The FEBS Journal

FEBS PRESS science publishing by scientists

MINIREVIEW

SARS-CoV-2 variants preferentially emerge at intrinsically disordered protein sites helping immune evasion

Federica Quaglia[1,2], Edoardo Salladini[2], Marco Carraro[2], Giovanni Minervini[2], Silvio C.E. Tosatto[2] and Philippe Le Mercier[3]

1 Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies, National Research Council (CNR-IBIOM), Bari, Italy
2 Department of Biomedical Sciences, University of Padova, Italy
3 Swiss-Prot group, SIB Swiss Institute of Bioinformatics, Geneva, Switzerland

Keywords
biocuration; DisProt; IDPs; immune escape; intrinsically disordered proteins; mutations; SARS-CoV-2; variants; ViralZone

Correspondence
P. Le Mercier, Swiss-Prot group, SIB Swiss Institute of Bioinformatics, CMU, 1 Michel Servet, 1211 Geneva 4, Switzerland
Tel: +41 22 379 58 70
E-mail: Philippe.Lemercier@sib.swiss
and
S. C. E. Tosatto, Department of Biomedical Sciences, University of Padova, Padova, Italy
Tel: +39 049 827 6269
E-mail: silvio.tosatto@unipd.it

Federica Quaglia and Edoardo Salladini contributed equally to this work

(Received 26 November 2021, revised 21 January 2022, accepted 31 January 2022)

doi:10.1111/febs.16379

The SARS-CoV-2 pandemic is maintained by the emergence of successive variants, highlighting the flexibility of the protein sequences of the virus. We show that experimentally determined intrinsically disordered regions (IDRs) are abundant in the SARS-CoV-2 viral proteins, making up to 28% of disorder content for the S1 subunit of spike and up to 51% for the nucleoprotein, with the vast majority of mutations occurring in the 13 major variants mapped to these IDRs. Strikingly, antigenic sites are enriched in IDRs, in the receptor-binding domain (RBD) and in the N-terminal domain (NTD), suggesting a key role of structural flexibility in the antigenicity of the SARS-CoV-2 protein surface. Mutations occurring in the S1 subunit and nucleoprotein (N) IDRs are critical for immune evasion and antibody escape, suggesting potential additional implications for vaccines and monoclonal therapeutic strategies. Overall, this suggests the presence of variable regions on S1 and N protein surfaces, which confer sequence and antigenic flexibility to the virus without altering its protein functions.

Introduction

Intrinsically Disordered Proteins (IDPs) are a widespread class of diverse proteins characterized by lack of a fixed 3D structure [1]. IDPs are well known players of multiple biological processes, such as nucleic acid binding, signalling, cell cycle regulation, and play a central role in a large number of physiological and pathological processes [2]. Although widely distributed in eukaryotes, the widest content is found among viruses [3], where IDPs have evolved to support virus-related biological functions [4,5]. Disordered proteins represent an important class of antigens in a variety of human pathogens and can be targets of protective antibody responses [6].

Abbreviations
IDP, intrinsically disordered protein; IDR, intrinsically disordered region; N, nucleoprotein; NTD, N-terminal domain; RBD, receptor-binding domain; RBM, receptor-binding motif; S, spike glycoprotein; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; VOC, variant of concern; VOI, variant of interest.

The presence of protein intrinsic disorder was also highlighted in the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteome [7–9]. In particular, both spike glycoprotein (S) and nucleoprotein (N) are nowadays well known to contain functionally relevant disordered regions (IDRs) [7–9]. Since the onset of the COVID-19 pandemic, several SARS-CoV-2 variants have been identified worldwide [10], affecting the epidemiology of the virus, and playing an important role in pandemic surveillance and control [11,12]. Mutations that affect the viral genome and potentially impact disease transmission and severity are referred to as variants of concern (VOC) and variants of interest (VOI), and the scientific community is increasingly dedicated to monitoring the emergence of new viral lineages worldwide. The most variable proteins are spike and nucleoprotein, which are also the major antigenic proteins [13].

In this work, we use manually curated structural data to describe the disordered regions of SARS-CoV-2—as a collaboration between leading data resources, UniProt [14], ViralZone [15] and DisProt [16,17]—focusing on the spike protein and nucleoprotein. Many different SARS-CoV-2 variants have been observed: there are 1737 lineages described in PANGO (https://cov-lineages.org/index.html/cite) as of December 2021. We chose to analyse the 13 Variants Of Concern (VOC) and the Variants Of Interest (VOI)—including Omicron—as they represent the most widespread and best adapted to humans (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/). We analyse mutation localization for these 13 major variants of the SARS-CoV-2 virus and uncover hotspots that correlate not only with disordered regions but also with immune evasion. Finally, we highlight the role of flexible regions in the major antigenic site of the spike protein, suggesting a role of intrinsic disorder in escaping the host immune response.

