De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2

Xiaoqiang Huang[1], Robin Pearce[1], Yang Zhang[1,2]

[1]Department of Computational Medicine and Bioinformatics, Ann Arbor, MI 48109, USA
[2]Department of Biological Chemistry, University of Michigan, Ann Arbor, MI 48109, USA

Correspondence to: Yang Zhang; **email:** zhng@umich.edu

ABSTRACT

The outbreak of COVID-19 has now become a global pandemic that has severely impacted lives and economic stability. There is, however, no effective antiviral drug that can be used to treat COVID-19 to date. Built on the fact that SARS-CoV-2 initiates its entry into human cells by the receptor binding domain (RBD) of its spike protein binding to the angiotensin-converting enzyme 2 (hACE2), we extended a recently developed approach, EvoDesign, to design multiple peptide sequences that can competitively bind to the SARS-CoV-2 RBD to inhibit the virus from entering human cells. The protocol starts with the construction of a hybrid peptidic scaffold by linking two fragments grafted from the interface of the hACE2 protein (a.a. 22-44 and 351-357) with a linker glycine, which is followed by the redesign and refinement simulations of the peptide sequence to optimize its binding affinity to the interface of the SARS-CoV-2 RBD. The binding experiment analyses showed that the designed peptides exhibited a significantly stronger binding potency to hACE2 than the wild-type hACE2 receptor (with -53.35 vs. -46.46 EvoEF2 energy unit scores for the top designed and wild-type peptides, respectively). This study demonstrates a new avenue to utilize computationally designed peptide motifs to treat the COVID-19 disease by blocking the critical spike-RBD and hACE2 interactions.

INTRODUCTION

The continuing pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, previously known as 2019-nCoV) has now become an international public health threat, causing inconceivable loss of lives and economic instability [1]. As of May 17, 2020, there have been more than 4500000 confirmed cases and over 300000 deaths caused by COVID-19 worldwide [2]. Exacerbating the problem, there is no specific antiviral medication toward COVID-19, though development efforts are underway [3–6]. Although vaccines are thought to be the most powerful weapon to fight against virus invasion, it may take quite a long time to develop and clinically test the safety of a vaccine. Moreover, vaccines are usually limited as preventative measures

given to uninfected individuals. Thus, as an emergency measure, it is desirable to develop effective antiviral therapeutics that can take effect rapidly not only to treat COVID-19, but also to prevent its further transmission.

It has been confirmed that SARS-CoV-2 initiates its entry into host cells by binding to the angiotensin-converting enzyme 2 (ACE2) via the receptor binding domain (RBD) of its spike protein [7, 8]. Therefore, it is possible to develop new therapeutics to block SARS-CoV-2 from binding to ACE2. Although small molecule compounds are commonly preferred as therapeutics, they are not effective at blocking protein-protein interactions (PPIs) where a deep binding pocket may be missing at the interface [9]. On the contrary, peptide binders are more suitable for disrupting PPIs by specifically binding to the interface binding region [10]. Also of importance, small

peptides have reduced immunogenicity [11]. These positive features make peptides great candidates to serve as therapeutics [12, 13]. Recently, Zhang et al. [14] reported that the natural 23-mer peptide (a.a. 21-43) cut from the human ACE2 (hACE2) α1 helix can strongly bind to SARS-CoV-2 RBD with a disassociation constant (K_d) of 47 nM, which was comparable to that of the full-length hACE2 binding to SARS-CoV-2 RBD [15]; they also showed that a shorter 12-mer peptide (a.a. 27-38) from the same helix was not able to bind the virus RBD. In an earlier report, Han et al. [16] performed a study to identify the critical determinants on hACE2 for SARS-CoV entry, and they found that two natural peptides from hACE2 (a.a. 22-44 and 22-57) exhibited a modest antiviral activity and inhibited the binding of SARS-CoV RBD to hACE2 with IC50 values of about 50 μM and 6 μM, respectively, implying that the peptide composed of residues 22-57 had a stronger binding affinity for SARS-CoV RBD. They also generated a peptide by linking two discontinuous fragments from hACE2 (a.a. 22-44 and 351-357) with a glycine, and this 31-mer exhibited a potent antiviral activity with an IC50 of about 0.1 μM, indicating that this artificial peptide had a much stronger binding affinity for SARS-CoV RBD than the peptides composed of residues 22-44 or 22-57. Due to the high similarity of the binding interfaces between SARS-CoV RBD/hACE2 and SARS-CoV-2 RBD/hACE2, we hypothesize that this artificial peptide may also bind to SARS-CoV-2 more strongly than the peptide 21-43 tested by Zhang et al. [14], which is similar to the peptide 22-44 from Han et al. [16]. Although the natural peptides are promising, it has been argued that the sequence of hACE2 is suboptimal for binding the S protein of SARS-CoV-2 [17]. Therefore, further redesign of the natural peptides may significantly enhance its binding affinity to the virus RBD and the improved peptide binders may have the potential to inhibit SARS-CoV-2 from entering human cells and hinder its rapid transmission.

In this work, we computationally designed thousands of peptide binders that exhibited a stronger binding affinity for SARS-CoV-2 than the natural peptides through computational experiments. Based on the crystal structure of the SARS-CoV-2 RBD/hACE2 complex, we constructed a hybrid peptide by linking two peptidic fragments from hACE2 (a.a. 22-44 and 351-357) with a glycine. Starting from the peptide-protein complex, we used our protein design approaches, EvoEF2 [18] and EvoDesign [19], to completely redesign the amino acid sequences that match the peptide scaffold while enhancing its binding affinity for SARS-CoV-2. Detailed analyses support the strong binding potency of the designed binders, which not only recapitulated the critical native binding interactions but also introduced new favorable interactions to enhance binding. Due to the urgency caused by COVID-19, we share these computational peptides to the community, which may be helpful for further developing antiviral peptide therapeutics to combat this pandemic.

RESULTS

Initial peptide scaffold construction

Several experimental SARS-CoV-2 RBD/hACE2 complex structures have been reported [20–22] and deposited in the Protein Data Bank (PDB) [23]. Specifically, PDB ID 6m17 is a 2.9 Å structure of the SARS-CoV-2 RBD/ACE2-B0AT1 complex determined using cryogenic electron microscopy (Cryo-EM) [22]. Furthermore, PDB ID 6m0j is a 2.45 Å X-ray crystal structure of SARS-CoV-2 RBD/hACE2 [20], while 6vw1 is a 2.68 Å X-ray structure of SARS-CoV-2 chimeric RBD/hACE2 [21], where the chimeric RBD is comprised of the receptor binding motif (RBM) from SARS-CoV-2 S and the core from SARS-CoV, with the mutation N439R. The three experimental complex structures are quite similar to each other in terms of global folds (Figure 1A). Since 6vw1 does not contain the wild-type SARS-CoV-2 RBD, we did not use it as a template. Based on a preliminary examination, we found that the structure quality of 6m0j was better than 6m17 (see below), and therefore we only considered 6m0j as the template complex.

