SARS-CoV-2 RBD *in vitro* evolution follows contagious mutation spread, yet generates an able infection inhibitor

Jiří Zahradník[1], Shir Marciano[1], Maya Shemesh[1], Eyal Zoler[1], Jeanne Chiaravalli[2], Björn Meyer[3] Orly Dym[4], Nadav Elad[5] and Gideon Schreiber[1,6]

[1] Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel

[2] Chemogenomic and Biological Screening Core Facility Institut Pasteur, 75724 Paris, France

[3] Viral Populations and Pathogenesis Unit CNRS UMR 3569 Institut Pasteur, 75724 Paris, France

[4] Department of Life Sciences Core Facilities, Weizmann Institute of Science, Rehovot 7610001, Israel

[4] Department of Chemical Research Support, Weizmann Institute of Science, Rehovot 7610001, Israel

[6] Corresponding author: gideon.schreiber@weizmann.ac.il

Short Title: RBD *in vitro* evolution

Abstract

SARS-CoV-2 is constantly evolving, with more contagious mutations spreading rapidly. Using *in vitro* evolution to affinity maturate the receptor-binding domain (RBD) of the spike protein towards ACE2, resulted in the more contagious mutations, S477N, E484K, and N501Y to be among the first selected. This includes the British and South-African variants. Plotting the binding affinity to ACE2 of selected RBD mutations against their incidence in the population shows a strong correlation between the two. Further *in vitro* evolution enhancing binding by 600-fold provides guidelines towards potentially new evolving mutations with even higher infectivity. Yet, the high-affinity RBD is also an efficient drug, inhibiting SARS-CoV-2 infection. The 2.9Å Cryo-EM structure of the high-affinity complex, including all rapidly spreading mutations provides structural basis for future drug development.

SARS-CoV-2, which causes COVID-19, resulted in an epidemic of global reach. It infects people through inhalation of viral particles, airborne, in droplets, or by touching infected surfaces. Structural and functional studies have shown that a single receptor-binding domain (RBD) of the SARS-CoV-2 homotrimer spike glycoprotein interacts with ACE2, which serves as its receptor (*1, 2*). Its binding and subsequent cleavage by the host protease TMPRSS2 results in the fusion between cell and viral membranes and cell entry (*1*). Blocking the ACE2 receptors by specific antibodies voids viral entry (*1, 3, 4*). *In vitro* binding measurements have shown that SARS-CoV-2 S-protein binds ACE2 with ~10 nM affinity, which is about 10-fold tighter compared to the binding of the SARS-CoV S-protein (*2, 3, 5*). It has been suggested that the higher affinity of SARS-CoV-2 is, at least partially, responsible for its higher infectivity (*6*). Recently evolved SARS-CoV-2 mutations in the Spike protein´s RBD have further strengthened this hypothesis. The "British" mutation (N501Y; variant B.1.1.7) was suggested from deep sequencing mutation analysis to enhance binding to ACE2 (*6*). The "South African" variant (501.V2), which includes three altered residues in the ACE2 binding site (K417N, E484K, and N501Y) is spreading extremely rapidly, becoming the dominant lineage in the Eastern Cape and Western Cape Provinces (*7*). Another variant that seems to enhance SARS-CoV-2 infectivity is S477N, which became dominant in many regions (*8*). ACE2 and TMPRSS2 express in lung, trachea, and nasal tissue (*9, 10*). The inhaled virus likely binds to epithelial cells in the nasal cavity and starts replicating. The virus propagates and migrates down the respiratory tract along the conducting airways, and a more robust innate immune response is triggered, which in some cases leads to severe disease. Recently, a number of efficient vaccines, based on presenting the spike protein or by administrating an inactivated virus were approved for clinical use (*11*). Still, due to less than 100% protection, particularly for high-risk populations and the continuously mutating virus, the development of drugs should continue. Potential therapeutic targets blocking the viral entry in cells include molecules blocking the spike protein, the TMPRSS2 protease, or the ACE2 receptor (*12*). Most prominently, multiple high-affinity neutralizing antibodies have been developed (*13*). Alternatives to the antibodies, the soluble forms of the ACE2 protein (*14*) or engineered parts or mimics have also been shown to work (*15, 16*). TMPRSS2, inhibitors were already previously developed, and are repurposed for COVID-19 (*1*). The development of molecules blocking the ACE2 protein did not receive as much attention as the other targets. One potential civet with this approach is the

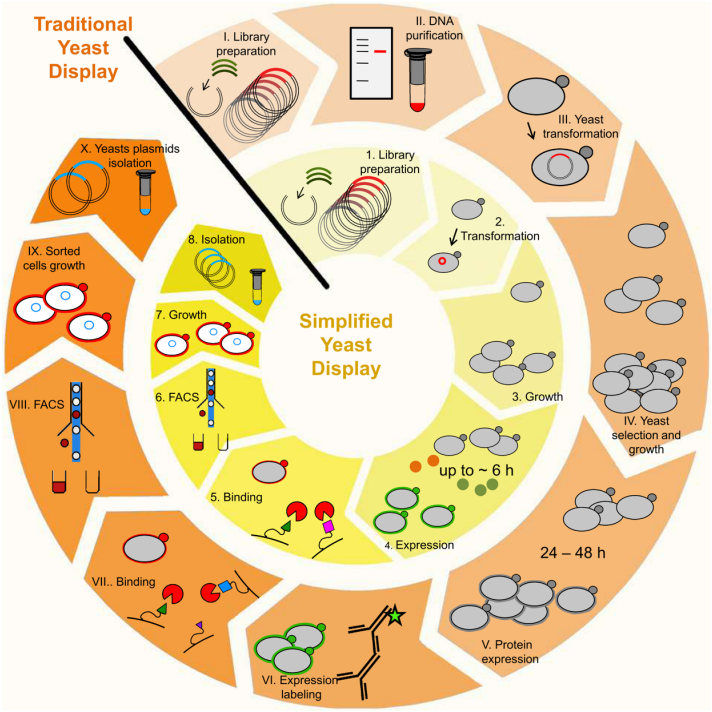

Fig. 1 Enhanced yeast display benefits over traditional method. The use of enhanced yeast display enables elimination of DNA purification procedures between libraries (step II.); exclusion of antibody-based expression labeling procedure (step VI.), and the bright reporters eUnaG2 (orange points, step 4.) or DnbALFA (green points, step 4.) allow for ultra-tight binding selection, with reduced background and increased sensitivity in a reduced time frame (*20*).

importance of the ACE2 activity in humans, which could be hampered by an inhibitor. ACE2 functions as a carboxypeptidase, removing a single C-terminal amino acid from Ang II to generate Ang-(1-7), which is important in blood pressure regulation. In addition, ACE2 is fused to a collectrin-like domain, regulating amino acid transport and pancreatic insulin secretion (*17, 18*). Through these processes, ACE2 also appears to regulate inflammation, which downregulation relates to increased COVID-19 severity. Dalbavancin is one drug that has been shown to block the spike protein–ACE2 interaction, however with low affinity (~130 nM) (*19*).

Notably, the RBD domain itself can be used as a competitive inhibitor of the ACE2 receptor binding site. However, for this to work, its affinity has to be significantly optimized, to reach pM affinity. We have recently developed an enhanced strategy for yeast display, based on C and N-terminal fusions of extremely bright fluorescence colors that can monitor expression at minute levels, allowing for selection to proceed down to pM bait concentrations (*20*). Here, we demonstrate, how this enhanced method allowed us to reach pM affinity between a mutant RBD and ACE2, based on multiple-steps of selection that combine enhanced binding with increase RBD protein thermostability. Fig. 1 shows step by step the selection process. We took advantage of using two different detection strategies, eUnaG2 and DnbALFA, and eliminate the DNA purification step, which can be tedious (*20, 21*) (Fig. 1, steps 2 and 4). Preceding library construction, we tested varied sizes of the RBD, for optimal surface expression (Table S1), and decided to continue using RBDcon2 for selection and RBDcon3 for protein expression (Supplementary Material Text).

RBD domain affinity maturation recapitulates multiple steps in the virus evolution

A

RBD residue	354	358	445	446	448	460	468	470	477	478	481	483	484	490	493	494	498	501
RBD wild-type	N	I	V	G	N	N	I	T	S	T	N	V	E	F	Q	S	Q	N
Library S2		F																
Library B3		F			S		T		N		Y	E	K	S				Y
Library B4		F		R		K	V		N	S			K	Y		P	R	Y
Library B5		F		R		K	V		N				K	Y	H	P	R	Y
Library FA	E	F	K			K	T	M	N				K			P	R	Y
Clone B62		F	K			K	T	M	N				K				R	Y
Circulating variants			A,I,F	A,D,V, S					N,I,R, G,T,K	A,I,R,K			K,Q,A,D, G,R,V	S,L,V	L,K,H, R	P,L,A	H,P	Y,R,T, S
Location/ lineage or cluster									Europe/ GR and GH				JAR 501Y.V2					UK B.1.1.7 JAR 501Y.V2)

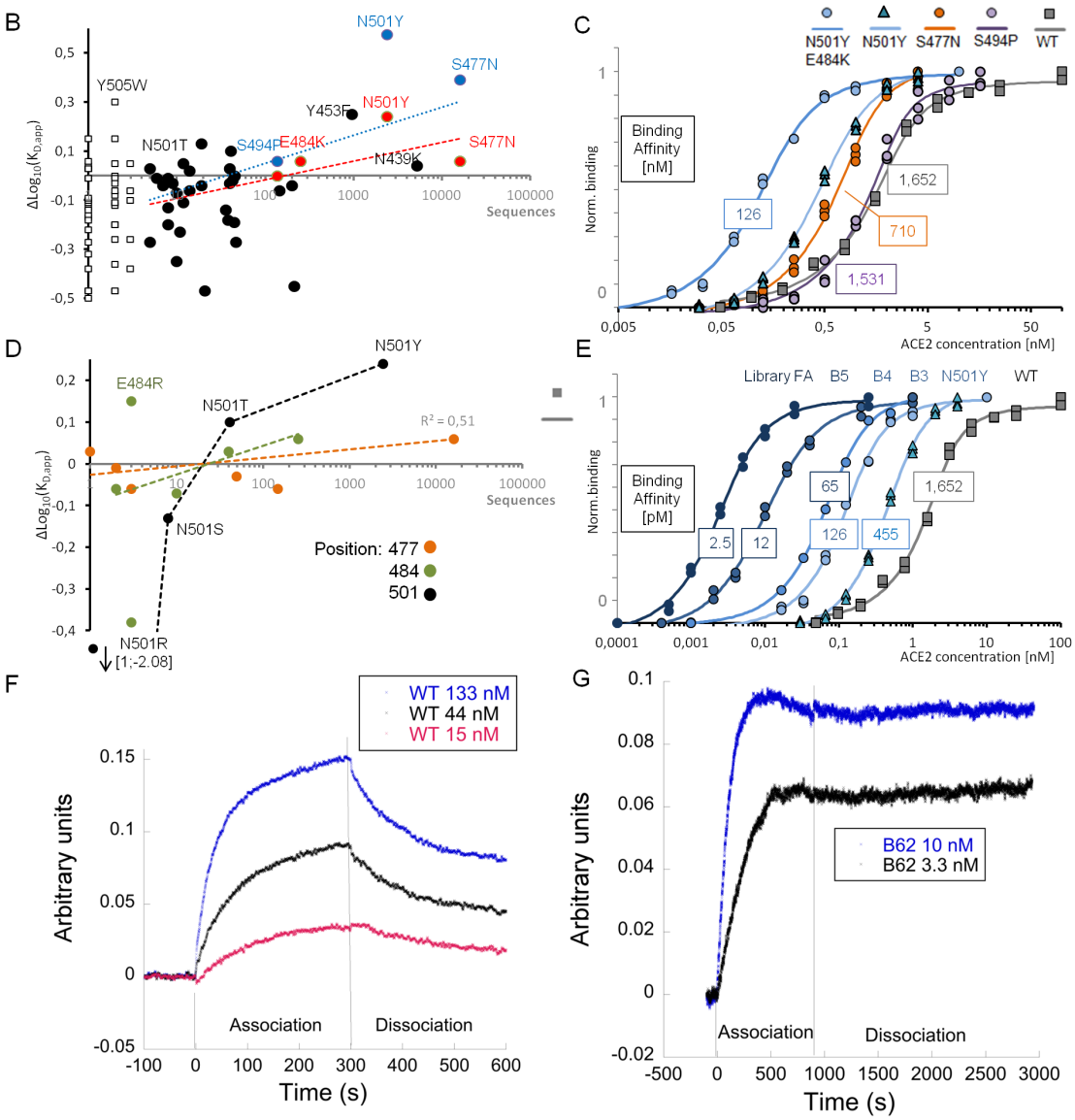

Fig. 2 *In vitro* evolution of Spike protein RBD using yeast display and the emergence of mutations in SARS-CoV-2 over time. (A) An overview of mutations identified during the yeast display affinity maturation process. The red and grey colored amino-acids are dominant (>50 %) or minor (<50 %) at a given position. Red and orange background highlight the emerging mutations both in clinical samples and yeast display, with a high and low impact to binding affinity, respectively. The bottom of the table shows the naturally evolved mutations at the same positions. (B) The relation between inferred affinity changes and occurrence. Red for prevalent mutations, black for others and empty squares for occurrence <5 sequences (*6,22*). Blue dots are values from binding titration curves shown in (C), which were selected also by yeast display affinity maturation. (D) Affinity changes and occurrence in population (as in (B) for different mutations at positions 477, 484 and 501. (E) Binding titration curves for the best binding variant in each successive yeast library, bound to ACE2-CF640R at the given concentration. Binding of additional clones from each

