# Designing a broad-spectrum four-helix bundle targeting different strains of SARS-CoV-2 Spike Receptor Binding Domains with ACE2-like binding interface

**Qinghui Nie[1†], Hongmei Yin[1†], Ke Chen[1,2,3†], Chao Zhang[4†], Weiwei Li[5], Jianxun Qi[5], Jinle Tang[1*], John Z.H. Zhang[4,6-9*], Jian Zhan[1,10*], Yaoqi Zhou[1*]**

[1]Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China

[2]University of Science and Technology of China, Hefei 230026, China

[3]Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China

[4]Faculty of Synthetic Biology, Shenzhen University of Advanced Technology, Shenzhen 518055, China

[5]Laboratory of Pathogen Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[6]State Key Laboratory of Quantitative Synthetic Biology, Shenzhen Institute of Synthetic Biology, Shenzhen Institute of advanced technology, Chinese Academy of Science, Shenzhen 518055, China

[7]NYU-ECNU Center for Computational Chemistry and Shanghai Frontiers Science Center of AI and DL, NYU Shanghai, 567 West Yangsi Road, Shanghai 200124, China

[8]Department of Chemistry, New York University, NY, NY 10003, USA

[9]Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China

[10]Ribopeutic Inc, Guangzhou International Bio Island, Guangdong, 510320, China

†Co-first authors.

*Corresponding authors.

## Abstract

One major strategy for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to evade antibody drugs or preventive vaccines is high mutation rate of the spike receptor binding domain (RBD). Because variable RBDs of different SARS-CoV-2 strains must bind to the same human receptor angiotensin-converting enzyme 2 (hACE2) for viral cell entry and infection, we hypothesize that designing a protein with

the same or very similar hACE2 binding interface might have a broad-spectrum effect against various SARS-CoV-2 strains. The designed protein binds specifically to the WT-RBD (with micromolar affinity) but not to RBDs from other SARS-CoV-2 strains. However, two rounds of the *E. coli* display and Magnetic Cell Sorting （MACS） selection are sufficient to yield a protein named CYN1 with nanomolar binding affinities not only to the WT-RBD but also to those of Omicron BA.1, XBB.1.16, and JN.1. Molecular dynamics simulations and free-energy hotspot analysis revealed that CYN1's broader spectrum capability stems from its engagement of essentially all ACE2 hotspot residues critical for WT-RBD binding, unlike the designed protein. The discovery of CYN1, differing by only four mutations from the designed protein, confirms that targeting the small interface of human viral receptors—rather than the entire receptor—offers a viable strategy for developing broad-spectrum inhibitors. This approach minimizes potential off-target effects arising from receptor multifunctionality.

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) quickly became a worldwide pandemic after it was first identified in December 2019. As of December 22th 2024, this virus has infected nearly 1 billion people and caused over seven million deaths according to the World Health Organization (COVID-19 cases | WHO COVID-19 dashboard, https://data.who.int/dashboards/covid19/cases?n=c). The SARS-CoV-2 virus enters the host cell through the binding of its trimeric spike protein with Angiotensin Converting Enzyme 2 (ACE2) on the human cell surface. The receptor-binding domain (RBD) on the S1 subunit of the spike protein is responsible for binding to ACE2 [1]. The main strategy for the virus to escape the human's immunity (either from previous infection or from the vaccines [2–6]) is evolved mutation on the spike RBD. Many broad-spectrum antibodies and vaccines were proposed and tested but yet to be successful [7,8].

While the spike RBD constantly mutates, our human receptor hACE2 does not change. Thus, it is conceivable that employing hACE2 to bind with the spike RBD might offer broad-spectrum binders of spike RBDs from different virus strains. Indeed, the engineered hACE2 [9–11] were proposed with a demonstrated broad-spectrum effect. However, hACE2, as a hydrophobic transmembrane protein, relied on a solubilizing label [9–13] to be effective. Moreover, ACE2 has many functions [14] and direct use of ACE2

may have unintended side effects. De novo designed proteins/peptides for blocking the binding of spike to ACE2 were also proposed [15–17], but they are not expected to have the broad-spectrum capability of ACE2 as they have significantly modified ACE2.

Here, we designed a four-helix bundle protein based on the short 22-residue double-helical region on ACE2 that interacts with spike's RBD. We subsequently improved it by directed evolution with a demonstrated strong binding to several spike's RBD variants. The mechanism of this broad-spectrum capability was elucidated by conducting molecular dynamics simulations and binding free energy analysis by combining enthalpy and entropy components derived from MM/GBSA[18–20] and interaction entropy (IE) methods[21] with computational Alanine scanning (AS) (ASGBIE method[22–25]) for hotspot analysis[26–29]. The result confirms that improved capture of hotspot residues of ACE2 by the four-helix bundle through directed evolution is the reason behind its improved broad-spectrum capability.

## Materials and Methods

### SPalign

SPalign [30] is a structural alignment program developed by us that compares two protein structures using a size-independent scoring function SP-score. Here it was employed to locate the best natural backbone structure for a manually constructed four-helix bundle (H4) based on two helices in ACE2 located at the hACE2-SARS-CoV-2 RBD interface.

### OSCAR-loop

OSCAR-loop[31] is a program for building loops for a protein region with missing amino acid residues. OSCAR-loop was kindly provided by Dr. Shide Liang. Here the program was employed to build the loop regions for the four-helix bundle H4.

### OSCAR-Design

OSCAR-Design[32] is a protein design program that was developed by us to design protein sequences for a given protein backbone. This program employed an energy function that was based on series expansion with coefficients optimized for protein design. Here the program was employed to design protein sequences given a backbone

97 structure and fixed sidechain conformations of RBD-binding interfacial residues copied
98 from hACE2.
99
100 **DbD2**

101 DbD2 (Disulfide by Design) [33] is a web-based program for designing disulfide bonds
102 in proteins, which is provided at http://cptweb.cpt.wayne.edu/DbD2/. Here the
103 program was used to add a disulfide bond to improve the stability of a designed protein
104 as a monomer.

105 **Plasmids**

106 **Protein expressions**: The H4 sequences designed and a positive control AHB2[15] (A
107 reported RBD binder) were optimized for *Escherichia coli* (*E. Coli*) expression system
108 and directly synthesized and subcloned into pET-21a (+) vector (BGI Genomics,
109 Beijing).
110 ***E. coli* display**: error prone-PCR (EP-PCR) was used to introduce mutations into a
111 target sequence using QuickMutation™ Random Mutagenesis Kit (D0219S, Beyotime)
112 with about 0.8% mutation rate. The mutated DNA fragments were subcloned into *E.coli*
113 display vector pDSG323-Null-Myc (115594, Addgene) between the *Kpn*I and *Bam*HI
114 restriction sites for construction of the mutation library.
115 **Gene knockout:** In order to reduce the interference of fimbriate on *E. coli* display, we
116 knocked out the *fimA-H* from MG1655 strain using pEcCas/pEcgRNA system[34].
117 pEcCas (73227, Addgene) and pEcgRNA (166581, Addgen) were purchased from
118 Addgene. The specific sgRNA sequence was designed using CRISPOR[35] with MG1655
119 genetic background and then constructed into the pEcgRNA vector, named as
120 pEcgRNA-1. The homology arm fragments (sequence in Table S6) were directly
121 synthesized as donor DNA and subcloned into between the EcoRV and PciI restriction
122 sites of pUC57-1.8k vector (General Biol, China).
123 All amino acid sequences and DNA sequences were listed in Tables S5 and S6,
124 respectively.
125
126 **Construction of MG1655$^{\Delta fimA\text{-}H}$ and Generation of *E.coli* display libraries**

127 The modified CRISPR/Cas9 system (pEcCas/pEcgRNA) [34] was used to knock out the
128 *fimA-H* gene for construction of the MG1655$^{\Delta fimA\text{-}H}$ strain. The pEcCas was first
129 transformed into the MG1655 strain. Then the MG1655-pEcCas strain was co-
130 transformed with the pEcgRNA plasmid and donor DNA by electroporation, and
131 recovered for 1 h in 2 mL LB at 37 °C. The clones were randomly picked for colony
132 PCR and Sanger sequencing for verifying the knockout results. Correct colonies were
133 subjected to the plasmid curing process to obtain empty MG1655$^{\Delta fimA\text{-}H}$ strains as
134 described in the reference[34].