Two peptide fragments (a.a. 22-44 and 351-357) from hACE2 (6m0j, chain A) were extracted because they formed extensive contacts with the SARS-CoV-2 RBD (6m0j, chain E). The positions 44 and 351 were chosen because the distance between their Cα atoms was only 5.5 Å (see Supplementary Figure 1), and therefore only one residue was required to link them. To reduce the interference to the surrounding amino acids, the linker residue was initially chosen as glycine. The small loop, 44S-glycine-351L, was then reconstructed using MODELLER [24], while the other parts of the whole peptide were kept constant. Five similar loop conformations were produced and the one with the best DOPE [25] score was selected, where DOPE is a built-in scoring function in the MODELLER package for model assessment and loop modeling. For the sake of simplifying the discussion, the initial hybrid peptide constructed in this manner was denoted as the wild-type (note that it was not a truly native peptide), and the complex structure of SARS-CoV-2 RBD/hACE2 hybrid peptide was used as the template for computational peptide design (Figure 1B).

Evaluation of EvoEF2 score on experimental complexes

At the very beginning of the outbreak of SARS-CoV-2, to determine its relative infectivity, many computational

studies were performed to compare the binding affinity of SARS-CoV-2 RBD for hACE2 with that of SARS-CoV RBD for hACE2 based on homology modeling structures; all these studies came up with the conclusion that SARS-CoV-2 showed much weaker binding affinity to hACE2 than SARS-CoV and SARS-CoV-2 might not be as infectious as SARS-CoV [26–28]. However, recent biochemical studies demonstrated that SARS-CoV-2 exhibits much stronger binding affinity to hACE2 than SARS-CoV [3, 15, 21], implying that the homology models may not have been sufficiently accurate for binding affinity assessment based on atomic-level scoring functions, although the global folds of these models were correct.

Here, we used the EvoEF2 energy function to evaluate the binding affinity of SARS-CoV and/or SARS-CoV-2 (chimeric) RBD for hACE2 based on the experimental structures described above. As shown in Table 1, SARS-CoV-2 RBD showed stronger binding potency (lower EvoEF2 scores indicate stronger binding affinity) to hACE2 than SARS-CoV based on the calculations performed on two X-ray crystal structures (PDB IDs: 2ajf and 6m0j), regardless of whether or not the residues at the protein-protein interfaces were repacked; the computational estimations were consistent with the experimental results (Table 1). However, the EvoEF2 binding scores calculated using the Cryo-EM structure (i.e. 6m17) were much higher than those obtained from the X-ray structure 6m0j, suggesting that the Cryo-EM structure might not be as high quality as its X-ray counterparts. We examined the possible steric clashes in these experimental structures using a criterion of $d_{ij} < 0.7(R_i+R_j)$, where d_{ij} is the distance between non-hydrogen atoms i and j, R_i and R_j are the van der Waals radii for i and j, respectively. A clash was counted if the formula holds. The d_{ij} values were calculated from the atom coordinates in the experimental structures and the

van der Waals radii were adapted from the EvoEF2 force field [18]. Five clashes were detected in 6m17 but none in 6m0j or 2ajf according to this criterion. Moreover, Shang et al. [21] demonstrated that the artificial SARS-CoV-2 chimeric RBD showed improved binding affinity to hACE2, compared to the wild-type SARS-CoV-2, and this improvement was also somewhat captured by EvoEF2 (Table 1). Thus, out of the two wild-type SARS-CoV-2 RBD/hACE2 structures (6m0j and 6m17), only 6m0j was used as a template structure for the peptide design study because it was better refined.

Peptide design based on the physical score

Eight out of the 1000 low-energy sequences that were designed using the EvoEF2 energy function were duplicates, resulting in 992 non-redundant designs. The EvoEF2 total energy values of the designed protein complex structures ranged from -829 to -816 EvoEF2 energy units (EEU), the majority of which varied from -827 to -822 EEU (Figure 2A). The EvoEF2 binding energies of the 992 designed peptides to SARS-CoV-2 RBD ranged from -53 to -40 EEU, centering around -50 to -47 EEU (Figure 2B). The sequence identities between the designed peptides and the wild-type peptide was diversely distributed, varying from 15% to 50% and centering around 37% (Figure 2C), which was much higher than the sequence recapitulation rate obtained for the protein surface residues during the benchmarking of EvoEF2 [18]. Although the peptide residues were considered to be highly exposed, the high sequence identity revealed that a large number of critical binding residues should be correctly predicted, indicating that the designed peptides are reasonable.

The wild-type peptide showed an EvoEF2 binding energy of -46.46 EEU, whereas the total energy of the
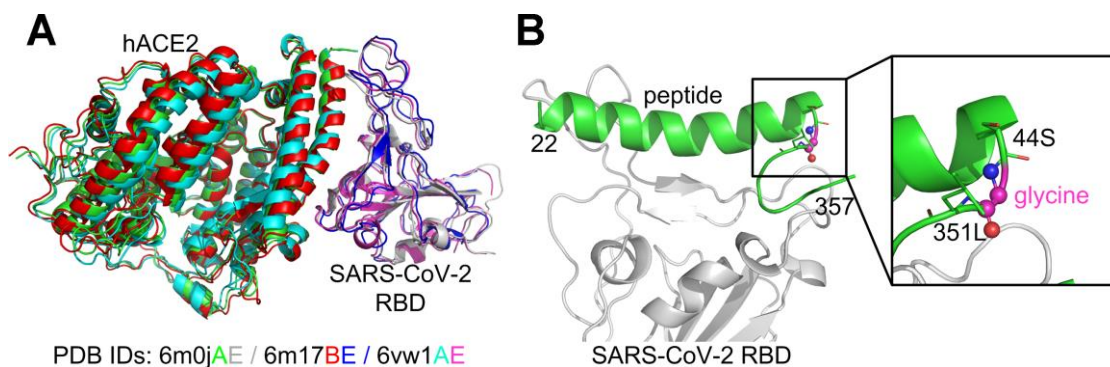
Figure 1. Comparison of the SARS-CoV-2 RBD/hACE2 complex structures (**A**) and the constructed SARS-CoV-2 RBD/hACE2 peptide complex (**B**). The superposition of the three complex structures was performed using MM-align [45]; the TM-score [46] between each complex pair was >0.98.

Table 1. Comparison of binding affinities for different PPIs.

