

# Mining peptides against SARS-CoV-2 entry from constructed RBMs-hACE2 isomer libraries using machine learning and molecular docking

Xiaotong He (✉ [xiaotong.he@manchester.ac.uk](mailto:xiaotong.he@manchester.ac.uk))

The University of Manchester

---

## Method Article

**Keywords:** SARS-CoV-2, COVID19, hACE2, Molecular docking, Machine learning

**Posted Date:** December 18th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-130055/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# **Mining peptides against SARS-CoV-2 entry from constructed RBMs-hACE2 isomer libraries using machine learning and molecular docking**

Xiaotong He

Division of Cancer Sciences, The University of Manchester, Manchester, UK

xiaotong.he@manchester.ac.uk

Keywords: SARS-CoV-2, COVID19, hACE2, Molecular docking, Machine learning

## **Abstract**

Cellular entry of SARS-CoV-2 initiates from the protein-protein interactions (PPIs) between viral surface protein S and human angiotensin converting enzyme 2 (hACE2). Peptide-based drugs have the advantage of small molecule compounds to block such viral-host PPIs. Thus the viral target-regions on hACE2 have been believed as promising templates for designing specific inhibitory peptides against SARS-CoV-2 infection. However, starting from a few potential templates, *in silico* design and prediction between binding affinity and bioactivities *in vivo* are very challenging, herein a novel design strategy was implemented by mining constructed template isomer libraries using feature filters, supervised classifier and peptide protein docking.

Applying these methods and the isomer libraries, 4 peptides were identified from 12 millions candidates owing to their distinct stability, interaction activity, inhibitory specificity, binding affinity, transmembrane potentials and effective conformation. These results have supplied a panel of specific anti-COVID19 leads for further drug development, supporting a new feasible antiviral strategy for targeting both intracellular and extracellular SARS-CoV-2 S proteins simultaneously. The methods have provided a useful tool for mining antiviral-peptides against viral diseases.

## Introduction

The viral disease COVID-19 has severely threatened global health and life with high morbidity and mortality because there are currently no approved drugs or vaccines that specifically target severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)<sup>1-4</sup>. Relying on its envelope glycoprotein spike (S), this single stranded positive-sense RNA virus is able to rapidly locate on and effectively attach to host cell surface, and consequently initiates cellular entry and triggers organ injury<sup>5,6</sup>. At molecular level, the multidimensional process of the virus invasion has been proven as protein-protein interactions (PPIs) between the viral protein S and human cellular angiotensin converting enzyme 2 (hACE2)<sup>7-10</sup>, and indeed such essential ligand-receptor binding has readily gained broad attention as critical targets for blocking the virus infection<sup>11-14</sup>.

In terms of the viral ligand, each monomer of the trimeric S protein contains approximately 1253 amino acids and is proteolytically processed into S1 and S2 domains, of which 194 amino acids in the S1 constitute the receptor binding motif (RBM, S438-Q506), where the S1 protein is functionally bound to the hACE2. Recent crystallographic analysis has detailed that the spatial interface mainly covers 17 residues of K417, G446, Y449, Y453, L455, F456, A475, F486, N487, Y489, Q493, G496, Q498, T500, N501, G502 and Y505<sup>7-9</sup>. On the receptor side, hACE2, a type I transmembrane protein comprises of an extracellular heavily N-glycosylated N-terminal domain containing the carboxypeptidase site and a short intracellular C-terminal cytoplasmic tail<sup>10,15,16</sup>. Corresponding to the RBM in S1, 20 residues in the N-terminal domain of ACE2 join to contact with above 17 residues, including Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L79, M82, Y83, N330, K353, G354, D355, R357 and R393<sup>7-9</sup>. These experimental findings plus atomic-level predictions indicate that SARS-CoV-2 exhibits a stronger affinity to hACE2 compared to SARS-CoV, which may contribute to its high contagiousity in humans and evidently enhances both clinical and epidemical significances in developing treatments to prevent the S-hACE2 binding<sup>8,9</sup>.

To date, many vaccines, repurposing FDA approved drugs and various monoclonal antibodies have been tested and relevant trials have been conducted<sup>3,17-23</sup>, whereas the prophylactic vaccines of ongoing trials do not play role in therapeutic intervention to the patients with SARS-CoV-2 infection<sup>18,19</sup> and there are no independent clinical reports so far to prove the efficacy of drugs for the antiviral therapy<sup>3,17,22,23</sup>, thus specific targeted treatments are urgently needed to combat the pandemic. Recently, several experimental and computational studies have been involved in the development of peptides as hACE2 blockaders or SARS-CoV-2 S inhibitors based on the primary sequence of S-ACE2 binding domains<sup>11-14,16,24</sup>. Particularly SARS-CoV-2 S RBM-primed regions on hACE2 (RBM-hACE2) have been believed as specific templates for in silico design of the potential inhibitors against CoV-2 infection<sup>11-14</sup>.

Clinically, peptide derived drugs, for instance insulin, ciclosporin and many others, have represented a new class of therapeutics for different type of diseases<sup>25-27</sup>. The molecular mechanisms of action in peptide drugs have received more attention recently for their antiviral potentials<sup>28,29</sup>. Peptide drugs have the advantage of small molecule compounds to block PPIs, especially in different kinds of viral-host PPIs, because peptides are capable of binding to the larger grooves or clefts on an interface, which are hard to target with traditional small molecule drugs due to large size, simple surface or shallow pocket<sup>28-30</sup>. Extensive studies demonstrated peptides effectively blocked virus attachment or the entry of a virus into host cells<sup>31-33</sup>. These efforts along with recent studies have offered a great hope for discovery of the peptide-based prophylactic and therapeutic agents against COVID-19<sup>11-14,31,32</sup>, however, starting from a few potential templates, in silico peptide design and properties prediction across binding affinity, specificity, stability and membrane permeability are still very challenging. Herein a novel design strategy was introduced by mining constructed RBMs-hACE2 isomer libraries using feature filters, supervised classifier and peptide-protein docking<sup>34-36</sup>.

Applying these methods and the isomer libraries, 4 peptides were identified from 12 millions of candidates owing to their distinct bioactivities and inhibitory potentials against SARS-CoV-2 entry. These results have supplied a panel of specific anti COVID19 leads for further drug development, which promisingly support a new feasible antiviral strategy for targeting both intracellular and extracellular SARS-CoV 2 S proteins simultaneously. The pipelined-protocols have provided a useful tool for mining antiviral-peptides against other viral diseases.

## **Results**

### **Development and application of a traversal method for collecting RBMs-hACE2 homolog templates**

As shown in Figure 1, three RBMs-hACE2 segments, covering 20 specific binding sites, were selected as primary templates for designing the inhibitors due to the corresponding druggable sites on the 3D S. With desired linkage at N and C terminal, the 3 templates were extracted as: 24QAKTFLDKFNHEAEDLFYQ42, 72FLKEQSTLAQMYPLQ86 and 347TAWDLGKGDFRIL-M360. Using a traversal method, all homolog residues within these 3 extracts ranging in length from 5 to 19 amino acids were collected. As indicated in Figure 2-3, a total of 120, 66 and 55 homolog templates were generated from the RBMs-hACE2 of Q24-Q42, F72-Q86 and T347-M360 respectively. The sequences of these homolog templates were detailed in Table S1, Supplementary data. Obviously the possibility for mining satisfactory inhibitors was increased 80 times since the number of templates was enlarged from 3 to 241.

### **Development and application of a feature filter for selecting potential inhibitor against SARS-CoV-2 S protein**

For screening the stability and interaction activity of the homolog templates, a feature filter was set up based on computing the values of two physico-chemical properties, Boman and Instaindex<sup>37,38</sup>.

Through this filter, a total of 12 homolog templates were selected from above 241 peptide segments, including 5 from Q24-Q42 and 7 from D347-M360 (Figure 4). All peptide homologs from Q72-Q86 were excluded for downstream analysis due to low stability and interaction activities (Figure 3-4 and Table S1 Supplementary data). In parallel, structural datasets of these homologs were generated from the extracts of hACE2 chain E (Figure 4). These 12 filtered-templates possessed high stabilities and interaction activities.

### **Development and application of a permutation program for constructing RBMs-hACE derived isomer libraries**

In order to retain the binding specificity and minimize undesired bioactivities from the wild type template, an isomer library of each template generated using a permutation method. One million isomers were sampled from billions of isomers randomly, this approach provided a large number of candidates for further optimisation. As summarized in Figure 3, a total of  $10^6$  inhibitory candidates were settled by the computational sieving again, and hence promisingly possessed not only high stability and interaction activity but also low potentials of unwanted bioactivities.

### **Application of a machine learning tool for predicting transmembrane potentials of the inhibitors**

Doubtlessly in vivo distribution, a key element of the well-known ADMET, has critically affected clinical application of the peptide drugs and has strongly challenged all steps in the development of effective peptides<sup>39</sup>. To address this issue, transmembrane potentials were predicted among the selected peptide isomers by SVM modeling. As shown in Figure 3 and Table S2 Supplementary data, 618 peptides, labelled with either extracellular or transmembrane potentials, were selected from  $10^5$  isomers of 12 wild type homologs for further molecular docking assay. This classification provided fundamental support to carry out a new antiviral strategy by using a panel of

the peptides for targeting both intracellular and extracellular SARS-CoV-2 S proteins simultaneously.

### **Application of a peptide-protein docking approach for validating binding affinity and effective conformation between the inhibitors and SARS-CoV-2 S**

The inhibitory potentials of the 618 selected peptides were finally validated by *in silico* docking each of them with the known binding sites on SARS-CoV-2 S protein (Figure 1). Since these peptide sequences were originally derived from 12 different homolog templates (Figure 4D-E), each templates tertiary structure was extracted as a reference ligand for the relevant docking. Using the reference ligands and the docking receptor SARS-CoV-2 S chain E (Figure 4B, 4D-E), individual affinity maps were computed by the ARDF model in AutoDock CrankPep. As shown in Figure 5 and in Table S2 Supplementary data, following 618 docking assays (all dots), 4 peptide ligands (red dots) were finally determined as the most promising leads against SARS-CoV-2 S because the mean values of all atoms Root-Mean-Square Deviation (RMSD) and binding affinity were less than 2.5 Å and -10 kcal/mol respectively among all the first 10 docking ranks<sup>36,40</sup>. These outputs were the first demonstration of the effective binding activities between RBMs-hACE2 isomers and SARS-CoV-2 S and suggested that these 4 peptide isomers retained inhibitory potentials against SARS-CoV-2 infection.

### **Interface analysis of the docking complexes**

Based on the spatial extract of SARS-CoV-2 S (Figure 4B) and docking conformation of the isomer ligand, the interface contacts within each docking complex were produced by a Python program using the Pymol package. In order to understand the significant docking outputs mostly and facilitate the comparisons with previous studies, data was generated from the contacts within 3.5 (Figure 6 and Table 1) and 3.8 angstroms separately (Table S3 Supplementary data). As shown in

Figure 6, the total contacts exhibited higher density and more evenly distributed between L4 and SARS-CoV-2 S RBM compared with that in wild type W2. Similar differences were discovered in each pairwise comparison between W1 and L1, L2, L3. Specifically, on the SARS-CoV-2 S sides, all contacted residues belonged to the known binding sites (Table 1 and Table S3 Supplementary data).

In line with previous findings<sup>8</sup>, conventional hydrogen bonds were found between the inhibitory isomers and SARS-CoV-2 S. These N-O hydrogen bonds included D1(N)-N487(OD1), L5(N)-Y489(OH) in Lead 1 binding, H5(N)-Y489(OH) in Lead 3 binding and Q1(O)-Y489(OH), T5(N)-Y489(OH), T5(OG1)-Y489(OH) in Lead 4 binding. Moreover, many other types of interactions were involved into the contacts as well, for instance, salt bridges E7(CG)-K417(CE) and E7(CG)-K417(NZ) in L1 and L2, E7(CG)-K417(NZ) in L3, and K8(CG)-E484(OE1) in L4 (Table 1 and Table S3 Supplementary data). Notably, these observations provided the evidence that all networks among the contacts, including all strengths, contributed to the distinct RMSD and binding affinity of these 4 isomer leads.