135 The constructed library plasmid was transformed into electrocompetent
136 MG1655$^{\Delta fimA\text{-}H}$ strain by electroporation. The size of each library was $1\times10^6$ clones, as
137 determined by plating on LB-Cm agar plates with 2% w/v glucose incubated at 30 °C.
138

139 ***E. coli* display and Magnetic Cell Sorting (MACS)**

140 We followed the previously reported *E. coli* display protocol[36]. More specifically, the
141 *E. coli* library was induced by 1 μg/mL anhydrotetracycline (ATc) at 37 °C for
142 overnight and harvested by centrifugation (3000 × g, 3 min). It is then washed five
143 times with 2 mL PBS and resuspended in 1 mL of PBS. Afterwards, 100 μL of the
144 bacteria was mixed with 50 nM Biotinylated RBD and adjusted to the final volume of
145 200 μL with PBS-BSA (PBS supplemented with 0.5% w/v BSA, sterile filtered and
146 degassed), which was incubated at room temperature for 30 min. Thereafter, the
147 bacteria were washed five times with 1 mL of PBS-BSA, resuspended in 100 μL of the
148 same buffer containing 20 μL of anti-biotin paramagnetic beads (Miltenyi Biotec), and
149 incubated at 4 °C for 20 min. Next, the bacteria were washed three times with 1 mL of
150 PBS-BSA and resuspended in 500 μL of the same buffer, of which 10 μL was kept
151 aside to calculate the input bacteria before the MACS procedure and the remaining
152 portion (490 μL) was applied onto a MACS MS column (Miltenyi Biotec), that was
153 previously equilibrated with 500 μL of PBS-BSA and placed on the OctoMACS
154 Separator (Miltenyi Biotec).

155 The unbound cells (flow-through) were collected, and the column was washed five
156 times with 2 mL of PBS-BSA. The wash was combined with the flow-through as the

157 “Unbound fraction”. Next, the column was removed from the OctoMACS Separator
158 and placed onto a new collection tube to free the bound fraction by washing with 2 mL
159 of PBS-BSA. The bound fraction was applied onto a new equilibrated MACS MS
160 column for reduced bacterial non-specific binding. The column was washed five times
161 with 2 mL of PBS-BSA[36]. Next, the low-affinity mutants were washed away by 10 mM
162 Glycine-HCl pH 2.0 and pH 3.0 (two times with 2 mL each time, and three times for
163 second-round evolution), and the high-affinity mutants were still on the column.
164 Afterwards, the column was removed from the OctoMACS Separator and placed onto
165 a new collection tube, 2 mL of LB was added, and the cells were eluted out. This
166 fraction was labeled as the “Screen fraction”. Serial dilutions of screen fractions were
167 plated to determine CFU and perform the Sanger sequencing.
168
169 **Expression and purification of recombinant proteins**
170 Expression plasmids were transformed into chemically competent *E. coli* Shuffle T7-
171 B. Protein expression was then induced with 0.1 mM of isopropyl β-D-
172 thiogalactopyranoside (IPTG) at 16 ℃. After overnight induction, the cells were
173 harvested by centrifugation at 10,000 × g for 30 min and then resuspended in lysis
174 buffer (150 mM NaCl, 20 mM Tris-HCl (pH 8.0), with 0.1 % triton X-100). The cells
175 were lysed with an Ultrasonic Homogenizer (JY92-IIN, Scientz) for 20 minutes (2 sec
176 on-3 sec off) at an amplitude of 30 %, and then centrifuged at 30,000 × g for 30 min.
177 The precipitate was then removed, and the supernatant was collected. The supernatant
178 was purified by affinity chromatography (HisTrap FF crude 5 mL, Cytiva) and followed
179 by size exclusion chromatography (HiLoad 16/600 Superdex 75 pg, or Superdex 75
180 Increase 10/300 GL, Cytiva) on the ÄKTA pure™ 25M system (Cytiva). All protein
181 samples were characterized using 4-20 % gradient SDS-PAGE and their purity was
182 higher than 95%. Protein concentrations were determined using a NanoDrop
183 spectrophotometer (Thermo Scientific) with absorbance at 280 nm using predicted
184 extinction coefficients (https://web.expasy.org/protparam/).
185

186 **Circular dichroism spectroscopy**

187 Far-ultraviolet CD measurements were carried out with a JASCO-1500 equipped with
188 a temperature-controlled multi-cell holder. Wavelength scans were recorded from a
189 wavelength region of 198 - 255 nm at 25 °C, 95 °C, and 25 °C after melt, and taking the
190 average of four repeats. Temperature melts monitored dichroism signal at 222 nm in
191 steps of 2 °C/minute from 16 to 95 °C with 30 s of equilibration time [15]. Wavelength
192 scans and temperature melts were performed using 0.025 mg/mL protein in PBS buffer
193 (1.8 mM $KH_2PO_4$, 10 mM $Na_2HPO_4$, 137 mM NaCl, 2.7 mM KCl, pH 7.4) with a 1
194 mm path-length cuvette.

195 Melting temperatures were determined by fitting the data with a sigmoid curve
196 equation by the instrument-integrated software. Helical contents were analyzed
197 according to the reported method[37]: negative ellipticity at 208 and 222 nm of the CD
198 spectrum. The α-helical content was predicted using K2D3 web server (K2D3:
199 Estimating protein secondary structure from CD spectra) [38] with the following
200 parameters: (1) the CD spectra from 198 to 240 nm was converted to molar residue
201 ellipticity (MRE, $deg \cdot cm^2 \cdot dmol^{-1}$) units prior to analysis; (2) structural prediction
202 constraints include specifying protein length as 146 amino acid residues.

203 **Biolayer interferometry**

204 Biolayer interferometry (BLI) binding data were collected in Octet RED384 (ForteBio)
205 at 30 °C and processed using the instrument-integrated software. All RBDs were
206 biotinylated by the biotinylating kit (GENEMORE, Cat. G-MM-IGT) according to the
207 kit's manual. The biotinylated RBDs were loaded onto streptavidin-coated biosensors
208 (SA ForteBio, Cat. 18-5019) in PBS (pH 7.4) buffer. The buffer of analyte proteins was
209 exchanged to binding buffer (PBS, 0.05% surfactant P20) by Amicon® Ultra
210 Centrifugal Filter (Millipore) and then diluted to the desired concentrations. After
211 baseline measurement with the binding buffer alone, the binding kinetics were
212 monitored by dipping the biosensors in wells containing the target protein at the
213 indicated concentration (association step) and then dipping the sensors back into
214 baseline/buffer (dissociation). The data were processed using ForteBio Data Analysis

software 12.0, and the equilibrium dissociation constant ($K_D$) values were obtained by global fitting with 1:2 and 1:1 binding models.

**SARS-CoV 2 RBDs**

The RBDs of SARS-CoV 2 variants (WT, XBB.1.16 and JN.1) were purchased from Sino Biological (Cat. 40592-V08B, 40592-V08H136, 40592-V08H155), and the RBD of BA.1 was kindly provided by Haitao Yang's Group. We chose the Omicron BA.1, XBB.1.16 and JN.1 as the representatives of virial strains for testing our designed proteins because these strains were a part of the important SARS-CoV-2 evolution nodes, and have a high immune evasion ability to the anti-spike antibodies or vaccines that were developed based on their strain's ancestors [7].

**Molecular Dynamics Simulations**

The complex structure of ACE2 and WT SARS-CoV-2 Spike RBD (WT-RBD) was obtained from PDB (PDB ID: 6M0J). As each designed protein has two binding interface to WT-RBD, the complex structures of WT-RBD with D2-133 in either its helices 1-2 or helices 3-4 regions were predicted by OSCAR-Design (Figure 1). Other complex structures of CYN1, E6, and E9 in their helices 1-2 or helices 3-4 regions with WT-RBD were mutated from D2-133. All complexes were filled by the leap module in AMBER20 package[39]. All MD simulations were performed using pmemd.cuda in AMBER20 with ff14SB force field for the protein in a truncated octahedron box formed by TIP3P explicit water molecules. The closest distance between any atoms originally present in solute and the edge of the periodic box was 12 Å. The particle mesh Ewald (PME) was used to treat the long-range electrostatic interactions. The nonbonded interactions were truncated with 10 Å cutoff. Periodic boundary condition (PBC) was imposed on the system during the calculation of nonbonded interactions. The time-step was set at 2 fs and SHAKE[40] was used to constrain the bonds involving hydrogen atom. Langevin thermostat with the collision frequency of 2.0 was applied to control the temperature. Firstly, the system was minimized with the protein constrained to equilibrate the solvent. Then, the protein was released to minimize the whole simulation system. Afterward, the system was slowly heated to 300 K, followed by a 40 ns equilibration of the whole system in an NPT ensemble at an interval of every 10 fs, and

during this session, no restrain was performed. Then, a 10ns production run was performed to obtain 5000 snapshots for free-energy and hotspot analysis by ASGBIE.

