

## An evolutionary RGD motif in the spike protein of SARS-CoV-2 may serve as a potential high risk factor for virus infection?

Shijia Yan,<sup>1</sup> Haixia Sun,<sup>1</sup> Xianzhang Bu,<sup>1,\*</sup> Guohui Wan<sup>1,\*</sup>

1. National-Local Joint Engineering Laboratory of Druggability and New Drug Evaluation, National Engineering Research Center for New Drug and Druggability (cultivation), Guangdong Province Key Laboratory of Chiral Molecule and Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-Sen University, Guangzhou, 510006, China

\* Correspondence to phsbxzh@mail.sysu.edu.cn (X.B.) and wanguoh@mail.sysu.edu.cn (G.W.)

### Abstract:

Pneumonia caused by a new coronavirus SARS-CoV-2 has caused serious harm to people's lives and health in Wuhan, China. By February 26, 2020, over 80,000 people were infected and 2,814 died from the infection. The initial route of infection is the binding of the spike protein (S protein) of the virus to the angiotensin-converting enzyme 2 (ACE2). From bioinformatics analysis, we found that the S protein of SARS-CoV-2 produced an evolutionary mutation of K403R compared with the S protein of SARS-CoV, forming an adjacent RGD motif at the interaction surface. As the RGD motif is considered as a ligand for many cell surface integrins, we proposed that the binding of S protein of SARS-CoV-2 with integrins may facilitate the infection process of the virus. Therefore, high-throughput virtual screening was performed by choosing the key residues of S protein interface of SARS-CoV-2 and the adjacent RGD motif as potential binding site, to search for the potential agents targeting interaction of S protein of SARS-CoV-2 with both ACE2 and integrins as potential therapeutic drugs. Various libraries including the FDA-approved drugs etc. were screened, and Nadide, Losartan, 9"-Methylthiospermate B and Leonurine etc. were identified as representative potential drugs candidate for COVID-19.

**Key words:** SARS-CoV-2, ACE2, RGD, Spike protein, therapeutic drugs

## Introduction

In December 2019, a pneumonia outbreak, termed COVID-19 by WHO, associated with the 2019 novel coronavirus, named SARS-CoV-2, occurred in Wuhan, Hubei province, China. By 26 February 2020, the epidemic has caused 78,631 laboratory-confirmed infections with 2,747 fatal cases in China. Cases are also reported in several other countries including Korean (1,766 cases), Japan (912 cases), Italy (528 cases), Iran (245 cases), Singapore (93 cases) and United States (60 cases) (1, 2). Since the outbreak represents a pandemic threat, WHO (World Health Organization) has declared it a public health emergency of international concern on 30 January 2020. Most of the early cases had contact history with the original seafood market, and human-to-human transmission of the disease is mainly through respiratory droplets (3). SARS-CoV-2 is more likely to affect older males with comorbidities, and can result in severe and even fatal respiratory diseases such as acute respiratory distress syndrome. However, the viral mutation rate and transmission, infection dynamics and the pathogenicity associated with the virus infection *in vivo* remain unclear.

Coronaviruses are enveloped unsegmented single-stranded positive sense RNA viruses, derived from the Coronaviridae family Nidovirales. The coronavirus isolated from the lower respiratory tract of the patients, SARS-CoV-2, is the 7<sup>th</sup> coronavirus that can infect humans. Other six coronavirus species known to cause diseases in humans include 229E, NL63, OC43, HKU1, Middle East respiratory syndrome coronavirus (MERS-CoV) and severe acute respiratory syndrome coronavirus (SARS-CoV). SARS-CoV-2 belongs to the sarbecoviruses subtype of betacoronavirus (4), and its full-length genome sequences was confirmed to share 79.5% sequence identity to SARS coronavirus (SARS-CoV). Furthermore, it was reported that SARS-CoV-2 is 96.2% identical at the whole-genome level to a bat coronavirus RaTG13 (5-7). The close phylogenetic relationship to RaTG13 provides evidence for a bat origin of SARS-CoV-2, while the intermediate host is currently unclear.

The S glycoprotein of SARS-CoV is a class I viral fusion protein, which is indispensable for virus–cell receptor interactions during viral entry. The S glycoprotein of SARS-CoV binds the cellular receptor angiotensin-converting enzyme 2 (ACE2) and mediates fusion of the viral and cellular membranes. Yang et al. pointed out that ACE2 is the presumptive receptor for SARS-CoV-2 entry into host cells since SARS-CoV-2 shares a similar receptor-binding domain structure with SARS-CoV according to homology modeling (8, 9). Subsequent structure studies have confirmed this presumption (10, 11). Recent cryo-EM structure studies further deciphered the S protein of SARS-CoV-2 and ACE2 interaction at Angstrom resolution level and revealed that the overall ACE2-binding mode of S protein of SARS-CoV-2 is almost identical to the mode of S protein of SARS-CoV (12, 13). However, the

