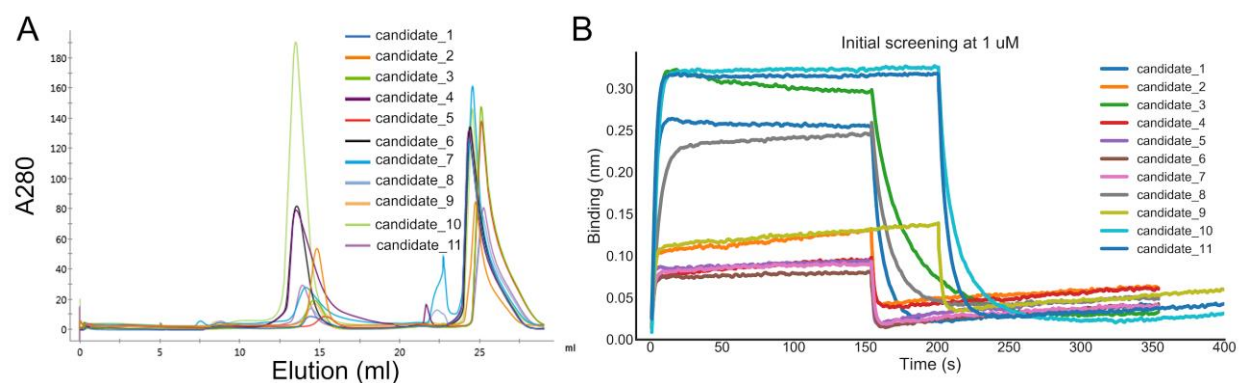
544

545

546
547
548
549
550
551
552
553
554

555

Figure S3 | Expression and characterization of *de novo* designed miniprotein binders. (A) Size Exclusion chromatography profiles of the E. coli expressed initial hits by the de novo approach. (B) Binding profiles of de novo miniproteins to RBD in the biolayer interferometry experiments. The biotinylated RBD protein was loaded onto the Streptavidin (SA) biosensors and allowed to equilibrate before setting the baseline to zero at t=0. The BLI tips were then placed into different concentrations of candidate minibinders (1 uM) as indicated for 150-200 seconds. The tips were then placed into buffer solution and the dissociation was monitored for an additional 600 seconds.
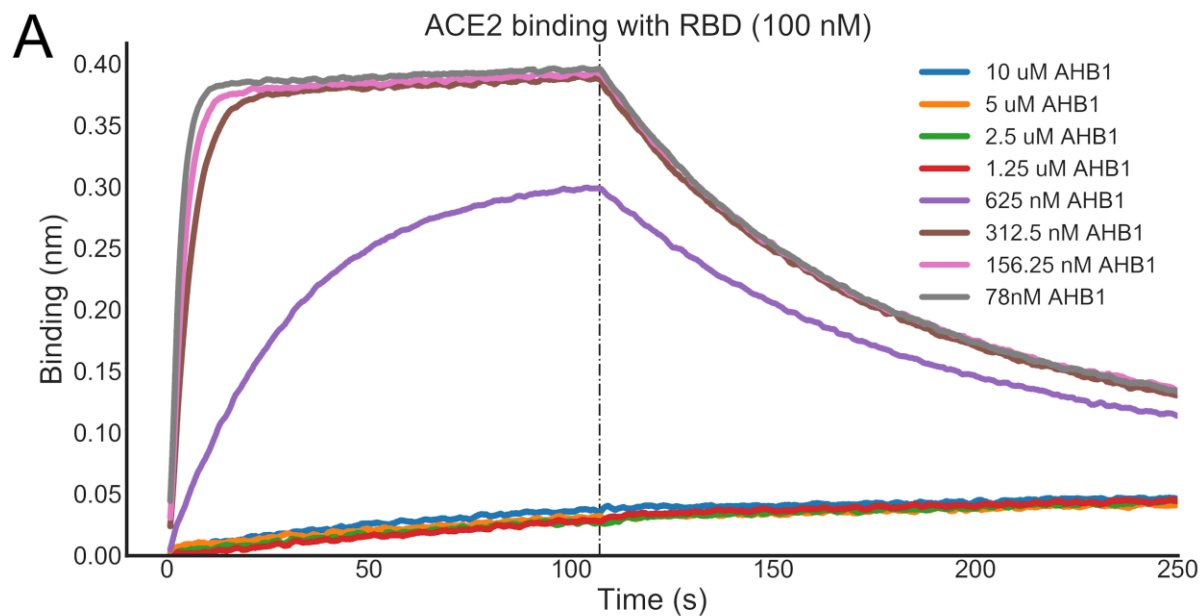
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577