Results

SARS-CoV-2 spike and nucleoprotein are enriched in IDRs

Intrinsically disordered proteins are characterized by the presence of unstructured segments, that is, intrinsically disordered regions (IDRs), that lack a stable tertiary structure. Intrinsic disorder in proteins can be identified by several experimental techniques, including biophysical and biochemical methods, the most widely used being X-ray crystallography, nuclear magnetic resonance (NMR), circular dichroism and small-angle X-ray scattering [18,19]. Using the information available in DisProt, the major repository of manually curated data of IDPs and IDRs from literature data, we investigated the presence of IDRs in the SARS-CoV-2 proteins, along with their interactions and functions [16,17]. By analysing published structures and raw experimental data, we investigated IDR regions in nucleoprotein, spike, E protein, ORF1ab, ORF3a and ORF7a proteins. We focused our analysis on those proteins playing a crucial role in the virus–host interaction, and targets of vaccines and antibodies development, that is, proteins spike and nucleoprotein [20,21].

Analysis revealed that several regions are omitted in the structures of SARS-CoV-2 spike glycoprotein (protein S, DisProt: DP02772) due to their flexibility. No apparent density can be detected for region 455-490 [7]: this region of the Receptor-Binding Motif (RBM) is indeed unstructured and flexible in the unbound conformation [7,8] and undergoes folding-upon-binding in the ACE2-bound form [22,23].

The IDR between S1 and S2 (673–686) [7] is required for the proteolytic processing essential for the viral entry into host cells [24]. An insertion at position 680–687, that includes the specific furin-like cleavage motif RRxR, has been shown to be absent in other beta coronaviruses such as SARS-CoV [25].

Several sterically accessible complex-type glycans were identified inside the IDRs of SARS-CoV-2 spike glycoprotein (N74, N149 and three positions in the unstructured C terminus, N1158, N1173, N1194) as characterized by mass spectrometry experiments [26]. As protein glycosylation is a well-established strategy adopted by viruses to evade host immunity [27], molecular dynamic simulations highlighted that glycans extensively shield the spike protein surface from antibody recognition [28]. Nevertheless, we found no significant correlation between glycan sites and IDR in spike protein.

SARS-CoV-2 nucleoprotein (protein N, DisProt: DP03212) is a 419-residue multidomain protein characterized by 52% of disorder content that include the unstructured N- and C-termini, along with a disordered flexible linker connecting the RNA-binding domain (RBD) and the dimerization domain [29]. The disordered N terminus plays a role in liquid–liquid phase separation of protein N, indeed its deletion strongly decreases phase separation in the presence of RNA, while slightly increasing turbidity and droplet formation in the absence of RNA [30]. Similarly, a deletion of the flexible linker (region 174-247) strongly reduces LLPS-associated droplet formation and turbidity [30]. NMR titration experiments characterizing the interaction of polyU with the protein N SR-peptide, region 182-197 inside the flexible linker that connects the two

globular domains, indicate that the interaction strength decreases in the phosphorylated form. Moreover, phosphorylation of full-length nucleoprotein affects its RNA-induced phase separation, resulting in a weaker interaction of protein N with RNA and an increased diffusion of the phosphorylated species inside polyU-induced droplets [31]. The C-terminal IDR, instead, is not required for nucleoprotein condensation with RNA via LLPS [31]. The N-terminal and C-terminal IDRs were also found to be involved in the binding of nucleocapsid-targeting single-domain antibodies (sdAbs), sdAbs-N5 and sdAb-N6, whose interaction with the nucleoprotein requires the presence of its intrinsically disordered termini [32]. Size-exclusion chromatography studies of the nucleoprotein in RNA-bound states and RNA-free state showed that truncations of its N-terminal IDR impair the RNA binding and that both the N-terminal and C-terminal IDRs contribute to RNA-binding activity of the SARS-CoV-2 nucleoprotein [33]. Finally, the C-terminal disordered region seems to play a role in droplet formation [33].