Multiple consecutive libraries were constructed, S - stability-enhancing, B - ACE2 binding, and FA for the fast association. The whole *RBD* library (S1, nucleotides -152 – 621) was constructed by random mutagenesis, introducing 1-5 mutations per clone. The best expressing clones were selected after expression at 30 °C. Subsequently, library S2 was selected after expression at 37 °C. The most significant mutation, which dominated the second library was I358F, which nicely fits inside the hydrophobic pocket formed in the RBD domain (Fig. S1). This mutation led to nearly doubling the fluorescence signal intensity and was used for the construction of the subsequent affinity selection library (B3). B3 library was constructed by 3 components homologous recombination to preferentially incorporate the mutations in the binding interface area. The random mutagenesis was limited to nucleotides 260 – 621. The library was expressed at 37 °C to keep the pressure on protein stability, and selected by FACS sorter against decreasing concentration of ACE2 labeled with CF®640R succinimidyl ester (1000, 800, and 600 pM; 4 h of incubation). To isolate a low number of RBD variants with the strongest phenotype effect, library enrichment was done by selecting the top 3% of binding cells and in subsequent rounds, the top 0.1 – 1% yeast cells (Fig. S2). Plasmid DNA was isolated from growing selected yeasts of the sorted lib rary and used for *E.coli* cell transformation and the preparation of a new library (B4). This approach enriched the subsequent library with multiple selected mutations and enabled the screening of wider sequence-space and cooperative mutations, as multiple trajectories are sampled. 30 single colony isolates (SCI) of transformed bacteria were used for sequencing to monitor the enrichment process and subsequently for binding affinity screening (Fig. S3). The library B4 was selected with the same schema using 600, 400, and 200 pM ACE2 receptor. Analysis of the selected B3 library yielded two dominant mutations appearing at > 70 % of clones: E484K and N501Y. In addition, multiple minor mutations: V483E, N481Y, I468T, S477N, N448S, and F490S were found (Fig. 2A). The analysis of library B4 showed the absolute domination of E484K and N501Y. Besides the dominant clones, the N460K, Q498R, and S477N mutations rose to frequencies > 20 %, and new minor populated mutations were identified: G446R, I468V, T478S, F490Y, and S494P. To validate our results, we choose clones with different mutation profiles, expressed them in Expi293F™ cells, and subjected them for further analyses (Figs. 2A, C, E, S4, Table 1 and Table S2).

We noted, that among the mutations selected and fixed in the yeast population during these initial steps of affinity maturation were three mutations that strongly emerged in clinical samples of SARS-CoV-2: S477N, E484K, and N501Y (*6*, *7*, *8*). It was already shown that an increase in RBD binding affinity increases pseudovirus entry (*6*). To validate the relation

between binding affinity and occurrence of specific mutations in the population, we combined our data with those obtained by deep-mutational scanning of the RBD domain (*6*), and the GISAID database (*22*). Fig. 2A shows the mutations selected by us, and the evolving, circulating SARS-CoV-2 variants at the same positions. In red are the most common mutations emerging in either (S477N, E484K, and N501Y). In addition, the yeast selection probed the most abundant naturally occurring variants in positions 490 and 493, which were lost during further rounds of yeast selection. S494P, which also occurs in nature but did not rapidly spread (Fig. 2B), was selected in round 4. Further analysis of this clone (Table 1, compare RBD-52 to RBD-521) shows it to increase the thermostability but decrease the association rate constant of the RBD to ACE2. Finally, some mutations were found in SARS-CoV-2 (albeit at low frequency, Fig. 2A and B) and not in the yeast selection (445, 446, 478, and 498). To evaluate why some mutations were prevalent in SARS-CoV-2 and in yeast display selection, while others were not, we plotted the occurrence of all mutations in the GISAID database (*22*) in respect to the apparent change in the RBD-ACE2 binding affinity (K_{Dapp}) as estimated by the frequency of given amino acids within a mutant library at the given concentration (so-called deep mutational scanning approach (*6*)). Figure 2B (red and black dots) shows that the more prevalent mutations have a higher binding affinity. To quantify these results, we measured the binding of re-cloned isogenic variants of the most prevalent mutations (Fig. 2C). The here calculated K_D values are shown as blue dots in Fig. 2B. The highest binding affinity was measured for the South-African variant (E484K, N501Y), which is the tightest binding clone of library B3 (Fig. 2C and Table S2), followed by the "British" (N501Y) and the European emerging S477N mutations (Figs. 2B, C and Table 1). The K_D of the South-African variant is 126 pM, the British 455 pM and for S477N a K_D of 710 pM was measured (compared to 1.6 nM for the WT). The here measured affinity data show an even stronger relation between binding affinity and spread in the population. To further test the lack of randomness in the selection of these mutations, we compared the occurrence of mutations for these three residues to other amino-acids in the population with the apparent binding affinity. Fig. 2D shows that indeed, in all cases (except E484R) the binding affinity of the most abundant variant in the population has the highest binding affinity at the given position. In respect to E484R, the mutation of Glu to Arg requires two nucleotide changes in the same codon, making this mutation reachable only by multiple rounds of random mutagenesis, which will delay its occurrence and may explain its low frequency (however, will not stop its spread). Next, we monitored whether the spread of mutation in the population also relates to the protein-stability of

the RBD. Here, we used the level of yeast surface expression as a proxy to estimate protein stability (*20*). Fig. S5 shows, that mutations which occurrence is increasing in the population did not affect protein stability, corroborating that maintaining protein-stability is an important evolutionary constraint.

The most abundant naturally occurring mutations in RBD have been selected by yeast display, already in the first affinity maturation library (B3) (*7, 8, 23*). Next, we aimed to explore whether much higher affinity binding can be achieved.

Exploring the affinity limits for ACE2-RBD interaction

A further selection of better binders can demonstrate the future path of SARS-CoV-2 evolution. In parallel, an ultra-tight binder can be used as an effective ACE2 blocker for inhibiting SARS-CoV-2 infection. We used the same approach as for B3 and B4 and created the subsequent library B5. The library B5 was enriched by using 200 pM ACE2 as bait, followed by 50 pM, and finally at 30 pM. Sorting with less than 100 pM bait was done after overnight incubation in 50 ml solution to prevent ligand depletion effect (as the number of ACE2 molecules becomes much lower than the number of RBD molecules). Round 5 resulted in the fixation of mutations N460K, E484K, Q498R, and N501Y in all sequenced clones. Mutations S477N and S494P were present with frequencies > 20 %. Additional mutations identified were G446R, I468V, and F490Y. Representative clones with different mutational profiles were subjected to detailed analyses (Table 1). In the next selection step, we targeted for faster association-rates by using pre-equilibrium selection (*24*). The new library FA (fast association) was created by randomization of the whole *RBD* gene population from the enriched B5 library. The library was pre-selected with 30 pM ACE2 for 8 hrs (reaching equilibrium after ON incubation) followed by 1 hr and 30 min incubation before selection. This resulted in the accumulation of additional mutations: V445K, I468T, T470M and also the fixation of the previously observed mutation S477N in all sequences cloned. 4 minor mutations N354E, K417F, V367W, and S494P, with only a single sequence each, were identified. One should note that V445K and T470M require two nucleotide mutations to be reached, demonstrating the efficiency of using multiple rounds of library creation on top of previous libraries (and not single clones). Interestingly, these mutations were not located at the binding interface but rather in the peripheries, which is in line with previously described computational fast association design, where periphery mutations were central (*25*). From the FA library, we determined the isogenic binding for 5 different clones with clone RBD-62 being the

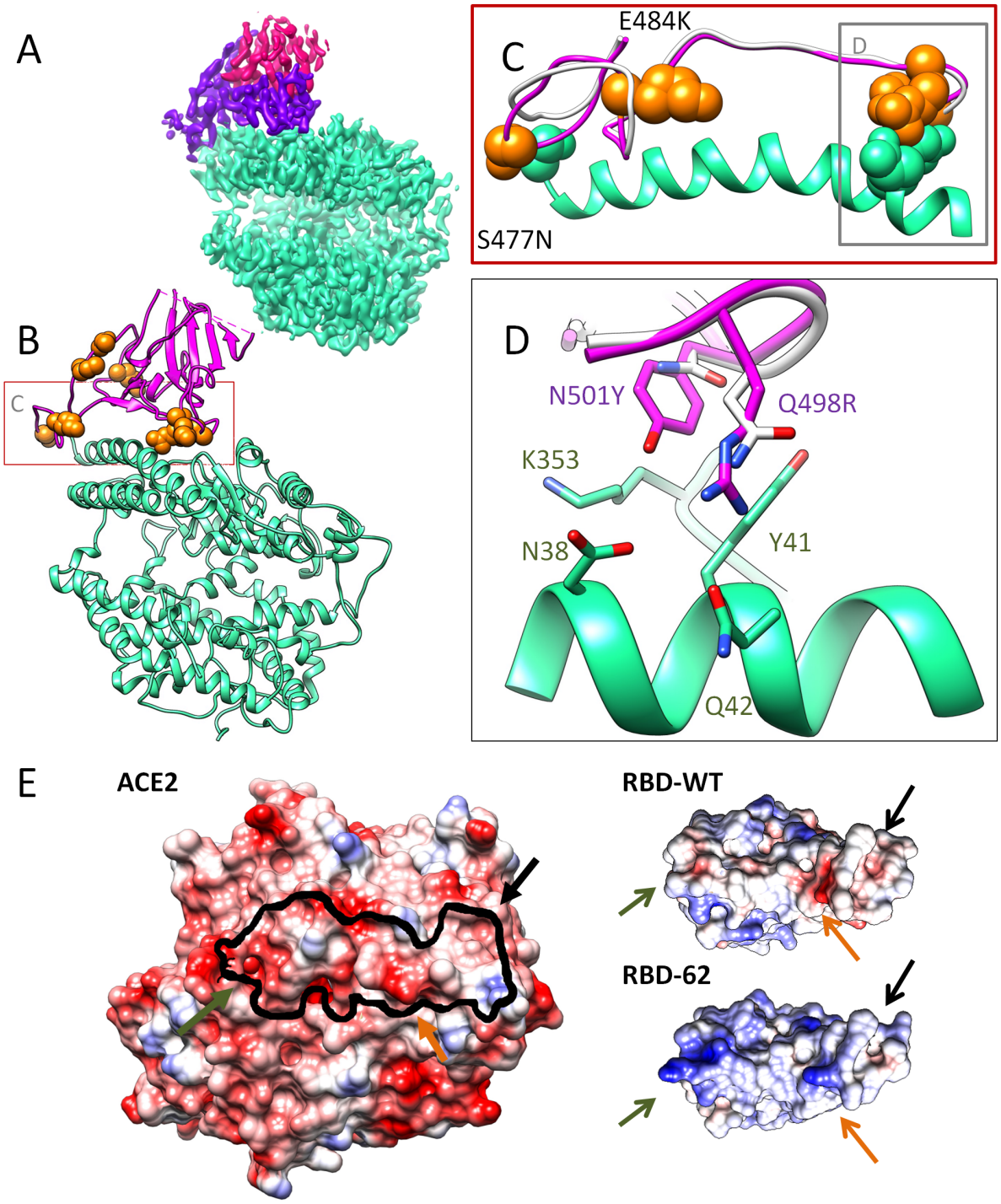

Fig. 3 Cryo-EM structure of the ACE2-RBD-62 complex at 2.9 Å resolutions. A) The Cryo-EM electron density map with ACE2 (cyan), RBD-62 RBM (magenta), and RBD core (pink). B) Cartoon representation of the ACE2-RBD-62 model with eight mutations resolved in the electron density map (orange). C) The S477N, Q498R and N501Y mutations depicted in RBM (orange spheres) interacting with S19, Q42 and K353 of ACE2 respectively (cyan spheres) are situated at the two extremes of the RBD-ACE2 interface, suggested to stabilizing the complex. D) The interaction network formed between RBD-62 mutations and ACE2. RBD-WT residues are in white (heteroatom coloring schema). E) Electrostatic complementarity between RBD and ACE2 is strengthened in RBD-62 by positive charges at positions N460K, E484K, and Q498R. The black line one ACE2 indicates the RBD binding site.

Table 1 – Biophysical parameters of the mutant clones selected by yeast display. For more details see Table S2.

Clone	Library	Plasmid[a]	Mutations	Tm[b] [°C]	Yeast display[c] $K_{D,app}$ (pM)	Octet RED[d] K_D, (pM)	k_{on}[e] $M^{-1}s^{-1}$ x10^5
RBD-WT	-	pJYDC1	AA 336 – 528	53.5	1600 ± 200	38000 ± 10000	1.7 ± 0.06
RBD-32	B3	pJYDC1	I358F, S477N, N501Y	ND	936 ± 60	ND	
RBD-33	B3	pJYDC1	I358F, E484K, N501Y	ND	126 ± 1.5	ND	
RBD-36	B3	pJYDC1	I358F, I468T, N481Y, N501Y	54.6	184.4 ± 1.9	ND	
RBD-48	B4	pJYDC3	I358F, S477N, Q498R, N501Y		65.3 ± 2.1	ND	
RBD-52	B5	pJYDC1	I358F, N460K, E484K, S494P, Q498R, N501Y	61.9	59 ± 6.2	3000 ± 1000	0.52 ± 0.07
RBD-521	B5	pJYDC1	I358F, N460K, E484K, Q498R, N501Y	58.2	11.96 ± 2.4	340 ± 40	5.7 ± 0.03
RBD-55	B5	pJYDC1	I358F, E484K, Q498R, N501Y	54.8	18.5 ± 0.6	ND	
RBD-62	FA	pJYDC3	I358F, V445K, N460K, I468T, T470M, S477N, E484K, Q498R, N501Y	57.9	2.5 ± 0.2	60 ± 16	13 ± 1
RBD-71	-	pJYDC3	I358F, V367W, R408D, K417V, V445K, N460K, I468T, T470M, S477N, E484K, Q498R, N501Y	63	8.5 ± 1.5	200 ± 100	16 ± 1

[a] pJYDC1 plasmid is using intrinsic eUnaG2 reporter; pJYDC3 plasmid contains DnbALFA reporter; see (*20*)

[b] Melting temperature as measured by differential scanning fluorimeter *Tycho NT.6* (NanoTemper Technologies GmbH)

[c] K_D values measured between yeast surface-exposed RBD variants and the monomeric extracellular portion of ACE2 receptor.

