ACS OMEGA

http://pubs.acs.org/journal/acsodf

Review

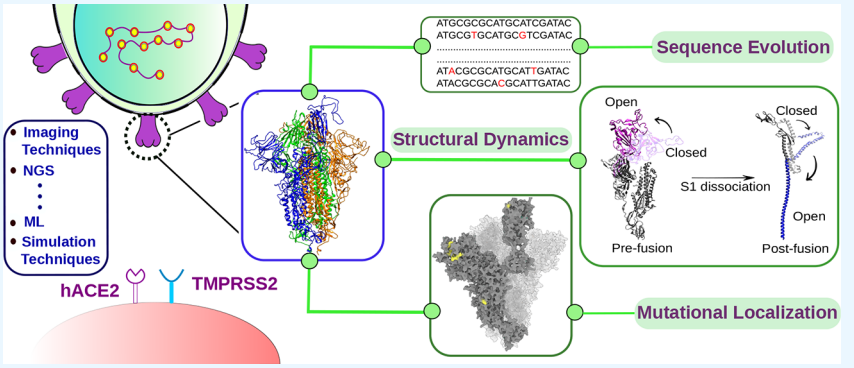
Evolution of Sequence and Structure of SARS-CoV-2 Spike Protein: A Dynamic Perspective

Anushree Sinha,[§] Satyam Sangeet,[§] and Susmita Roy*

Cite This: *ACS Omega* 2023, 8, 23283−23304

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Novel coronavirus (SARS-CoV-2) enters its host cell through a surface spike protein. The viral spike protein has undergone several modifications/mutations at the genomic level, through which it modulated its structure−function and passed through several variants of concern. Recent advances in high-resolution structure determination and multiscale imaging techniques, cost-effective next-generation sequencing, and development of new computational methods (including information theory, statistical methods, machine learning, and many other artificial intelligence-based techniques) have hugely contributed to the characterization of sequence, structure, function of spike proteins, and its different variants to understand viral pathogenesis, evolutions, and transmission. Laying on the foundation of the sequence−structure−function paradigm, this review summarizes not only the important findings on structure/function but also the structural dynamics of different spike components, highlighting the effects of mutations on them. As dynamic fluctuations of three-dimensional spike structure often provide important clues for functional modulation, quantifying time-dependent fluctuations of mutational events over spike structure and its genetic/amino acidic sequence helps identify alarming functional transitions having implications for enhanced fusogenicity and pathogenicity of the virus. Although these dynamic events are more difficult to capture than quantifying a static, average property, this review encompasses those challenging aspects of characterizing the evolutionary dynamics of spike sequence and structure and their implications for functions.

INTRODUCTION

Severe acute respiratory syndrome (SARS) was first identified to cause severe respiratory illness and detected in the Guangdong province of China in November 2002, which turned into a global outbreak.[1,2] In late 2019, it again emerged as SARS-CoV-2 in the city of Wuhan, China, causing a pandemic of acute respiratory disease known as coronavirus disease 2019 (COVID-19).[3] Global research efforts have been spurred to develop specific therapeutics for COVID-19 and vaccines targeting various SARS-CoV-2 encoded proteins. Among other proteins, the SARS-CoV-2 spike (S) protein is a key target for the induction of neutralizing antibodies and the development of T-cell responses during infection. With next-generation sequencing techniques, it is now known that after the Wuhan variant, the viral spike protein has undergone several modifications throughout the course of months of alterations, increasing the virus's ability to infect and elude the immune system at a much faster pace than we could follow.[4−8]

The exceptional and unanticipated genome flexibility of SARS-CoV-2 is frequently linked to the faster rate of mutational adaptation of the viral genome.[9] While evolution is normally a slow time scale phenomenon, such a faster mutational pace of SARS-CoV-2 allows researchers to assort a good collection of mutational data in a relatively less time frame. Thus, depending on the country of origin/1st detection, a number of variants of this virus have been detected, leaving their fresh footprints over its evolutionary trajectory intervened by several variants of interest (VOIs) and variants of concern (VOCs). This

Received: February 12, 2023
Accepted: June 1, 2023
Published: June 21, 2023

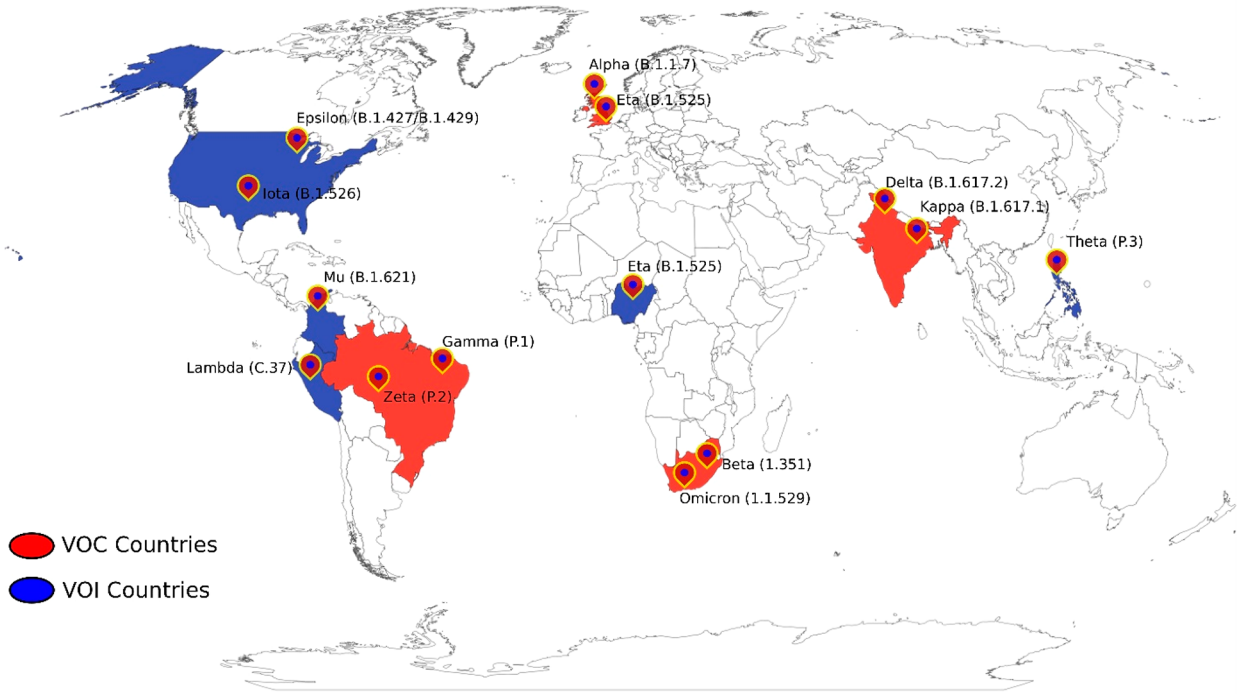


Figure 1. Origin (addressing first detection) map of variant of concern (VOCs) and variant of interest (VOIs). World map representing the country of the emergence of different VOCs and VOIs. The red represents the VOC emergent countries, and the blue represents the VOI emergent countries.

trajectory may carry a critical characteristic evolutionary signal, and a thorough statistical analysis of this dynamic trajectory is of the utmost necessity. Unfortunately, the dynamic processes of mutational changes have been less investigated over the entire viral genome and protein sequence. When this 1D sequence information would translate into 3D structure, how the mutational changes in VOCs would change the dynamic behavior of the corresponding structure/function needs more attention. To understand the multistep SARS-CoV-2 entry process, structural biology and chemistry have been dedicated to determining static snapshots of biomolecules at an atomistic level. However, the dynamic fluctuation within biomolecules retains more information that helps elucidate the functional mechanism through which the virus evolves by changing its infectivity and transmission. Thus, we need to focus on developing and exploiting more such techniques that can accurately quantify dynamic fluctuations within biomolecules, apart from extracting only static structures. Being cognizant of the above facts, this review not only provides a detailed connection between the dynamic evolution of sequence and structure of spike protein but also highlights recent research advancements that characterize both.

EVOLUTIONARY DYNAMICS OF SPIKE PROTEIN AND THE EMERGENCE OF DIFFERENT VOCS

To facilitate the membrane fusion, spike glycoprotein (S protein), the primary surface antigen of SARS-CoV-2, binds to the host-cell receptor hACE2 (human angiotensin-converting enzyme 2) to perform its job.[10−17] The viral spike protein and its other nonstructured proteins (NSPs) have undergone several modifications so far, increasing the virus's ability to infect and elude the immune system. As a result, based on the Pango lineage identification, more than 1300 lineages have been established.[18] 14 divergent clades are used to phylogenetically describe the genetic relationships between this lineages.[19] In addition, SARS-CoV-2's rapid evolutionary divergence during this pandemic is a feature that may be greatly explained by natural selection. Given the observed high rate of mutations, it is imperative to study the dynamics of the virus's transmission and comprehend how the virus changes in concert with the host population and the factors guiding this evolution. It is crucial to look into the mutational pattern and viral spread of over 30,000 base-pair-long genomic sequences of SARS-CoV-2[20] and its variants. Due to the lack of a proofread mechanism and lack of variety, the overall genome of SARS-CoV-2 also seems to exhibit an abnormally high rate of recombination.[21,22] Therefore, genomic level data on various viral proteins including surface glycoprotein domains aids in examining the viral transmission pattern to comprehend the progression of the illness and the effectiveness of administered vaccines and medications.[23,24]

Many methods have been used to study the dynamics of viral transmission and evolution. Here, we emphasize the integration of various methods with genomic research to understand the evolutionary history of the virus.

GENOMIC ANALYSIS OF SARS-COV-2

Genome sequencing enables researchers to categorize a virus as a specific variant and establish its lineage, going beyond testing for SARS-CoV-2. With the emergence of different variants throughout the globe (Figure 1), genomic surveillance has been a crucial aspect of public health initiatives. Numerous single nucleotide variations (SNVs) have been found[25] in the millions of SARS-CoV-2 genome sequences that have been deposited in databases, such as the Global Initiative on Sharing All Influenza Data (GISAID; https://www.epicov.org/),[26,27] National Genomic Data Center of China (https://ngdc.cncb.ac.cn/) and National Center for Biotechnology Information

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

(NCBI; https://www.ncbi.nlm.nih.gov/) databases. Numerous researchers have turned their attention to the functions of certain SNVs in the zoonotic origin, development, and transmission of SARS-CoV-2.[3,28−32] Researchers discovered that the SARS-CoV-2 can be separated into the L and S main lineages during the early COVID-19 epidemic.[25] Because of the presence of leucine and serine residues at positions 8782 and 28144, respectively, the "L" and "S" lineages received their distinct names. Two SNV pairs at sites 8782 and 28144, with practically perfect linkages, were used to distinguish between the L and S lineages. These SNV pairs were C8782/U28144 for L and U8782/C28144 for S, with the reference genome (NC_045512) corresponding to the L lineage. It was discovered that whereas the U-to-C substitution at position 28144 causes leucine to mutate into serine, the C-to-U mutation at position 8782 did not affect the final amino acid. It is crucial to note that when the phylogenetic tree is rooted using the bat and pangolin CoVs as the outgroup, the S lineage is categorized as the ancestral sequence.[33] Based on Forster's nomenclature, SARS-CoV-2 is divided into three classes: A, B, and C, where class A corresponds to the S lineage, and L lineage is further divided into classes B and C. Additionally, GISAID divides the virus into four primary groups: S, L, V, and G based on the S/L and other SNVs. Although a significant portion of the viral genome has been extensively investigated, the distinction between the L and S strains remains a persistent phenomenon. For instance, Tang and colleagues[34] examined nearly 127,119 high-quality SARS-CoV-2 genomic data in 2021 and found that 120,958 (95.15%) belonged to the L lineage, 5950 (4.68%) belonged to the S lineage, and just 211 (0.17%) could not be accurately classified into either of the categories. Moreover, the community transmission of SARS-CoV-2 has been traced using genomic epidemiology, which has shown that the epidemics in Oceania,[35,36] Europe,[37−43] and the Americas[44,45] were the result of numerous distinct introductions followed by the local spread of specific virus strains. These analyses show that by grouping related variants, the genomic sequences of SARS-CoV-2 may be effectively used to capture the lineage evolution as the epidemic develops.

Previous studies[46−49] have shown that the S protein, which is under natural selection and where recombination signals have been recorded, is one of the most variable areas of the SARS-CoV-2 genome. This suggests that this protein evolves continuously and plays a crucial role in adapting to humans. This variation in the S protein is frequently linked to the virus's evolutionary background. Flores-Alanis and groups[50] research on the S gene suggests that the original CoV that gave rise to SARS-CoV-2 most likely originated from a bat rather than a pangolin. Regarding the receptor binding domain (RBD), the study shows that the region required for hACE2 (-RBD) binding is a hybrid between RaTG13 and MP789 CoVs. The SARS-CoV-2 likely acquired the fundamental function of the RBD from the MP789 CoV by recombination with an ancestral CoV, which had a RaTG13 genomic background. The SARS-CoV-2 ancestor then transitioned to humans, where the RBD has undergone extensive conservation and the S protein has been preserved by positive selection. This demonstrates the intricacy of the dynamics of CoV cross-species infection and the significance of CoV genetic exchange explored via genomic analysis. Additionally, Dash and colleagues[51] demonstrated that RBD mutations in the whole genome isolates of 35 patients sequenced by the Regional Medical Research Center (RMRC), Bhubaneswar, India, occurred frequently, indicating a rise in the virus' antigenicity. They discovered through genetic analysis that the frequency with which various mutations in the RBD occur improves the binding affinity of the specific isolate toward the ACE2 receptor. Moreover, Liu and colleagues[52] used the SARS-CoV-2 virus to methodically evaluate the activity of ACE2 orthologs from 48 mammalian hosts, which shed light on the virus's probable host range and interspecies transmission. This study raised the possibility that SARS-CoV-2 may be far more widespread than previously believed, highlighting the need to keep an eye on vulnerable hosts, particularly those with the capacity to transmit zoonoses to stop further outbreaks. Another study[53] showed that the RBD's two sequence sections, "VGGNY" and "EIYQAGSTPCNGV", as well as a disulfide bridge joining 480C and 488C in the extended loop, serve as a strong structural cues for the identification of the hACE2 receptor.

From these studies, one can realize the necessity of data analysis and the preprocessing of genomic sequences to prepare an enormous amount of genetic data for future research. Some studies entirely focus on creating a workflow to preprocess the SARS-CoV-2 genome sequences.[48] There, it was found that the alpha and omicron variants had sequences with 99.76% and 98.47% similarity with respect to the wild-type sequence. However, the sequence similarity of the gamma variant is 96.72%. In addition, the study found that the omicron variant has increased protein diversity only in the spike and nucleocapsid regions. The lower virulence of the gamma variant compared with other VOCs was caused by a higher level of sequence divergence throughout the genomic sequences. Padhan and colleagues[54] showed that even though SARS-CoV-2 evolved from SARS-CoV, it is more closely linked to bat-SL-CoV according to sequence analysis. Their analysis of the early epidemic phase uncovered a number of novel mutations, including those that the receptor binding motif (RBM) in the RBD acquired and those that resulted in the generation of a furin-like cleavage site in the spike protein with the insertion of a proline-arginine-arginine-alanine (PRRA) segment that was not seen in earlier bat-SL-CoV. The presence of a furin-like cleavage site may indicate that SARS-CoV-2 is highly contagious in people due to easy entrance and improved cell attachment.

STUDIES OF 1D SEQUENCE EVOLUTION USING INFORMATION THEORY

In addition to genetic studies, information theory has been used to understand the evolutionary dynamics of SARS-CoV-2. To distinguish between the SARS-CoV-2 subtypes, highly informative genomic variation sites known as informative subtype markers (ISM) have been identified.[55] A genome-wide Shannon entropy and site selection (dN-dS) analysis was carried out by Ghanchi and associates[56] to assess the genomic diversity between SARS-CoV-2 genomes. In early pandemic strains compared with later strains, they found more entropy and diversity, which suggests that the genomes became more stable in succeeding COVID-19 waves. This saturating entropic behavior would probably result in the virus choosing site-specific alterations that are beneficial to it.

To follow the evolution of SARS-CoV-2, various efforts have been made to set up an epidemiological surveillance system using information theory due to the S protein gaining many mutations. Recently, we showed how crucial it is to analyze domain-wise mutational entropy to comprehend the evolution

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

of viruses.[57] We described a new responsive quantity called mutational response function (MRF) that, based on its evolutionary database, accurately quantifies domain-wise average entropy-fluctuation in the spike glycoprotein sequence of SARS-CoV-2. We found that the evolutionary crossover from a particular variety to a VOC is accompanied by a significant shift in MRF, maintaining the hallmark of a dynamic phase transition. The details of the MRF description can be found elsewhere.[57] Using viral genomic and protein sequences, our method creates a pipeline that can be used to follow the evolution of any variant. Additionally, VOCs have been predicted using mutational entropy. Zhao and colleagues[58] showed how to model competition between many variations using mutational entropy to determine which lineage will develop into a VOC and when this would emerge. For the alpha variation (B.1.1.7), they utilized a threshold value of mutational entropy (4.25) to gauge the rate of evolution for various variants. Instead of using phylogenetic analysis, their approach relies on sequence mutation, which sheds light on the rate of evolution. They also divided the formation of various VOCs into stages, allowing them to analyze the rivalry between several versions throughout the history of a certain VOC.

Besides the 1D sequence, VOCs have left their impression on their respective 3D structures and structural dynamics. Therefore, how the characteristic mutations of VOC change the structure−function mechanism from that of the Wuhan variant remains an extensive area of research since the emergence of this virus. The following sections discuss a 3D structural overview of S/spike protein and mutational response on the spike structure.