S1 and N mutation hotspots cluster in unstructured regions

Since late 2020, the SARS-CoV-2 pandemic has been driven by the emergence of variants [34]. These lineages carry fixed mutations that increase the viral fitness while enhancing the spread of the virus at population level. Our analysis reveals that nonsynonymous mutations tend to cluster in hotspots (Fig. 1,2),

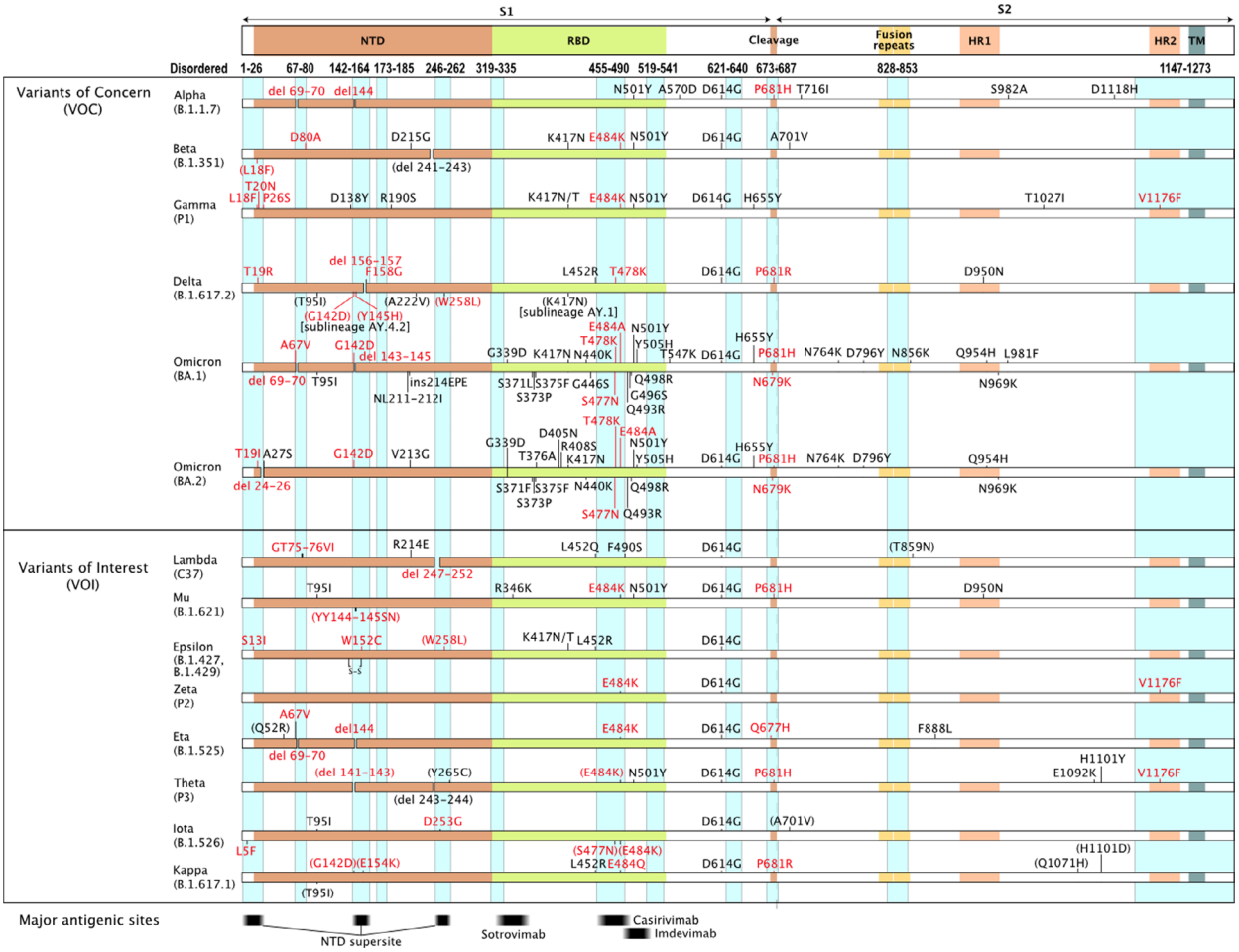

Fig. 1. Amino acid changes in the spike protein of Variants of Concern (VOC) Alpha, Beta, Gamma, Delta, Omicron BA.1 and BA.2; Variants of Interest (VOI) Lambda, Mu, Epsilon, Zeta, Eta, Theta, Iota and Kappa. Disordered regions are indicated by cyan columns, and variants in disordered regions are coloured in red. Parentheses indicate variants whose prevalence is < 80% but > 10%. The main regions are annotated: S1 with N-terminal domain (NTD) and receptor-binding domain (RBD); S2 with fusion peptides, heptad repeat 1 (HR1) and 2 (HR2) and the transmembrane domain (TM) [73]. Major antigenic sites are shown below with the NTD supersite [56], and monoclonal antibody-binding regions for sotrovimab [74], casirivimab and imdevimab [75,76].