[d,e] Measured by Octet RED96 system (ForteBio) by using AR2G biosensors. For details see Materials and Methods.

best. Yeast display titration showed an affinity of 2.5 ± 0.2 pM (Fig. 2E and Table 1). The other clones tested from the FA library had affinities between 5 to 10 pM (Fig. S3 and Table S2).

ACE2 receptor and clones RBD-52, RBD-521, and RBD-62 were expressed and purified (Fig. S4). Measuring the binding affinity to ACE2 using the Octet RED96 System showed a systematically lower binding affinity in comparison to yeast titration (Table 1). For WT, yeast titration was reduced from 1.6 to 38 nM and for RBD-62 the affinity was reduced from 2.5 to 60 pM. However, the improvement in affinity is similar for both methods (~600-fold). While most of the improvement came from reduced k_{off} (Fig. 2F and G) k_{on} increased 8-fold, from 1.7x10^5 to 13 x10^5 $M^{-1}s^{-1}$ for RBD-62 (Table 1). In addition, RBD-62 is 4 ^{0}C more stable than WT, probably due to the introduction of the I358F stabilizing mutation (Fig. S1). To further increase the RBD-62 affinity we prepared a site-directed mutational library on top of RBD-62, including the 15 mutations suggested from deep mutational scanning (*6*), which require more than one nucleotide

change to be reached (Fig. S6). Surprisingly, these mutations did not significantly increase the affinity towards RBD-62 as they did for wild-type (*6*). Yet, a combination of three of them stabilized RBD-62 by 5 ^{0}C, creating RBD-71, but at the cost decreased binding affinity (Table 1, Fig. S6). This demonstrates the limitation of the use of single amino-acid changes from deep mutational scanning to obtain high-affinity binders.

RBD-62-ACE2 structure

We determined the cryo-EM structure of the N-terminal peptidase domain of the ACE2 (G17-Y613) receptor bound to the RBD-62 (T333-K528) (Fig. 3A), including nine mutations (I358F, V445K, N460K, I468T, T470M, S477N, E484K, Q498R, N501Y; Table 1, Figs. 2A and 3B). Structure comparison of the ACE2-RBD-62 complex and the WT complex (PDB ID: 6M0J) revealed their overall similarity with rmsd of 0.97 Å across 586 amino acids of the ACE2 and 0.66 Å among 143 amino acids of the RBD (Fig. S10A). Three segments; R357-S371 (β2, α2), G381-V395 (α3), and F515-H534 (β11**)** are disordered in RBD-62, and thus not visible in the electron density map (Fig. S9 and blue cartoon in Fig. S10B). These segments are situated opposite to ACE2 binding interface and therefore not stabilized and rigidified by ACE2 contacts. All mutations, except I358F, are present in the electron density map. Details of cryo-sample preparation, data acquisition, and structural determination are given in the Supplementary Materials Methods. The cryo-EM data collection and refinement statistics are summarized in Fig. S7, S8, S9, and Table S3. Mutations V445K, N460K, I468T, T470M, S477N, E484K, Q498R, and N501Y are part of the receptor-binding motif (RBM) that interacts directly with ACE2 (orange spheres Fig. 3B and C) (*3*). The RBM including residues S438-Q506 shows the most pronounced conformational differences in comparison to the RBD-WT (the black circle in Fig. S10A). Out of the nine mutations in the RBM four involve intramolecular interactions, stabilizing the RBD-62 structure, including hydrogen contacts between K460 and D420, T468 and R466, and M470 and Y351. The mutations S477N, Q498R, N501Y are forming new contacts with ACE2. The Arg at position 498 makes a salt bridge to Q42 and hydrogen contact to Y41 of ACE2 making together with mutation N501Y (Y has contact with K353) a strong network of new interactions supporting the impact of these two residues (Fig. 3D). Calculating the electrostatic potential of the RBD-62 in comparison to RBD-WT shows a much more positive surface of the former, which is complementary to the negatively charged RBD binding surface on ACE2 (Fig. 3E). In addition, the mutation N477

interacts with S19 of ACE2 (Fig. 3C). Interestingly the interface of ACE2-RBD involves the interaction of amino acid residues from the N-terminal segment Q24-Q42 , K353, and D355 of the ACE2 domain and residues from the RBM domain of the RBD. The S477N, Q498R, and N501Y mutations in RBD-62 are situated at the two extremes of the RBD-ACE2 interface therefore stabilizing the complex (Fig. 3C).

RBD-62 inhibits SARS-CoV-2 infection without affecting ACE2 enzymatic activity

The main driver of this study was to generate a tight inhibitor of ACE2 for medicinal purposes, which will be administered to the nose and lungs through inhalation. Therefore, we had to verify that the evolved RBD does not interfere with the ACE2 enzymatic activity, which is important in the Renin-Angiotensin-Aldosterone system (*17, 18*). We assayed the impact of RBD-WT and RBD-62 proteins on ACE2 activity. Both the *in vitro* assay and assays done on various cells expressing ACE2 did not show much difference in ACE2 activity with and without RBD-WT or RBD-62 added (Figs. 4A, S5).

Finally, we explored the inhibition of RBD-WT and RBD-62 on viral entry. Initially, we used Lentivirus pseudotyped with spike protein variant SΔC19 (*26*). This spike variant lacked the last 19 amino acids that are responsible for its retention in the endoplasmic reticulum. The relative cellular entry was analyzed by flow-cytometry of Lentivirus infection promoting GFP

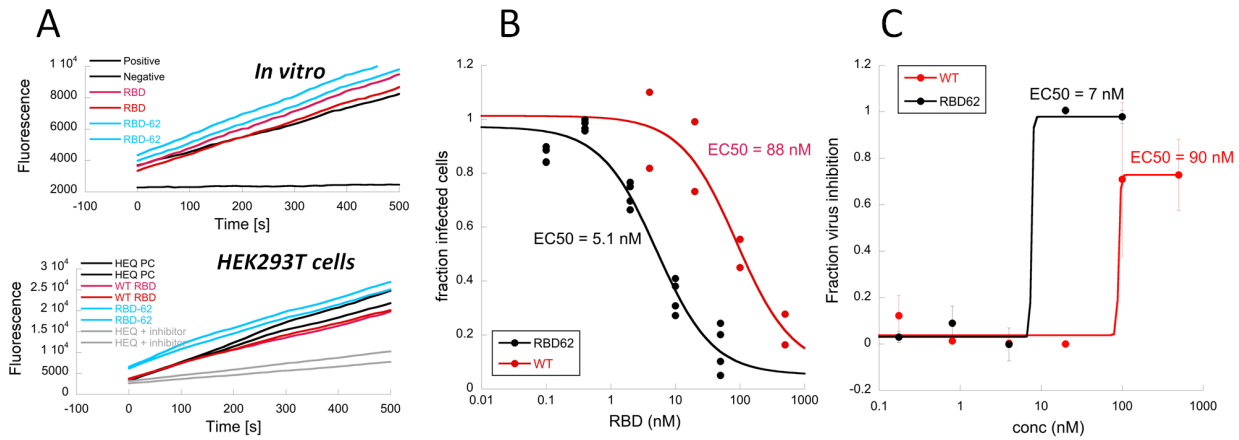

Fig. 4 Inhibition of RBD-WT and RBD-62 on ACE2 activity and their potential to inhibit viral entry and infection. (A) ACE2 activity (*in vitro* or on cells) assayed using SensoLyte® 390 ACE2 Activity Assay Kit. Fluorogenic peptide cleavage by ACE2 was measured in 10 seconds intervals over 30 minutes. The activity rate is indicated by the slope of the plot [product/time]. An Ace2 inhibitor (Inh.), provided with the kit, was used as the negative control. **(A)**. ACE2 activity was measured *in vitro* after the addition of 100 nM of RBD-WT or RBD-62 to purified ACE2. **(**upper panel**).** ACE2 activity was measured following incubation with RBD-WT or RBD-62 on HeLa cells transiently transfected with human full-length ACE2 (bottom panel). **(B)** Inhibition of infection of HEK-293T cells stably expressing ACE2 by Lentivirus pseudotyped with SARS-CoV-2 spike protein. (C) Inhibition of SARS-CoV-2 infection by RBD-WT and RBD-62 proteins.

signal. The HEK-293T cells stably expressing hACE2 were pre-incubated with serial dilutions of the two RBDs for 1 h and then the pseudovirus was added for 48 hrs. Results in Fig. 4B show that the EC50 was reduced from 88 nM for RBD-WT to 5.1 nM for RBD-62. Next, RBD-WT and RBD-62 were evaluated for their potency in inhibiting SARS-CoV-2 infection to VeroE6 cells (Fig. 4C). Similar to the pseudovirus, also here the EC50 was reduced from 90 to 6.8 nM for RBD-WT and RBD-62 respectively. More significantly, RBD-62 blocked >99% of viral entry and replication, while RBD-WT blocked only ~75% of viral replication. The complete blockage of viral replication, using a low nM concentration of RBD-62 makes it a promising drug candidate.

Discussion

The SARS-CoV-2 pandemic is an ongoing event, with the virus constantly acquiring new mutations. Intriguingly, the naturally selected mutations S477N, E484K, and N501Y of the Spike protein RBD, which show higher infectivity, were selected by yeast surface display affinity maturation already in the first round, giving rise to the South-African, E484K, N501Y, and British variants that bind ACE2 13 and 3.5-fold tighter than RBD-WT. Following three additional rounds of yeast display selection resulted in 600-fold tighter binding in comparison to RBD-WT. The selection process took advantage of combinatorial selection, without compromising protein-stability. The high-affinity binder, RBD-62 was evaluated as a potential drug and showed to efficiently block ACE2, without affecting its important enzymatic activity.

While natural virus selection is not as efficient as *in vitro* selection, the gained information on the more critical mutations can be used as a tool to identify emerging mutations. We hypothesize that E484R will continue to spread and will become more dominant, especially in combination with N501Y. In contrast, we do not expect the rapid spread of S494P. Importantly, the mutation Q498R appeared in the library B4 after the incorporation of Tyr at position 501. This combination dramatically increased the affinity below 100 pM as is shown by the difference between RBD-32 and RBD-44 (Table 1). Notably, the wild-type RBD codon at position 498 is CAA, allowing for direct change to arginine codon CGA. R498 was not sampled yet by the virus (Fig. 2A) but its appearance should be carefully monitored. Moreover, R498 is located in a hypervariable location of the RBD (Fig. S11), which makes its appearance more plausible.

We successfully solved the Cryo-EM structure of RBD-62 to high resolution. The structure shows that RBD-62 has much improved electrostatic complementarity with ACE2, in relation to RBD-WT (Figure 3). This can be attributed to the use of the Fast Association protocol. The structure contains many of the currently evolving mutations (S47N, E484K, N501Y) and can serve not only as valuable source of information but also as a "crystal ball" to predict future virus evolution steps. To evaluate the effect of mutations in the RBD on antibody binding, we manually inspected 92 antibody-RBD (nanobody, Spike) structures for clashes. 28 of the antibodies bind outside the RBM and 8 interactions are similar with RBD-62 and RBD-WT. However, for 56 antibodies, a decrease in the number of contacts was observed and in 9 cases major clashes with RBD-62 are observed (Fig. S12). Notably, E484R and Q498R caused most of the observed effects. These findings suggest the need for close monitoring of the efficiency of drugs and vaccines for current and future mutations.

An intriguing question is whether the spreading of the tighter binding SARS-CoV-2 variants in humans is accidental. From the similarity to yeast display selection, where stringent conditions are used, one may hypothesize that stringent selection is also driving the rapid spread of these mutations. Face masks of low quality (which are by far the most abundant) would provide such selection conditions, as they reduce exhaled viral titers, given tighter binding variants an advantage over WT to spread rapidly in the population (as a result of R_0 of mutated viruses being >1, while <1 for WT viruses). This should be urgently investigated, as one may consider the mandatory use of higher quality face-masks, which will reduce viral titer to bellow infection levels (as indeed seen with medical personal who use such masks) and stop spreading these tighter binding virus mutations.

Acknowledgments

Funding: This research was supported by the Israel Science Foundation (grants No. 3814/19 and 1268/18) within the KillCorona – Curbing Coronavirus Research Program and by the Ben B. and Joyce E. Eisenberg Foundation. **Authors contribution**: J.Z. and G.S. conceived the project; J.Z., S.M., M.S., E.Z., J.C., B.M and G.S. performed experiments; N.E. prepared cryo-EM samples and built atomic models and refined structures with O.D. J.Z, N.E, O.D and G.S wrote the manuscript. **Competing interests:** The authors J.Z. and G.S. declare the US Provisional Patent Application No. 63/125,984 (Yeda Ref.: 2020-091). **Data and materials availability:** maps and atomic coordinates have been deposited in the Protein Data Bank (www.rcsb.org) and the

Electron Microscopy Data Bank (www.ebi.ac.uk/pdbe/emdb with accession codes: XXX, XXX, respectively.

