THE JOURNAL OF PHYSICAL CHEMISTRY LETTERS
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

This article is made available via the ACS COVID-19 subset for unrestricted RESEARCH re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for the duration of the World Health Organization (WHO) declaration of COVID-19 as a global pandemic.

pubs.acs.org/JPCL

Letter

Mechanisms of SARS-CoV-2 Evolution Revealing Vaccine-Resistant Mutations in Europe and America

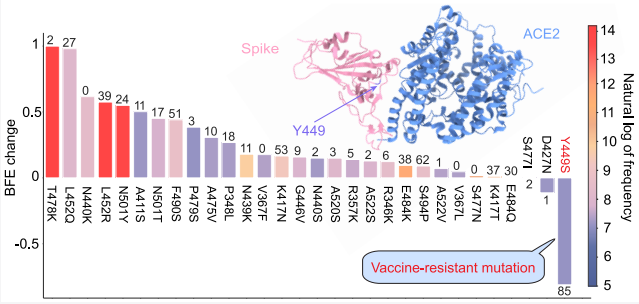
Rui Wang, Jiahui Chen, and Guo-Wei Wei*

Cite This: *J. Phys. Chem. Lett.* 2021, 12, 11850−11857

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The importance of understanding SARS-CoV-2 evolution cannot be overlooked. Recent studies confirm that natural selection is the dominating mechanism of SARS-CoV-2 evolution, which favors mutations that strengthen viral infectivity. Here, we demonstrate that vaccine-breakthrough or antibody-resistant mutations provide a new mechanism of viral evolution. Specifically, vaccine-resistant mutation Y449S in the spike (S) protein receptor-binding domain, which occurred in co-mutations Y449S and N501Y, has reduced infectivity compared to that of the original SARS-CoV-2 but can disrupt existing antibodies that neutralize the virus. By tracking the evolutionary trajectories of vaccine-resistant mutations in more than 2.2 million SARS-CoV-2 genomes, we reveal that the occurrence and frequency of vaccine-resistant mutations correlate strongly with the vaccination rates in Europe and America. We anticipate that as a complementary transmission pathway, vaccine-breakthrough or antibody-resistant mutations, like those in Omicron, will become a dominating mechanism of SARS-CoV-2 evolution when most of the world's population is either vaccinated or infected. Our study sheds light on SARS-CoV-2 evolution and transmission and enables the design of the next-generation mutation-proof vaccines and antibody drugs.

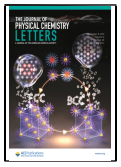


Started in late 2019, the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has had devastating impacts worldwide, plunging the world into an economic recession. Although several authorized vaccines have offered promise for controlling the disease in early 2021, the emergence of multiple variants of SARS-CoV-2 indicates that the combat with SARS-CoV-2 will be protracted. At this stage, almost all SARS-CoV-2 vaccines and monoclonal antibodies (mAbs) are targeted at the spike (S) protein,[1] while mutations on the S protein have been verified to compromise the efficacy of existing vaccines and mAbs.[2−4] Therefore, it is imperative to understand the mechanisms of viral mutations, especially on the S gene of SARS-CoV-2, which will promote the development of mutation-proof vaccines and mAbs.

The mechanism of mutagenesis is driven by various competitive processes,[5−9] which can be categorized into three different scales with many factors as illustrated in Figure 1a: (1) the molecular scale, (2) the organism scale, and (3) the population scale. From the molecular-scale perspective, the reading frame shifts, replication errors, transcription errors, translation errors, viral proofreading, and viral recombination are the main driven sources. Moreover, the host gene editing induced by the adaptive immune response[9] and the recombination between the host and virus are the key-driven factors at the organism level. Finally, the natural selection popularized by Charles Darwin is a critical population-level process, which favors mutations that have reproductive advantages for the virus to have adaptive traits in evolution. Such complicated mechanisms of viral mutagenesis make the comprehension of viral transmission and evolution a grand challenge.

Although there are 28,912 unique single mutations distributed widely on the whole SARS-CoV-2 genome, the mutations on the S gene stand out among all 29 genes on SARS-CoV-2 due to the mechanism of viral infection. With the assistance of host transmembrane protease, serine 2 (TMPRSS2), SARS-CoV-2 enters the host cell by interacting with its S protein and the host angiotensin-converting enzyme 2 (ACE2)[10] (see Figure 1b). Later, antibodies will be generated by the host immune system, aiming to eliminate the invading virus through direct neutralization or non-neutralizing binding,[11,12] which makes the S protein the main target for the current vaccines. Specifically, there is a short immunogenic fragment located on the S protein of SARS-CoV-2 that can facilitate the binding of SARS-CoV-2 S protein to ACE2, which is called the receptor-binding domain (RBD).[13] Studies have shown that the

Received: October 14, 2021
Accepted: December 2, 2021
Published: December 7, 2021



