bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Some mechanistic underpinnings of molecular adaptations of SARS-COV-2 spike protein by integrating candidate adaptive polymorphisms with protein dynamics

Nicholas J. Ose[1], Paul Campitelli[1], Tushar Modi[1], I. Can Kazan[1], Sudhir Kumar[2,3,4,*] and S. Banu Ozkan[1,*]

[1]Department of Physics and Center for Biological Physics, Arizona State University, Tempe, Arizona, United States of America

[2]Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania, United States of America

[3]Department of Biology, Temple University, Philadelphia, Pennsylvania, United States of America

[4]Center for Genomic Medicine Research, King Abdulaziz University, Jeddah, Saudi Arabia

* s.kumar@temple.edu (SK); * Banu.Ozkan@asu.edu (SBO)

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

We integrate evolutionary predictions based on the neutral theory of molecular evolution with protein dynamics to generate mechanistic insight into the molecular adaptations of the SARS-COV-2 Spike (S) protein. With this approach, we first identified Candidate Adaptive Polymorphisms (CAPs) of the SARS-CoV-2 Spike protein and assessed the impact of these CAPs through dynamics analysis. Not only have we found that CAPs frequently overlap with well-known functional sites, but also, using several different dynamics-based metrics, we reveal the critical allosteric interplay between SARS-CoV-2 CAPs and the S protein binding sites with the human ACE2 (hACE2) protein. CAPs interact far differently with the hACE2 binding site residues in the open conformation of S protein compared to the closed form. In particular, the CAP sites control the dynamics binding residues in the open state, suggesting an allosteric control of hACE2 binding. We also explored the characteristic mutations of different SARS-CoV-2 strains to find dynamic hallmarks and potential effects of future mutations. Our analyses reveal that Delta strain-specific variants have non-additive (i.e., epistatic) interactions with CAP sites, whereas the less pathogenic Omicron strains have mostly compensatory variants. Finally, our dynamics-based analysis suggests that the novel mutations observed in the Omicron strain epistatically interact with the CAP sites to help escape antibody binding.

Introduction

Since 2019, the evolution of SARS-CoV-2 in humans has been characterized by the spread of mutations, many notably found within the Spike (S) glycoprotein. The S protein is directly related to the human immune response to COVID-19 and, as such, has been one of the most studied and targeted proteins in the SARS-CoV-2 research (Shang et al. 2020; Harvey, Carabelli, Jackson, Gupta, Thomson, Harrison, Ludden, Reeve, Rambaut, Consortium, et al. 2021; Jackson et al. 2022; Carabelli et al. 2023; Markov et al. 2023). Subsequently, research into the biophysical properties and mutational patterns associated with S

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

protein evolution not only remains critical to understanding the pandemic but also emerges as a useful system to understand the mechanics of molecular adaptation within viruses.

For successful infection of a human host, the S glycoprotein of SARS-CoV-2 binds to the human ACE2 (hACE2) receptor through its receptor-binding domain (RBD). Evidence indicates that fine-tuning S protein interactions with hACE2 significantly affects viral reproduction (Rehman et al. 2020; Saputri et al. 2020; Rochman et al. 2021). Previous evolutionary studies show a complex network of interactions among mutated residues (Changeux and Edelstein 2005; Doshi et al. 2016; O'Rourke et al. 2016; Mishra and Jernigan 2018). Therefore, there has been a vast effort to uncover which mutations are important steps of adaptation for the S protein (Cagliani et al. 2020; Damas et al. 2020; Singh and Yi 2021; Kistler et al. 2022; Maher et al. 2022; Neher 2022). In particular, a significant aspect of many such studies was a focus on understanding adaptive mutations of SARS-CoV-2 that contributed to the leap to human hosts (Cagliani et al. 2020; Damas et al. 2020; Singh and Yi 2021; Starr, Zepeda, et al. 2022). This is because SARS-CoV-2 has continuously mutated since its early detection (Kistler et al. 2022), causing the emergence of CDC-designated "variants of concern" (VOCs) that an accelerated substitution rate may drive (Tay et al. 2022).

Predicting how new mutations impact the biophysical properties of the S protein remains a challenge, let alone explaining their complex interactions with one-another and how they might affect hACE2 binding because many factors affect hACE2 interactions. Binding affinity with hACE2 can be enhanced directly through stronger receptor interactions or mediated through changes in RBD opening (Teruel et al. 2021; Zhang et al. 2021; Díaz-Salinas et al. 2022). The RBD exhibits both 'closed' and 'open' conformational states. In the closed state, the RBD is shielded from receptor binding. In the open state, the RBD is accessible for hACE2 binding (Kirchdoerfer et al. 2016; Gur et al. 2020; Henderson et al. 2020; Hoffmann et al. 2020). While some mutations may affect the transition between these states (Henderson et al. 2020; Yurkovetskiy et al. 2020; Gobeil, Janowska, McDowell, Mansouri, Parks, Manne, et al. 2021; Sztain et al. 2021; Zhang et al. 2021; Shoemark et al. 2022), the other mutations may allosterically regulate RBD

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

openings through Furin cleavage site (residue ID range: 681-695) interactions to regulate hACE2 binding (Deng et al. 2021; Gobeil, Janowska, McDowell, Mansouri, Parks, Stalls, et al. 2021; Khan et al. 2021; Laiton-Donato et al. 2021).

Moreover, as new mutations accumulate, culminating in the emergence of a new VOC, these mutations must occur on varied sequence backgrounds containing neutral, nearly-neutral, and adaptive mutations. While many studies have explored the impacts of individual mutations, VOCs result in a substantial difference in protein function compared to their individual effects (Moulana et al. 2022a; Starr, Greaney, Hannon, et al. 2022; Moulana et al. 2023; Witte et al. 2023). Here we integrate an evolutionary approach with protein dynamics analysis to address the fundamental mechanisms of mutations dictating the VOCs and the impact of their epistatic interaction on the function of the S protein. Many earlier studies have combined phylogeny and evolutionary theory to identify adaptive mutations as well as to see how the viral sequence has changed over time (Frost et al. 2018; Boni et al. 2020; Cagliani et al. 2020; Damas et al. 2020; Tang et al. 2020). Similarly, we first use a well-established Evolutionary Probability (EP) approach (Liu et al. 2016) that utilizes phylogenetic trees in combination with the neutral theory of molecular evolution to determine Candidate Adaptive Polymorphisms (CAPs) determined using the early Wuhan sequence as a variant. CAPs are substitutions in SARS-CoV-2 that are rarely observed in other closely related sequences (Figure 1A), which implies a degree of functional importance and makes them candidates for adaptation (Liu et al. 2016). In support of this method, we find an overlap between our list of sites containing CAPs and putative adaptive sites identified by others (Cagliani et al. 2020; Singh and Yi 2021; Starr, Zepeda, et al. 2022). Second, we use a suite of computational tools to analyze how CAPs that arose in the early and late phases of the COVID-19 pandemic modulate the dynamics of the S protein. We also explore the complex interactions between these sets of CAPs to gain mechanistic insight into the behavior of molecular adaptation involving the S protein. In particular, we focused on how mutations modulate protein dynamics, as we and others have previously found that rather than changing a protein's

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

structure solely, mutations modulate conformational dynamics leading to changes in biophysical properties such as stability, flexibility, and allosteric dynamic coupling, any of which may affect protein binding (Swint-Kruse et al. 1998; Keskin et al. 2000; Bhabha et al. 2013; Nussinov, R., Tsai, C.-J 2013; Campbell, E. et al. 2016; Ma and Nussinov 2016: 201; Saavedra et al. 2018; Kuzmanic et al. 2020).

With this evolutionary-dynamics unified approach, we aim to answer the following questions about VOCs: Are all the mutations in VOCs adaptive in nature? Are they coupled to one another or provide some measure of biophysical, dynamical, or mechanical compensation? While many of these mutations are found within the RBD domain, numerous others are located distal to this region; hence, we aim to investigate the functional roles of distal mutations, particularly from a protein dynamics perspective. Our integrated analysis revealed that protein dynamics play a significant role in the evolution of the S protein. The flexibility of sites withing the S protein shows a strong, direct correlation with substitution rate, and newly evolving CAPS are mostly compensatory (i.e., additive) mutations that modulate the dynamics of the hACE2 binding site. Yet other CAPs, 346R, 486F, and 498Q, show highly epistatic (i.e., non-additive) modulation of the hACE2 binding site and provide immune escape benefits.

Results and Discussions

Candidate Adaptive mutations in the Spike protein are recognized via Evolutionary Probabilities.

SARS-CoV-2 is part of a family of coronaviruses, many of which infect mainly animals and are less capable of infecting humans (Dicken et al. 2021). Therefore, to identify the most likely mutations responsible for the infection of human hosts (i.e., putative adaptive mutations for humans), we estimated the (neutral) evolutionary probability (EP) scores of mutations found within the S protein (Liu et al. 2016). EP scores of the amino acid variants of S protein are obtained by constructing a maximum likelihood phylogenetic tree (Goldman 1990) containing 19 orthologous coronavirus sequences, which were selected based on the

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

amount of divergence over evolutionary history to ensure that each amino acid position had ample time to experience purifying selection (Patel et al. 2018) (Figure 1A). The likelihood of finding a particular amino acid in the sequence is then determined using a Bayesian framework, with calculations carried out by MEGA X software (Kumar et al. 2018). As apparent in the name, EP scores obtained for the amino acids in the sequence provide information regarding the likelihood of finding them at their position, given the history of the sequence. Amino acid residues receiving low EP scores (<0.05) at a position are less likely to be found in a given position within the sequence because they are non-neutral. Generally, positions with low EP amino acids are far less common than those containing mutations with high EP, a trend also realized in the CoV-2 S protein (Figure 1B).

Of particular interest is an observed evolutionary change where an amino acid with high EP is replaced by an amino acid residue with low EP. While amino acids with low EP should be harmful or deleterious to viral fitness due to functional disruption or change, fixation of a low EP amino acid at a position suggests an underlying mechanism for natural selection to operate. These fixed, low EP mutations are called candidate adaptive polymorphisms (CAPs) as they are predicted to alter protein function, and adaptive pressures may drive their prevalence (Patel et al. 2018). Indeed, there is an overlap between these CAPs and the mutations suggested by other methods to be adaptive for the S protein (Cagliani et al. 2020; Singh and Yi 2021; Starr, Zepeda, et al. 2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

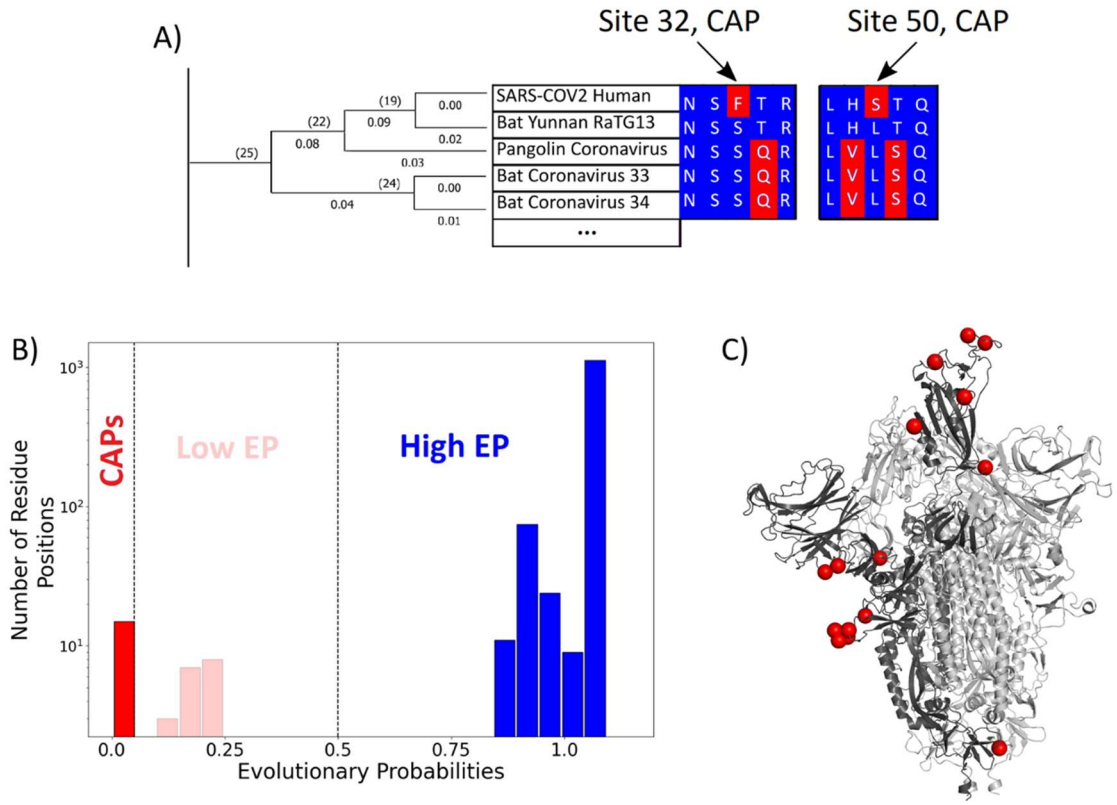


Figure 1. (A) The evolutionary probabilities (EP) of each amino acid in the S protein sequence are calculated by taking the multiple sequence alignment of the S proteins through their evolutionary tree and using Bayesian inferences to determine the likelihood of finding a particular residue at a particular location within a given sequence. Simply, if the residue is found at a location 'x' in closely related sequences, it will have a higher EP at location 'x' in the target sequence. Residues with an EP <0.05 in the target sequence are CAPs (Red). (B) The distribution of EP scores of the wild-type residues in the S protein. Here, lower EP scores are shown in red, and higher EP scores in blue. While the vast majority of the wild-type (reference) protein consists of high EP residues, a few residues have low EP. (C) The CAPs are also highlighted as red spheres in the open configuration of the S protein, with the open chain in a darker shade. We observe that a majority of the CAP positions reside at the receptor binding domain (RBD) and the Furin cleavage site (676-689; (Wrobel et al. 2020)) shown as transparent light gray spheres.