Supplementary Materials

Materials and Methods

Supplementary text

Table S1 – S3

Figs. S1 – S12

References (27 – 41)

References

1. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181, 271-280.e278 (2020).
2. J. Lan *et al.*, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581, 215-220 (2020).
3. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260-1263 (2020).
4. W. Tai *et al.*, Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cellular & Molecular Immunology* 17, 613-620 (2020).
5. A. C. Walls *et al.*, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181, 281-292.e286 (2020).
6. T. N. Starr *et al.*, Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182, 1295-1310.e1220 (2020).
7. H. Tegally *et al.*, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*, 2020.2012.2021.20248640 (2020).
8. J. Chen, R. Wang, M. Wang, G.-W. Wei, Mutations Strengthened SARS-CoV-2 Infectivity. *Journal of Molecular Biology* 432, 5212-5226 (2020).
9. S. Lukassen *et al.*, SARS-CoV-2 receptor ACE2 and TMPRSS2 are primarily expressed in bronchial transient secretory cells. *EMBO J* 39, e105114-e105114 (2020).
10. C. G. K. Ziegler *et al.*, SARS-CoV-2 Receptor ACE2 Is an Interferon-Stimulated Gene in Human Airway Epithelial Cells and Is Detected in Specific Cell Subsets across Tissues. *Cell* 181, 1016-1035.e1019 (2020).
11. L. Dai, G. F. Gao, Viral targets for vaccines against COVID-19. *Nature reviews. Immunology*, 1-10 (2020).
12. S. H. Nile *et al.*, COVID-19: Pathogenesis, cytokine storm and therapeutic potential of interferons. *Cytokine & Growth Factor Reviews* 53, 66-70 (2020).
13. C. O. Barnes *et al.*, SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* 588, 682-687 (2020).

14. T. M. Abd El-Aziz, A. Al-Sabi, J. D. Stockand, Human recombinant soluble ACE2 (hrsACE2) shows promise for treating severe COVID-19. *Signal Transduction and Targeted Therapy* 5, 258 (2020).
15. D. Schütz *et al.*, Peptide and peptide-based inhibitors of SARS-CoV-2 entry. *Adv Drug Deliv Rev* 167, 47-65 (2020).
16. L. Cao *et al.*, De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* 370, 426-431 (2020).
17. D. d. F. Lelis, D. F. d. Freitas, A. S. Machado, T. S. Crespo, S. H. S. Santos, Angiotensin-(1-7), Adipokines and Inflammation. *Metabolism* 95, 36-45 (2019).
18. H. Zhang, J. M. Penninger, Y. Li, N. Zhong, A. S. Slutsky, Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Medicine* 46, 586-590 (2020).
19. G. Wang *et al.*, Dalbavancin binds ACE2 to block its interaction with SARS-CoV-2 spike protein and is effective in inhibiting SARS-CoV-2 infection in animal models. *Cell Research*, (2020).
20. J. Zahradník, D. Dey, S. Marciano, G. Schreiber, An enhanced yeast display platform demonstrates the binding plasticity under various selection pressures. *bioRxiv*, 2020.2012.2016.423176 (2020).
21. G. Chao *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nature Protocols* 1, 755-768 (2006).
22. S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* 1, 33-46 (2017).
23. S. Kemp *et al.*, Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/ΔV70. *bioRxiv*, 2020.2012.2014.422555 (2020).
24. R. Cohen-Khait, G. Schreiber, Selecting for Fast Protein–Protein Association As Demonstrated on a Random TEM1 Yeast Library Binding BLIP. *Biochemistry* 57, 4644-4650 (2018).
25. T. Selzer, S. Albeck, G. Schreiber, Rational design of faster associating and tighter binding protein complexes. *Nature structural biology* 7, 537-541 (2000).
26. H. Cohen-Dvashi *et al.*, Coronacept – a potent immunoadhesin against SARS-CoV-2. *bioRxiv*, 2020.2008.2012.247940 (2020).
27. L. Benatuil, J. M. Perez, J. Belk, C.-M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Engineering, Design and Selection* 23, 155-159 (2010).
28. A. R. Aricescu, W. Lu, E. Y. Jones, A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta crystallographica. Section D, Biological crystallography* 62, 1243-1250 (2006).
29. Y. Peleg, T. Unger, Application of the Restriction-Free (RF) cloning for multicomponents assembly. *Methods in molecular biology (Clifton, N.J.)* 1116, 73-87 (2014).
30. D. S. Wilson, A. D. Keefe, Random mutagenesis by PCR. *Current protocols in molecular biology* Chapter 8, Unit8.3 (2001).
31. R. D. Gietz, Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods in molecular biology (Clifton, N.J.)* 1163, 33-44 (2014).
32. D. N. Mastronarde, Automated electron microscope tomography using robust prediction of specimen movements. *Journal of structural biology* 152, 36-51 (2005).
33. A. Punjani, J. L. Rubinstein, D. J. Fleet, M. A. Brubaker, cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature Methods* 14, 290-296 (2017).

34. A. Punjani, H. Zhang, D. J. Fleet, Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nature Methods* 17, 1214-1221 (2020).
35. A. Punjani, D. J. Fleet, 3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *bioRxiv*, 2020.2004.2008.032466 (2020).
36. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66, 213-221 (2010).
37. B. P. Klaholz, Deriving and refining atomic models in crystallography and cryo-EM: the latest Phenix tools to facilitate structure analysis. *Acta crystallographica. Section D, Structural biology* 75, 878-881 (2019).
38. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 60, 2126-2132 (2004).
39. V. B. Chen *et al.*, MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* 66, 12-21 (2010).
40. E. F. Pettersen *et al.*, UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry* 25, 1605-1612 (2004).
41. F. Amanat *et al.*, A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nature Medicine* 26, 1033-1036 (2020).

Supplementary Materials

Materials and methods

Cloning and DNA manipulations

The RBD domain variants (see Table S1) were PCR amplified (KAPA HiFi HotStart ReadyMix, Roche, Switzerland) from codon-optimized SARS-CoV-2 Spike protein gene (Sino Biological, SARS-CoV-2 (2019-nCoV) Cat: VG40589-UT, GenBank: QHD43416.1) by using appropriate primers. Amplicons were purified by using NucleoSpin® Gel and PCR Clean-up kit (Nacherey-Nagel, Germany) and eluted in DDW. Yeast surface display plasmid pJYDC1 (Adgene ID: 162458) and pJYDC3 (162460) were cleaved by *Nde*I and *Bam*HI (NEB, USA) restriction enzymes, purified, and tested for non-cleaved plasmids *via* transformation to *E.coli* Cloni® 10G cells (Lucigen, USA). Each amplicon was mixed with cleaved plasmid in the ratio: 4 µg insert: 1 µg plasmid per construct, electroporated in *S.cerevisiae* EBY100 (*27*), and selected by growth on SD-W plates.

Cloning of ACE2 extracellular domain (AA G17-Y613) gene and *RBD*s into vectors pHL-sec (*28*) were done in two steps. Initially, the RBD gene was inserted in helper vector pCA by restriction-free cloning (*29*). pCA is a pHL-sec derivative lacking 862 bp in the GC rich region (nt 672 - 1534). In the second step, the correctly inserted, verified by sequencing, *RBDs* with flanking sequences were cleaved by using restriction enzymes *Xba*I and *Xho*I (NEB, USA) and ligated (T4 DNA ligase, NEB, USA) in cleaved full-length plasmid pHL-sec.

Site-directed mutagenesis of *RBDs* was performed by restriction-free cloning procedure (*29*). Megaprimers were amplified by KAPA HiFi HotStart ReadyMix (Roche, Switzerland), purified with NucleoSpin™ Gel and PCR Clean-up Kit (Nachery-Nagel, Germany), and subsequently inserted by PCR in the destination using high fidelity Phusion® (NEB, USA) or KAPA polymerases. The parental plasmid molecules were inactivated by *Dpn*I treatment (1 h, NEB, USA) and the crude reaction mixture was transformed to electrocompetent *E. coli* Cloni® 10G cells (Lucigen, USA). The clones were screened by colony PCR and their correctness was verified by sequencing.

DNA libraries preparation

SARS-CoV-2 RBD gene (RBD) libraries were prepared by $MnCl_2$ error-prone mutagenesis (*30*) using Taq Ready-mix (Hylabs, Israel). The mutagenic PCR reactions (50 µl) were supplemented

with increasing $MnCl_2$ concentrations: 0.05, 0.1, 0.2, 0.4, 0.6, 0.8 and 1.0 nM. Template DNA concentration ranged between 100 and 400 ng per reaction and 20 – 30 reaction cycles were applied. The amplified DNA was purified, pooled, and used directly for yeast transformation via electroporation. The whole gene randomization amplicon comprised RBD and linker between it and Aga2p protein (nucleotides -152 – 621, pJYDC1 vector). Libraries B3, B4, and B5 were prepared by homologous recombination of an invariant fragment of RBD with necessary overlaps (1 – 321) and the mutagenized library fragment (260 – 621). The mutagenic fragments were prepared by the same error-prone PCR procedure (20 cycles).

Yeast transformation, cultivation, and expression procedures

The detailed description of all the procedures and our enhanced yeast display platform itself was described in details (*20*). Briefly, plasmids were transformed into the EBY100 *Saccharomyces cerevisiae* (*27, 31*). Single colonies were inoculated into 1.0 ml liquid SD-CAA media (*20*), and grown overnight at 30°C (220 rpm). The overnight cultures were spun down (3000 g, 3 min) and the exhausted culture media was removed before dilution in the expression media 1/9 (*20*) to OD ~ 1. The expression cultures were grown at different temperatures 20, 30, and 37 °C for 8 – 24 h at 220 rpm, depending on the experimental setup. The expression co-cultivation labeling was achieved by the addition of 1 nM DMSO solubilized bilirubin (pJYDC1, eUnaG2 reporter holo-form formation, green/yellow fluorescence (Ex. 498 nm, Em. 527 nm)) or 5 nM ALFA-tagged mNeonGreen (pJYDC3, DnbALFA). Aliquots of cells (100 ul) were collected by centrifugation (3000 g, 3 min) resuspended in ice-cold PBSB buffer (PBS with 1 g/L BSA), passed through cell strainer nylon membrane (40 µM, SPL Life Sciences, Korea), and analyzed.

Binding assays and affinity determination using yeast surface display

Aliquots of yeast expressed and labeled cells ready for flow-cytometry analysis were resuspended in analysis solution with a series of labeled ACE2 concentrations. The concentration range was of CF®640R succinimidyl ester labeled (Biotium, USA) ACE2 extracellular domain (AA Q18 – S740) was dependent on the protein analyzed (0.1 pM – 50 nM). The analysis solution volume was adjusted (1 – 100 ml) to avoid the ligand depletion greater than 10% as well as the time needed to reach the equilibrium (1 h – 12 h, 5 rpm, 4 °C) (*21*). After the incubation, samples were collected (3000 g, 3 min), resuspended in 200 ul of ice-cold PBSB buffer (200 µl), passed through a cell strainer, and analyzed. The expression and binding signals were determined by

flow cytometry using BD Accuri™ C6 Flow Cytometer (BD Biosciences, USA). The cell analysis and sorting were done by S3e Cell Sorter (BioRad, USA). The analysis was done by single-cell event gating (Fig. S2), green fluorescence channel (FL1-A) was used to detect RBD expression positive cells (RBD+) via eUnaG2 or DnbALFA, and far-red fluorescent channel (FL4-A) recorded CF®640R labeled ACE2 binding signals (CF640+). The eUnaG2 signals were automatically compensated by the ProSort™ Software and pJYDNp positive control plasmid (Adgene ID 162451 (*20*)). The mean FL4-A fluorescence signal values of RBD+ cells, subtracted by RBD-, were used for determination of binding constant K_D. The standard non-cooperative Hill equation was fitted by nonlinear least-squares regression using Python 3.7. The total concentration of yeast exposed protein was fitted together with two additional parameters describing the given titration curve (*6*).

Production and purification of RBD and ACE2 proteins

The extracellular part of ACE2 (Q18 – S740) and RBD protein variants (Table S1) were produced in Expi293F cells (ThermoFisher). Pure DNA was transfected using ExpiFectamine 293 Transfection Kit (ThermoFisher) using the manufacturer protocol. 72 hours post-transfection, the cells were centrifuged at 1500 rpm for 15 minutes. The supernatant was filtered using 0.45 µm Nalgene, ThermoFisher filter and the pellet was discarded. The filtered supernatant was loaded onto a 5 ml of HisTrap Fast Flow column (Cytivia (GE, USA), cat 17-5255-01). ÄKTA pure (Cytivia, USA) was used to purify the protein. The column was washed in 25 mM Tris, 200 mM NaCl 20 mM imidazole, then, the protein was eluted using gradient elution with elution buffer containing 25 mM Tris, 200 mM NaCl 1M imidazole. Buffer exchange to PBS and the concentration of the protein were done by using amicons® (Merck Millipore Ltd, cat:UFC900324).

Cryo-Electron Microscopy

Sample preparation: 2.5 µl of ACE2-RBD-62 complex at 3.5 mg/ml concentration was transferred to glow discharged UltrAuFoil R 1.2/1.3 300 mesh grids (Quantifoil), blotted for 2.5 seconds at 4°C, 100% humidity, and plunge frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot plunger (Thermo Fisher Scientific).

Cryo-EM image acquisition: Cryo-EM data were collected on a Titan Krios G3i transmission electron microscope (Thermo Fisher Scientific) operated at 300 kV. Movies were recorded on a K3 direct detector (Gatan) installed behind a BioQuantum energy filter (Gatan), using a slit of 20 eV. Movies were recorded in counting mode at a nominal magnification of 165,000x, corresponding to a physical pixel size of 0.53 Å. The dose rate was set to 16.2 e^-/pixel/sec, and the total exposure time was 1.214 sec, resulting in an accumulated dose of 70 $e^-/Å^2$. Each movie was split into 57 frames of 0.021 sec. The nominal defocus range was -0.7 to -1.1 µm, however, the actual defocus range was larger. Imaging was done using an automated low dose procedure implemented in SerialEM (*32*). A single image was collected from the center of each hole using image shift to navigate within hole arrays and stage shift to move between arrays. The 'Multiple Record Setup' together with the 'Multiple Hole Combiner' dialogs were used to map hole arrays of up to 3x3 holes. Beam tilt was adjusted to achieve coma-free alignment when applying image shift.