### **Retrospective comparison of the isomer inhibitor and the corresponding wild type template**

The structure of each selected isomer was folded and generated from sequential input by Autodock modeling. As displayed in Figure 7A, each helical backbone of the isomer is similar to the structure of wild type extract, and it is indicated as reliable right-handed alpha helix by further Ramachandran plots<sup>41</sup> since 100% of residues were outlined in the blue areas at -180~0 degrees and 98.7% (1/72) of them were located within the dark blue islands (Figure 7B). Indeed these short helical structures were well retained in all isomers according to the primary compositions of wild types despite the random permutation process. Typical hydrophobic face is not seen in Lead 1, 3 and 4 though it seems to be in W1, W2 and L2 (Figure 7C). As emphasized before, the 4 distinct

isomers were determined and selected by the experiential thresholds of RMSD and binding affinity<sup>36,40</sup> though there were no statistical differences of RMSD and binding affinity among the isomers and wild types (Figure 8). In line with early studies<sup>42</sup>, various physicochemical properties were found in the millions of isomers (data not shown). In particular, the stability of isomers L1-L4 was increased 30-50 times compared to the wild types. Isomers with or without transmembrane potentials were obtained from the constructed libraries (Table 2), which were failed to select from wild types due to the limited sample size (data not shown).

## Discussion

Preventing the binding between SARS-CoV-2 S and hACE2 is an effective action to fight COVID-19, however, to fit this, there are neither new approved compounds nor repurposing existing drugs from large number of basic studies and clinical trials<sup>3,4,20,21,23</sup>. This insufficiency might suffer from an essential difficulty that small molecule compounds were not effective at blocking PPIs where a deep binding pocket might be missing at the interface<sup>11,25,43</sup>. While other studies brought advantages of peptide-based inhibitor for preventing or interfering the viral host PPIs<sup>31-33</sup>, which has been further extended by recent researches of peptide-based inhibitors against SARS-CoV 2 entry<sup>11-14</sup>. However, along with the structure to binding affinity, *in silico* peptide design across bioactivity, stability and distribution profile *in vivo* are critical elements for the development of effective inhibitors<sup>44,45</sup>. To address these challenges, a novel design strategy was carried out in this study.

In line with the attempts of previous studies<sup>11-14</sup>, three sequential peptide segments were initially extracted from the significant druggable regions of the hACE2 interface, including Q24-Q42, F72-Q86 and D347-M360 (Figure 1). These 3 wild type residues were promising primary templates for starting the design because of the potential binding affinity and specificity<sup>8,11-14</sup>. Furthermore, necessary optimizations were made to tackle three major difficulties in the template-based design: (i)

Properties of the wild type origins could cause undesirable effects to host or be affected by endogenous ACE2 substrates since both membrane bound and circulating form of hACE2 play important enzymatic roles in vivo<sup>15,44,45</sup> (ii) A panel of antiviral peptides with and without membrane permeability are indispensable for effectively targeting the viral protein S because functional exposures of this surface protein are involved into both intracellular and extracellular distribution through whole viral life cycle<sup>39,46,47</sup>. (iii) The small sample size of templates could limit the selection of satisfactory interaction activity, binding affinity, membrane permeability and stability.

Logically all peptide homolog segments within these 3 wild type templates may retain specific binding affinity since the full-length templates possess the known binding sites<sup>8,9</sup>, thus a total of 241 homologous sequences ranging in length from 5 to 19 amino acids were collected and pooled for the selection by feature filter. In theory, position change of amino acids in peptide isomers strongly modulated the relative stabilities of topologically similar regions in the energy landscape, rather than redefined the topology space<sup>48</sup>. Experimentally, certain isomeric versions of antimicrobial peptide significantly increased in vitro activity against bacterial pathogens<sup>49</sup>. These demonstrations raised a hypothesis that the isomer pools established from wild type templates could reserve a certain number of candidates retaining desired target specificity, binding affinity, distribution properties and stability. This was finally verified by mining constructed isomer libraries using a SVM classifier and peptide docking<sup>34-36</sup>(Figure 3 and 5).

As indicated in Figure 2 and 3, based on the 3D structures of RBMs-hACE2, millions of the homolog isomers were generated by computational traversal and permutation methods. These candidates were screened through a feature filter, a Python program using modlamp and AutoDock CrankPep<sup>35,36,40</sup>, 4 peptide leads with potential inhibitory activity against SARS CoV-2 S were identified by their distinct binding activity, stability, binding affinity, effective conformation and

transmembrane potentials (Table 1-2, Figure 3 and 5). Interface analysis unveiled that these peptide leads exhibited specific binding activities against SARS-CoV-2 S, encompassing whole RBM regions through K417, Y453, L455, F456, A475, N487, Y489, Q493, S494 to Y505. It was also found that the molecular-atomic actions of these bindings were involved the formation of N O hydrogen bonds, salt bridges and many other bonds. Most significantly, these salt bridges were all linked to the unique residue K417 in SARS-CoV-2 S, which has been thought to contribute to the high contagiousity of SARS-CoV-2 in humans compared to SARS-CoV<sup>8,9</sup>, therefore, the salt bridge associated bindings between the druggable target and the inhibitory isomers might enhance the blockage of viral entry. In addition, these findings correlate with the early studies<sup>50</sup> and thus support that either strong or weak bonds can make an important contribution to the association and stability of protein complexes<sup>50,51</sup>.

In sum, this work has provided specific and effective leads as therapeutic agents against COVID-19 before the synthesis process and wet experiments. The peptides with different transmembrane potentials promisingly support a combination antiviral therapy for targeting both intracellular and extracellular SARS-CoV-2 S. The computational design has combined multiple approaches, exploring a new way to identify effective and specific antiviral peptides by mining viral-target isomers.

## **Methods**

### **Extraction of sequential and spatial RBMs-hACE2**

Initially, an X-ray crystal structure data of SARS-CoV-2 RBD / hACE2 complex was retrieved from RCSB Protein Data Bank (PDB, <https://www.rcsb.org/structure/6MOJ>)<sup>8</sup>. Three segments of RBMs-hACE2, Q24-Q42 / F72-Q86 / T347-M360, were determined and selected as primary templates for the peptide design according to a literature search and the definition of sequential and spatial motifs

<sup>8,52</sup> (Figure 1-3). All sequential residues from three original RBMs-hACE2 templates were denoted as wild-type segments in this study. Both primary sequence and tertiary structure of each wild type, SARS-CoV-2 S chain E and hACE2 chain A, were extracted respectively by a Python program using libraries such as PyMol, Biopython and Pandas<sup>53-55</sup>

### **Search and collection of RBMs-hACE2 homolog by traversal method**

All homolog segments from RBMS-hACE2 were mined to maximise the chances of finding potential bioactivity and binding affinity. Another Python program using a custom traversal method was developed. Using this program, fragments within the sequence RBMS-hACE2 ranging in length from 5 to 19 amino acids were filtered and collected as primary templates (Figure 2 and 3).

### **Construction of RBMs-hACE2 homolog isomer libraries by permutation method**

In order to use the selected templates to generate isomers with different physico-chemical properties, RBMs-hACE2 homolog isomer libraries were built using a permutation method. Using a random permutation algorithm, millions of peptide segments were generated by reordering amino acid positions of the RBMs-hACE2 homolog. Datasets were specifically sampled according to the wild type template origins (Figure 2 and 3).

### **Selection of inhibitory candidates against SARS-CoV-2 S via a feature filter**

A custom feature filter was written using the R package Peptides. With this feature filter, each peptide sequence from the isomer library was mapped to numerical data which incorporated two physico-chemical properties, Boman and Instaindex. The candidate was selected for downstream analysis by having a Boman value of over 2.0 and an Instaindex value less than 0.0 or 40.0<sup>37,38,56</sup>.

### **Prediction of transmembrane potentials for the inhibitory candidates using SVM classifier**

In order to obtain the inhibitory candidates with and without membrane permeability, transmembrane potentials of each peptide isomer were predicted by a supervised machine learning tool. A support vector machine was written in Python and optimised using the Modlamp package.<sup>34,35</sup> Peptides with or without transmembrane potentials were classified by inputting the selected peptide sequences with the built-in training dataset TM-AMP provided by the Modlamp package<sup>35</sup>. The threshold of possibility was defined as 0.99.

### **Validation of the inhibitory specificity and affinity by peptide-protein docking**

According to the known binding sites, the selected peptides were validated by peptide-protein docking using Autodock CrankPep 1.1<sup>36</sup>. Each docking assay was performed between the receptor SARS-CoV-2 S and the ligand peptide using a corresponding reference grid box. Each docking grid box (the affinity map) was constructed using the S receptor and an individual reference ligand. The reference ligand was determined by the wild type template and extracted from the structure of hACE2 chain A<sup>8</sup>. Missing side chain atoms in both receptor and ligand were completed using the Reduce module in the AutoDock CrankPep<sup>36,57</sup>. Large batch docking assays were achieved by Bash scripts. Large datasets of docking outputs were processed by a Python program, which analyzed the docking interface using PyMol-API.

### **Statistical analysis and data visualization**

Statistical comparison of significance was performed by a pairwise permutation test implemented in R with the coin package<sup>58</sup>. A p-value of less than 0.05 was considered statistically significant. Data figures were plotted generally using Matplotlib in Python<sup>59</sup>. The Wenxiang diagram was utilised to visualize alpha helical sequences by in R with Helixvis<sup>60</sup>. The Ramachandran diagram was plotted in Python using MDanalysis<sup>41</sup>.