Interestingly, most of the CAP residues are in functionally critical sites, including the receptor binding domain (RBD) and the Furin cleavage site (Figure 1C). As mentioned earlier, the RBD plays a key role in initiating the infection of a healthy cell by binding it with the host organism's ACE2 protein. Before ACE2

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

binding, one chain of the homotrimer comprising the S protein must open to expose the RBD (Kirchdoerfer et al. 2016; Henderson et al. 2020; Hoffmann et al. 2020; Sztain et al. 2021). The Furin cleavage site plays a key role in the opening process as it aids in the cleavage of the S protein into two domains: S1 and S2 (Wrobel et al. 2020: 13). Similar cleavage sites have been found in related coronaviruses, including HKU1 and Middle East respiratory syndrome coronavirus (MERS-CoV), which infect humans (Chan et al. 2008: 1; Millet and Whittaker 2014; Millet and Whittaker 2015), and the acquisition of similar cleavage sites is associated with increased pathogenicity in other viruses such as the influenza virus (Steinhauer 1999). Interestingly, however, CAPs do not display such an overwhelming tendency to occur at well-known critical sites within human proteins studied with similar methods (Ose, Campitelli, et al. 2022), yet mutations at those sites are associated with disease, indicating their critical role in inducing functional change. Therefore, the identified CAPs in the S protein, which are signs of recent evolution, can provide mechanistic insights regarding the molecular adaptation of the virus. In particular, we aimed to analyze how these CAP positions in the S protein modulate the interaction with hACE2 using our protein dynamics-based analysis (Gerek and Ozkan 2011; Nevin Gerek, Z., Kumar, S., Banu Ozkan, S. 2013; Larrimore et al. 2017; Kumar, A., Glembo, T.J., Ozkan, S.B. 2015b).

Asymmetry in communications among the network of interactions in Spike describes how CAPs regulate the dynamics of the Spike protein.

A mutation at a given amino acid position inevitably not only alters local interactions, but this change cascades through the residue-residue interaction network, which gives rise to a variation in native ensemble dynamics to modulate function (Dror et al. 2012; Labbadia and Morimoto 2015; Sekhar and Kay 2019; Campitelli et al.). Thus, we analyze the internal dynamics of the system to understand the functional role of CAPs in S proteins. This analysis allows us to gain a mechanistic understanding of the relationship

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between CAP mutations and biophysical outcomes (Teruel et al. 2021). First, we implement the Dynamic Coupling Index (DCI) approach to study long-distance coupling between the CAPs and the hACE2 binding sites emerging from the 3D network of interactions across the S protein system. The DCI parameter combines Perturbation Response Scanning and Linear Response Theory to capture the strength of a displacement response for position *i* upon perturbation of position *j*, relative to the average fluctuation response of position *i* to all other positions in the protein. It represents the strength of dynamic coupling between positions *i* and *j* upon perturbation to *j.*

Further, asymmetry can be captured in the DCI values, as dynamic coupling is not necessarily due to an anisotropic network. That is, each amino acid has a set of positions to which it is highly coupled, and this anisotropy in connections gives rise to unique differences in coupling between a given *i, j* pair of amino acids which do not have direct interactions. By calculating the coupling of the hACE2 binding interface in the RBD with respect to the CAP residue positions and vice versa, we can generate DCI_{asym} (Figure 2A) as the difference between the normalized displacement response of position *j* upon a perturbation to position *i* (DCI_{ij}) and the normalized displacement response of position *i* upon a perturbation to position *j* (DCI_{ji}) (See Methods). If the DCI_{asym} values significantly differ from zero, it shows asymmetry in coupling and presents a cause-effect relationship between the *i, j* pair in terms of force/signal propagation. This metric has been used previously in a variety of systems to analyze the unique behavior of positions within a protein and a given position's propensity to effect biophysical changes upon mutation, particularly at long distances (Modi and Ozkan 2018; Campitelli and Ozkan 2020; Kolbaba-Kartchner et al. 2021; Ose, Butler, et al. 2022; Kazan et al. 2023; Campitelli et al. 2020b).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

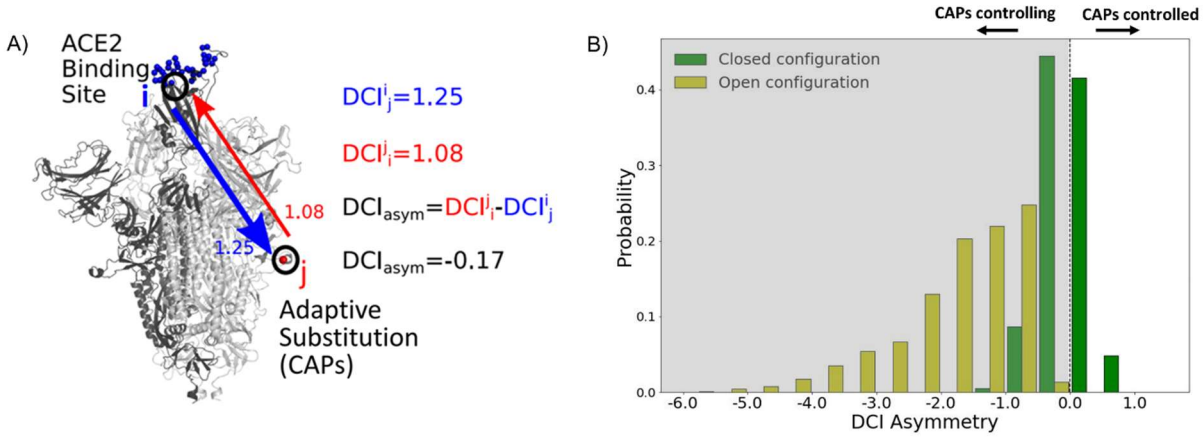


Figure 2. (A) Schematic representation of DCI asymmetry. (B) DCI asymmetry of CAP residue positions with the binding interface of RBD in the open chain. Residues in the closed chains with a low EP amino acid in the reference sequence dominate the binding site interface of RBD in the open chain. There is a significant difference between the asymmetry profiles of the closed (M = -0.06, SD = 0.33) and open (M = -1.68, SD = 0.89) conformations ($p < .001$).

Recent work from our group has shown an enhancement in cross-chain communication within the main protease of SARS COV-2 compared to SARS COV-1 (Campitelli et al. 2022). Furthermore, previous studies have shown that allosteric inter-chain communication is important to S protein function (Zhou et al. 2020; Spinello et al. 2021; Tan et al. 2022; Xue et al. 2022). In support of these findings, we observe through DCI_{asym} that when the S protein is in its pre-fusion conformation with one chain open, the CAPs in the closed chains have negative coupling asymmetry with respect to the hACE2 binding site interface in the RBD-open chain. This indicates an allosteric control where the hACE2 binding site is dominated by the dynamics of the CAPs in closed chains (Figure 2B, yellow bars). As this open-state RBD is critical for the viral infection of host cells (Kirchdoerfer et al. 2016), our results suggest that this type of closed-to-open cross-chain interaction is important for viral proliferation. Our prior studies on DCI_{asym} show a similar trend in Lactose Inhibitor (LacI), a protein with a functional role in gene expression through binding DNA. The allosteric mutations (i.e., mutations on the sites that are far from the DNA binding sites) that alter DNA

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

binding affinity not only exhibited unique asymmetry profiles with the DNA binding sites of LacI, but also regulated the dynamics of these binding sites (Campitelli et al. 2020b).

Similarly, it is possible that mutations to such residue positions within the S protein can be used to regulate the dynamics of the hACE2 bindings sites of the open RBD state. We, therefore, propose that the residue positions with CAP substitutions hold the potential for mutations in the spike sequence which can alter the opening and closing dynamics of the RBD domain. This hypothesis is further supported by many mutations already observed at these residue positions which alter the infection rate (Brister et al. 2015). Interestingly, residues responsible for extremely low asymmetry values (< -4) lie overwhelmingly in the region 476–486. These same residues were suggested to stabilize S protein dynamics and prime it for host Furin proteolysis (Raghuvamsi et al. 2021).

Moreover, as a control, we performed the same analysis on the S protein with the RBD domains of all chains in the closed configuration. In this case, we observed that the DCI_{asym} of the CAPs residue positions with respect to the hACE2 interface in the other chains yields a largely symmetric distribution about 0 (Figure 2B, green bars). This verifies that the asymmetry in the coupling of CAPs with the exposed binding site interface in pre-fusion configuration results from one of the RBDs opening up and further suggests the allosteric role played by CAPs in locking the S protein in the RBD open state.

Dynamic analysis shows that rigid sites tend to be more highly conserved than flexible sites.

CAPs represent important S protein amino acid changes between related coronaviruses across multiple species and the Wuhan-Hu-1 reference sequence (MN908947). Since SARS-CoV-2 first spread to humans, it has continued to mutate and evolve rapidly, particularly regarding the S protein (Amicone et al. 2022; Liu et al. 2022; Tay et al. 2022). Like the mutations leading to the Wuhan strain caused an increase in binding affinity to hACE2, continued evolution in human hosts has resulted in further altered binding

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

affinities as well as different phenotypic outcomes for those infected (Ali et al. 2021; Barton et al. 2021; Ozono et al. 2021).

We explore whether protein dynamics has played a role in the selection of mutational sites during the evolution of the S protein since 2019. Our previous work has indicated that rate of evolution per positional site exhibits a positive correlation with positional flexibility; generally, positions that exhibit higher flexibility are also sites that experience a higher number of amino acid substitutions (Liu and Bahar 2012; Maguid et al. 2008; Maguid et al. 2006; Mikulska-Ruminska et al. 2019; Nevin Gerek, Z., Kumar, S., Banu Ozkan, S. 2013). To confirm these findings for the evolution of the S protein using the sequenced variants of infected humans, we analyze the site-specific amino acid flexibility using the Dynamic Flexibility Index (DFI). Using the same mathematical foundation as DCI, DFI evaluates each position's displacement response to random force perturbations at other locations in the protein (Gerek and Ozkan 2011; Nevin Gerek, Z., Kumar, S., Banu Ozkan, S. 2013), and it can be considered a measure of a given position's ability to explore its local conformational space. We found that the covid-19 S protein shows the expected high correlation between the occurrence of mutations and site flexibility (Figure 3) when we compare %DFI (DFI ranked by percentile) to the average number of variants per position found within a given %DFI bin. Previous studies have indicated that rigid residues are critical for functional dynamics, thus more likely to impact function if mutated and, generally, can lead to a loss of function and thus more conserved (Kim, H. et al. 2015; Butler et al. 2018; Modi, Risso, et al. 2021; Modi, Campitelli, et al. 2021; Kazan et al. 2022; Ose, Butler, et al. 2022; Stevens et al. 2022; Campitelli et al. 2020a; Kumar, A., Glembo, T.J., Ozkan, S.B. 2015b). This analysis also agrees with these previous studies and highlights the power of negative selection, in line with the neutral theory of molecular evolution, stating that deleterious mutations (i.e., those on the rigid positions) should be eliminated and therefore not observed (Kimura, Motoo 1983).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

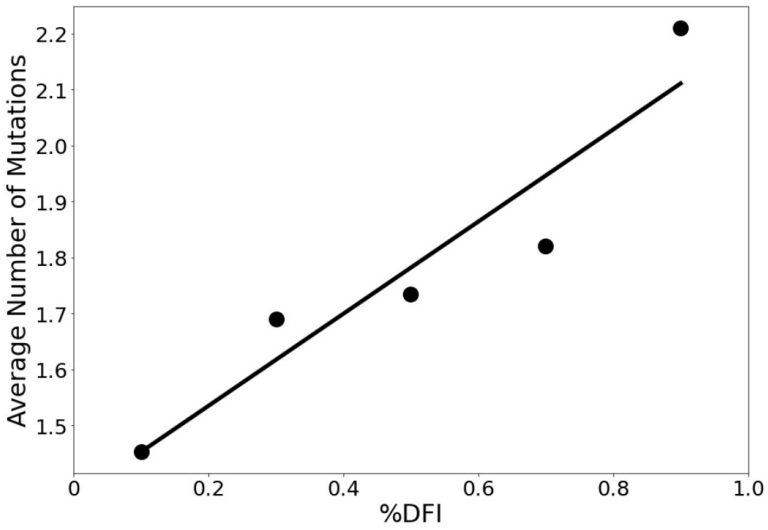


Figure 3. The average number of variants observed among residues of different flexibility. Residues were sorted into one of five bins based on flexibility. After that, the average number of variants for residues within that bin was calculated. Here, the number of variants is defined as the number of different amino acid varieties found at that site. Mutational data was calculated across approximately 24,000 SARS-CoV-2 S protein sequences from the NCBI Datasets Project (Brister et al. 2015). Residue flexibility, as reported here via %DFI, was computed using structure id 6vsb from the Protein DataBank (Berman, H.M. et al. 2000). More rigid residues tend to have fewer variants. (r = 0.94).