outbreak of SARS-CoV-2 started from December 2019 is much worse than that of SARS-CoV in 2013 according to the daily increasing scale of accumulated cases reported worldwide. Though similarity in structure and sequence supports the convergent evolution between SARS-CoV-2 and SARS-CoV, the underlying mechanisms of distinct infection capability still remain unclear. The variation of crucial residues in S protein of SARS-CoV-2 and their counterpart receptors may contribute to the infection efficiency of the virus. Based on this rationale, we performed the bioinformatics analysis of the S protein sequences of coronaviruses and found that SARS-CoV-2 produced an evolutionary mutation of K403R in S protein of SARS-CoV-2 compared with that in S protein of SARS-CoV, forming an adjacent RGD motif at the interaction surface. The RGD motif is the cell attachment site of a large amount of adhesive extracellular matrix and cell surface proteins and recognized by integrins. The evolutionary obtainment of the RGD motif in SARS-CoV-2 may play an important role in promoting rapid transmission from people to people. Therefore, we performed high-throughput virtual screening by choosing the key residues of S protein interface of SARS-CoV-2 and the adjacent RGD motif as potential binding site, to screen out the potential agents targeting interaction of S protein of SARS-CoV-2 with both ACE2 and integrins as potential therapeutic drugs.

## Results

### *Bioinformatics analysis of the S protein of SARS-CoV-2*

We performed multiple sequences alignment of S proteins from SARS-CoV-2 and other coronavirus that infect humans, including SARS, OC43, MERS, NL63, 229E, HKU1, RaTG13, Bat-SARS-like coronavirus, was used as a control. Our results showed that the S protein of SARS-CoV-2 maintains the highest homology (96%) sequence with the S protein of RaTG13, and 75% of homology sequence with the S protein of SARS-CoV, but is significantly different from other coronavirus (Fig. 1A-B). Evolutionarily, SARS-CoV-2 has obtained some features for its adaption in human other than RaTG13, and followed the path as SARS-CoV did. Previous group identified an "RXXR" furin site in SARS-CoV-2, indicating the similar packaging system the virus may use as avian and human influenza viruses (14, 15). We used the most recent crystal structure of S protein of SARS-CoV-2 and ACE2 complex to locate the binding surface of S protein to ACE2, and found G482, V483, E484, Q493, S494, Q498 and N501 varied from the ones of S protein of SARS-CoV, where G482 was shown as an insertion compared to S protein of SARS-CoV (Fig. 1C). But the binding affinity alone between S protein and ACE in SARS-CoV-2 was suggested to vary in the range from 10 nM to 60 nM, failing to explain the unusual high transmissibility of SARS-CoV-2 (12, 13). Notably, we found that S protein of SARS-CoV-2 produced an evolutionary mutation K403R, located at site 403, forming an RGD motif (Arg-Gly-Asp) in

the S protein next to the binding site, which has neither been found in RaTG13 nor in SARS-CoV (Fig. 1D). Tracing back to its genome nucleotide sequences, we found this RGD motif was derived from the point mutation of the second codon of “ACA” into “AGA”, creating an Arg from Thr compared with the sequence in RaTG13, while the counterpart in SARS is “AAG” coding for Lys (Fig. 1E).

### *The potential clinical significance of the S protein harboring the RGD motif in SARS-CoV-2*

The RGD sequence is the cell attachment site of a large number of adhesive extracellular matrix, blood, and cell surface proteins, which can recognize integrin on the surface of various epithelial cells and is important to promote cell adhesion (16). Integrin is a transmembrane heterodimer consisting of two noncovalently bound transmembrane subunits  $\alpha$  and  $\beta$ , many of which recognize RGD sequences displayed on the exposed loops of viral capsid proteins and induces conformational changes in integrin quaternary structure to promote virus internalization. There are 8 integrins that can recognize the RGD sequence including  $\alpha 5\beta 1$ ,  $\alpha 8\beta 1$ ,  $\alpha v\beta 1$ ,  $\alpha v\beta 3$ ,  $\alpha v\beta 5$ ,  $\alpha v\beta 6$ ,  $\alpha v\beta 8$  and  $\alpha IIb\beta 3$  (Fig. 2A), which hold potential to interact with S protein of SARS-CoV-2. Previously, ACE2 was found to bind with integrin, regulate cardiac remodeling signaling pathway and affect cell survival and proliferation (17, 18). A recent study revealed that integrin monomers regulate CCL2 levels in alveolar epithelial cells, recruiting monocytes to induce an inflammatory response (19), which suggests that in addition to interacting with ACE2, the RGD sequence of S protein of SARS-CoV-2 may also be recognized and recruited by integrins in alveolar epithelial cells to accelerate the infection process.