## Reference

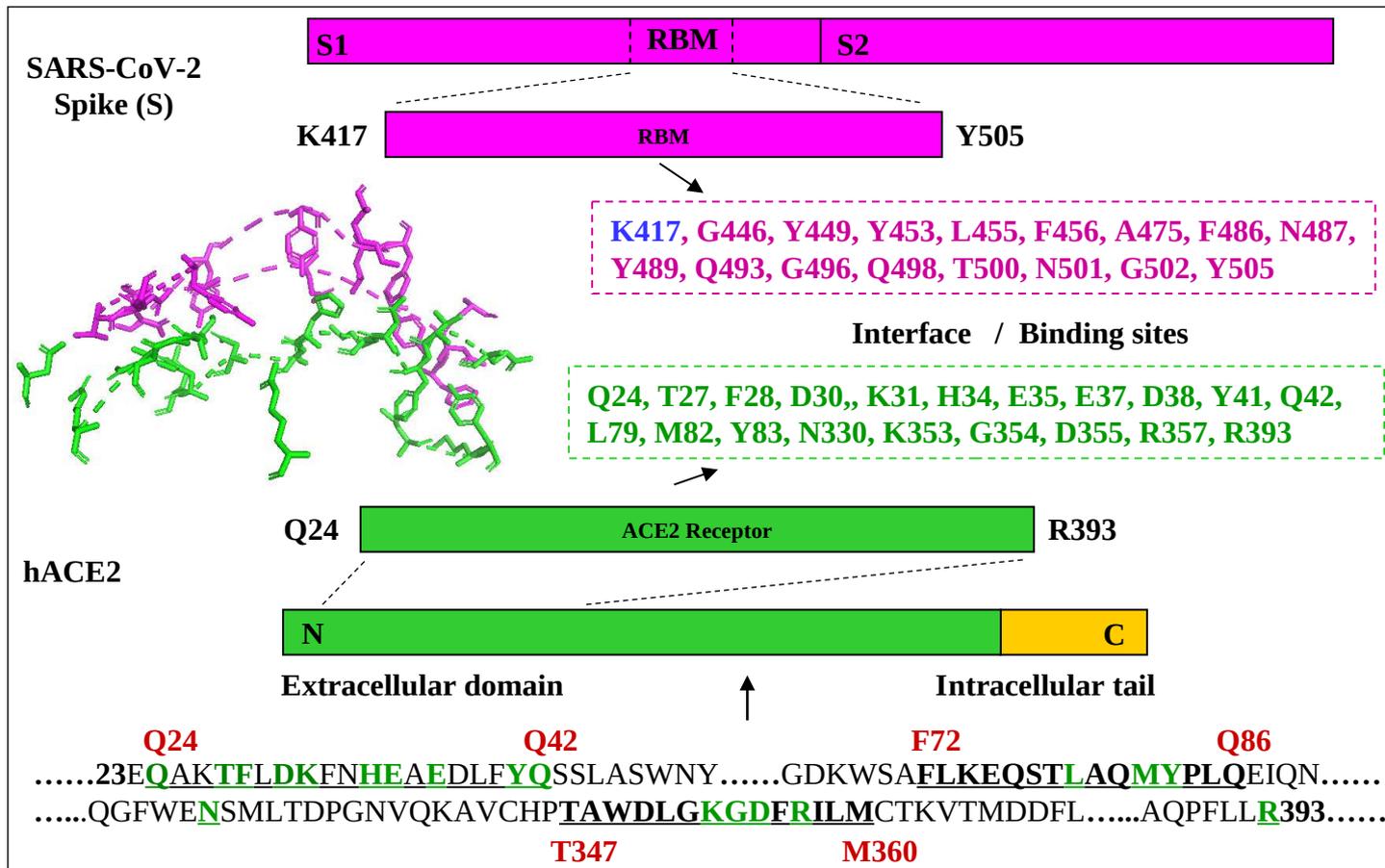
1. Ferretti, L. *et al.* Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936 (2020).
2. Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect. Dis.* **20**, e238–e244 (2020).
3. The RECOVERY Collaborative Group. Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19. *N. Engl. J. Med.* NEJMoa2022926 (2020) doi:10.1056/NEJMoa2022926.
4. Horby, P. W. *et al.* Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *The Lancet* **396**, 1345–1352 (2020).
5. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565–574 (2020).
6. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).
7. Yan, R. *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **367**, 1444–1448 (2020).
8. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
9. Wang, Q. *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904.e9 (2020).
10. Zhao, P. *et al.* Virus-Receptor Interactions of Glycosylated Sars-Cov-2 Spike and Human Ace2 Receptor. *SSRN Electron. J.* (2020) doi:10.2139/ssrn.3641894.
11. Zhang, G. *et al.* Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD. <http://biorxiv.org/lookup/doi/10.1101/2020.03.19.999318> (2020) doi:10.1101/2020.03.19.999318.
12. Huang, X., Pearce, R. & Zhang, Y. De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. *Aging* **12**, 11263–11276 (2020).
13. Han, Y. & Král, P. Computational Design of ACE2-Based Peptide Inhibitors of SARS-CoV-2. *ACS Nano* **14**, 5143–5147 (2020).

14. Baig, M. S., Alagumuthu, M., Rajpoot, S. & Saqib, U. Identification of a Potential Peptide Inhibitor of SARS-CoV-2 Targeting its Entry into the Host Cells. *Drugs RD* **20**, 161–169 (2020).
15. Xiao, L., Sakagami, H. & Miwa, N. ACE2: The key Molecule for Understanding the Pathophysiology of Severe and Critical Conditions of COVID-19: Demon or Angel? *Viruses* **12**, 491 (2020).
16. Hoffmann, M. *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.e8 (2020).
17. Horby, P. W. *et al.* Lopinavir–ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *The Lancet* **396**, 1345–1352 (2020).
18. Jackson, L. A. *et al.* An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N. Engl. J. Med.* NEJMoa2022483 (2020) doi:10.1056/NEJMoa2022483.
19. Tregoning, J. S. *et al.* Vaccines for COVID 19. *Clin. Exp. Immunol.* **202**, 162–192 (2020).
20. Jiang, Y.-P., Zhao, X.-X., Lv, H.-Q. & Wen, C.-P. Drug screening and development from the affinity of S protein of new coronavirus with ACE2. *Eur. J. Clin. Microbiol. Infect. Dis.* (2020) doi:10.1007/s10096-020-04048-7.
21. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* NEJMoa2007764 (2020) doi:10.1056/NEJMoa2007764.
22. Brouwer, P. J. M. *et al.* Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Science* **369**, 643–650 (2020).
23. Hung, I. F.-N. *et al.* Triple combination of interferon beta-1b, lopinavir–ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial. *The Lancet* **395**, 1695–1704 (2020).
24. Souza, P. F. N., Lopes, F. E. S., Amaral, J. L., Freitas, C. D. T. & Oliveira, J. T. A. A molecular docking study revealed that synthetic peptides induced conformational changes in the structure of SARS-CoV-2 spike glycoprotein, disrupting the interaction with human ACE2 receptor. *Int. J. Biol. Macromol.* **164**, 66–76 (2020).
25. Bruzzoni-Giovanelli, H. *et al.* Interfering peptides targeting protein–protein interactions: the next generation of drugs? *Drug Discov. Today* **23**, 272–285 (2018).

26. Hamada, H. *et al.* Efficacy of primary treatment with immunoglobulin plus ciclosporin for prevention of coronary artery abnormalities in patients with Kawasaki disease predicted to be at increased risk of non-response to intravenous immunoglobulin (KAICA): a randomised controlled, open-label, blinded-endpoints, phase 3 trial. *The Lancet* **393**, 1128–1137 (2019).
27. Rosenstock, J. *et al.* Once-Weekly Insulin for Type 2 Diabetes without Previous Insulin Treatment. *N. Engl. J. Med.* NEJMoa2022474 (2020) doi:10.1056/NEJMoa2022474.
28. Sorolla, A. *et al.* Precision medicine by designer interference peptides: applications in oncology and molecular therapeutics. *Oncogene* **39**, 1167–1184 (2020).
29. Nyanguile, O. Peptide Antiviral Strategies as an Alternative to Treat Lower Respiratory Viral Infections. *Front. Immunol.* **10**, 1366 (2019).
30. Skalickova, S. *et al.* Perspective of Use of Antiviral Peptides against Influenza Virus. *Viruses* **7**, 5428–5442 (2015).
31. Bai, F. *et al.* Antiviral Peptides Targeting the West Nile Virus Envelope Protein. *J. Virol.* **81**, 2047–2055 (2007).
32. Zhao, H. *et al.* A novel peptide with potent and broad-spectrum antiviral activities against multiple respiratory viruses. *Sci. Rep.* **6**, 22008 (2016).
33. Kadam, R. U. *et al.* Potent peptidic fusion inhibitors of influenza virus. *Science* **358**, 496–502 (2017).
34. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *ArXiv12010490 Cs* (2018).
35. Müller, A. T., Gabernet, G., Hiss, J. A. & Schneider, G. modLAMP: Python for antimicrobial peptides. *Bioinformatics* **33**, 2753–2755 (2017).
36. Zhang, Y. & Sanner, M. F. AutoDock CrankPep: combining folding and docking to predict protein–peptide complexes. *Bioinformatics* **35**, 5121–5127 (2019).
37. Guruprasad, K., Reddy, B. V. B. & Pandit, M. W. Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Protein Eng. Des. Sel.* **4**, 155–161 (1990).
38. Boman, H. G. Antibacterial peptides: basic facts and emerging concepts. *J. Intern. Med.* **254**, 197–215 (2003).

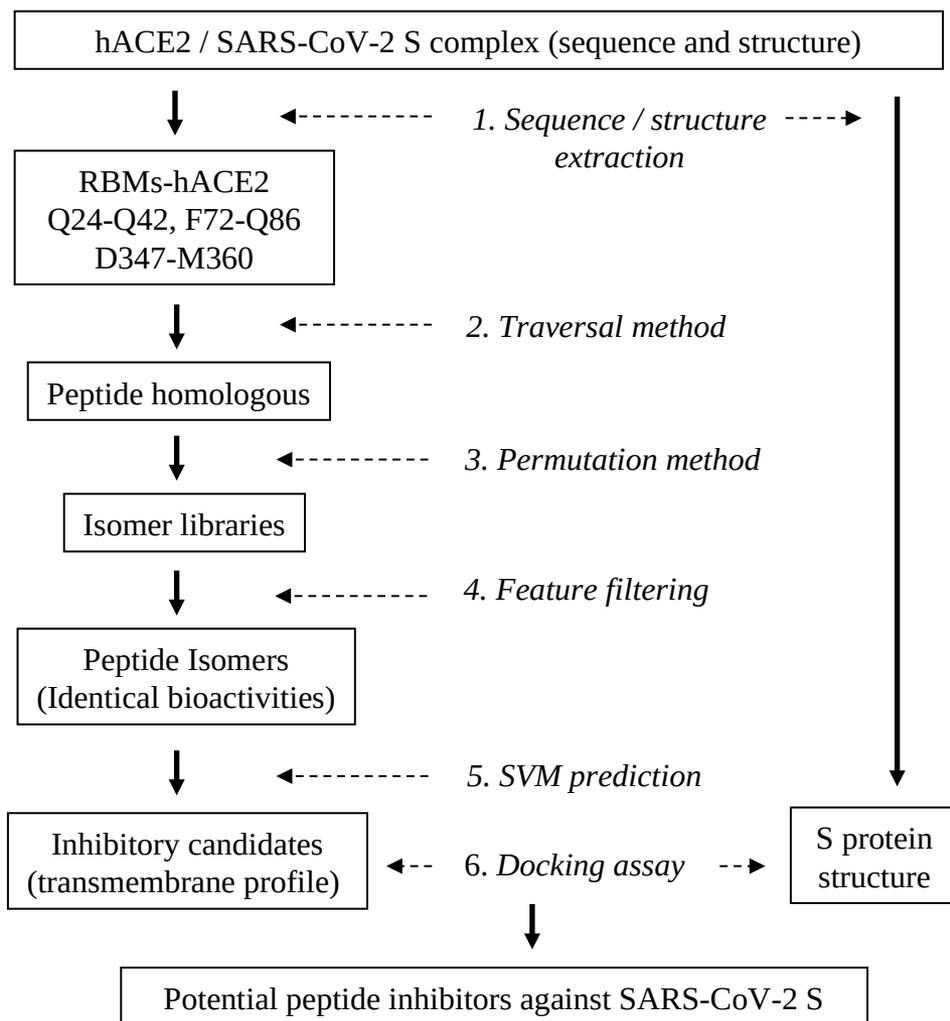
39. Santos, G. B., Ganesan, A. & Emery, F. S. Oral Administration of Peptide-Based Drugs: Beyond Lipinski's Rule. *ChemMedChem* **11**, 2245–2251 (2016).
40. Weng, G. *et al.* Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes. *J. Chem. Theory Comput.* **16**, 3959–3969 (2020).
41. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
42. Goodman, M. Peptide homologs, isosteres, and isomers: A general approach to structure-activity relationships. *Biopolymers* **24**, 137–155 (1985).
43. Upadhyaya, P. *et al.* Inhibition of Ras Signaling by Blocking Ras-Effector Interactions with Cyclic Peptides. *Angew. Chem. Int. Ed.* **54**, 7602–7606 (2015).
44. Santos, G. B., Ganesan, A. & Emery, F. S. Oral Administration of Peptide-Based Drugs: Beyond Lipinski's Rule. *ChemMedChem* **11**, 2245–2251 (2016).
45. Vilas Boas, L. C. P., Campos, M. L., Berlanda, R. L. A., de Carvalho Neves, N. & Franco, O. L. Antiviral peptides as promising therapeutic drugs. *Cell. Mol. Life Sci.* **76**, 3525–3542 (2019).
46. Lew, R. A. *et al.* Angiotensin-converting enzyme 2 catalytic activity in human plasma is masked by an endogenous inhibitor: Endogenous inhibitor masks plasma ACE2 activity. *Exp. Physiol.* **93**, 685–693 (2008).
47. Payandeh, Z. *et al.* Design of an engineered ACE2 as a novel therapeutics against COVID-19. *J. Theor. Biol.* **505**, 110425 (2020).
48. Ma, B. & Nussinov, R. Energy landscape and dynamics of the  $\beta$ -hairpin G peptide and its isomers: Topology and sequences. *Protein Sci.* **12**, 1882–1893 (2003).
49. Falciani, C. *et al.* Isomerization of an Antimicrobial Peptide Broadens Antimicrobial Spectrum to Gram-Positive Bacterial Pathogens. *PLoS ONE* **7**, e46259 (2012).
50. Jiang, L. & Lai, L. CH $\cdots$ O Hydrogen Bonds at Protein-Protein Interfaces. *J. Biol. Chem.* **277**, 37732–37740 (2002).
51. Itoh, Y. *et al.* N $^+$ -C-H $\cdots$ O Hydrogen bonds in protein-ligand complexes. *Sci. Rep.* **9**, 767 (2019).