Continued mutations within human hosts have resulted in a multitude of variants. Indeed, by fitting various molecular clock models to genome sequence data, VOC emergence is punctuated by an episodic period of rapid evolution, with a substitution rate of up to 4-fold greater than the background substitution rate (Kumar et al. 2021; Tay et al. 2022). With such an aggressive evolutionary rate, we are finding VOCs to consist of a number of different characteristic mutations, almost all of which are CAPs.

To explore the dynamic effects of the evolution of the Spike in humans, we examine asymmetry with these new potentially adaptive sites, namely the low EP (CAP) characteristic mutation sites observed in the Delta variant, the widely dominant variant from December 2021 to January 2022 (Thye et al. 2021), and the Omicron variant, a highly transmissible variant whose lineages have remained dominant since January 2022) (Kim et al. 2021) (Figure 4). This analysis revealed a mechanism similar to that for the CAPs in the

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reference protein (Figure 2), as the open-chain binding interface is also allosterically controlled by these potentially new adaptive sites. Regarding this, we see that the asymmetry is much more pronounced in observed mutations of Omicron variants suggesting that these new mutations have a stronger power in controlling the dynamics of open chain hACE2 binding interface compared to those observed in Delta variants. We can speculate that the difference in virulence and infection rates between Omicron and Delta (Earnest et al. 2022; Bager et al. 2021; Sheikh et al. 2021; Twohig et al. 2022;(Houhamdi et al. 2022; Menni et al. 2022) might be due to these specific CAPs within each variant and their differences in allosterically controlling the dynamics of open RBD binding sites as observed in the DCI_{asym} analysis.

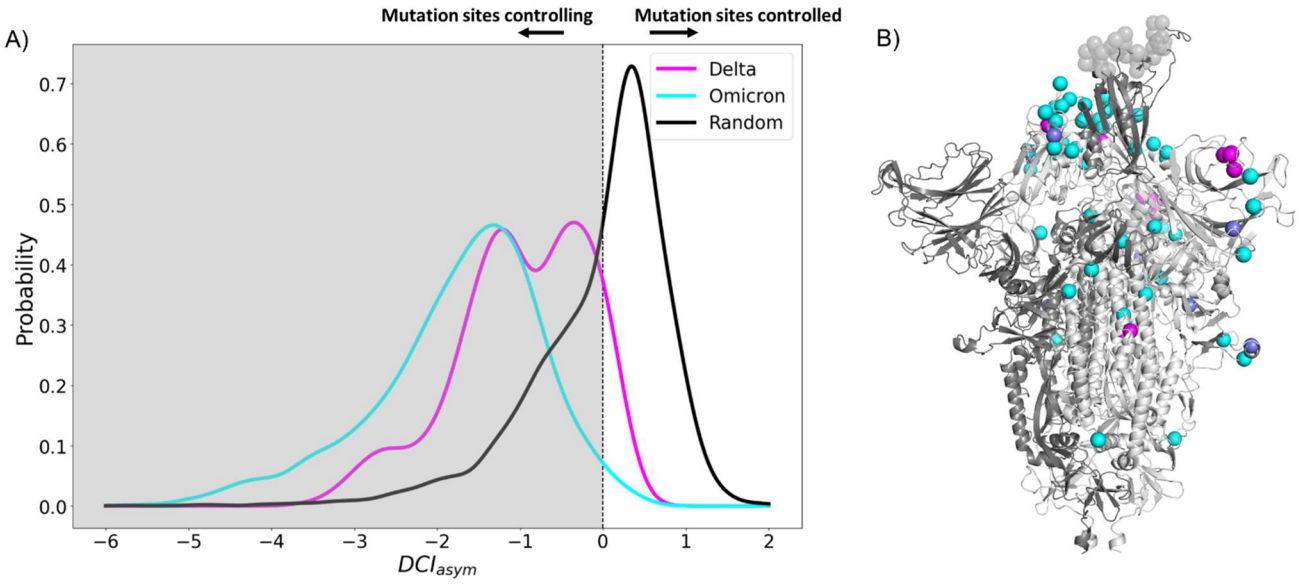


Figure 4. (A) DCI asymmetry with low EP characteristic mutation sites of Delta or Omicron strains in the closed chains and the binding interface of RBD in the open chain. Delta displays a second peak closer to zero, suggesting that Delta mutation sites (M = -0.98, SD = 0.80) have less allosteric control over the hACE2 binding sites than Omicron mutation sites (M = -1.74, SD = 1.00) ($p < .001$). However, both sets of sites have far more control over hACE2 binding sites than expected, based on a random control group (M = 0.03, SD = 0.85) ($p < .001$). (B) S protein structure showing binding interface sites (transparent gray), Delta mutation sites (magenta), Omicron mutation sites (cyan), and sites mutated in both Omicron and Delta (Blue).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Experimental results motivate the use of EpiScore within the SARS-CoV-2 Spike protein.

The fact that the identified CAPs in the reference protein and the more recently evolved CAPs of Delta and Omicron variants both show a high degree of control over the active sites begs the question: what is the complex interaction between these previous and new CAP sites? Motivated by this concept, we explore the interplay of mutational pairs to understand the effects of the specific amino acid backgrounds associated with these two predominant variants. Some CAP sites in Delta and Omicron have already been considered adaptive (Kemp et al. 2021; Kistler et al. 2022; Maher et al. 2022; Neher 2022).

It is well understood that the impact of even a single mutation to a protein sequence can sometimes dramatically alter the biophysical behavior of the system. However, the mechanistic impact of point mutations can only be fully understood when the sequence background upon which it is made is accounted for. This means that, in the case of strains with multiple mutations, the interplay between mutated positions will ultimately impact a protein as an aggregate behavior, where the presence of previous mutations may strongly (or weakly) influence some mutations. This concept of non-additivity is known as epistasis. In fact, studies of evolutionary pathways of mutations have suggested that a majority of the mutations have a second or a higher order epistasis among them (Bershtein et al. 2006). Nature exploits this higher order complex relationship between the mutations to evolve their function.

To computationally capture and interpret the pairwise effects of mutations, we have developed an in-house computational tool called EpiScore (Figure 5A). Here, we evaluate how a given position pair *i j* may affect other critical positions *k* of the protein. EpiScore is the relative coupling strength to a position *k* when positions *i* and *j* are perturbed *simultaneously* compared to the average dynamic coupling strength of *i* to *k* and *j* to *k.* EpiScore has previously been used successfully to capture overarching trends in GB1 deep mutational scan data as well as specific instances of the development of antibiotic resistance in various enzymatic systems (Campitelli and Ozkan 2020). An EpiScore of 1 indicates perfect coupling additivity, and deviations from this value represent non-additive behavior between position pairs and

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functionally important sites. Prior EpiScore work has shown a difference in EpiScore between the sites of compensatory and non-compensatory mutations, where both yield distributions with peaks around 1, but non-compensatory mutations show higher deviation in their EpiScore distribution (Ose, Campitelli, et al. 2022).

Many studies have confirmed epistasis between residues within the S protein (Moulana et al. 2022a; Starr, Greaney, Hannon, et al. 2022; Moulana et al. 2023; Witte et al. 2023). These epistatic residues can have various effects on hACE2 or antibody binding. To further motivate our use of EpiScore within SARS-CoV-2, we calculate the EpiScore (Figure 5B) of a set of mutation pairs used by Moulana et al. (Nature comm 2022) and compare our results to quantified epistatic effects determined by the experimental hACE2 binding affinity of "first-order" single mutation variants compared to "second-order" mutation pair. Our EpiScore results and the experimentally determined epistasis have a reasonable similarity. Both methods captured highly epistatic behavior among residues 493, 496, 498, 501, and 505, as well as a lack of epistatic behavior for residues 339, 371, 373, and 375.

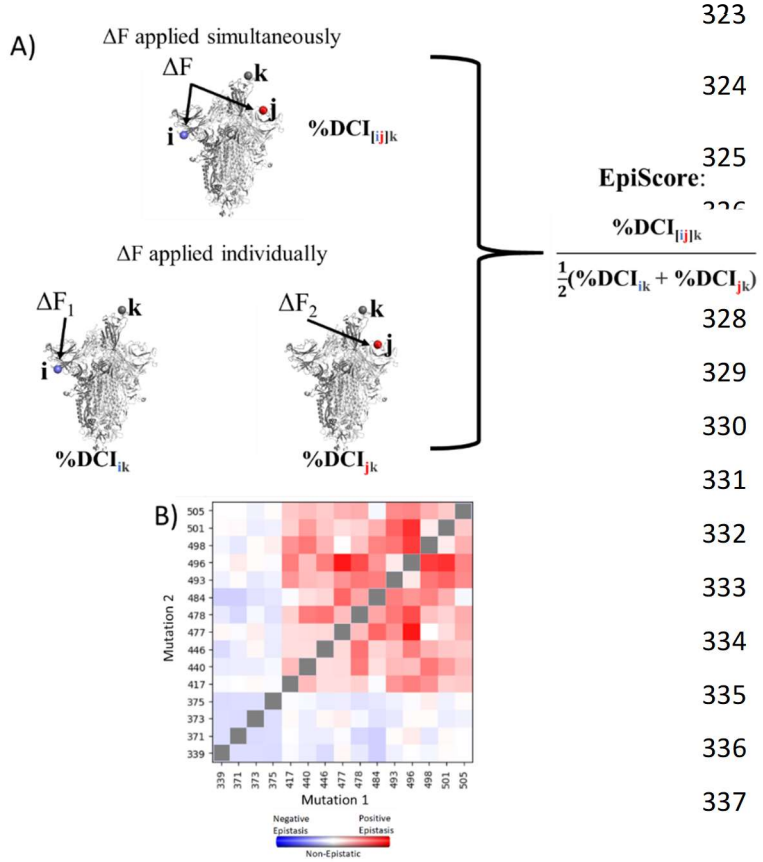


Figure 5. (A) Schematic representation of cross-chain EpiScore, describing i, j, in chain B and C respectively and its impact in RBD binding position k in the open RBD conformer chain A. (B) Colors indicate EpiScore values for given mutation pairs, averaged over hACE2 binding sites. Cross-chain residue pairs in the upper right tend to be highly epistatic, similar to pairwise second-order interaction coefficients from Moulana et al. (Nature comm 2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Episcore highlights the epistatic relationship between the recent adaptive mutations in VOCs and the CAPs of the Wuhan reference.

Seeking further to understand the role of epistasis within S protein variants, we explored the possibility of epistatic relationships between the CAPs of the Wuhan variant and the new CAPs in VOCs. Thus, we computed the Episcore of these CAP positions in the closed RBD domains (i.e., chain B and C) with respect to functional hACE2 binding interface sites of the open RBD domain chain (chain A) (Figure 5B) and obtained Episcore distributions.

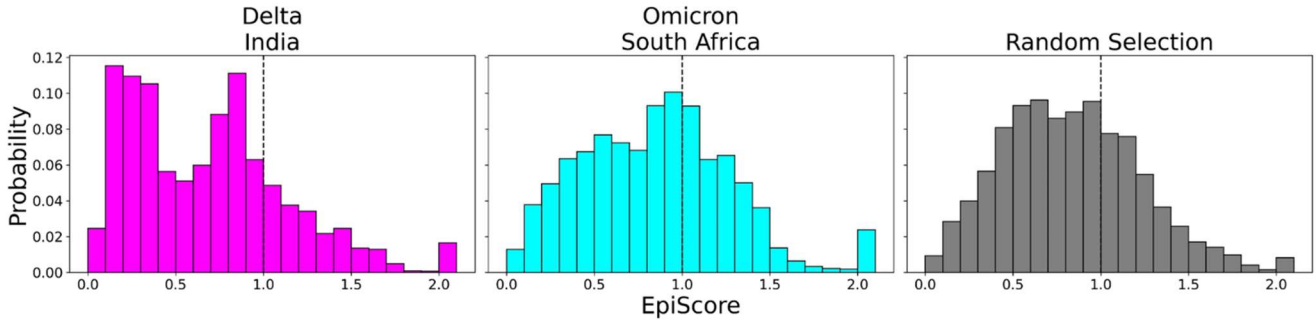


Figure 6. EpiScores with I = low EP Delta mutation sites (magenta), low EP Omicron mutation sites (Cyan), and a random selection of sites (gray), j = low EP sites in the Wuhan variant, and k = the binding interface of the open chain. EpiScores using sites of either variant (Delta: M = 0.70, SD = 0.50, Omicron M = 0.86, SD = 0.46) are significantly different (p<.001) from a set of EpiScores using random sites (M = 0.84, SD = 0.41), but the distribution for Delta variants differs much more from the other two. EpiScores for other variants can be found in Supplementary Figure S1.