STRUCTURAL OVERVIEW OF THE SPIKE PROTEIN OF SARS-COV-2

Spike protein of SARS-CoV-2 is a homotrimer, in which each chain consists of 1273 residues. In the prefusion conformation, the spike protein comprises the S1 head domain and S2 subunit, where S1 wraps around the twisted S2 bundle. The interaction of the spike and the host cell receptor is mediated by the RBD which is an intrinsic part of the S1 head. Apart from RBD, the other critical functional components are the N-terminal domain (NTD), carboxy-terminal domain-1 (CTD1), carboxy-terminal domain-2 (CTD2), a loop called 630 loop which functions as a connector between these two domains (CTD1 and CTD2). The S2 subunit consists of the FPPR (fusion peptide proximal region) loop, heptad repeat 1 (HR1), central helix (CH), heptad repeat 2 and transmembrane. This HR1 bends back toward the viral membrane, forming a helix bundle in conjunction with CH (Figure 2). Table 1 shows the residue range of different domains of a single chain of the spike protein.

The structural features of these different components of the spike protein of SARS-CoV-2 and the characteristic functionalities are summarized in Table 2.

STRUCTURAL DYNAMICS OF SPIKE PROTEIN

Each monomer of the trimeric spike structure dynamically switches between the S1-head-up and S1-head-down conformational states.[11,14,66−69] For the trimer, altogether, four distinct conformational states have been found in the prefusion state: (i) 3 down, (ii) 1 up-2 down, (iii) 2 up-1 down, and (iv) 3 up. These four states are primarily driven by distinctive intra-

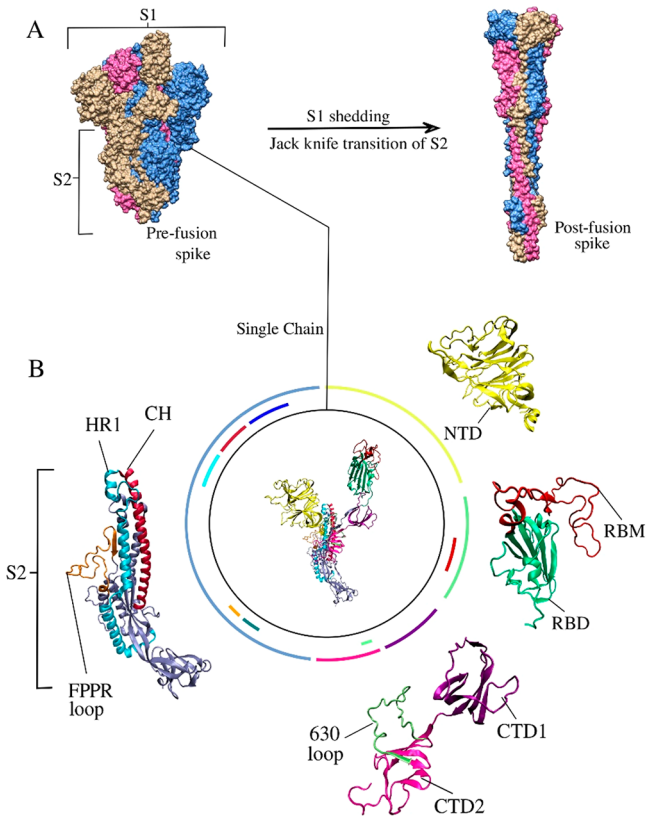


Figure 2. Structural overview of spike protein with several of its key functional segments. (A) Structural transformation of the spike protein during its transition from prefusion to postfusion (having S1 dissociation followed by S2 elongation). (B) The entire circle corresponds to a single chain full-length spike protein. Each of the colored arcs in the circle represents different segments of the spike including NTD (in yellow), RBD (in light green), RBM (in crimson red), CTD1 (in purple), CTD2 (in magenta), 630 loop (in mint green), FPPR loop (in orange), HR1 (in cyan), and CH (in deep red).

Table 1. Domain Segmentation of Spike Glycoprotein of SARS-CoV-2

Domain	Residue Range
NTD	13−304
RBD	319−521
CTD1	522−590
630 loop	620−640
CTD2	591−671
Junction	672−709
S2	710−1273
FPPR loop	828−853

and interchain interactions. The symmetric states are defined as 3 up and 3 down, whereas the remaining states are referred to be asymmetric. These findings are supported by several computational studies.[59,70−75] The spike uses hACE2 as the host receptor.[68,69,76−79] hACE2 accessibility is mediated by the S1 head-up conformation, whereas the hACE2 inaccessible state is mediated by the S1 head-down conformation. Such a connection between RBD and ACE2 has also been characterized by several experimental and computational studies.[12,80−84] In a recent study,[85] the conservation of ACE2 and its ability to act as a receptor by SARS-CoV-2 has been

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

Table 2. Overview of Different Components of the Spike Protein of SARS-CoV-2 with Their Respective Characteristics and Functionalities

Components	Characteristics	Functionality
N-terminal domain (NTD)	• Comprises of four stacked-β-sheet, linked together with bendable loops.	• Provides assistance to the up conformation of RBD of neighboring protomer.[59] Serves as a target for significant neutralizing antibodies.[60,61]
Receptor binding domain (RBD)	• Consists of one of the components as an extended loop called the receptor-binding motif (RBM). • RBM wraps around one terminal of the core structure of the RBD. • The core component of RBD is comprised of five-stranded antiparallel β sheets.	• RBD controls the accessibility of the spike to the host cell receptors. RBM directly potentiates engagement with ACE2.[11] Indeed, a main target for antibodies.[62]
Carboxy-terminal domain 1 (CTD1)	• Primarily formed with β-sheets. • Packed beneath RBD.	• Adjusts to accommodate RBD movements. Act as a structural relay between RBD of the same protomer and FPPR loop from neighboring protomer.[10]
Carboxy-terminal domain 2 (CTD2)	• Comprises two stacked four-stranded β sheets. • Connects NTD and RBD. • Consists of a furin cleavage site at the S1−S2 boundary.	• Crucial component for the structural changes in the S protein needed for membrane fusion.[63]
630 loop	• An IDR-like loop in CTD2 joining CTD1 and CTD2. • Disordered in Wuhan-1-Hu strain, ordered in D614G variant. • Has order−disorder transition in a few VOCs.	• Protects hydrophobic interface of CTD2 from solution exposure. Prevents premature S1 shedding. Increases stability of the trimer. The ordered state gives rise to an RBD intermediate in a few variants.[63]
FPPR loop	• An IDR-like component of S2. • Interacts with neighboring chain CTD1. • Ordered in Wuhan-1-Hu strain, also ordered in D614G variant. • Has order−disorder transition in a few VOCs.	• Ordered loop helps to clamp RBD down. New interactions between the FPPR loop and CTD1, in strains other than Wuhan-1-Hu, strengthen RBD up posture.[64]
S2	• In prefusion a coiled-coil, HR1, and another helix stay together to form a nine-helix bundle. The bottom of the coiled coil is encompassed by HR2 and the connector region, which together protect the fusion peptide in part. • In postfusion the central three-stranded coiled coil formed by HR1 and the central helix is exceptionally long. The C-terminal region of HR2 adopts a longer helix that forms a six-helix bundle structure with the rest of the HR1 coiled-coil, resulting in a rigid postfusion structure. The N-terminal region of HR2 adopts a helix with one turn conformation and packs against the groove of the HR1 coiled coil structure.	• The same strategy employed by other coronaviruses[16,65] to induce viral and host cell membrane fusion is used in the postfusion of SARS-CoV-2, where the extended six helix bundle[10] aids in injecting the host cell membrane with fusion peptide.

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

studied extensively, where ACE2 sequence of 410 vertebrates and 252 mammals has been included in the cohort. Our study mainly focuses on the interaction of the virus with the hACE2 receptor.

Coronaviruses gain entry into cells through the binding of viral spike proteins to the host-cell receptors, followed by the subsequent priming of the spike protein by host-cell protease. The binding is also followed by cleavage at two cut sites: S1/S2 (PRRAR685) and S2′. This cleavage causes conformational changes in the protein's structure, enabling the virus to fuse with the host cell membrane. However, TMPRSS2 is a type of host cell serine protease with an extracellular protease segment that can cleave the spike protein of SARS-CoV-2, thereby initiating membrane fusion attacking through the S2′ region.[86−89] Therefore, understanding the interaction between the spike protein, TMPRSS2, and ACE2 is crucial in the development of future drugs and therapeutics to combat this virus.

The structure of TMPRSS2 consists of two distinct domains: a N-terminal activation domain and a C-terminal proteolytic or catalytic domain. The N-terminal segment faces the transmembrane of the host cell, while the C-terminal section is directed outward and is designed to interact with target peptides, the spike protein. TMPRSS2 encourages cell-to-cell fusion, even in the absence of furin-mediated cleavage at the S1/S2 site. In a recent study,[90] the structure of TMPRSS2 has been found with its individual catalytic triad, which potentially contribute to immune evasion and cell infectivity. Along with the recognized and functionally active S2′ site (residue: 806−814, an extended loop region in the outward direction and solvent exposed), two additional potential cleavage sites in the S2 domain are T1 (residue: 837−845) and T2 (residue: 976−986). These two domains have been identified as TMPRSS2 serine protease target sites. It has been observed that among these two, TMPRSS2 cleaves the spike protein predominantly in the S2′ area (KPSKR8425SFIED). The primary interaction between the spike protein and TMPRSS2 is facilitated by the classical protease binding mode, where this interaction involves a burial of the interfaces.

Research has shown that the furin protease enzyme is responsible for initiating the cleavage of the spike protein in the S1/S2 region, resulting in the cleavage of the spike protein into S1 and S2. This cleavage leads to exposure of the S2 domain and TMPRSS2 recognition sites. The cleavage sites that have been detected are crucial for TMPRSS2 to cleave the S2 domain of the spike protein, which enables viral entry into the host cell through membrane fusion. Since this epitope remains conserved, it can be a potential target for developing vaccines and therapeutic drugs. Moreover, in order to inhibit the TMPRSS2 activity and its interaction with the S2 subunit of the spike protein, researchers have tested various clinically approved protease inhibitors such as Chemostat, Upamostat, Nafmostat, and Bromhexine.[90]

Thus, by preventing the cleavage activity in the particular segments of the spike proteins by blocking the interaction of the spike with TMPRSS2 using TMPRSS2 inhibitors, the engagement of the protein with the host cell can be terminated. These insights may contribute to the development of TMPRSS2 inhibition-related medications as well as potential intervention techniques.

THE ROLE OF GLYCANS IN THE STRUCTURE−FUNCTIONS OF SPIKE PROTEIN

With the internal structural components of the spike protein, it also becomes crucial to recognize how the structure is protected by N-linked glycans.[66] It is proven that robust glycosylation coating on the SARS-CoV-2 spike is essential for the spike to evade the immune system and activate the prefusion mechanism, facilitating protein folding.[91−93] The glycan shield plays an important role in the virus's pathogenesis, and the evolution of glycosylation sites may be linked to protein sequence evolution to optimize its function. Each trimer of the spike protein has 66 N-linked glycosylation sites comprising of over 50% of the protein's molecular weight. Using a combined site-specific mass spectrometric approach and Cryo-EM analysis, Watanabe and his group[66] demonstrated how the N-linked glycans obstruct particular regions of the spike protein's surface and also provides a comprehensive precedent of site-specific glycan, highlighting the features in a trimeric spike that is natively folded. The SARS-CoV-2 virus contains a gene that is supposed to direct the production of 22 N-linked glycan sequences for each protomer. Interestingly, none of these glycan sites in the spike have undergone any mutations. Out of the 22 sites, 8 are primarily composed of oligomannose-type glycans, while the remaining 14 sites mostly contain complex-type glycans that have been processed. Approximately 52% of these glycosylation sites have fucose molecules attached, and 15% of the glycans comprise at least one sialic acid residue. The spike structure of SARS-CoV-2 contains a combination of oligomannose and complex-type glycans at six specific sites (N61, N122, N606, N717, N801, and N1074). The proximity of certain glycosylation sites (N165, N234, N343) can shield the receptor binding sites on the spike, particularly when the RBD is in a down conformation. Oligomannose-type glycans are found throughout the S1 and S2 subunits of the spike and are likely protected by the protein component, as seen in the N234 glycan, which is partially wedged between the NTD and RBD. The membrane-proximal C terminus (N1158, N1173, N1194) and the extended flexible loops structures (N74 and N149), which were not resolved in the Cryo-EM microscopy maps, have both been classified as having N-linked glycans.[66]

Studies have reported that some of the N-glycan sites within the spike structure have been observed to play a role in regulating the functionality of the spike. The glycan shielding has been investigated for both the up and down states of the receptor binding domain.[94] During the release of the receptor binding domain, a sudden decrease in glycan shielding has been observed. The N343 on the receptor binding domain has been identified as a crucial residue that participates in the RBD opening mechanism by acting as a "glycan gate" that pushes the RBD from the down to the up conformation. Conversely, when the RBD is in the down state, glycans at positions N165, N234, and N343 shield the RBD of the SARS-CoV-2 spike.[66] The receptor binding motif (RBM) is consistently shielded by the N165 and N234 glycans, while the shielding by the N343 glycan decreases with RBD opening.[94] In addition to the shielding, the N165 and N234 have been found to play a crucial role in maintaining the RBD up confirmation through their structural support, as demonstrated by the biolayer interferometry experiment.[72] The removal of these two glycans via N165A and N234A mutations has been shown to result in a significant reduction in ACE2 binding, which results in the

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

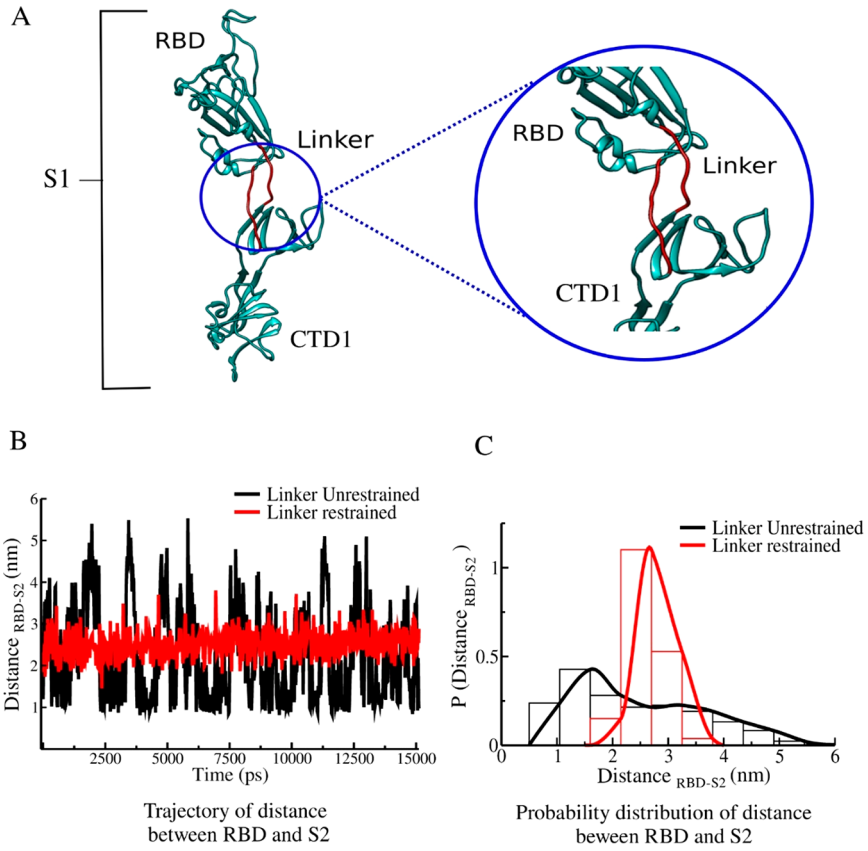


Figure 3. The linker region and its control over the up−down movement of the spike head. (A) Structural overview of the linker joining the RBD and CTD1 in the S1 head. (B) Trajectory of the distance between the RBD and S2 representing the unrestrained linker allows the transition of the RBD between two states, up and down (black curve), whereas the restrained linker inhibits the swing of the RBD keeping it stay at one specific conformation (red). (C) The probability distribution of the distance trajectory representing the event of RBD transition in linker unrestrained (black, bimodal nature indicating a transition between up and down state) and linker restrained condition (red, unimodal nature signifies one specific conformation of the RBD, up/down).

conformational shift of the RBD toward the down state. It has been observed that the alteration in the topology of the glycan shield around N370 and the acquisition of N-glycosylation at N234 may contribute to the enhanced infectivity of the SARS-CoV-2 and make it more contagious compared to other closely related coronaviruses.[95]

Furthermore, a recent study[96] demonstrates a noteworthy characteristic of these glycans, namely that their steric composition can cause a delay in the spike protein's conformational shift on the way to the postfusion pathway. The fusion peptides have a crucial chance to seize the host cell during this glycan-induced delay. On the other hand, the entry of the viral particle into the host is unlikely to occur in the absence of glycans. Their approach provides a significant structural foundation for understanding the dynamics of this widespread disease, while demonstrating how the glycosylation state might influence infectiousness.

LINKERS CONTROLLING THE UP−DOWN MOVEMENT OF RBD

The junction between the two lobes of the S1-head domain is a connector region that accommodates two linkers (linker 1: 329−337, linker 2: 526−535) (Figure 3A). The S1 head's up−down dynamics are regulated by these noteworthy linkers. If the position of two flexible linkers is restrained, providing a harmonic bias potential (simulation details and modeling information have been provided in the Supporting Information), the spike head's dynamics can be blocked. It will stay in its up state if the initial state is an up state, and vice versa. This simple but significant finding can help to understand how such linkers in a spike chain can govern a crucial event of the biomolecular prefusion machinery.