578
579
580
581
582
583
584
585

586
Figure S4 | The ACE2 scaffolded helix design AHB1 blocks binding of ACE2 to the RBD.
The Fc-tagged ACE2 protein was loaded onto the Protein A biosensors in the BLI assay, and allowed to equilibrate before setting the baseline to zero at t=0. The BLI tips were then then placed for 100 seconds into different concentrations of AHB 1 minibinder as indicated, together with a constant concentration of 100 nM of RBD. The tips were then placed into buffer solution and the dissociation was monitored for an additional 150 seconds.
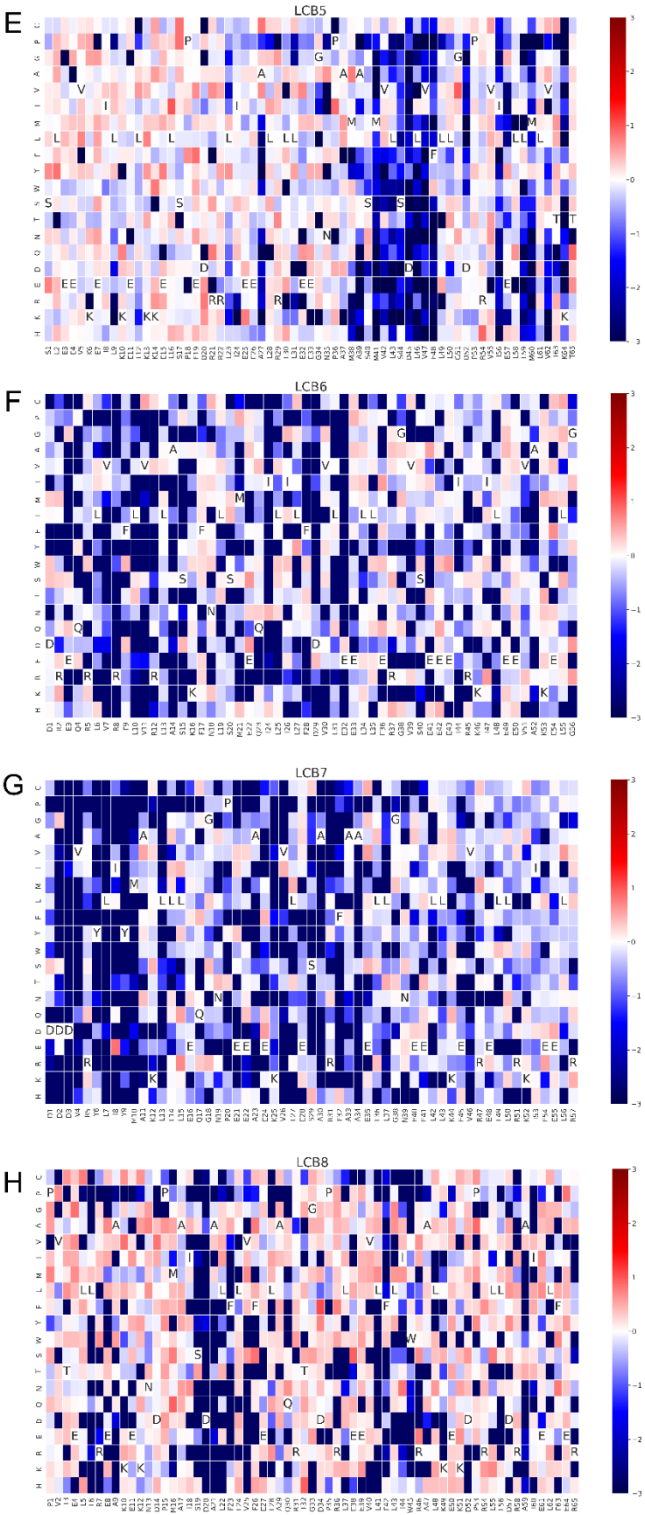
593
594
595
596
597
598
599
600