ACS Publications © 2021 American Chemical Society

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

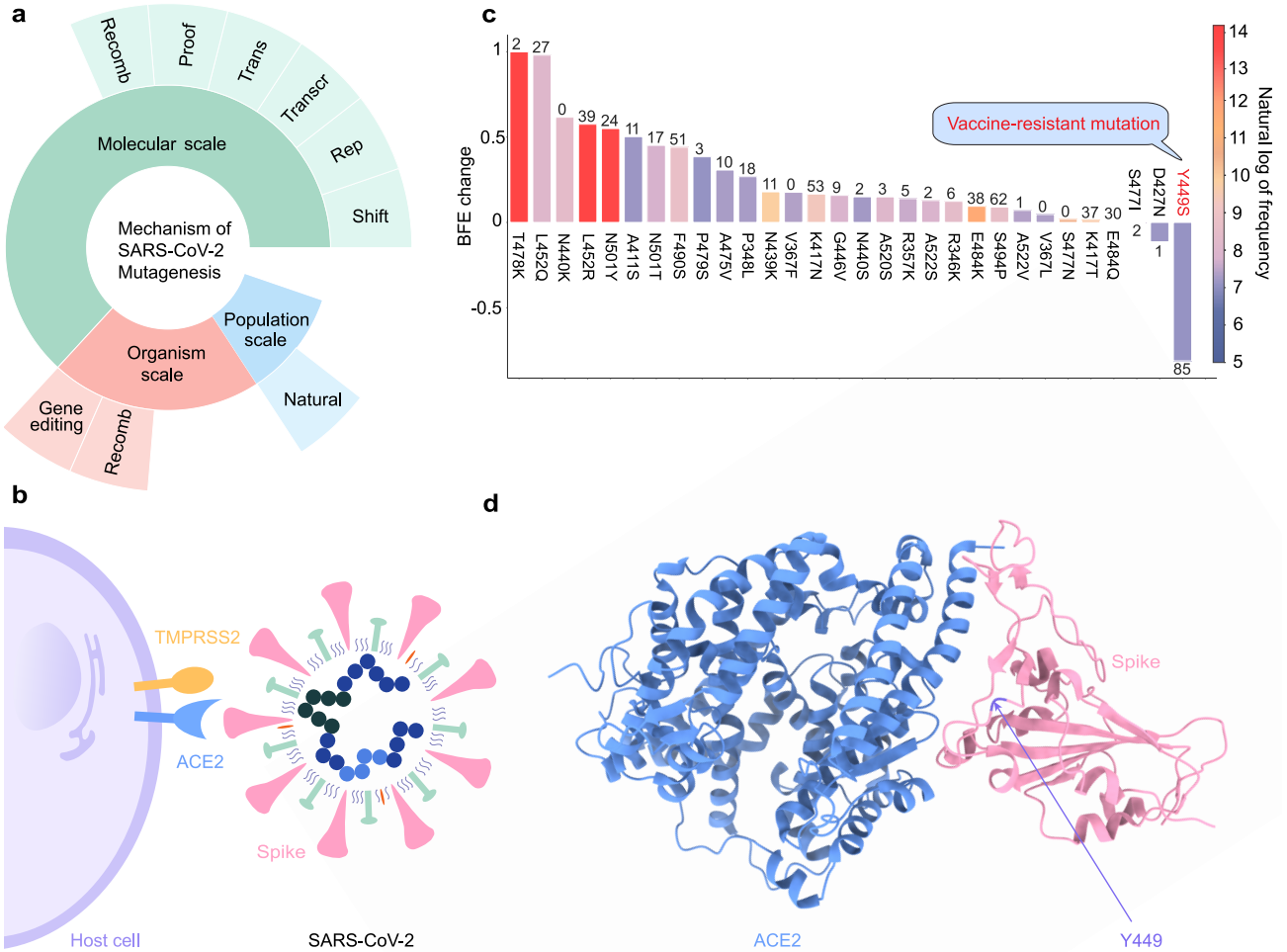


Figure 1. (a) Mechanism of mutagenesis. Nine mechanisms are grouped into three scales: (1) molecule-based mechanism (green), (2) organism-based mechanism (red), and (3) population-based mechanism (blue). The reading frame shifts (Shift), replication error (Rep), transcription error (Transcr), translation error (Trans), viral proofreading (Proof), and recombination (Recomb) are the six molecule-based mechanisms. Gene editing and host−virus recombination are the organism-based mechanism. In addition, the natural selection (Natural) is the population-based mechanism, which is the mainly driven source in the transmission of SARS-CoV-2. (b) Sketch of SARS-CoV-2 and its interaction with a host cell. (c) Illustration of 30 single-site RBD mutations with the top frequencies. The height of each bar shows the BFE change of each mutation; the color of each bar represents the natural log of the frequency of each mutation, and the number at the top of each bar means the AI-predicted number of antibody and RBD complexes that may be significantly disrupted by a single-site mutation. (d) Illustration of SARS-CoV-2 S protein with human ACE2. The blue chain represents the human ACE2; the pink chain represents the S protein, and the purple fragment on the S protein points out the two vaccine-resistant mutations Y449S and Y449H.

binding free energy (BFE) between the S RBD and the ACE2 is proportional to the infectivity.[10,14−17] Therefore, tracking and monitoring the RBD mutations and their corresponding BFE changes will expedite the understanding of the infectivity, transmission, and evolution of SARS-CoV-2, especially for the new SARS-CoV-2 variants, such as Alpha, Beta, Gamma, Delta, Lambda, etc.[18] Specifically, a positive BFE change between S and ACE2 induced by the mutation of a given variant indicates an infectivity-strengthened capacity, while a negative BFE change between S and ACE2 suggests an infectivity-weakened variant.

The current prevailing variants Alpha, Beta, Gamma, Delta, Kappa, Theta, Lambda, Mu, and Omicron carry at least one vital mutation at residues 452 and 501 on the S protein RBD. Notably, in early 2020, we successfully predicted that residues 452 and 501 "have high changes to mutate into significantly more infectious COVID-19 strains".[19] In the same work, we hypothesized that "natural selection favors those mutations that enhance the viral transmission" and provided the first evidence for infectivity-based natural selection. In other words, we revealed the mechanism of SARS-CoV-2 evolution and transmission based on very limited genome data in June 2020.[19] Additionally, we predicted three categories of RBD mutations: (1) most likely (1149 mutations), (2) likely (1912 mutations), and (3) unlikely (625 mutations).[19] To date, almost all of the RBD mutations we detected fall into our first category.[3,20] Moreover, all of the top 100 most observed RBD mutations have a BFE change greater than the average BFE changes of −0.28 kcal/mol (the average BFE changes for all RBD mutations[21]). It is an extremely low odd [i.e., $1/(1.27 \times 10^{30})$] for 100 RBD mutations to accidentally have BFE changes simultaneously above the average value, which confirms our hypothesis that the transmission and evolution of new SARS-CoV-2 variants are governed by infectivity-based natural selection, despite all other competing mechanisms.[19] Our predictions rely on algebraic topology[22−24]-assisted deep learning[19,25] but have been extensively validated.[3,4] However, infectivity is not the only transmission pathway that governs viral