The correlation between linker dynamics and RBD movement is demonstrated by the aforementioned finding (Figure 3B,C). When the linker was unrestrained, the RBD was able to provide two states with bistability in its population. However, a constrained linker restricts RBD movement, which causes the latter to have a stiff, fixed conformation.

POSTFUSION MECHANISM

After ACE2 binds to the RBD, S1 dissociation occurs as a result of cleavage at the S2′ site, a second cleavage site of S2 stalk,[12,97] followed by severance at the S1/S2 boundary[98] which leads to the initiation of the membrane fusion mechanism.[96,97] The SARS-CoV-2 spike adheres to the same postfusion mechanism as it was identified in SARS-CoV.[65] In postfusion, the spike protein functions as a loaded spring that is released following S1 dissociation at the S2′ location. According to investigations where the Cryo-EM structure of the postfusion state has been resolved, the S2 segment of the structure goes through a "jack knife"-style transition (Figure 2A). Here, HR1 (heptad repeat-1, a vital segment of S2) and CH (central helix) get elongated and combine to create an exceptionally long, three-stranded coil (∼80A), which causes

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

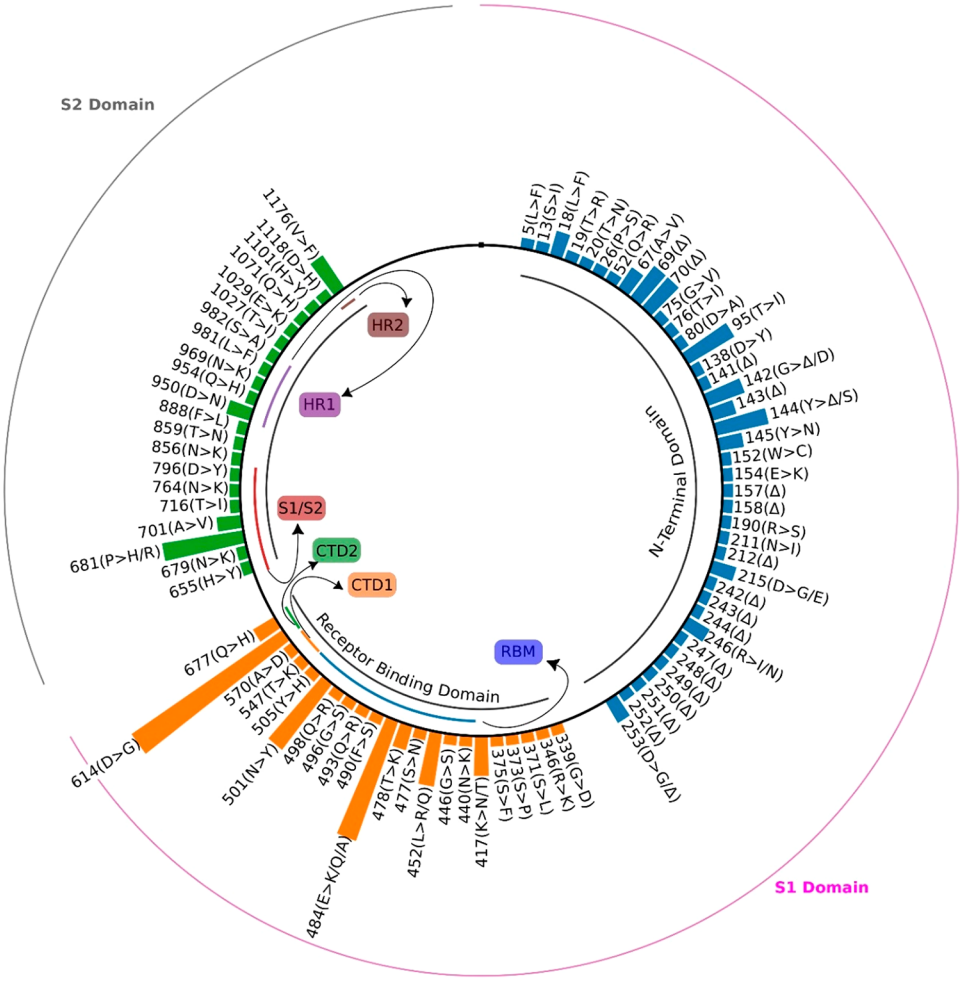


Figure 4. Circos plot of the occurrence of mutations represented in histograms. Each color signifies several motifs of the spike as labeled in the figure. The histogram bars correspond to the frequency of occurrence of a specific mutation in the spike counted in all of the variants. This plot shows largely occurring mutations for particular amino acids such as Δ69/70, T95I, G142Δ/D, Y142Δ/S, L452R/Q, E484K/Q/A, N501Y, D614G, P618H/R.

the firmly bound stem region of S2 to lengthen. The three-stranded β sheet that is formed by the connector domain and a section of the S1/S2−S2′ fragment is constant in both the prefusion and postfusion structures. Residues 1127 to 1135 link the connector β-sheet in the postfusion state, expanding it into four strands and directing the C-terminal HR2 toward the viral membrane. Three helical sections in the S1/S2−S2′ fragment (residues 737−769), which pack against the groove of the CH part of the coiled coil to form a brief six-helix bundle structure, are bound together by two disulfide bonds. The C-terminal section of HR2 forms a longer helix that, along with the rest of the HR1 coiled-coil, forms the second six-helix bundle structure. The N-terminal region of HR2 adopts a one-turn helical conformation and packs against the groove of the HR1 coiled coil. The result is an extremely rigid architecture that has the long central coiled-coil reinforced repeatedly along its long axis.[10] Additionally, it has been determined through comparative sequence analysis, structural evaluation, and molecular dynamics simulation[99] that an immunologically significant epitope (residue range 662−671, sequence CDIPIGAGIC) in the spike protein, which has been shown to be the target of neutralizing antibodies,[100,101] is structurally conserved across the substantial variants of the SARS-CoV-2. Moreover, it has been revealed that the characteristics of this epitope are unaffected by mutations linked to the VOCs. The epitope is located outside of both the NTD and RBD,[102] more precisely it resides extremely close to the furin cleavage site. A vaccine that targets this domain may be beneficial in halting the S1 dissociation process, which may therefore prevent the initiation of the postfusion mechanism and reduce the virus's ability to inject the viral genome into the host cell. The epitope is perforated to solvent in both the close and open conformation of the spike protein, which is reinforced by SASA (solvent accessible surface area) and is indeed a structural feature that is crucial for vaccine development because epitope access is necessary for antibody binding. The empirical results suggest that the 662−671 epitope may be a favored motif for the development of the COVID-19 vaccine and perhaps a prospective target for irresistible antibodies.

MUTATIONAL RESPONSE ON THE STRUCTURE−FUNCTION OF SPIKE PROTEIN

The spike glycoprotein of SARS-CoV-2 has undergone multiple mutational evolutions, leading to the emergence of several VOCs and VOIs.[103] Each variant spike evolved with additional significant structural modifications likely to affect virus pathogenesis. These mutation-mediated structural changes affect not only antigenic properties but also the

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

Table 3. Detailed Overview of Mutational Data Pertaining to the Spike Protein of Various Variants of SARS-CoV-2[a]

VOC	Origin of first detection	NTD	RBD	CTD1	CTD2	Junction	S2
Alpha (B.1.1.7)	United Kingdom, November 2020	Δ69, Δ70, Δ144	N501Y	A570D	D614G	P681H	T716I, S982A, D1118H
Beta (B.1.351)	South Africa, October 2020	L18F, D80A, D215G, Δ242, Δ243, Δ244, R246I	K417N, E484K, N501Y	NIL	D614G	A701V	NIL
Gamma (P.1)	Japan, November 2020	L18F, T20N, P26S, D138Y, R190S	K417T, E484K, N501Y	K417T, E484K, N501Y	D614G, H655Y	NIL	T1027I, V1176F
Delta (B.1.617.2)	India, October 2020	T19R, Δ157, Δ158	L452R, T478K	NIL	D614G	P681R	D950N
Omicron (B.1.1.529)	South Africa, November 2021	A67V, Δ69, Δ70, *T95I*, *G142D*, Δ143, Δ144, Δ145, *N211I*, Δ212, 215EPEins	G339D, S371L, S373P, S375F, K417N, N440 K, G446S, S477N, T478 K, E484A, Q493R, G496S, Q498R, Y505H	T547K	D614G, H655Y	N679K, P681H	N764K, D796Y, N856 K, Q954H, N969K, L981F
Omicron Subvariant (*BQ*.1.8)[b]	India, October 2022		K444T, N460K				
Omicron Subvariant (*BF*.7.1)[b]	India, July 2022	G261V	R346T				
Omicron Subvariant (*XBB*)[b]	Singapore, August, 2022	V83A, H146Q, G252V	F486S				

VOI	Origin of first detection	NTD	RBD	CTD1	CTD2	JUNCTION	S2
Epsilon (B.1.427, B.1.429)	USA, March, 2022	S13I, W152C	L452R	NIL	D614G	NIL	NIL
Zeta (P.2)	Brazil, April, 2020	NIL	E484K	NIL	D614G	NIL	V1176F
Eta (B.1.525)	UK and Nigeria, December, 2020	Q52R, A67V, Δ69, Δ70, Δ144	E484K	NIL	D614G	Q677H	F888L
Theta (P.3)	Philippines, February, 2021	Δ141, Δ142, Δ143	E484K, N501Y	NIL	D614G	P681H	E1092K, H1101Y, V1176F
Iota (B.1.526)	New York City, November, 2020	T95I, D253G	E484K	NIL	D614G	A701V	NIL
Kappa (B.1.617.1)	India, December, 2020	T95I, G142D, E154K	L452R, E484Q	NIL	D614G	P681R	Q1071H
Lambda (C.37)	Peru, August, 2020	G75V, T76I, Δ246 Δ247, Δ248, Δ249 Δ250, Δ251, Δ252	L452Q, F490S	NIL	D614G	NIL	T859N
Mu (B.1.621)	Colombia, January, 2021	T95I, Y144S, Y145N	R346K, E484K, N501Y	NIL	D614G	P681H	D950N

[a]Δ = Deletion at that position. Nil = No mutation. L18F = Leucine (L) at position 18 is getting mutated to phenylalanine (F) and the same for other mutations. [b]The mutational data shown in these variants include all the mutations that have emerged with omicron (B.1.529), only the new characteristic key mutations are added to the table.

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

functionality of the virus. They have functional roles in enhancing transmissibility by making it more resistant to antibody neutralization.[104] Moreover, the RBD and ACE2 binding affinity has also changed due to some effective mutational alterations that eventually result in increased binding capacity of the RBD in the prefusion state.[105,106]

The variants that appeared to spread more efficiently than the one in the initial outbreak (the strain Wuhan-Hu-1[107]) were revealed to be more lethal. We displayed the distribution of all of the mutations (Figure 4) that have occurred since the genesis of the variants and noted that certain mutations occur very often. We term those crucial residual positions as recurrent mutation regions, as they have a high probability of getting mutated among the variants. In this section, we will discuss the mutations that most frequently occurred among the variants and their impact from a structural standpoint. The residue range of important domains of SARS-CoV-2 is listed in Table 1. We have tabulated the mutational data corresponding to VOCs and VOIs (Table 3).

A few key mutations, including deletion mutations, have been examined and determined to significantly impact the structural and functional aspects, which are discussed in the following sections.

Δ69/Δ70. The 6-nucleotide deletion at positions 69 (histidine) and 70 (valine) in the spike protein is caused by the 69/70del mutation in the SARS-CoV-2 spike gene. These two residues are in a disordered loop in the NTD of the spike structure. The 69/70del mutation, primarily observed in the alpha variant (B1.1.7 lineage), has been recently found in the omicron variant (lineage B1.1.529). In a study of the B.1.1.7 lineage, it was found that H69/V70 deletion boosts the spike infectivity and cleavage of S2, it does not grant immunity to antibodies. These deletions also increase the incorporation of cleaved spikes into virions and cause them to be more infectious.[108] BA.1 and BA.3 are the omicron groups with the 69−70 deletion in the spike protein's NTD, leading to S-gene target failure. In recent studies, this advantageous property has been reported for the omicron variant to gain a strong escape tendency against the defenses of the immune system which results in it being more transmissible.[109,110]

D614G. The spike protein of SARS-CoV-2 appears to promote rapid viral spread by replacing aspartic acid (D) at position 614 with glycine (G) in the carboxy(C)-terminal region of the S1 domain. The D614G mutation is possibly one of the best examples of the effects of mutational alterations in the SARS-CoV-2 spike protein. It is retained in all the VOCs and VOIs of SARS-CoV-2 that have emerged to date. After the Wuhan-Hu-1 strain, this single point mutation D614G alone caused a threat with increased infectivity and resulted in the variant B.1, known as the Italy variant of the SARS-CoV-2 virus. This was first identified in March 2020.[111,112] Although there was no significant change in the 614G spike structure, this point mutation affected the functions of a few structural segments called the 630 loop and FPPR loop, thereby modulating the structural rearrangement and the stability of the trimer.[113] Each protomer adopts three distinct prefusion conformations in these mutants due to this mutation, which differ principally in the position of one RBD.[63,114]

In contrast to 614D S trimer, which has a disordered 630 loop, the 614G S trimer has a wedge-shaped 630 loop that inserts between the same protomer's NTD and CTD1. This insertion causes shielding of the hydrophobic residual interface of CTD2, along with the interaction between residues from NTD and the loop itself. In 614D S, this interface was solvent exposed; however, in 614G, it is buried. The CTD2 is stabilized due to the blocking of the exposed hydrophobic surface, inhibiting the release of the N-terminal segment of S2, which helps in halting premature shedding of S1, thereby improving the stability of the S trimeric protein.[63,115] In addition, this phenotype of the D614G mutation causes the FPPR to have an order−disorder transition unlike in D614. Furthermore, when ACE2 binds to the RBD-up conformation, then both the 630 loop and the FPPR loop are expelled from their locations. This FPPR loop shift may assist in exposing the S2′ site adjacent to the fusion peptide for proteolytic cleavage in the closed S trimer conformation which appears to influence the membrane fusion event.[10,115,116]

E484K. The RBD of the spike protein contains the glutamate (E) to lysine (K) substitutions at position 484 (E484K), which is found in the rapidly spread VOCs beta and gamma, also in a few VOIs. The E484K mutant can avoid vaccination-induced antibodies. This enables resistance to neutralizing antibodies, potentially weakening the protective effects of vaccine-induced immunity. Monitoring the E484K mutation, which is found in the viral receptor binding domain and has been independently surfacing in various SARS-CoV-2 variants, is crucial.[117,118]

Nonaka and coauthors have also studied the E484K spike mutation. They reveal a case of SARS-CoV-2 reinfection from different lineages in Brazil, including the E484K mutation, a variant linked to the virus evading neutralizing antibodies. Their findings of SARS-CoV-2 reinfection with an E484K variant support in vitro and in silico studies that predict the potential for lineages carrying this mutation to evade neutralizing antibodies. These studies emphasize the significance of genomic surveillance in finding and tracking new viral lineages with potential implications for public health policies and immunization strategies.[119]

N501Y. It has been demonstrated that several genetic alterations in the receptor-binding domain has caused the spread rate of the SARS-CoV-2 virus to become significantly faster.[120] One of the residues in the RBD−ACE2 interaction region, asparagine at position 501, has been changed to Tyrosine (N501Y) as a result of these mutations.[121] Experiments suggest that the N501Y could increase the SARS-CoV2 affinity of the RBD of spike protein for binding to ACE2.[122,123] It is known that different protein−protein electrostatic interactions play a major role in how well viruses attach to host receptors.[124,125] A study demonstrates how the SARS-CoV-2 N501Y mutation can impact the binding of the RBD with ACE2, with an emphasis on the influence of electrostatic interactions on the binding energy.[105] Moreover, additional hydrogen bonds were observed between the spike RBD domain side residues Y500, Y501, G502, V503, and Y505, enhancing the interaction between the ACE2 receptor and the RBD.[126−128] Due to the considerable change in the electrostatic interaction, it was demonstrated that the binding affinity of SARS-CoV-2 to human ACE2 is higher in the N501Y mutant structure than in the wild type.[129]

P681R/H. As the pandemic progressed, a number of sequenced SARS-CoV-2 genomes have shown mutations at P681. Fusogenicity and the generation of syncytia are determined by the 681 position in the viral spikes.[130] Mutations at this location are capable of enhancing furin cleavage, as shown by the structural modeling of P681 mutations.[131] Moreover, P681R alone has been shown to

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

boost cellular infectivity through furin cleavage in pseudovirus models, whereas P681H does not appear to have a significant effect on furin cleavage or viral infectivity independently.[132,133] The furin cleavage site has a significant impact on the viral infectivity. Studies have revealed that the O-glycosylated sites near the furin cleavage site are governed by the GALNT enzyme family, resulting in a reduction in furin cleavage. Additionally, research demonstrates that the O-glycosylation process is dependent on a unique proline located at position 681 (P681), which disrupts the O-glycosylation and results in an enhancement of furin cleavage. The host's O-glycosylation may affect the infectivity and tropism of a virus by altering the furin cleavage of the spike and provides a clearer understanding of the significance of the P681 mutations present in the highly transmissible alpha and delta variants.[134−136] The delta spike mutation P681R of SARS-CoV-2 plays a crucial part in the substitution of the alpha-to-delta variant. In basic human airway tissues and human lung epithelial cells, delta SARS-CoV-2 effectively outperforms the alpha variant. The replication of the delta variant is drastically decreased to a level lower than the alpha version when the P681R mutation is restored to wild-type P681.[133,137−141]

L452R. The L452R mutation is linked to immune escape and may lead to stronger viral cell attachment; both effects are anticipated to increase the pathogenicity, transmissibility, and infectivity of the virus. According to one study, 14 of 34 RBD-specific monoclonal antibodies had less neutralizing effectiveness when carrying the L452R mutation.[142]

Human leukocyte antigen (HLA)-restricted cellular immunity is mediated by cytotoxic T cells, and L452R confers viral escape from this defense.[136] It was predicted that changes to residue L452, which is close to the RBD-ACE2 interaction interface, may increase the affinity of receptor binding at least marginally, increasing the rate of infectivity of human cells.[143−145] The L452 mutants' effective immune evasion and improved virus-cell attachment might increase the virulence, transmissibility, and/or infectiousness of SARS-CoV-2.[146]

Furthermore, the epsilon variant B.1.429 has the L452R mutation, which is linked to an increase in viral load and a 20% increase in transmissibility.[142] In trials using pseudotyped virus (PV) particles, this variant was found to enhance ACE2 binding, increasing infectivity[39] and 3 to 6-fold decrease in neutralization sensitivity to vaccine-elicited sera. It has been discovered that the SARS-CoV-2 spike RBM containing the L452R and Y453F alterations allows the virus to replicate more readily.[145]

Notably, L452R, one of the most prevalent mutations in the RBD, appears in the delta variant but is absent in the omicron variant. According to reports, the L452R mutation improves the infectivity and fusogenicity of SARS-CoV-2.[145] Zhang and group created an L452R-mutated version of the omicron (omicron-L452R) and discovered that it may cause high infectivity and restore fusogenicity via improving spike protein cleavage. The omicron-L452R mutation of a L452R also significantly increases glycolysis in host cells. Their findings imply that the absence of the L452R mutation from the delta variant accounts for the omicron variant's reduced fusogenicity.[147] The L452R mutation needs to be investigated further to characterize its targeted therapeutic antibodies and vaccines.