601

602

603
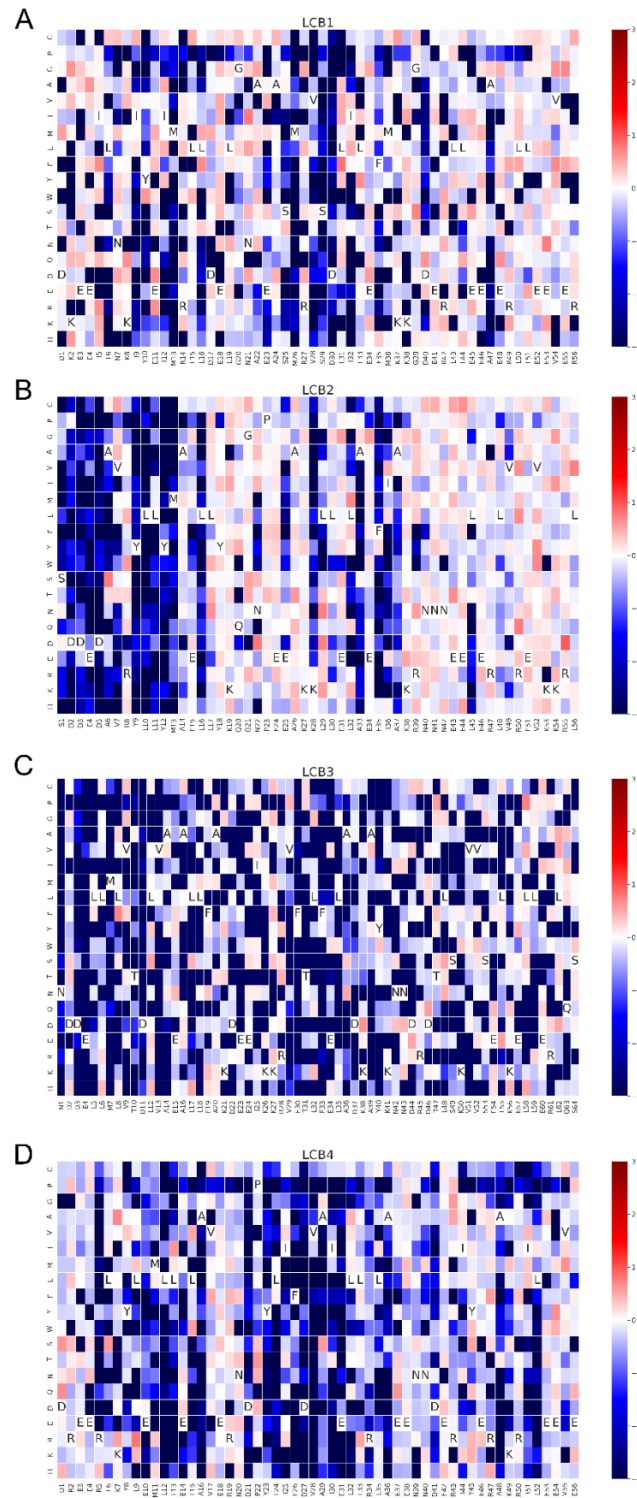
604 **Figure S5 | SSM maps for all eight de novo miniprotein binders.** Colors indicate relative
605 enrichment or depletion of a substitution relative to the unselected population.
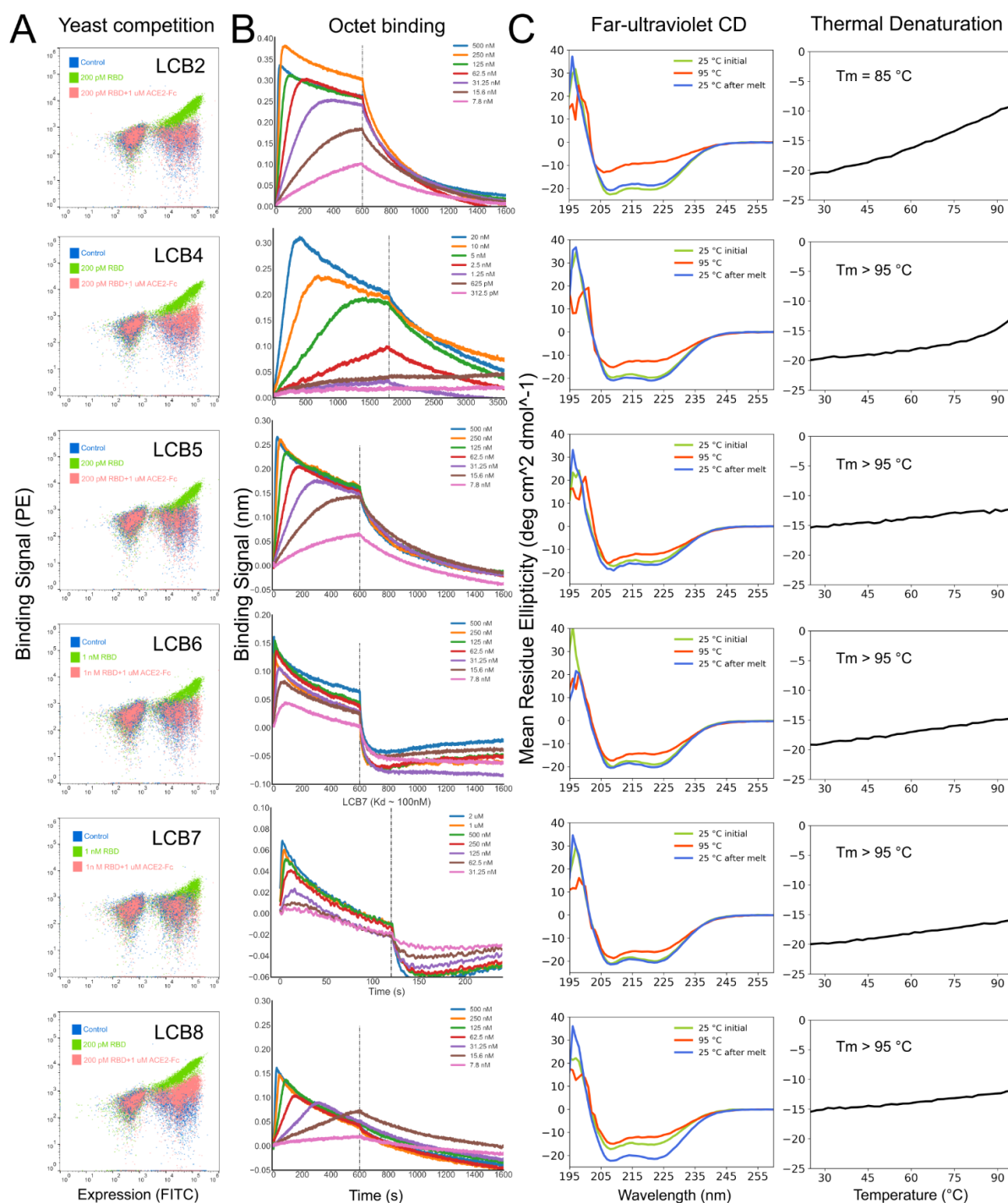
Figure S6 | The optimized designs bind with high affinity to the RBD, compete with ACE2, and are thermostable. A. ACE2 competes with the designs for binding to the RBD. Yeast cells displaying the indicated design were incubated with 200pM RBD in the presence or absence of 1uM ACE2, and RBD binding to cells (Y axis) was monitored by flow cytometry. B. Binding of purified miniproteins to the RBD monitored by biolayer interferometry. The biotinylated RBD