52. Saidi, R., Dhifli, W., Maddouri, M. & Mephu Nguifo, E. Efficiently Mining Recurrent Substructures from Protein Three-Dimensional Structure Graphs. *J. Comput. Biol.* **26**, 561–571 (2019).
53. Schrödinger, LLC. The PyMOL Molecular Graphics System, Version 1.8. (2015).
54. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
55. McKinney, W. Data Structures for Statistical Computing in Python. in 56–61 (2010). doi:10.25080/Majora-92bf1922-00a.
56. Osorio, D., Rondón-Villarreal, P. & Torres, R. Peptides: A Package for Data Mining of Antimicrobial Peptides. *R J.* **7**, 4 (2015).
57. Word, J. M., Lovell, S. C., Richardson, J. S. & Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation 1 1 Edited by J. Thornton. *J. Mol. Biol.* **285**, 1735–1747 (1999).
58. Hothorn, T., Hornik, K., Wiel, M. A. van de & Zeileis, A. Implementing a Class of Permutation Tests: The **coin** Package. *J. Stat. Softw.* **28**, (2008).
59. Barrett, P., Hunter, J., Miller, J. T., Hsu, J.-C. & Greenfield, P. matplotlib – A Portable Python Plotting Package. in *Astronomical Data Analysis Software and Systems XIV* (eds. Shopbell, P., Britton, M. & Ebert, R.) vol. 347 91 (2005).
60. Wadhwa, R., Subramanian, V. & Stevens-Truss, R. Visualizing alpha-helical peptides in R with helixvis. *J. Open Source Softw.* **3**, 1008 (2018).



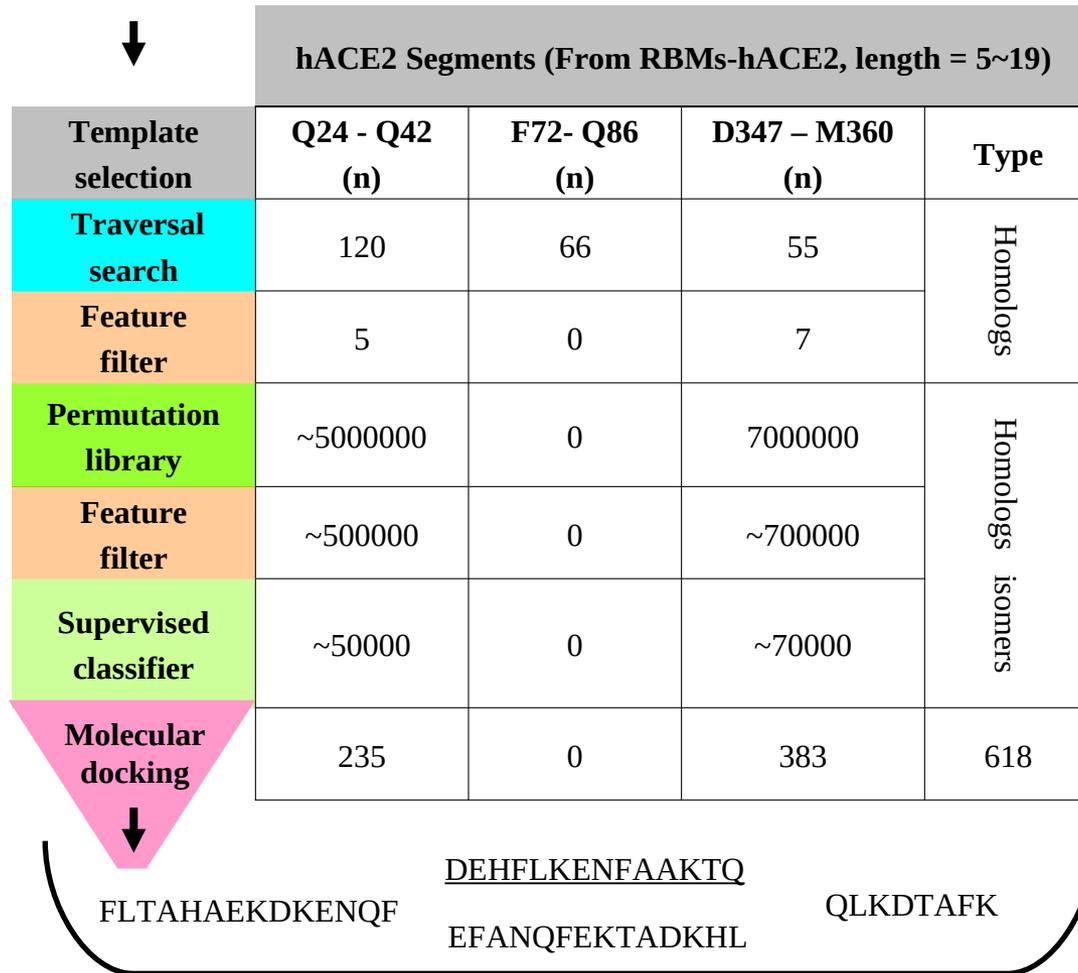
**Figure 1. Selection of primary template for designing inhibitory peptides against SARS-CoV-2 S from the druggable RBMs-hACE2**

The primary structure of SARS-CoV-2 S RBM and hACE2 protein were diagrammatized. As highlighted in pink and green in the middle, datasets of both binding sites on hACE2 and SARS-CoV-2 S were collected. This provided a fundamental direction to determine the specific primary templates for designing anti-SARS-CoV-2 S peptides. On the hACE2 side, 3 underscored segments, including Q24-Q42, F72-Q86 and T347-M360 were selected as starting templates based on the known binding sites.



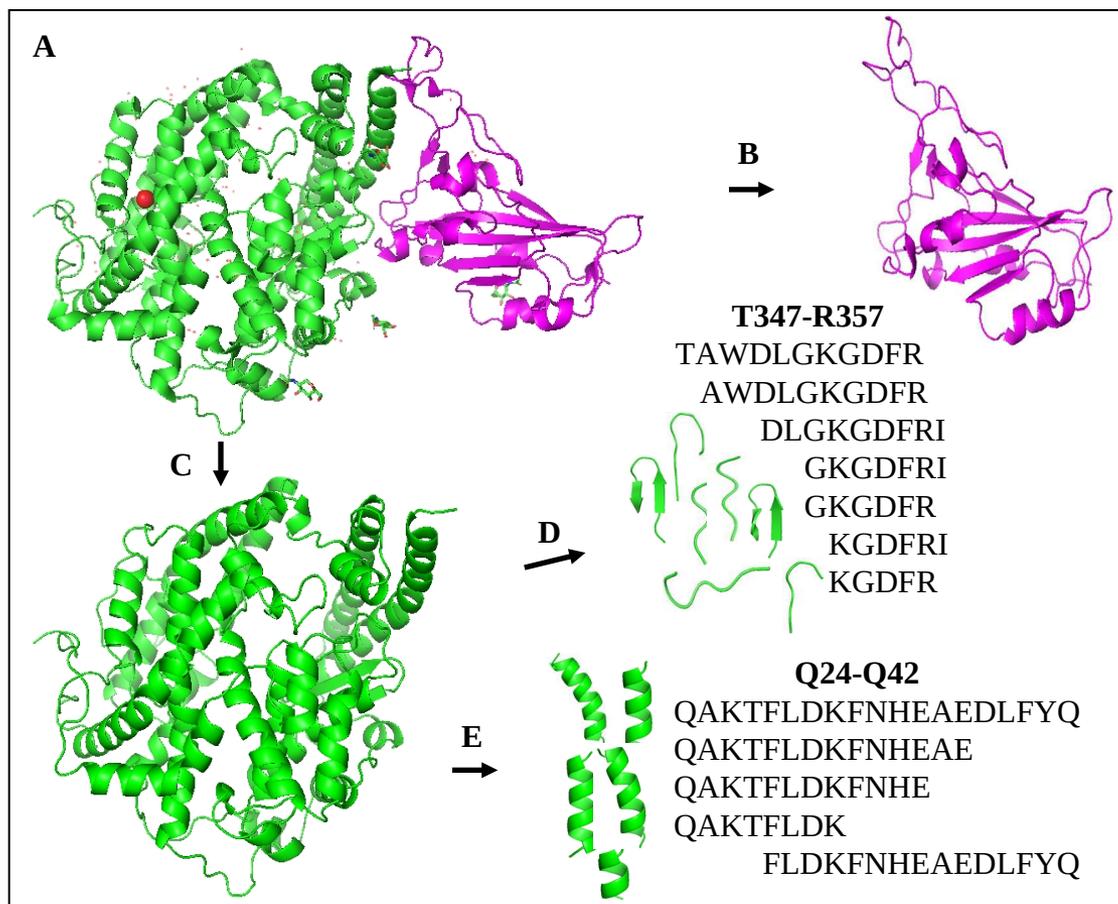
**Figure 2. Work pipeline of the novel design strategy for mining anti-COVID-19 peptides from constructed RBMs-hACE2 isomer libraries**

As shown in the flowchart, hACE2 / SARS-CoV-2 S complex dataset was downloaded and used for further extraction of primary templates and SARS-CoV-2 S. Via traversal and permutation methods, peptide homolog isomer libraries were successively generated from the three RBMs-hACE2 templates Q24-Q42, F72-Q86 and D347-M360. The isomer candidates were selected by the high stability, interaction activity and transmembrane potentials. The specificity and binding affinity of the selected inhibitors were validated by final peptide-protein docking assays of each RBM-hACE2 isomer and SARS-CoV-2 S.



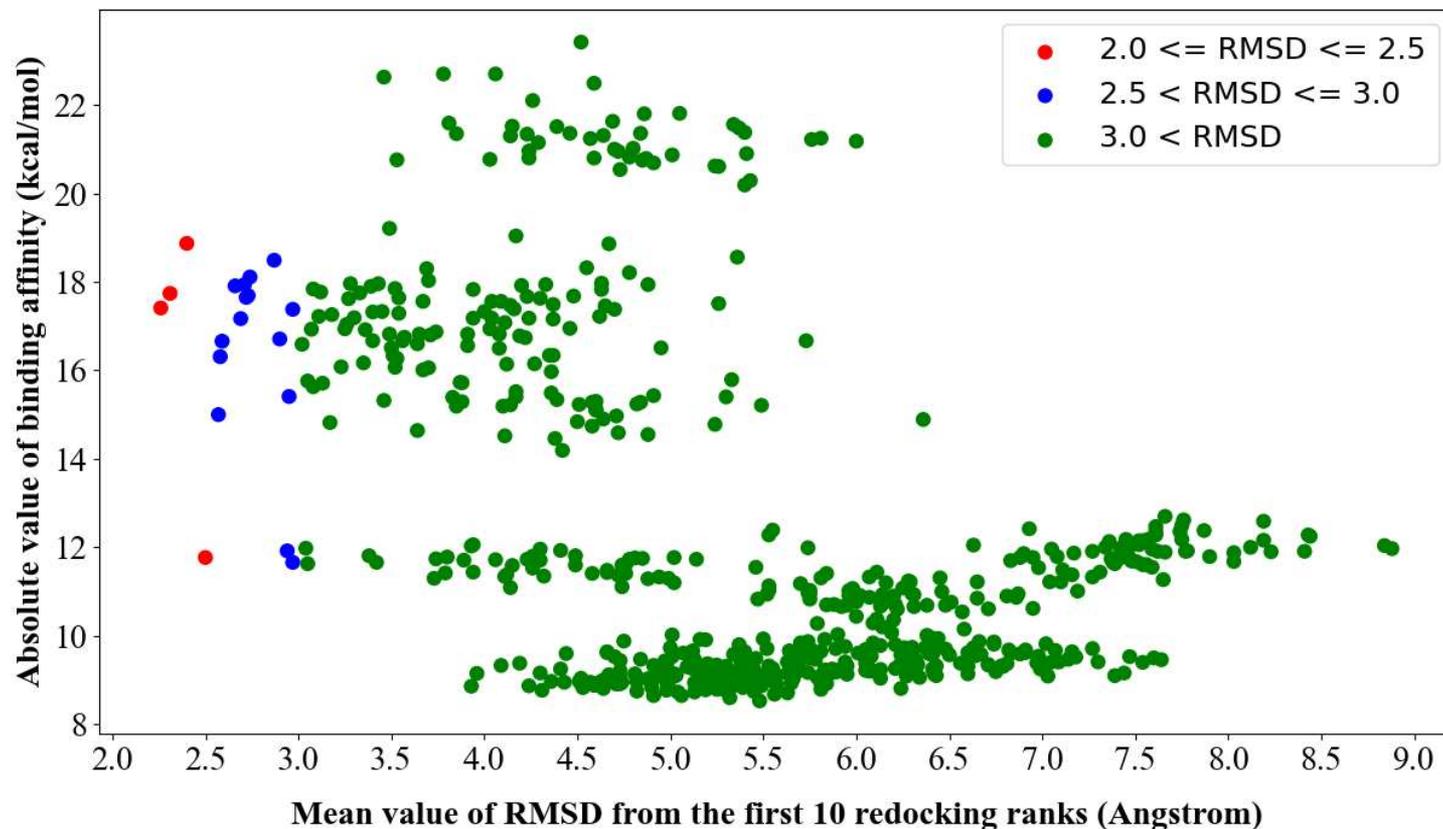
**Figure 3. Identification of 4 anti-SARS-CoV-2 S peptides by multi-step approaches**