Therefore, taking S protein of SARS-CoV-2\_RBD-ACE2 as the template, we analyzed the 3-dimensional structure of S1 domain in S protein and its interaction interface with ACE2. Our results showed that S protein of SARS-CoV-2 adopts a conformation and interaction mode similar to that of S protein of SARS-CoV when interacting with ACE2 (5) (Fig. 2B). The key interaction region of S protein of SARS-CoV-2 locates in sites 487-505, referring to the region of sites 473-490 in S protein of SARS-CoV which was part of epitope (472-502) that was used for generating antibody against infection of SARS-CoV as previously described (20). The spatial structure of the RGD motif (403-405) is located outside of the S protein and adjacent to its interaction interface with ACE2 (Fig. 2B), which may enhance infection efficiency through recruiting ACE2 by binding integrin. Previous study reported that the S protein processes a dynamic prefusion conformation during fusing into the host cell membrane (10, 21). When the RBD of S1 subunit undergoes hinge-like conformational shifts, the change exposes or hides the key region of binding domain to access to ACE2 by controlling the “up” and “down” conformation, respectively (10). In this scenario, the

RGD motif would be exposed to the surface of the host cell membrane in conjunction with the key binding region during prefusion conformation movements. Since the RGD sequence is not contained in the receptor (ACE2)-binding domain, it is possible that the variation amino acid lead to virus-integrin-interaction, which enhances the efficiency of cell invasion and promote respiratory tract and lung infection. Previous study reported that the binding kinetics for ACE2 and S protein of SARS-CoV-2 showed much higher affinity than that of S protein of SARS-CoV-2 at the molecular level by SPR sensorgram (22). Based on this, the presence of the RGD motif may account for potential mechanisms of S protein of SARS-CoV-2 binding to ACE2 with higher affinity than S protein of SARS-CoV (Fig. 2C, right panel). Upon prefusion conformation change, the RGD motif is exposed to the cell surface and facilitates the binding of the S protein with integrins in host cells and accelerate the fusion of the virus. Once interacting with integrins, ACE2 may be recruited to the binding complex by integrins and facilitate the invasion of the virus. While we understand that this increasing affinity may attribute to the binding of S protein with ACE2 and integrins simultaneously, and it may rarely occur in virus with high transmissibility. Another possible mechanism is proposed that the RGD motif may bind to integrins parallelly or sequentially in an ACE2-independent manner (Fig. 2D). This is supported by previous study that ACE2 can serve as a cell adhesion substrate and regulate integrin signaling (18). They showed that interactions of ACE2 with  $\alpha 5\beta 1$  and  $\alpha 2\beta 1$  occurred independently of an RGD motif, and the RGD motif was inaccessible in ACE2, suggesting the redundancy of ACE2 in integrin binding (18). In this scenario, after integrins in lung alveolar epithelial cells bind to the RGD motif in S protein of SARS-CoV-2 independently, the frequent interaction between integrins and ACE2 may promote the contact of S protein to ACE2, increasing the fusion of the virus in the cells. Phenomenally, by comparing the infection efficiency in clinical data, accumulated infected cases of SAR-CoV-2 are significantly higher than those of SARS in 2013. For the first month of the outbreak, over 80,000 people were infected by SARS-CoV-2, while only 3,389 people were infected by SARS in 2013 (Fig. 2E). A recent study suggested that obese people and some cancer patients with high levels of ACE2 might be more vulnerable to SARS-CoV-2 infection (23).  $\alpha 5$  intergrin was found to be expressed at a high level in preadipocytes but decreased in adipocytes, which might support the role of integrin in mediating virus infection. With facilitation of binding with integrins, it may be one of the explanations for the higher transmissibility of SARS-CoV-2, though these mechanisms need to be verified in vitro and in vivo models in future.

#### *Identification of chemical compounds as potential inhibitors for interaction of S protein of SARS-CoV-2 and ACE2.*

To screen the potential drug candidates, we have built different libraries that include the FDA-approved drug entities (2040 species), our own medicine food homology natural

products derivatives entities (1500 species) and cyclic peptides entities (230 species), and along with virtual bioactive and natural products libraries, adding up to about 15000 species (Fig. 3A). Therefore, we adopted the S protein of SARS-CoV-2 structure extracted from the recent crystal structure of S protein/human ACE2 complex (NMDCS0000001), and chose the key residues in the S protein interface (Q493, Y495, Q498, N501 and Y505) and the R403, D405 as the potential binding site (Fig. 3B) for virtual screening by Sybyl X using Surflex-Dock Geom (SFXC) approach. Several representative agents, including Nadide (NDC active ingredient), Losartan (the angiotensin receptor blockers), and Adenosine phosphate (the veterinary drug for vasodilator) from the FDA-approved drug library; Difludionone-119 and Methyl-benzyloxychadone-844 from natural products derivatives library; GR8-1 [Cyclo (f-P-V-R-f-P-R-L-)], MGR-7 [Cyclo (L-f-P-V-R-L-f-P0-V-R-)], and GR6-2 [(Cyclo (R-f-P-R-f-P-))] from our own cyclic peptide library, and S-9"-Methylithospermate B, and the S-Leonurine from bioactive and natural product library respectively, were hit. The illustrated detail interaction modes between S protein and these representative compounds were shown as indicated, including the S-Nadide interactions (Fig. 4A), the S-Losartan interactions (Fig. 4B), the S-Difludionone-119 interactions (Fig. 4C), the S-GR8-1 interactions (Fig. 4D), the S-9"-Methylithospermate B interactions (Fig. 4E), and the S-Leonurine interactions (Fig. 4F). These compounds are well docked into a pocket formed by the selected key residues in the S protein interface and the RGD motif by Hydrogen bonds, and/or  $\pi$ - $\pi$ / $\rho$ - $\pi$  interactions respectively, indicating that these hits may potentially block interaction of S protein with ACE2 and integrin.