PPI	Experiment K_d (nM)	EvoEF2 score (EEU)	
		Interface not repacked	Interface repacked
SARS-CoV RBD/hACE2	325.8 [15] 185 [21]	-40.73 (2ajfAE)	-51.12 (2ajfAE)
SARS-CoV-2 RBD/hACE2	14.7 [15] 44.2 [21]	-49.95 (6m0jAE) -19.84 (6m17BE) -19.84 (6m17DF)	-55.67 (6m0jAE) -30.50 (6m17BE) -30.50 (6m17DF)
SARS-CoV-2 chimeric RBD/hACE2	23.2 [21]	-53.15 (6vw1AE)	-58.81 (6vw1AE)

EEU stands for EvoEF2 energy unit.

wild-type peptide/SARS-CoV-2 RBD complex was -802 EEU (Figure 2D). 757 out of the 992 designs exhibited better binding affinities to SARS-CoV-2 RBD and showed lower total energies than the wild-type, and some designs showed good binding and stability simultaneously (Figure 2D), indicating that the wild-type peptide can be improved through design. Figure 2E illustrates the binding energy as a function of sequence identity for the designed peptides; it illustrates that a majority of the designs showed weaker binding affinity to SARS-CoV-2 than the wild-type peptide when the sequence identity was <25%, whereas most of the designs with sequence identities >35% exhibited stronger binding to SARS-CoV-2. These results suggest that, in general, low sequence identity designs may not be as good as high sequence identity designs. However, we can also see from Figure 2E that it does not necessarily mean that higher sequence identity always ensures better designs, since the two designs with the highest sequence identity (15/31=48.4%) did not always show stronger binding than those with sequence identities around 35%. Thus, the results suggest that good binders showed a high similarity to the wild-type, but the similarity should not be too high in order to leave room for the designs to be improved. This is in line with the common thinking that the critical binding residues (i.e. hot spot residues) should be conserved while some other residues can be mutated to enhance binding. Note that the wild-type peptide was comprised of a helix (a.a. 22-44) and a short loop (a.a. 351-357) with a glycine linker. To ensure good binding to SARS-CoV-2 RBD, the designed peptides should be able to preserve the secondary structure of this motif. To check this point, we used an artificial neural network-based secondary structure predictor [29] implemented in EvoDesign to predict the secondary structure of the designed peptides; the predictor that we used here was much faster than some other state-of-the-art predictors, e.g. PSIPRED [30] and PSSpred [31], but showed similar performance [29]. To quantify the similarity between the secondary structure of a designed peptide and that of the wild-type, we calculated the secondary structure match rate, which was defined as the ratio of the number of residues with correctly assigned

secondary structure elements (i.e. helix, strand, and coil) to the total number of residues (i.e. 31). As shown in Figure 2F, 892 out of the 992 designed peptides had >90% secondary structure elements predicted to be identical to that of the wild-type peptide, indicating the high accuracy of the designs, although the EvoEF2 scoring function does not include any explicit secondary structure-related energy terms [18].

We used WebLogo [32] to perform a sequence logo analysis for the 992 designed sequences to investigate the residue substitutions and the results are shown in Figure 3A. 16 residues from the initial peptide scaffold were at the protein-peptide surface in contact with residues from SARS-CoV-2 RBD; these residues were Q24, T27, F28, D30, K31, H34, E35, E37, D38, F40, Y41, Q42, K353, G354, D355, and R357. Of these residues, Q24, D30, E35, E37, D38, Y41, Q42, and K353 formed hydrogen bonds or ion bridges with the binding partner (i.e. SARS-CoV-2 RBD) and the designed residues at these positions maintained favorable binding interactions. As shown in Figure 3A, the native residue types at these positions were top ranked out of all 20 canonical amino acids, suggesting that these residues may play critical roles in binding. For the nonpolar residues that were originally buried in the hACE2 structure (e.g. A25, L29, F32, L39, L351, and F356), they were likely to be mutated into polar or charged amino acids (Figure 3A), because they were largely exposed to the bulk solvent. The three glycine residues, including the one that was artificially introduced, were conserved, probably due to the narrow space at these positions.

To further examine what interactions improved the binding affinity of most designs, we carried out a detailed examination of some designed structures. We found that favorable hydrogen bonds or hydrophobic interactions were introduced in the binder that had the lowest EvoEF2 binding score (Figure 3B, 3C); the amino acid sequence of this binder was "EQEERI QQDKRKNEQEDKRYQRYGRGKGHQP". For this design, T27 was mutated to isoleucine (Figure 3B). In the wild-type structure, the threonine was enveloped by

four hydrophobic residues on SARS-CoV-2 RBD (i.e. Y489, F456, Y473 and A475), but its hydroxyl group did not form any hydrogen bonds with the hydroxyl group of either Y489 or Y473, and the mutation enhanced the favorable burial of nonpolar groups. The interface residue H34 was substituted for asparagine (Figure 3B), introducing a hydrogen bond to Y453 on SARS-CoV-2 RBD. Additionally, two mutations,

F28Q and Q24E, simultaneously formed hydrogen bonds with the amide group of N487 from SARS-CoV-2 RBD (Figure 3C). Although the mutation D355H did not form hydrogen bonds with any residues from SARS-CoV-2, it simultaneously formed two hydrogen bonds with the hydroxyl group of Y41 and the main-chain carbonyl group of G45 on the peptide, which may help stabilize the loop region (a.a. 351-357).
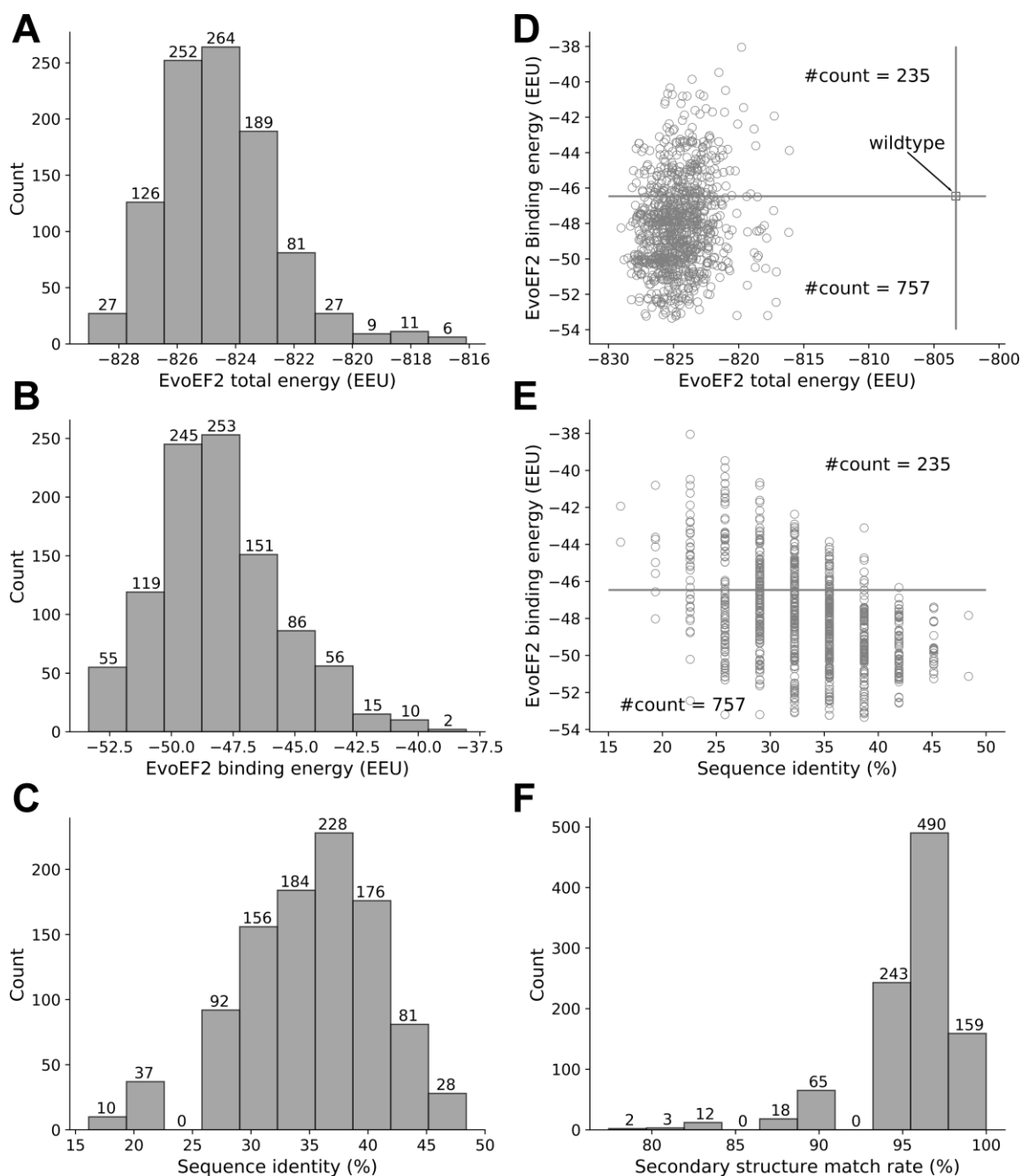