The left multi-layers filter shows all steps of antiviral peptide mining in this study. The right table indicates the number of peptides analyzed by the method in the same row and the outputs from the corresponding wild type in the same column. Finally 4 distinct peptides were selected from 12 millions candidates by these methods. The underscored sequence represents transmembrane potential.



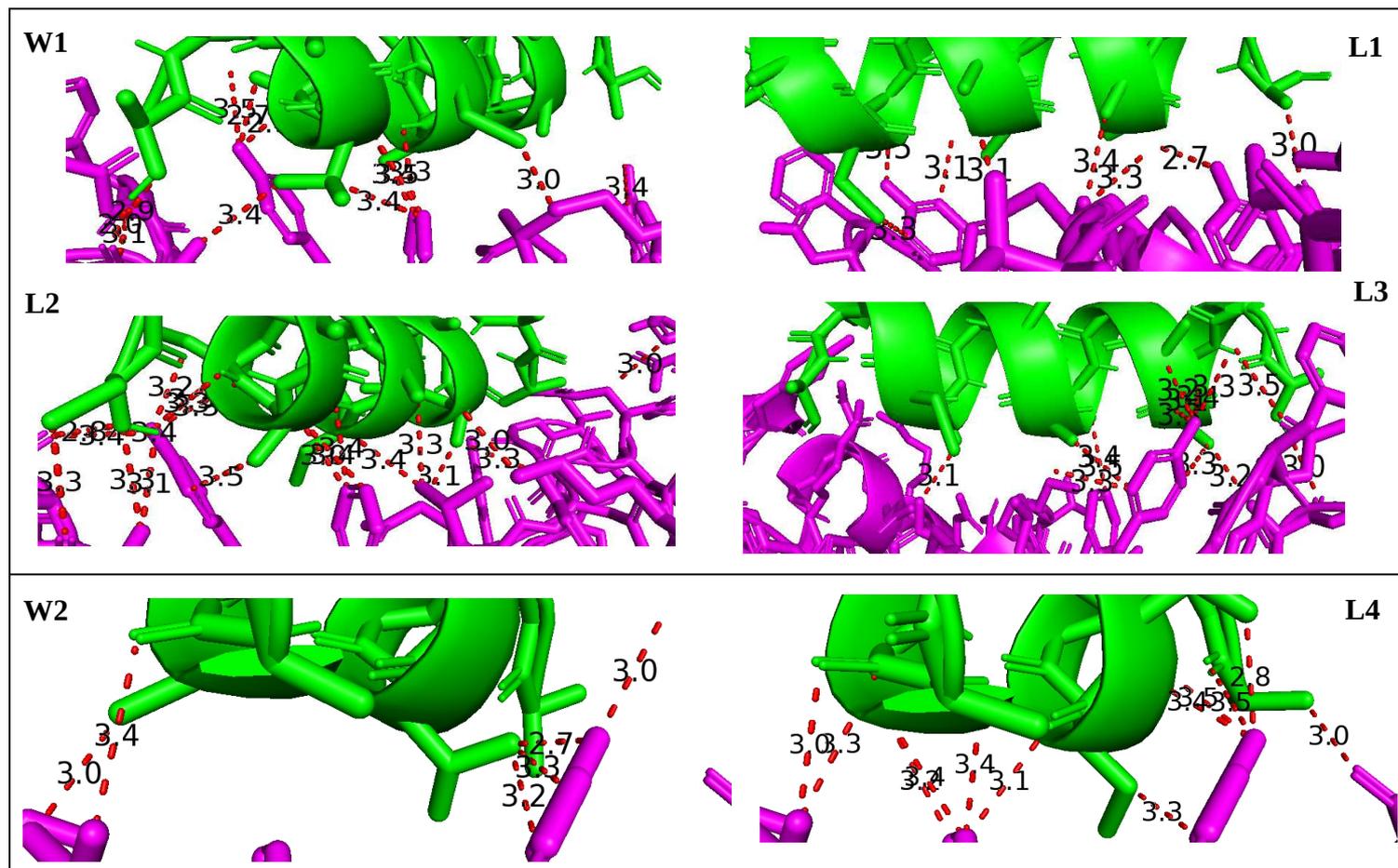
**Figure 4. Extraction of sequential and spatial RBMs-hACE2 template homologs**

Twelve RBMs-hACE2 homolog segments were obtained from the wild type templates of T347-R357 and Q24-Q42. Both sequences and structures of these segments (D-E), SARS-CoV-2 S (B) and hACE2 (C) were extracted from hACE2-SARS-CoV-2 S complex (A) by PyMol. Sequences from D and E were input for constructing isomer libraries. Structures of peptides and SARS-CoV-2 were prepared for generating docking maps.



**Figure 5. Peptide-protein docking assays between 618 RBMs-hACE2 isomers and SARS-CoV-2 S**

As shown in Figure 5, outputs of 618 docking assays were plotted by RMSD against binding affinity. Each dot represents the docking result from each peptide candidate. As shown in left, 4 isomers in red were determined as promising inhibitors against SARS-CoV-2 S because both values of RMSD and binding affinity were  $< 2.5$  and  $< 0$ .



**Figure 6. Evaluation of molecular / atomic contacts of the SARS-CoV-2 spike-peptide docking complexes**

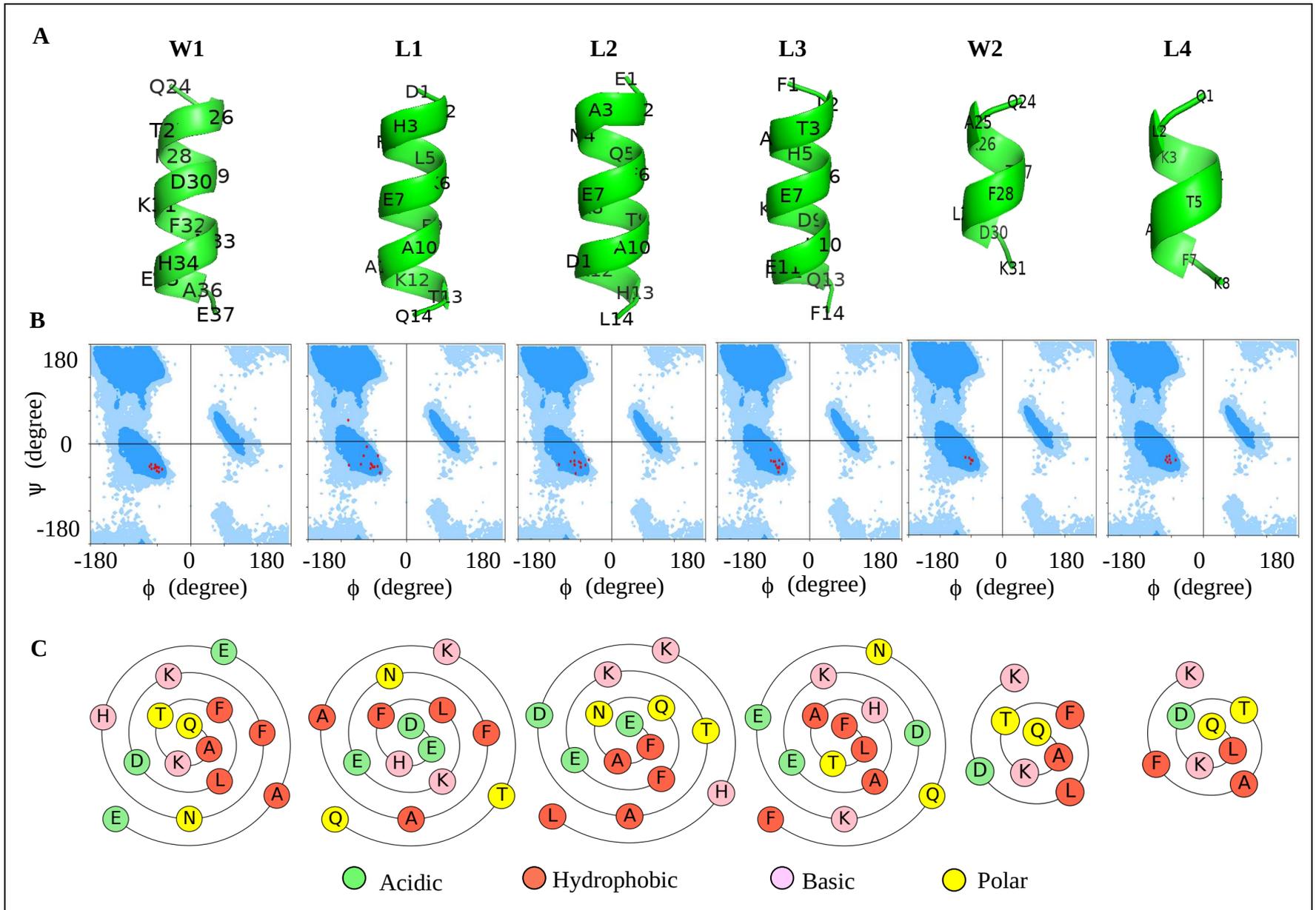
These images show the general contacts between peptides (green) and SARS CoV-2 S (pink) within 3.5 angstroms. Data represent the contact as red-dotted line and the distance label in black. The two group images include the selected Lead 1-3 (L1-3) derived from wild type 1 (W1) and Lead 4 (L4) derived from wild type 2 (W2).