To contrast these variants, Omicron (Figure 6, cyan) shows a high proportion of additive, potentially compensatory, mutations compared to the Delta variant (Figure 6, magenta), with a peak centered on 1. The comparatively more pathogenic Delta variant exhibited many non-additive, suspected non-compensatory mutations with EpiScores below one. This again suggests that the cross-communication between the open and closed chain of the S protein is important for regulating the function. Four out of

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

seven low EP Delta mutation sites used in this analysis often resulted in EpiScores below 1. Each of those is found in the N-terminal domain (NTD) on or near the N3 loop and is implicated in antibody escape in recent studies (Chi et al. 2020; Weisblum et al. 2020; Harvey, Carabelli, Jackson, Gupta, Thomson, Harrison, Ludden, Reeve, Rambaut, COVID-19 Genomics UK (COG-UK) Consortium, et al. 2021; Klinakis et al. 2021; Cantoni et al. 2022). The low EpiScores of NTD mutations suggest that they dampen the control of Wuhan variant CAPs over the active site in addition to their effects on antibody binding. It is possible that what the Delta variant gained in transmission rate also came with being more harmfully pathogenic due in part to negatively epistatic interactions. It follows that the mutations leading to the development of the Omicron strain were compensatory in nature, possibly leading to a lower pathogenicity and higher effective immune escape, resulting in a higher transmission rate. It is worth noting that other variants contain NTD mutations which result in low EpiScores, however the proportion of these mutations within the set is considerably less than in Delta (Supplementary Figure S2).

One of the more notable similarities of generated EpiScore distributions is a tail of EpiScore values upward of 2.0, indicating highly epistatic behavior. Interestingly, these tails are largely due to three different CAPs: 346R, 486F, and 498Q. Those residues are nearby one another within the protein structure and have been reported to play a role in antibody binding, either being known antibody binding sites (346R and 486F) or having received very high antibody accessibility scores (498Q) (Harvey, Carabelli, Jackson, Gupta, Thomson, Harrison, Ludden, Reeve, Rambaut, COVID-19 Genomics UK (COG-UK) Consortium, et al. 2021; Raghuvamsi et al. 2021). These observed high Episcore values also support the other studies that the epistatic interactions between these CAPs and the mutations of the VOCs within the S protein are crucial for maintaining binding affinity of hACE2 whilst evading immunity (Hong et al. 2022; Moulana et al. 2022b; Starr, Greaney, Stewart, et al. 2022).

Viewing EpiScores of Delta and Omicron potentially adaptive mutation sites with only CAP site 486F (a binding site for both hACE2 and antibodies) (Huang et al. 2020; Ali et al. 2021; Harvey, Carabelli, Jackson,

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Gupta, Thomson, Harrison, Ludden, Reeve, Rambaut, COVID-19 Genomics UK (COG-UK) Consortium, et al. 2021; Raghuvamsi et al. 2021) shows highly epistatic interactions at other hACE2 binding sites (Figure 6). However, within a recent and rapidly spreading subvariant of Omicron, XBB 1.5, we see a mutation of F to S, a rare double nucleotide mutation, at site 486. This new variant has unprecedented immune escape capabilities, resisting neutralizing antibodies almost entirely (Qu et al. 2023). EpiScores of other XBB 1.5 specific mutation sites with 486S are almost entirely greater than 1, showing an even higher degree of epistasis with the binding sites of RBD (Figure 7). These results present a threefold importance for the F486S mutation: Not only does this residue alter antibody (i.e., immune escape) and hACE2 binding by directly modifying a binding site, but it may also be responsible for modifying hACE2 binding via epistatic cooperation with other co-occurring mutations.

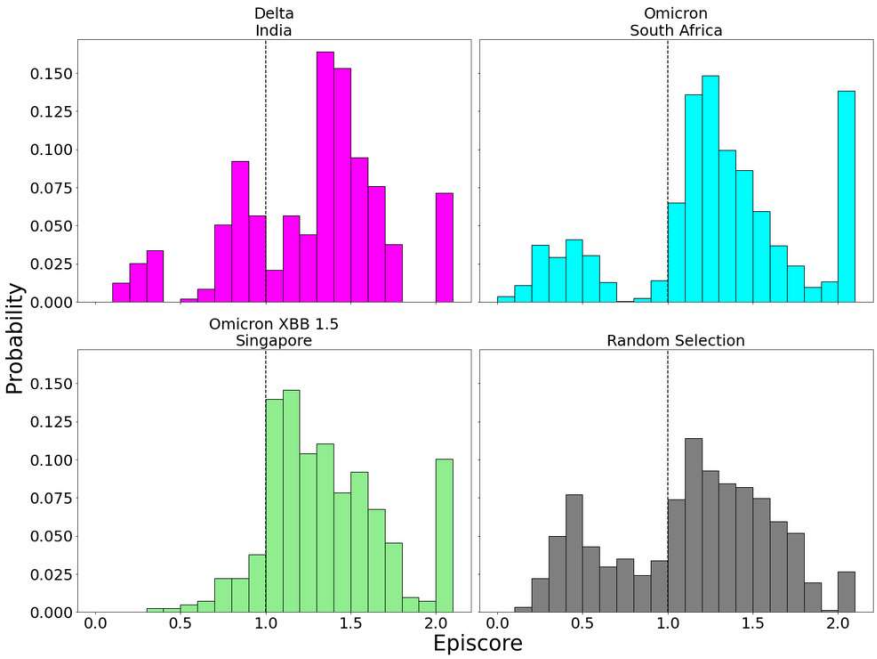


Figure 7. EpiScores with i = low EP Delta mutation sites (magenta), low EP Omicron mutation sites (Cyan), and a random selection of sites (gray), j = site 486, and k = the binding interface of the open chain. CAP and hACE2 and antibody binding site 486 displays epistasis with almost all XBB 1.5 variant sites at almost every hACE2 binding site (M = 1.40, SD = 0.46) and presents a significantly different profile from other variant sites (p <.01). EpiScores involving 486 for other variants can be found in Supplementary Figure S1.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Change in flexibility of RBD binding site correlates with experimental binding affinities for Omicron and Omicron XBB variants.

Experimental studies have tracked hACE2 binding for different variants since the virus first spread (Ali et al. 2021; Barton et al. 2021; Ozono et al. 2021; Wu et al. 2022). Within the Omicron variant, for example, characteristic mutations on the RBD are shown to increase the overall binding affinity of the virus to the ACE2 receptor, which is suspected to allow it to spread more easily (Kim et al. 2021). Furthermore, the new Omicron XBB and Omicron XBB 1.5 variants contain additional mutations in the RBD and antibody binding residues, which may further impact their dynamics and interactions with the host.

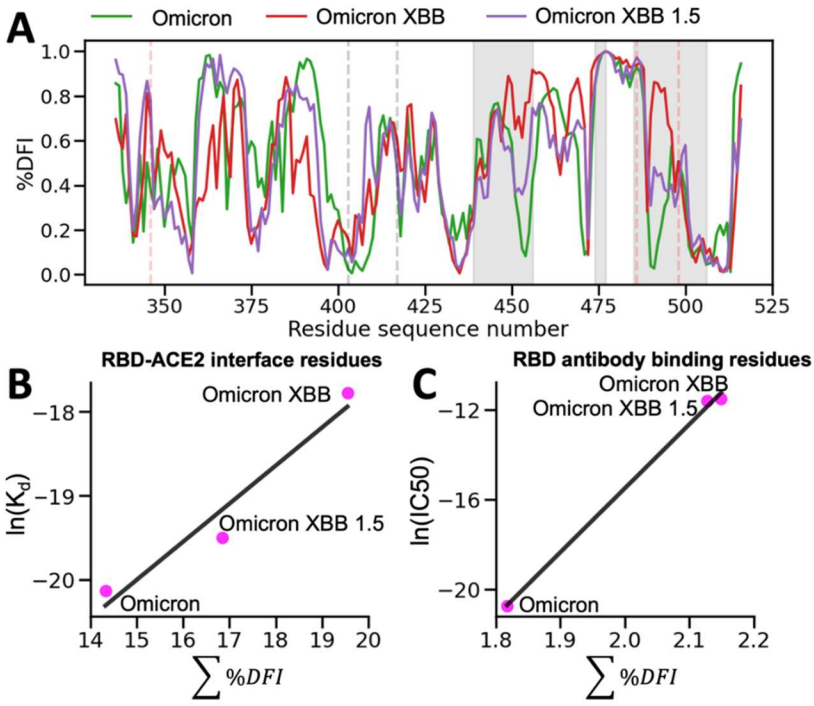


Figure 8. The %DFI calculations for variants Omicron, XBB, and XBB 1.5. (A) %DFI profile of the variants are plotted in the same panel. The grey shaded areas and dashed lines indicate the ACE2 binding regions, whereas the red dashed lines show the antibody binding residues. (B) The sum of %DFI values of RBD-ACE2 interface residues. The trend of total %DFI with the log of K_d values overlaps with the one seen with the experiments (R=0.97). (C) The RBD antibody binding residues are used to calculate the sum of %DFI. The ranking captured with the total %DFI agrees with the log of IC50 values from the experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To gain deeper insights into the impact of dynamics on the binding affinity of hACE2 and antibodies with the recent Omicron XBB variants, we conducted molecular dynamics (MD) simulations. By analyzing the resulting trajectories, we investigated how these mutations influence the flexibility and rigidity of the RBD and antibody binding residues, consequently affecting their binding affinity and potential for immune evasion (Figure 8). To understand the overall flexibility changes, we measured the sum of DFI of the ACE2 binding residues, as well as the sum of DFI of the antibody binding residues, calculated from the MD trajectories and compared then with experimental viral binding (disassociation constants) and immunity evasion antibody IC50 values (Yue et al. 2023).

This investigation elucidated the impact of mutations in the receptor-binding domain (RBD) and antibody binding residues on the binding affinity of the S protein and immune evasion by modulating their flexibility and rigidity (Figure 8A). The Omicron XBB variant exhibits heightened flexibility in hACE2 and antibody binding residues, reducing infectivity and enhancing immune evasion. Conversely, the Omicron XBB 1.5 variant induces distinct dynamics in these regions, rendering the RBD-ACE2 interface more rigid while increasing flexibility in antibody binding residues. These effects indicate that Omicron XBB1.5 retains its antibody escape capabilities while regaining ACE2 binding affinity comparable to previous Omicron variants, in accordance with experimental findings (Yue et al. 2023). These findings suggest that mutations in the RBD and antibody binding residues can have complex effects on the dynamics of the protein and, ultimately, on the virus's ability to infect and evade the host immune system through an alteration of biding site dynamics.

Conclusion:

We analyzed the evolutionary trajectory of the CoV-2 S protein in humans to understand the dynamic and epistatic interactions of the mutations defining specific VOCs within the S protein. We first obtain the phylogenetic tree of the COV-2 S protein and identify the sites of certain recent mutations known as

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

candidate adaptive polymorphisms (CAPs). CAPs are considered adaptive because mutations rarely tolerated in closely related sequences have suddenly become fixed, implying a degree of functional importance or evolution (Liu et al. 2016). In addition, our earlier work has shown that CAPS can also be compensatory as multiple CAPs may dynamically compensate for one another, changing the dynamic landscape and allowing for different mutations (Ose, Campitelli, et al. 2022). We then explored the mechanistic insights and epistatic relationship between the observed mutations in different VOCs and the CAP sites, and, particularly, the relationship between CAP sites and the functionally critical RBD domain using our dynamic coupling analysis (Kumar, A., Glembo, T.J., Ozkan, S.B. 2015b).

We find a mechanistic pattern in the S protein evolution that is common amongst previously studied systems, where allosteric sites exert control over the dynamics of the active site, and mutations of these allosteric sites modulate function. Within the S protein, these critical regulatory sites are CAPs and VOC-defining new putative adaptive mutations. However, this regulation is significantly less in Delta-defining mutations than in the Omicron variant. We speculate that this may cause some differences in behavior between the strains. Simply looking at the connection between hACE2 binding and pathogenicity, we know that the Omicron variant is less pathogenic and has a lower binding affinity (Wu et al. 2022) than previous variants. Our dynamics analysis provides a mechanistic insight where the Omicron defining sites have greater control over the active sites than the Delta variant, and compensate for the functional advantage of CAPs, thus, the greatest infectivity may not be a coincidence.

Specifically, we find that the interactions between CAP sites and VOC-defining mutations show fingerprints of non-compensatory dynamics within the Delta variant. In contrast, mutations leading to the Omicron variant are largely compensatory, driving critical dynamical behavior closer to the patterns observed within the wild-type. These interactions may drive observed behavior similar between the reference and Omicron strains yet differ in the delta strain, such as the severity of infection as evidenced by hospitalization rates (Houhamdi et al. 2022; Menni et al. 2022).