A FEW OTHER NOTABLE MUTATIONS

Along with these significant mutations, many other observed mutations appear to have a sizable number of occurrences in different segments of the spike protein. NTD carries a noticeable number of crucial deletions and mutations in the new variants that play an essential role in the spike's functionality.[142,148−155] In addition, mutation such as T95I can have a detrimental effect on antibody neutralization.[156] G142D, Del144, R246I and Del242−244 mutations are considered to promote mAb resistance.[157,158] Recognition of monoclonal antibodies might be diminished by these mutations.[159] D215G also helps in antibody neutralization and immune escape.[155] An extensive analysis of these mutational impacts needs assured investigation.

NEW CUTTING-EDGE EXPERIMENTAL AND COMPUTATIONAL TECHNIQUES OVERCOMING THE CHALLENGING ASPECT OF SPIKE STRUCTURE, FUNCTION, AND SEQUENCE EVOLUTION

Getting a thorough understanding of the virus's structure was one of the biggest problems during the early stages of the pandemic. In order to allow researchers from all over the world to take action before the corresponding study was published, researchers from Fudan University sequenced the viral genome and made it publicly available.[160] Understanding the viral process, the viral protein's interaction with the human host, and the following interactions of ligands and small molecules with the viral protein have all benefited greatly from structural biology. We now understand the viral dynamics better because of structural biologists' extensive structural analysis of the viral proteins.[161,162] Within weeks of the initial SARS-CoV-2 genome becoming accessible to the public, the first 3D structure of the main protease was determined using X-ray crystallography and uploaded to the Protein Databank (PDB ID: 6LU7).[163] After X-ray crystallography, the spike protein's initial crystal structure was finally determined using cryo-electron microscopy (cryo-EM) spectroscopy.[7]

In the early stages of the pandemic, researchers concentrated on using X-ray crystallography to understand the structure of the virus. The size and complexity of the molecule under study are not constrained a priori, making it one of the most effective methods for macromolecular structure determination, often achieving a quasi-atomic resolution.[77,164] Free electron lasers are used to create extremely brief X-ray pulses with peak brilliance 9−10 orders of magnitude higher than that of third generation synchrotrons. These pulses have a duration of tens of femtoseconds.[165−167] Over the years, X-ray crystallography has consistently been the method of choice for structural biologists. A protein's 3D structure can be determined by X-ray crystallography if enough protein can be produced and crystallized at the same time in a well-organized lattice. The crystal is rotated using an X-ray beam, and the diffraction pattern of those results is converted into a map of electron density that reveals the protein's structural viewpoint. This information, along with the protein sequence, is frequently used by structural biologists to comprehend how proteins fold into various secondary structures such as alpha helices and beta sheets. Although X-ray crystallography remains the most effective tool in contemporary structural biology and has the potential to resolve macromolecule structures at the atomic level, recent technological advancements have sparked a

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

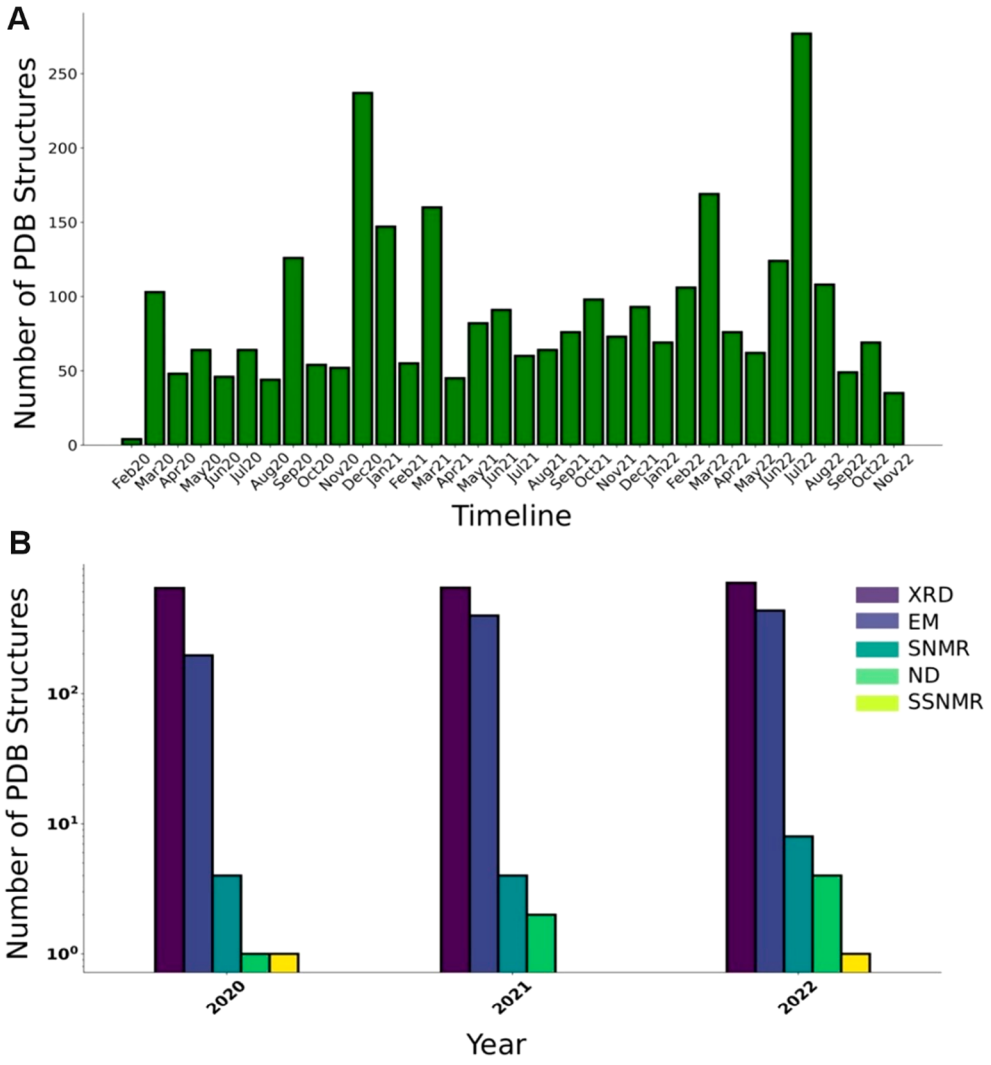


Figure 5. Protein Databank Analysis for PDB structures of SARS-CoV-2. (A) Month wise SARS-CoV-2 structure submission analysis shows the increase in structure submission postepidemic onset with July 2022 corresponding to the largest structure submission month. (B) Year wise Protein Databank Analysis for different structure submission methodology. X-ray Diffraction and electron microscopy dominate the characterization techniques followed for characterizing the crystal structure of SARS-CoV-2. Solution NMR (SNMR) and neutron diffraction (ND) methods have also garnered attention as potential characterization techniques with their usage increasing post 2020.

resolution revolution in the single-particle cryo-EM approach. Researchers opt for cryo-EM, an alternative crystallographic technique, when dealing with proteins that are incapable of being crystallized. Cryo-EM has experienced significant technical advancements recently and is taking the lead in high-resolution macromolecular structure determination, particularly supra-assembles.[168] Using the cryo-EM method, a protein is thinly frozen onto a metal grid. Low-energy electrons are used to bombard the layer in place of X-rays, creating a 2D picture of the individual proteins. The 3D image of the protein is created by computationally sorting and reconstructing thousands of these noisy images. Cryo-EM's ability to produce images from multiple copies of the protein also has added the advantage of enabling researchers to observe how a protein might move in various conformations.

Until the middle of the 2010s, conventional transmission electron microscopy (with resolutions on the order of nanometers) and X-ray crystallography were typically separated by a significant gap in feasible imaging resolutions due to the constraints of these techniques (with resolutions on the order of angstroms). High-resolution structure determination by any approach remained, for the most part, inaccessible to the structure of an entire virion, whether it be by itself or in the context of a host cell. With the Nobel Prize-winning invention of cryogenic vitrification of aqueous materials, including viruses, for cryo-EM imaging by Dubochet and colleagues,[169] this started to alter. Although these early cryo-EM micrographs were virtually free of preservation artifacts, they were nevertheless distorted and had low signal-to-noise ratios. The area of cryo-EM crystallography has also advanced, thanks to innovations like electron detectors, enhanced tomographic reconstruction, and subtomogram averaging techniques. These scientific developments make it simpler for structural biologists to recreate the atomic-level features of the structure of giant asymmetric macromolecules.[170−172] While conventional electron microscopy with contrast agents and chemically fixed samples remains the method of choice for quickly identifying viruses, particularly those that are present in infected cells or tissues, cryo-EM increases our capacity to collect high-resolution data. When SARS-CoV-2 first appeared, these technologies were already developed and could be used in conjunction with traditional electron microscopy, cryo-EM,

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

and X-ray crystallography to quickly apply to the virus and its components on a continuous resolution spectrum.

We examined the Protein Databank to determine the rate of data submission and the kind of crystallographic methodology used for structure elucidation after a large number of crystallographic structures of SARS-CoV-2 were submitted. We discovered that the number of submitted constructions climbed quickly, from 842 in the year 2020 to 1044 in 2021 to 1144 in 2022 (Figure 5A). Additionally, we looked at the methods used for structure prediction and discovered that X-ray crystallography is still the go-to way for doing so. However, the speed at which the structure was resolved by electron microscopy increased. The distinctive outcome we discovered was the rise to 8 in the use of solution NMR for structural refinement by 2022 (Figure 5B). There were other techniques utilized as well, such as neutron diffraction and solid-state NMR, although X-ray crystallography and electron microscopy contributed the majority of the structure refinement.

The best course of action is to expand the number of intelligent multiscale imaging studies that combine various methodologies to provide a complete picture of the biological process as more complex structures emerge in nature. These gaps in the imaging modalities' resolution spectrum are filled by cryo-EM and cryo-electron tomography (cryo-ET) techniques. Cryo-EM enables the visualization of more complicated macromolecular assemblies, whereas cryo-ET enables the imaging of complete biological structures in situ. Additionally, both methodologies' handled problems are constantly evolving as a result of growing technical improvements. On the other hand, X-ray crystallography continues to be the industry standard for applications needing the utmost resolution, such as structure-based drug discovery. However, more recent methods have the potential to be used for future contributions. One such method is nuclear magnetic resonance (NMR), a potent method for determining the binding interfaces and analyzing the dynamics of proteins.[173] The work of Bartenschlager and colleagues,[174] in which they combined fluorescent microscopy, transmission electron microscopy (TEM), scanning electron microscopy (SEM), and cryo-ET image reconstruction to visualize the structural changes in cells induced by SARS-CoV-2 infections, is one of the best examples of multiscale imaging applied to SARS-CoV-2.

In addition to experimental methods, computational methods have drawn a great deal of interest. To comprehend the evolutionary development of SARS-CoV-2,[53,57,175−180] determine the potency of natural phytochemicals as potent drugs against the virulence proteins of SARS-CoV-2,[181−189] and track the viral progression in the general population via mathematical modeling,[190−197] a variety of computational techniques have been developed. In our earlier research,[57] we created a machine learning feedforward model that could forecast the possible mutational locations in the delta variant. We created several features to train our model, with the most crucial feature being the mutational entropy of the virus's protein sequence. In the spike protein, our model identified probable mutation-prone regions that could change in the upcoming generation. A deep learning methodology was put forth by Chimmula and Zhang[198] to forecast the pattern of the COVID outbreak and the potential halting point. They made future predictions about the COVID-19 response using long-term memory (LSTM) networks. The LSTM models were also used by Tomar and Gupta[199] to forecast the number of recovered cases, daily positive cases, and deceased patients in India for the upcoming 30 days. A quick COVID-19 screening tool called nCOVnet was created by Panwar and his team.[200] They created a CNN model based on deep learning and trained it using chest X-ray pictures. Their algorithm had a 97% accuracy rate in predicting a patient who might test positive for COVID-19.

Moreover, mathematical models also played a crucial role in understanding the spread of COVID-19. Using complex mathematical models, researchers can estimate the impact of different interventions and predict the potential outcomes of different scenarios, providing valuable insights and information for policy makers and public health officials. Wang and colleagues[197] developed a computational model to understand the diverse progression of COVID-19 patients infected with SARS-CoV-2. The model integrated the clinical data and analyzed the intracellular viral dynamics, multicellular infection process, and immune responses by using a combination of differential equations and stochastic modeling. The study found that increasing the host antiviral state or virus-induced type I interferon production rate can delay the transition from asymptomatic to symptomatic outcomes. The study also found that T cell exhaustion plays a crucial role in the transition between mild, moderate, and severe symptoms, and patients with severe symptoms lack naïve T cells. The study highlights the importance of interferons and T cell responses in regulating the stage transition during COVID-19 progression and provides potential guidance for personalized therapy in COVID-19 patients. Adam and colleagues[201] used the mathematical model to investigate the role of superspreading events (events where a small number of individuals infect a large number of people) in the transmission of the virus. The researchers found that superspreading events were responsible for a large proportion of infections and that controlling them would be the key to reducing the overall spread of the virus. Lai and colleagues[202] developed a stochastic susceptible-exposed-infectious-removed modeling framework based on travel networks to simulate the spread of COVID-19 in mainland China. They parametrized the model using epidemiological data on COVID-19 and historical and near-real-time anonymized data on human movement. The researchers used the model to conduct before and after comparable analyses to quantify the relative effect of three major groups of nonpharmaceutical interventions (NPI): the restriction of intercity population movement, the identification and isolation of cases, and the reduction of travel and contact within cities to increase social distance. The study found the implementation of a combination of NPIs significantly impacted the reduction of the transmission of COVID-19 across China. The researchers found that if NPIs were implemented earlier, the magnitude and geographical range of the outbreak would have been notably reduced.

SUMMARY

The Covid-19 pandemic has made us realize the dire need to understand the sequence/structure/function paradigm of the spike protein of SARS-CoV-2 and its variant of concerns. The current situation is like this fast-evolving virus left its afresh evolutionary footprint in the genetic and amino acidic sequence through which it modulated its structure−function whenever required.[57] Although short-term and this evolutionary trajectory may or may not carry a critical characteristic signal, a thorough analysis is of utmost necessity. For the past 3

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

years, the vast quantity of current data associated with the SARS-CoV-2 spike protein poses a higher degree of challenges connected to the sequence/structure/function paradigm. In fact, this remains a central challenge of computational structural biology: to decode the language of biological sequences and to decipher the structural, functional, and evolutionary clues hidden in it. In the investigation of decoding the meaning of sequence, two following approaches are usually employed: (i) pattern recognition techniques to check for sequence similarity and connect to the related structures and functions; (ii) ab initio prediction methods to resolve 3D structure from 1D sequence and then connect with structure−function relation.[203] The first approach inspires us to understand the unconventional evolutionary strategies adopted by RNA viruses that help them evolve from a certain variant to a VOC. Once a specific variant evolves faster, one can track its mutational trajectory from its country of origin which sheds light on several mutationally fluctuating sites at both the genomic and amino acidic levels. Using information entropy and statistical mechanical approaches, it is then possible to distinguish the nature of the mutational fluctuation. Depending on the nature of fluctuation, one can identify the transition from one variant to VOC. This review also summarizes how information entropy and mutational response function help characterize the changes both at a single residual level and in a collective domain-wise manner.

With the new cutting-edge sequencing technology, a diversity of database analyses, taxonomic classification at the molecular level, various approaches of pattern recognition, and numerous structure-based algorithms, including 3D structure prediction from 1D sequence information have been developed and evolved.[48,204−209] Although these developments indeed help connect the above complex notion of sequence-structure annotation, much remains to be understood from the functional aspects. To communicate the functionalities, it is rather critical to understand the structural and conformational dynamics of the protein at a microscopic level. This poses a next-level challenge in dealing with complex structural dynamics associated with flexible domains, linkers, and highly flexible loop regions that actively control the functional mechanism. For instance, in this review, we have highlighted the linker between the two lobes of the S1-head domain, which actively controls the up−down dynamics of the spike head and its accessibility to the host-cell receptors. We have also summarized the control mechanism of two other critical loops, 630 and FPPR loops. It is interesting to note that these dynamic regions are highly prone to mutations. D614G is one of the best examples that bear the mutational response over these loop regions (IDRs) and retains almost in all variants of concerns and interests as shown in Table 3. This fatal mutation highly affects the functions of 630 and FPPR loops and, thereby, the stability of the spike trimer.