Cryo-EM image processing: Image processing was performed using CryoSPARC software v3.0.1 (*33*). The processing scheme is outlined in Fig. S7. A total of 4470 acquired movies were subjected to patch motion correction, followed by patch CTF estimation. Of these, 3357 micrographs having CTF fit resolution better than 5 Å and relative ice thickness lower than 1.07, were selected for further processing. Initial particle picking was done using the 'Blob Picker' job on a subset of 100 micrographs. Extracted particles were iteratively classified in 2D and their class averages were used as templates for automated particle picking from all selected micrographs, resulting in 2,419,995 picked particles. Particles were extracted, binned 6x6 (60-pixel box size, 3.18 Å/pixel), and cleaned by multiple rounds of 2D classification, resulting in 1,649,355 particles. These particles were used for *ab initio* 3D reconstruction with 5 classes. Out of the 5 classes only one, containing 552,575 particles, refined to high resolution. Two additional classes may show ACE2 in a closed conformation (containing 249,841 and 503,670 particles), however, they did not refine, partially because of preferred orientation. The 3D class containing 552,575 particles was refined as follows: Particles were re-extracted only from micrographs with defocus lower than 1.7 µm, binned 2x2, and subjected to homogeneous refinement (355,891 particles, 200-pixel box size, 1.06 Å/pixel). The particles were then sub-classified into 2 classes, and particles from the higher-resolution class were re-extracted without binning in 680-pixel boxes, subjected to per particle motion correction, followed by non-uniform refinement (*34*) with per-particle defocus optimization. The final map, at a resolution of 2.9 Å (Fig. S8), was

sharpened with a B-factor of -83 before atomic model building. In the final map, the RBD is only partially resolved at the distal region from the ACE2 interface. To better understand the reason for the missing density, we subjected the particles from the well-refined 3D class (355,891 particles) to variability analysis (*35*), with a binary mask imposed on the RBD region (Fig. S9). Classification into 5 distinct classes based on 3 eigenvectors, revealed variable density at the RBD distal region, which could not be modeled reliably. The cryo-EM data collection process and refinement statistics are summarized and visualized in Fig. S7, S8, S9, and Table S3.

Model building: The atomic model of the ACE2-RBD-62 was solved by docking into the Cryo-EM maps the homologous refined structure of the SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB-ID 6M0J) as a model, using the Dock-in-Map program in PHENIX (*36*). All steps of atomic refinements were carried out with the Real-space refinement in PHENIX (*37*). The model was built into the cryo-EM map by using the COOT program (*38*). The ACE2-RBD-62 model was evaluated with the MOLPROBIDITY program (*39*). The ACE2 (G17-Y613) contains one zinc ion linked to H374, H378, and E402 and three N-acetyl-β-glucosaminide (NAG) glycans linked to N53, N90, and N546. In the RBD-62 structure (T333-K528) three fragments; R357-S371 (β2, α2), G381-V395 (α3), and F515-H534 (β11) are disordered, and thus not visible in the electron density map. Details of the refinement statistics of the ACE2-RBD62 structure are described in Table S3. 3D visualization and analyses were performed using UCSF Chimera (*40*) and PyMol (Schrödinger, Inc.; 2.4.0).

Analysis of RBD circulating virus variants

All amino acid substitutions in the RBD (116) were downloaded from the GISAID database (23 December 2020) (*22*) with the corresponding numbers of sequences and regions and plotted against the binding ($\Delta Log_{10}(K_{D,App})$) or expression ($\Delta Log_{10}MFI$) extracted from the RBD deep mutational scanning dataset (*6*). We gratefully acknowledge all GISAID contributors and Starr et all for sharing their data.

Octet RED binding analysis

Octet RED96 System (forte´BIO, Pall Corp., USA) was used for real-time binding determination. Briefly, 10 µg/ml of ACE2 diluted in 10 mM NaAcetate pH5.5 was immobilization to an amine-reactive 2G biosensor using standard procedure. The purified RBD was diluted in a sample buffer

(PBS+0.1% BSA+0.02% Tween20). Analyte concentrations, association, and dissociation times were adjusted per sample. Data Analysis v10 software (forte´BIO, Pall Corp., USA) was used for data fitting, with the mathematical model assuming a simple 1:1 stoichiometry.

Pseudo-virus production and inhibition of infection by RBD

Pseudo-virus production: SARS-CoV-2-Spike pseudotyped Lentivirus was produced by co-transfection of Hek293T cells pCMV ΔR8.2, pGIPZ-GFP, (*26*) and pCMV3 SΔC19 at a ratio of 1:1:1. 24 hours before the transfection 1 x 10^6 cells were seeded into a 10 cm plate. On the day of the transfection cells were washed by Dulbecco's Modified Eagle's Medium (DMEM) (Gibco 11965092) and 5 ml of Opti-MEM (Gibco 11058021) was added to the plate. 10 µg of plasmids mix was transfect using lipofectamine 2000 transfection reagent (Thermo Fisher 11668027) according to the manufacturer's instructions. After 4 hours, the media was replaced by 9 ml of fresh media. The supernatant was harvested 72 h post-transfection, centrifuged (1000 g, 5 min), and filtered to remove all residual debris (Millex-HV Syringe Filter Unit, 0.45 µm).

RBD inhibition assay: HEK-293T cells stably expressing hACE2 (GenScript M00770) were seeded into 24-well plate at an initial density of 6 x 10^4 cells per well. The following day cells were pre-incubated with serial dilutions of RBDs (1 h) and then the pseudotyped Lentivirus was added. After 24 h, the cell medium was replaced with fresh DMEM, and cells were grown for an additional 24 h. After this procedure, cells were harvested and the GFP signal was analyzed by flow cytometry (BD Accuri™ C6 Plus Flow Cytometer, BD Biosciences, USA).

Inhibition of SARS-CoV-2 infection

The strain 2019-nCoV/IDF0372/2020 was supplied by the National Reference Centre for Respiratory Viruses hosted by Institute Pasteur (Paris, France) and headed by Dr. Sylvie van der Werf. The human sample from which strain 2019-nCoV/IDF0372/2020 was isolated has been provided by Dr. X. Lescure and Pr. Y. Yazdanpanah from the Bichat Hospital. The experiments were done by Institute Pasteur. VeroE6 (C1008) cells were grown in DMEM with 10% serum and 1% penicillin to 50% confluence in 384 well format and incubated with RBDs at given concentration for 2 hrs before 0.1 MOI of SARS-CoV-2 was added for one hour. The inoculum was subsequently removed and a medium with the RBD was added. After 48 hrs of incubation, the supernatant was recovered and viral load was measured using RT-PCR with forward primer: TAATCAGACAAGGAACTGATTA, reverse primer: CGAAGGTGTGACTTCCATG. In

parallel, cell viability was assessed after 48 hrs incubation using the CellTiter Glo kit from Promega. Raw data are normalized against appropriate negative and positive controls and are expressed as the fraction of virus inhibition. The curve fit was performed using the variable Hill slope model of four parameters logistic curve:

$$\text{Response} = \text{Baseline} + (\text{Max} - \text{Baseline})/(1+10^{(\text{logEC50}-\text{Log}(C)+\text{Hill})})$$

ACE2 activity assay

Human ACE2 activity was evaluated using SensoLyte® 390 ACE2 Activity Assay Kit (ANASPEC; cat# 72086) according to manufacturer's protocol, with the following changes - assay was performed in 384 well plates with a ratio of 1:5 of the recommended volume of buffer, substrate, and inhibitor. The activity was measured on either purified ACE2 (0.75 ng; Abcam, ab151852) or on the following cell lines - HeLa transiently transfected with ACE2 (6000 cells per assay), HEK-293T stable transfected with ACE2 (GenScript M00770, 8000 cells per assay), Caco2 cells (40,000 cells per assay). To assess the effect of RBD on ACE2 activity 10 nM WT RBD or RBD-B62 were added before activity measurement. The activity rate is indicated by the slope of the plot [product/time].

Supplementary text

Optimizing the RBD domain length for yeast display and protein expression

To optimize the RBD for yeast display, we screened multiple different constructs for yeast surface expression. RBDs of different starting and termination positions were cloned in a pJYDC1 vector and their impact on expression, stability, and ACE2 binding were determined (Table S1). The RBDcon1 was the shortest construct lacking the last C-terminal loop of the RBD domain (516 – 528) and including one unpaired cysteine. This resulted in poor expression and binding. The RBDcon2 and con3 included this loop, resulting in domain stabilization and an increase both in binding and expression. Although RBDcon4 (*41*) construct demonstrated high expression yields both in yeast and Expi293F™ cells, as well as good thermo-stability, we decided not to use it in yeast display since one unpaired cysteine (C538) is close to its C-terminus and the construct contains part of the neighboring domain. We continued with the RBDcon2 and RBDcon3 constructs for yeast display and protein expression in Expi293F™ cells respectively.

Supplementary tables

Table S1 – Comparison of different RBD domains for yeast display and protein expression.

Construct	Position[a]	Number of AA	Size [kDa]	Yeast expression [mean FL1*10^3][b]	Yeast display estimated K_D [nM][c]	Melting temperature [°C]
RBDcon1	336-516	181	20.5	16.4	3.2 ± 0.1	
RBDcon2	336-528	193	21.7	32.9	1.6 ± 0.3	53.5 ± 0.3
RBDcon2b	333-528	196	22.1	37.9	1.1 ± 0.3	
RBDcon3	330-528	199	22.3	38.2	1.0 ± 0.1	
RBDcon4	319-541	223	25.1	67.6	1.2 ± 0.2	53.8 ± 0.2

[a] numbers are according to UniProtKB- P0DTC2

[b] measured in pJYDC1 (eUnaG2 fluorescence signal)

[c] Binding affinity against ACE2 was determined by FACS, with the relevant construct expressed on yeast surface.

Table S2 Analysis of mutant clones selected by yeast display.

	445	446	448	460	468	470	477	478	481	483	484	490	494	498	501	514	scr*	full*
WT	V	G	N	N	I	T	S	T	N	V	E	F	S	Q	N	S	1600	1650
477							N											710
494													P					1531
501															Y			455
31			S												Y		762	884
32							N								Y		884	936
33											K				Y		99	126
34					T						K						553	
36					T				Y						Y		184	268
39															Y		179	448
310			S								K				Y			130
312										E					Y		202	
316			S								K	S			Y		620	
41											K			R	Y		38	
42			S								K		P		Y		103	
47				K							K				Y		78	
48							N							R	Y		57	65
51											K				Y		5	
52				K							K		P	R	Y		9	59
53				K				S			K			R	Y		6	
55											K			R	Y		12	18,5
57		R									K	Y		R	Y		10	
513															Y		13	
520					V						K			R	Y		4	
521				K							K			R	Y			12
522				K							K			R	Y		4	
FA2	K			K	T	M	N				K			R	Y		1	2.5
FA10				K			N				K			R	Y		3	
FA11				K							K		P	R	Y	T	3	
FA13	K		S	K	T	M	N				K			R	Y		2	

* - The yeast display affinity was determined using 4 different concentrations of ACE2 (scr – see Fig. S3) or by full titration curve (full).

Table S3: Cryo-EM data collection and refinement statistics of ACE2-RBD62

Data collection	
EM equipment	Titan Krios (Thermo Fisher Scientific)
Voltage (kV)	300
Detector	K3 (Gatan)
Energy filter	BioQuantum (Gatan), 20 eV slit
Pixel size (Å)	0.53
Electron dose (e-/Å^2)	70
Defocus range (μm)	-0.4 to -1.7
Number of collected micrographs	4,470
Number of selected micrographs	2,535
3D Reconstruction	
Software	CryoSPRAC
Number of used particles	164,636
Resolution (Å)	2.9
Symmetry	C1
Map sharpening B factor (Å^2)	-83
PDB code	XXX
Refinement	
Software	Phenix
Cell dimensions (Å)	313.056
Model composition	
Protein residues	729
Atoms	5,873
Sugar	3
Zn	1
RMSD	
Bonds length (Å)	0.007
Bonds Angle (°)	0.687
Ramachandran plot statistics (%)	
Preferred	95.12
Allowed	4.88
Outlier	0.1

Supplementary Material Figures:

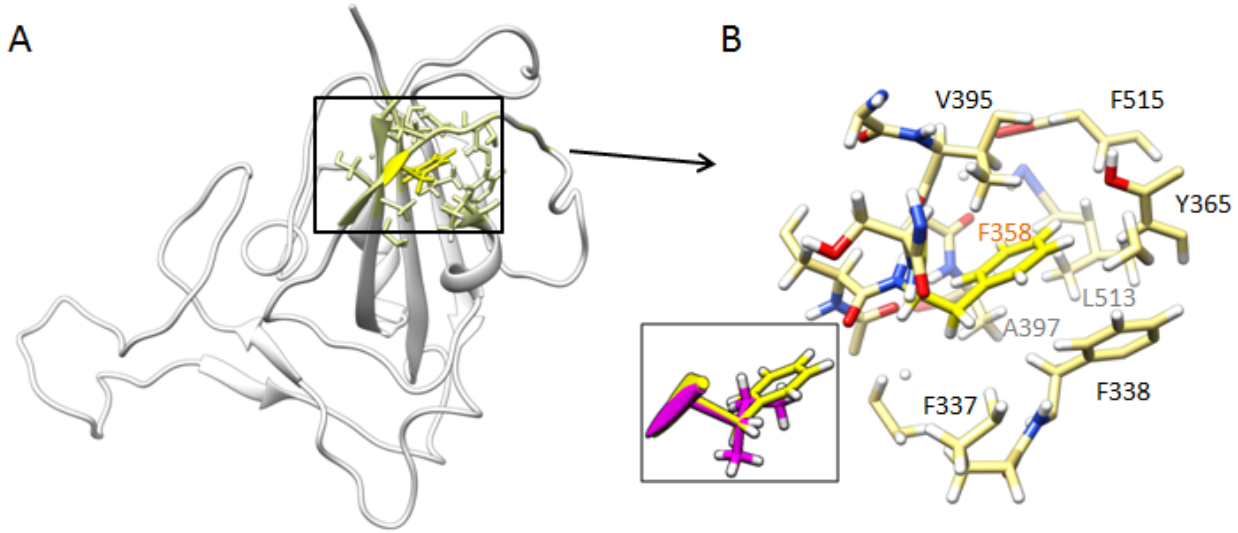

Fig. S1 The I358F mutation, selected by yeast surface display, increases protein stability and expression. A) The position of I358F (bright yellow) mutation in the RBD structure (PDB ID 6M17) and the neighboring residues within 5 Å distance (pale yellow). B) Shows the residues involved in the formation of the hydrophobic cavity around I358F mutation predicted from the X-ray structure. Additional residues that are involved: K356, R357, S359, V395, Y396. Inset – the wild-type residue (isoleucine in magenta) overlaid with the phenylalanine mutant.