Importantly, we found that Nadide was scored with high grade (10.7719), which is superior to other hits, implying it may serve as a promising drug candidate for COVID-19. Once the risk of the RGD motif in S protein of SARS-CoV-2 for infection efficiency is confirmed in vivo, novel therapeutic screening strategy to block the interaction between the RGD motif and ACE2 protein is highly recommended in anti-SARS-CoV-2 drug discovery (Fig. 4G).

## Discussion

Since coronaviruses are under extensive mutagenesis and the mutation in key proteins are crucial to the virus, the potential clinical significance of the S protein harboring the RGD motif in SARS-CoV-2 is notable. By 26 January 2020, a total of 8,866 patients including 4,021 laboratory confirmed patients were reported from 30 provinces in China. Chen et al. reported that the R0 (the basic reproductive number) was estimated to be 3.77 (95% CI 3.51-4.05) by comparing with the R0 of SARS (around 3), indicating higher infection efficiency though with a lower estimated mortality risk of ~2% (24, 25). Compared with SARS-CoV, SARS-CoV-2 has comparable, even higher, transmissibility that urging us to uncover its infection mechanism and to develop specific drugs against SARS-CoV-2 to

alleviate the current epidemic.

In our study, we shed light on a new potential mechanism for SARS-CoV-2 infection that the RGD motif on the S glycoprotein may bind to the integrin on the surface of host cells, resulting in higher affinity with the host cells in comparison with SARS-CoV. Although a variety of viruses utilize viron-integrin-interaction for cell entry, there is no evidence that beta-coronaviruses share similar mechanism for their infection. Further investigation should be done to verify whether viron-integrin-interaction plays a critical role in the invasion to the host cells, and to screen the subtype of the binding integrins. Moreover, virus like adenovirus has been proved to activate signaling events via ligation of integrin (26), but the molecular mechanisms and signaling pathways may vary in different viruses. Therefore, signaling events of SARS-CoV-2 that facilitate receptor-mediated endocytosis of virus particles after interacting with integrin needs more clarification.

The S glycoprotein of the coronavirus is crucial to the viral life cycle and is a major target for antiviral drugs and vaccines. Coronavirus entry is a multi-step process involving multiple, distinct domains in S protein that mediate virus attachment to the cell surface, receptor engagement, membrane fusion and protease processing. Infection blockers can be designed to block either ACE2 binding or integrin binding by designing molecules that are highly compatible with S protein. Griffithsin, a lectin extracted from red algae, has been previously reported to bind to oligosaccharides on the surface of various viral glycoproteins, including HIV glycoprotein 120 and SARS-CoV glycoproteins. However its efficacy and delivery system still need to be reassessed as a means of treating or preventing neocoronavirus (27, 28). In our study, we performed high-throughput virtual screening on the basis of libraries including the FDA-approved drugs, medicine food homology natural products and derivatives, and our own cyclic peptides etc. A list of representative compounds from each library was identified to potentially block S protein from ACE2 or integrins. Particularly Nadide was shown to block the interaction of the RGD motif and its unknown integrin counterpart at the same time, which may serve as a promising drug candidate for COVID-19. Integrin-targeted drugs might also modulate virus–ligand affinity and signaling, which is useful in controlling infectious diseases. We anticipate that further investigation of virus–integrin interactions may not only contribute to a broader understanding of viral pathogenesis but also provide with knowledge which eventually leads to the development of novel antiviral therapeutic strategy to block the interaction between the RGD motif and ACE2 protein.

## Materials and Methods

### *Data resources*

The Spike protein sequence of SARS-CoV-2 (QHD43416.1), SARS (AAP50485.1), OC43 (AIX10763.1), MERS (AMO03401.1), NL63 (AGT51394.1), 229E (ABB90529.1), HKU1 (AAT98580.1) and RaTG13 (QHR63300.2) were obtained from NCBI database. The structures and sequences of ACE2 (PDB code: 2AJF) and S protein of SARS-CoV (PDB code: 2AJF) were downloaded from PDB database. Physical and virtual libraries including the FDA-approved drugs, medicine food homology natural products and derivatives, bioactive compounds and active cyclic peptides are established in our group, adding up to 15,000 entities for the screening.

