Biophysical Journal
New and Notable

# Computational methods to predict the mutational landscape of the spike protein

Emanuele Paci[1,*] and James F. Ross[1]
[1]Astbury Centre and School of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom

"Spike mutations" are on everybody's mind. More than one year into the coronavirus disease 2019 (COVID-19) pandemic, the reason that these mutations are much talked about is the deep concern that, through them, the virus will be here to stay (1). Binding of the virus spike (S) protein to the human receptor protein angiotensin-converting enzyme II (hACE2) is the first necessary step for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to infect its host. Spontaneous mutations in the receptor binding domain (RBD) of the viral S protein observed in the emerging SARS-CoV-2 variants (https://www.gisaid.org) may reduce the effectiveness of vaccines and foreseeable therapeutic drugs targeting the spike protein. Mutations observed so far have not obviously affected clinical outcomes, although current variants of concern, including Alpha (UK, B.1.1.7), Beta (South Africa, B.1.351), Gamma (Brazil, P.1) and Delta (India, B.1.617.2), appear to be more transmissible. It is all but certain that SARS-CoV-2 will evolve to escape neutralization by the immune system and that the rate of concerning mutations will increase as a larger fraction of the population becomes (temporarily) immune. Unprecedented research efforts across the globe led to the sequencing of >1,100,000 genomes by April 2021 (https://www.gisaid.org). Sequencing results show that, although SARS-CoV-2 has a relatively low rate, the spike protein varies more than any other SARS-CoV-2 protein (https://nextstrain.org/groups/blab/sars-like-cov).

Knowledge of the "mutational landscape" of the hACE2-S protein complex may turn out to be crucial, for example, to design broad spectrum vaccines effective for predictable variants, to find drugs that prevent infection by binding to hACE2 more strongly than the viral S protein or, alternatively, to design soluble variants of hACE2 that bind the S protein more strongly than the wild-type, which could be engineered into "receptor traps" (2,3).

The simplest metric from a mutational landscape corresponds to the difference in free energy between bound and unbound states for each of the possible variants of all amino acids involved in binding. A small glimpse into the mutational landscape, namely a map of mutations in the RBD of the S protein that do not affect the binding affinity or increase it, is being provided by a large number of high-throughput experimental and computational studies produced since the sequence and structure of the spike protein have been in the public domain. Experimentally, deep mutational scanning makes it possible to characterize thousands of nondeleterious mutations (i.e., those that do not impact replicative fitness) from a single selection experiment.

Several high-resolution structures of SARS-CoV-2-related proteins have been solved. The availability of structures of the proteins when bound together in a complex is necessary to attempt to estimate the effect of any mutations, based on physical principles, through computational methods. Determining the variation in binding affinity upon mutation for a large number of mutants can only be done approximately because mathematical generalizations are required to complete the calculations in a timely manner.

In the current issue, a computational design study investigates mutations in the spike RBD that may increase the virus hACE2 affinity (4). Much of the work was performed before new variants of SARS-CoV-2 were observed. Since then, a large amount of research has been performed that confirms some of the findings in the work and expands the knowledge of the effect of mutations on the binding affinity of the S protein to hACE2.

In their article, Polydorides and Archontis (4) first compare the affinities of the SARS-CoV-2 and earlier SARS-CoV S protein RBDs for hACE2 via all-atom MD simulations and binding free energy calculations (Fig. 1). Affinities of the two spike proteins are similar, despite considerable differences in the sequence of the spike RBD, which

*Submitted March 25, 2021, and accepted for publication May 5, 2021.*

*Correspondence: e.paci@leeds.ac.uk

Editor: Alexandr Kornev.
https://doi.org/10.1016/j.bpj.2021.05.001

© 2021 Biophysical Society.