**Table 1. Significant interface of the SARS-CoV-2 spike-peptide docking complexes**

W1	Length	CoV2-RBD	L1	Length	CoV2-RBD	L2	Legth	CoV2-RBD	L3	Length	CoV2-RBD
Q24(CD)	3.5	N487(ND2)	D1(N)	3.3	A475(CB)	N4(CG)	3.3	Y489(CE2)	F1(H)	3.0	A475(O)
Q24(OE1)	2.7	N487(ND2)	D1(N)	2.8	N487(OD1)	Q5(CG)	3.5	Y489(OH)	A4(C)	3.3	Y489(OH)
T27(CG2)	3.5	F456(CE1)	D1(CG)	3.3	A475(O)	E7(CG)	3.5	K417(CE)	A4(O)	3.5	Y489(CZ)
F28(CA)	3.4	Y489(OH)	D1(H)	3.0	A475(CB)	E7(CG)	3.1	K417(NZ)	A4(O)	3.5	Y489(OH)
F28(CB)	3.5	Y489(OH)	D1(H)	3.4	N487(OD1)	K8(CG)	3.1	Y489(CE1)	A4(CB)	3.2	A475(CB)
D30(OD2)	2.9	K417(NZ)	D1(H)	3.4	Y489(OH)	D11(CB)	3.4	L455(CD2)	A4(CB)	3.3	Y489(CE2)
K31(NZ)	3.4	Q493(CD)	F4(O)	3.4	F456(CE1)	D11(CG)	2.7	Y453(OH)	H5(N)	3.0	Y489(OH)
K31(NZ)	2.9	Q493(NE2)	F4(O)	3.0	F456(CZ)	D11(CG)	3.3	L455(CD2)	H5(CA)	3.2	Y489(OH)
H34(CD2)	2.9	Y453(OH)	F4(CB)	3.5	Y489(CE2)				H5(CG)	3.5	N487(OD1)
E35(OE2)	3.5	Q493(NE2)	L5(N)	3.3	Y489(OH)				H5(CG)	3.3	Y489(OH)
E37(OE2)	3.5	Y505(OH)	L5(CA)	3.3	Y489(OH)				H5(H)	3.4	Y489(OH)
			L5(CB)	3.2	Y489(OH)				E7(CB)	3.4	F456(CE1)
			L5(H)	3.4	Y489(OH)				K8(CB)	3.5	Y489(CD1)
			E7(O)	3.3	L455(CD2)				K8(CG)	3.3	Y489(CD1)
			E7(CG)	3.0	K417(NZ)				E11(CG)	3.1	Y453(OH)
			N8(N)	3.5	F456(CZ)						
			N8(CA)	3.4	L455(CD2)						
			N8(H)	3.4	F456(CZ)						
			A11(CB)	3.3	Y453(OH)						
			A11(CB)	3.1	Q493(CG)						

W2	Length	CoV2-RBD	L4	Length	CoV2-RBD
Q24(CD)	3.5	N487(ND2)	Q1(O)	3.5	Y489(OH)
Q24(OE1)	2.7	N487(ND2)	Q1(CB)	3.0	N487(OD1)
T27(CG2)	3.5	F456(CE1)	D4(O)	3.1	F456(CZ)
F28(CA)	3.4	Y489(OH)	D4(CB)	3.2	Y489(CE2)
F28(CB)	3.5	Y489(OH)	T5(N)	3.4	Y489(OH)
D30(OD2)	2.9	K417(NZ)	T5(OG1)	2.8	Y489(OH)
K31(NZ)	3.4	Q493(CD)	T5(H)	3.5	Y489(OH)
K31(NZ)	2.9	Q493(NE2)	F7(C)	3.4	L455(CD2)
			F7(O)	3.0	L455(CD2)
			F7(CB)	3.2	F456(CE1)
			F7(CB)	3.4	F456(CZ)
			K8(H)	3.4	F456(CZ)

As shown in Table 1., the general contacts within 3.5 angstroms between the peptide ligand and SARS-CoV-2 S receptor were generated. In order to compare the contacts among the isomer inhibitors and wild types, data represent docking complex in each column group and residue of corresponding number, atom and length in each row.



**Figure 7. Structural comparison between each selected RBMs-hACE2 isomer and the wild type origin**

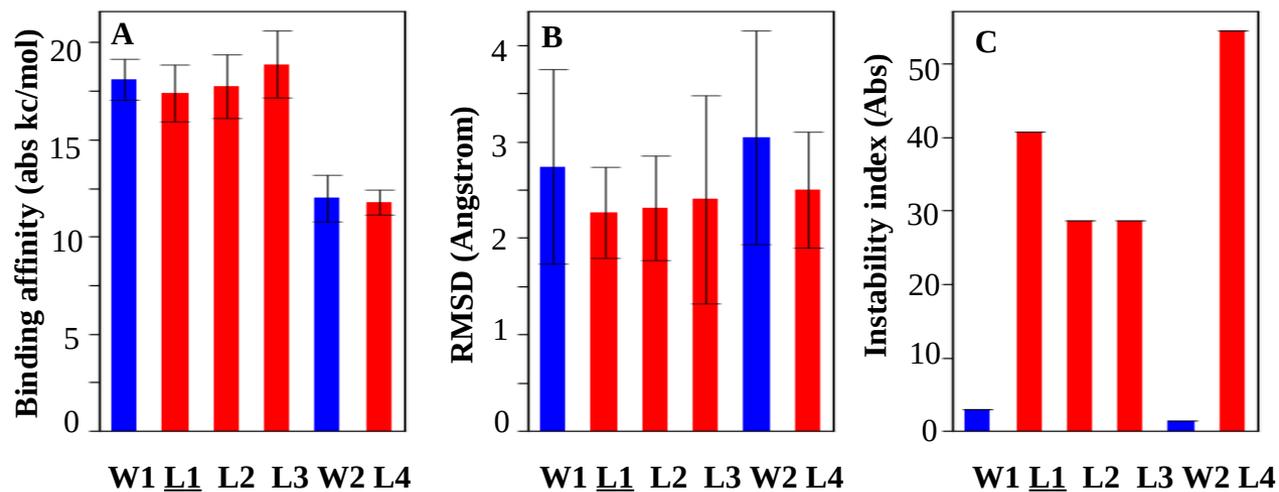
### Figure 7. Structural comparison between each selected RBMs-hACE2 isomer and the wild type origin

The tertiary, secondary and primary structure of each selected inhibitor and wild type origin were visualized by PyMol spatial extraction, Ramachandran and Wenxiang plots. As shown in A, each helical backbone of the isomer is similar to the structure of wild type extract. These 3D structures are indicated as reliable right-handed alpha helices by the Ramachandran plots (B). The well-retained helical backbone can be observed in all isomers by Wenxiang plots (C). Typical hydrophobic face is not seen in Lead 1, 3 and 4 though seems to be in W1, W2 and L2.

**Table 2. Peptide sequences of the selected isomers and wild type origins**

<b>RBM-hACE2</b>	<b>Q24-E37</b>	<b>Q24-K31</b>
<b>Wild types (W)</b>	W1 QAKTFLDKFNHEAE	W2 QAKTFLDK
<b>Isomers (leads)</b>	L1 DEHFLKENFAAKTQ L2 EFANQFEKTADKHL L3 FLTAHAEKDKENQF	L4 QLKDTAFK

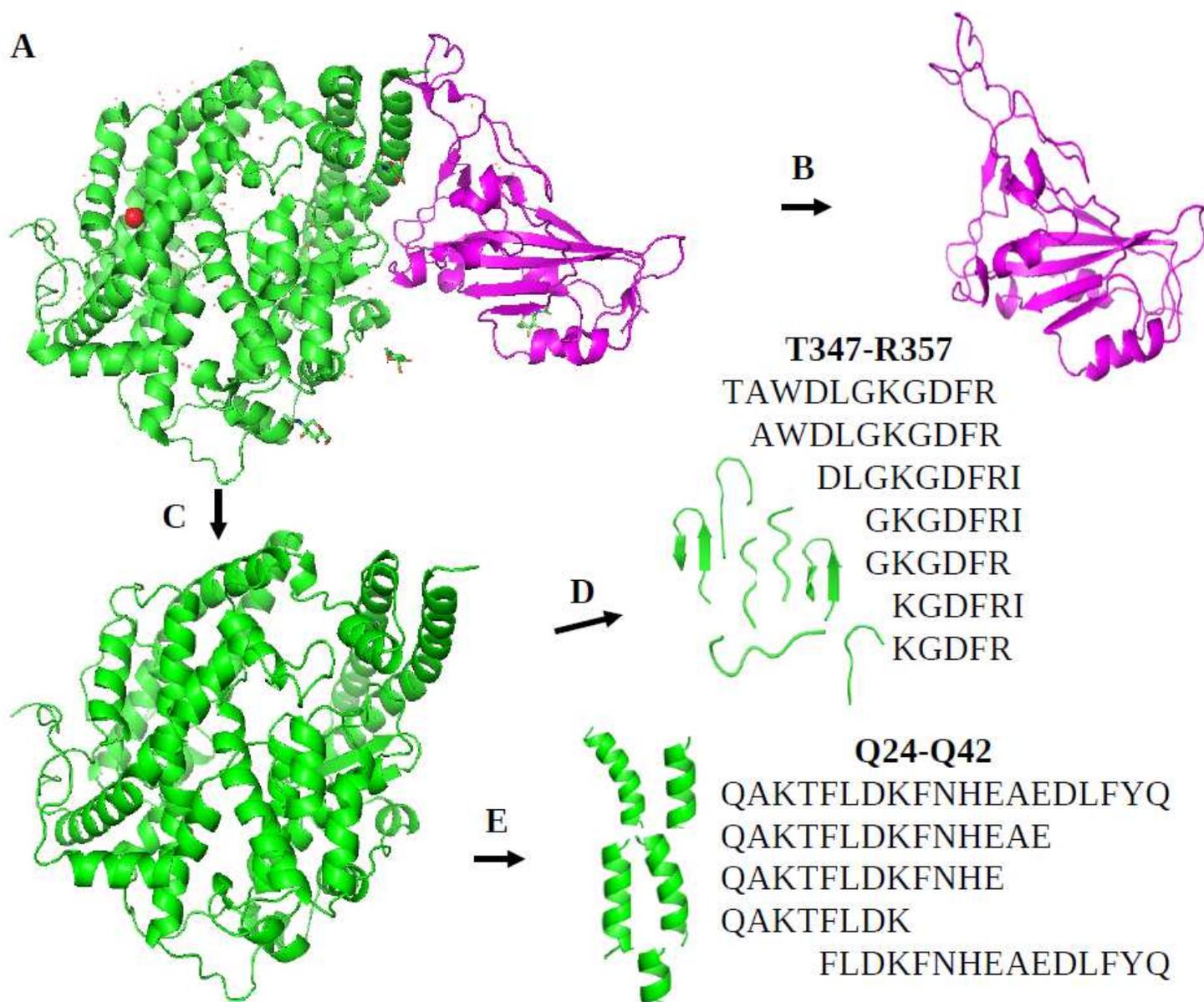
Four RBMs-hACE2 isomers L1-L4 were identified as the potential inhibitors against SARS-CoV-2 S. Leads L1-L3 were derived from the wild type W1 and L4 from W2.