**Free-energy and hotspot analysis by ASGBIE**

The ASGBIE method[22–25]combines enthalpy and entropy components derived from MM/GBSA[18–20] and interaction entropy (IE) methods[21] with computational Alanine scanning (AS) for hotspot analysis[26–29]. Here, we simply followed the same method utilized in Ref. [22–25].

## Results

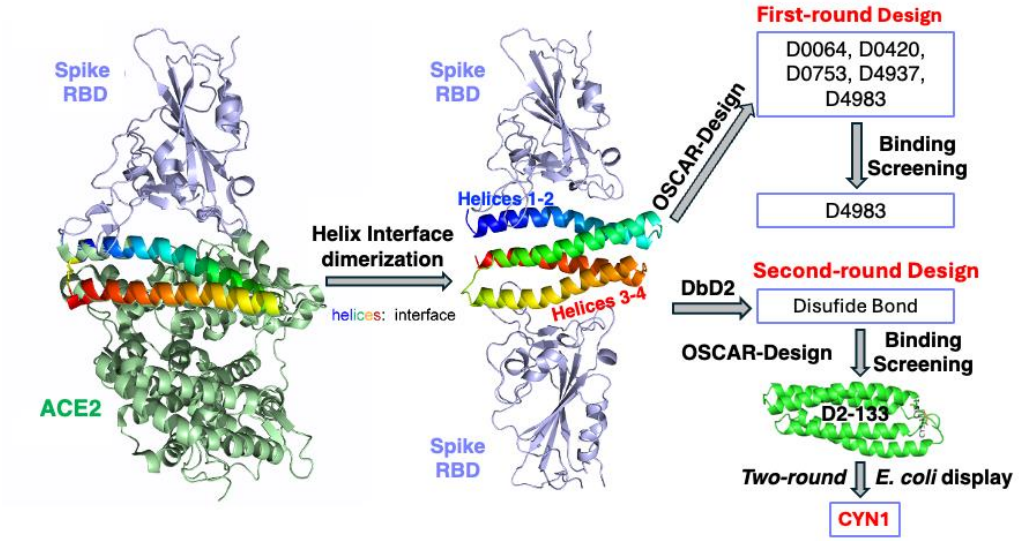



**Figure 1. The flowchart of the iterative design and improvement of a four-helix bundle (H4).** The designed protein has two copies of the two-helix binding interface of hACE2 (helices 1-2 and 3-4) to SARS-CoV-2 Spike receptor binding domain (RBD), whose backbone structure was modeled after the closest four-helix template in the protein databank. The sequences for this backbone structure along with fixed binding residues were designed by OSCAR-Design and five were tested for binding affinity. To improve monomeric stability, the second-round design was performed after adding a disulfide bond. The best design (D2-133) was improved by directed evolution using *E. coli* display and Magnetic Cell Sorting (MACS) to obtain the final protein CYN1.

The flowchart of the overall design, subsequent screening and redesign as well as final improvement of four-helix ACE2-interface mimics by directed evolution is shown in Figure 3. Specific details are described as below.

268 **Initial design of a four-helix bundle (H4) with two ACE2-Spike binding interfaces**
269
270 **Backbone Structure**: We obtained the interface residues between human ACE2
271 (hACE2) and wild-type SARS-CoV-2 Spike protein according to their complex
272 structure (PDB 6M0J) based on contact distances and the reduction of solvent
273 accessible surface area (SASA) per residue after complex formation. Most interface
274 residues in hACE2 were found located on the two N-terminal helices. After careful
275 inspection of the conformation of the two helices, we built a four-helix bundle prototype
276 by manually dimerizing the two N-term helices of hACE2. This manually built dimeric
277 structure together with binding Spike proteins was further relaxed using ROSETTA[41].
278 The relaxed structure was then compared to all existing four-helix bundle structures in
279 RCSB PDB with SP-align[30]. All structurally aligned PDBs were sorted according to the
280 SPe-alignment score and top 10 ranked structures were manually inspected and the
281 structure of the serine-rich domain from Crk-associated substrate (PDB 1Z23) was
282 found as the closest match and employed to adjust the backbone structure of the
283 prototype by modifying the relative distance and orientation between dimerized two-
284 helix bundles.
285 **Sequence Design**: We copied the conformations of ACE2 interface residues onto
286 the prototype of the four-helical bundle: residue 22-84 from ACE2 in 6M0J were used
287 to build the N-term half of the four-helix bundle, with P84 modified to G84, and
288 residues 22-88 from ACE2 in 6M0J were extracted to build the C-term half. Residues
289 85-88 were included in building the C-term half, because they mimicked well the
290 terminal cap of the template H4 in PDB 1Z23, which may contribute to the stability of
291 designed H4. The gap between the two halves was connected by building the loops of
292 various lengths (5-9) modeled using OSCAR-loop [31]. Note that residue 84 from N-term
293 and residue 22 from C-term were freely mutated during loop modeling and were
294 included in loop modelling. All residues were then re-numbered starting from 1. The
295 H4 structure was built to have two binding interfaces to interact with the Spike protein.
296 The above constructed structures with different loop lengths were further
297 optimized with AmberTools [42] and inspected with Procheck [43]. One backbone structure,

298 which has a modeled loop of 9 residues passed the check. Thus, this backbone structure
299 was employed to redesign by OSCAR-Design [32] with 2x22 fixed interface residues (N-
300 term half: E2, Q3, T6, F7, D9, K10, H13, E14, E16, D17, Y20, Q21, L24, N40, N43,
301 K47, F51, E54, Q55, L58, M61 and Y62; C-term half: E72, Q73, T76, F77, D79, K80,
302 H83, E84, E86, D87, Y90, Q91, L94, N110, N113, K117, F121, E124, Q125, L128,
303 M131 and Y132) in the presence of two fixed Spike proteins (Figure 1).

304 **Sequence Selection**: Five thousand sequences designed by OSCAR-Design were
305 clustered with CD-HIT [44] and filtered according to the area of the largest hydrophobic
306 surface-patch $S_{HPA}$, the number of low-complexity areas $N_{LCA}$ and the number of
307 atomic clashes $N_{clash}$. The areas of hydrophobic surface-patches were calculated by
308 QUILT [45]. The number of low-complexity areas was calculated by segmasker from the
309 NCBI BLAST+ package [46]. We found that there were five sequences with sequence
310 identity lower than 75% between each other, $S_{HPA} \leq 550$Å$^2$ (509 Å$^2$ for the original
311 sequence), $N_{LCA} = 0$ and $N_{clash} = 0$. These sequences (D0064, D0420, D0753,
312 D4937, and D4983) were selected for the first-round experimental test (Table S1).
313 Three of the five designs (D0753, D4937, and D4983) were expressed soluble and
314 purified (Figure S1). All have micromolar binding affinities to WT-RBD with D4983
315 having the strongest binding affinity (Figure S2). However, these three designed
316 proteins all existed in aggregated forms on a chromatographic column (Figure S3).

317 **Adding Disulfide-bond and Redesigning Sequences**: To make the monomeric
318 folding of the designed proteins more stable, we introduced one disulfide bond to the
319 backbone structure in its non-interfacial areas by using the program DbD2 (Disulfide
320 by Design) [33]. This process produced several designs with the disulfide bond added to
321 various locations. These designs were further structurally optimized and sequence-
322 redesigned in the same way as in the first-round design with the disulfide-bond fixed.
323 We collected 1000 designs and further filtered with $S_{HPA} \leq 385$Å2, $N_{LCA} = 0$, and
324 $N_{clash} = 0$. This results in six second-round sequences with disulfide bonds located at
325 CYS1-CYS133 (D2-133), CYS22-CYS48 (D2-128), CYS25-CYS44 (D2-477),
326 CYS70-CYS136 (D2-247), CYS92-CYS118 (D2-830) and CYS95-CYS114 (D2-099),
327 respectively (Figure 1 and Supplementary Table S2). Among 6 new designs (D2-099,

D2-128, D2-133, D2-247, D2-477, D2-830), only D2-133 was monomeric, D2-099, D2-477 were partly monomeric, whereas D2-128, D2-247, and D2-830 were mostly multimeric (Figure 2A, 2B and 2C). Thus, only D2-133 was employed for the binding affinity test.