### *Sequence alignment*

The amino acid sequences were aligned by Clustal X 2.1 software using default parameters, and refined using GeneDoc software. The ACE2-interaction region in the spike protein was determined by MOE2015 software (Chemical Computing Group, USA) (29).

### *Virtual screening*

High-throughput virtual screening was performed by Sybyl X2.0 software (Tripos, DE). The S protein of SARS-CoV-2 structure extracted from the recent crystal structure of S protein/human ACE2 complex (NMDCS0000001) was adopted, and key residues in the S protein interface and RGD motif were chosen as the potential binding site to generate the protocol for virtual screening using Surflex-Dock Geom (SFXC) approach. Docking solution was visually inspected and the pose was refined with energy minimization and molecular dynamics (MD) simulations.

## Acknowledgement

This work is supported in part by grants from the National Natural Science Foundation of China 31701114 (GW), 21672266 (XB); and Science and Technology Planning Project of Guangdong Province (2019A050510019, 2019B030301005).

## Reference

1. C. Rothe *et al.*, Transmission of 2019-nCoV Infection from an Asymptomatic Contact in



- Germany. *N Engl J Med*, (2020).
2. M. L. Holshue *et al.*, First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med*, (2020).
  3. J. F. Chan *et al.*, A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* **395**, 514-523 (2020).
  4. N. Zhu *et al.*, A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med*, (2020).
  5. X. Xu *et al.*, 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).
  6. D. Paraskevis *et al.*, Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. *Infection, Genetics and Evolution* **79**, (2020).
  7. P. Zhou *et al.*, Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *bioRxiv*, (2020).
  8. R. Lu *et al.*, Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, (2020).
  9. F. Yang *et al.*, [Pulmonary rehabilitation guidelines in the principle of 4S for patients infected with 2019 novel coronavirus (2019-nCoV)]. *Zhonghua Jie He He Hu Xi Za Zhi* **43**, E004 (2020).
  10. D. Wrapp *et al.*, Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. *bioRxiv*, (2020).
  11. Q. Zhou, R. Yan, Y. Zhang, Y. Li, L. Xia, Structure of dimeric full-length human ACE2 in complex with BOAT1. *bioRxiv*, (2020).
  12. R. Yan, Y. Zhang, Y. Guo, L. Xia, Q. Zhou, Structural basis for the recognition of the 2019-nCoV by human ACE2. *bioRxiv*, (2020).
  13. J. Lan *et al.*, Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *bioRxiv*, (2020).
  14. X. Liu *et al.*, A furin cleavage site was discovered in the Wuhan 2019 human coronavirus S protein. *Chinese Journal of Bioinformatics*, (2020).
  15. J. Chen *et al.*, Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell* **95**, 409-417 (1998).
  16. E. Ruoslahti, RGD and other recognition sequences for integrins. *Annu Rev Cell Dev Biol* **12**, 697-715 (1996).
  17. Q. Lin, R. S. Keller, B. Weaver, L. S. Zisman, Interaction of ACE2 and integrin beta1 in failing human heart. *Biochim Biophys Acta* **1689**, 175-178 (2004).
  18. N. E. Clarke, M. J. Fisher, K. E. Porter, D. W. Lambert, A. J. Turner, Angiotensin converting enzyme (ACE) and ACE2 bind integrins and ACE2 regulates integrin signalling. *PLoS One* **7**, e34747 (2012).
  19. E. J. Plosa *et al.*, beta1 Integrin regulates adult lung alveolar epithelial cell inflammation. *JCI Insight* **5**, (2020).
  20. W. Song, M. Gui, X. Wang, Y. Xiang, Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathog* **14**, e1007236 (2018).
  21. F. Li, Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **3**, 237-261 (2016).

22. X. Liu, B. Zhang, Z. Jin, H. Yang, Z. Rao, The crystal structure of 2019-nCoV main protease in complex with an inhibitor N3. *Protein Data Bank*, (2020).
23. X. Jia *et al.*, Two Things about COVID-19 Might Need Attention. *Preprints*, (2020).
24. N. Chen *et al.*, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507-513 (2020).
25. Q. Li *et al.*, Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*, (2020).
26. P. L. Stewart, G. R. Nemerow, Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol* **15**, 500-507 (2007).
27. A. Zumla, J. F. Chan, E. I. Azhar, D. S. Hui, K. Y. Yuen, Coronaviruses - drug discovery and therapeutic options. *Nat Rev Drug Discov* **15**, 327-347 (2016).
28. C. Lee, Griffithsin, a Highly Potent Broad-Spectrum Antiviral Lectin from Red Algae: From Discovery to Clinical Application. *Mar Drugs* **17**, (2019).
29. G. Wen *et al.*, Stabilizing the monomeric amyloid-beta peptide by tyrocidine A prevents and reverses amyloidogenesis without the accumulation of oligomers. *Chem Commun (Camb)* **53**, 3886-3889 (2017).