613   protein was loaded onto the Streptavidin (SA) biosensors and allowed to equilibrate before
614   setting the baseline to zero at t=0. The BLI tips were then placed into different concentrations of
615   candidate minibinders (at concentrations noted) as indicated for 120 or 600 seconds. The tips
616   were then placed into buffer solution and the dissociation was monitored for an additional 300 or
617   1,000 seconds. C. Circular dichroism (CD) spectra at different temperatures, and D. CD signal
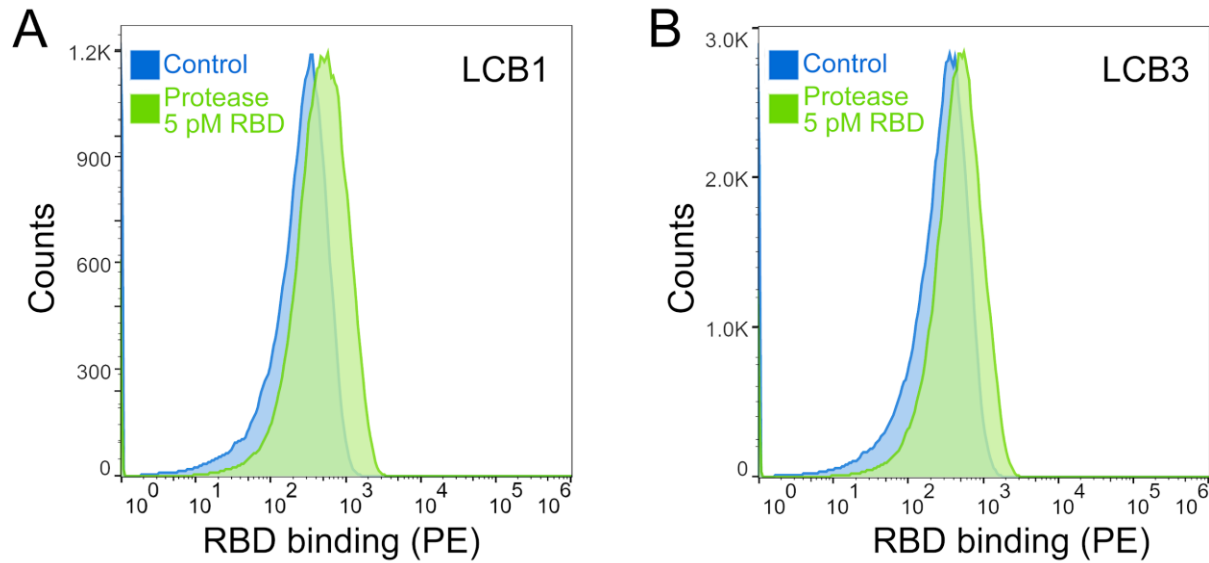618   at 222 nm wavelength as a function of temperature.

619

620

621

622

623

624

625

626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654

655
656
657
658
659
660

661
662 **Figure S7 | LCB1 and LCB3 bind to RBD at 5 pM following protease treatment.** Yeast cells
663 displaying LCB1 and LCB3 were incubated with 1uM trypsin and 0.5 uM chymotrypsin, washed,
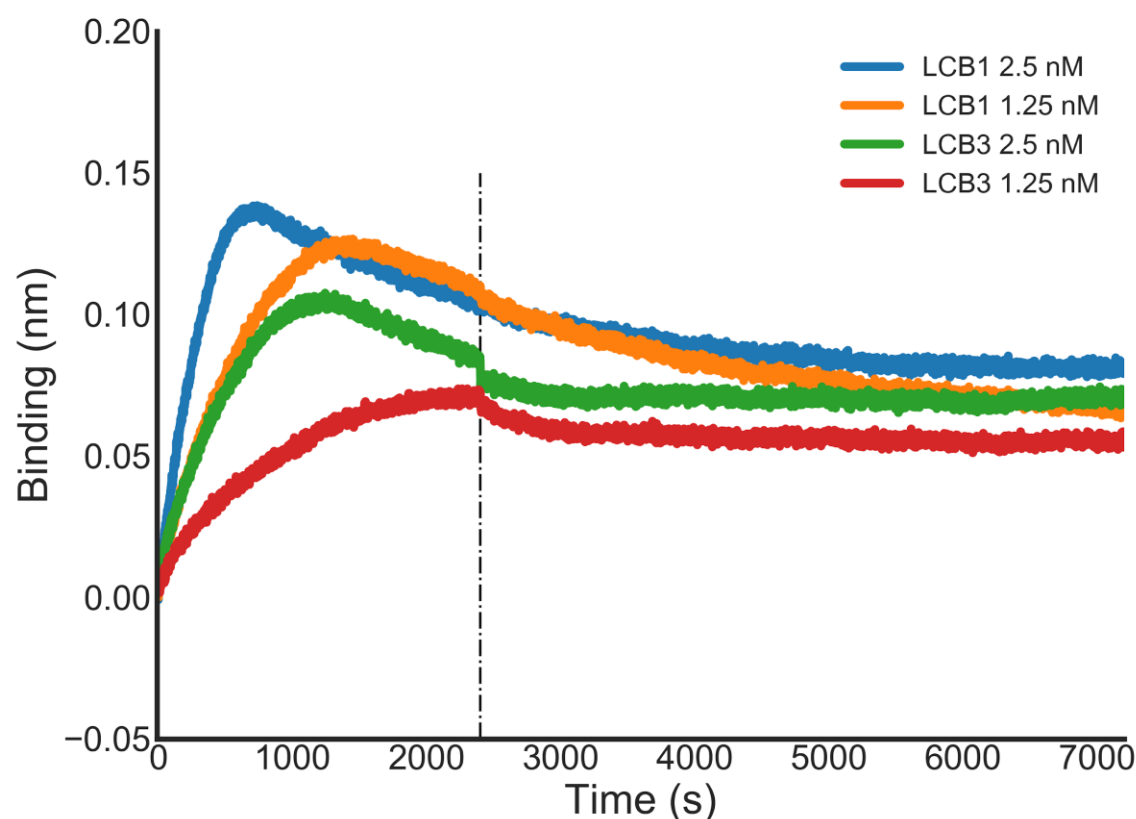664 incubated with 5 pM biotinylated RBD, and analyzed by flow cytometry.