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

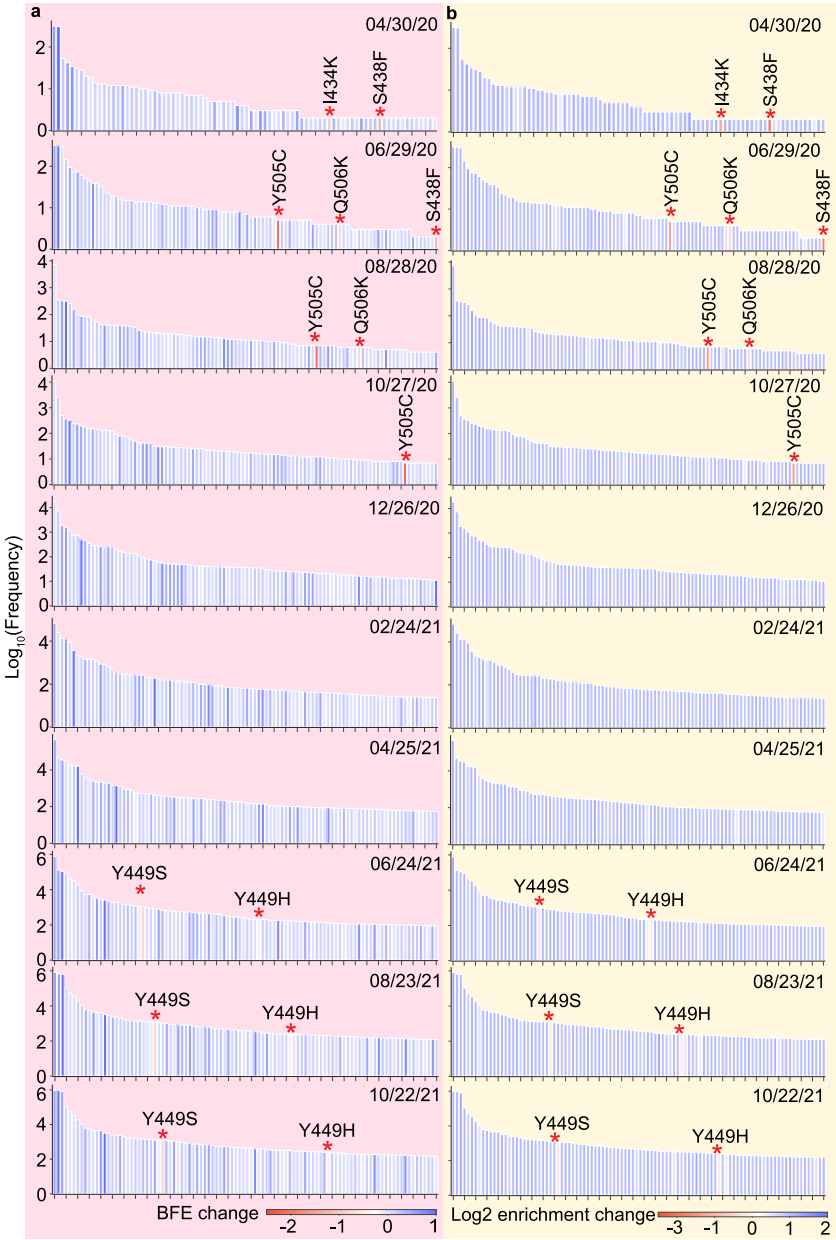


Figure 2. Most significant RBD mutations. (a) Time evolution of RBD mutations with its mutation-induced BFE changes per 60 days from April 30, 2020, to October 22, 2021. Here, only the top 100 most observed RBD mutations are displayed. Each bar represents a RBD single mutation. The height and color of each bar represent the log frequency and ACE-S BFE change induced by a given RBD mutation. The red star marks the vaccine-resistant mutations with significantly negative BFE changes. (b) Time evolution of RBD mutations with its experimental mutation-induced log2 enrichment ratio changes per 60 days from April 30, 2020, to October 22, 2021. The height and color of each bar represent the log frequency and enrichment ratio change induced by a given RBD mutation. The red star marks vaccine-resistant mutations with significantly negative BFE changes.

evolution. Vaccine-resistant mutations or, more precisely, antibody-resistant mutations that can disrupt the protection of antibodies have become a viable mechanism for new variants to transmit among the vaccinated population since the vaccine was put on the market. In early January 2021, we have predicted that RBD mutations W353R, I401N, Y449D, Y449S, P491R, P491L, Q493P, etc., will weaken the binding of most antibodies to the S protein.[3] Later, we provided a list of most likely vaccine escape RBD mutations with high frequency, including S494P, Q493L, K417N, F490S, F486L, R403K, E484K, L452R, K417T, F490L, E484Q, and A475S.[20] Moreover, we have pointed out that Y449S and Y449H are two vaccine-resistant mutations, and "Y449S, S494P, K417N, F490S, L452R, E484K, K417T, E484Q, L452Q, and N501Y" are the top 10 mutations that will disrupt most antibodies with high frequency.[21] As mentioned in ref 26, RBD mutations such as E484K/A, Y489H, Q493K, and N501Y found in late-stage evolved S variants "confer resistance to a common class of SARS-CoV-2 neutralizing antibodies", which suggests the viral evolution is also regulated by vaccine-resistant mutations. Interestingly, experimental results confirm that Y449, L455, F456, E484, F486, N487, Y489, Q493, S494, and Y505 are important for antibody binding, which means that mutations on these residues may enable the virus to escape antibodies.[27] Notably, the most common mode of binding between antibodies and S protein is through hydrophobic contacts, and Y449 is located at the receptor-binding motif with hydrophobic