## Figure legends

### Figure 1. Bioinformatics analysis of the S protein of the SARS-CoV-2.

- A. Amino acid sequence alignment of the S proteins in the coronaviruses infected human.
- B. Homology of SARS-CoV-2 to other coronaviruses based on the S protein alignment.
- C. Amino acid sequence alignment of the surface domain interacting with ACE2 protein between SARS-Cov-2 and SARS-CoV.
- D. Evolution of the RGD motif in the S protein of SARS-CoV-2 compared with the S protein of SARS-CoV.
- E. Nucleotide mutation produces RGD in SARS-CoV-2.

### Figure 2. The potential clinical significance of the S protein harboring the RGD motif in SARS-CoV-2.

- A. RGD recognizing integrins hold potential to interact with S protein of SARS-CoV-2.
- B. ACE2 binding interface regions of the S protein of SARS-CoV2. Location of the RGD motif was shown in green, and the yellow residues are the interaction surfaces binding with ACE2.
- C. Potential mechanism of S protein of SARS-CoV-2 binding ACE2 with higher affinity than S protein of SARS-CoV. The binding kinetics for ACE2 and S proteins are cited from

previous report by SPR sensorgram (10).

D. Possible mechanism of the RGD motif in S protein binds to integrins in an ACE2-independent manner.

E. Outbreak of SARS-CoV-2 with much higher infection efficiency compared with outbreak of SARS-CoV in 2013.

**Figure 3. Identification of chemical compounds as potential inhibitors for interaction of S protein of SARS-CoV-2 and ACE2.**

A. Flowchart of interaction surface structure-based virtual high-throughput screening.

B. The key residues in the S protein interface and the RGD motif were chosen as the potential binding site to generate the protocol for virtual screening by using Surflex-Dock Geom (SFXC) approach.

**Figure 4. The binding pose of representative hits to S protein of SARS-CoV-2.**

A. Nadide (blue), where carbon atoms of binding site residues are colored in yellow and the RGD motif is shown in green. The hydrogen bonds are illustrated with red dashed lines.

B. Losartan (blue), where carbon atoms of binding site residues are colored in yellow and the RGD motif is shown in green. The hydrogen bonds are illustrated with red dashed lines.

C. Difludionone-119 (yellow), where carbon atoms of binding site residues are colored in gray and the hydrogen bonds are illustrated with red dashed lines.

D. GR8-1 (yellow), where carbon atoms of binding site residues are colored in gray and the hydrogen bonds are illustrated with red dashed lines.

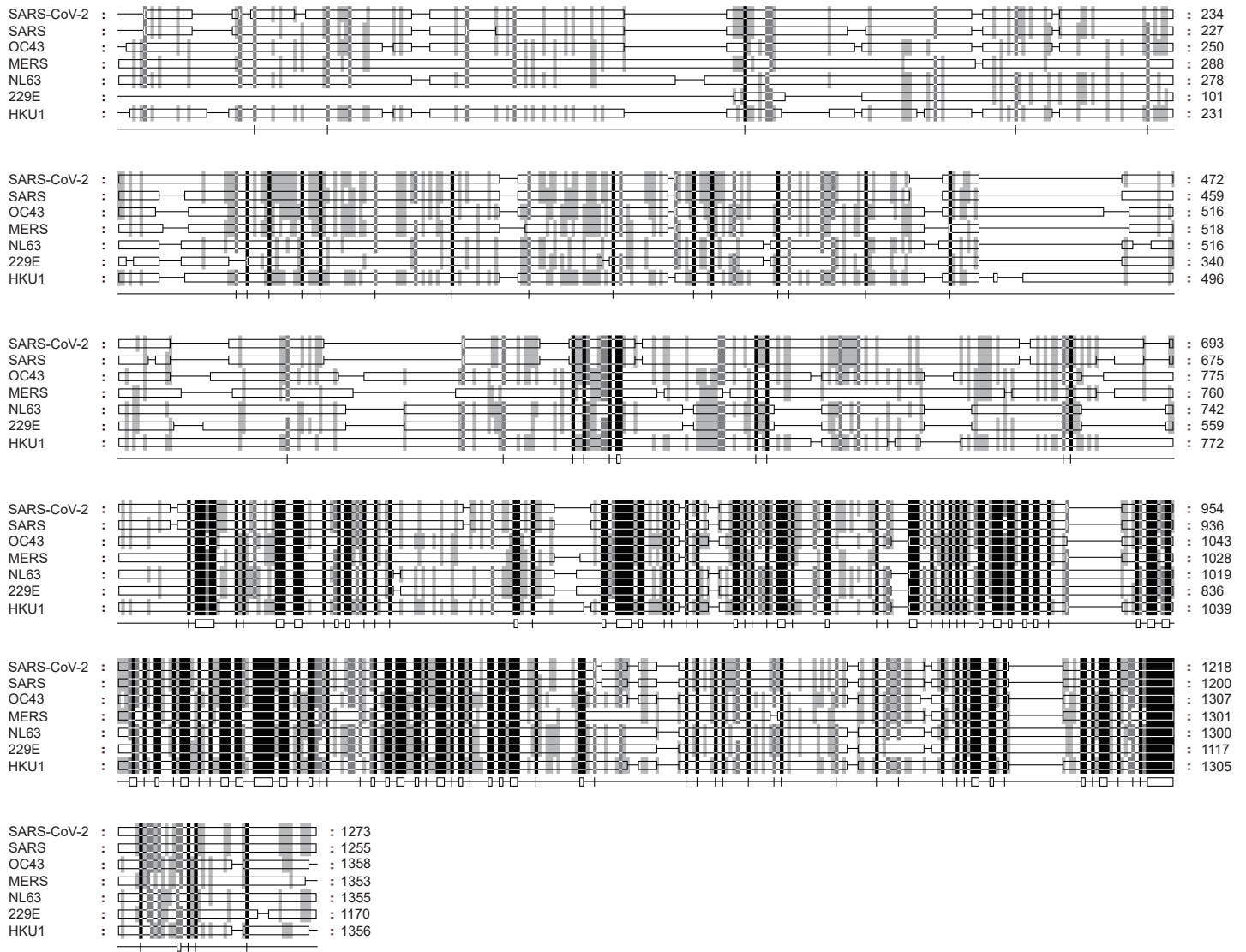
E. 9"-Methylthiospermate B (yellow), where carbon atoms of binding site residues are colored in gray and the hydrogen bonds are illustrated with red dashed lines.