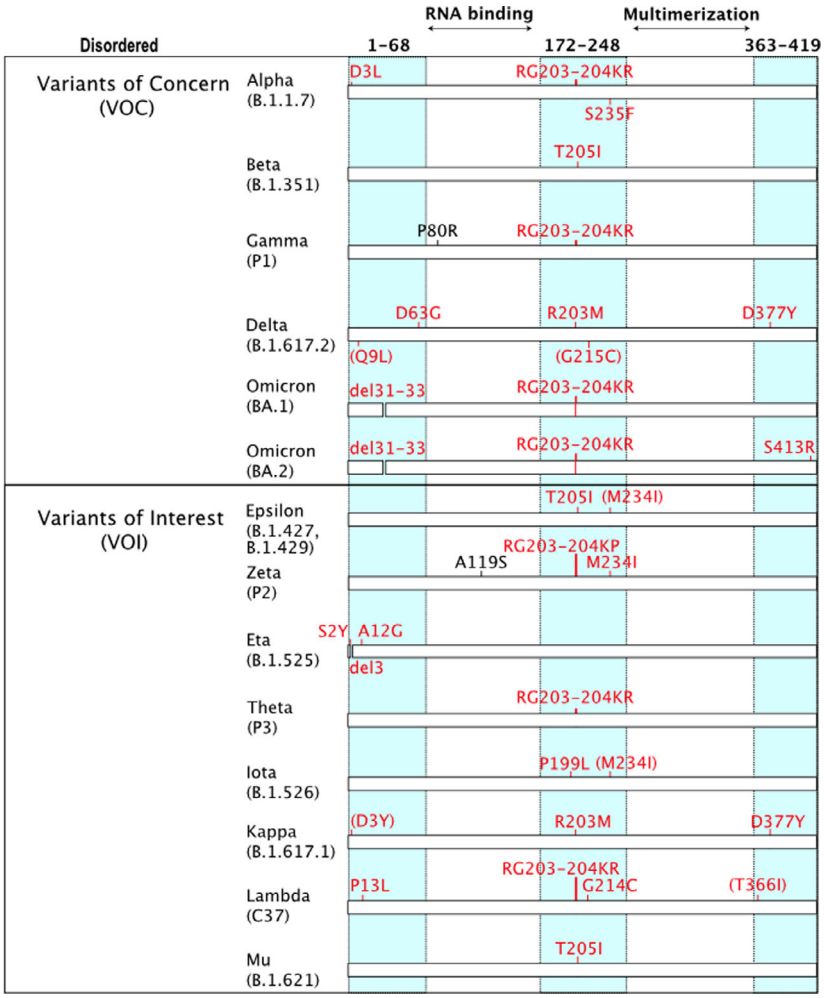

Fig. 2. Amino acid changes in the nucleoprotein of Variants of Concern Alpha, Beta, Gamma, Delta, Omicron BA.1 and BA.2; Variants of Interest (VOI) Lambda, Mu, Epsilon, Zeta, Eta, Theta, Iota and Kappa. Disordered regions are indicated by cyan columns, and variants in disordered regions are coloured in red. Parentheses indicate variants whose prevalence is < 80% but > 10%.

suggesting the presence of variable disordered regions. Such features in viral surface proteins may influence viral antigenicity and/or tropism. The external loop domain III of dengue virus envelope protein is disordered and plays a role in selective host binding ([35], DisProt: DP00876). Moreover, it is the major target of highly neutralizing and protective serotype-specific antibodies [36]. Similarly, the HIV-1 glycoprotein is characterized by multiple variable loops that are intrinsically disordered [37] and play a role in immune evasion [38] and coreceptor binding [39]. To assess the presence of variable disordered regions in SARS-CoV-2, we compared the substitutions/deletions found in the 13 major variants classified by WHO (January 2022) (https://www.who.int/en/activities/tracking-SARS- CoV-2-variants/) with the experimentally determined IDRs (Fig. 1,2,3), identifying a strong correlation among mutations and disordered regions in SARS-CoV-2 spike protein and nucleoprotein. For instance, mutations in the S1 subunit of the spike glycoprotein tend to cluster in hotspots at the N terminus and occur in its unstructured regions—32 out of 45 mutated positions accounting for 71% of variants are localized inside S1 IDRs, whereas the S2 chain variants do not (Table 1). Similarly, 16 out of 18 mutated positions in SARS-CoV-2 nucleoprotein (N) are localized inside its IDRs, accounting for 89% of variants affecting protein N (Table 1).