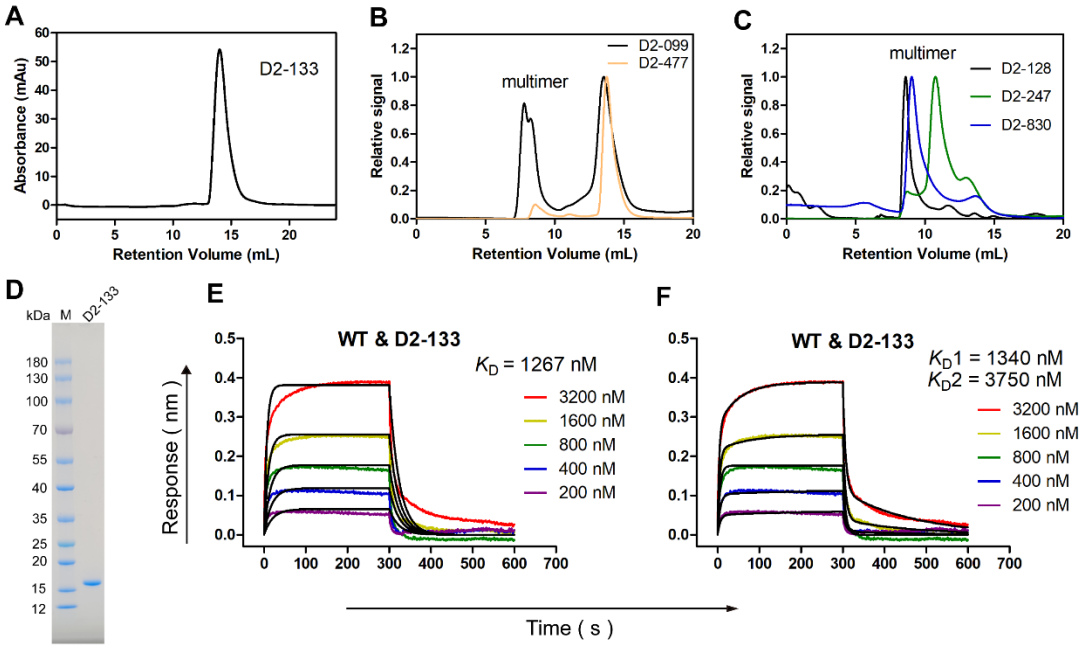



**Figure 2. Expression and purification of the second-round designs of H4 along with affinity measurement**. Size-exclusion chromatography for D2-133 (**A**), D2-099 and D2-477 (**B**), and D2-128, D2-247, and D2-830 (**C**) indicates that only the D2-133 design is purely monomeric. (**D**) Purified D2-133 was shown on a 4-20% SDS-polyacrylamide gel and stained with Coomassie brilliant blue. The binding affinity between D2-133 and WT SARS-CoV-2 Spike RBD at the micromolar level according to the Biolayer-interferometry (BLI) measurement (Fitted by 1:1 model (**E**) and 1:2 model (**F**)).

## The designed ACE2 mimic D2-133 exhibited micromolar affinity against RBD domain of Spike protein

The purified monomeric D2-133 was obtained from Size-exclusion chromatography (Figure 2D). Its binding affinity to RBD of wild type Spike was determined using BLI (Figure 2E and 2F). The D2-233 showed a micromolar affinity (1.3 μM). This affinity was similar to the AHB2 (affinities ranging from 100 nM to 2 μM) [15] or DBR3_02 ($K_D$: 4.6 μM), previously designed RBD binders [17]. Compared with the 1:1 fitting model

351 (Figure 2E), D2-133 showed a better fitting with the 1:2 fitting model (Figure 2F),
352 consistent with our inclusion of two hACE2-RBD binding interfaces from hACE2.
353
354 **Improving D2-133 by *E. coli* display and MACS selection.**
355
356 The *E. coli* display was used to improve the D2-133 affinity. To exclude the influence
357 of fimbriae on display[36], *fimA-H* was knocked out from MG1655 using
358 pEcCas/pEcgRNA system [34] and the MG1655$^{\Delta fimA\text{-}H}$ strain was successfully constructed
359 for bacterial display (see Figure S4 and Methods). EP-PCR was used to introduce
360 mutations on the D2-133 sequence for the construction of *E. coli* display library with
361 about 0.8% mutation rate.
362 The D2-133 library can bind to RBD-labeled beads but not to empty beads as
363 shown in Figure 3A (Group 1 vs Group 2). However, neither strong acid (Figure 3A,
364 Group 3) nor high salt (Figure 3A, Group 4) conditions can remove low-affinity
365 mutants from the library. Thus, we increased the selection pressure by lowering the pH
366 in round 1 (Figure 3B, Group C and D). Two clones (labeled as E-6 and E-9) from
367 round 1 were selected for expression and purification for affinity evaluation using BLI.
368 Both E-6 and E-9 have double mutations over D2-133 (K70E and K90R for E-6 and
369 N103I and E125G for E-9). Both improve over D2-133 in binding affinity to WT-RBD
370 with $K_D$ of 99 nM and 501 nM, respectively (Figure 3B and 3C).
371 Employing the best-affinity sequence E-6, we performed the second round of *E.*
372 *coli* display and MACS selection (Figure 3A). One clone CYN1 was selected. It has
373 four mutations over D2-133 (L59T, K70E, K90R, L129T). CYN1 was expressed and
374 purified in the monomeric state (Figure 3D). It has a binding affinity of 22 nM to WT
375 Spike RBD, about 60 times stronger than that of D2-133 (Figure 2 F vs Figure 3 E) and
376 4 times stronger than that of hACE2 with WT Spike RBD [47–49].
377