F. Leonurine (yellow), where carbon atoms of binding site residues are colored in gray and the hydrogen bonds are illustrated with red dashed lines.

G. Potential therapeutic strategy to block the interaction between the RGD motif and ACE2 protein.

## Figure 1

A



B

	RaTG13	SARS	OC43	MERS	NL63	229E	HKU1
SARS-CoV-2	97%	75%	26%	24%	19%	19%	24%

C

	Interaction surface			
SARS-CoV-2	474	QAGSTPCNGVEGFNCYFPLQSYGFQPTNGVG	504	
RaTG13	474	QAGSKPCNGQTLNLCYPLRYGFYPTDGVG	504	
SARS	461	SPDGKPCCT-PPALNCYWPLNDYGFYTTGIG	490	
		.....** .:***:** *** * **:*		

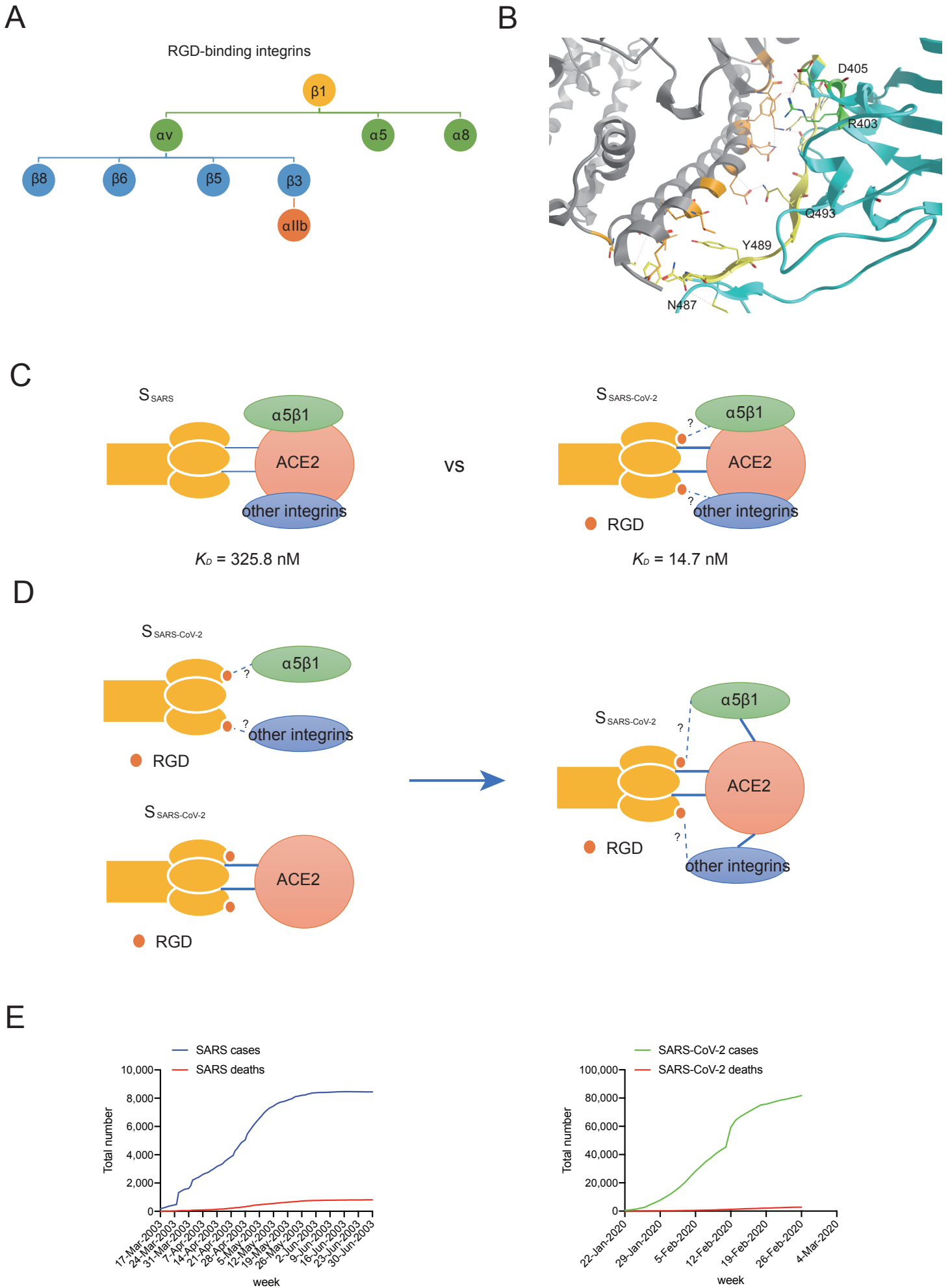