665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683

684
685
686
687

688
689 **Figure S8 | LCB1 and LCB3 retain activity after a few days at room temperature.** LCB1 and
690 LCB3 were left on the lab bench at room temperature for 14 days and the binding with RBD was
691 checked using the BLI assay.   Fc-tagged RBD protein was loaded onto Protein A biosensors,
692 and allowed to equilibrate before setting the baseline to zero at t=0. The BLI tips were then
693 placed into different concentrations LCB1 and LCB3 as indicated for 2500 seconds. The tips
694 were then placed into buffer, and the dissociation was monitored for an additional 4500
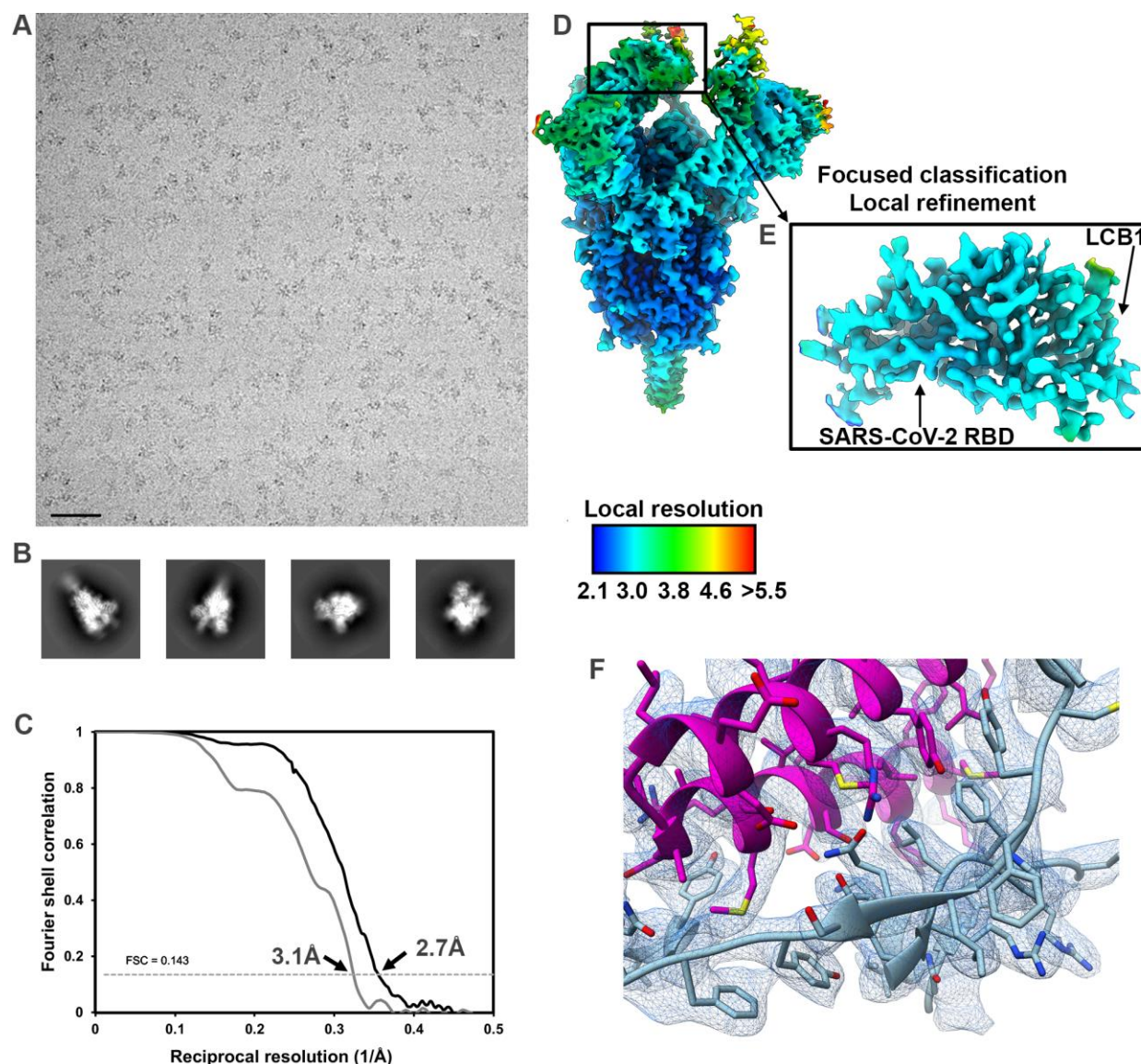695 seconds.
696