Figure 2. Overview of the characteristics of the EvoEF2 designs. (A) Distribution of total energy, (B) distribution of binding energy, (C) distribution of sequence identity, (D) binding energy as a function of total energy, (E) binding energy as a function of sequence identity, and (F) distribution of secondary structure match rate.

Peptide design based on the physical and evolutionary score

In previous studies, we found that evolutionary information can facilitate the design of proteins, improving their ability to fold into desired structures [29, 33]. To examine whether the evolutionary profile is important for peptide design here, we also performed four sets of designs with different weight settings for the evolution energy; for each design set, 1000 independent design simulation trajectories were carried out and the unique sequences out of the 1000 lowest energy designs were analyzed (Table 2). In general, giving a higher weight to the evolutionary energy facilitated the convergence of the design simulations, as indicated by the reduced number of unique designed sequences. It also helped identify sequences that were closer to the wild-type peptide as demonstrated by the higher sequence identities and the lower average evolutionary energy, which were both much more similar to those of the wild-type than the designs created using the physical score alone. We also found that incorporation of the profile energy moderately increased the ability of the designed sequences to
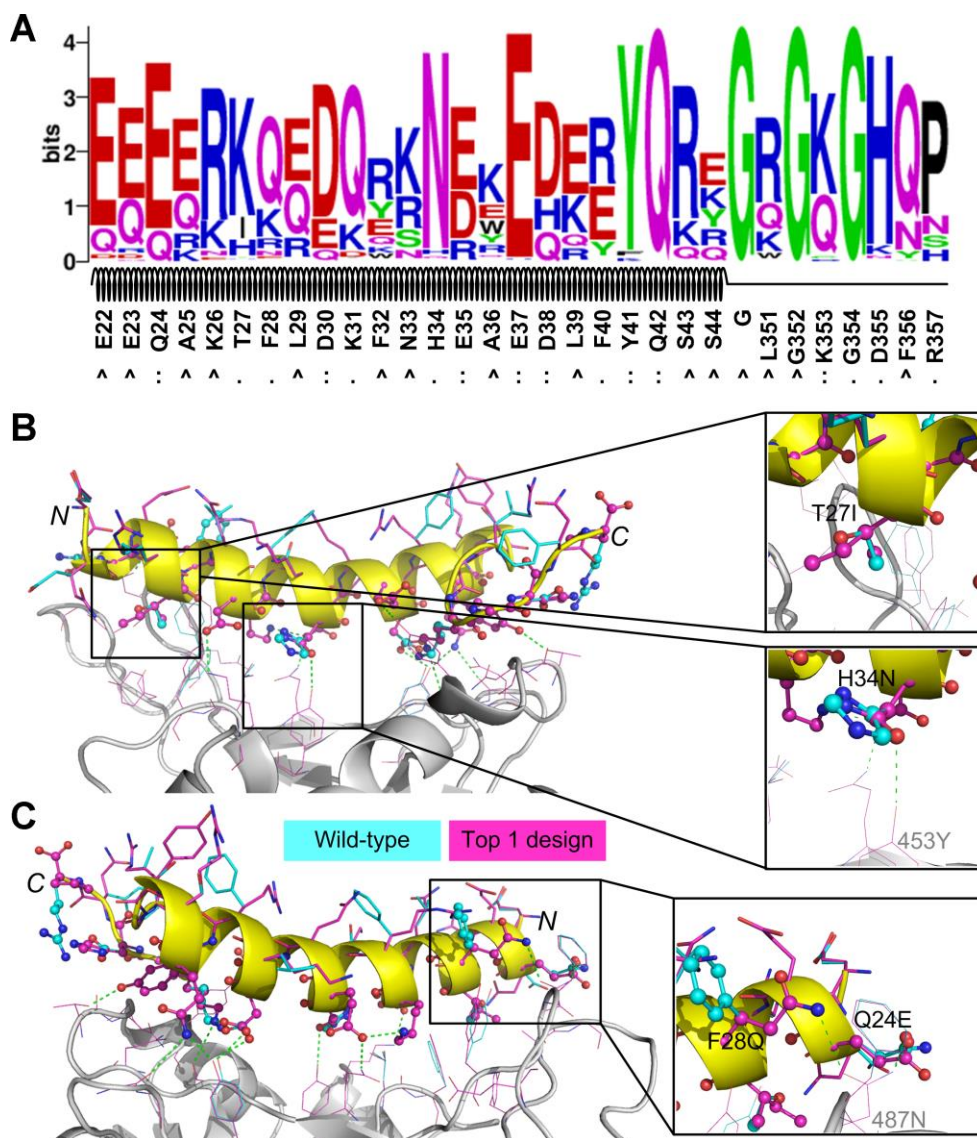
Figure 3. Sequence logo analysis of 992 unique peptide binders designed by EvoEF2 (**A**) and favorable interactions introduced in the top binder (B and C). In figure (A), the interface residues on the wild-type peptide are marked with ':' if hydrogen bonds or ion bridges exist, or '.' otherwise; non-interface residues are marked with '^'. In figures (**B**) and (**C**), the residues on the wild-type and designed structures are colored in cyan and magenta, respectively; interface and non-interface residues on the peptide are shown in ball-and-stick and stick models, respectively, while residues on SARS-CoV-2 RBD are shown in lines. Hydrogen bonds and/or ion bridges are shown using green-dashed lines.