The above examples imply that although experimental and computational capabilities have significantly advanced to address the structure, function, and evolution of SARS-CoV-2, we still need a better way to understand their dynamic perspective. Thus, among other experimental techniques to resolve structural information, NMR spectroscopy has emerged as the method of choice for studying both the protein structure and dynamics in solution. It is interesting to follow Figure 5, which elucidates a lot about the need of the hour. Immediately after the emergence of SARS-CoV-2, if we follow the usage pattern of different structure resolving techniques, it is reflected in how the usage of NMR is rising from 2020 to 2022. Despite major progress in experimental and computational methods, obtaining reliable information about the structure and dynamics from experimental NMR data remains challenging. Thus, molecular dynamics simulations have been emerged as an indispensable tool in the analysis of NMR results.[210] To understand the structural dynamics of highly flexible regions, several sophisticated knowledge-based and physics-based in silico sampling techniques have been developed.[211,212] Apart from the classical molecular dynamics simulation approach, in this review, we have also highlighted how coarse-grained molecular dynamics simulations help discern the broad conformational phase space of the dynamic regions and discuss how the conformational preferences may modulate the function in spike protein.[96] We are optimistic that the ongoing development of improved technologies will make it possible to decipher not only the sequence−structure−function annotation but also their dynamic demeanor.

Needless to say, through all the above developments, we would like to characterize better the spike protein (a critical target for vaccine development) which may help invent new therapeutics and vaccine development strategies as new VOCs would emerge. Recently, two targeted phage-based vaccination strategies against SARS-COV-2, dual ligand peptide-targeted phage and adeno-associated virus/phage (AAVP) particles, have been developed and have undergone preliminary evaluation.[213] Surveying vaccine development strategies is beyond the scope of this review. It indeed needs another review to tour that world of discoveries.

■ ASSOCIATED CONTENT

Supporting Information

Supporting Information includes The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c00944.

Structure-based coarse-grained simulation methods of SARS-CoV-2 spike protein; Restrained simulation details to understand linker's control on the up−down movement of RBD (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Susmita Roy − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246 West Bengal, India;* orcid.org/0000-0001-6411-4347; Email: susmita.roy@iiserkol.ac.in

Authors

Anushree Sinha − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246 West Bengal, India*

Satyam Sangeet − *Department of Chemical Sciences, Indian Institute of Science Education and Research Kolkata, Mohanpur 741246 West Bengal, India*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c00944

Author Contributions

[§]These authors contributed equally.

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

■ ACKNOWLEDGMENTS

S.R. acknowledges support from the Department of Biotechnology (DBT) (grant no. BT/12/IYBA/2019/12) and Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Govt. of India (grant no. SRG/2020/001295). A.S. acknowledges DST Inspire Fellowship.

■ REFERENCES

(1) Fehr, A. R.; Perlman, S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **2015**, *1282*, 1−23.
(2) Zhong, N. S.; Zheng, B. J.; Li, Y. M.; Poon, L. L. M.; Xie, Z. H.; Chan, K. H.; Li, P. H.; Tan, S. Y.; Chang, Q.; Xie, J. P.; Liu, X. Q.; Xu, J.; Li, D. X.; Yuen, K. Y.; Peiris, J. S. M.; Guan, Y. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **2003**, *362*, 1353−1358.
(3) Hu, B.; Guo, H.; Zhou, P.; Shi, Z. L. Characteristics of SARS-CoV-2 and COVID-19. *Nat. Rev. Microbiol* **2021**, *19* (3), 141−154.
(4) Jahan, M.; Nasif, M. A. O.; Rahmat, R.; Ul Islam, S. M. R.; Munshi, S. U.; Ahmed, M. S. Genome Sequencing of Omicron Variants of SARS-CoV-2 circulating in Bangladesh during the third wave of the COVID-19 Pandemic. *Microbiol Resour Announc* **2022**, *11* (7), No. e0038122.
(5) Canessa, E.; Tenze, L. GenomeBits insight into omicron and delta variants of coronavirus pathogen. *PLoS One* **2022**, *17* (7), No. e0271039.
(6) Chan, J. F.; Kok, K. H.; Zhu, Z.; Chu, H.; To, K. K.; Yuan, S.; Yuen, K. Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect* **2020**, *9* (1), 221−236.
(7) Jackson, C. B.; Farzan, M.; Chen, B.; Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **2022**, *23* (1), 3−20.
(8) Ntagereka, P. B.; Oyola, S. O.; Baenyi, S. P.; Rono, G. K.; Birindwa, A. B.; Shukuru, D. W.; Baharanyi, T. C.; Kashosi, T. M.; Buhendwa, J. C.; Bisimwa, P. B.; Kusinza, A. B.; Basengere, R. A.; Mukwege, D. Whole-genome sequencing of SARS-CoV-2 reveals diverse mutations in circulating Alpha and Delta variants during the first, second, and third waves of COVID-19 in South Kivu, east of the Democratic Republic of the Congo. *Int. J. Infect Dis* **2022**, *122*, 136−143.
(9) Garvin, M. R.; E, T. P.; Pavicic, M.; Jones, P.; Amos, B. K.; Geiger, A.; Shah, M. B.; Streich, J.; Felipe Machado Gazolla, J. G.; Kainer, D.; Cliff, A.; Romero, J.; Keith, N.; Brown, J. B.; Jacobson, D. Potentially adaptive SARS-CoV-2 mutations discovered with novel spatiotemporal and explainable AI models. *Genome Biol.* **2020**, *21* (1), 304.
(10) Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S. M.; Walsh, R. M., Jr; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020**, *369* (6511), 1586−1592.
(11) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367* (6483), 1260−1263.
(12) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **2020**, *588* (7837), 327−330.
(13) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Krausslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G. Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **2020**, *588* (7838), 498−502.
(14) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2), 281−292.
(15) Walls, A. C.; Tortorici, M. A.; Bosch, B. J.; Frenz, B.; Rottier, P. J. M.; DiMaio, F.; Rey, F. A.; Veesler, D. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. *Nature* **2016**, *531* (7592), 114−117.
(16) Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Veesler, D. Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (42), 11157−11162.
(17) Wrobel, A. G.; Benton, D. J.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct Mol. Biol.* **2020**, *27* (8), 763−767.
(18) O'Toole, A.; Pybus, O. G.; Abram, M. E.; Kelly, E. J.; Rambaut, A. Pango lineage designation and assignment using SARS-CoV-2 spike gene nucleotide sequences. *BMC Genomics* **2022**, *23* (1), 121.
(19) Khare, S.; Gurry, C.; Freitas, L.; Schultz, M. B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Lee, R. T.; Yeo, W.; Curation Team, G. C.; Maurer-Stroh, S. GISAID's Role in Pandemic Response. *China CDC Wkly* **2021**, *3* (49), 1049−1051.
(20) Shishir, T. A.; Naser, I. B.; Faruque, S. M. In silico comparative genomics of SARS-CoV-2 to determine the source and diversity of the pathogen in Bangladesh. *PLoS One* **2021**, *16* (1), No. e0245584.
(21) Mandal, S.; Roychowdhury, T.; Bhattacharya, A. Pattern of genomic variation in SARS-CoV-2 (COVID-19) suggests restricted nonrandom changes: Analysis using Shewhart control charts. *J. Biosci* **2021**, *46* (1), 1.
(22) Rouchka, E. C.; Chariker, J. H.; Chung, D. Variant analysis of 1,040 SARS-CoV-2 genomes. *PLoS One* **2020**, *15* (11), No. e0241535.
(23) Li, Q.; Wu, J.; Nie, J.; Zhang, L.; Hao, H.; Liu, S.; Zhao, C.; Zhang, Q.; Liu, H.; Nie, L.; Qin, H.; Wang, M.; Lu, Q.; Li, X.; Sun, Q.; Liu, J.; Zhang, L.; Li, X.; Huang, W.; Wang, Y. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. *Cell* **2020**, *182* (5), 1284−1294.
(24) Srivastava, S.; Banu, S.; Singh, P.; Sowpati, D. T.; Mishra, R. K. SARS-CoV-2 genomics: An Indian perspective on sequencing viral variants. *J. Biosci* **2021**, *46* (1), 1.
(25) Tang, X.; Wu, C.; Li, X.; Song, Y.; Yao, X.; Wu, X.; Duan, Y.; Zhang, H.; Wang, Y.; Qian, Z.; Cui, J.; Lu, J. On the origin and continuing evolution of SARS-CoV-2. *Natl. Sci. Rev.* **2020**, *7* (6), 1012−1023.
(26) Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **2017**, *1* (1), 33−46.
(27) Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **2017**, *22* (13), 1.
(28) *WHO-convened global study of origins of SARS-CoV-2: China Part*; World Health Organization, 2021.
(29) Banerjee, A.; Doxey, A. C.; Mossman, K.; Irving, A. T. Unraveling the Zoonotic Origin and Transmission of SARS-CoV-2. *Trends Ecol Evol* **2021**, *36* (3), 180−184.
(30) van Dorp, L.; Acman, M.; Richard, D.; Shaw, L. P.; Ford, C. E.; Ormond, L.; Owen, C. J.; Pang, J.; Tan, C. C. S.; Boshier, F. A. T.; Ortiz, A. T.; Balloux, F. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect Genet Evol* **2020**, *83*, 104351.
(31) Wong, G.; Bi, Y. H.; Wang, Q. H.; Chen, X. W.; Zhang, Z. G.; Yao, Y. G. Zoonotic origins of human coronavirus 2019 (HCoV-19/SARS-CoV-2): why is this work important? *Zool Res.* **2020**, *41* (3), 213−219.
(32) Zhang, Y. Z.; Holmes, E. C. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell* **2020**, *181* (2), 223−227.
(33) Forster, P.; Forster, L.; Renfrew, C.; Forster, M. Phylogenetic network analysis of SARS-CoV-2 genomes. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (17), 9241−9243.
(34) Tang, X.; Ying, R.; Yao, X.; Li, G.; Wu, C.; Tang, Y.; Li, Z.; Kuang, B.; Wu, F.; Chi, C.; Du, X.; Qin, Y.; Gao, S.; Hu, S.; Ma, J.;

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

Liu, T.; Pang, X.; Wang, J.; Zhao, G.; Tan, W.; Zhang, Y.; Lu, X.; Lu, J. Evolutionary analysis and lineage designation of SARS-CoV-2 genomes. *Sci. Bull. (Beijing)* **2021**, *66* (22), 2297−2311.
(35) Rockett, R. J.; Arnott, A.; Lam, C.; Sadsad, R.; Timms, V.; Gray, K. A.; Eden, J. S.; Chang, S.; Gall, M.; Draper, J.; Sim, E. M.; Bachmann, N. L.; Carter, I.; Basile, K.; Byun, R.; O'Sullivan, M. V.; Chen, S. C.; Maddocks, S.; Sorrell, T. C.; Dwyer, D. E.; Holmes, E. C.; Kok, J.; Prokopenko, M.; Sintchenko, V. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nat. Med.* **2020**, *26* (9), 1398−1404.
(36) Seemann, T.; Lane, C. R.; Sherry, N. L.; Duchene, S.; Goncalves da Silva, A.; Caly, L.; Sait, M.; Ballard, S. A.; Horan, K.; Schultz, M. B.; Hoang, T.; Easton, M.; Dougall, S.; Stinear, T. P.; Druce, J.; Catton, M.; Sutton, B.; van Diemen, A.; Alpren, C.; Williamson, D. A.; Howden, B. P. Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **2020**, *11* (1), 4376.
(37) Bluhm, A.; Christandl, M.; Gesmundo, F.; Ravn Klausen, F.; Mancinska, L.; Steffan, V.; Stilck Franca, D.; Werner, A. H. SARS-CoV-2 transmission routes from genetic data: A Danish case study. *PLoS One* **2020**, *15* (10), No. e0241405.
(38) Dellicour, S.; Durkin, K.; Hong, S. L.; Vanmechelen, B.; Marti-Carreras, J.; Gill, M. S.; Meex, C.; Bontems, S.; Andre, E.; Gilbert, M.; Walker, C.; Maio, N.; Faria, N. R.; Hadfield, J.; Hayette, M. P.; Bours, V.; Wawina-Bokalanga, T.; Artesi, M.; Baele, G.; Maes, P. A Phylodynamic Workflow to Rapidly Gain Insights into the Dispersal History and Dynamics of SARS-CoV-2 Lineages. *Mol. Biol. Evol.* **2021**, *38* (4), 1608−1613.
(39) Deng, X.; Garcia-Knight, M. A.; Khalid, M. M.; Servellita, V.; Wang, C.; Morris, M. K.; Sotomayor-Gonzalez, A.; Glasner, D. R.; Reyes, K. R.; Gliwa, A. S.; Reddy, N. P.; Martin, C. S. S.; Federman, S.; Cheng, J.; Balcerek, J.; Taylor, J.; Streithorst, J. A.; Miller, S.; Kumar, G. R.; Sreekumar, B.; Chen, P. Y.; Schulze-Gahmen, U.; Taha, T. Y.; Hayashi, J.; Simoneau, C. R.; McMahon, S.; Lidsky, P. V.; Xiao, Y.; Hemarajata, P.; Green, N. M.; Espinosa, A.; Kath, C.; Haw, M.; Bell, J.; Hacker, J. K.; Hanson, C.; Wadford, D. A.; Anaya, C.; Ferguson, D.; Lareau, L. F.; Frankino, P. A.; Shivram, H.; Wyman, S. K.; Ott, M.; Andino, R.; Chiu, C. Y., Transmission, infectivity, and antibody neutralization of an emerging SARS-CoV-2 variant in California carrying a L452R spike protein mutation. *medRxiv*, March 9, 2021, ver. 1. DOI: 10.1101/2021.03.07.21252647 (accessed Dec 15, 2022).
(40) Gambaro, F.; Behillil, S.; Baidaliuk, A.; Donati, F.; Albert, M.; Alexandru, A.; Vanpeene, M.; Bizard, M.; Brisebarre, A.; Barbet, M.; Derrar, F.; van der Werf, S.; Enouf, V.; Simon-Loriere, E. Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Euro Surveill* **2020**, *25* (26), 2001200.
(41) Gudbjartsson, D. F.; Helgason, A.; Jonsson, H.; Magnusson, O. T.; Melsted, P.; Norddahl, G. L.; Saemundsdottir, J.; Sigurdsson, A.; Sulem, P.; Agustsdottir, A. B.; Eiriksdottir, B.; Fridriksdottir, R.; Gardarsdottir, E. E.; Georgsson, G.; Gretarsdottir, O. S.; Gudmundsson, K. R.; Gunnarsdottir, T. R.; Gylfason, A.; Holm, H.; Jensson, B. O.; Jonasdottir, A.; Jonsson, F.; Josefsdottir, K. S.; Kristjansson, T.; Magnusdottir, D. N.; le Roux, L.; Sigmundsdottir, G.; Sveinbjornsson, G.; Sveinsdottir, K. E.; Sveinsdottir, M.; Thorarensen, E. A.; Thorbjornsson, B.; Love, A.; Masson, G.; Jonsdottir, I.; Möller, A. D.; Gudnason, T.; Kristinsson, K. G.; Thorsteinsdottir, U.; Stefansson, K. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J. Med.* **2020**, *382* (24), 2302−2315.
(42) Miller, D.; Martin, M. A.; Harel, N.; Tirosh, O.; Kustin, T.; Meir, M.; Sorek, N.; Gefen-Halevi, S.; Amit, S.; Vorontsov, O.; Shaag, A.; Wolf, D.; Peretz, A.; Shemer-Avni, Y.; Roif-Kaminsky, D.; Kopelman, N. M.; Huppert, A.; Koelle, K.; Stern, A. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nat. Commun.* **2020**, *11* (1), 5518.
(43) Oude Munnink, B. B.; Nieuwenhuijse, D. F.; Stein, M.; O'Toole, A.; Haverkate, M.; M?llers, M.; Kamga, S. K.; Schapendonk, C.; Pronk, M.; Lexmond, P.; van der Linden, A.; Bestebroer, T.; Chestakova, I.; Overmars, R. J.; van Nieuwkoop, S.; Molenkamp, R.; van der Eijk, A. A.; GeurtsvanKessel, C.; Vennema, H.; Meijer, A.; Rambaut, A.; van Dissel, J.; Sikkema, R. S.; Timen, A.; Koopmans, M.; et al. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. *Nat. Med.* **2020**, *26* (9), 1405−1410.
(44) Gonzalez-Reiche, A. S.; Hernandez, M. M.; Sullivan, M. J.; Ciferri, B.; Alshammary, H.; Obla, A.; Fabre, S.; Kleiner, G.; Polanco, J.; Khan, Z.; et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **2020**, *369* (6501), 297−301.
(45) Worobey, M.; Pekar, J.; Larsen, B. B.; Nelson, M. I.; Hill, V.; Joy, J. B.; Rambaut, A.; Suchard, M. A.; Wertheim, J. O.; Lemey, P. The emergence of sars-cov-2 in europe and north america. *Science* **2020**, *370* (6516), 564−570.
(46) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The proximal origin of SARS-CoV-2. *Nat. Med.* **2020**, *26* (4), 450−452.
(47) Liu, P.; Jiang, J. Z.; Wan, X. F.; Hua, Y.; Li, L.; Zhou, J.; Wang, X.; Hou, F.; Chen, J.; Zou, J.; Chen, J. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathog* **2020**, *16* (5), e1008421.
(48) Sangeet, S.; Khan, A. Exploratory Data Analysis of Genomic Sequence of Variants of SARS-CoV-2 Reveals Sequence Divergence and Mutational Localization. *Bioinform Biol. Insights* **2022**, *16*, 117793222211262.
(49) Zhou, H.; Chen, X.; Hu, T.; Li, J.; Song, H.; Liu, Y.; Wang, P.; Liu, D.; Yang, J.; Holmes, E. C.; Hughes, A. C.; Bi, Y.; Shi, W. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr. Biol.* **2020**, *30* (11), 2196−2203.
(50) Flores-Alanis, A.; Sandner-Miranda, L.; Delgado, G.; Cravioto, A.; Morales-Espinosa, R. The receptor binding domain of SARS-CoV-2 spike protein is the result of an ancestral recombination between the bat-CoV RaTG13 and the pangolin-CoV MP789. *BMC Res. Notes* **2020**, *13* (1), 398.
(51) Dash, P.; Turuk, J.; Behera, S. K.; Palo, S. K.; Raghav, S. K.; Ghosh, A.; Sabat, J.; Rath, S.; Subhadra, S.; Rana, K.; Bhattacharya, D.; Kanungo, S.; Kshatri, J. S.; Mishra, B. K.; Dash, S.; Parida, A.; Pati, S. Sequence analysis of Indian SARS-CoV-2 isolates shows a stronger interaction of mutant receptor-binding domain with ACE2. *Int. J. Infect Dis* **2021**, *104*, 491−500.
(52) Liu, Y.; Hu, G.; Wang, Y.; Ren, W.; Zhao, X.; Ji, F.; Zhu, Y.; Feng, F.; Gong, M.; Ju, X.; Zhu, Y.; Cai, X.; Lan, J.; Guo, J.; Xie, M.; Dong, L.; Zhu, Z.; Na, J.; Wu, J.; Lan, X.; Xie, Y.; Wang, X.; Yuan, Z.; Zhang, R.; Ding, Q. Functional and genetic analysis of viral receptor ACE2 orthologs reveals a broad potential host range of SARS-CoV-2. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (12), 1 DOI: 10.1073/pnas.2025373118.
(53) Guruprasad, L. Evolutionary relationships and sequence-structure determinants in human SARS coronavirus-2 spike proteins for host receptor recognition. *Proteins* **2020**, *88* (11), 1387−1393.
(54) Padhan, K.; Parvez, M. K.; Al-Dosari, M. S. Comparative sequence analysis of SARS-CoV-2 suggests its high transmissibility and pathogenicity. *Future Virology* **2021**, *16* (3), 245−254.
(55) Zhao, Z.; Sokhansanj, B. A.; Malhotra, C.; Zheng, K.; Rosen, G. L. Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization. *PLoS Comput. Biol.* **2020**, *16* (9), e1008269.
(56) Ghanchi, N. K.; Nasir, A.; Masood, K. I.; Abidi, S. H.; Mahmood, S. F.; Kanji, A.; Razzak, S.; Khan, W.; Shahid, S.; Yameen, M.; Raza, A.; Ashraf, J.; Ansar, Z.; Dharejo, M. B.; Islam, N.; Hasan, Z.; Hasan, R. Higher entropy observed in SARS-CoV-2 genomes from the first COVID-19 wave in Pakistan. *PLoS One* **2021**, *16* (8), No. e0256451.
(57) Sangeet, S.; Sarkar, R.; Mohanty, S. K.; Roy, S. Quantifying Mutational Response to Track the Evolution of SARS-CoV-2 Spike Variants: Introducing a Statistical-Mechanics-Guided Machine Learning Method. *J. Phys. Chem. B* **2022**, *126* (40), 7895−7905.
(58) Zhao, H.; Han, K.; Gao, C.; Madhira, V.; Topaloglu, U.; Lu, Y.; Jin, G. VOC-alarm: mutation-based prediction of SARS-CoV-2 variants of concern. *Bioinformatics* **2022**, *38* (14), 3549−3556.