**Figure 8. Pairwise comparison of docking output and stability between each selected RBMs-hACE2 isomer and the wild type origin**

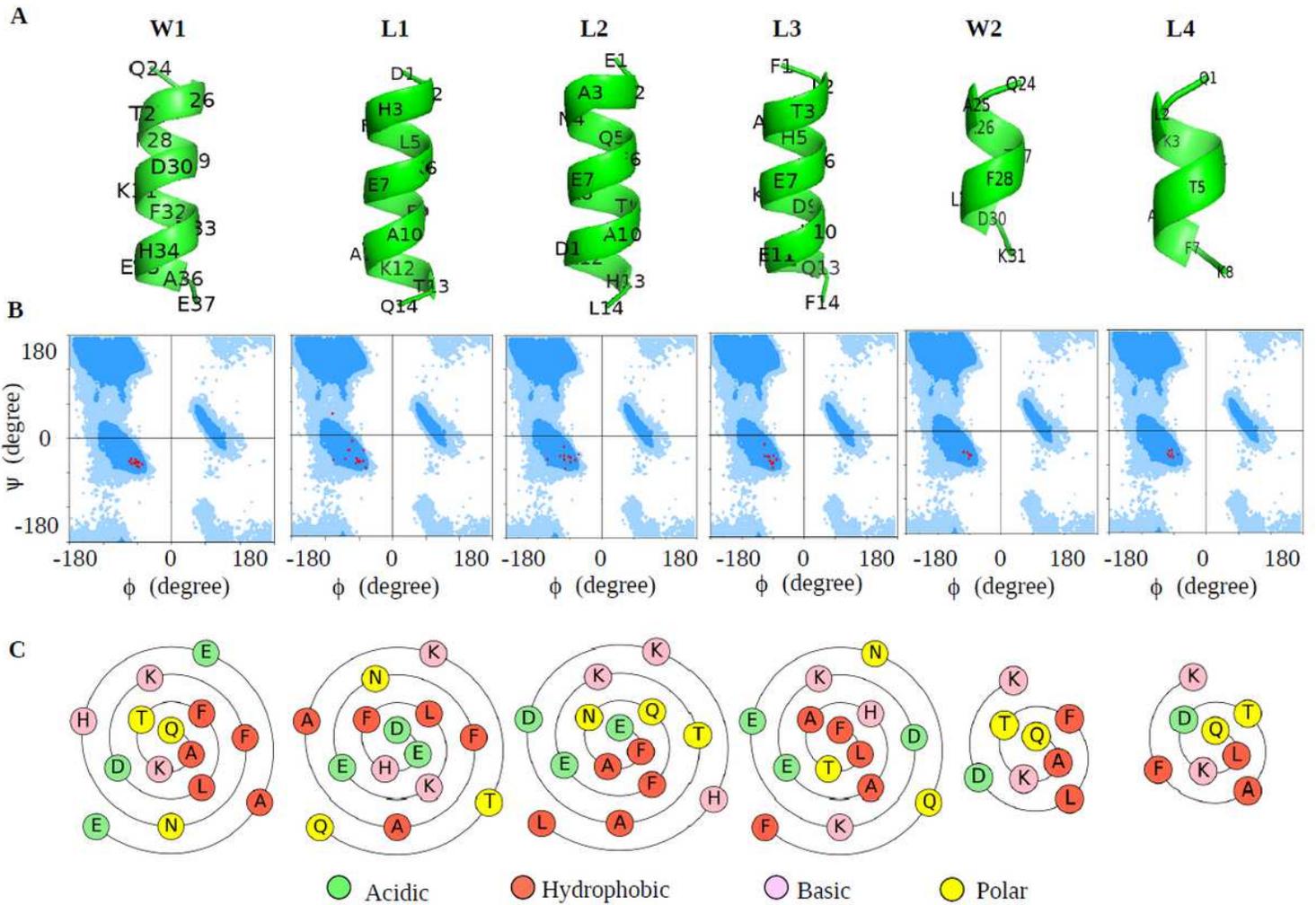
Four distinct isomers were selected via the multi-layers filter and finally validated by experiential thresholds of RMSD and binding affinity though there were no statistical differences of RMSD and binding affinity among the isomers and wild types (all p values > 0.05) (A-B). As shown in C, the stability of isomers L1-L4 was increased 30-50 times compared to the corresponding wild type.

## Figures



**Figure 1**

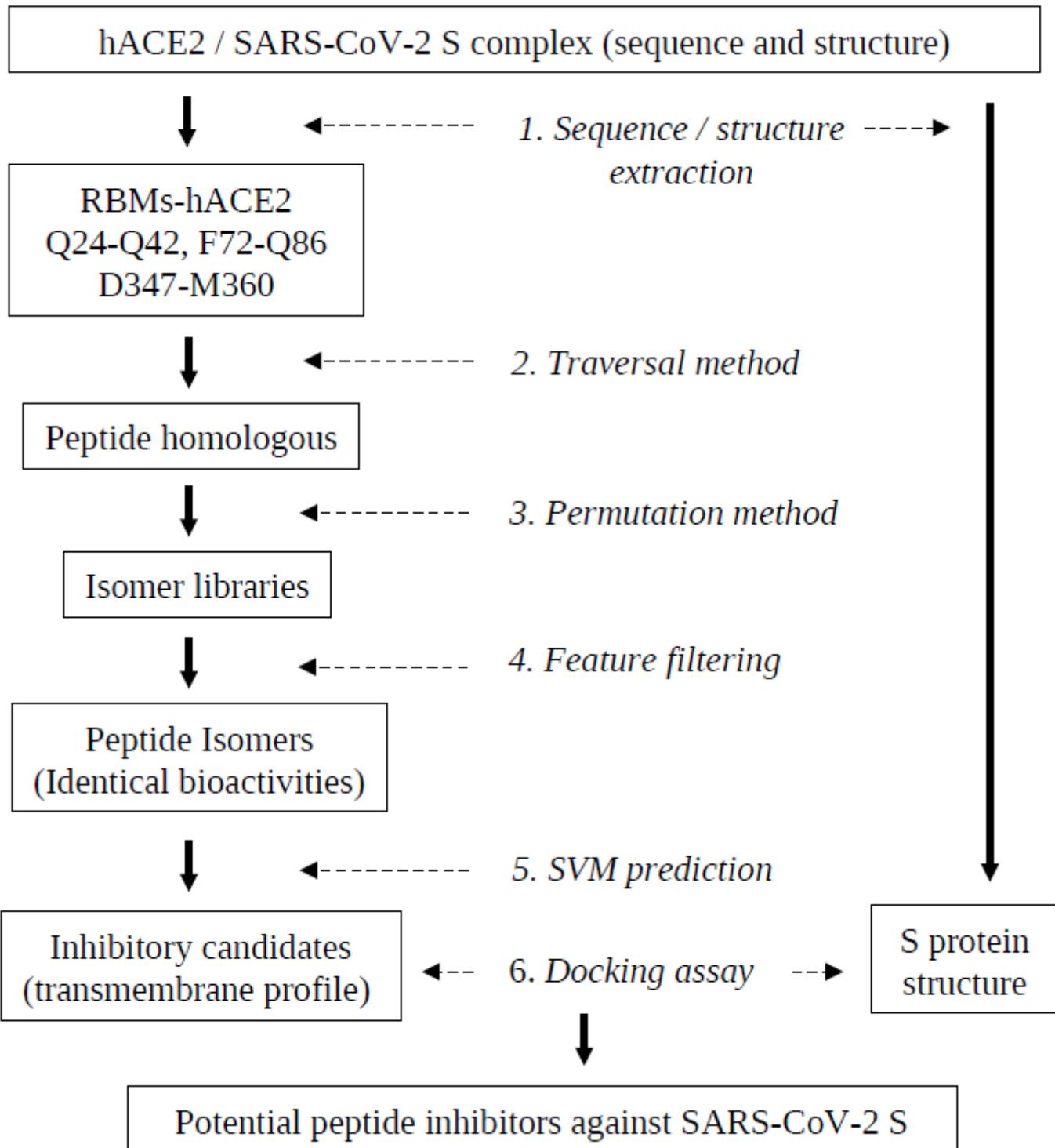
Extraction of sequential and spatial RBMs-hACE2 template homologs Twelve RBMs-hACE2 homolog segments were obtained from the wild type templates of T347-R357 and Q24-Q42. Both sequences and structures of these segments (D-E), SARS-CoV-2 S (B) and hACE2 (C) were extracted from hACE2-SARS-CoV-2 S complex (A) by PyMol. Sequences from D and E were input for constructing isomer libraries. Structures of peptides and SARS-CoV-2 were prepared for generating docking maps.



**Figure 1**

Structural comparison between each selected RBMs-hACE2 isomer and the wild type origin. The tertiary, secondary and primary structure of each selected inhibitor and wild type origin were visualized by PyMol spatial extraction, Ramachandran and Wenxiang plots. As shown in A, each helical backbone of the isomer is similar to the structure of wild type extract. These 3D structures are indicated as reliable right-handed alpha helices by the Ramachandran plots (B). The well-retained helical backbone can be observed in all isomers by Wenxiang plots (C). Typical hydrophobic face is not seen in Lead 1, 3 and 4 though seems to be in W1, W2 and L2.