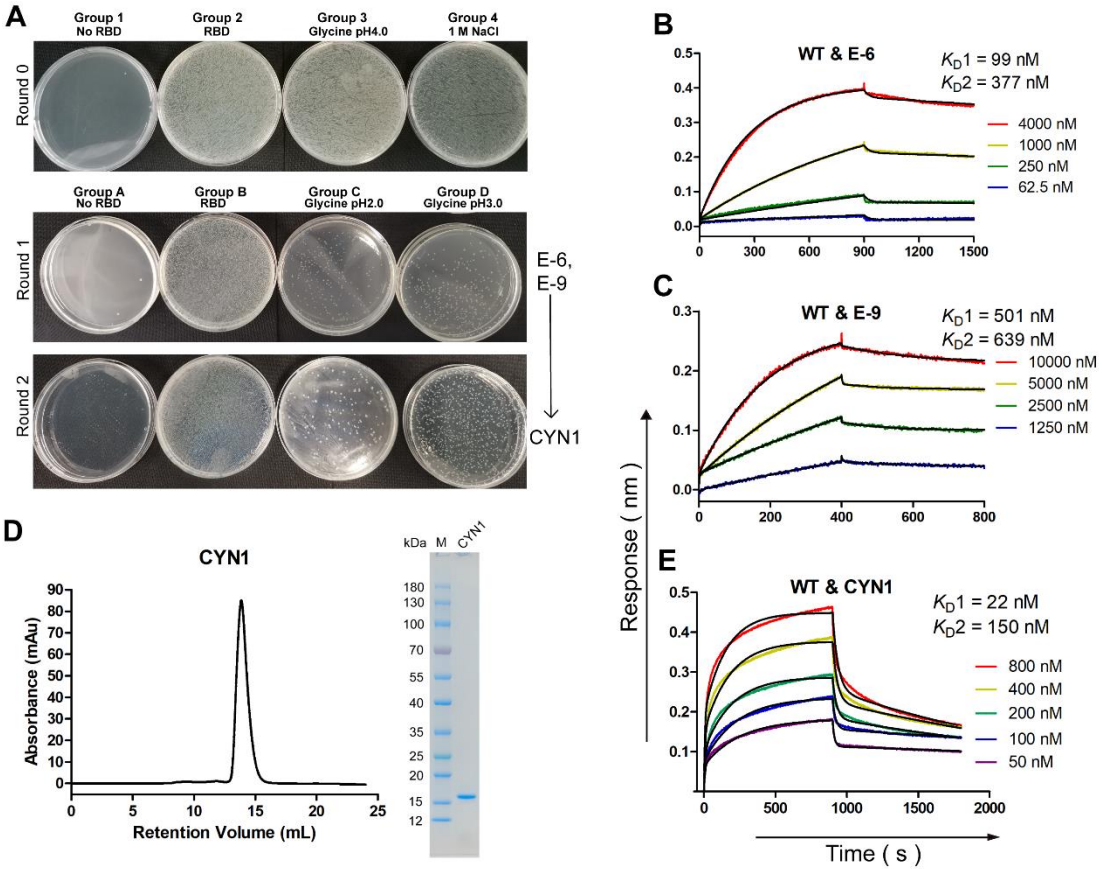



**Figure 3. Improved binding affinity of D2-133 by *E. Coli* display and MACS.** (**A**) Round 0 indicates poor selection of D2-133 mutants with pH 4.0 (Group 3) and 1M NaCl (Group 4), compared to no selection pressure (Group 2). Increased selection pressure at pH 2.0 and 3.0 in Round 1 led to E-6 and E-9. In Round 2, the selection of E-6 mutants led to CYN1. (**B**) and (**C**) The binding affinities of E-6 and E-9 with WT-RBD were measured by BLI. (**D**) Size-exclusion chromatography for CYN1. (**E**) The binding affinity of CYN1 with WT-RBD measured by BLI.

## High thermal stability of CYN1

To investigate the secondary structural content and thermal stability of D2-133 and CYN1 at far-UV region, Circular dichroism (CD) experiments were carried out using Circular Dichroism Spectrophotometer J-1500. The results of CD spectrum indicated D2-133 and CYN1 showed typical characteristic peaks of the α-helical secondary structure with negative ellipticity at 222 and 208 nm (Figure 4A and 4B), suggested

393 that their secondary structure was dominated by α-helix which was consistent with our
394 design of four-helix bundle. Although the CD intensity decreased as the temperature
395 increased from 25 ºC to 95 ºC, the negative peaks at 222 and 208 nm were retained,
396 indicating that the helical structure of the protein sample at room temperature retained
397 part of the secondary helical structure at high temperatures. Figure 4A and 4B further
398 showed the CD spectrum of the initial 25 ºC spectrum (black) compared with the
399 spectrum measured at 95 ºC (red) and at 25 ºC after the melt (green). The spectra before
400 and after refolding were very similar, indicating that the protein does refold once the
401 temperature reduced, confirming that the folding was reversible. The CD-spectrum-
402 derived helical contents according to the K2D3 web server is 52.2 % for D2-133 and
403 61.42 % for CYN1, suggesting that CYN1 has a more stable structural fold around the
404 designed helical conformation (Figure 4E, F). However, this is still lower than 91%
405 residues in helical regions for the designed protein.

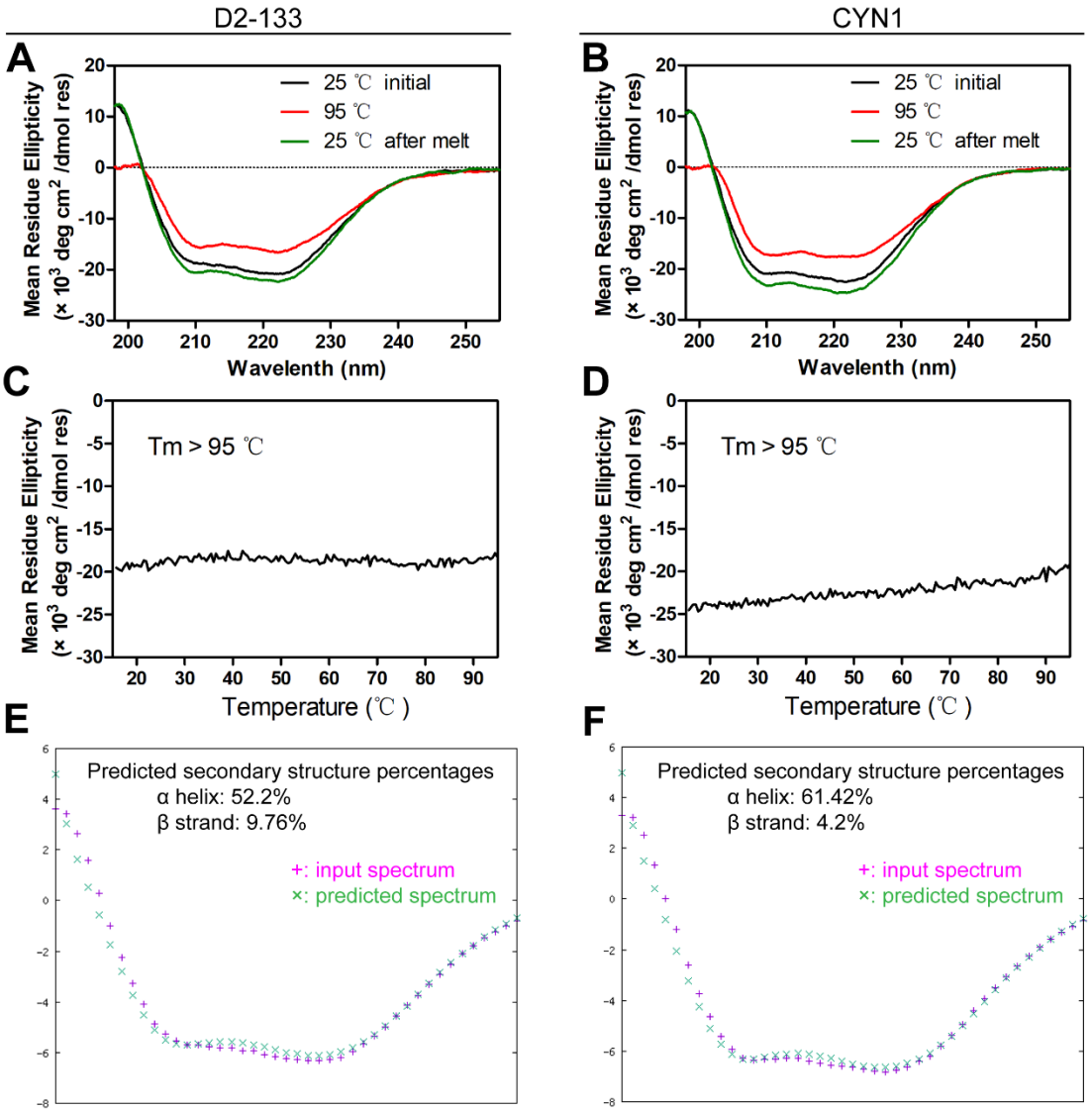
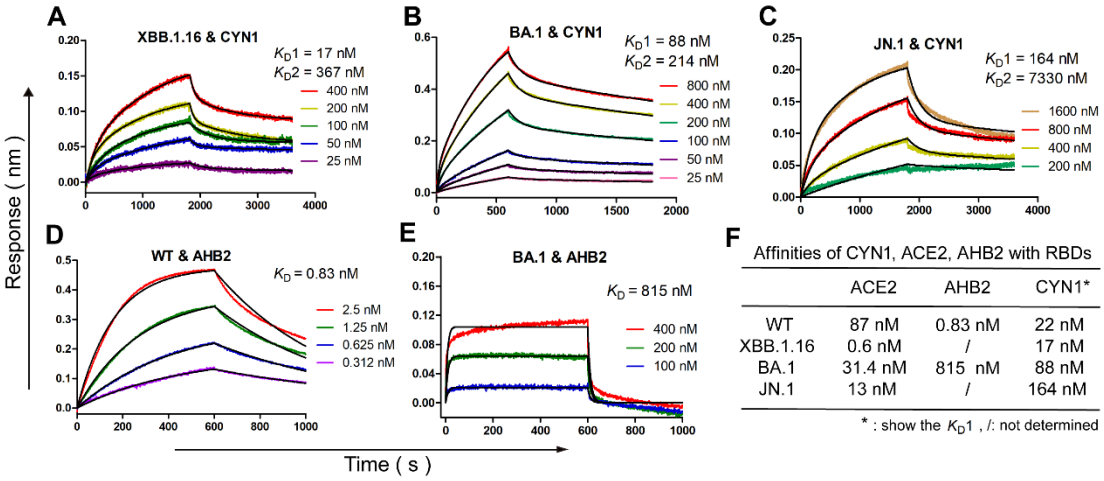



**Figure 4. The Circular Dichroism Spectrum indicates high thermal and structural stability of D2-133 and CYN1.** The Circular Dichroism Spectrum of D2-133 (**A**) and CYN1 (**B**) at 95 ºC and 25 ºC at native and refolded conditions as labeled. The CD signal of D2-133 (**C**) and CYN1 (**D**) at 222-nm wavelength, as a function of temperature. The predicted secondary structure contents of D2-133 (**E**) and CYN1(**F**) by K2D3 web server.