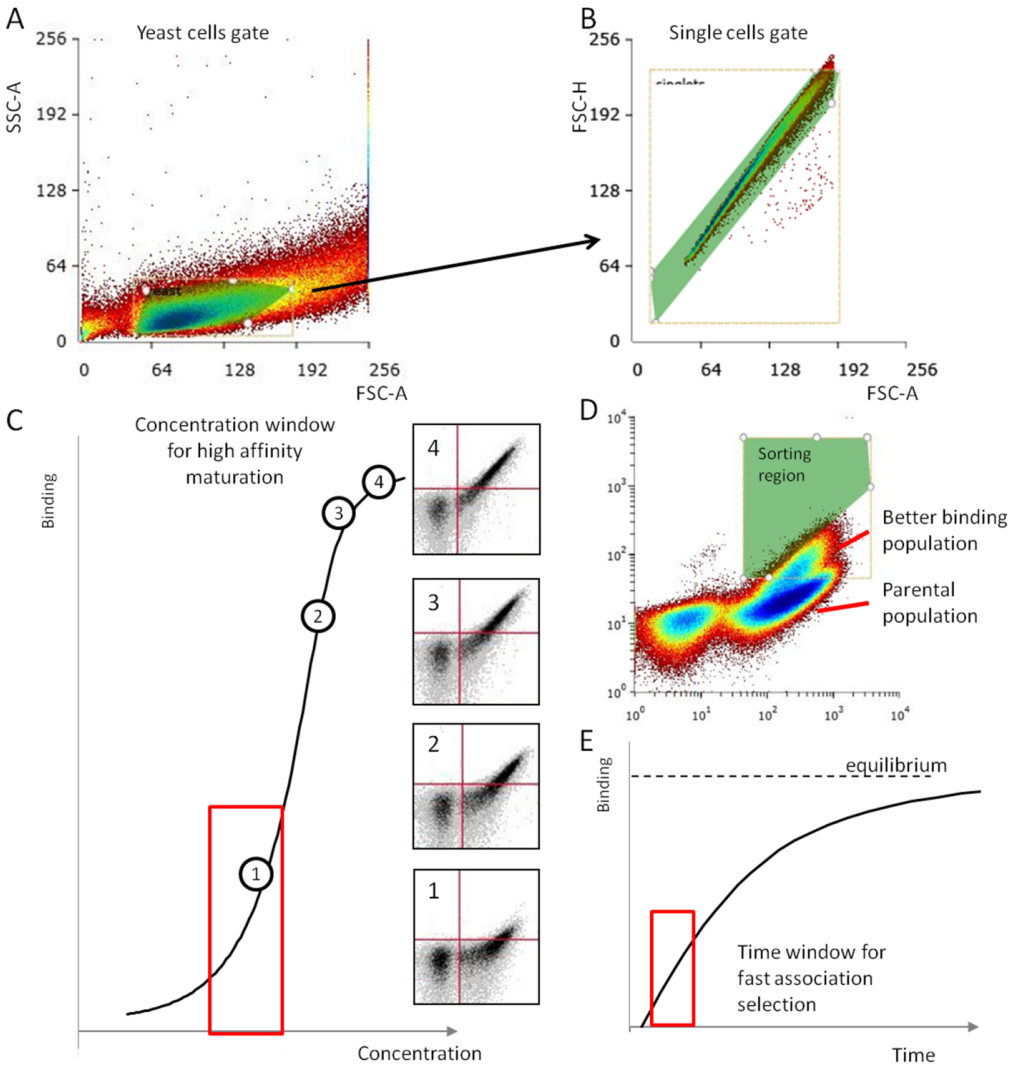

Fig. S2 Gaiting and selection strategies for *in vitro* evolution of SARS-CoV-2 RBD domain. A, B) Gating strategy for FACS sorting. In the first step, yeast cells are isolated by their FSC-A and SSC-A properties (A). In the second step (B), single cells are isolated by their FSC properties (area and height) on the diagonal plot. The Green area represents the gated region. C) Selection strategy for affinity maturation. The library was titrated with a range of ACE2 concentrations to select the concentration with limited signal (inset 1). Under such conditions, the tighter binding clones gain the highest advantage over the parental population. Using less stringent selection (insets 2 – 4) reduces the advantage of the tighter binders. Using too low concentrations of ACE2 protein will also result in loss of selectivity. D) Affinity maturation library after 3 sorts, where the separation between parental and tighter binding population is well defined. The top 0.1 – 0.3 % of cells were sorted – green region. E) Fast association selection strategy. The library was incubated with a constant concentration (30 pM) of ACE2 for a different times. The time with minimal signal was determined and used for the selection of clones with faster association. The same shape of the sorting region as in (D) was applied.

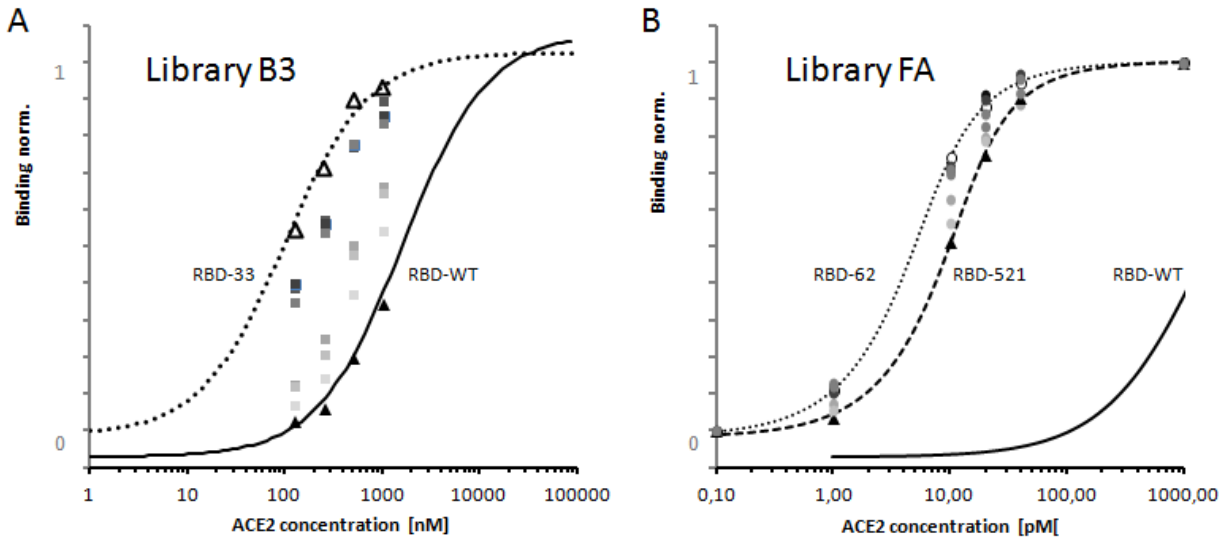

Fig. S3 Evaluating the binding affinity of 5 individual clones, from libraries B3 and FA. Five single-clones were evaluated for binding to ACE2 from each library, to determine the range of affinity maturations after FACS selection. Each clone was incubated with four (library B3) or six (FA) different concentrations of ACE2. The binding curve was fitted using additional parameters describing the curve minimum and maximum as determined from the RBD-WT titration curve. Calculated affinities are in Table S2.

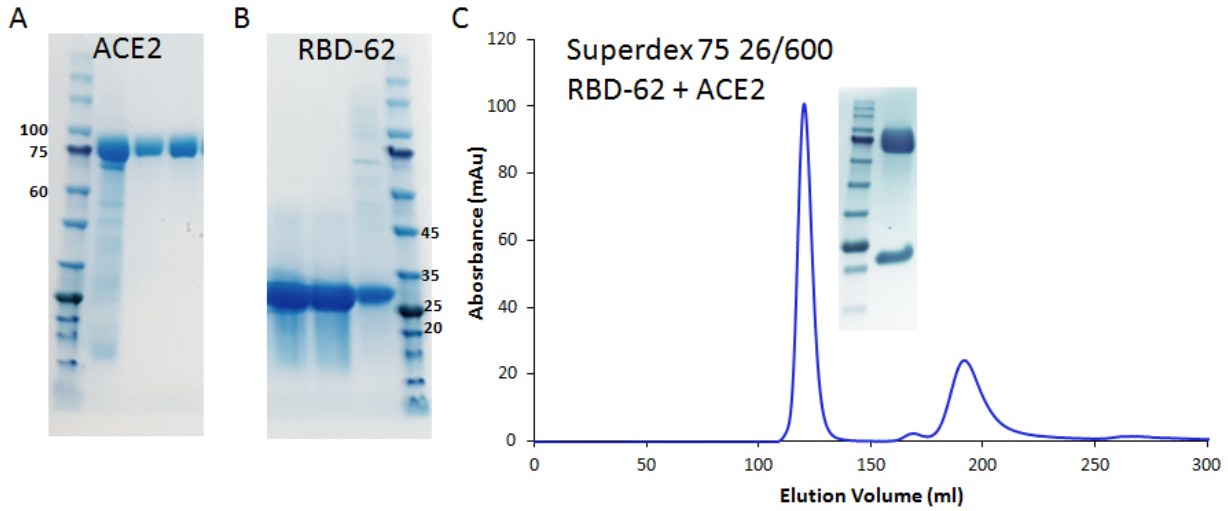

Fig. S4 Protein purification of ACE2, RBD-62 and the complex between the two. Both proteins were expressed in Expi293F cells and secreted to cell culture media. A) SDS-PAGE analysis after NiNTA agarose purification of ACE2 receptor extracellular portion (AA Q18 – S740). B) SDS-PAGE analysis from NiNTA agarose purification of RBD-62 (AA 333-528). C) The ACE2 + RBD-62 complex was purified by gel filtration chromatography column prior to CryoEM. ACE2 protein was mixed with an excess of RBD-62 (1:1.5), incubated 1h on ice, and applied on the chromatography column by using ÄKTA pure FPLC system. The first peak corresponds to the complex (SDS-gel inset) and the second peak represents excess RBD-62.

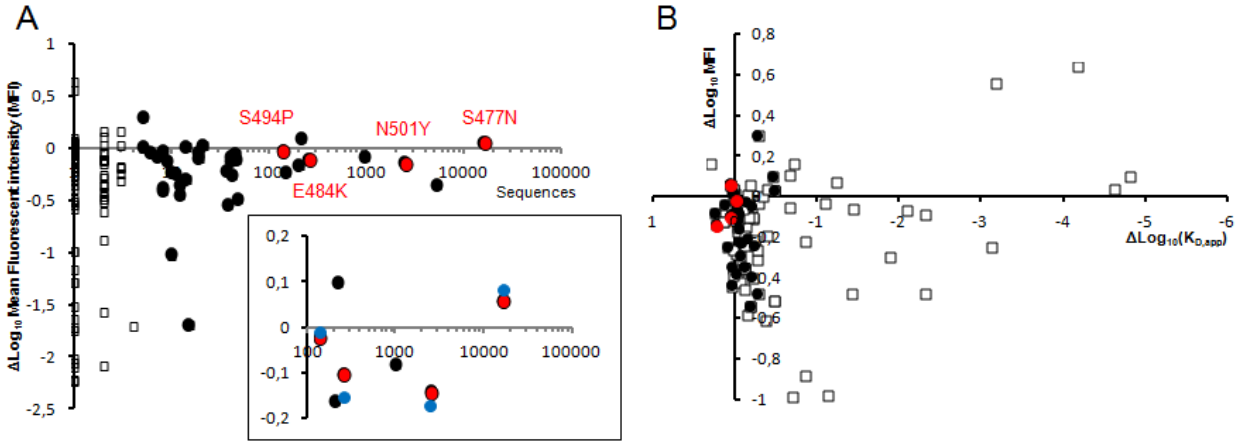

Fig. S5 SARS-CoV-2 RBD mutations in the population and their expression level. A) Relation between the impact of mutations on yeast surface expression and their occurrence in the population. Expression was measured as the mean fluorescence intensity (MFI) of the specific clone expressed on the yeast surface by Star ret al. (*6*) (black and red) or by us (blue, inset). Empty squares and black dots are showing data with < 5 or > 5 sequences recorded, respectively. The emerging mutations in the population are shown in red. The graph shows that the variance in expression decreases with higher occurrence in the population. B) Relation between the affinity (x-axis), expression (y-axis), and the occurrence in population: Empty squares < 5 sequences; black dots > 5 sequences; red dots represent four emerging mutants (all with more than 100 sequences). Based on A and B, rapidly spreading mutations have increased ACE2 binding affinity without compromising protein stability.