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

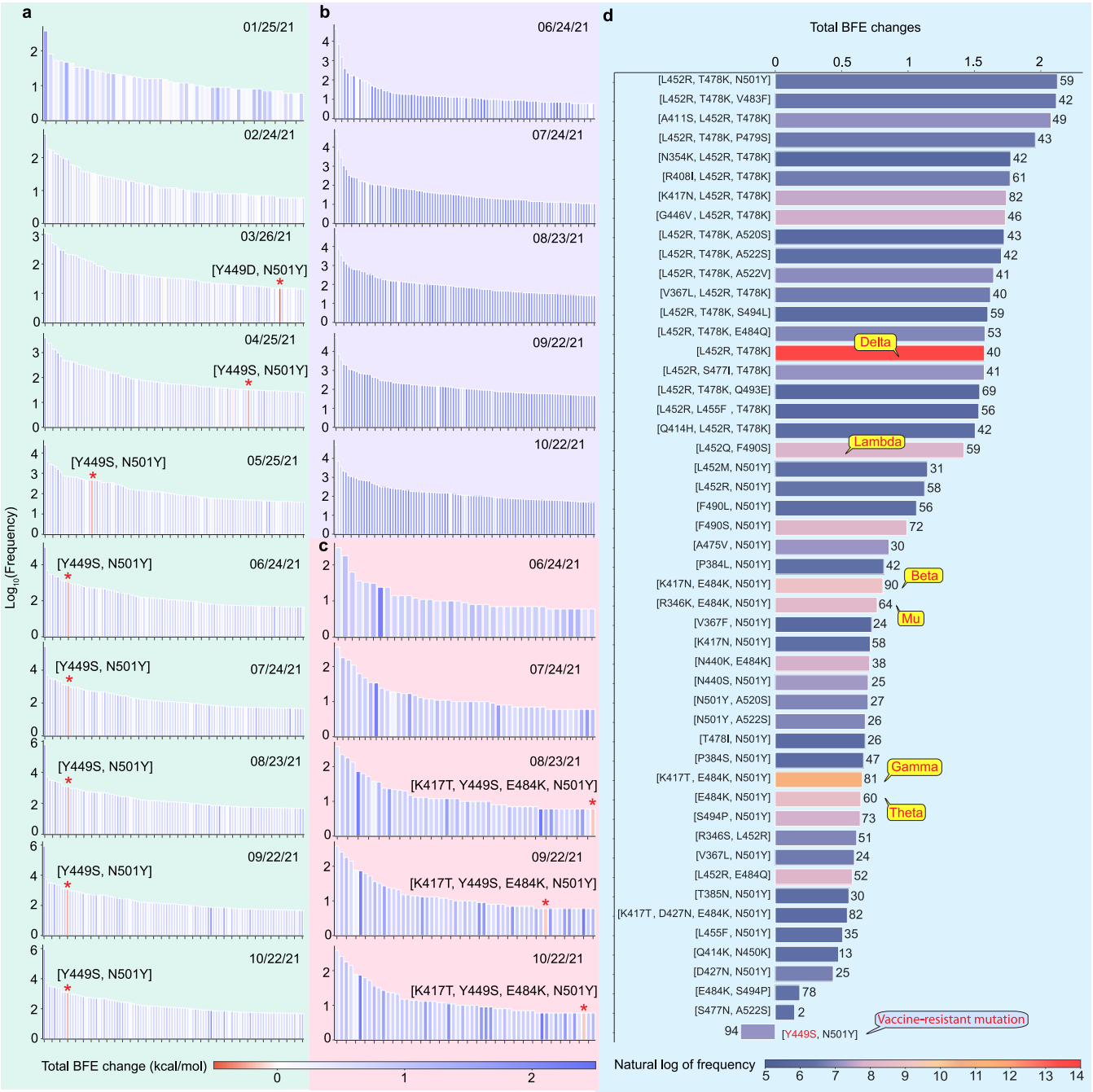


Figure 3. RBD co-mutation analysis. (a) Time evolutionary trajectory of two RBD co-mutations with its mutation-induced BFE changes per 30 days from January 25, 2021, to October 22, 2021. Each bar represents a pair of RBD co-mutations. The height and color of each bar represent the log frequency and ACE-S BFE change induced by a given RBD mutation. Red stars mark the two co-mutations with significantly negative BFE changes. (b) Time evolutionary trajectory of three RBD co-mutations with its mutation-induced BFE changes per 30 days from June 24, 2021, to October 22, 2021. Each bar represents a RBD co-mutation. The height and color of each bar represent the log frequency and ACE-S BFE change induced by a given RBD mutation. (c) Time evolutionary trajectory of four RBD co-mutations with its mutation-induced BFE changes per 30 days from June 24, 2021, to October 22, 2021. Each bar represents a RBD co-mutation. The height and color of each bar represent the log frequency and ACE-S BFE change induced by a given RBD mutation. (d) Illustration of the top 50 most observed RBD co-mutations. Here, the length of each bar represents the total ACE2-S BFE changes induced by a specific RBD co-mutation; the color of each bar represents the natural log frequency of each co-mutation, and the number at the side of each bar is the AI-predicted antibody disruption count.

side chains, indicating it is one of the vital residues for the binding between antibodies and S protein.[27,28]

The objective of this work is to analyze the evolution of the mechanisms of SARS-CoV-2 evolution, driven by complementary viral transmission pathways. We demonstrate how the interplay among molecular-scale, organism-scale, and population-scale mechanisms of SARS-CoV-2 mutations has affected the evolution of SARS-CoV-2. As a primary driven source of mutagenesis, the molecule-based mechanisms such as reading frame shifts, proofreading, etc., change the genetic information initially. Next, gene editing takes charge of the organism-based mechanism, suggesting the immune response of the host to the

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

virus.[9] Then, the population-level mechanism governs the transmission pathways of viral evolution. Two complementary pathways (infectivity and vaccine resistance) regulated by natural selection become the preponderance of evolution-driven force. The RBD mutations regulated by infectivity-based pathways exist in the prevailing variants, while the mutations regulated by the vaccine-resistant pathway start to emerge in countries with relatively high vaccination rates. In this work, 2,298,349 complete SARS-CoV-2 genomes isolated from patients are decoded by single-nucleotide polymorphism (SNP) calling, from where a total of 28,912 unique single mutations are detected. Among them, 774 RBD mutations were discovered by October 20, 2021 (detailed information can be found in section S6 of the Supporting Information). On the basis of our comprehensive topology-based artificial intelligence (AI) model for predicting RBD mutation-induced BFE changes of RBD and ACE2/antibody complexes,[3,19] the transmission trajectory of vaccine-resistant RBD mutations will be analyzed (detailed information about the methods and model can be found in sections S1 and S2 of the Supporting Information). Moreover, vaccine-resistant RBD mutation Y449S that has been found in more than 1000 isolates is discussed. Furthermore, the vaccination rates of 12 countries where Y449S is distributed are also analyzed, which provides a sound explanation of the relation between the emergence of vaccine-resistant mutations and the vaccination rate. Such an understanding of two complementary transmission pathways will shed light on the long-term efficacy of S-targeted antibody countermeasures and benefit the development of next-generation mutation-proof vaccines and mAbs.