Table 2. Summary of evolution-based peptide design results.

Comparison items [a]	Weight of evolutionary profile energy				
	0.00	0.25	0.50	0.75	1.00
Number of unique designs	992	991	966	877	695
Number of better binders [b]	757	636	392	340	226
EvoEF2 binding energy	-48.1±2.5	-47.2±2.2	-46.1±1.7	-45.8±1.6	-45.5±1.6
EvoEF2 total energy	-824.6±2.0	-823.4±2.2	-818.4±2.3	-813.3±1.9	-809.7±2.3
Profile energy [c]	6.7±2.7	-0.8±3.6	-13.3±3.3	-21.6±2.0	-25.6±1.6
EvoEF2+profile energy	-824.6±2.0	-823.6±1.8	-825.0±1.4	-829.5±1.2	-835.3±1.4
Sequence identity (%)	33.7±5.6	39.1±5.5	44.2±5.1	46.2±5.6	48.3±6.0
Sec. Str. match rate (%) [d]	95.7±3.3	96.2±3.0	97.5±2.7	97.5±2.5	97.7±2.4

[a] The units for the EvoEF2 and profile energies are EEU. [b] The EvoEF2 binding energy of the wild-type peptide binder was -46.46 EEU; this row shows the number of designed peptide binders with EvoEF2 binding energies lower than -46.46 EEU. [c] The profile energy of wild-type peptide binder was -22.2 EEU. [d] Secondary structure match rate.

maintain the original secondary structure. However, despite these improvements, giving a higher value to the profile weight hindered the identification of binders that exhibited better binding energy than the wild-type.

We performed sequence logo analyses of the four sets of designs obtained from the evolution-based method and the results are illustrated in Figure 4. Overall, the evolutionary profile did not have a dramatic effect on most interface residues (e.g. Q24, K31, H34, E35, E37, D38, Y41, Q42 and K353), because the dominating residue types identified in the EvoEF2-based designs were also top ranked (Figure 3A and Figure 4). However, some interface residues were indeed influenced. For instance, T27 could be substituted for either lysine or isoleucine without evolution (Figure 3A), but it was only mutated to lysine when the evolutionary weight was ≥0.75 (Figure 4C–4D). Additionally, without evolutionary profiles, F28 preferred glutamine over all other residues (Figure 3A), but it was conserved as phenylalanine when the evolutionary weight was ≥0.5 (Figure 4B–4D). The naturally occurring residues, glutamic acid, and arginine never appeared at positions 335 and 337, respectively, without evolutionary profile-guided design (Figure 3A); however, both of them were ranked second when a weight of 1.0 was given to the profiles. The residues that were most affected by evolution were those nonpolar residues that were not at the interface (e.g. A25, L29, F32, A36, L39, L351, and F356); without the evolutionary profile, polar or charged residue types were preferred at these positions (Figure 3A), while nonpolar residues were more frequently chosen for most of them when the weight of the profile energy was high (Figure 4B–4D). As discussed above, most of these residues were buried in the original hACE2 structure, but they were solvent exposed in the peptide, and therefore it might not be necessary to maintain the hydrophobic nature at these positions.

DISCUSSION

Although many different strategies are being employed to develop therapeutics or vaccines to treat COVID-19, there are, however, no effective antiviral drugs to combat the pandemic at present. Built on the fact that SARS-CoV-2 initiates its entry into human cells by the RBD of its spike protein binding to hACE2 [7, 8], we believe that molecules that can effectively block association of the SARS-CoV-2 spike protein with hACE2 may have the potential to treat COVID-19. In this regard, we extended a recently developed protein design approach, EvoDesign [19], to design novel peptides that can competitively bind to the SARS-CoV-2 RBD to inhibit the virus from entering human cells.

We constructed a novel hybrid peptide by linking two discontinuous peptide fragments from hACE2 with a linker glycine (denoted as 22-44G351-357), and utilized it as a template for designing new sequences with enhanced binding affinities for SARS-CoV-2 RBD. Based on the previous work by Han et al. [16], a peptide constructed using a similar approach exhibited a potent antiviral activity with an IC50 of about 0.1 μM when inhibiting the binding of SARS-CoV to hACE2, which was much higher that of two other peptides (a.a. 22-44 and 22-57). Since both SARS-CoV-2 and SARS-CoV use hACE2 as the receptor for entry into human cells and SARS-CoV-2 has much stronger binding toward hACE2 than SARS-CoV [15, 21], we believe that the wild-type hybrid peptide may also possess a high antiviral activity for inhibiting SARS-CoV-2 from binding to hACE2. Recently, Zhang et al. [14] reported that a natural hACE2 peptide (a.a. 21-43) can strongly bind to SARS-CoV-2 RBD with a K_d of 47 nM. We believe that the binding affinity of this peptide to SARS-CoV-2 may be weaker than peptide 22-44G351-357, because essentially it is almost identical to the natural hACE2 peptide 22-44 with only one residue

shifted, and Han et al. [16] demonstrated that peptide 22-44 showed much weaker binding to SARS-CoV than peptide 22-44G351-357. Therefore, it may be more promising to perform *de novo* sequence design starting with 22-44G351-357.