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Long-ranged interactions between different sites within a given protein is critically important for protein function (Peters and Lively 1999; Bershtein et al. 2006; Collins et al. 2006; Ekeberg et al. 2013; Levy et al. 2017; Harrigan et al. 2018; Otten et al. 2018; Rojas Echenique et al. 2019; Shimagaki and Weigt 2019; de la Paz et al. 2020; Rizzato et al. 2020; Yang et al. 2020; Bisardi et al. 2022) and for the CoV-2 S protein in particular (Zeng et al. 2020; Castiglione et al. 2021; Dong et al. 2021; Garvin et al. 2021; Nielsen et al. 2022; Ramarao-Milne et al. 2022; Rochman et al. 2022; Rodriguez-Rivas et al. 2022). By showing dynamic differences between the interactions of CAPs, which have likely played a major role in allowing the virus to infect human hosts, the binding site, and the characteristic mutations of dominant Delta and Omicron strains, we see a "fine-tuning" of protein behavior. As variants continue to evolve, Omicron sub-variants are of growing concern due in large part to further increased immune evasion (Callaway 2022; Wang, Guo, et al. 2022; Wang, Iketani, et al. 2022), and we observe that the new mutations observed in antibody binding sites yield more epistatic interaction with the CAPs. In addition to supporting previous dynamic research on the S protein, this analysis provides the insight that CAP sites are of continued importance to protein function and should be given special attention when considering the impact of future mutations.

Methods

Dynamic Flexibility and Dynamic Coupling

The Dynamic Flexibility Index utilizes a Perturbation Response Scanning technique that combines the Elastic Network Model (ENM) and Linear Response Theory (LRT) (Gerek and Ozkan 2011; Nevin Gerek, Z., Kumar, S., Banu Ozkan, S. 2013). In ENM, the protein is considered as a network of beads at Cα positions interacting with each other via a harmonic spring potential. Using LRT, Δ**R** is calculated as the fluctuation response vector of residue *j* due to unit force's **F** perturbation on residue *i,* averaged over multiple unit force directions to simulate an isotropic perturbation.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$[\Delta\mathbf{R}]_{3N\times1} = [\mathbf{H}]^{-1}_{3N\times3N}[\mathbf{F}]_{3N\times1} \tag{1}$$

H is the Hessian, a 3N × 3N matrix that can be constructed from 3-D atomic coordinate information and is composed of the second derivatives of the harmonic potential with respect to the components of the position's vectors of length 3N. The hessian inverse in this equation may be replaced with the covariance matrix **G** obtained from MD simulations as follows.

$$[\Delta\mathbf{R}]_{3N\times1} = [\mathbf{G}]_{3N\times3N}[\mathbf{F}]_{3N\times1} \tag{2}$$

MD simulations were used to obtain the DFI profiles of Omicron, Omicron XBB, and Omicron XBB 1.5. In order to obtain DFI, each position in the structure was perturbed sequentially to generate a Perturbation Response Matrix ***A***

$$\mathbf{A}_{N\times N} = \begin{bmatrix} |\Delta R^1|_1 & \cdots & |\Delta R^N|_1 \\ \vdots & \ddots & \vdots \\ |\Delta R^1|_N & \cdots & |\Delta R^N|_N \end{bmatrix} \tag{3}$$

where $|\Delta R^j|_i = \sqrt{\langle(\Delta R)^2\rangle}$ is the magnitude of fluctuation response at position *i* due to perturbations at position *j. T*he DFI value of position *i* is then treated as the displacement response of position *i* relative to the net displacement response of the entire protein, which is calculated by sequentially perturbing each position in the structure.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$$\mathrm{DFI_i} = \frac{\sum_{j=1}^{N} \left|\Delta R^j\right|_i}{\sum_{i=1}^{N} \sum_{j=1}^{N} \left|\Delta R^j\right|_i} \tag{4}$$

It is also often useful to quantify position flexibility relative to the flexibility ranges unique to individual structures. To that end, DFI can be presented as a percentile rank, %DFI. All %DFI calculations present in this work used the DFI value of every residue of the full spike structure for ranking. The DFI parameter can be considered a measure of a given amino acid position's ability to explore its local conformational space.

Dynamic Coupling Index

Similar to DFI, the dynamic coupling index (DCI) (Larrimore et al. 2017; Kumar, A., Glembo, T.J., Ozkan, S.B. 2015b) also utilizes Perturbation Response Scanning with the Elastic Network Model and Linear Response Theory. DCI captures the strength of displacement response of a given position *i* upon perturbation to a single functionally important position (or subset of positions) *j*, relative to the average fluctuation response of position *i* when all of the positions within a structure are perturbed.

$$\mathrm{DCI_{ji}} = \frac{\sum_{j}^{N_{functional}} \left|\Delta R^j\right|_i / N_{functional}}{\sum_{j=1}^{N} \left|\Delta R^j\right|_i / N} \tag{5}$$

When only positional pairs are concerned, this expression reduces to:

$$\mathrm{DCI_{ji}} = \frac{\left|\Delta R^j\right|_i}{\sum_{j=1}^{N} \left|\Delta R^j\right|_i / N} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

As such, this parameter represents a measure of the dynamic coupling between *i* and *j* upon a perturbation to *j*. As with DFI, DCI_{ji} can also be presented as a percentile-ranked $\%DCI_{ji}$.

One of the most important aspects of DCI is that the entire network of interactions is explicitly included in subsequent calculations without the need for dimensionality reduction techniques. If one considers interactions such as communication directionality or dynamic coupling regulation between position pairs as inherent properties of an anisotropic interaction network, it is critical to include the interactions of the entire network to accurately model the effect one residue can have on another.

Here, we present two further extensions of DCI which allow us to uniquely model coupling directionality and epistatic effects: DCI_{asym} and EpiScore, respectively. Interestingly, we can capture asymmetry between different residues within a protein through DCI, as a coupling in and of itself is asymmetric within an anisotropic network. That is, each amino acid has a set of positions to which it is highly coupled, and this anisotropy in connections gives rise to unique differences in coupling between a given *i j* pair of amino acids which do not have direct interactions (Figure 2A). DCI_{asym}, then, is simply DCI_{ij} (the normalized displacement response of position *j* upon a perturbation to position *i*) − DCI_{ji} (Equation (7)). Using DCI_{asym} we can determine a cause-effect relationship between the *i j* pair in terms of force/signal propagation between these two positions.

$$\mathrm{DCI_{asym}} = \mathrm{DCI_{ij}} - \mathrm{DCI_{ji}} \tag{7}$$

$$\%\mathrm{DCI_{asym}} = \%\mathrm{DCI_{ij}} - \%\mathrm{DCI_{ji}} \tag{8}$$

where a positive DCI_{asym} value indicates communication from position *i* to position *j*.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

EpiScore can identify or describe potential non-additivity in substitution behavior between residue pairs. This metric can capture the differences in a normalized perturbation response to a position *k* when a force is applied at two residues *i* and *j* simultaneously versus the average additive perturbation response when each residue *i*, *j*, is perturbed individually (Figure 5A, Equation 9).

$$\mathrm{EpiScore} = \frac{\%\mathrm{DCI}_{[ij]k}}{\frac{1}{2}\left(\%\mathrm{DCI}_{ik} + \%\mathrm{DCI}_{jk}\right)} \tag{9}$$

EpiScore values < 1 (> 1) indicate that the additive perturbations of positions *i* and *j* generates a greater (lesser) response at position *k* than the effect of a simultaneous perturbation. This means that, when treated with a simultaneous perturbation at both sites i and j, the displacement response of k is lower (higher) than the average effect of individual perturbations to *i* and *j*, one at a time. As EpiScore is a linear scale, the further the value from 1, the greater the effect described above.

Molecular Dynamics (MD)

The production simulations of the variants Omicron, Omicron XBB, and Omicron XBB 1.5 were generated using the AMBER software package. The mutations in the variants were modeled using PYMOL taking the template as PDB 6M0J. The initial input proteins were parametrized utilizing the ff14SB force field (Maier et al. 2015). To ensure adequate solvation of the protein, the solvation box was defined to encompass the protein, maintaining a minimum distance of 16Å from the protein to the box edges, utilizing the explicit TIP3P water model (Sun 1995). The neutralization of the solvated system was achieved through the addition of sodium and chloride ions. The system was subjected to a steepest descent algorithm for 11000

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

steps for minimization purposes. The minimized system was then heated to 300K and subjected to a constant number of particles, pressure, and temperature ensemble (NPT) production simulations. These simulations were conducted using the Langevin thermostat (Hünenberger 2005) and Berendsen barostat (Berendsen et al. 1984). The hydrogens were constrained using the SHAKE algorithm (Pearlman et al. 1995). The production trajectories were simulated for 1µs at 300K and 1 bar.

Data and Resource Availability

The code to perform DFI and DCI analysis is available at https://github.com/SBOZKAN/DFI-DCI. Molecular Dynamics data are available upon request. The mutation sites and EP values are contained in the supporting information files as "Supplementary_mutation_info.csv". The alignment used to generate EP values is also contained within the supporting information files as "EP_alignment.fas". Protein Databank ID number 6VXX (Walls et al. 2020) was used for closed conformation DCI calculations. 6VSB (Wrapp et al. 2020) was used for DFI calculations, EpiScore calculations, and open conformation DCI calculations. 6M0J (Lan et al. 2020) was used in molecular dynamics simulations of the RBD.

Acknowledgements

Funding was provided to N.J.O., P.C., T.M., and I.C.K. by the Gordon and Betty Moore Foundation (award number AWD00034439) and to S.B.O. by the National Science Foundation (award numbers: 1715591 and 1901709) and the National Institutes of Health R01GM147635-01. S.K. acknowledges the National Science Foundation (GCR 1934848) and the National Institutes of Health (GM139540) grants.

References:

Ali F, Kasry A, Amin M. 2021. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. *Med. Drug Discov.* 10:100086.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Amicone M, Borges V, Alves MJ, Isidro J, Zé-Zé L, Duarte S, Vieira L, Guiomar R, Gomes JP, Gordo I. 2022. Mutation rate of SARS-CoV-2 and emergence of mutators during experimental evolution. *Evol. Med. Public Health* 10:142–155.

Barton MI, MacGowan SA, Kutuzov MA, Dushek O, Barton GJ, van der Merwe PA. 2021. Effects of common mutations in the SARS-CoV-2 Spike RBD and its ligand, the human ACE2 receptor on binding affinity and kinetics. *eLife* 10:e70658.

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235–242.

Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS. 2006. Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444:929–932.

Bhabha G, Ekiert DC, Jennewein M, Zmasek CM, Tuttle LM, Kroon G, Dyson HJ, Godzik A, Wilson IA, Wright PE. 2013. Divergent evolution of protein conformational dynamics in dihydrofolate reductase. *Nat. Struct. Mol. Biol.* 20:1243–1249.

Bisardi M, Rodriguez-Rivas J, Zamponi F, Weigt M. 2022. Modeling Sequence-Space Exploration and Emergence of Epistatic Signals in Protein Evolution.Ozkan B, editor. *Mol. Biol. Evol.* 39:msab321.

Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, Castoe TA, Rambaut A, Robertson DL. 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* 5:1408–1417.

Brister JR, Ako-adjei D, Bao Y, Blinkova O. 2015. NCBI Viral Genomes Resource. *Nucleic Acids Res.* 43:D571–D577.

Butler BM, Kazan IC, Kumar A, Ozkan SB. 2018. Coevolving residues inform protein dynamics profiles and disease susceptibility of nSNVs.Jernigan RL, editor. *PLOS Comput. Biol.* 14:e1006626.

Cagliani R, Forni D, Clerici M, Sironi M. 2020. Computational Inference of Selection Underlying the Evolution of the Novel Coronavirus, Severe Acute Respiratory Syndrome Coronavirus 2.Gallagher T, editor. *J. Virol.* 94:e00411-20.

Callaway E. 2022. COVID 'variant soup' is making winter surges hard to predict. *Nature* 611:213–214.

Campbell, E., Kaltenbach, M., Correy, G., Carr, P., Porebski, B.T., Livingstone, E., Jurnou, L., Hyvönen, M., Buckle, A.M., Weik, M., et al. 2016. The role of protein dynamics in the evolution of new enzyme function. *Nat Chem Biol*.

Campitelli P, Lu J, Ozkan SB. 2022. Dynamic allostery highlights the evolutionary differences between the CoV-1 and CoV-2 main proteases. *Biophys. J.* 121:1483–1492.

Campitelli P, Modi T, Kumar S, Ozkan SB. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* 49:267–288.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Campitelli P, Modi T, Kumar S, Ozkan SB. 2020a. The Role of Conformational Dynamics and Allostery in Modulating Protein Evolution. *Annu. Rev. Biophys.* 49:267–288.

Campitelli P, Ozkan SB. 2020. Allostery and Epistasis: Emergent Properties of Anisotropic Networks. *Entropy* 22:667.

Campitelli P, Swint-Kruse L, Ozkan SB. 2020b. Substitutions at Nonconserved Rheostat Positions Modulate Function by Rewiring Long-Range, Dynamic Interactions.Wilke C, editor. *Mol. Biol. Evol.* 38:201–214.