Studying the mechanisms of SARS-CoV-2 mutagenesis is beneficial to the understanding of viral transmission and evolution. The main driving force of viral evolution is regulated by natural selection, which is employed by two complementary transmission pathways: (1) infectivity-based pathway and (2) vaccine-resistant pathway. We have discussed the infectivity-based pathways in refs 21 and 29. This section focuses on the vaccine-resistant pathway and its impact on the transmission and evolution of SARS-CoV-2. To understand the mechanisms of vaccine-resistant mutations, we first analyze 2,298,349 complete SARS-CoV-2 genomes, and a total of 28,912 unique single mutations are decoded. Among them, there are 774 non-degenerate RBD mutations. The infectivity of SARS-CoV-2 is proportional to the BFE between the S RBD and ACE2.[10,14−17] Therefore, the BFE change induced by a specific RBD mutation reveals whether the RBD mutation is an infectivity-strengthening mutation or an infectivity-weakening one. Similarly, the BFE change between the S RBD and antibody induced by a given mutation reveals whether this mutation will strengthen the binding between S and the antibody. To date, we have collected 130 antibody structures (see section S6 of the Supporting Information), which includes Food and Drug Administration (FDA)-approved mAbs from Eli Lilly and Regeneron. For a specific RBD mutation, its antibody disruption count shows the number of antibodies that have antibody-S BFE changes of less than −0.3 kcal/mol. The ACE2-S and antibody-S BFE changes induced by RBD mutations are predicted from our TopNetTree model,[19] which is available at TopNetmAb. All of the predicted BFE changes induced by RBD mutations can be found at Mutation Analyzer (https://weilab.math.msu.edu/MutationAnalyzer/). Figure 1c illustrates the top 30 most observed RBD mutations. The height and color of each bar represent the ACE2-S BFE changes and the frequency of each RBD mutation. The number at the top of each bar shows the antibody disruption count of each mutation. The detailed information can be viewed in section S4 of the Supporting Information. One can see that 27 mutations have positive ACE2-S BFE changes, suggesting they are regulated by the infectivity-based transmission pathway. However, three RBD mutations (S477I, D427N, and Y449S) have negative BFE changes. Notably, the Y449S mutation has a significantly negative BFE change (−0.8112 kcal/mol) and a large antibody disruption count (85), revealing an atypical mechanism of mutagenesis. Such a mutation with a significantly negative ACE2-S BFE change together with a high antibody disruption count is called a vaccine-resistant or antibody-resistant mutation. Figure 1d is the illustration of SARS-CoV-2 S protein (pink color) with human ACE2 (blue color), and the Y449 residue (purple color) is located on the random coil of the S protein. Among all of the vaccine-resistant mutations, the Y449S mutation has the highest frequency (1193). In addition, at residue 449, mutations Y449H, Y449N, and Y449D are all vaccine-resistant mutations that have been observed in more than 20 SARS-CoV-2 genome isolates.

To track the evolution trajectory of vaccine-resistant mutations, the BFE changes, log 2 enrichment ratios,[a] and log 10 frequencies of RBD mutations are analyzed from April 30, 2020, to October 22, 2021, per 60 days, as illustrated in Figure 2. Here, the top 100 most observed RBD mutations are displayed. In Figure 2a, red stars mark the vaccine-resistant mutations that have negative BFE changes. A few vaccine-resistant mutations (S438F, I434K, Y505C, and Q506K) were detected before November 2020 with relatively low frequencies. Notably, since December 2020, such vaccine-resistant mutations were no longer in the list of the top 100 most observed RBD mutations, suggesting that in this period, the evolution of SARS-CoV-2 is mainly regulated by natural selection through the infectivity-based transmission pathway. Moreover, in May 2021, two vaccine-resistant mutations (Y449S and Y449H) came back to the top 100 most observed RBD mutation list. In addition, the Y449S mutation has a relatively high frequency. This finding indicates that natural selection favors not only those mutations that enhance the transmission but also those mutations that can disrupt plenty of antibodies since SARS-CoV-2 vaccination was administered to provide protection among populations in early May. Similarly, the patterns can be found in Figure 2b, suggesting our AI-predicted BFE changes are highly consistent with the deep mutational enrichment ratio from experiments.[30]

The vaccine-resistant mutations are usually found along with other RBD mutations. Therefore, analyzing the time evolution of RBD co-mutations offers a better understanding of the mechanisms of vaccine-resistant mutations. Panels a−c of Figure 3 illustrate the time evolution of two, three, and four RBD co-mutations, respectively, with their corresponding BFE changes every 30 days. Here, each bar represents a RBD co-mutation, and the height and color of each bar represent the log 10 frequency and total BFE change induced by a given RBD co-mutation, respectively. Considering the number of co-mutations is quite low in the year 2020, the time range of analysis is set to [January 25, 2021, October 22, 2021] for the time evolution analysis of two co-mutations. For three and four co-mutations, their time ranges are set to [June 24, 2021, October 22, 2021]. In Figure 3a, a red star marks the two co-mutations with significantly negative BFE changes.