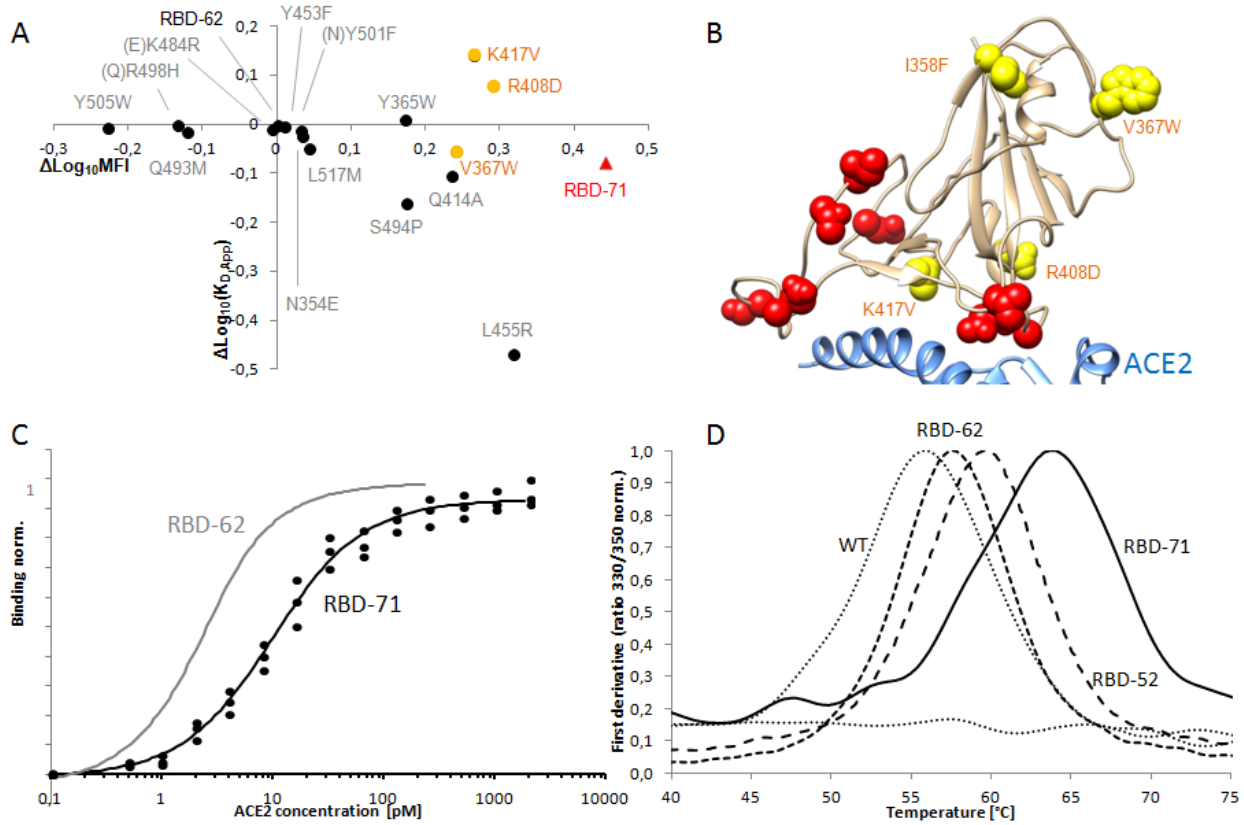

Fig. S6 Site-directed mutagenesis of RBD-62, using affinity enhancing mutations. 23 mutations were predicted to enhance RBD-ACE2 binding (*6*). These mutations were evaluated for enhancing the affinity of RBD-62 towards ACE2. A) Impact of mutations, on top of RBD-62 on ACE2 binding (y-axis) and yeast surface expression. Three mutations (orange circles), which have the highest impact on expression, were combined in RBD-71 (red triangle). B) Localization of stabilizing (yellow) and binding enhancing mutations depicted in the RBD structure (PDB ID 6m17, best rotamer is shown). C) Binding curve of RBD-71 with RBD-62 for comparison. D) Normalized protein melting curves for RBD-WT, RBD-62, RBD-52, and RBD-71 measured using the Tycho NT.6 (NanoTemper).

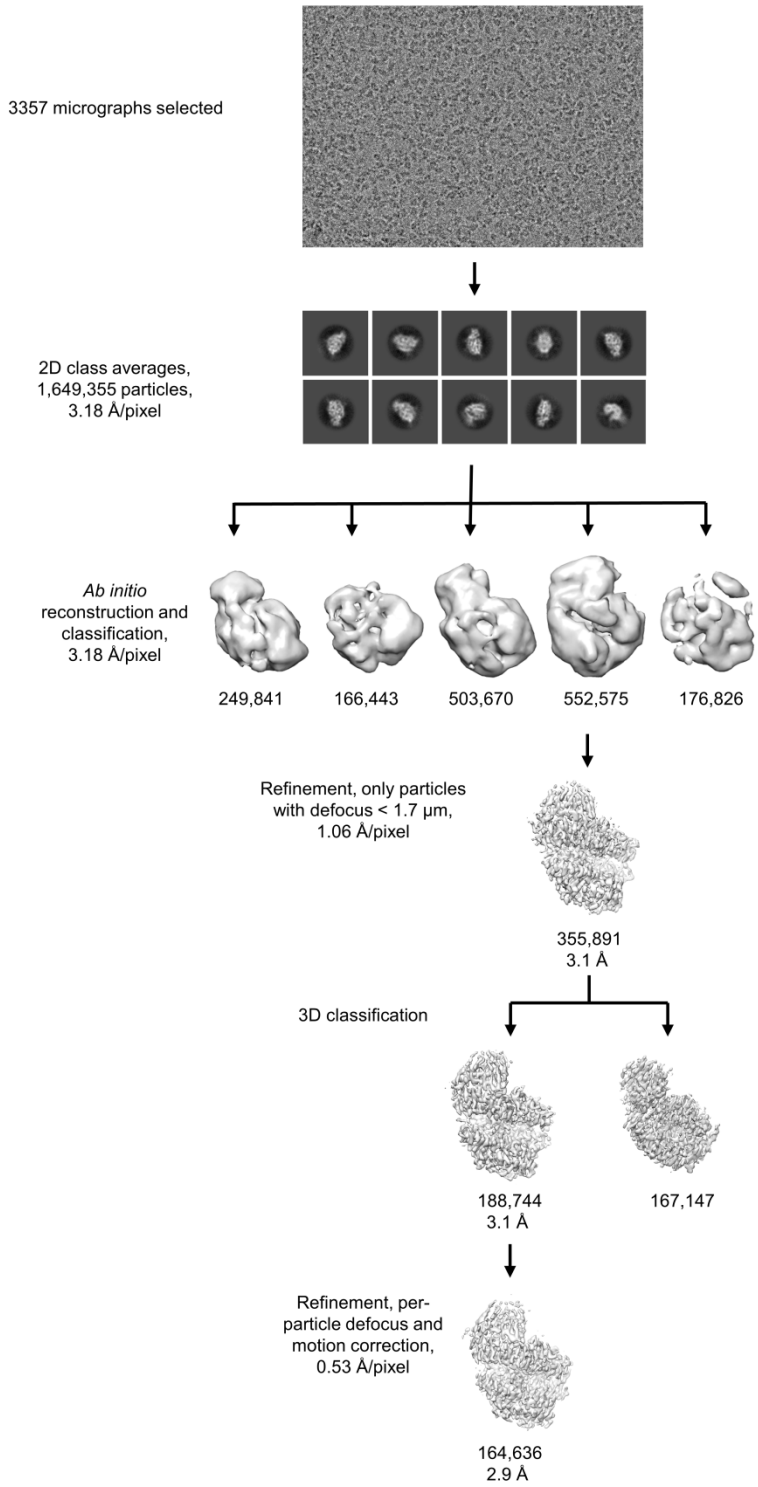

Fig. S7 Single-particle cryo-EM processing scheme. The details of the process are described in the Methods section under "Cryo-EM image processing". The number of particles in each map is indicated under the map's image, along with the map's resolution where relevant.

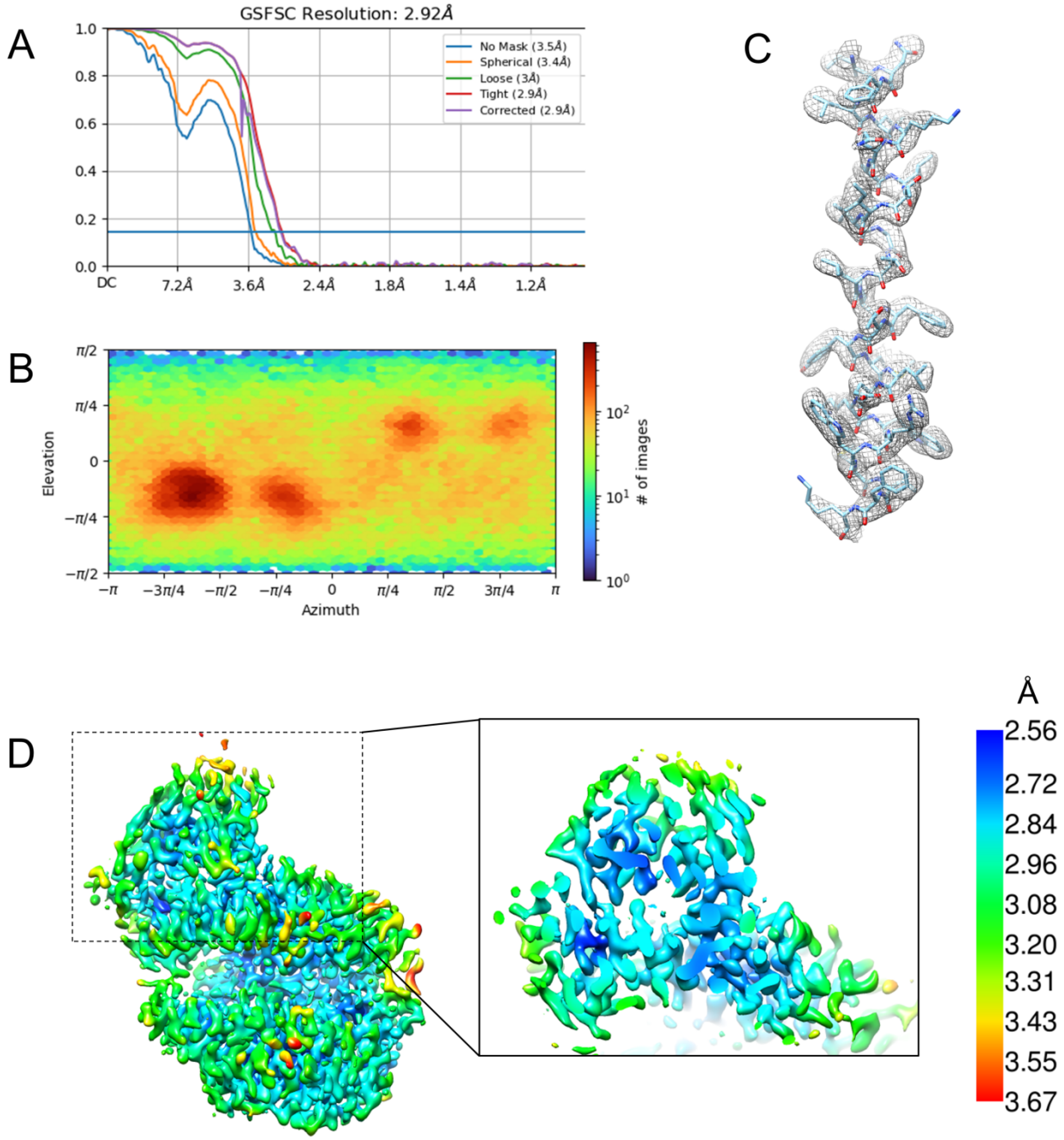

Fig. S8 Resolution estimate and angular distribution for the ACE2-RBD-62 cryo-EM map. (A) Fourier Shell Correlation (FSC) curves. (B) Angular distribution plot. (C) An alpha-helical segment showing the map density and fitted atomic coordinates. (D) Cryo-EM map colored according to local resolution estimate. The inset shows a slice through the RBD-ACE2 interface.

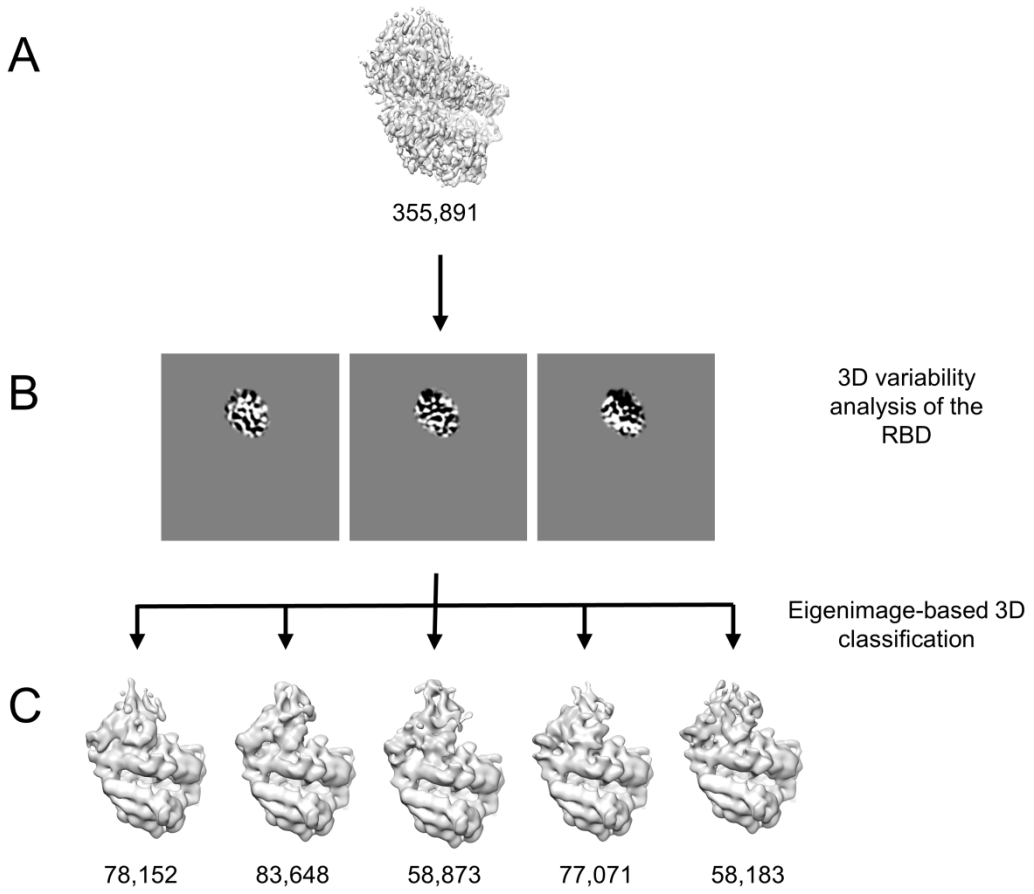

Fig. S9 Variability analysis of the RBD. (A) Particle images from the well-resolved 3D class were subjected to 3D variability analysis. (B) Central slices through the three eigenimages calculated with a binary mask around the RBD region. (C) Five 3D classes, which were calculated based on the eigenimages. The maps show variable density for the RBD.