The CD signal strength as a function of temperature was monitored at 222 nm for D2-133 and CYN1 (Figure 4C and 4D). However, there is a lack of significant temperature-dependent changes suggesting that the melting temperatures of CYN1 and D2-133 were both greater than 95 ºC. These results suggest that CYN1 exhibits the same high thermal stability as the designed H4 protein D2-133.

420 **The CYN1 showed broad-spectrum binding capability against RBD mutants from**
421 **three different virus strains**

422 To verify whether D2-133 and CYN1 bind broadly to RBDs of different SARS-CoV-2
423 strains, the RBDs of Omicron BA.1, XBB.1.16, and JN.1 virus strains were purchased
424 for measuring binding affinity, and they had 13 mutations, 14 mutations, 27 mutations
425 to the WT-RBD, respectively (Figure 5G). The results showed that CYN1 retains its
426 activity in binding to RBDs from different virus strains (Figure 3E vs Figure 5A, 5B
427 and 5C). Compared to the wild type RBD, the binding affinity of CYN1 to XBB.1.16
428 RBD was similar, at 17 nM, while the affinities of CYN1 to BA.1 and JN.1 RBDs were
429 weakened, at 88 nM and 164 nM, respectively. The previously reported RBD binder
430 AHB2 [15] was employed as a control, and its binding affinities to WT-RBD and BA.1
431 RBD were determined. As shown in Figure 5D, the binding affinity of AHB2 to WT-
432 RBD was at 0.83 nM, consistent with the previous study [15]. However, the affinity of
433 AHB2 to the Omicron BA.1 RBD was only at 815 nM (Figure 5E), reduced by 1000-
434 fold compared to the wildtype RBD (Figure 5F). By comparison, only a 4-fold
435 reduction of binding of CYN1 to BA.1 relative to the binding to WT was observed
436 (Figure 3E). It should be emphasized that the original design D2-133 does not display
437 significant binding with the RBDs of BA.1, XBB.1.16 and JN.1 (Figure S5), despite of
438 only 4 mutations away from CYN1.
439




F

Affinities of CYN1, ACE2, AHB2 with RBDs

| | ACE2 | AHB2 | CYN1* |
|---|---|---|---|
| WT | 87 nM | 0.83 nM | 22 nM |
| XBB.1.16 | 0.6 nM | / | 17 nM |
| BA.1 | 31.4 nM | 815 nM | 88 nM |
| JN.1 | 13 nM | / | 164 nM |

* : show the $K_D1$ , /: not determined

G

Mutation sites of the RBDs

| | sites |
|---|---|
| WT | No |
| BA.1 | G339D, R346K, S371L, S373P, S375F, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H |
| XBB.1.16 | G339H, R346T, L368I, V445P, G446S, N460K, S477N, T478R, E484A, F486P, F490S, Q498R, N501Y, Y505H |
| JN.1 | I332V, G339H, K356T, S371F, S373P, S375F, T376A, R403K, D405N, R408S, K417N, N440K, V445H, G446S, N450D, L452W, L455S, N460K, S477N, T478K, N481K, V483del, E484K, F486P, Q498R, N501Y, Y505H |

**Figure 5. Broad spectrum measurement of binding to RBDs of other SARS-CoV 2 strains.** Affinity of CYN1 with RBDs, fitted by 1:2 model to XBB.1.16 (**A**), BA.1 (**B**), JN.1 (**C**). Binding affinity of AHB2 fitted by 1:1 model to WT (**D**) and to BA.1 (**E**). (**F**) Table of Affinities of ACE2 [7,47,50,51], AHB2, and CYN1 with RBDs. (G) The mutation sites of the three Omicron RBDs to WT-RBD.

## Interpretation of binding free energy by molecular dynamics simulations and free energy calculations

To understand different behavior observed between D2-133 and CYN1, we performed molecular dynamics simulations of complexes of the WT-Spike RBD with ACE2, D2-133, E6, E9, and CYN1 and observed quite stable binding interfaces with the root mean square deviation (RMSD) values of the designed binding interfaces stable less than 3 Å for both 1-2 and 3-4 helices (Figures 6A and 6B). We then employed the ASGBIE method to calculate the interaction free energy between all residues within 10 Å of the binding interface and the other protein molecule. Fig. 6C shows a strong correlation with a Pearson correlation coefficient at 0.97 between experimental determined free energies and ASGBIE-calculated results according to simulated conformations. This consistency provides the basis for the following hot-spot analysis by computational alanine scanning.

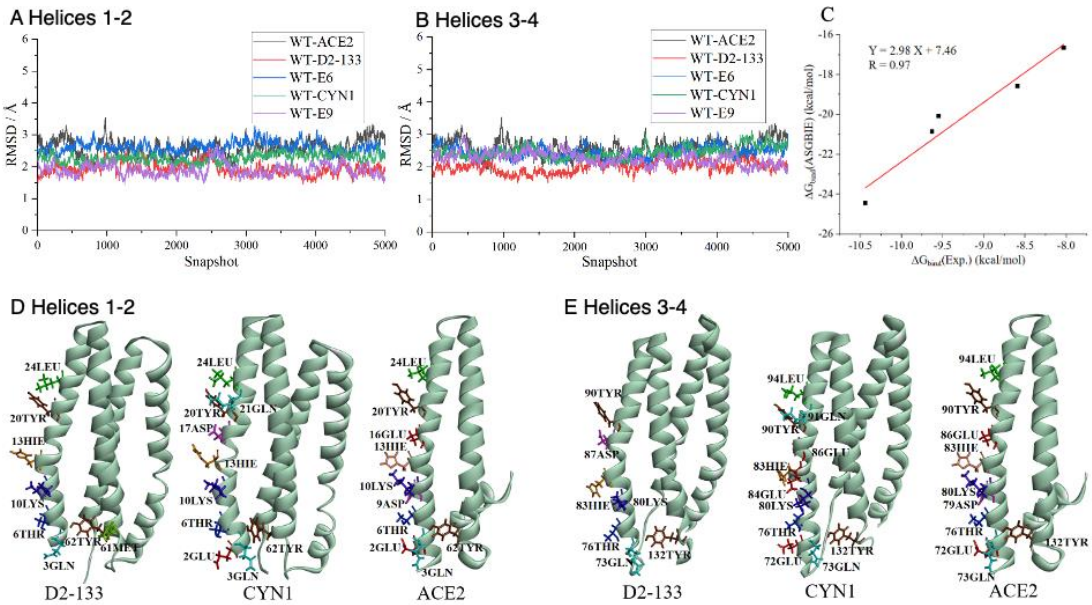



**Figure 6.** (A) Root-mean-squared deviation of helices 1-2 interface from the respective starting conformations by molecular dynamics simulations of ACE2, D2-133, E6, CYN1, and E9 in complex with a single WT-RBD of 5000 snap shots of 10 ns production runs after 40 ns equilibration run. (B) as in (A) but for helices 3-4 interface. (C) Calculated versus experimentally measured binding free energies of ACE2, D2-133, E6, CYN1, and E9 in complex with a single WT-RBD. (D) Comparison of hotspot residues identified by computational alanine scanning in helices 1-2 interface of D2-133 and CYN1 with the hotspot residues of ACE2 (E) As in (D) but for helices 3-4 interface. The sequence number in ACE2 was changed to match with those designed proteins for comparison.