Cantoni D, Murray MJ, Kalemera MD, Dicken SJ, Stejskal L, Brown G, Lytras S, Coey JD, McKenna J, Bridgett S, et al. 2022. Evolutionary remodelling of N-terminal domain loops fine-tunes SARS-CoV -2 spike. *EMBO Rep.* [Internet] 23. Available from: https://onlinelibrary.wiley.com/doi/10.15252/embr.202154322

Carabelli AM, Peacock TP, Thorne LG, Harvey WT, Hughes J, COVID-19 Genomics UK Consortium, De Silva TI, Peacock SJ, Barclay WS, De Silva TI, et al. 2023. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat. Rev. Microbiol.* [Internet]. Available from: https://www.nature.com/articles/s41579-022-00841-7

Castiglione GM, Zhou L, Xu Z, Neiman Z, Hung C-F, Duh EJ. 2021. Evolutionary pathways to SARS-CoV-2 resistance are opened and closed by epistasis acting on ACE2.Hejnol A, editor. *PLOS Biol.* 19:e3001510.

Chan C-M, Woo PCY, Lau SKP, Tse H, Chen H-L, Li F, Zheng B-J, Chen L, Huang J-D, Yuen K-Y. 2008. Spike Protein, S, of Human Coronavirus HKU1: Role in Viral Life Cycle and Application in Antibody Detection. *Exp. Biol. Med.* 233:1527–1536.

Changeux J-P, Edelstein SJ. 2005. Allosteric Mechanisms of Signal Transduction. *Science* 308:1424–1428.

Chi X, Yan R, Zhang Jun, Zhang G, Zhang Y, Hao M, Zhang Z, Fan P, Dong Y, Yang Y, et al. 2020. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* 369:650–655.

Collins SR, Schuldiner M, Krogan NJ, Weissman JS. 2006. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* 7:R63.

Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, Hiller M, Koepfli K-P, Pfenning AR, Zhao H, et al. 2020. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci.* 117:22311–22322.

Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Morris MK, Sotomayor-González A, Glasner DR, Reyes KR, Gliwa AS, et al. 2021. Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. Infectious Diseases (except HIV/AIDS) Available from: http://medrxiv.org/lookup/doi/10.1101/2021.03.07.21252647

Díaz-Salinas MA, Li Q, Ejemel M, Yurkovetskiy L, Luban J, Shen K, Wang Y, Munro JB. 2022. Conformational dynamics and allosteric modulation of the SARS-CoV-2 spike. *eLife* 11:e75433.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dicken SJ, Murray MJ, Thorne LG, Reuschl A-K, Forrest C, Ganeshalingham M, Muir L, Kalemera MD, Palor M, McCoy LE, et al. 2021. Characterisation of B.1.1.7 and Pangolin coronavirus spike provides insights on the evolutionary trajectory of SARS-CoV-2. Microbiology Available from: http://biorxiv.org/lookup/doi/10.1101/2021.03.22.436468

Dong A, Zhao J, Griffin C, Wu R. 2021. The Genomic Physics of COVID-19 Pathogenesis and Spread. *Cells* 11:80.

Doshi U, Holliday MJ, Eisenmesser EZ, Hamelberg D. 2016. Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci.* 113:4735–4740.

Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. 2012. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* 41:429–452.

Ekeberg M, Lövkvist C, Lan Y, Weigt M, Aurell E. 2013. Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys. Rev. E* 87:012707.

Frost SDW, Magalis BR, Kosakovsky Pond SL. 2018. Neutral Theory and Rapidly Evolving Viral Pathogens.Kumar S, editor. *Mol. Biol. Evol.* 35:1348–1354.

Garvin MR, Prates ET, Romero J, Cliff A, Machado Gazolla JGF, Pickholz M, Pavicic M, Jacobson D. 2021. Rapid expansion of SARS-CoV-2 variants of concern is a result of adaptive epistasis. Evolutionary Biology Available from: http://biorxiv.org/lookup/doi/10.1101/2021.08.03.454981

Gerek ZN, Ozkan SB. 2011. Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning.Nussinov R, editor. *PLoS Comput. Biol.* 7:e1002154.

Gobeil SM-C, Janowska K, McDowell S, Mansouri K, Parks R, Manne K, Stalls V, Kopp MF, Henderson R, Edwards RJ, et al. 2021. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. *Cell Rep.* 34:108630.

Gobeil SM-C, Janowska K, McDowell S, Mansouri K, Parks R, Stalls V, Kopp MF, Manne K, Li D, Wiehe K, et al. 2021. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. *Science* 373:eabi6226.

Goldman N. 1990. Maximum Likelihood Inference of Phylogenetic Trees, with Special Reference to a Poisson Process Model of DNA Substitution and to Parsimony Analyses. *Syst. Zool.* 39:345.

Gur M, Taka E, Yilmaz SZ, Kilinc C, Aktas U, Golcuk M. 2020. Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states. *J. Chem. Phys.* 153:075101.

Harrigan P, Madhani HD, El-Samad H. 2018. Real-Time Genetic Compensation Defines the Dynamic Demands of Feedback Control. *Cell* 175:877-886.e10.

Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, Consortium C-19 GU (COG-U, et al. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19:409–424.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Harvey WT, Carabelli AM, Jackson B, Gupta RK, Thomson EC, Harrison EM, Ludden C, Reeve R, Rambaut A, COVID-19 Genomics UK (COG-UK) Consortium, et al. 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19:409–424.

Henderson R, Edwards RJ, Mansouri K, Janowska K, Stalls V, Gobeil SMC, Kopp M, Li D, Parks R, Hsu AL, et al. 2020. Controlling the SARS-CoV-2 spike glycoprotein conformation. *Nat. Struct. Mol. Biol.* 27:925–933.

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, Schiergens TS, Herrler G, Wu N-H, Nitsche A, et al. 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* 181:271-280.e8.

Hong Q, Han W, Li J, Xu S, Wang Yifan, Xu C, Li Z, Wang Yanxing, Zhang C, Huang Z, et al. 2022. Molecular basis of receptor binding and antibody neutralization of Omicron. *Nature* 604:546–552.

Houhamdi L, Gautret P, Hoang VT, Fournier P, Colson P, Raoult D. 2022. Characteristics of the first 1119 SARS-CoV-2 Omicron variant cases, in Marseille, France, November−December 2021. *J. Med. Virol.* 94:2290–2295.

Huang Y, Yang C, Xu X, Xu W, Liu S. 2020. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* 41:1141–1149.

Hünenberger PH. 2005. Thermostat Algorithms for Molecular Dynamics Simulations. In: Dr. Holm C, Prof. Dr. Kremer K, editors. Advanced Computer Simulation. Vol. 173. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105–149. Available from: http://link.springer.com/10.1007/b99427

Jackson CB, Farzan M, Chen B, Choe H. 2022. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* 23:3–20.

Kazan IC, Mills JH, Ozkan SB. 2023. Allosteric regulatory control in dihydrofolate reductase is revealed by dynamic asymmetry. *Protein Sci.* 32:e4700.

Kazan IC, Sharma P, Rahman MI, Bobkov A, Fromme R, Ghirlanda G, Ozkan SB. 2022. Design of novel cyanovirin-N variants by modulation of binding dynamics through distal mutations. *eLife* 11:e67474.

Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, Hosmillo M, Rees-Spear C, Mlcochova P, Lumb IU, et al. 2021. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* 592:277–282.

Keskin O, Bahar I, Jernigan RL, Beutler JA, Shoemaker RH, Sausville EA, Covell DG. 2000. Characterization of anticancer agents by their growth inhibitory activity and relationships to mechanism of action and structure. *Anticancer. Drug Des.* 15:79–98.

Khan A, Zia T, Suleman M, Khan T, Ali SS, Abbasi AA, Mohammad A, Wei D. 2021. Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J. Cell. Physiol.* 236:7045–7057.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kim, H., Zou, T., Modi, C., Dörner, K., Grunkemeyer, T.J., Chen, L., Fromme, R., Matz, M. V., Ozkan, S.B., Wachter, R.M. 2015. A hinge migration mechanism unlocks the evolution of green-to-red photoconversion in GFP-like proteins. *Structure* 23:34–43.

Kim Sinae, Nguyen TT, Taitt AS, Jhun H, Park H-Y, Kim S-H, Kim Y-G, Song EY, Lee Y, Yum H, et al. 2021. SARS-CoV-2 Omicron Mutation Is Faster than the Chase: Multiple Mutations on Spike/ACE2 Interaction Residues. *Immune Netw.* 21:e38.

Kimura, Motoo. 1983. The neutral theory of molecular evolution. Cambridge University Press

Kirchdoerfer RN, Cottrell CA, Wang N, Pallesen J, Yassine HM, Turner HL, Corbett KS, Graham BS, McLellan JS, Ward AB. 2016. Pre-fusion structure of a human coronavirus spike protein. *Nature* 531:118–121.

Kistler KE, Huddleston J, Bedford T. 2022. Rapid and parallel adaptive mutations in spike S1 drive clade success in SARS-CoV-2. *Cell Host Microbe* 30:545-555.e4.

Klinakis A, Cournia Z, Rampias T. 2021. N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. *Comput. Struct. Biotechnol. J.* 19:5556–5567.

Kolbaba-Kartchner B, Kazan IC, Mills JH, Ozkan SB. 2021. The Role of Rigid Residues in Modulating TEM-1 β-Lactamase Function and Thermostability. *Int. J. Mol. Sci.* 22:2895.

Kumar, A., Glembo, T.J., Ozkan, S.B. 2015b. The Role of Conformational Dynamics and Allostery in the Disease Development of Human Ferritin. *Biophys J* 109:1273–1281.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.Battistuzzi FU, editor. *Mol. Biol. Evol.* 35:1547–1549.

Kumar S, Tao Q, Weaver S, Sanderford M, Caraballo-Ortiz MA, Sharma S, Pond SLK, Miura S. 2021. An Evolutionary Portrait of the Progenitor SARS-CoV-2 and Its Dominant Offshoots in COVID-19 Pandemic.Yeager M, editor. *Mol. Biol. Evol.* 38:3046–3059.

Kuzmanic A, Bowman GR, Juarez-Jimenez J, Michel J, Gervasio FL. 2020. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Acc. Chem. Res.* 53:654–661.

Labbadia J, Morimoto RI. 2015. The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.* 84:435–464.

Laiton-Donato K, Franco-Muñoz C, Álvarez-Díaz DA, Ruiz-Moreno HA, Usme-Ciro JA, Prada DA, Reales-González J, Corchuelo S, Herrera-Sepúlveda MT, Naizaque J, et al. 2021. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. *Infect. Genet. Evol.* 95:105038.

Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, Zhang Q, Shi X, Wang Q, Zhang L, et al. 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 581:215–220.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Larrimore KE, Kazan IC, Kannan L, Kendle RP, Jamal T, Barcus M, Bolia A, Brimijoin S, Zhan C-G, Ozkan SB, et al. 2017. Plant-expressed cocaine hydrolase variants of butyrylcholinesterase exhibit altered allosteric effects of cholinesterase activity and increased inhibitor sensitivity. *Sci. Rep.* 7:10419.

Levy RM, Haldane A, Flynn WF. 2017. Potts Hamiltonian models of protein co-variation, free energy landscapes, and evolutionary fitness. *Curr. Opin. Struct. Biol.* 43:55–62.

Liu L, Tamura K, Sanderford M, Gray VE, Kumar S. 2016. A Molecular Evolutionary Reference for the Human Variome. *Mol. Biol. Evol.* 33:245–254.

Liu Xueyan, Liu Xuan, Zhou J, Dong Y, Jiang Wen, Jiang Wenqing. 2022. Rampant C-to-U deamination accounts for the intrinsically high mutation rate in SARS-CoV-2 spike gene. *RNA* 28:917–926.

Liu Y, Bahar I. 2012. Sequence Evolution Correlates with Structural Dynamics. *Mol. Biol. Evol.* 29:2253–2263.

Ma B, Nussinov R. 2016. Conformational footprints. *Nat. Chem. Biol.* 12:890–891.

Maguid S, Fernandez-Alberti S, Echave J. 2008. Evolutionary conservation of protein vibrational dynamics. *Gene* 422:7–13.

Maguid S, Fernández-Alberti S, Parisi G, Echave J. 2006. Evolutionary Conservation of Protein Backbone Flexibility. *J. Mol. Evol.* 63:448–457.

Maher MC, Bartha I, Weaver S, Di Iulio J, Ferri E, Soriaga L, Lempp FA, Hie BL, Bryson B, Berger B, et al. 2022. Predicting the mutational drivers of future SARS-CoV-2 variants of concern. *Sci. Transl. Med.* 14:eabk3445.

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11:3696–3713.

Markov PV, Ghafari M, Beer M, Lythgoe K, Simmonds P, Stilianakis NI, Katzourakis A. 2023. The evolution of SARS-CoV-2. *Nat. Rev. Microbiol.* 21:361–379.

Menni C, Valdes AM, Polidori L, Antonelli M, Penamakuri S, Nogal A, Louca P, May A, Figueiredo JC, Hu C, et al. 2022. Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. *The Lancet* 399:1618–1624.

Mikulska-Ruminska K, Shrivastava I, Krieger J, Zhang S, Li H, Bayır H, Wenzel SE, VanDemark AP, Kagan VE, Bahar I. 2019. Characterization of Differential Dynamics, Specificity, and Allostery of Lipoxygenase Family Members. *J. Chem. Inf. Model.* 59:2496–2508.