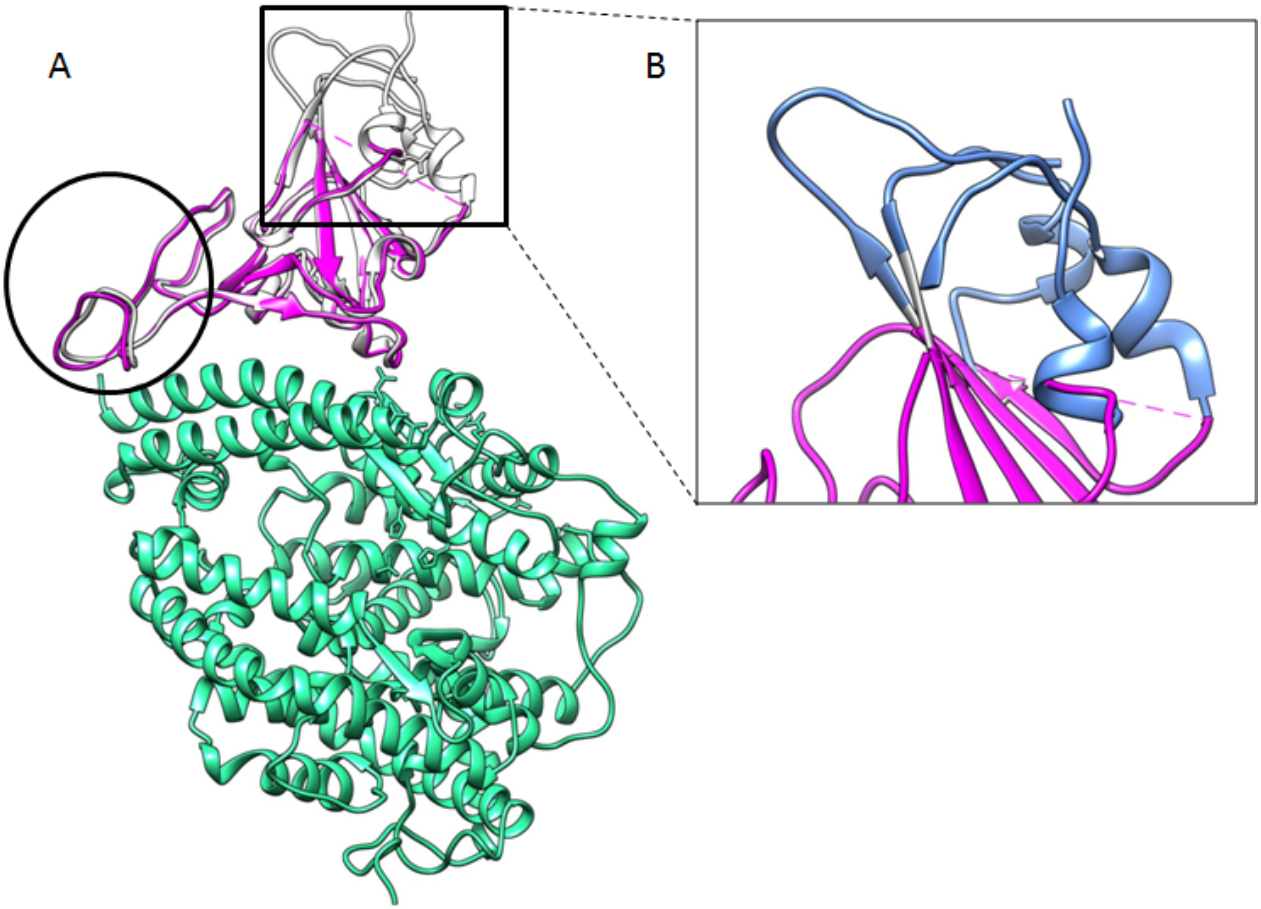

Fig. S10 Global comparison between RBD-WT and RBD-62. A) The RBD-62 preserves its typical twisted five-stranded antiparallel β sheet (β1, β3-β5, and β10) with an extended insertion containing the short β5-β9 strands, α4, and η3 helices and loops. The biggest differences are pronounced between M470 and F490 (black circle). B) The upper part comprised of three segments: R357-S371 (β2, α2), G381-V395 (α3), and F515-H534 (β11**)** is not resolved in the electron density map (blue ribbon, added from PDB ID: 6M0J).

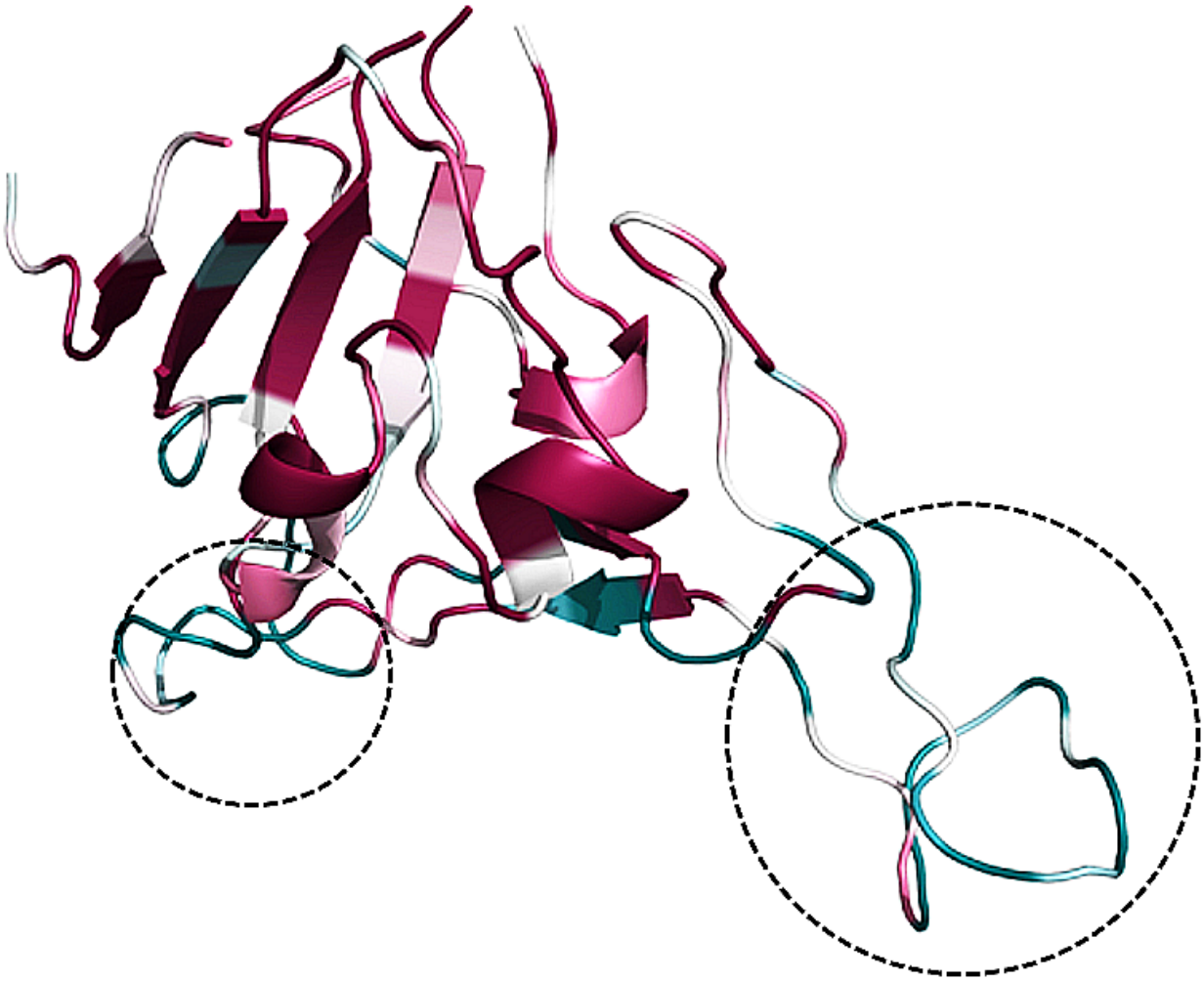

Fig. S11 An analysis of conserved positions computed by ConSurf server depicted on the RBD-62 structure. The amino acids are colored by their conservation grades with turquoise-through-maroon indicating variable-through-conserved.

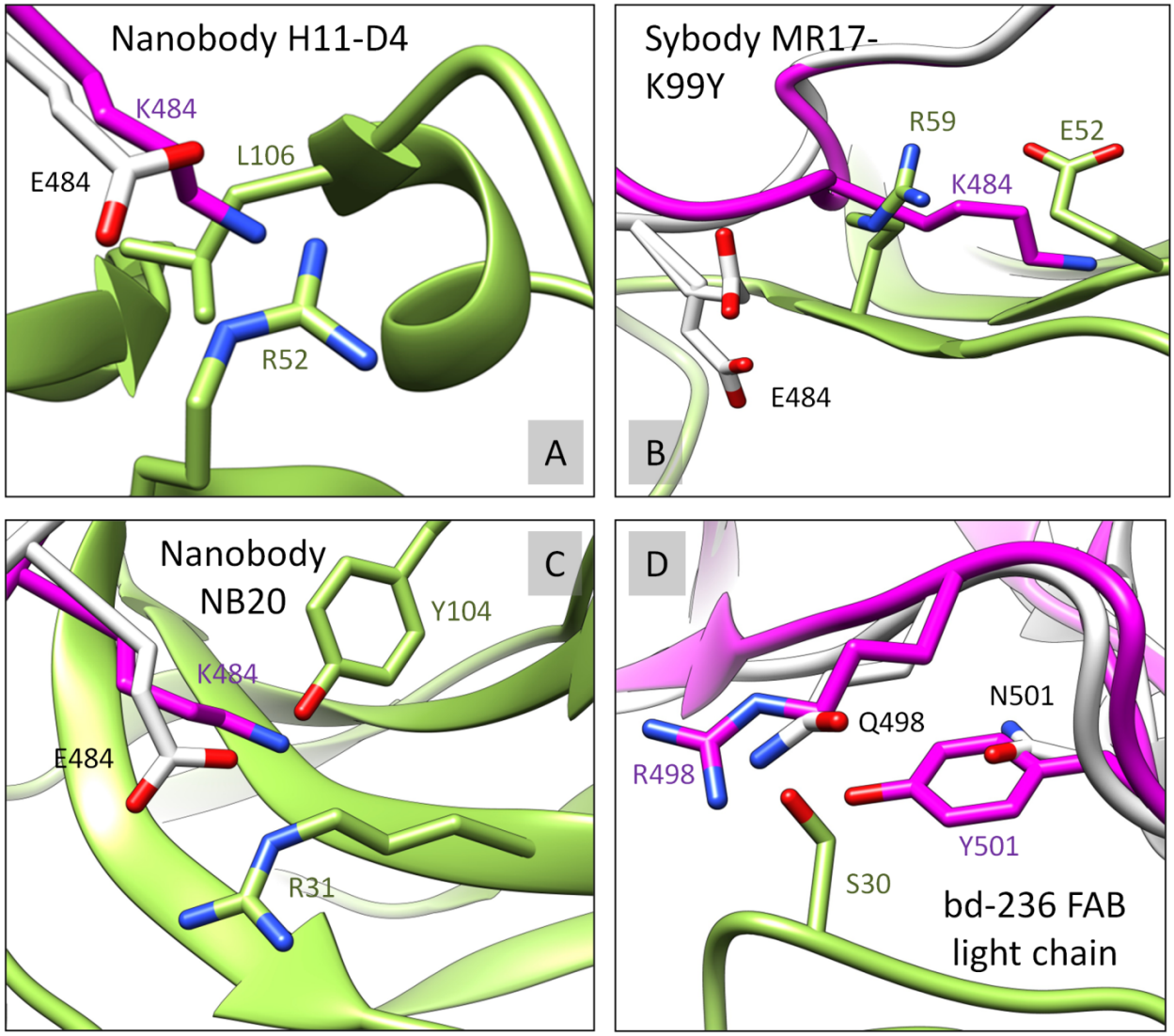

Fig. S12 RBD-62 mutations are interfering with binding to multiple antibodies. The RBD-62 (magenta) was structurally overlayed with RBD-WT (white). S477N, E484K, Q498R, and N501Y RBD mutated residues were analyzed for disruptive contacts/clashes with corresponding binding antibody/nanobody (green) in relation to RBD-WT. Four examples A) PDB ID: 6YZ5, B) PDB ID: 7CAN, C) PDB ID: 7JVB, D) PDB ID: 7CHE, where RBD-62 (but not RBD-WT) forms serious clashes with the second chain. Further experimental evaluation is needed to support our observation.