**Hotspot residues and their implication on broad-spectrum capability**

For binding of ACE2 to WT-RBD, we found a total of 8 hotspot residues, ranked by their contributions as follows: 20TYR, 13HIE, 10LYS, 16GLU, 62TYR, 3GLN, 2GLU, and 6THR. These hotspot residues were included as a part of fixed interfacial residues for designing ACE2 mimics. For our designed proteins, we have two surface regions mimicking the ACE2 interface to WT-RBD (helices 1-2 and helices 3-4). Thus, we performed two separated calculations to identify hotspots in two dual-helix regions. As shown in Figures 6D-E and Supplementary Tables S3-4), D2-133 has 6 out of 8

481 hotspot residues in the helices 1-2 interface and 6 out of 7 hotspot residues in helices
482 3-4 interface overlapping with ACE2 hotspot residues. By comparison, CYN1 has more
483 hotspot residue interactions with WT-RBD with stronger binding affinities (10 for helix
484 1-2 and 11 for helix 3-4). Moreover, these hotspots overlap more with 8 ACE2 hotspot
485 residues (7 for helix 1-2 and 8 for helix 3-4). This improved recovery of ACE2 hotspot
486 residues by CYN1 explains its broader spectrum capability over D2-133.

## 487 Discussion

488 SARS-CoV-2 has a fast mutation rate because it is an RNA virus with billions of hosts
489 [52,53]. Much research shows that the spike's mutation was the most important factor for
490 SARS-CoV-2 to escape the vaccine protection developed by pharmaceutical companies
491 [3,4,6,7,54]. The spike's mutations also lead to resistance to antibody drugs because many
492 antibody drugs were based on antigens on the spike proteins that are subsequently
493 mutated [55,56]. Here, we examined the possibility of designing a broad-spectrum protein
494 by copying the same hACE2 interface that can bind to the RBDs of the different viral
495 strains.

496 Despite the simplicity of the above idea, the design progress is not as
497 straightforward as originally planned. In the first-round design, 2 out of 5 top-ranked
498 designs did not have soluble expression, despite we have chosen the sequences with the
499 lowest continuous hydrophobic patch area, indicating that the largest hydrophobic
500 patch area is not necessarily the best indicator for solubility. Nevertheless, three of the
501 five designs were expressed soluble and purified (Figure S1) and all bind with WT-
502 RBD with micromolar binding affinity (Figure S2). Although the binding affinity is
503 much lower than the binding between ACE2 and WT-RBD, the rate of success in
504 achieving binding by OSCAR-Design is high (60%).

505 However, these three expressed proteins existed in aggregated forms on the
506 chromatographic column (Figure S3). The difficulty of expression and aggregation
507 faced here is likely due to the high hydrophobic patch area of the ACE2 binding
508 interface and the need to maintain two large hydrophobic patches for our design with
509 two ACE2 binding interfaces. To address this problem, we incorporated a disulfide

510 bond in our design. All six designs are now expressed and soluble. Moreover, three of
511 the six designs tested now exist in significant monomeric form with D2-133 purely
512 monomeric, corresponding well to the fact that it has the smallest hydrophobic patch
513 area (361Å$^2$) and two with the largest hydrophobic patch areas (382 Å$^2$ and 384 Å$^2$) are
514 all in multimeric form.

515 D2-133, however, continues to bind at the micromolar level. Moreover, it does not
516 bind to RBDs of other SARS-CoV-2 strains. This indicates that the D2-133 interface
517 grafted from ACE2 for RBD binding is only partially functional. That is, D2-133 must
518 not have all the original ACE2 interfacial residues in the right place to interact with
519 Spike RBD because it has a weak affinity to WT-RBD and no affinity to the RBDs of
520 other viral strains.

521 D2-133 was improved by a factor of 13 in binding affinity to WT-RBD with one
522 round of MACS selections. However, the only two mutations (K70E and K90R) in E6
523 are unexpectedly all outside the pre-defined ACE2-mimic interface region. The second
524 round of MACS selections added two more mutations. This time both mutations are in
525 the pre-defined ACE2 interfacial regions (L59T and L129T), but mutations are from
526 hydrophobic leucine to more hydrophilic threonine. One would expect the opposite as
527 less hydrophobic interface would lead to lower binding affinity.

528 To gain some understanding why these mutations yielded higher affinity and
529 improved resistance against RBD mutations, we conducted molecular dynamics
530 simulations on all systems to investigate the microstructural features of binding. Both
531 the 1-2 and 3-4 helical regions were simulated as binding sites, respectively. The
532 binding free energies of four-helix structures with WT-RBD obtained by using the
533 ASGBIE method have a near perfect correlation to experimental measured results. A
534 further hotspot analysis indicates that CYN1's hotspot residues (both helices 1-2 and
535 helices 2-4 regions) have more overlap with ACE2 hotspot residues than D2-133,
536 suggesting the likely reason behind improved broad-spectrum capability than D2-133.

537 CYN1 has not achieved a perfect broad-spectrum capability to constantly evolving
538 SARS-CoV-2. This is because the binding affinity of CYN1 to other strains' RBDs are
539 lower than those of hACE2 [7,57,58]. Moreover, there are many strains' RBDs we have not

tested yet due to the huge experiment cost. We had tried to determine the RBD-CYN1 complex structure, but the RBD-CYN1 complex was unstable at high concentration, which maybe contributed by the CYN1's high hydrophobic area or the 1: 2 binding style to RBD. Nevertheless, we demonstrated that mimicking the small interface of ACE2 has the potential to achieve broad-spectrum capability. We expected that this can be improved with constantly improved AI-based protein design techniques so that the interface can be more faithfully captured.

## Author Contribution

**Jian Zhan and Yaoqi Zhou:** Conceptualization, Investigation, Supervise, Writing-review & editing. **Qinghui Nie:** Protein expression and purification, *E. coli* display to evolve ACE2 mimic's affinity, Determined thermostability and affinity, Writing the original draft, Writing- review & editing. **Hongmei Yin**: Sequence design of the mimics, Writing the original draft. **Ke Chen**: Design pipeline development, backbone design of the ACE2 mimics, Writing the original draft. **Chao Zhao**: Performed computational simulations and binding analysis, Writing the original draft. **Weiwei Li:** sample preparation for structure determination. **Jianxun Qi**: Supervising, Writing review & editing**.** **Jinle Tang:** Supervise, Writing- review & editing. **John Z.H. Zhang:** Supervise, review & editing.

## Acknowledgements

The author would like to thank **Dr. Haitao Yang** and his doctoral student **Yuchi Lu** for their effort for helping determination of the RBD-CYN1 complex structure by Crystallization. This research paper was supported by the National Natural Science Foundation of China (Grant No. 92370202, 22333006, 92270001) and the National Key R&D Program of China (Grant No. 2021YFF1200400). We gratefully acknowledge the High Performance Computing Cluster at Shenzhen Bay Laboratory for participating in completing this study.

**Conflict of Interest**

All authors declare no financial interest. Zhan and Zhou are the CEO and the chair of the scientific advisor board for Ribopeutic, respectively.

**Reference**

1. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022).