Millet JK, Whittaker GR. 2014. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc. Natl. Acad. Sci.* 111:15214–15219.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Millet JK, Whittaker GR. 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* 202:120–134.

Mishra SK, Jernigan RL. 2018. Protein dynamic communities from elastic network models align closely to the communities defined by molecular dynamics.Zhang Y, editor. *PLOS ONE* 13:e0199225.

Modi T, Campitelli P, Kazan IC, Ozkan SB. 2021. Protein folding stability and binding interactions through the lens of evolution: a dynamical perspective. *Curr. Opin. Struct. Biol.* 66:207–215.

Modi T, Ozkan S. 2018. Mutations Utilize Dynamic Allostery to Confer Resistance in TEM-1 β-lactamase. *Int. J. Mol. Sci.* 19:3808.

Modi T, Risso VA, Martinez-Rodriguez S, Gavira JA, Mebrat MD, Van Horn WD, Sanchez-Ruiz JM, Banu Ozkan S. 2021. Hinge-shift mechanism as a protein design principle for the evolution of β-lactamases from substrate promiscuity to specificity. *Nat. Commun.* 12:1852.

Moulana A, Dupic T, Phillips AM, Chang J, Nieves S, Roffler AA, Greaney AJ, Starr TN, Bloom JD, Desai MM. 2022a. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat. Commun.* 13:7011.

Moulana A, Dupic T, Phillips AM, Chang J, Nieves S, Roffler AA, Greaney AJ, Starr TN, Bloom JD, Desai MM. 2022b. Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nat. Commun.* 13:7011.

Moulana A, Dupic T, Phillips AM, Chang J, Roffler AA, Greaney AJ, Starr TN, Bloom JD, Desai MM. 2023. The landscape of antibody binding affinity in SARS-CoV-2 Omicron BA.1 evolution. *eLife* 12:e83442.

Neher RA. 2022. Contributions of adaptation and purifying selection to SARS-CoV-2 evolution. *Virus Evol.* 8:veac113.

Nevin Gerek, Z., Kumar, S., Banu Ozkan, S. 2013. Structural dynamics flexibility informs function and evolution at a proteome scale. *Evol Appl*.

Nielsen BF, Li Y, Sneppen K, Simonsen L, Viboud C, Levin SA, Grenfell BT. 2022. Immune Heterogeneity and Epistasis Explain Punctuated Evolution of SARS-CoV-2. Epidemiology Available from: http://medrxiv.org/lookup/doi/10.1101/2022.07.27.22278129

Nussinov, R., Tsai, C.-J. 2013. Allostery in disease and in drug discovery. *Cell* 153:293–305.

O'Rourke KF, Gorman SD, Boehr DD. 2016. Biophysical and computational methods to analyze amino acid interaction networks in proteins. *Comput. Struct. Biotechnol. J.* 14:245–251.

Ose NJ, Butler BM, Kumar A, Kazan IC, Sanderford M, Kumar S, Ozkan SB. 2022. Dynamic coupling of residues within proteins as a mechanistic foundation of many enigmatic pathogenic missense variants.Wallqvist A, editor. *PLOS Comput. Biol.* 18:e1010006.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ose NJ, Campitelli P, Patel RP, Ozkan SB, Kumar S. 2022. Protein dynamics provide mechanistic insights about the epistatic relationships among highly observed potentially adaptive missense variants. *Biophys. J.* 121:456a.

Otten R, Liu L, Kenner LR, Clarkson MW, Mavor D, Tawfik DS, Kern D, Fraser JS. 2018. Rescue of conformational dynamics in enzyme catalysis by directed evolution. *Nat. Commun.* 9:1314.

Ozono S, Zhang Y, Ode H, Sano K, Tan TS, Imai K, Miyoshi K, Kishigami S, Ueno T, Iwatani Y, et al. 2021. SARS-CoV-2 D614G spike mutation increases entry efficiency with enhanced ACE2-binding affinity. *Nat. Commun.* 12:848.

Patel R, Scheinfeldt LB, Sanderford MD, Lanham TR, Tamura K, Platt A, Glicksberg BS, Xu K, Dudley JT, Kumar S. 2018. Adaptive Landscape of Protein Variation in Human Exomes.Yeager M, editor. *Mol. Biol. Evol.* 35:2015–2025.

de la Paz JA, Nartey CM, Yuvaraj M, Morcos F. 2020. Epistatic contributions promote the unification of incompatible models of neutral molecular evolution. *Proc. Natl. Acad. Sci.* 117:5873–5882.

Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham III TE, DeBolt S, Ferguson D, Seibel G, Kollman P. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput Phys Commun* 91:1–41.

Peters AD, Lively CM. 1999. The Red Queen and Fluctuating Epistasis: A Population Genetic Analysis of Antagonistic Coevolution. *Am. Nat.* 154:393–405.

Qu P, Faraone JN, Evans JP, Zheng Y-M, Carlin C, Anghelina M, Stevens P, Fernandez S, Jones D, Panchal A, et al. 2023. Extraordinary Evasion of Neutralizing Antibody Response by Omicron XBB.1.5, CH.1.1 and CA.3.1 Variants. Microbiology Available from: http://biorxiv.org/lookup/doi/10.1101/2023.01.16.524244

Raghuvamsi PV, Tulsian NK, Samsudin F, Qian X, Purushotorman K, Yue G, Kozma MM, Hwa WY, Lescar J, Bond PJ, et al. 2021. SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets. *eLife* 10:e63646.

Ramarao-Milne P, Jain Y, Sng LMF, Hosking B, Lee C, Bayat A, Kuiper M, Wilson LOW, Twine NA, Bauer DC. 2022. Data-driven platform for identifying variants of interest in COVID-19 virus. *Comput. Struct. Biotechnol. J.* 20:2942–2950.

Rehman S ur, Shafique L, Ihsan A, Liu Q. 2020. Evolutionary Trajectory for the Emergence of Novel Coronavirus SARS-CoV-2. *Pathogens* 9:240.

Rizzato F, Coucke A, de Leonardis E, Barton JP, Tubiana J, Monasson R, Cocco S. 2020. Inference of compressed Potts graphical models. *Phys. Rev. E* 101:012309.

Rochman ND, Faure G, Wolf YI, Freddolino PL, Zhang F, Koonin EV. 2022. Epistasis at the SARS-CoV-2 Receptor-Binding Domain Interface and the Propitiously Boring Implications for Vaccine Escape.Diamond MS, editor. *mBio* 13:e00135-22.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rochman ND, Wolf YI, Faure G, Mutz P, Zhang F, Koonin EV. 2021. Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc. Natl. Acad. Sci.* 118:e2104241118.

Rodriguez-Rivas J, Croce G, Muscat M, Weigt M. 2022. Epistatic models predict mutable sites in SARS-CoV-2 proteins and epitopes. *Proc. Natl. Acad. Sci.* 119:e2113118119.

Rojas Echenique JI, Kryazhimskiy S, Nguyen Ba AN, Desai MM. 2019. Modular epistasis and the compensatory evolution of gene deletion mutants.Butler G, editor. *PLOS Genet.* 15:e1007958.

Saavedra HG, Wrabl JO, Anderson JA, Li J, Hilser VJ. 2018. Dynamic allostery can drive cold adaptation in enzymes. *Nature* 558:324–328.

Saputri DS, Li S, van Eerden FJ, Rozewicki J, Xu Z, Ismanto HS, Davila A, Teraguchi S, Katoh K, Standley DM. 2020. Flexible, Functional, and Familiar: Characteristics of SARS-CoV-2 Spike Protein Evolution. *Front. Microbiol.* 11:2112.

Sekhar A, Kay LE. 2019. An NMR View of Protein Dynamics in Health and Disease. *Annu. Rev. Biophys.* 48:297–319.

Shang J, Wan Y, Luo C, Ye G, Geng Q, Auerbach A, Li F. 2020. Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* 117:11727–11734.

Shimagaki K, Weigt M. 2019. Selection of sequence motifs and generative Hopfield-Potts models for protein families. *Phys. Rev. E* 100:032128.

Shoemark DK, Oliveira ASF, Davidson AD, Berger I, Schaffitzel C, Mulholland AJ. 2022. Molecular dynamics of spike variants in the locked conformation: RBD interfaces, fatty acid binding and furin cleavage sites. Biochemistry Available from: http://biorxiv.org/lookup/doi/10.1101/2022.05.06.490927

Singh D, Yi SV. 2021. On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* 53:537–547.

Spinello A, Saltalamacchia A, Borišek J, Magistrato A. 2021. Allosteric Cross-Talk among Spike's Receptor-Binding Domain Mutations of the SARS-CoV-2 South African Variant Triggers an Effective Hijacking of Human Cell Receptor. *J. Phys. Chem. Lett.* 12:5987–5993.

Starr TN, Greaney AJ, Hannon WW, Loes AN, Hauser K, Dillen JR, Ferri E, Farrell AG, Dadonaite B, McCallum M, et al. 2022. Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. *Science* 377:420–424.

Starr TN, Greaney AJ, Stewart CM, Walls AC, Hannon WW, Veesler D, Bloom JD. 2022. Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains.Mok CKP, editor. *PLOS Pathog.* 18:e1010951.

Starr TN, Zepeda SK, Walls AC, Greaney AJ, Alkhovsky S, Veesler D, Bloom JD. 2022. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* 603:913–918.

Steinhauer DA. 1999. Role of Hemagglutinin Cleavage for the Pathogenicity of Influenza Virus. *Virology* 258:1–20.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Stevens AO, Kazan IC, Ozkan B, He Y. 2022. Investigating the allosteric response of the PICK1 PDZ domain to different ligands with all-atom simulations. *Protein Sci.* 31:e4474.

Sun PA Y, Kollman. 1995. Hydrophobic solvation of methane and nonbond parameters of the TIP3P water model. *J Comput Chem* 16:1164–1169.

Swint-Kruse L, Matthews KS, Smith PE, Pettitt BM. 1998. Comparison of Simulated and Experimentally Determined Dynamics for a Variant of the LacI DNA-Binding Domain, Nlac-P. *Biophys. J.* 74:413–421.

Sztain T, Ahn S-H, Bogetti AT, Casalino L, Goldsmith JA, Seitz E, McCool RS, Kearns FL, Acosta-Reyes F, Maji S, et al. 2021. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* 13:963–968.

Tan ZW, Tee W-V, Samsudin F, Guarnera E, Bond PJ, Berezovsky IN. 2022. Allosteric perspective on the mutability and druggability of the SARS-CoV-2 Spike protein. *Structure* 30:590-607.e4.

Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, et al. 2020. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* 7:1012–1023.

Tay JH, Porter AF, Wirth W, Duchene S. 2022. The Emergence of SARS-CoV-2 Variants of Concern Is Driven by Acceleration of the Substitution Rate.Leitner T, editor. *Mol. Biol. Evol.* 39:msac013.

Teruel N, Mailhot O, Najmanovich RJ. 2021. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants.Dunbrack RL, editor. *PLOS Comput. Biol.* 17:e1009286.

Thye AY-K, Law JW-F, Pusparajah P, Letchumanan V, Chan K-G, Lee L-H. 2021. Emerging SARS-CoV-2 Variants of Concern (VOCs): An Impending Global Crisis. *Biomedicines* 9:1303.

Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D. 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* 181:281-292.e6.

Wang Q, Guo Y, Iketani S, Nair MS, Li Z, Mohri H, Wang M, Yu J, Bowen AD, Chang JY, et al. 2022. Antibody evasion by SARS-CoV-2 Omicron subvariants BA.2.12.1, BA.4 and BA.5. *Nature* 608:603–608.

Wang Q, Iketani S, Li Z, Liu Liyuan, Guo Y, Huang Y, Bowen AD, Liu M, Wang M, Yu J, et al. 2022. Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. *Cell*:S0092867422015318.

Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann H-H, Michailidis E, et al. 2020. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* 9:e61312.

Witte L, Baharani VA, Schmidt F, Wang Z, Cho A, Raspe R, Guzman-Cardozo C, Muecksch F, Canis M, Park DJ, et al. 2023. Epistasis lowers the genetic barrier to SARS-CoV-2 neutralizing antibody escape. *Nat. Commun.* 14:302.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, Graham BS, McLellan JS. 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367:1260–1263.

Wrobel AG, Benton DJ, Xu P, Roustan C, Martin SR, Rosenthal PB, Skehel JJ, Gamblin SJ. 2020. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* 27:763–767.

Wu L, Zhou L, Mo M, Liu T, Wu C, Gong C, Lu K, Gong L, Zhu W, Xu Z. 2022. SARS-CoV-2 Omicron RBD shows weaker binding affinity than the currently dominant Delta variant to human ACE2. *Signal Transduct. Target. Ther.* 7:8.

Xue Q, Liu X, Pan W, Zhang A, Fu J, Jiang G. 2022. Computational Insights into the Allosteric Effect and Dynamic Structural Features of the SARS-COV-2 Spike Protein. *Chem. – Eur. J.* [Internet] 28. Available from: https://onlinelibrary.wiley.com/doi/10.1002/chem.202104215

Yang QE, MacLean C, Papkou A, Pritchard M, Powell L, Thomas D, Andrey DO, Li M, Spiller B, Yang W, et al. 2020. Compensatory mutations modulate the competitiveness and dynamics of plasmid-mediated colistin resistance in Escherichia coli clones. *ISME J.* 14:861–865.