At the end of March 2021, vaccine-resistant mutation Y449D showed up with mutation N501Y in some genome isolates, resulting in a negative BFE change (−0.473 kcal/mol) and a

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

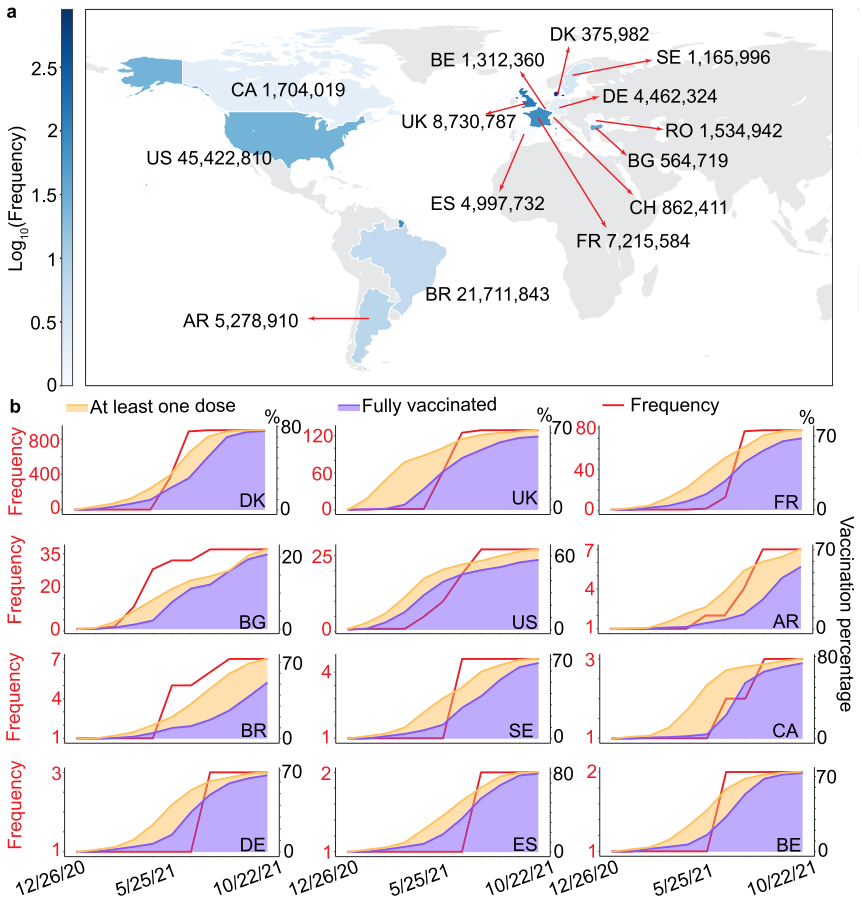


Figure 4. (a) Distribution of vaccine-resistant mutation Y449S. The color bar represents the log 10 frequency of Y449S in 14 countries: Denmark (DK), the United Kingdom (UK), France (FR), Bulgaria (BG), the United States (US), Argentina (AR), Brazil (BR), Sweden (SE), Canada (CA), Switzerland (CH), Germany (DE), Spain (ES), Romania (RO), and Belgium (BE). The number located at the side of the country shows the total number of positive SARS-CoV-2 cases by October 22. (b) Time evolution of the vaccination rate and the frequency of Y449S in top 12 countries from December 26, 2020, to October 22, 2021. The data are collected per 30 days. The red line shows the frequency of mutation Y449S. The orange and purple areas represent the rate of at least one dose and the rate of full vaccination, respectively, in each country.

high antibody disruption count (98) for a pair of RBD co-mutations (Y449D and N501Y). However, its global frequency is relatively low. Since late April 2021, vaccine-resistant mutation Y449S showed up with N501Y, making RBD co-mutations Y449S and N501Y some of the most prevament vaccine-resistant co-mutations. Figure 3d shows the top 50 most observed RBD co-mutations; the length and color of each bar represent the total BFE change and the natural log of frequency of an RBD co-mutation, respectively. The number at the side of each bar is the count of antibody disruption. Among the 50 most observed RBD co-mutations, the Y449S and N501Y co-mutation is the only co-mutation with a significantly negative BFE change and an extremely high antibody disruption count (94). Observing the evolution trajectory of Y449S and N501Y shows that the infectivity transmission pathway regulated by natural selection in the population level was the major evolution-driving force of SARS-CoV-2 mutagenesis before March 2021. Starting in January 2021, several vaccines were authorized for emergent use. Two months later, because many people had been protected by the vaccines, the mutations that disrupted the binding between the S and antibodies could be transmitted among vaccinated people, especially in countries with high vaccination rates. Such a vaccine-resistant pathway reduces the efficacy of vaccines and antibody therapies, indicating the combat with COVID-19 will be a prolonged battle.

Similar time evolution trajectories are drawn for three and four RBD co-mutations (see panels b and c, respectively, of Figure 3). There are no triple vaccine-resistant co-mutations at present, while a quadruple vaccine-resistant co-mutation (K417T, Y449S, E484K, and N501Y) appeared after late August 2021. Notably, Gamma variants, one of the variants of concern (VOC), carry three co-mutations (K417T, E484K, and N501Y) on the RBD, which indicates that four vaccine-resistant co-mutations (K417T, Y449S, E484K, and N501Y) may be a potential threat in the future.

Analysis of the vaccination trends and vaccine-resistant mutations leads to a fundamental understanding of the transmission and evolution of vaccine-resistant mutations. We investigate the distribution and time evolution of vaccine-resistant RBD mutation Y449 in 14 countries. As the most observed vaccine-resistant RBD mutation, Y449S has been detected in 14 countries, including Denmark (DK), the United Kingdom (UK), France (FR), Bulgaria (BG), the United States (US), Argentina (AR), Brazil (BR), Sweden (SE), Canada (CA), Switzerland (CH), Germany (DE), Spain (ES), Romania (RO), and Belgium (BE), as illustrated in Figure 4a. Here, 14 countries in which Y449S was found are colored blue. The darker the blue, the higher the frequency of Y449S. The number on the side of each country is the total positive cases up to October 22, 2021. Although DK has the smallest number of