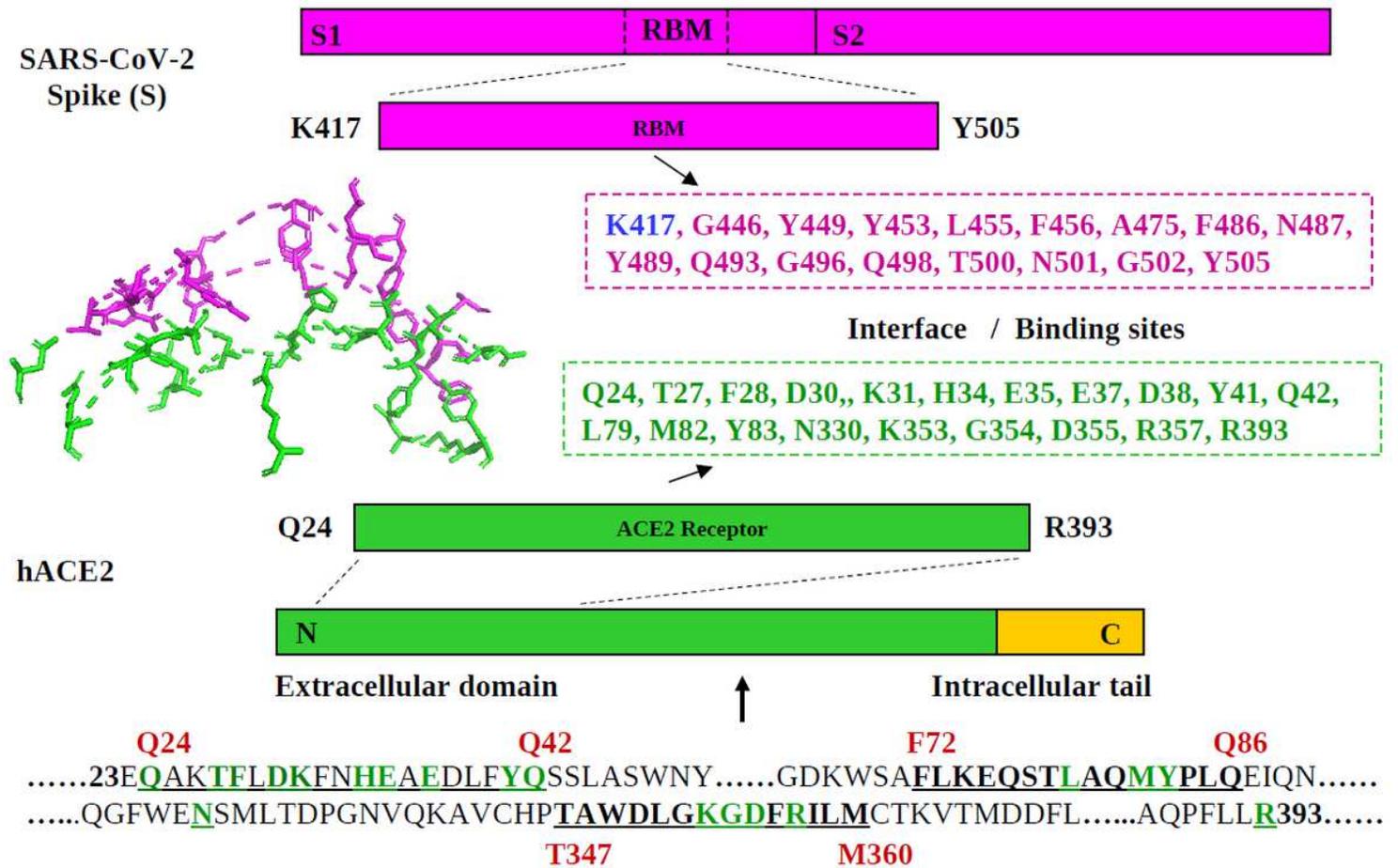




**Figure 1**

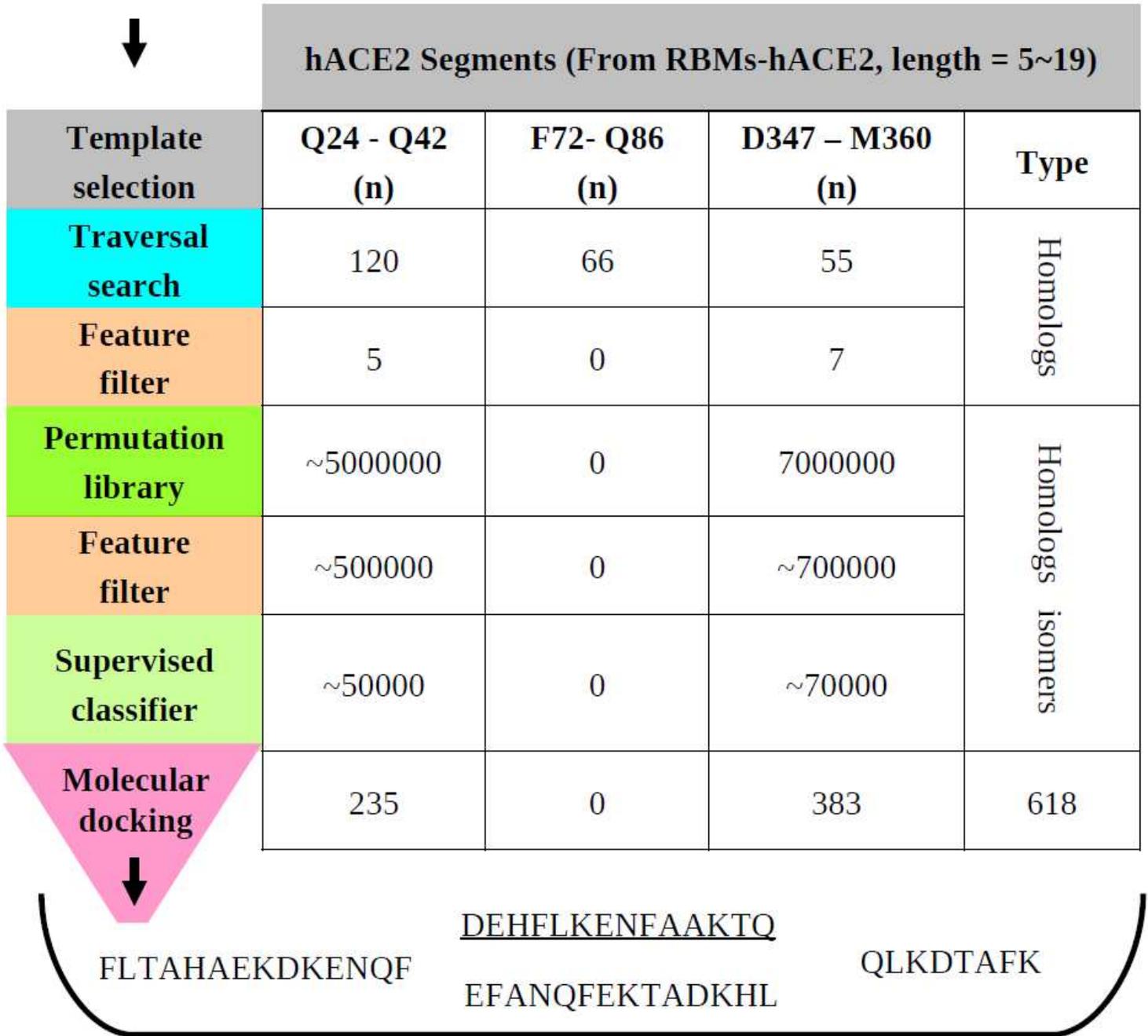
Work pipeline of the novel design strategy for mining anti-COVID-19 peptides from constructed RBMs-hACE2 isomer libraries. As shown in the flowchart, hACE2 / SARS-CoV-2 S complex dataset was downloaded and used for further extraction of primary templates and SARS-CoV-2 S. Via traversal and permutation methods, peptide homolog isomer libraries were successively generated from the three RBMs-hACE2 templates Q24-Q42, F72-Q86 and D347-M360. The isomer candidates were selected by the

high stability, interaction activity and transmembrane potentials. The specificity and binding affinity of the selected inhibitors were validated by final peptide-protein docking assays of each RBM-hACE2 isomer and SARS-CoV-2 S.



**Figure 1**

Selection of primary template for designing inhibitory peptides against SARS-CoV-2 S from the druggable RBMs-hACE2. The primary structure of SARS-CoV-2 S RBM and hACE2 protein were diagrammatized. As highlighted in pink and green in the middle, datasets of both binding sites on hACE2 and SARS-CoV-2 S were collected. This provided a fundamental direction to determine the specific primary templates for designing anti-SARS-CoV-2 S peptides. On the hACE2 side, 3 underscored segments, including Q24-Q42, F72-Q86 and T347-M360 were selected as starting templates based on the known binding sites.



**Figure 1**

Identification of 4 anti-SARS-CoV-2 S peptides by multi-step approaches. The left multi-layers filter shows all steps of antiviral peptide mining in this study. The right table indicates the number of peptides analyzed by the method in the same row and the outputs from the corresponding wild type in the same column. Finally 4 distinct peptides were selected from 12 millions candidates by these methods. The underscored sequence represents transmembrane potential.

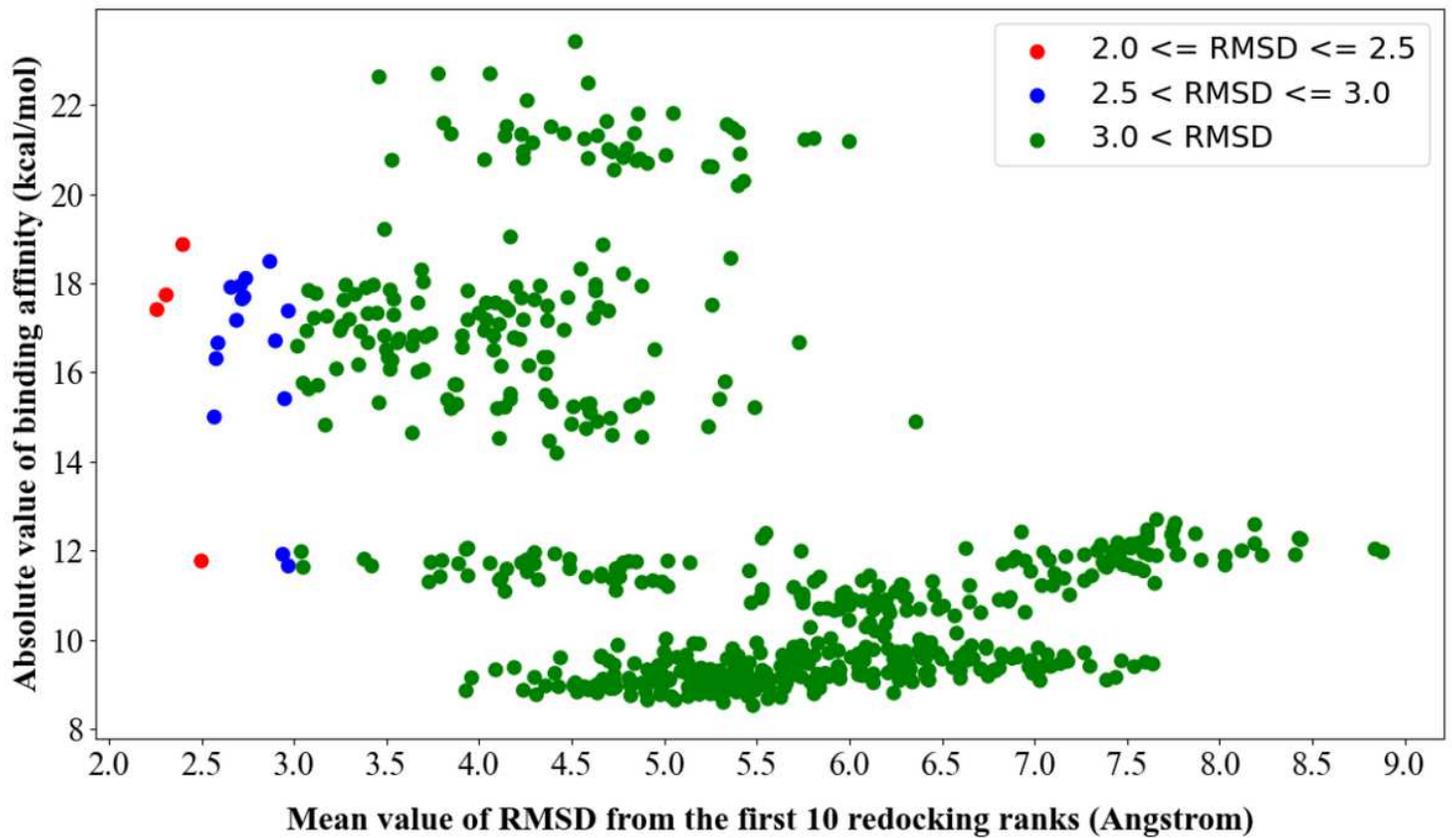
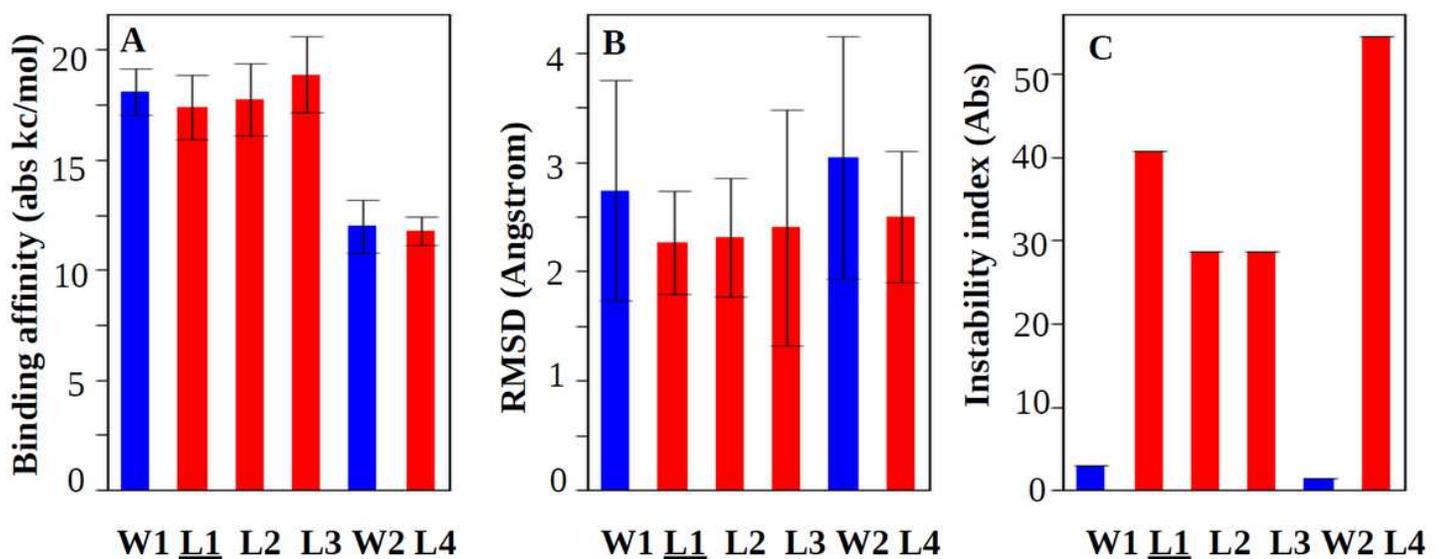


Figure 1

Peptide-protein docking assays between 618 RBMs-hACE2 isomers and SARS-CoV-2 S As shown in Figure 5, outputs of 618 docking assays were plotted by RMSD against binding affinity. Each dot represents the docking result from each peptide candidate. As shown in left, 4 isomers in red were determined as promising inhibitors against SARS-CoV-2 S because both values of RMSD and binding affinity were  $< 2.5$  and  $< 0$ .



## Figure 1

Pairwise comparison of docking output and stability between each selected RBMs-hACE2 isomer and the wild type origin. Four distinct isomers were selected via the multi-layers filter and finally validated by experiential thresholds of RMSD and binding affinity though there were no statistical differences of RMSD and binding affinity among the isomers and wild types (all p values > 0.05) (A-B). As shown in C, the stability of isomers L1-L4 was increased 30-50 times compared to the corresponding wild type.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS3supplementary.pdf](#)
- [TableS2supplementary.pdf](#)
- [TableS1supplementary.pdf](#)