Computational design experiments showed that the binding energy of the peptide for SARS-CoV-2 RBD could be significantly enhanced, though the wild-type peptide already attained a good binding affinity. For instance, the wild-type peptide had an EvoEF2 binding score of -46.46 EEU, while the top designed binder achieved a score of -53.35 EEU. In contrast, the peptide used by Zhang et al. [14] had a binding score of only -37.37 EEU in our computational experiment. In the EvoDesign procedure, new peptides were designed starting from randomly generated sequences, where no wild-type sequence information was used [19]. However, sequence logo analysis suggested that the wild-type amino acid types were quite conserved for a large number of positions at the protein-peptide interface (Figures 3 and 4), indicating that some residues were critical for binding and they were correctly recapitulated by our design approach. Detailed inspection confirmed this point and also revealed that some extra favorable interactions were introduced to enhance binding in the top designed binders. Most of the *de novo* designed peptide binders shared a sequence identity of >30% to the wild-type peptide. This, on the one hand, indicates that our protein design potential was of high accuracy, and on the other hand, implies that good binders should not be random, and interestingly
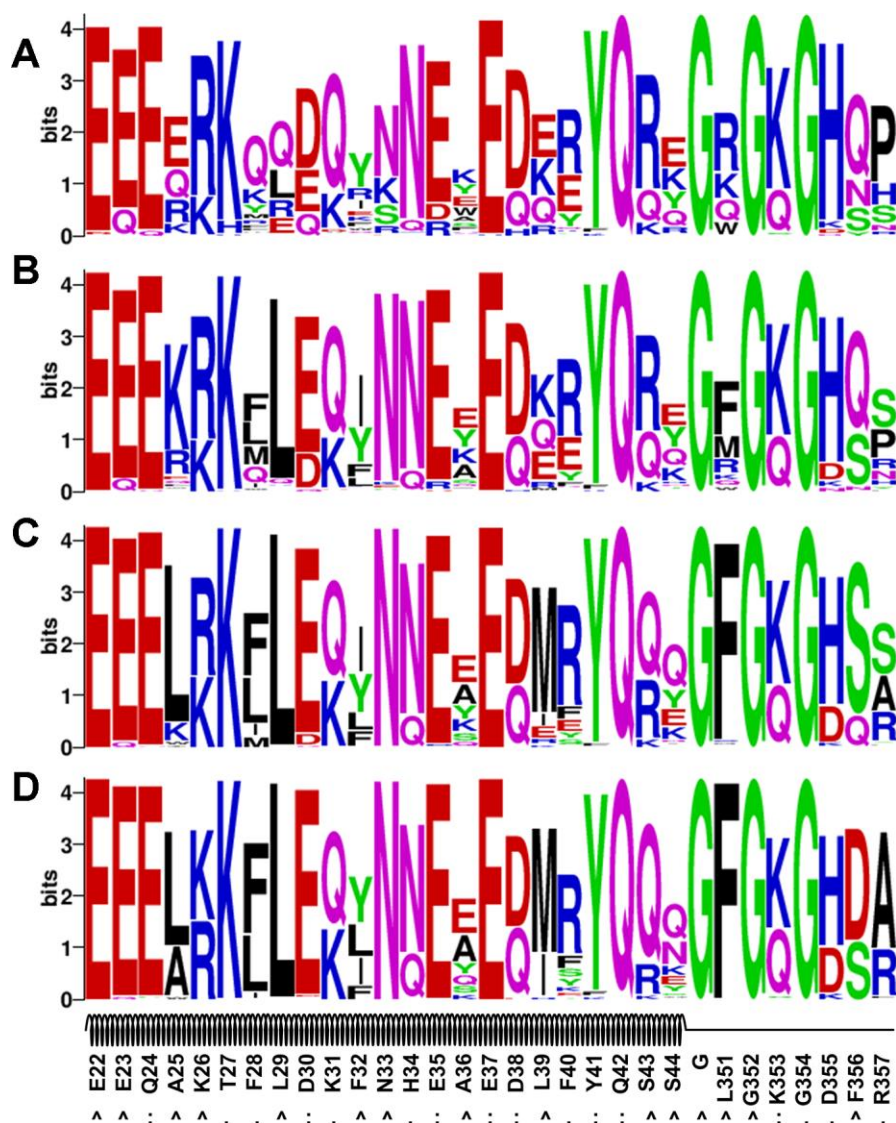
Figure 4. Sequence logo analysis of the evolution-based design results. Four sets of profile energy weight were used: 0.25 (**A**), 0.50 (**B**), 0.75 (**C**) and 1.00 (**D**).

they were somewhat similar to the wild-type peptide. Additionally, the machine-learning-based secondary structure prediction results showed that the *de novo* designed sequences should preserve the initial secondary structure topology of the peptide motif, which is important for facilitating the protein-peptide binding interaction.

In summary, we constructed a novel hybrid peptide from the interface of the natural hACE2 protein, and based on this peptide scaffold, we designed multiple novel peptide sequences with enhanced affinity toward SARS-CoV-2 RBD in computational binding experiments. Detailed analyses showed that the designed peptides were reasonable, as indicated by the recapitulation of critical binding interactions at the protein-peptide interface and the introduction of new favorable binding interactions, as well as the preservation of secondary structure to maintain the interactions. This work demonstrates the possibility of designing novel peptide therapeutics using computational algorithms. Other approaches can also be employed to engineer the hybrid peptide constructed based on the hACE2 protein, such as directed evolution [34, 35], which is widely used in the field of enzyme engineering [36–38]. Moreover, structure-based computational protein design can be combined with experiment-based approaches like directed evolution [39]. It is noteworthy that the experimental investigation of these designed peptides is of great importance for both methodology validation and drug design. We are working with our collaborators on the related experiments, which are still being conducted given that significantly more time is required for wet-lab experimental validation than a computational study. Due to the urgent situation caused by COVID-19 worldwide, we share our computational data with the community, which may help favorably combat the COVID-19 pandemic.

MATERIALS AND METHODS

Peptide design procedure

Based on the constructed protein-peptide complex structure (SARS-CoV-2 RBD/hACE2-22-44G351-357), we performed 1000 independent design trajectories individually, using (1) EvoEF2 [18], a physics- and knowledge-based energy function specifically designed for protein design and (2) a new version of EvoDesign [19], which combines EvoEF2 and evolutionary profiles for design scoring. A simulated annealing Monte Carlo (SAMC) [40] protocol was used to search for low total energy sequences as previously described [18]. For each trajectory, only the single lowest energy in that design simulation was selected, and therefore 1000 sequences

each were collected from the EvoEF2 and EvoDesign designs. The EvoEF2 and EvoDesign designs were separately analyzed to determine the impact of the physics- and profile-based scores. Since SAMC is a stochastic searching method, some of the 1000 sequences were duplicates and thus excluded from the analysis. The backbone conformations of the hACE2 peptide and SARS-CoV-2 RBD were held constant during the protein design simulations, all the residues on the peptide were redesigned, and the side-chains of the interface residues on the virus RBD were repacked without design. The non-redundant designed peptides are listed in Supplementary Tables 1–5, and the raw data and computational protein design tools are freely available at https://zhanglab.ccmb.med.umich.edu/EvoEF/.

Evolutionary profile construction

To construct reliable structural evolutionary profiles, we used the hACE2 protein structure instead of the hybrid peptide to search structural analogs against a non-redundant PDB library. Only structures with a TM-score ≥0.7 to the hACE2 scaffold were collected to build a pairwise multiple sequence alignment (MSA). A total of nine structural analogs were identified. The corresponding alignment for residues 22-44 and 351-357 were directly extracted from the full-length MSA and combined to build an MSA for the hybrid peptide. Since an arbitrary glycine was used to link positions 44 and 351, a gap '-' was inserted in the peptide MSA for the glycine position. The peptide MSA constructed in this manner is described in Supplementary Figure 2. The peptide MSA was used to construct the evolutionary profile position-specific scoring matrix (PSSM) as previously described [29].