697
698
699

Figure S9 | CryoEM data processing and validation of the SARS-CoV-2 structure in complex with LCB1. A-B. Representative electron micrograph (A) and 2D class averages (B) of SARS-CoV-2 S in complex with LCB1 embedded in vitreous ice. Scale bar: 400Å. **C.** Gold-standard Fourier shell correlation curves for the LCB1-bound trimer (black solid line) and locally refined RBD/LCB1 (grey solid line). The 0.143 cutoff is indicated by horizontal dashed lines. **D-E.** Local resolution maps calculated using cryoSPARC for LCB1/S (D) and the locally refined RBD/LCB1 region (E). **F.** Zoomed-in view of the interface between LCB1 (magenta) and the SARS-CoV-2 RBD (cyan) with the corresponding region of density shown as blue mesh.
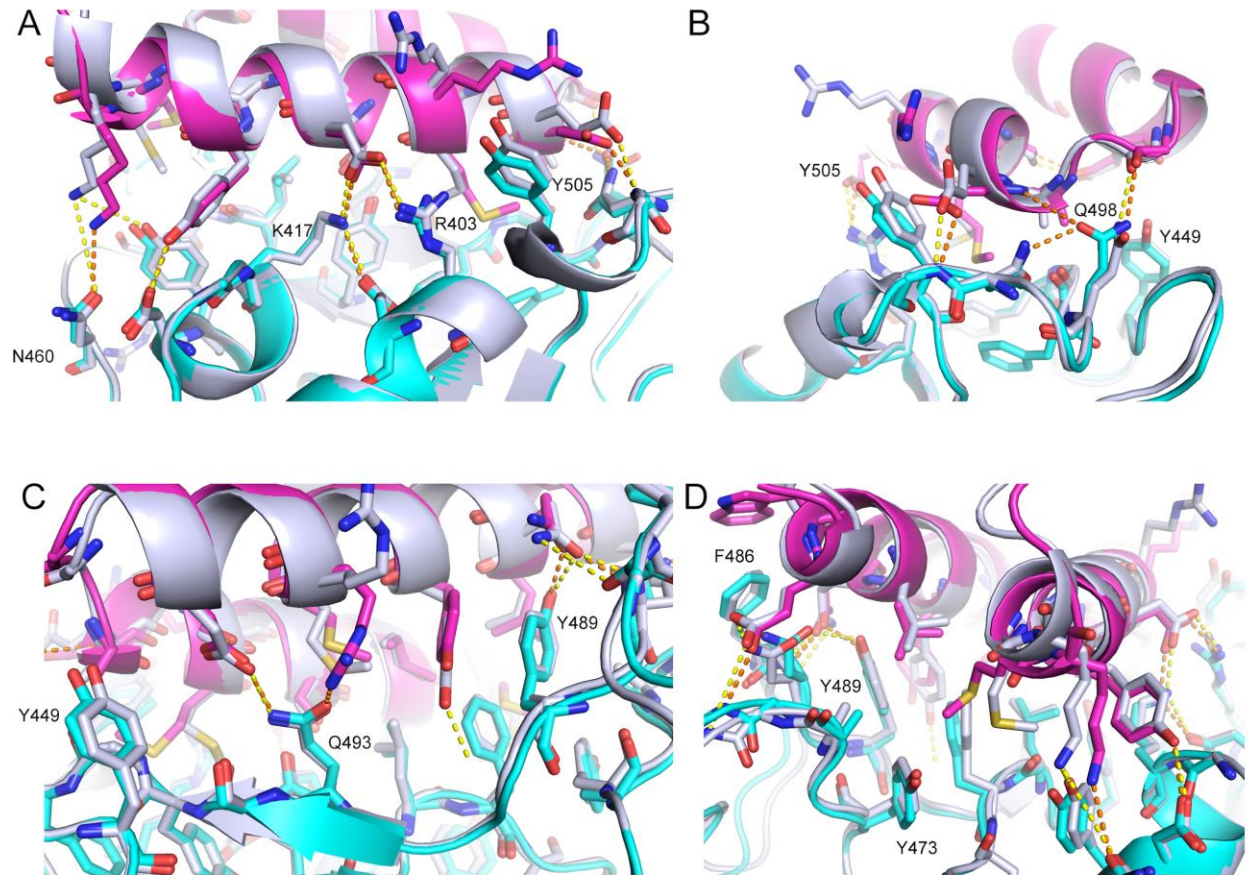
708
709
710