Yue C, Song W, Wang L, Jian F, Chen X, Gao F, Shen Z, Wang Y, Wang X, Cao Y. 2023. Enhanced transmissibility of XBB.1.5 is contributed by both strong ACE2 binding and antibody evasion. Immunology Available from: http://biorxiv.org/lookup/doi/10.1101/2023.01.03.522427

Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, Wang Y, Baum A, Diehl WE, Dauphin A, Carbone C, et al. 2020. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* 183:739-751.e8.

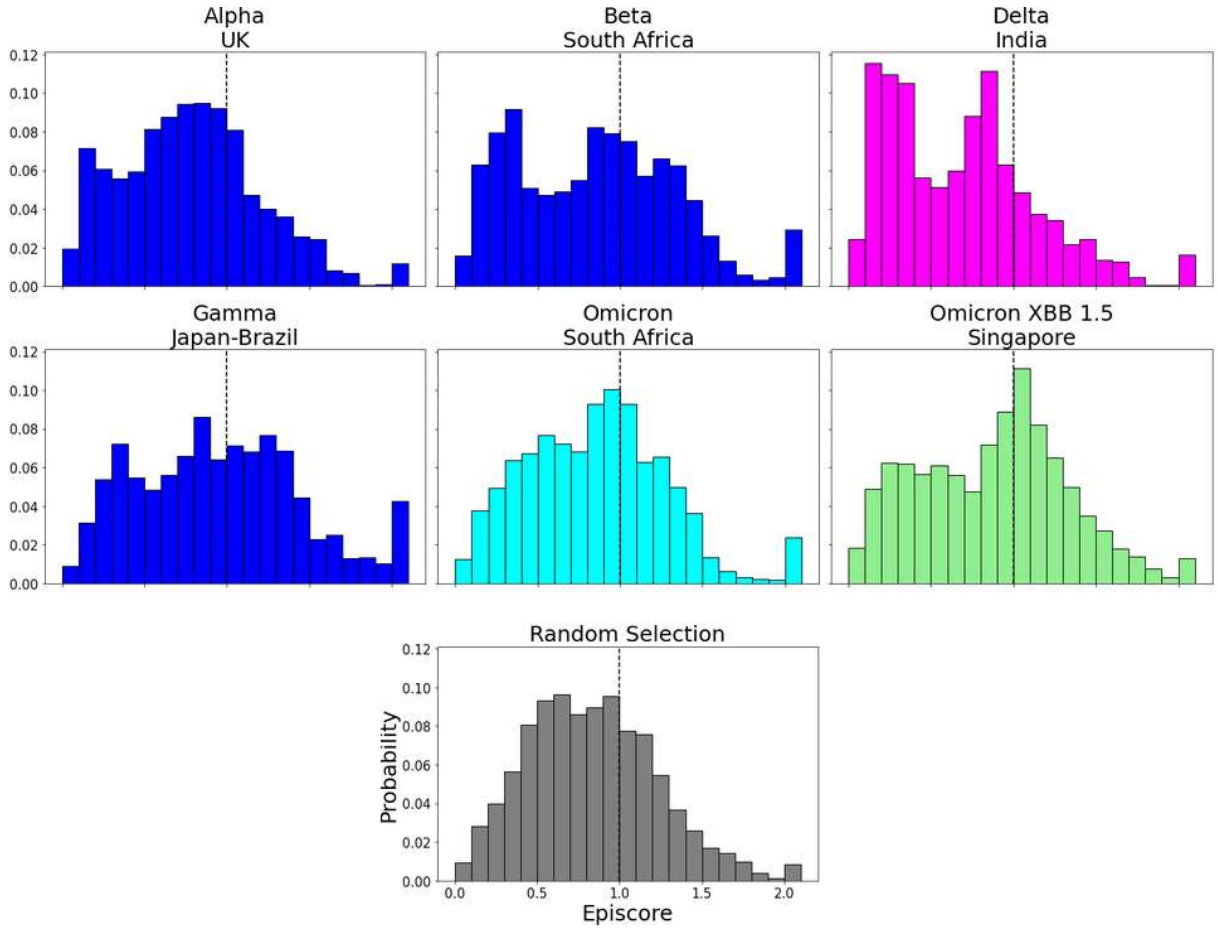
Zeng H-L, Dichio V, Rodríguez Horta E, Thorell K, Aurell E. 2020. Global analysis of more than 50,000 SARS-CoV-2 genomes reveals epistasis between eight viral genes. *Proc. Natl. Acad. Sci.* 117:31519–31526.

Zhang J, Cai Y, Xiao T, Lu J, Peng H, Sterling SM, Walsh RM, Rits-Volloch S, Zhu H, Woosley AN, et al. 2021. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* 372:525–530.

Zhou T, Tsybovsky Y, Gorman J, Rapp M, Cerutti G, Chuang G-Y, Katsamba PS, Sampson JM, Schön A, Bimela J, et al. 2020. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* 28:867-879.e5.

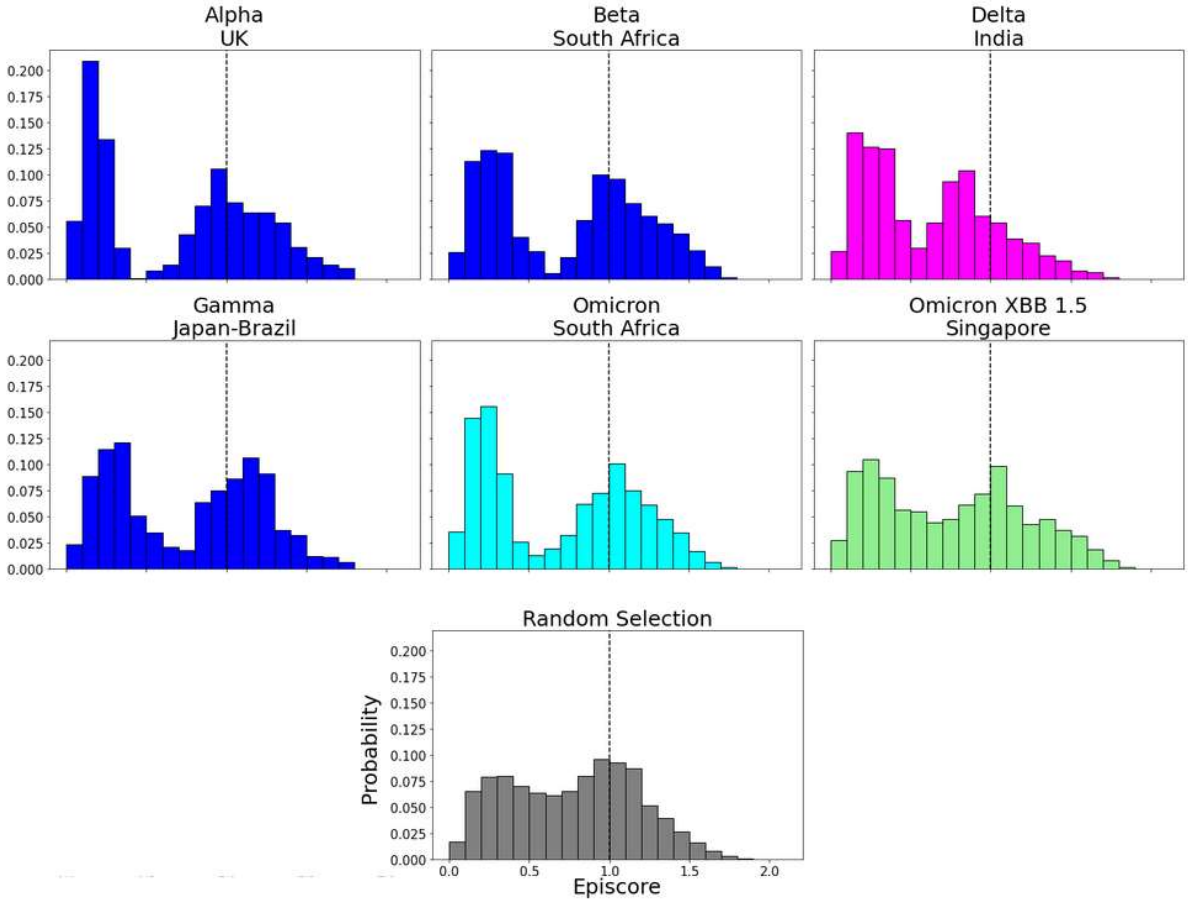
Supplemental Figures

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



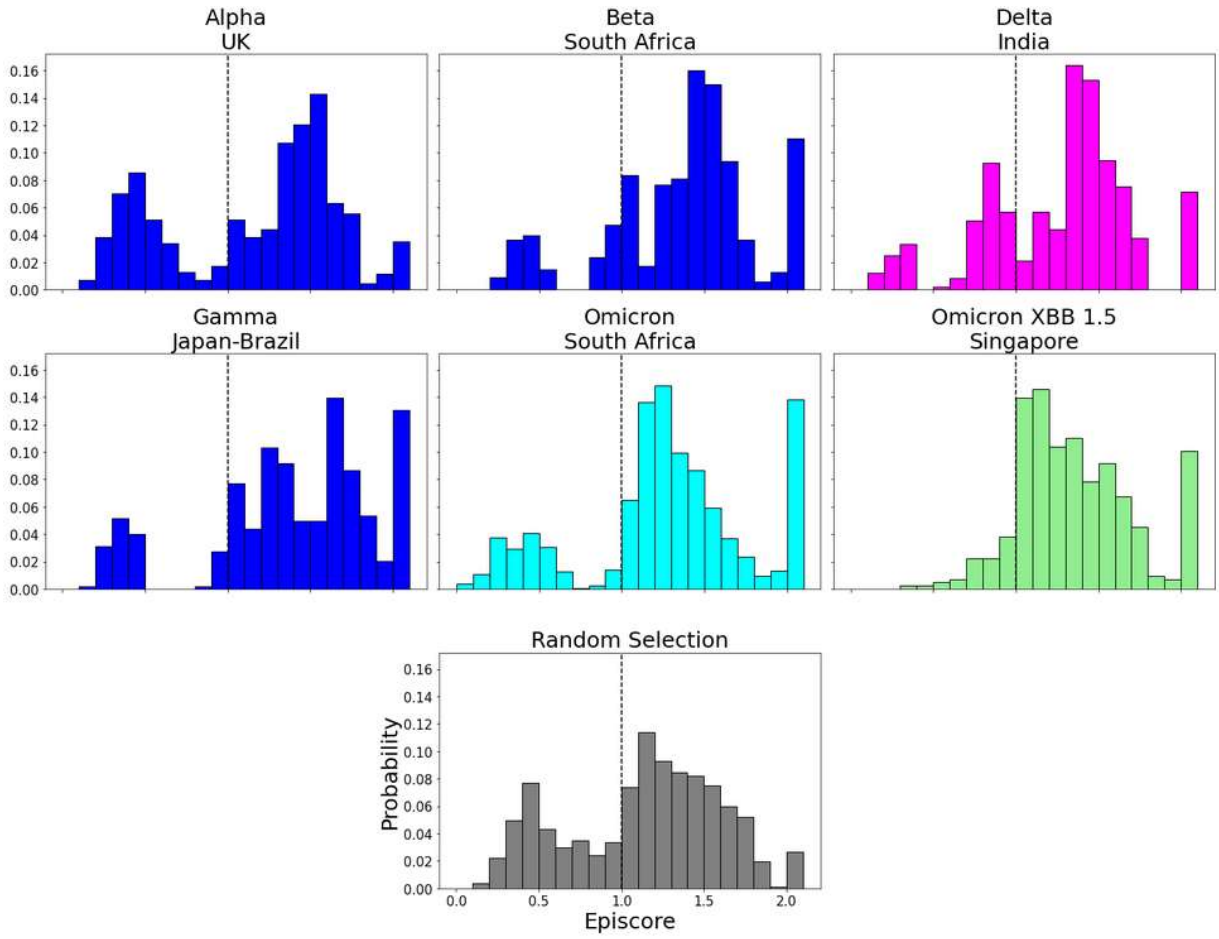
Supplementary figure S1. EpiScores with i = characteristic mutation sites, j = low EP sites, and k = the binding interface of the open chain. EpiScores using variant sites are significantly different ($p<.001$) from a set of EpiScores using random sites.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary figure S2. EpiScores with i = characteristic mutation sites within the NTD, j = low EP sites, and k = the binding interface of the open chain. NTD domain mutation sites result in markedly lower EpiScores compared to elsewhere.

bioRxiv preprint doi: https://doi.org/10.1101/2023.09.14.557827; this version posted September 15, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Supplementary figure S3. EpiScores with i = characteristic mutation sites, j = site 486, and k = the binding interface of the open chain. CAP and hACE2 and antibody binding site 486 displays epistasis with almost all XBB 1.5 variant sites at almost every hACE2 binding site. EpiScores using variant sites are significantly different ($p<.001$) from a set of EpiScores using random sites.