In previous studies, we also proposed incorporating protein-protein interface evolutionary profiles to model PPIs [19, 41, 42]. However, no interface structural analogs were identified from the non-redundant interface library (NIL) [42], and no interface sequence analogs were found from the STRING [43] database with a PPI link score ≥0.8. Therefore, the interface evolutionary profile scoring was excluded from the design.

Abbreviations

ACE2: angiotensin-converting enzyme 2; COVID-19: coronavirus disease 2019; Cryo-EM: cryogenic electron microscopy; EEU: EvoEF2 energy unit; hACE2: human angiotensin-converting enzyme 2; MSA: multiple sequence alignment; NIL: non-redundant interface library; PDB: protein data bank; PPI: protein-protein interaction; PSSM: position-specific scoring matrix; RBD: receptor binding domain; RBM: receptor binding

motif; SAMC: simulated annealing Monte Carlo; SARS-CoV: severe acute respiratory syndrome coronavirus; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2.

AUTHOR CONTRIBUTIONS

X.H performed computational experiments, analyzed data, and drafted the manuscript; R.P advised and edited the manuscript; Y.Z. conceived the project, acquired funding, designed computational experiments, assumed supervision, and edited the manuscript.

ACKNOWLEDGMENTS

CONFLICTS OF INTEREST

The authors declare that no conflict interest exists.

FUNDING

REFERENCES

1. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, Mchugh TD, Memish ZA, Drosten C, Zumla A, Petersen E. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in wuhan, China. Int J Infect Dis. 2020; 91:264–66.
https://doi.org/10.1016/j.ijid.2020.01.009
PMID:31953166

2. WHO. Coronavirus disease (COVID-2019) situation reports. 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.

3. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, Lu L, Jiang S, Yang Z, Wu Y, Ying T. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. Emerg Microbes Infect. 2020; 9:382–85.
https://doi.org/10.1080/22221751.2020.1729069
PMID:32065055

4. Liu X, Wang XJ. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. J Genet Genomics. 2020; 47:119–21.
https://doi.org/10.1016/j.jgg.2020.02.001
PMID:32173287

5. Lin S, Shen R, Guo X. Molecular Modeling Evaluation of the Binding Abilities of Ritonavir and Lopinavir to Wuhan Pneumonia Coronavirus Proteases. bioRxiv. 2020.
https://doi.org/10.1101/2020.01.31.929695

6. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020; 30:269–71.
https://doi.org/10.1038/s41422-020-0282-0
PMID:32020029

7. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579:270–73.
https://doi.org/10.1038/s41586-020-2012-7
PMID:32015507

8. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu NH, Nitsche A, Müller MA, Drosten C, Pöhlmann S. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell. 2020; 181:271–80.e8.
https://doi.org/10.1016/j.cell.2020.02.052
PMID:32142651

9. Upadhyaya P, Qian Z, Selner NG, Clippinger SR, Wu Z, Briesewitz R, Pei D. Inhibition of ras signaling by blocking ras-effector interactions with cyclic peptides. Angew Chem Int Ed Engl. 2015; 54:7602–06.
https://doi.org/10.1002/anie.201502763
PMID:25950772

10. Kaspar AA, Reichert JM. Future directions for peptide therapeutics development. Drug Discov Today. 2013; 18:807–17.
https://doi.org/10.1016/j.drudis.2013.05.011
PMID:23726889

11. Fosgerau K, Hoffmann T. Peptide therapeutics: current status and future directions. Drug Discov Today. 2015; 20:122–28.
https://doi.org/10.1016/j.drudis.2014.10.003
PMID:25450771

12. Naider F, Anglister J. Peptides in the treatment of AIDS. Curr Opin Struct Biol. 2009; 19:473–82.
https://doi.org/10.1016/j.sbi.2009.07.003
PMID:19632107

13. Schneider JA, Craven TW, Kasper AC, Yun C, Haugbro M, Briggs EM, Svetlov V, Nudler E, Knaut H, Bonneau R,

Garabedian MJ, Kirshenbaum K, Logan SK. Design of peptoid-peptide macrocycles to inhibit the β-catenin TCF interaction in prostate cancer. Nat Commun. 2018; 9:4396.
https://doi.org/10.1038/s41467-018-06845-3
PMID:30352998

14. Zhang G, Pomplun S, Loftis AR, Loas A, Pentelute BL. The first-in-class peptide binder to the SARS-CoV-2 spike protein. bioRxiv. 2020.
https://doi.org/10.1101/2020.03.19.999318

15. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh CL, Abiona O, Graham BS, McLellan JS. cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science. 2020; 367:1260–63.
https://doi.org/10.1126/science.abb2507
PMID:32075877

16. Han DP, Penn-Nicholson A, Cho MW. Identification of critical determinants on ACE2 for SARS-CoV entry and development of a potent entry inhibitor. Virology. 2006; 350:15–25.
https://doi.org/10.1016/j.virol.2006.01.029
PMID:16510163

17. Procko E. The sequence of human ACE2 is suboptimal for binding the S spike protein of SARS coronavirus 2. bioRxiv. 2020.
https://doi.org/10.1101/2020.03.16.994236

18. Huang X, Pearce R, Zhang Y. EvoEF2: accurate and fast energy function for computational protein design. Bioinformatics. 2020; 36:1135–42.
https://doi.org/10.1093/bioinformatics/btz740
PMID:31588495

19. Pearce R, Huang X, Setiawan D, Zhang Y. EvoDesign: designing protein-protein binding interactions using evolutionary interface profiles in conjunction with an optimized physical energy function. J Mol Biol. 2019; 431:2467–76.
https://doi.org/10.1016/j.jmb.2019.02.028
PMID:30851277

20. Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, Wang X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature. 2020; 581:215–20.
https://doi.org/10.1038/s41586-020-2180-5
PMID:32225176

21. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, Geng Q, Auerbach A, Li F. Structural basis of receptor recognition by SARS-CoV-2. Nature. 2020; 581:221–24.
https://doi.org/10.1038/s41586-020-2179-y
PMID:32225175

22. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science. 2020; 367:1444–48.
https://doi.org/10.1126/science.abb2762
PMID:32132184

23. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, et al. The protein data bank. Acta Crystallogr D Biol Crystallogr. 2002; 58:899–907.
https://doi.org/10.1107/s0907444902003451
PMID:12037327

24. Fiser A, Sali A. Modeller: generation and refinement of homology-based protein structure models. Methods Enzymol. 2003; 374:461–91.
https://doi.org/10.1016/S0076-6879(03)74020-8
PMID:14696385

25. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci. 2006; 15:2507–24.
https://doi.org/10.1110/ps.062416606
PMID:17075131

26. Xu X, Chen P, Wang J, Feng J, Zhou H, Li X, Zhong W, Hao P. Evolution of the novel coronavirus from the ongoing wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci. 2020; 63:457–60.
https://doi.org/10.1007/s11427-020-1637-5
PMID:32009228