(59) Roy, S.; Jaiswar, A.; Sarkar, R. Dynamic Asymmetry Exposes 2019-nCoV Prefusion Spike. *J. Phys. Chem. Lett.* **2020**, *11* (17), 7021−7027.
(60) Chi, X.; Yan, R.; Zhang, J.; Zhang, G.; Zhang, Y.; Hao, M.; Zhang, Z.; Fan, P.; Dong, Y.; Yang, Y.; et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **2020**, *369* (6504), 650−655.
(61) Liu, L.; Wang, P.; Nair, M. S.; Yu, J.; Rapp, M.; Wang, Q.; Luo, Y.; Chan, J. F.; Sahi, V.; Figueroa, A.; Guo, X. V.; Cerutti, G.; Bimela, J.; Gorman, J.; Zhou, T.; Chen, Z.; Yuen, K. Y.; Kwong, P. D.; Sodroski, J. G.; Yin, M. T.; Sheng, Z.; Huang, Y.; Shapiro, L.; Ho, D. D. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **2020**, *584* (7821), 450−456.
(62) Min, L.; Sun, Q. Antibodies and Vaccines Target RBD of SARS-CoV-2. *Front Mol. Biosci* **2021**, *8*, 671633.
(63) Zhang, J.; Cai, Y.; Xiao, T.; Lu, J.; Peng, H.; Sterling, S. M.; Walsh, R. M., Jr; Rits-Volloch, S.; Zhu, H.; Woosley, A. N.; et al. Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **2021**, *372* (6541), 525−530.
(64) Yang, T. J.; Yu, P. Y.; Chang, Y. C.; Liang, K. H.; Tso, H. C.; Ho, M. R.; Chen, W. Y.; Lin, H. T.; Wu, H. C.; Hsu, S. D. Effect of SARS-CoV-2 B.1.1.7 mutations on spike protein structure and function. *Nat. Struct Mol. Biol.* **2021**, *28* (9), 731−739.
(65) Fan, X.; Cao, D.; Kong, L.; Zhang, X. Cryo-EM analysis of the post-fusion structure of the SARS-CoV spike glycoprotein. *Nat. Commun.* **2020**, *11* (1), 3618.
(66) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369* (6501), 330−333.
(67) Yuan, M.; Wu, N.; Zhu, X.; Lee, C.; So, R.; Lv, H.; Mok, C.; Wilson, I. A. A Highly Conserved Cryptic Epitope in the Receptor-Binding Domains of SARS-CoV-2 and SARS-CoV. *Science.* **2020**, *368*, 630.
(68) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581* (7807), 215−220.
(69) Wang, Q.; Zhang, Y.; Wu, L.; Niu, S.; Song, C.; Zhang, Z.; Lu, G.; Qiao, C.; Hu, Y.; Yuen, K. Y.; Wang, Q.; Zhou, H.; Yan, J.; Qi, J. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **2020**, *181* (4), 894−904.
(70) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J.; Welsch, S.; Blanc, F. E.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, *370* (6513), 203−208.
(71) Ali, A.; Vijayan, R. Dynamics of the ACE2-SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Sci. Rep* **2020**, *10* (1), 14214.
(72) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci.* **2020**, *6* (10), 1722−1734.
(73) Schlick, T.; Zhu, Q.; Jain, S.; Yan, S. Structure-altering mutations of the SARS-CoV-2 frameshifting RNA element. *Biophys. J.* **2021**, *120* (6), 1040−1053.
(74) Shin, D.; Mukherjee, R.; Grewe, D.; Bojkova, D.; Baek, K.; Bhattacharya, A.; Schulz, L.; Widera, M.; Mehdipour, A. R.; Tascher, G.; Geurink, P. P.; Wilhelm, A.; van der Heden van Noort, G. J.; Ovaa, H.; Muller, S.; Knobeloch, K. P.; Rajalingam, K.; Schulman, B. A.; Cinatl, J.; Hummer, G.; Ciesek, S.; Dikic, I. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* **2020**, *587* (7835), 657−662.
(75) Verkhivker, G. M. Molecular Simulations and Network Modeling Reveal an Allosteric Signaling in the SARS-CoV-2 Spike Proteins. *J. Proteome Res.* **2020**, *19* (11), 4587−4608.
(76) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* **2020**, *367* (6485), 1444−1448.
(77) Shang, J.; Ye, G.; Shi, K.; Wan, Y.; Luo, C.; Aihara, H.; Geng, Q.; Auerbach, A.; Li, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature* **2020**, *581* (7807), 221−224.
(78) Sharifkashani, S.; Bafrani, M. A.; Khaboushan, A. S.; Pirzadeh, M.; Kheirandish, A.; Yavarpour Bali, H.; Hessami, A.; Saghazadeh, A.; Rezaei, N. Angiotensin-converting enzyme 2 (ACE2) receptor and SARS-CoV-2: Potential therapeutic targeting. *Eur. J. Pharmacol.* **2020**, *884*, 173455.
(79) Yi, C.; Sun, X.; Ye, J.; Ding, L.; Liu, M.; Yang, Z.; Lu, X.; Zhang, Y.; Ma, L.; Gu, W.; Qu, A.; Xu, J.; Shi, Z.; Ling, Z.; Sun, B. Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies. *Cell Mol. Immunol* **2020**, *17* (6), 621−630.
(80) Xu, C.; Wang, Y.; Liu, C.; Zhang, C.; Han, W.; Hong, X.; Wang, Y.; Hong, Q.; Wang, S.; Zhao, Q.; et al. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Science advances* **2021**, *7* (1), eabe5575.
(81) Rui, L.; Haonan, L.; Wanyi, C. Silico analysis of interaction between full-length SARS-CoV2 S protein with human Ace2 receptor: Modelling, docking, MD simulation. *Biophys Chem.* **2020**, *267*, 106472.
(82) Wang, Y.; Xu, C.; Wang, Y.; Hong, Q.; Zhang, C.; Li, Z.; Xu, S.; Zuo, Q.; Liu, C.; Huang, Z.; Cong, Y. Conformational dynamics of the Beta and Kappa SARS-CoV-2 spike proteins and their complexes with ACE2 receptor revealed by cryo-EM. *Nat. Commun.* **2021**, *12* (1), 7345.
(83) Xiao, T.; Lu, J.; Zhang, J.; Johnson, R. I.; McKay, L. G. A.; Storm, N.; Lavine, C. L.; Peng, H.; Cai, Y.; Rits-Volloch, S.; Lu, S.; Quinlan, B. D.; Farzan, M.; Seaman, M. S.; Griffiths, A.; Chen, B. A trimeric human angiotensin-converting enzyme 2 as an anti-SARS-CoV-2 agent. *Nat. Struct Mol. Biol.* **2021**, *28* (2), 202−209.
(84) Dutta, S.; Panthi, B.; Chandra, A. All-Atom Simulations of Human ACE2-Spike Protein RBD Complexes for SARS-CoV-2 and Some of its Variants: Nature of Interactions and Free Energy Diagrams for Dissociation of the Protein Complexes. *J. Phys. Chem. B* **2022**, *126* (29), 5375−5389.
(85) Damas, J.; Hughes, G. M.; Keough, K. C.; Painter, C. A.; Persky, N. S.; Corbo, M.; Hiller, M.; Koepfli, K. P.; Pfenning, A. R.; Zhao, H.; Genereux, D. P.; Swofford, R.; Pollard, K. S.; Ryder, O. A.; Nweeia, M. T.; Lindblad-Toh, K.; Teeling, E. C.; Karlsson, E. K.; Lewin, H. A. Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (36), 22311−22322.
(86) Essalmani, R.; Jain, J.; Susan-Resiga, D.; Andreo, U.; Evagelidis, A.; Derbali, R. M.; Hyunh, D. N.; Dallaire, F.; Laporte, M.; Delpal, A.; Sutto-Ortiz, P.; Coutard, B.; Mapa, C.; Wilcoxen, K.; Decroly, E.; Pham, T. N. Q.; Cohen, E. A.; Seidah, N. G. Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity. *J. Virol* **2022**, *96* (8), No. e0012822.
(87) Mollica, V.; Rizzo, A.; Massari, F. The pivotal role of TMPRSS2 in coronavirus disease 2019 and prostate cancer. *Future Oncol* **2020**, *16* (27), 2029−2033.
(88) Hoffmann, M.; Kleine-Weber, H.; Schroeder, S.; Kruger, N.; Herrler, T.; Erichsen, S.; Schiergens, T. S.; Herrler, G.; Wu, N. H.; Nitsche, A.; Muller, M. A.; Drosten, C.; Pohlmann, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **2020**, *181* (2), 271−280.
(89) Shrimp, J. H.; Kales, S. C.; Sanderson, P. E.; Simeonov, A.; Shen, M.; Hall, M. D. An Enzymatic TMPRSS2 Assay for Assessment of Clinical Candidates and Discovery of Inhibitors as Potential Treatment of COVID-19. *ACS Pharmacol Transl Sci.* **2020**, *3* (5), 997−1007.
(90) Vankadari, N.; Ketavarapu, V.; Mitnala, S.; Vishnubotla, R.; Reddy, D. N.; Ghosal, D. Structure of Human TMPRSS2 in Complex with SARS-CoV-2 Spike Glycoprotein and Implications for Potential Therapeutics. *J. Phys. Chem. Lett.* **2022**, *13* (23), 5324−5333.