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

positive cases among 14 countries, the frequency of the Y449S mutation is the highest. More than 800 patients carry vaccine-resistant mutation Y449S in DK. All of the Y449S-related cases are found in Europe and America, where the vaccination rates in those areas are relatively high. Figure 4b shows the time evolution of the vaccination ratio and the frequency of Y449S in the top 12 countries as mentioned above in 30-day periods. An illustration of CH and RO can be found in section S5 of the Supporting Information. The *x*-axis records the date, which ranges from December 26, 2020, to October 22, 2021. The left-hand side *y*-axis shows the frequency of Y499S (red lines), and the right-hand side *y*-axis shows the vaccination ratio. In addition, the orange region shows at least one dose ratio, while the purple region means the fully vaccinated ratio. One can see that the Y449S mutation was first found in BG and the US in December 2020. However, the frequency of the Y449S mutation in BG and the US is quite low before April 2021. After April 2021, the Y449S mutation quickly spread to 10 other countries. Among them, the total number of cases related to Y449S has a tendency to increase rapidly, especially in DK, the UK, and FR. Notably, all three countries have relatively high vaccination ratios (>70% up to late October 2021). It is worth mentioning that the frequency of the Y449S mutation is low in DE, ES, BE, etc., which is mainly due to the first Y449-related case in these countries being detected after June 2021. Since then, Delta variants dominated among the prevailing variants, which gave the Y449S mutation a limited chance to spread rapidly. Moreover, from Figure 4, one can see that the frequency of the Y449S mutation has a tendency to increase similar to that of the fully vaccinated ratio, suggesting that the vaccine-resistant mutations will gradually become one of the main evolution-driving forces of SARS-CoV-2, especially in those areas with high vaccination rates.

Due to the appearance of multiple mutations known to reduce the efficacy of antibody neutralization generated by vaccines, it is vital to better comprehend the mechanisms of SARS-CoV-2 mutagenesis, which will be of paramount importance to the understanding of SARS-CoV-2 transmission and evolution. The driving forces of mutagenesis can be categorized into three groups: (1) molecular-scale mechanisms, (2) organism-scale mechanisms, and (3) population-level mechanisms. As an initial driving force of mutagenesis, the genetic information is changed by reading frame shifts, viral proofreading, etc., which all belong to the group of molecular-scale mechanisms. Also, regulated by the host immune system, host gene editing and rarely occurring host−viral recombination are two organism-scale mechanisms. The molecular- and organism-scale mechanisms provide a large number of candidate mutations in the SARS-CoV-2 genome, while it is the population-scale mechanisms that determine what mutations become dominant.

Natural selection is a population-scale mechanism, which promotes the surge of the emerging SARS-CoV-2 variants by two complementary pathways: infectivity and vaccine resistance. The early stage of SARS-CoV-2 evolution was entirely dominated by infectivity-strengthening mutations. However, since late March 2021, once vaccines had provided protection to highly vaccinated populations, several vaccine-resistant mutations such as Y449S and Y449H have been observed relatively frequently. Considering that a good portion of the population is still not vaccinated, infectivity-strengthening mutations still dominate among the prevailing and future variants. However, vaccine-breakthrough or antibody-resistant mutations, like many RBD mutations associated with the Omicron variant, will become a major mechanism of transmission once most of the populations are carrying antibodies through either vaccination or infection. Our studies are valuable to the development of the next-generation vaccines and mAbs, which are greatly important for the long-term combat with SARS-CoV-2.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.1c03380.

Names of antibodies disrupted by mutations (S6.0.1: antibodies_disruptmutation.csv), PDB entries for all of the 130 SARS-CoV-2 antibodies (S6.0.2: antibodies.csv), all of the SNPs of RBD co-mutations by October 20, 2021 (S6.0.3: RBD_comutation_residue_10202021.csv), and all of the nondegenerate RBD co-mutations with their frequencies, antibody disruption counts, total BFE changes, and first detection dates and countries (S6.0.4: Track_Comutation_10202021.xlsx) (ZIP)

Data preprocessing and feature generation methods (section S1), machine learning methods (section S2), validations of our machine learning predictions with experimental data (section S3), the top 30 most observed S protein RBD mutations by October 20, 2021 (section S4), time evolution of the vaccination rate and the frequency of the Y449S mutation in CH and RO from December 26, 2020, to October 22, 2021 (section S5), and titles of tables of supplementary data (section S6) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Guo-Wei Wei − *Department of Mathematics, Department of Biochemistry and Molecular Biology, and Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-5781-2937; Email: wei@math.msu.edu

Authors

Rui Wang − *Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0002-7402-6372

Jiahui Chen − *Department of Mathematics, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-5416-6231

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpclett.1c03380

Notes

The authors declare no competing financial interest.

The world's SARS-CoV-2 SNP data are available at Mutation Tracker (https://users.math.msu.edu/users/weig/SARS-CoV-2_Mutation_Tracker.html). The most observed SARS-CoV-2 RBD mutations are available at Mutaton Analyzer (https://weilab.math.msu.edu/MutationAnalyzer/). The TopNetTree model is available at TopNetmAb (https://github.com/WeilabMSU/TopNetmAb).

https://doi.org/10.1021/acs.jpclett.1c03380
J. Phys. Chem. Lett. 2021, 12, 11850−11857

■ ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health Grant GM126189, National Science Foundation Grants DMS-2052983, DMS-1761320, and IIS-1900473, National Aeronautics and Space Administration Grant 80NSSC21M0023, Michigan Economic Development Corp., the Michigan State University Foundation, Bristol-Myers Squibb 65109, and Pfizer.

■ ADDITIONAL NOTE

[a]The log 2 enrichment ratio is collected from the experimental deep mutation enrichment data in ref 30.

■ REFERENCES

(1) Malik, J. A.; Mulla, A. H.; Farooqi, T.; Pottoo, F. H.; Anwar, S.; Rengasamy, K. R. Targets and strategies for vaccine development against SARS-CoV-2. *Biomed. Pharmacother.* **2021**, *137*, 111254.

(2) Annavajhala, M. K.; Mohri, H.; Zucker, J. E.; Sheng, Z.; Wang, P.; Gomez-Simmonds, A.; Ho, D. D.; Uhlemann, A.-C. A novel SARS-CoV-2 variant of concern, B. 1.526, identified in New York. *medRxiv* **2021**, DOI: 10.1101/2021.02.23.21252259.