Figure S10 | The CryoEM structure of LCB1 matches *de novo* designed model with atomic accuracy. The computational model (silver grey) of LCB1/RBD complex is overlaid with its CryoEM structure (cyan for RBD and pink for LCB1). Interface interaction details are shown from 4 different views (A-D). Polar interactions are highlighted as dashed lines in orange for the CryoEM structure and in yellow for the design model. Representative residue indices are labeled according to the PDB 6M0J.
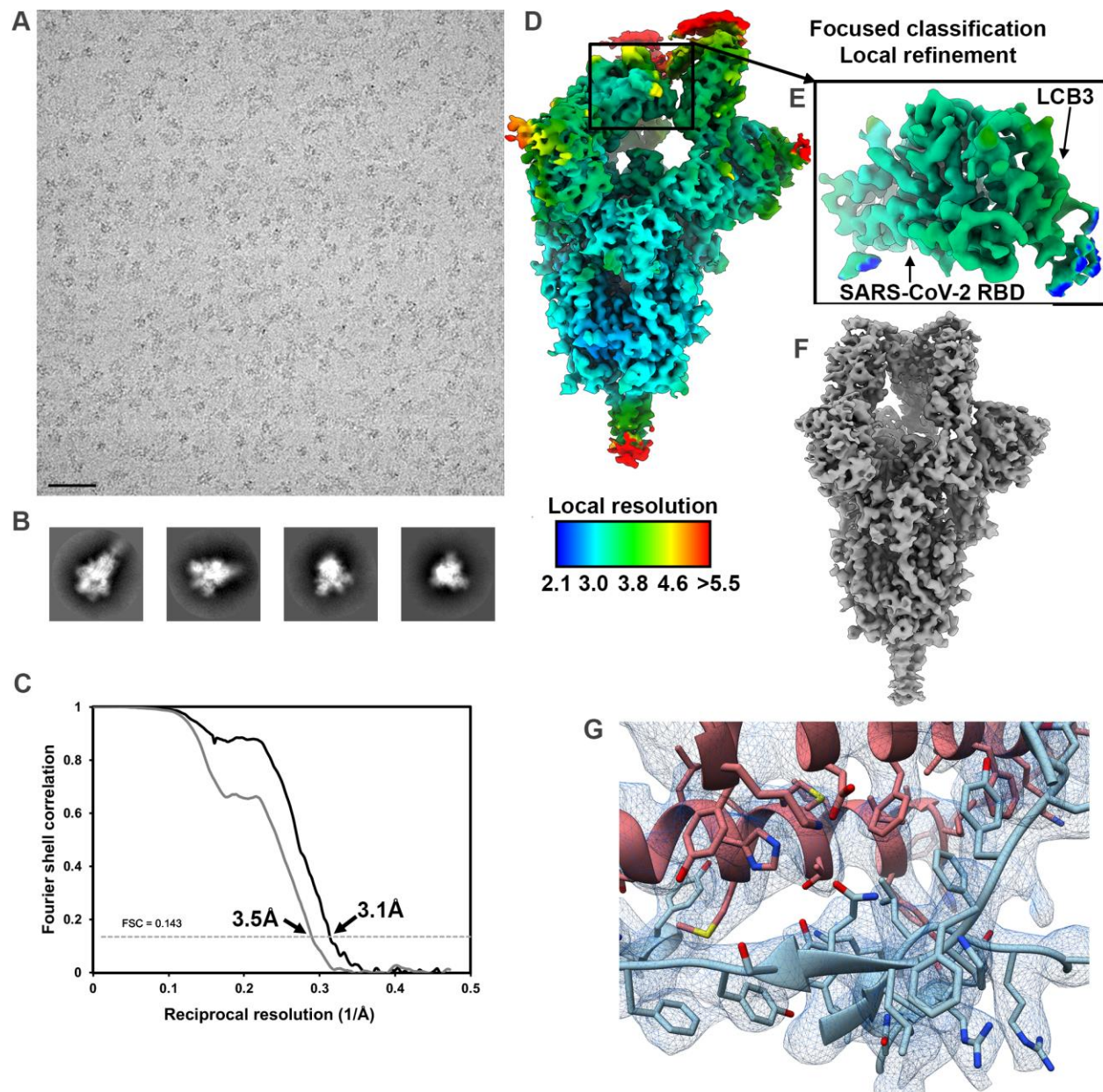
720
721
722

Figure S11 | CryoEM data processing and validation of the SARS-CoV-2 structure in complex with LCB3. A-B. Representative electron micrograph (A) and 2D class averages (B) of SARS-CoV-2 S in complex with LCB3 embedded in vitreous ice. Scale bar: 400Å. **C.** Gold-standard Fourier shell correlation curves for the LCB3-bound trimer (black solid line) and locally refined RBD/LCB3 (grey solid line). The 0.143 cutoff is indicated by horizontal dashed lines. **D-E.** Local resolution maps calculated using cryoSPARC for LCB3/S (D) and the locally refined RBD/LCB3 region (E). **F.** CryoEM reconstruction of LCB3 in complex with SARS-CoV-2 S with three RBDs open. **G.** Zoomed-in view of the interface between LCB3 (salmon) and the SARS-CoV-2 RBD (cyan) with the corresponding region of density shown as blue mesh.
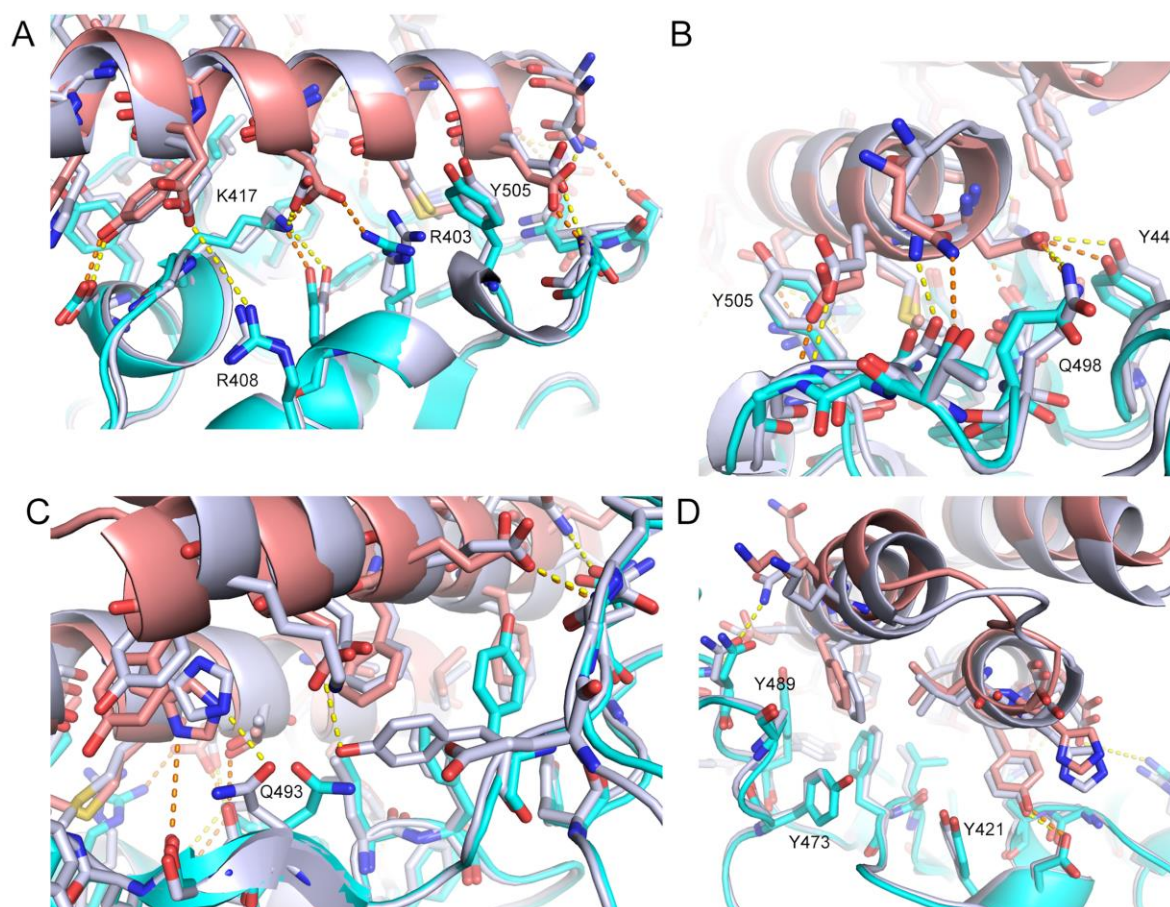
732