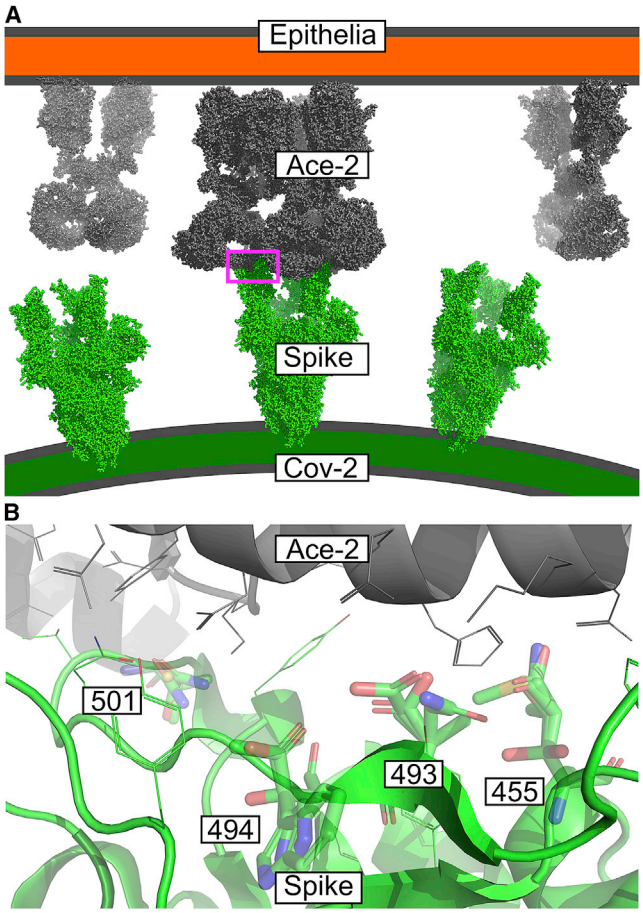


FIGURE 1 (*A*) Schematic of the binding interaction between SARS-CoV-2 and the lung epithelium, showing bound and unbound ACE2 receptors and spike proteins. The magenta box shows an individual spike to ACE2 receptor interaction, which is expanded in (*B*). (*B*) Selection of the best-predicted mutations from Fig. 3 b (in (4)) at residues 455, 493, 494, and 501 of the spike protein. To see this figure in color, go online.

suggests a relatively flat mutational landscape. Amino acid differences between SARS-CoV and SARS-CoV-2 at one set of positions in the RBD where a positive charge is inserted or a negative charge is eliminated improve interactions with negatively charged hACE2 residues, whereas another group of mutations have the opposite effect. Interestingly, the recent Alpha, Beta, Gamma, and Delta SARS-CoV-2 variants reintroduce positively charged residues at positions of the latter group. A charge mutation (N439K) occurs in the Alpha 20I/501Y.V1 (B.1.1.1.7) variant; the Beta 20H/501Y.V2 (B.1.351) (5,6) and Gamma 20J/501Y.V3 variants (7) contain another charge-changing mutation identified here (E484K). The mutation N439K might augment affinity by restoring a salt bridge with hACE2 E329 and has been shown to maintain the virus fitness and elicit immune escape (6). The mutation E484K may also enhance virus affinity for its human receptor (8). Of interest, since submitting their article, this trend of mutations, which increase the global positive charge of the spike protein, has continued in the more recent Delta variant (B.1.617.2), with a 19R insert and the substitutions L452R, T478K, D614G, P681R, and D950N.

The computational method adopted in this research employs a physics-based energy function and an efficient sampling method that populates sequences according to binding free energies. The design samples mutations at four positions (455, 493, 494, and 501) in the spike protein sequence: 18 chemical types in the first three positions and 14 types in the last position, corresponding to a total of $\sim 10^5$ possible sequences.

Accurate predictions of binding affinities and the effect of mutations rely inevitably on empirical methods with limited accuracy, particularly when dealing with a large number of mutants. For example, to increase efficiency, the protein backbone is kept fixed at the conformation in the x-ray crystal structure of the SARS-CoV-2 complex (9). Also, mutations to aromatic residues are not considered because of a limitation of the approach; thus, they do not assess the N501Y mutation that is contained in the recent SARS-CoV-2 variants (5–7,10). Yet, the design is reasonably successful in the light of most recent findings.

Methods to computationally design or redesign protein-protein interactions are becoming increasingly more reliable. Despite their limited accuracy, they also provide a structural rationale for the effect of the mutation on the binding affinity and information on structural constraints that affect the propensity of specific amino acids to mutate spontaneously. Particularly promising are approaches that combine the COVID-19 computational design with random mutagenesis and selection (2). The intense focus on COVID-19, specifically on the determinants of the interaction between the two proteins responsible for the onset of the infection, provides an opportunity to test different computational approaches to predict the effect of mutations and contribute to the design of vaccines and therapies.

## REFERENCES

1. Lavine, J. S., O. N. Bjornstad, and R. Antia. 2021. Immunological characteristics govern the transition of COVID-19 to endemicity. *Science*. 371:741–745.

2. Glasgow, A., J. Glasgow, …, J. A. Wells. 2020. Engineered ACE2 receptor traps

potently neutralize SARS-CoV-2. *Proc. Natl. Acad. Sci. USA*. 117:28046–28055.

3. Chan, K. K., D. Dorosky, …, E. Procko. 2020. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science*. 369:1261–1265.
4. Polydorides, D., and G. Archontis. 2021. Computational optimization of the SARS-CoV-2 receptor-binding-motif affinity for human ACE2. *Biophys. J.* 120:2859–2871.
5. Tegally, H., E. Wilkinson, …, T. de Oliveira. 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* https://doi.org/10.1101/2020.12.21.20248640.
6. Thomson, E. C., L. E. Rosen, …, G. Snell; ISARIC4C Investigators; COVID-19 Genomics UK (COG-UK) Consortium. 2021. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell.* 184:1171–1187.e20.
7. Faria, N. R., I. M. Claro, …, E. C. S. Xue. 2021. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.
8. Starr, T. N., A. J. Greaney, …, J. D. Bloom. 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell.* 182:1295–1310.e20.
9. Lan, J., J. Ge, …, X. Wang. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 581:215–220.
10. Kemp, S. A., B. Meng, …, R. K. Gupta; The COVID-19 UK (COG-UK) Consortium. 2021. Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion H69/V70. *bioRxiv* https://doi.org/10.1101/2020.12.14.422555.