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304

(91) Dalziel, M.; Crispin, M.; Scanlan, C. N.; Zitzmann, N.; Dwek, R. A. Emerging principles for the therapeutic exploitation of glycosylation. *Science* **2014**, *343* (6166), 1235681.
(92) Scanlan, C. N.; Offer, J.; Zitzmann, N.; Dwek, R. A. Exploiting the defensive sugars of HIV-1 for drug and vaccine design. *Nature* **2007**, *446* (7139), 1038−45.
(93) Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M. Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj* **2019**, *1863* (10), 1480−1497.
(94) Sztain, T.; Ahn, S. H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* **2021**, *13* (10), 963−968.
(95) Harbison, A. M.; Fogarty, C. A.; Phung, T. K.; Satheesan, A.; Schulz, B. L.; Fadda, E. Fine-tuning the spike: role of the nature and topology of the glycan shield in the structure and dynamics of the SARS-CoV-2 S. *Chem. Sci.* **2022**, *13* (2), 386−395.
(96) Dodero-Rojas, E.; Onuchic, J. N.; Whitford, P. C. Sterically confined rearrangements of SARS-CoV-2 Spike protein control cell invasion. *Elife* **2021**, *10*, 1.
(97) Jaimes, J. A.; Millet, J. K.; Whittaker, G. R. Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. *iScience* **2020**, *23* (6), 101212.
(98) Madu, I. G.; Roth, S. L.; Belouzard, S.; Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol* **2009**, *83* (15), 7411−21.
(99) Markosian, C.; Staquicini, D. I.; Dogra, P.; Dodero-Rojas, E.; Lubin, J. H.; Tang, F. H. F.; Smith, T. L.; Contessoto, V. G.; Libutti, S. K.; Wang, Z.; Cristini, V.; Khare, S. D.; Whitford, P. C.; Burley, S. K.; Onuchic, J. N.; Pasqualini, R.; Arap, W. Genetic and Structural Analysis of SARS-CoV-2 Spike Protein for Universal Epitope Selection. *Mol. Biol. Evol.* **2022**, *39* (5), 1.
(100) Heffron, A. S.; McIlwain, S. J.; Amjadi, M. F.; Baker, D. A.; Khullar, S.; Armbrust, T.; Halfmann, P. J.; Kawaoka, Y.; Sethi, A. K.; Palmenberg, A. C.; Shelef, M. A.; O'Connor, D. H.; Ong, I. M. The landscape of antibody binding in SARS-CoV-2 infection. *PLoS Biol.* **2021**, *19* (6), e3001265.
(101) Li, Y.; Lai, D. Y.; Zhang, H. N.; Jiang, H. W.; Tian, X.; Ma, M. L.; Qi, H.; Meng, Q. F.; Guo, S. J.; Wu, Y.; Wang, W.; Yang, X.; Shi, D. W.; Dai, J. B.; Ying, T.; Zhou, J.; Tao, S. C. Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol. Immunol* **2020**, *17* (10), 1095−1097.
(102) Farrera-Soler, L.; Daguer, J. P.; Barluenga, S.; Vadas, O.; Cohen, P.; Pagano, S.; Yerly, S.; Kaiser, L.; Vuilleumier, N.; Winssinger, N. Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *PLoS One* **2020**, *15* (9), No. e0238089.
(103) Aleem, A.; Samad, A. B. A.; Slenker, A. K., Emerging variants of SARS-CoV-2 and novel therapeutics against coronavirus (COVID-19). In *StatPearls (internet)*; StatPearls Publishing: 2022.
(104) Cai, Y.; Zhang, J.; wang, T.; Lavine, C. L.; Rawson, S.; Peng, H.; Zhu, H.; Anand, K.; Tong, P.; Gautam, A.; et al. Structural basis for enhanced infectivity and immune evasion of SARS-CoV-2 variants. *Science* **2021**, *373* (6555), 642−648.
(105) Ali, F.; Kasry, A.; Amin, M. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. *Med. Drug Discov* **2021**, *10*, 100086.
(106) Chakraborty, S. Evolutionary and structural analysis elucidates mutations on SARS-CoV2 spike protein with altered human ACE2 binding affinity. *Biochem. Biophys. Res. Commun.* **2021**, *534*, 374−380.
(107) Zhou, P.; Yang, X. L.; Wang, X. G.; Hu, B.; Zhang, L.; Zhang, W.; Si, H. R.; Zhu, Y.; Li, B.; Huang, C. L.; Chen, H. D.; Chen, J.; Luo, Y.; Guo, H.; Jiang, R. D.; Liu, M. Q.; Chen, Y.; Shen, X. R.; Wang, X.; Zheng, X. S.; Zhao, K.; Chen, Q. J.; Deng, F.; Liu, L. L.; Yan, B.; Zhan, F. X.; Wang, Y. Y.; Xiao, G. F.; Shi, Z. L. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**, *579* (7798), 270−273.
(108) Meng, B.; Kemp, S. A.; Papa, G.; Datir, R.; Ferreira, I.; Marelli, S.; Harvey, W. T.; Lytras, S.; Mohamed, A.; Gallo, G.; Thakur, N.; Collier, D. A.; Mlcochova, P.; Consortium, C.-G. U.; Duncan, L. M.; Carabelli, A. M.; Kenyon, J. C.; Lever, A. M.; De Marco, A.; Saliba, C.; Culap, K.; Cameroni, E.; Matheson, N. J.; Piccoli, L.; Corti, D.; James, L. C.; Robertson, D. L.; Bailey, D.; Gupta, R. K.; et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. *Cell Rep* **2021**, *35* (13), 109292.
(109) Cober, O.; Cober, S. Omicron S:69-70 deletion and SARS-CoV-2 Detection Test Evasion. Literature Review, *ResearchGate*, February, 2022 DOI: 10.13140/RG.2.2.21976.78088 (accessed Dec 26, 2022).
(110) Ren, S. Y.; Wang, W. B.; Gao, R. D.; Zhou, A. M. Omicron variant (B.1.1.529) of SARS-CoV-2: Mutation, infectivity, transmission, and vaccine resistance. *World J. Clin Cases* **2022**, *10* (1), 1−11.
(111) Hou, Y. J.; Okuda, K.; Edwards, C. E.; Martinez, D. R.; Asakura, T.; Dinnon, K. H., 3rd; Kato, T.; Lee, R. E.; Yount, B. L.; Mascenik, T. M.; Chen, G.; Olivier, K. N.; Ghio, A.; Tse, L. V.; Leist, S. R.; Gralinski, L. E.; Schafer, A.; Dang, H.; Gilmore, R.; Nakano, S.; Sun, L.; Fulcher, M. L.; Livraghi-Butrico, A.; Nicely, N. I.; Cameron, M.; Cameron, C.; Kelvin, D. J.; de Silva, A.; Margolis, D. M.; Markmann, A.; Bartelt, L.; Zumwalt, R.; Martinez, F. J.; Salvatore, S. P.; Borczuk, A.; Tata, P. R.; Sontake, V.; Kimple, A.; Jaspers, I.; O'Neal, W. K.; Randell, S. H.; Boucher, R. C.; Baric, R. S. SARS-CoV-2 Reverse Genetics Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* **2020**, *182* (2), 429−446.
(112) Korber, B.; Fischer, W. M.; Gnanakaran, S.; Yoon, H.; Theiler, J.; Abfalterer, W.; Hengartner, N.; Giorgi, E. E.; Bhattacharya, T.; Foley, B.; Hastie, K. M.; Parker, M. D.; Partridge, D. G.; Evans, C. M.; Freeman, T. M.; de Silva, T. I.; Sheffield, C.-G. G.; McDanal, C.; Perez, L. G.; Tang, H.; Moon-Walker, A.; Whelan, S. P.; LaBranche, C. C.; Saphire, E. O.; Montefiori, D. C.; et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **2020**, *182* (4), 812−827.
(113) Zhang, Y.; Doruker, P.; Kaynak, B.; Zhang, S.; Krieger, J.; Li, H.; Bahar, I. Intrinsic dynamics is evolutionarily optimized to enable allosteric behavior. *Curr. Opin Struct Biol.* **2020**, *62*, 14−21.
(114) Yurkovetskiy, L.; Wang, X.; Pascal, K. E.; Tomkins-Tinch, C.; Nyalile, T. P.; Wang, Y.; Baum, A.; Diehl, W. E.; Dauphin, A.; Carbone, C.; Veinotte, K.; Egri, S. B.; Schaffner, S. F.; Lemieux, J. E.; Munro, J. B.; Rafique, A.; Barve, A.; Sabeti, P. C.; Kyratsous, C. A.; Dudkina, N. V.; Shen, K.; Luban, J. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020**, *183* (3), 739−751.
(115) Zhang, L.; Jackson, C. B.; Mou, H.; Ojha, A.; Rangarajan, E. S.; Izard, T.; Farzan, M.; Choe, H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv*, June 12, 2020 DOI: 10.1101/2020.06.12.148726 (accessed Dec 29, 2022).
(116) Daniloski, Z.; Jordan, T. X.; Ilmain, J. K.; Guo, X.; Bhabha, G.; tenOever, B. R.; Sanjana, N. E. The Spike D614G mutation increases SARS-CoV-2 infection of multiple human cell types. *Elife* **2021**, *10*, 1.
(117) Bai, C.; Wang, J.; Chen, G.; Zhang, H.; An, K.; Xu, P.; Du, Y.; Ye, R. D.; Saha, A.; Zhang, A.; Warshel, A. Predicting Mutational Effects on Receptor Binding of the Spike Protein of SARS-CoV-2 Variants. *J. Am. Chem. Soc.* **2021**, *143* (42), 17646−17654.
(118) Jangra, S.; Ye, C.; Rathnasinghe, R.; Stadlbauer, D.; Krammer, F.; Simon, V.; Martinez-Sobrido, L.; Garcia-Sastre, A.; Schotsaert, M. The E484K mutation in the SARS-CoV-2 spike protein reduces but does not abolish neutralizing activity of human convalescent and post-vaccination sera. medRxiv, January 29, 2021 DOI: 10.1101/2021.01.26.21250543 (accessed Dec 29, 2022).
(119) Nonaka, C. K. V.; Franco, M. M.; Graf, T.; de Lorenzo Barcia, C. A.; de Avila Mendonca, R. N.; de Sousa, K. A. F.; Neiva, L. M. C.; Fosenca, V.; Mendes, A. V. A.; de Aguiar, R. S.; Giovanetti, M.; de Freitas Souza, B. S. Genomic Evidence of SARS-CoV-2 Reinfection

Involving E484K Spike Mutation, Brazil. *Emerg Infect Dis* **2021**, *27* (5), 1522−1524.
(120) Rambaut, A.; Loman, N.; Pybus, O.; Barclay, W.; Barrett, J.; Carabelli, A.; Connor, T.; Peacock, T.; Robertson, D. L.; Volz, E.; et al. Preliminary genomic characterization of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations, 2020, https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed Dec 29, 2022).
(121) Leung, K.; Shum, M. H.; Leung, G. M.; Lam, T. T.; Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill* **2021**, *26* (1), 1.
(122) Gu, H.; Chen, Q.; Yang, G.; He, L.; Fan, H.; Deng, Y.-Q.; Wang, Y.; Teng, Y.; Zhao, Z.; Cui, Y.; et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **2020**, *369* (6511), 1603−1607.
(123) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H. D.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; King, N. P.; Veesler, D.; Bloom, J. D. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182* (5), 1295−1310.
(124) Norel, R.; Sheinerman, F.; Petrey, D.; Honig, B. Electrostatic contributions to protein−protein interactions: fast energetic filters for docking and their physical basis. *Protein Sci.* **2001**, *10* (11), 2147−2161.
(125) Sheinerman, F. B.; Norel, R.; Honig, B. Electrostatic aspects of protein−protein interactions. *Curr. Opin. Struct. Biol.* **2000**, *10* (2), 153−159.
(126) Fratev, F. N501Y and K417N Mutations in the Spike Protein of SARS-CoV-2 Alter the Interactions with Both hACE2 and Human-Derived Antibody: A Free Energy of Perturbation Retrospective Study. *J. Chem. Inf Model* **2021**, *61* (12), 6079−6084.
(127) Luan, B.; Wang, H.; Huynh, T. Enhanced binding of the N501Y-mutated SARS-CoV-2 spike protein to the human ACE2 receptor: insights from molecular dynamics simulations. *FEBS Lett.* **2021**, *595* (10), 1454−1461.
(128) Santos, J. C.; Passos, G. A. The high infectivity of SARS-CoV-2 B. 1.1. 7 is associated with increased interaction force between Spike-ACE2 caused by the viral N501Y mutation. *BioRxiv*, January 1, 2021, ver. 1 DOI: 10.1101/2020.12.29.424708 (accessed Jan 2, 2023).
(129) Liu, Y.; Liu, J.; Plante, K. S.; Plante, J. A.; Xie, X.; Zhang, X.; Ku, Z.; An, Z.; Scharton, D.; Schindewolf, C.; Widen, S. G.; Menachery, V. D.; Shi, P. Y.; Weaver, S. C. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. *Nature* **2022**, *602* (7896), 294−299.
(130) Kuzmina, A.; Korovin, D.; Lass, I. C.; Atari, N.; Ottolenghi, A.; Hu, P.; Mandelboim, M.; Rosental, B.; Rosenberg, E.; Diaz-Griffero, F.; Taube, R. Changes within the P681 residue of spike dictate cell fusion and syncytia formation of Delta and Omicron variants of SARS-CoV-2 with no effects on neutralization or infectivity. *Heliyon* **2023**, *9* (6), No. e16750.
(131) Mohammad, A.; Abubaker, J.; Al-Mulla, F. Structural modelling of SARS-CoV-2 alpha variant (B.1.1.7) suggests enhanced furin binding and infectivity. *Virus Res.* **2021**, *303*, 198522.
(132) Lubinski, B.; Fernandes, M. H. V.; Frazier, L.; Tang, T.; Daniel, S.; Diel, D. G.; Jaimes, J. A.; Whittaker, G. R. Functional evaluation of the P681H mutation on the proteolytic activation of the SARS-CoV-2 variant B.1.1.7 (Alpha) spike. *iScience* **2022**, *25* (1), 103589.
(133) Saito, A.; Irie, T.; Suzuki, R.; Maemura, T.; Nasser, H.; Uriu, K.; Kosugi, Y.; Shirakawa, K.; Sadamasu, K.; Kimura, I.; Ito, J.; Wu, J.; Iwatsuki-Horimoto, K.; Ito, M.; Yamayoshi, S.; Loeber, S.; Tsuda, M.; Wang, L.; Ozono, S.; Butlertanaka, E. P.; Tanaka, Y. L.; Shimizu, R.; Shimizu, K.; Yoshimatsu, K.; Kawabata, R.; Sakaguchi, T.; Tokunaga, K.; Yoshida, I.; Asakura, H.; Nagashima, M.; Kazuma, Y.; Nomura, R.; Horisawa, Y.; Yoshimura, K.; Takaori-Kondo, A.; Imai, M.; et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* **2022**, *602* (7896), 300−306.
(134) Sanda, M.; Morrison, L.; Goldman, R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal. Chem.* **2021**, *93* (4), 2003−2009.
(135) Tran, D. T.; Ten Hagen, K. G. Mucin-type O-glycosylation during development. *J. Biol. Chem.* **2013**, *288* (10), 6921−9.
(136) Zhang, L.; Mann, M.; Syed, Z. A.; Reynolds, H. M.; Tian, E.; Samara, N. L.; Zeldin, D. C.; Tabak, L. A.; Ten Hagen, K. G. Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (47), 2109905118 DOI: 10.1073/pnas.2109905118.
(137) McCallum, M.; Bassi, J.; De Marco, A.; Chen, A.; Walls, A. C.; Di Iulio, J.; Tortorici, M. A.; Navarro, M.-J.; Silacci-Fregni, C.; Saliba, C.; et al. SARS-CoV-2 immune evasion by the B. 1.427/B. 1.429 variant of concern. *Science* **2021**, *373* (6555), 648−654.
(138) Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N. G.; Decroly, E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* **2020**, *176*, 104742.
(139) Escalera, A.; Gonzalez-Reiche, A. S.; Aslam, S.; Mena, I.; Laporte, M.; Pearl, R. L.; Fossati, A.; Rathnasinghe, R.; Alshammary, H.; van de Guchte, A.; Farrugia, K.; Qin, Y.; Bouhaddou, M.; Kehrer, T.; Zuliani-Alvarez, L.; Meekins, D. A.; Balaraman, V.; McDowell, C.; Richt, J. A.; Bajic, G.; Sordillo, E. M.; Dejosez, M.; Zwaka, T. P.; Krogan, N. J.; Simon, V.; Albrecht, R. A.; van Bakel, H.; Garcia-Sastre, A.; Aydillo, T. Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. *Cell Host Microbe* **2022**, *30* (3), 373−387.
(140) Johnson, B. A.; Xie, X.; Bailey, A. L.; Kalveram, B.; Lokugamage, K. G.; Muruato, A.; Zou, J.; Zhang, X.; Juelich, T.; Smith, J. K.; Zhang, L.; Bopp, N.; Schindewolf, C.; Vu, M.; Vanderheiden, A.; Winkler, E. S.; Swetnam, D.; Plante, J. A.; Aguilar, P.; Plante, K. S.; Popov, V.; Lee, B.; Weaver, S. C.; Suthar, M. S.; Routh, A. L.; Ren, P.; Ku, Z.; An, Z.; Debbink, K.; Diamond, M. S.; Shi, P. Y.; Freiberg, A. N.; Menachery, V. D. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **2021**, *591* (7849), 293−299.
(141) So, M. K.; Park, S.; Lee, K.; Kim, S. K.; Chung, H. S.; Lee, M. Variant Prediction by Analyzing RdRp/S Gene Double or Low Amplification Pattern in Allplex SARS-CoV-2 Assay. *Diagnostics (Basel)* **2021**, *11* (10), 1854.
(142) McCallum, M.; De Marco, A.; Lempp, F. A.; Tortorici, M. A.; Pinto, D.; Walls, A. C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; Zepeda, S.; di Iulio, J.; Bowen, J. E.; Montiel-Ruiz, M.; Zhou, J.; Rosen, L. E.; Bianchi, S.; Guarino, B.; Fregni, C. S.; Abdelnabi, R.; Foo, S. C.; Rothlauf, P. W.; Bloyet, L. M.; Benigni, F.; Cameroni, E.; Neyts, J.; Riva, A.; Snell, G.; Telenti, A.; Whelan, S. P. J.; Virgin, H. W.; Corti, D.; Pizzuto, M. S.; Veesler, D. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **2021**, *184* (9), 2332−2347.
(143) Chen, J.; Wang, R.; Wang, M.; Wei, G. W. Mutations Strengthened SARS-CoV-2 Infectivity. *J. Mol. Biol.* **2020**, *432* (19), 5212−5226.
(144) Deshpande, A.; Harris, B. D.; Martinez-Sobrido, L.; Kobie, J. J.; Walter, M. R. Epitope Classification and RBD Binding Properties of Neutralizing Antibodies Against SARS-CoV-2 Variants of Concern. *Front Immunol* **2021**, *12*, 691715.
(145) Motozono, C.; Toyoda, M.; Zahradnik, J.; Saito, A.; Nasser, H.; Tan, T. S.; Ngare, I.; Kimura, I.; Uriu, K.; Kosugi, Y.; Yue, Y.; Shimizu, R.; Ito, J.; Torii, S.; Yonekawa, A.; Shimono, N.; Nagasaki, Y.; Minami, R.; Toya, T.; Sekiya, N.; Fukuhara, T.; Matsuura, Y.; Schreiber, G.; Ikeda, T.; Nakagawa, S.; Ueno, T.; Sato, K. SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **2021**, *29* (7), 1124−1136.
(146) Tchesnokova, V.; Kulasekara, H.; Larson, L.; Bowers, V.; Rechkina, E.; Kisiela, D.; Sledneva, Y.; Choudhury, D.; Maslova, I.; Deng, K.; et al. Acquisition of the L452R mutation in the ACE2-binding interface of Spike protein triggers recent massive expansion of