(3) Chen, J.; Gao, K.; Wang, R.; Wei, G.-W. Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chemical Science* **2021**, *12*, 6929−6948.

(4) Chen, J.; Gao, K.; Wang, R.; Wei, G.-W. Revealing the threat of emerging SARS-CoV-2 mutations to antibody therapies. *J. Mol. Biol.* **2021**, *433*, 167155.

(5) Sanjuán, R.; Domingo-Calap, P. Mechanisms of viral mutation. *Cell. Mol. Life Sci.* **2016**, *73*, 4433−4448.

(6) Grubaugh, N. D.; Hanage, W. P.; Rasmussen, A. L. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell* **2020**, *182*, 794−795.

(7) Kucukkal, T. G.; Petukh, M.; Li, L.; Alexov, E. Structural and physico-chemical effects of disease and non-disease nsSNPs on proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 18−24.

(8) Yue, P.; Li, Z.; Moult, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* **2005**, *353*, 459−473.

(9) Wang, R.; Hozumi, Y.; Zheng, Y.-H.; Yin, C.; Wei, G.-W. Host immune response driving SARS-CoV-2 evolution. *Viruses* **2020**, *12*, 1095.

(10) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Krüger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N.-H.; Nitsche, A.; et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **2020**, *181*, 271−280.

(11) Chen, J.; Gao, K.; Wang, R.; Nguyen, D. D.; Wei, G.-W. Review of COVID-19 antibody therapies. *Annu. Rev. Biophys.* **2021**, *50*, 1−30.

(12) Chen, P.; Nirula, A.; Heller, B.; Gottlieb, R. L.; Boscia, J.; Morris, J.; Huhn, G.; Cardona, J.; Mocherla, B.; Stosor, V.; et al. SARS-CoV-2 neutralizing antibody LY-CoV555 in outpatients with COVID-19. *N. Engl. J. Med.* **2021**, *384*, 229−237.

(13) Tai, W.; He, L.; Zhang, X.; Pu, J.; Voronin, D.; Jiang, S.; Zhou, Y.; Du, L. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell. Mol. Immunol.* **2020**, *17*, 613−620.

(14) Li, W.; Shi, Z.; Yu, M.; Ren, W.; Smith, C.; Epstein, J. H.; Wang, H.; Crameri, G.; Hu, Z.; Zhang, H.; et al. Bats are natural reservoirs of SARS-like coronaviruses. *Science* **2005**, *310*, 676−679.

(15) Qu, X.-X.; Hao, P.; Song, X.-J.; Jiang, S.-M.; Liu, Y.-X.; Wang, P.-G.; Rao, X.; Song, H.-D.; Wang, S.-Y.; Zuo, Y.; et al. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *J. Biol. Chem.* **2005**, *280*, 29588−29595.

(16) Song, H.-D.; Tu, C.-C.; Zhang, G.-W.; Wang, S.-Y.; Zheng, K.; Lei, L.-C.; Chen, Q.-X.; Gao, Y.-W.; Zhou, H.-Q.; Xiang, H.; et al. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2430−2435.

(17) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* **2020**, *183*, 1735.

(18) Yin, C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics* **2020**, *112*, 3588−3596.

(19) Chen, J.; Wang, R.; Wang, M.; Wei, G.-W. Mutations strengthened SARS-CoV-2 infectivity. *J. Mol. Biol.* **2020**, *432*, 5212−5226.

(20) Wang, R.; Chen, J.; Gao, K.; Wei, G.-W. Vaccine-escape and fast-growing mutations in the United Kingdom, the United States, Singapore, Spain, India, and other COVID-19-devastated countries. *Genomics* **2021**, *113*, 2158−2170.

(21) Wang, R.; Chen, J.; Hozumi, Y.; Yin, C.; Wei, G.-W. Emerging vaccine-breakthrough SARS-CoV-2 variants. *arXiv* **2021**, 2103.08023.

(22) Carlsson, G. Topology and data. *Bulletin of the American Mathematical Society* **2009**, *46*, 255−308.

(23) Edelsbrunner, H.; Letscher, D.; Zomorodian, A. Topological persistence and simplification. *Proceedings of the 41st annual symposium on foundations of computer science*; 2000; pp 454−463.

(24) Xia, K.; Wei, G.-W. Persistent homology analysis of protein structure, flexibility, and folding. *International journal for numerical methods in biomedical engineering* **2014**, *30*, 814−844.

(25) Wang, M.; Cang, Z.; Wei, G.-W. A topology-based network tree for the prediction of protein−protein binding affinity changes following mutation. *Nature Machine Intelligence* **2020**, *2*, 116−123.

(26) Clark, S. A.; Clark, L. E.; Pan, J.; Coscia, A.; McKay, L. G.; Shankar, S.; Johnson, R. I.; Brusic, V.; Choudhary, M. C.; Regan, J.; et al. SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms. *Cell* **2021**, *184*, 2605−2617.

(27) Alenquer, M.; Ferreira, F.; Lousa, D.; Valério, M.; Medina-Lopes, M.; Bergman, M.-L.; Gonçalves, J.; Demengeot, J.; Leite, R. B.; Lilue, J.; et al. Signatures in SARS-CoV-2 spike protein conferring escape to neutralizing antibodies. *PLoS Pathog.* **2021**, *17*, e1009772.

(28) Ju, B.; Zhang, Q.; Ge, J.; Wang, R.; Sun, J.; Ge, X.; Yu, J.; Shan, S.; Zhou, B.; Song, S.; et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **2020**, *584*, 115−119.

(29) Chen, J.; Wang, R.; Wei, G.-W. Review of the mechanisms of SARS-CoV-2 evolution and transmission. *arXiv* **2021**, 2109.08148.

(30) Linsky, T. W.; Vergara, R.; Codina, N.; Nelson, J. W.; Walker, M. J.; Su, W.; Barnes, C. O.; Hsiang, T.-Y.; Esser-Nobis, K.; Yu, K.; et al. De novo design of potent and resilient hACE2 decoys to neutralize SARS-CoV-2. *Science* **2020**, *370*, 1208−1214.