For all the other SARS-CoV-2 proteins for which we gathered intrinsic disorder data, the observed mutations either did not correlate with known IDRs, or there were too few mutations to be significant. Here, we provide an insight on the intrinsic disorder and mutation content of SARS-CoV-2 ORF3a, E protein, ORF7a and ORF1ab (Table 2, Fig. 4,5).

ORF3a (DisProt: DP03003): electron cryomicroscopy experiments of the protein shed light on the intrinsic disorder of its N- and C-terminal regions [40]. Point mutations disrupting the TRAF-binding region of ORF3a (residues 36–40) lack the ability to activate

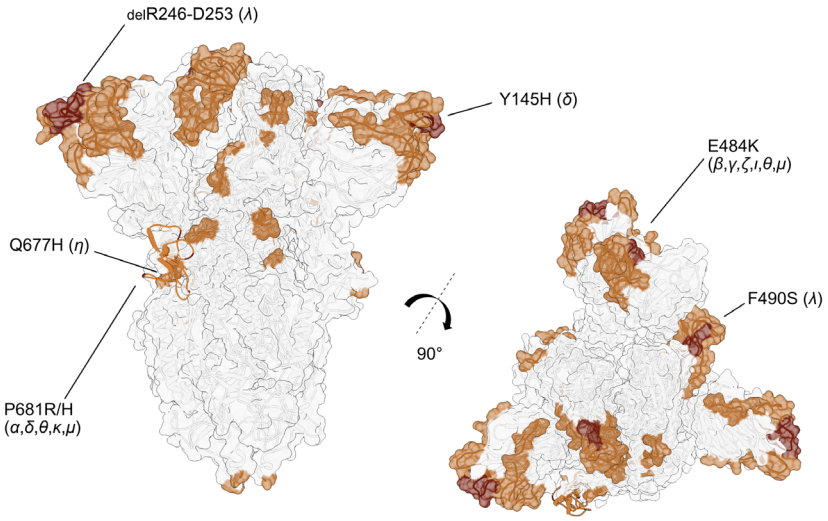

Fig. 3. Immune escape-related mutations mapped on the IDRs of the spike protein (structure in closed conformation) [61]. The disordered regions—according to the DisProt database (protein S, DisProt: DP02772) - are coloured in light brown on the structure, while mutations are highlighted in dark brown. Molecular graphics were performed using UCSF Chimera [71].

Table 1. Disorder content in SARS-CoV-2 proteins according to DisProt, mutation prevalence across 12 VOC and VOI lineages (except Omicron) (*mut*) and the mutations mapped to the IDRs of spike and nucleoprotein (mut_{IDR}/mut). Mutations and variants data retrieved from https://outbreak.info/, intrinsic disorder data from https://disprot.org/.

	disorder content (%)	*mut*	mut_{IDR}	mut_{IDR}/mut
Spike (S1)	28	45	32	0.71
Spike (S2)	26	10	1	0.10
Nucleoprotein (N)	52	18	16	0.89

Table 2. Disorder content in SARS-CoV-2 proteins according to DisProt, mutation prevalence across VOC and VOI lineages (*mut*) and the mutations mapped to the IDRs of ORF3a, E protein, ORF7a and ORF1ab (mut_{IDR}/mut). Mutations and variants data retrieved from https://outbreak.info/, intrinsic disorder data from https://disprot.org/.

	disorder content (%)	*mut*	mut_{IDR}	mut_{IDR}/mut
ORF3a	28	12	5	0.42
E protein	20	3	1	0.33
ORF7a	11.6	3	0	0
ORF1ab	3.9	55	1	0.02

either IL-1β or IL-8–Luc secretion, highlighting the role of ORF3a in NF-κB and NLRP3 inflammasome activation [41]. The ORF3a unstructured N terminus is also responsible for its subcellular localization, for instance a deletion of the first 41 residues increases its expression in the plasma membrane while impairing localization to internal membranes [40]. Finally, 42% of the mutations affecting ORF3a in the variants here described are localized in its disordered N- and C termini: T9I (peculiar to Omicron variant), I20M (Mu), S26L (Delta and Kappa), S253P (Gamma), del257 and V259L (Mu).

E protein (DisProt: DP03450): NMR data indicate that E, a 75-residue-long protein, exhibits a higher mobility in its N-terminal (2–7) and C-terminal (61–75) regions. The central region is characterized by structured elements, that is, a transmembrane helix (8–43) and a cytoplasmic helix (53–60) [42]. A single mutation, P71L in the Beta variant, is localized in the highly mobile C-terminal region of the E protein.