27. Dong N, Yang X, Ye L, Chen K, Chan EW-C, Yang M, Chen S. Genomic and protein structure modelling analysis depicts the origin and infectivity of 2019-nCoV, a new coronavirus which caused a pneumonia outbreak in Wuhan, China. bioRxiv. 2020.
https://doi.org/10.1101/2020.01.20.913368

28. Huang Q, Herrmann A. Fast assessment of human receptor-binding capability of 2019 novel coronavirus (2019-nCoV). bioRxiv. 2020.
https://doi.org/10.1101/2020.02.01.930537

29. Mitra P, Shultis D, Brender JR, Czajka J, Marsh D, Gray F, Cierpicki T, Zhang Y. An evolution-based approach to de novo protein design and case study on mycobacterium tuberculosis. PLoS Comput Biol. 2013; 9:e1003298.
https://doi.org/10.1371/journal.pcbi.1003298
PMID:24204234

30. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. Bioinformatics. 2000; 16:404–05.
https://doi.org/10.1093/bioinformatics/16.4.404
PMID:10869041

31. Yan R, Xu D, Yang J, Walker S, Zhang Y. A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction. Sci Rep. 2013; 3:2619.

https://doi.org/10.1038/srep02619
PMID:24018415

32. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004; 14:1188–90.
https://doi.org/10.1101/gr.849004
PMID:15173120

33. Shultis D, Dodge G, Zhang Y. Crystal structure of designed PX domain from cytokine-independent survival kinase and implications on evolution-based protein engineering. J Struct Biol. 2015; 191:197–206.
https://doi.org/10.1016/j.jsb.2015.06.009
PMID:26073968

34. Arnold FH. Design by directed evolution. Acc Chem Res. 1998; 31:125–131.
https://doi.org/10.1021/ar960017f

35. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. Nat Rev Mol Cell Biol. 2009; 10:866–76.
https://doi.org/10.1038/nrm2805
PMID:19935669

36. Turner NJ. Directed evolution drives the next generation of biocatalysts. Nat Chem Biol. 2009; 5:567–73.
https://doi.org/10.1038/nchembio.203
PMID:19620998

37. Kan SB, Lewis RD, Chen K, Arnold FH. Directed evolution of cytochrome C for carbon-silicon bond formation: bringing silicon to life. Science. 2016; 354:1048–51.
https://doi.org/10.1126/science.aah6219
PMID:27885032

38. Khersonsky O, Röthlisberger D, Wollacott AM, Murphy P, Dym O, Albeck S, Kiss G, Houk KN, Baker D, Tawfik DS. Optimization of the in-silico-designed kemp eliminase KE70 by computational design and directed evolution. J Mol Biol. 2011; 407:391–412.
https://doi.org/10.1016/j.jmb.2011.01.041
PMID:21277311

39. Ward TR. Artificial enzymes made to order: combination of computational design and directed evolution. Angew Chem Int Ed Engl. 2008; 47:7802–03.

https://doi.org/10.1002/anie.200802865
PMID:18780363

40. Kirkpatrick S, Gelatt CD Jr, Vecchi MP. Optimization by simulated annealing. Science. 1983; 220:671–80.
https://doi.org/10.1126/science.220.4598.671
PMID:17813860

41. Xiong P, Zhang C, Zheng W, Zhang Y. BindProfX: assessing mutation-induced binding affinity change by protein interface profiles with pseudo-counts. J Mol Biol. 2017; 429:426–34.
https://doi.org/10.1016/j.jmb.2016.11.022
PMID:27899282

42. Huang X, Zheng W, Pearce R, Zhang Y. SSIPe: accurately estimating protein-protein binding affinity change upon mutations using evolutionary profiles in combination with an optimized physical energy function. Bioinformatics. 2020; 36:2429–37.
https://doi.org/10.1093/bioinformatics/btz926
PMID:31830252

43. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ, von Mering C. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362–68.
https://doi.org/10.1093/nar/gkw937
PMID:27924014

44. Towns J, Cockerill T, Dahan M, Foster I, Gaither K, Grimshaw A, Hazlewood V, Lathrop S, Lifka D, Peterson GD. XSEDE: accelerating scientific discovery. Comput Sci Eng. 2014; 16:62–74.
https://doi.org/10.1109/MCSE.2014.80

45. Mukherjee S, Zhang Y. MM-align: a quick algorithm for aligning multiple-chain protein complex structures using iterative dynamic programming. Nucleic Acids Res. 2009; 37:e83.
https://doi.org/10.1093/nar/gkp318 PMID:19443443

46. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. Proteins. 2004; 57:702–10.
https://doi.org/10.1002/prot.20264
PMID:15476259

Supplementary Figures
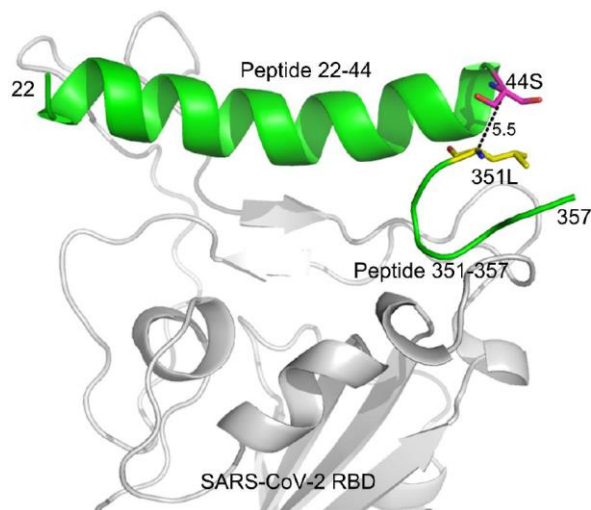
Supplementary Figure 1. Illustration of the scaffold fragments used for design. Two discontinuous peptide fragments (a.a. 22-44 and a.a. 351-357) were used for peptide scaffold construction. The hACE2 peptides and SARS-CoV-2 RBD are shown in green and gray, respectively. The residues Ser44 and Leu351 are shown in magenta and yellow sticks, respectively. The distance between the Cα atoms of Ser44 and Leu351 is 5.5 Å.

Supplementary Figure 2. Peptide multiple sequence alignment for evolutionary profile construction.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1 to 5.

Supplementary Table 1. Summary of 992 peptide sequences designed using EvoEF2 only.

Supplementary Table 2. Summary of 991 peptide sequences designed using EvoEF2 and the evolutionary profile (weight = 0.25).

Supplementary Table 3. Summary of 966 peptide sequences designed using EvoEF2 and the evolutionary profile (weight = 0.50).

Supplementary Table 4. Summary of 877 peptide sequences designed using EvoEF2 and the evolutionary profile (weight = 0.75).

Supplementary Table 5. Summary of 695 peptide sequences designed using EvoEF2 and the evolutionary profile (weight = 1.00).