2. Wang, Q. *et al.* Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature* **608**, 603–608 (2022).
3. Liu, L. *et al.* Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **602**, 676–681 (2022).
4. Cao, Y. *et al.* BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **608**, 593–602 (2022).
5. Tong, Z. *et al.* Deciphering a reliable synergistic bispecific strategy of rescuing antibodies for SARS-CoV-2 escape variants, including BA.2.86, EG.5.1, and JN.1. *Cell Rep.* **43**, 114338 (2024).
6. Dejnirattisai, W. *et al.* SARS-CoV-2 Omicron-B.1.1.529 leads to widespread escape from neutralizing antibody responses. *Cell* **185**, 467-484.e15 (2022).
7. Yamasoba, D. *et al.* Virological characteristics of the SARS-CoV-2 omicron XBB.1.16 variant. *Lancet Infect. Dis.* **23**, 655–656 (2023).
8. Cao, Y. *et al.* Rational identification of potent and broad sarbecovirus-neutralizing antibody cocktails from SARS convalescents. *Cell Rep.* **41**, 111845 (2022).
9. Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261–1265 (2020).
10. Urano, E. *et al.* An inhaled ACE2 decoy confers protection against SARS-CoV-2 infection in preclinical models. *Sci. Transl. Med.* **15**, eadi2623 (2023).
11. Lu, M. *et al.* Broadly Effective ACE2 Decoy Proteins Protect Mice from Lethal SARS-CoV-2 Infection. *Microbiol. Spectr.* **11**, e0110023 (2023).
12. Liu, P., Xie, X., Gao, L. & Jin, J. Designed variants of ACE2-Fc that decouple anti-SARS-CoV-2 activities from unwanted cardiovascular effects. *Int. J. Biol. Macromol.* **165**, 1626–1633 (2020).
13. Tanaka, S. *et al.* An ACE2 Triple Decoy that neutralizes SARS-CoV-2 shows enhanced affinity for virus variants. *Sci. Rep.* **11**, 12740 (2021).
14. Wang, K., Gheblawi, M. & Oudit, G. Y. Angiotensin Converting Enzyme 2: A Double-Edged Sword. *Circulation* **142**, 426–428 (2020).
15. Cao, L. *et al.* De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
16. Linsky, T. W. *et al.* De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science* **370**, 1208–1214 (2020).
17. Gainza, P. *et al.* De novo design of protein interactions with learned surface fingerprints. *Nature* **617**, 176–184 (2023).
18. Sanner, M. F., Olson, A. J. & Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **38**, 305–320 (1996).
19. Massova, I.; Kollman, P. A., Computational alanine scanning to probe protein-protein interactions: A novel approach to evaluate binding free energies. J. Am. Chem. Soc. 1999, 121, 8133-8143.
20. Kollman, P. A. *et al.* Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000).

620 21. Duan, L., Liu, X. & Zhang, J. Z. H. Interaction Entropy: A New Paradigm for
621 Highly Efficient and Reliable Computation of Protein-Ligand Binding Free Energy. *J.*
622 *Am. Chem. Soc.* **138**, 5722–5728 (2016).
623 22. Yan, Y., Yang, M., Ji, C. G. & Zhang, J. Z. H. Interaction Entropy for
624 Computational Alanine Scanning. *J. Chem. Inf. Model.* **57**, 1112–1122 (2017).
625 23. Liu, X., Peng, L., Zhou, Y., Zhang, Y. & Zhang, J. Z. H. Computational Alanine
626 Scanning with Interaction Entropy for Protein-Ligand Binding Free Energies. *J.*
627 *Chem. Theory Comput.* **14**, 1772–1780 (2018).
628 24. He, L. *et al.* Study of SHMT2 Inhibitors and Their Binding Mechanism by
629 Computational Alanine Scanning. *J. Chem. Inf. Model.* **59**, 3871–3878 (2019).
630 25. Liu, X., Peng, L. & Zhang, J. Z. H. Accurate and Efficient Calculation of Protein-
631 Protein Binding Free Energy-Interaction Entropy with Residue Type-Specific
632 Dielectric Constants. *J. Chem. Inf. Model.* **59**, 272–281 (2019).
633 26. Song, J., Qiu, L. & Zhang, J. Z. H. An efficient method for computing excess free
634 energy of liquid. *Sci. China Chem.* **61**, 135–140 (2018).
635 27. Sun, H. *et al.* Assessing the performance of MM/PBSA and MM/GBSA methods.
636 7. Entropy effects on the performance of end-point binding free energy calculation
637 approaches. *Phys. Chem. Chem. Phys. PCCP* **20**, 14450–14460 (2018).
638 28. Cong, Y. *et al.* Entropic effect and residue specific entropic contribution to the
639 cooperativity in streptavidin-biotin binding. *Nanoscale* **12**, 7134–7145 (2020).
640 29. Cong, Y. *et al.* Anchor-Locker Binding Mechanism of the Coronavirus Spike
641 Protein to Human ACE2: Insights from Computational Analysis. *J. Chem. Inf. Model.*
642 **61**, 3529–3542 (2021).
643 30. Yang, Y., Zhan, J., Zhao, H. & Zhou, Y. A new size-independent score for
644 pairwise protein structure alignment and its application to structure classification and
645 nucleic-acid binding prediction. *Proteins* **80**, 2080–2088 (2012).
646 31. Liang, S., Zhang, C., Sarmiento, J. & Standley, D. M. Protein Loop Modeling
647 with Optimized Backbone Potential Functions. *J. Chem. Theory Comput.* **8**, 1820–
648 1827 (2012).
649 32. Liang, S., Li, Z., Zhan, J. & Zhou, Y. De novo protein design by an energy
650 function based on series expansion in distance and orientation dependence.
651 *Bioinforma. Oxf. Engl.* **38**, 86–93 (2021).
652 33. Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for
653 disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346 (2013).
654 34. Li, Q. *et al.* A modified pCas/pTargetF system for CRISPR-Cas9-assisted
655 genome editing in Escherichia coli. *Acta Biochim. Biophys. Sin.* **53**, 620–627 (2021).
656 35. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for
657 CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* **46**,
658 W242–W245 (2018).
659 36. Salema, V. *et al.* Selection of single domain antibodies from immune libraries
660 displayed on the surface of E. coli cells with two β-domains of opposite topologies.
661 *PloS One* **8**, e75126 (2013).
662 37. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary
663 structure. *Nat. Protoc.* **1**, 2876–2890 (2006).

38. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins* **80**, 374–381 (2012).
39. Case, D. A. *et al.* AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).
40. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-integration of cartesian equations of motion of a system with constraints-molecular-dynamics of N-alkanes. J. Comput. Phys. 1977, 23, 327-341.
41. Conway, P., Tyka, M. D., DiMaio, F., Konerding, D. E. & Baker, D. Relaxation of backbone bond geometry improves protein energy landscape modeling. *Protein Sci. Publ. Protein Soc.* **23**, 47–55 (2014).
42. D.A. Case et al. AMBER 2020. University of California (2020).
43. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography* https://doi.org/10.1107/S0021889892009944 (1993) doi:10.1107/s0021889892009944.
44. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinforma. Oxf. Engl.* **22**, 1658–1659 (2006).
45. Lijnzaad, P., Berendsen, H. J. & Argos, P. A method for detecting hydrophobic patches on protein surfaces. *Proteins* **26**, 192–203 (1996).
46. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
47. Xu, Z. *et al.* Binding and structural basis of equine ACE2 to RBDs from SARS-CoV, SARS-CoV-2 and related coronaviruses. *Nat. Commun.* **13**, 3547 (2022).
48. Lin, S. *et al.* Characterization of SARS-CoV-2 Omicron spike RBD reveals significantly decreased stability, severe evasion of neutralizing-antibody recognition but unaffected engagement by decoy ACE2 modified for enhanced RBD binding. *Signal Transduct. Target. Ther.* **7**, 56 (2022).
49. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
50. Han, P. *et al.* Receptor binding and complex structures of human ACE2 to spike RBD from omicron and delta SARS-CoV-2. *Cell* **185**, 630-640.e10 (2022).
51. Jian, F. *et al.* Evolving antibody response to SARS-CoV-2 antigenic shift from XBB to JN.1. *Nature* **637**, 921–929 (2025).
52. Hu, B., Guo, H., Zhou, P. & Shi, Z.-L. Author Correction: Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol.* **20**, 315 (2022).
53. Steiner, S. *et al.* SARS-CoV-2 biology and host interactions. *Nat. Rev. Microbiol.* **22**, 206–225 (2024).
54. Planas, D. *et al.* Distinct evolution of SARS-CoV-2 Omicron XBB and BA.2.86/JN.1 lineages combining increased fitness and antibody evasion. *Nat. Commun.* **15**, 2254 (2024).
55. Hattori, T. *et al.* The ACE2-binding Interface of SARS-CoV-2 Spike Inherently Deflects Immune Recognition. *J. Mol. Biol.* **433**, 166748 (2021).
56. Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).

57. Li, L. *et al.* Structural basis of human ACE2 higher binding affinity to currently circulating Omicron SARS-CoV-2 sub-variants BA.2 and BA.1.1. *Cell* **185**, 2952-2960.e10 (2022).
58. Kaku, Y. *et al.* Virological characteristics of the SARS-CoV-2 JN.1 variant. *Lancet Infect. Dis.* **24**, e82 (2024).