ORF7a protein (DisProt: DP03460): X-ray crystallography of the SARS-CoV-2 ORF7a ectodomain (PDB: 7CI3, residues 14–96) shows that this protein (121 aa) is characterized by a well-defined structure and visible electron density from residues 14 to 82. Residues 83–96 are instead not visible in the electron density map, indicating the presence of structural disorder in the ORF7 protein, followed by a transmembrane domain (97–116) and an ER-retention signal (117–121) not included in the crystal structure [43]. No mutations are found inside the IDR of ORF7a identified so far.

ORF1ab (DisProt: DP02925): Several unstructured regions were identified in the replicase polyprotein 1ab, although the structural characterization of several of its regions is still missing in the scientific literature. Residues 1–147 of ORF1ab:NSP1 are unstructured and include a flexible linker, spanning region 129–147, that connects the disordered N-terminal domain of Nsp1 and its C-terminal domain [44]. Similarly, IDRs are found in ORF1ab:NSP3 (residues 1782–1796), ORF1ab:NSP8 (residues 3931–4020) and ORF1ab:NSP10 (residues 4254–4271) [45–48]. To date, only mutation S135R in the Omicron BA.2 lineage maps to an IDR.

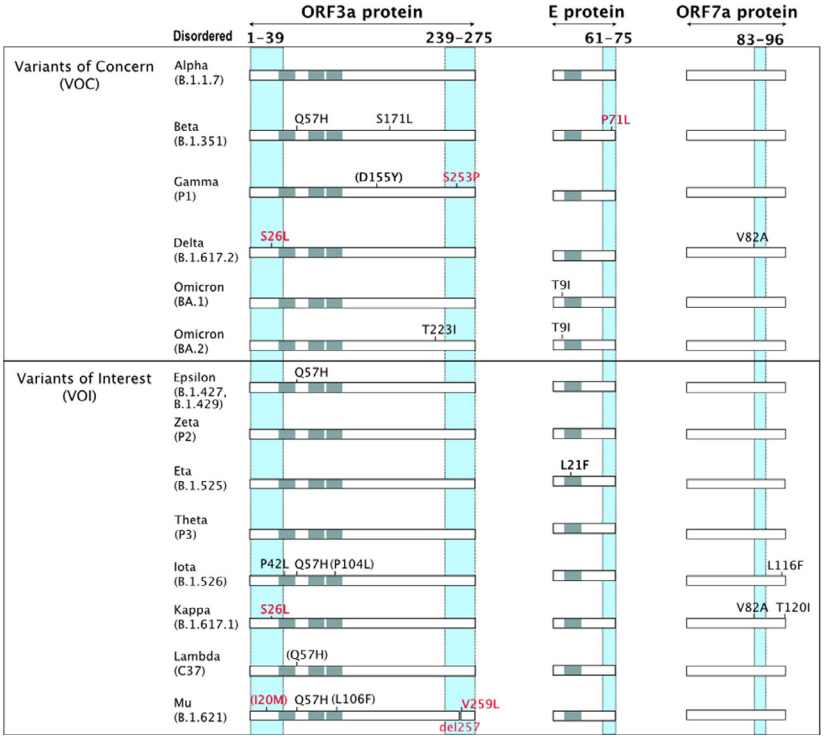

Fig. 4. Mutations in the VOCs and VOI lineages mapped on the sequences of ORF3a, E protein and ORF7a. VOCs and VOIs lineages are represented, along with the mutations falling inside (red) and outside (black) IDRs. IDRs are represented as cyan columns while transmembrane regions are in grey.