SARS-Cov-2 variants. *Journal of clinical microbiology* **2021**, *59* (11), e00921−21.
(147) Zhang, Y.; Zhang, T.; Fang, Y.; Liu, J.; Ye, Q.; Ding, L. SARS-CoV-2 spike L452R mutation increases Omicron variant fusogenicity and infectivity as well as host glycolysis. *Signal Transduct Target Ther* **2022**, *7* (1), 76.
(148) McCarthy, K. R.; Rennick, L. J.; Nambulli, S.; Robinson-McCarthy, L. R.; Bain, W. G.; Haidar, G.; Duprex, W. P. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **2021**, *371* (6534), 1139−1142.
(149) Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N. V.; Paciello, I.; Dal Monego, S.; Pantano, E.; Manganaro, N.; Manenti, A.; Manna, R.; Casa, E.; Hyseni, I.; Benincasa, L.; Montomoli, E.; Amaro, R. E.; McLellan, J. S.; Rappuoli, R. SARS-CoV-2 escape from a highly neutralizing COVID-19 convalescent plasma. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (36), e2103154118.
(150) Grabowski, F.; Preibisch, G.; Gizinski, S.; Kochanczyk, M.; Lipniacki, T. SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations. *Viruses* **2021**, *13* (3), 392.
(151) Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M. M.; Planchais, C.; Porrot, F.; Robillard, N.; Puech, J.; Prot, M.; Gallais, F.; Gantner, P.; Velay, A.; Le Guen, J.; Kassis-Chikhani, N.; Edriss, D.; Belec, L.; Seve, A.; Courtellemont, L.; Pere, H.; Hocqueloux, L.; Fafi-Kremer, S.; Prazuck, T.; Mouquet, H.; Bruel, T.; Simon-Loriere, E.; Rey, F. A.; Schwartz, O. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **2021**, *596* (7871), 276−280.
(152) Samarakoon, U.; Alvarez-Arango, S.; Blumenthal, K. G. Delayed Large Local Reactions to mRNA Covid-19 Vaccines in Blacks, Indigenous Persons, and People of Color. *N Engl J. Med.* **2021**, *385* (7), 662−664.
(153) Voloch, C. M.; da Silva Francisco, R.; de Almeida, L. G. P.; Cardoso, C. C.; Brustolini, O. J.; Gerber, A. L.; Guimaraes, A. P. C.; Mariani, D.; da Costa, R. M.; Ferreira, O. C.; Frauches, T. S.; de Mello, C. M. B.; Leitao, I. C.; Galliez, R. M.; Faffe, D. S.; Castineiras, T.; Tanuri, A.; de Vasconcelos, A. T. R. Genomic characterization of a novel SARS-CoV-2 lineage from Rio de Janeiro, Brazil. *J. Virol* **2021**, *95* (10), 1 DOI: 10.1128/JVI.00119-21.
(154) Wang, P.; Nair, M. S.; Liu, L.; Iketani, S.; Luo, Y.; Guo, Y.; Wang, M.; Yu, J.; Zhang, B.; Kwong, P. D.; Graham, B. S.; Mascola, J. R.; Chang, J. Y.; Yin, M. T.; Sobieszczyk, M.; Kyratsous, C. A.; Shapiro, L.; Sheng, Z.; Huang, Y.; Ho, D. D. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **2021**, *593* (7857), 130−135.
(155) Wibmer, C. K.; Ayres, F.; Hermanus, T.; Madzivhandila, M.; Kgagudi, P.; Oosthuysen, B.; Lambson, B. E.; de Oliveira, T.; Vermeulen, M.; van der Berg, K.; Rossouw, T.; Boswell, M.; Ueckermann, V.; Meiring, S.; von Gottberg, A.; Cohen, C.; Morris, L.; Bhiman, J. N.; Moore, P. L. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **2021**, *27* (4), 622−625.
(156) West, A. P., Jr.; Wertheim, J. O.; Wang, J. C.; Vasylyeva, T. I.; Havens, J. L.; Chowdhury, M. A.; Gonzalez, E.; Fang, C. E.; Di Lonardo, S. S.; Hughes, S.; Rakeman, J. L.; Lee, H. H.; Barnes, C. O.; Gnanapragasam, P. N. P.; Yang, Z.; Gaebler, C.; Caskey, M.; Nussenzweig, M. C.; Keeffe, J. R.; Bjorkman, P. J. Detection and characterization of the SARS-CoV-2 lineage B.1.526 in New York. *Nat. Commun.* **2021**, *12* (1), 4886.
(157) Suryadevara, N.; Shrihari, S.; Gilchuk, P.; VanBlargan, L. A.; Binshtein, E.; Zost, S. J.; Nargi, R. S.; Sutton, R. E.; Winkler, E. S.; Chen, E. C.; Fouch, M. E.; Davidson, E.; Doranz, B. J.; Chen, R. E.; Shi, P. Y.; Carnahan, R. H.; Thackray, L. B.; Diamond, M. S.; Crowe, J. E., Jr. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* **2021**, *184* (9), 2316−2331.
(158) Wang, P.; Casner, R. G.; Nair, M. S.; Wang, M.; Yu, J.; Cerutti, G.; Liu, L.; Kwong, P. D.; Huang, Y.; Shapiro, L.; Ho, D. D. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. *Cell Host Microbe* **2021**, *29* (5), 747−751.
(159) Hsu, S.; Parker, M. D.; Lindsey, B. B.; State, A.; Zhang, P.; Foulkes, B. H.; Louka, S. F.; Wolverson, P.; Gallis, M.; Hornsby, H. R., Detection of Spike Mutations; D80G, T95I, G142D, △144, N439K, E484K, P681H, I1130V, and D1139H, in B.1.1.482 Lineage (AV.1) Samples from South Yorkshire, UK, 2021. https://virological.org/t/detection-of-spike-mutations-d80g-t95i-g142d-144-n439k-e484k-p681h-i1130v-and-d1139h-in-b-1-1-482-lineage-av-1-samples-from-south-yorkshire-uk/699 (accessed Jan 14, 2023).
(160) Wu, F.; Zhao, S.; Yu, B.; Chen, Y. M.; Wang, W.; Song, Z. G.; Hu, Y.; Tao, Z. W.; Tian, J. H.; Pei, Y. Y.; Yuan, M. L.; Zhang, Y. L.; Dai, F. H.; Liu, Y.; Wang, Q. M.; Zheng, J. J.; Xu, L.; Holmes, E. C.; Zhang, Y. Z. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579* (7798), 265−269.
(161) Baker, E. N. Visualizing an unseen enemy; mobilizing structural biology to counter COVID-19. *Acta Crystallogr. F Struct Biol. Commun.* **2020**, *76* (4), 158−159.
(162) Subramaniam, S. COVID-19 and cryo-EM. *IUCrJ.* **2020**, *7* (4), 575−576.
(163) Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Bai, F.; Liu, H.; Liu, X.; Guddat, L. W.; Xu, W.; Xiao, G.; Qin, C.; Shi, Z.; Jiang, H.; Rao, Z.; Yang, H. Structure of M(pro) from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582* (7811), 289−293.
(164) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science* **2020**, *368* (6489), 409−412.
(165) Ishchenko, A.; Gati, C.; Cherezov, V. Structural biology of G protein-coupled receptors: new opportunities from XFELs and cryoEM. *Curr. Opin Struct Biol.* **2018**, *51*, 44−52.
(166) Maia, F. R. N. C.; White, T. A.; Loh, N. D.; Hajdu, J. CCP-FEL: a collection of computer programs for free-electron laser research. *J. Appl. Crystallogr.* **2016**, *49* (4), 1117−1120.
(167) Moreno, A. Advanced Methods of Protein Crystallization. *Methods Mol. Biol.* **2017**, *1607*, 51−76.
(168) Kuhlbrandt, W. Biochemistry. The resolution revolution. *Science* **2014**, *343* (6178), 1443−4.
(169) Adrian, M.; Dubochet, J.; Lepault, J.; McDowall, A. W. Cryo-electron microscopy of viruses. *Nature* **1984**, *308* (5954), 32−36.
(170) Crowther, R. A., The resolution revolution: recent advances in cryoEM. In *Methods in Ezymology*, 1st ed.; Crowther, R., Ed.; Academic Press: New York, 2016.
(171) Blundell, T. L.; Chaplin, A. K. The resolution revolution in X-ray diffraction, Cryo-EM and other Technologies. *Prog. Biophys. Mol. Biol.* **2021**, *160*, 2−4.
(172) Merk, A.; Bartesaghi, A.; Banerjee, S.; Falconieri, V.; Rao, P.; Davis, M. I.; Pragani, R.; Boxer, M. B.; Earl, L. A.; Milne, J. L. S.; Subramaniam, S. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. *Cell* **2016**, *165* (7), 1698−1707.
(173) *Welcome to the COVID-19 NMR Project*. https://covid19-nmr.de/ (accessed Feb 8, 2023).
(174) Cortese, M.; Lee, J. Y.; Cerikan, B.; Neufeldt, C. J.; Oorschot, V. M. J.; Kohrer, S.; Hennies, J.; Schieber, N. L.; Ronchi, P.; Mizzon, G.; Romero-Brey, I.; Santarella-Mellwig, R.; Schorb, M.; Boermel, M.; Mocaer, K.; Beckwith, M. S.; Templin, R. M.; Gross, V.; Pape, C.; Tischer, C.; Frankish, J.; Horvat, N. K.; Laketa, V.; Stanifer, M.; Boulant, S.; Ruggieri, A.; Chatel-Chaix, L.; Schwab, Y.; Bartenschlager, R. Integrative Imaging Reveals SARS-CoV-2-Induced Reshaping of Subcellular Morphologies. *Cell Host Microbe* **2020**, *28* (6), 853−866.
(175) Cariou, M.; Picard, L.; Gueguen, L.; Jacquet, S.; Cimarelli, A.; Fregoso, O. I.; Molaro, A.; Navratil, V.; Etienne, L. Distinct evolutionary trajectories of SARS-CoV-2-interacting proteins in bats and primates identify important host determinants of COVID-19. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119* (35), e2206610119.

(176) Grifoni, A.; Sidney, J.; Zhang, Y.; Scheuermann, R. H.; Peters, B.; Sette, A. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **2020**, *27* (4), 671−680.
(177) Hussain, I.; Pervaiz, N.; Khan, A.; Saleem, S.; Shireen, H.; Wei, D. Q.; Labrie, V.; Bao, Y.; Abbasi, A. A. Evolutionary and structural analysis of SARS-CoV-2 specific evasion of host immunity. *Genes Immun* **2020**, *21* (6−8), 409−419.
(178) Jungreis, I.; Sealfon, R.; Kellis, M. SARS-CoV-2 gene content and COVID-19 mutation impact by comparing 44 Sarbecovirus genomes. *Nat. Commun.* **2021**, *12* (1), 2642.
(179) Nunes, D. R.; Braconi, C. T.; Ludwig-Begall, L. F.; Arns, C. W.; Duraes-Carvalho, R. Deep phylogenetic-based clustering analysis uncovers new and shared mutations in SARS-CoV-2 variants as a result of directional and convergent evolution. *PLoS One* **2022**, *17* (5), No. e0268389.
(180) Singh, J.; Pandit, P.; McArthur, A. G.; Banerjee, A.; Mossman, K. Evolutionary trajectory of SARS-CoV-2 and emerging variants. *Virol J.* **2021**, *18* (1), 166.
(181) Sangeet, S.; Khan, A. Drug Discovery Against COVID-19 from Indian Medicinal Plants - Computational Approach. *Int. J. Eng. Appl. Sci. Technol.* **2020**, *5* (6), 324−330.
(182) Abel, R.; Paredes Ramos, M.; Chen, Q.; Perez-Sanchez, H.; Coluzzi, F.; Rocco, M.; Marchetti, P.; Mura, C.; Simmaco, M.; Bourne, P. E.; Preissner, R.; Banerjee, P. Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front Chem.* **2020**, *8*, 590263.
(183) Gil, C.; Ginex, T.; Maestro, I.; Nozal, V.; Barrado-Gil, L.; Cuesta-Geijo, M. A.; Urquiza, J.; Ramirez, D.; Alonso, C.; Campillo, N. E.; Martinez, A. COVID-19: Drug Targets and Potential Treatments. *J. Med. Chem.* **2020**, *63* (21), 12359−12386.
(184) Kumar, S.; Singh, B.; Kumari, P.; Kumar, P. V.; Agnihotri, G.; Khan, S.; Kant Beuria, T.; Syed, G. H.; Dixit, A. Identification of multipotent drugs for COVID-19 therapeutics with the evaluation of their SARS-CoV2 inhibitory activity. *Comput. Struct Biotechnol J.* **2021**, *19*, 1998−2017.
(185) Mirza, M. U.; Froeyen, M. Structural elucidation of SARS-CoV-2 vital proteins: Computational methods reveal potential drug candidates against main protease, Nsp12 polymerase and Nsp13 helicase. *J. Pharm. Anal* **2020**, *10* (4), 320−328.
(186) Rafi, M. O.; Al-Khafaji, K.; Sarker, M. T.; Taskin-Tok, T.; Rana, A. S.; Rahman, M. S. Design of a multi-epitope vaccine against SARS-CoV-2: immunoinformatic and computational methods. *RSC Adv.* **2022**, *12* (7), 4288−4310.
(187) Swiderek, K.; Moliner, V. Revealing the molecular mechanisms of proteolysis of SARS-CoV-2 M(pro) by QM/MM computational methods. *Chem. Sci.* **2020**, *11* (39), 10626−10630.
(188) Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; Zheng, M.; Chen, L.; Li, H. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin B* **2020**, *10* (5), 766−788.
(189) Zhou, Y.; Hou, Y.; Shen, J.; Huang, Y.; Martin, W.; Cheng, F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discov* **2020**, *6*, 14.
(190) Bastolla, U. Mathematical Model of SARS-Cov-2 Propagation Versus ACE2 Fits COVID-19 Lethality Across Age and Sex and Predicts That of SARS. *Front Mol. Biosci* **2021**, *8*, 706122.
(191) Davies, N. G.; Abbott, S.; Barnard, R. C.; Jarvis, C. I.; Kucharski, A. J.; Munday, J. D.; Pearson, C. A. B.; Russell, T. W.; Tully, D. C.; Washburne, A. D.; Wenseleers, T.; Gimma, A.; Waites, W.; Wong, K. L. M.; van Zandvoort, K.; Silverman, J. D.; Diaz-Ordaz, K.; Keogh, R.; Eggo, R. M.; Funk, S.; Jit, M.; Atkins, K. E.; Edmunds, W. J. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* **2021**, *372* (6538), 1 DOI: 10.1126/science.abg3055.
(192) Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Di Filippo, A.; Di Matteo, A.; Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **2020**, *26* (6), 855−860.
(193) Hoze, N.; Paireau, J.; Lapidus, N.; Tran Kiem, C.; Salje, H.; Severi, G.; Touvier, M.; Zins, M.; de Lamballerie, X.; Levy-Bruhl, D.; Carrat, F.; Cauchemez, S. Monitoring the proportion of the population infected by SARS-CoV-2 using age-stratified hospitalisation and serological data: a modelling study. *Lancet Public Health* **2021**, *6* (6), e408−e415.
(194) Sasmita, N. R.; Ikhwan, M.; Suyanto, S.; Chongsuvivatwong, V. Optimal control on a mathematical model to pattern the progression of coronavirus disease 2019 (COVID-19) in Indonesia. *Glob Health Res. Policy* **2020**, *5*, 38.
(195) Schioler, H.; Knudsen, T.; Brondum, R. F.; Stoustrup, J.; Bogsted, M. Mathematical modelling of SARS-CoV-2 variant outbreaks reveals their probability of extinction. *Sci. Rep* **2021**, *11* (1), 24498.
(196) Smith, M. R.; Trofimova, M.; Weber, A.; Duport, Y.; Kuhnert, D.; von Kleist, M. Rapid incidence estimation from SARS-CoV-2 genomes reveals decreased case detection in Europe during summer 2020. *Nat. Commun.* **2021**, *12* (1), 6009.
(197) Wang, S.; Hao, M.; Pan, Z.; Lei, J.; Zou, X. Data-driven multiscale mathematical modeling of SARS-CoV-2 infection reveals heterogeneity among COVID-19 patients. *PLoS Comput. Biol.* **2021**, *17* (11), e1009587.
(198) Chimmula, V. K. R.; Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos Solitons Fractals* **2020**, *135*, 109864.
(199) Tomar, A.; Gupta, N. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures. *Sci. Total Environ.* **2020**, *728*, 138762.
(200) Panwar, H.; Gupta, P. K.; Siddiqui, M. K.; Morales-Menendez, R.; Singh, V. Application of deep learning for fast detection of COVID-19 in X-Rays using nCOVnet. *Chaos Solitons Fractals* **2020**, *138*, 109944.
(201) Adam, D. C.; Wu, P.; Wong, J. Y.; Lau, E. H. Y.; Tsang, T. K.; Cauchemez, S.; Leung, G. M.; Cowling, B. J. Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **2020**, *26* (11), 1714−1719.
(202) Lai, S.; Ruktanonchai, N. W.; Zhou, L.; Prosper, O.; Luo, W.; Floyd, J. R.; Wesolowski, A.; Santillana, M.; Zhang, C.; Du, X.; Yu, H.; Tatem, A. J. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature* **2020**, *585* (7825), 410−413.
(203) Pearce, R.; Zhang, Y. Toward the solution of the protein structure prediction problem. *J. Biol. Chem.* **2021**, *297* (1), 100870.
(204) Modrek, B.; Lee, C. A genomic view of alternative splicing. *Nat. Genet.* **2002**, *30* (1), 13−9.
(205) Yu, Y.; Santat, L. A.; Choi, S. Bioinformatics Packages for Sequence Analysis. *Applied Mycology and Biotechnology* **2006**, *6*, 143−160.
(206) Brent, M. R.; Guigo, R. Recent advances in gene structure prediction. *Curr. Opin Struct Biol.* **2004**, *14* (3), 264−72.
(207) Ejigu, G. F.; Jung, J. Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology (Basel)* **2020**, *9* (9), 295.
(208) Steward, C. A.; Parker, A. P. J.; Minassian, B. A.; Sisodiya, S. M.; Frankish, A.; Harrow, J. Genome annotation for clinical genomic diagnostics: strengths and weaknesses. *Genome Med.* **2017**, *9* (1), 49.
(209) Yakimovich, A. Machine Learning and Artificial Intelligence for the Prediction of Host-Pathogen Interactions: A Viral Case. *Infect Drug Resist* **2021**, *14*, 3319−3326.
(210) Staquicini, D. I.; Tang, F. H. F.; Markosian, C.; Yao, V. J.; Staquicini, F. I.; Dodero-Rojas, E.; Contessoto, V. G.; Davis, D.; O'Brien, P.; Habib, N.; Smith, T. L.; Bruiners, N.; Sidman, R. L.; Gennaro, M. L.; Lattime, E. C.; Libutti, S. K.; Whitford, P. C.; Burley, S. K.; Onuchic, J. N.; Arap, W.; Pasqualini, R. Design and proof-of-concept for targeted phage-based COVID-19 vaccination strategies with a streamlined cold-free supply chain. *PNAS* **2021**, *118* (30), e2105739118 DOI: 10.1073/pnas.2105739118.

(211) Markwick, P. R.; Malliavin, T.; Nilges, M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS Comput. Biol.* **2008**, *4* (9), e1000168.
(212) Oliveira Junior, A. B.; Lin, X.; Kulkarni, P.; Onuchic, J. N.; Roy, S.; Leite, V. B. P. Exploring Energy Landscapes of Intrinsically Disordered Proteins: Insights into Functional Mechanisms. *J. Chem. Theory Comput* **2021**, *17* (5), 3178−3187.
(213) Wu, H.; Wolynes, P. G.; Papoian, G. A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2018**, *122* (49), 11115−11125.

https://doi.org/10.1021/acsomega.3c00944
ACS Omega 2023, 8, 23283−23304