D

	RGD mutation			
SARS-CoV-2	394	NVYADSFVIRGDEVRQIAPGQTG	416	
RaTG13	394	NVYADSFVITGDEVRQIAPGQTG	416	
SARS	381	NVYADSFVVKGDDVRQIAPGQTG	403	
		*****: **:*****		

E

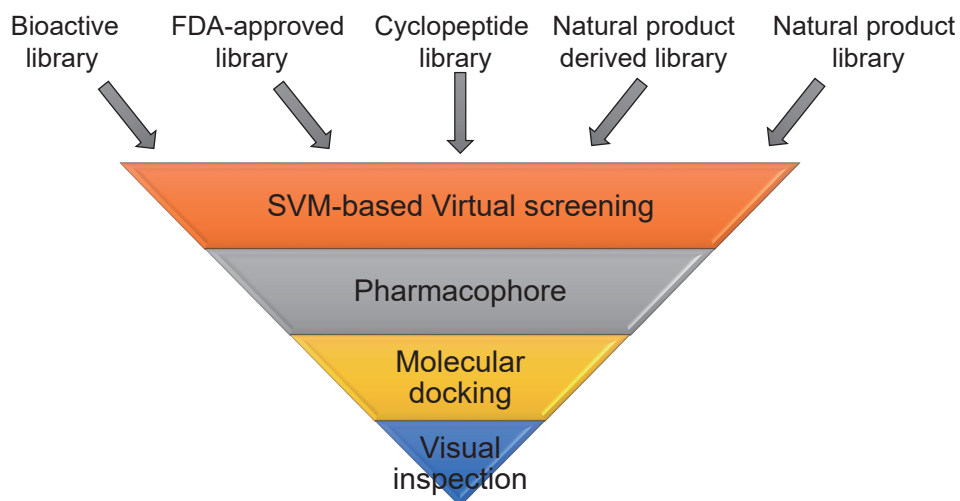
SARS-CoV-2	22,742	AATGTCATGCAGATTCATTTGTAATTAGAGGTGATGAAGTCAGACAAATCGCTCCAGGGCAAACCTGGA	22,810
RaTG13	22,724	AATGTTTATGCAGACTCATTTGTCATTACAGGTGATGAAGTCAGACAAATTCGCCAGGCAAACCTGGA	22,792
SARS	22,632	AATGTCATGCAGATTCATTTGTAAGTCAGGAGATGATGTAAGACAAATAAGCCAGGCAAACCTGCT	22,700

# Figure 2

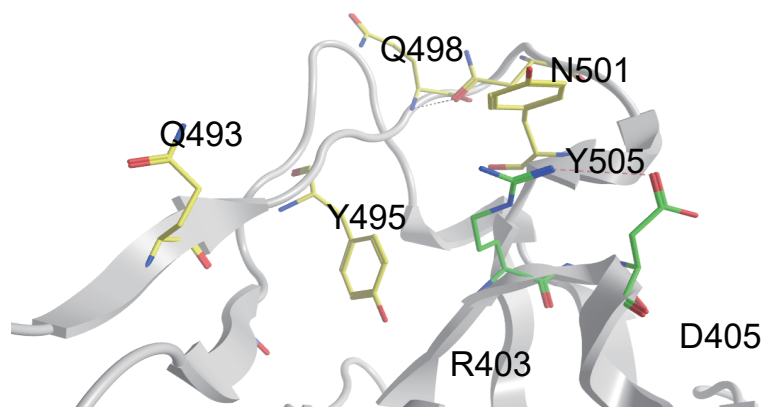


## Figure 3

A



B



## Figure 4