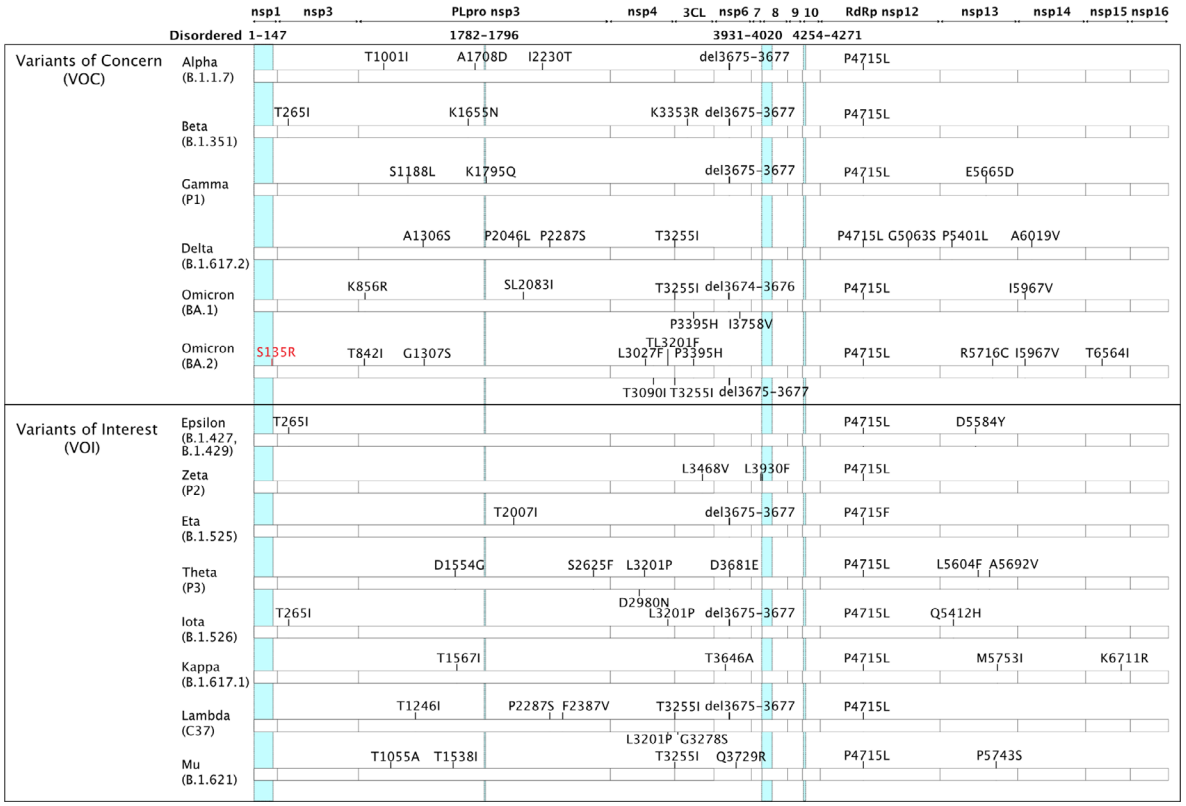

Fig. 5. Mutations in the VOCs and VOI lineages mapped on the sequences of ORF1ab polyprotein. VOCs and VOIs lineages are represented, along with the mutations falling inside (red) and outside (black) IDRs. IDRs are represented as cyan columns.

Omicron variant

During the time this paper was submitted, the Omicron variant appeared [49]. This variant is unusual in that it has more than 30 mutations localized in the spike glycoprotein, so many that it escapes most therapeutic monoclonal antibodies and, to a large extent, vaccine-triggered antibodies [50,51]. The variant presents a large number of mutated positions in the S1 region ($n = 31$), with a significant number mapping to disordered regions (53%) although less than the 12 previous variants (71%) (Table 3). This may be due to the tremendous acceleration of evolution that has led to omicron emergence, not yet completely understood [52]. Interestingly, in the Omicron variant and its lineages, all the mutated positions in the nucleoprotein are found in disordered regions. Specifically, P13L and del31-33 are localized in the unstructured N terminus, while R203K and G204R are inside the intrinsically disordered linker connecting the N-terminal domain with the C-terminal domain. Finally, although the P13L, R203K and G204R substitutions have already been identified in other variants, the deletion affecting positions 31–33 and S413R missense mutation are peculiar to Omicron (https://outbreak.info/compare-lineages?pango = Omicron).

Antigenic drift is closely associated with SARS-CoV-2 IDRs

The major SARS-CoV-2-specific antibody responses target the spike glycoprotein (S1 subunit) [8,53]. Two major antigenic regions are present in the S1 subunit: the receptor-binding domain (RBD) and the N-terminal domain (NTD) [54].

The RBD is the main antigenic site to which neutralizing antibodies bind, and this region includes three IDRs. Many neutralizing antibodies target the receptor-binding motif (RBM, pos. 438-506) in the RBD [8,55]. They act by preventing binding to the host receptor or reducing attachment to the host cell [54,55]. The inner part of this region is unstructured (pos. 455–490) [7,8] and it folds when interacting with the ACE2 receptor [22,23].

Table 3. Disorder content in Omicron BA.1 and BA.2 SARS-CoV-2 proteins according to DisProt, mutation prevalence (*mut*) and the mutations mapped to the IDRs of spike and nucleoprotein (mut_{IDR}/*mut*). Mutations and variants data retrieved from https://outbreak.info/, intrinsic disorder data from https://disprot.org/.

	disorder content (%)	*mut*	mut_{IDR}	mut_{IDR}/*mut*
Spike (S1)	28	39	20	0.51
Spike (S2)	26	8	0	0
Nucleoprotein (N)	52	6	6	1

The NTD contains an antigenic supersite to which neutralizing antibodies bind [56]. Interestingly, this supersite corresponds to the first three IDRs where most of the variation occurs [54,57]. These three regions behave similarly to the variable loops in flavivirus envelope or HIV gp120: unstructured regions that allow the virus to escape immunity through a high potential for variation [56,58].