733
734
735 **Figure S12 | The CryoEM structure of LCB3 matches with the *de novo* designed model**
736 **with atomic accuracy.** The computational model (silver grey) of LCB3/RBD complex is overlaid
737 with its CryoEM structure (cyan for RBD and light pink for LCB3). Interface interaction details
738 are shown from 4 different views (A-D). Polar interactions are highlighted as dashed lines in
739 orange for the CryoEM structure and in yellow for the design model. Representative residue
740 indices are labeled according to the PDB 6M0J.
741

742 1. Y. J. Hou *et al.*, SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in
743   the Respiratory Tract. *Cell*, (2020).
744 2. R. Shi *et al.*, A human neutralizing antibody targets the receptor binding site of SARS-
745   CoV-2. *Nature*, (2020).
746 3. D. Pinto *et al.*, Structural and functional analysis of a potent sarbecovirus neutralizing
747   antibody. *bioRxiv*, 2020.2004.2007.023903 (2020).
748 4. J. Lan *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the
749   ACE2 receptor. *Nature* **581**, 215-220 (2020).
750 5. M. Yuan *et al.*, A highly conserved cryptic epitope in the receptor binding domains of
751   SARS-CoV-2 and SARS-CoV. *Science* **368**, 630-633 (2020).
752 6. Y. Wu *et al.*, A noncompeting pair of human neutralizing antibodies block COVID-19
753   virus binding to its receptor ACE2. *Science* **368**, 1274-1278 (2020).
754 7. K. L. Winarski *et al.*, Antibody-dependent enhancement of influenza disease promoted
755   by increase in hemagglutinin stem flexibility and virus fusion kinetics. *Proceedings of the
756   National Academy of Sciences* **116**, 15194-15199 (2019).

757 8.   A. Taylor *et al.*, Fc receptors in antibody-dependent enhancement of viral infections.
758   *Immunol Rev* **268**, 340-364 (2015).
759 9.   B. S. Graham, Rapid COVID-19 vaccine development. *Science*, eabb8923 (2020).
760 10. A. C. Walls *et al.*, Structure, function and antigenicity of the SARS-CoV-2 spike
761   glycoprotein. *bioRxiv*, 2020.2002.2019.956581 (2020).
762 11. A. Chevalier *et al.*, Massively parallel de novo protein design for targeted therapeutics.
763   *Nature* **550**, 74-79 (2017).
764 12. J. Dou *et al.*, De novo design of a fluorescence-activating beta-barrel. *Nature* **561**, 485-
765   491 (2018).
766 13. W. B. Alsoussi *et al.*, A Potently Neutralizing Antibody Protects Mice against SARS-CoV-
767   2 Infection. *J Immunol*, (2020).
768 14. D. A. Silva *et al.*, De novo design of potent and selective mimics of IL-2 and IL-15.
769   *Nature* **565**, 186-191 (2019).
770 15. P. S. Huang *et al.*, RosettaRemodel: a generalized framework for flexible backbone
771   protein design. *PLoS One* **6**, e24109 (2011).
772 16. N. Koga *et al.*, Principles for designing ideal protein structures. *Nature* **491**, 222-227
773   (2012).
774 17. D. M. Hoover, J. Lubkowski, DNAWorks: an automated method for designing
775   oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res* **30**, e43 (2002).
776 18. C. L. Hsieh *et al.*, Structure-based Design of Prefusion-stabilized SARS-CoV-2 Spikes.
777   *bioRxiv*, (2020).
778