Antibody recognition of disordered epitopes is particularly sensitive to epitope variation [6]. A recent study analysed viral mutations that occurred in immunocompromised patients, and found out that most mutations are observed in either the NTD supersite or the RBM [59]. The flexibility of the IDR regions allows variants to escape neutralization by many antibodies, as shown by the resistance of Beta and Gamma variants to bamlanivimab and casirivimab treatments [50]. In particular, E484K substitution—localized in the IDR within the RBM—triggers immune evasion against casirivimab monoclonal antibodies [60]. In addition, Q677H and deletion 246-253 in the eta and lambda variants confer a better resistance to neutralizing antibodies [61].

A superantigen-like motif—absent in other SARS family beta coronaviruses—has been identified in the spike of SARS-CoV-2. This motif, corresponding to the furin cleavage site at position 681–684 (PRRA) [62], was proposed to be a high-affinity site for T-cell receptor (TCR) β-Chain and may play a crucial role in the immune inflammation responsible for severe cases of COVID [63]. Strikingly this motif at position 681–684 maps to an intrinsically disordered region of the spike protein, moreover P681 is a mutational hotspot in SARS-CoV-2 variants Alpha, Delta, Kappa, Mu (Fig. 1,3).

The nucleocapsid is the second major antigen of SARS-CoV-2 [64]. Early studies on SARS-CoV showed that the immunodominant epitopes are located in regions 1–69, 153–235 and 354–422 [65], corresponding to the three disordered domains conserved in both SARS-CoV and SARS-CoV-2.

Collectively, these findings suggest that the immunodominant epitopes of the S1 subunit and of the N protein are closely associated with the disordered regions in the SARS-CoV-2 proteins.

Discussion

Intrinsically disordered regions (IDRs), protein regions characterized by a lack of stable three-dimensional structure, are present and abundant in native SARS-

CoV-2 proteins. The IDRs described here were identified by screening the associated scientific literature and the data retrieved were subsequently manually curated into DisProt and integrated with information from ViralZone. These IDRs have been shown to be associated with hotspots of mutations in spike S1 protein and nucleoprotein. Substitutions and deletions falling inside unstructured regions are likely to have a minor impact on the protein folding [66,67]. Moreover we show that these disordered regions overlap with major antigenic sites. IDRs are known to be specific targets of antibody recognition [6] and this variability might have an impact on antibodies' binding specificity. Our results suggest that SARS-CoV-2 displays disordered regions (IDRs) on the spike S1 subunit and on the N protein, and that these regions are enriched in mutations that could provide the virus with an advantage both for genetic and antigenic drift.

These findings are particularly important in light of emerging variants, such as the delta subvariant AY.4.2, which is being monitored by the European Centre for Disease Prevention and Control (ECDC, https://www.ecdc.europa.eu/) and the World Health Organization (WHO, https://www.who.int/). The major mutation associated with the AY.4.2 variant, Y145H, is located in an IDR of the spike glycoprotein and is structurally close to the known immunodominant epitope at position 153–235 (Fig. 1,3), possibly playing a role in viral immune defence. Omicron variants have a higher amount of mutations in S1 IDRs (20) than any other variants. It combines all the high-consequence mutations identified in previous variants and has an unexpected ability to evade vaccine protection. In addition, it has an enormous number of mutations (19) in structured regions of the protein, making it distinctly different from previous variants. This suggests that Omicron arose under different selective pressures. Indeed, early studies suggest that the Omicron may have arisen in chronically infected COVID-19 patients [52] or infected animals [68].

The proposed correlation between intrinsic disorder with mutational hotspots and major antigenic sites may have potential implications for the management of the SARS-CoV-2 pandemic and associated disease. Treatment of severe COVID patients depends on monoclonal antibodies, which in turn relies on their ability to recognize specific epitopes. Mutations in the targeted epitopes may inhibit the binding of monoclonal antibodies and reduce the therapeutic effect of this treatment [69]. Given the established link between IDR and mutation hotspot, it may be beneficial in the long term to select monoclonal antibodies that target ordered regions. Similarly, vaccine development could benefit from knowing where the key variable regions of the spike protein are located.

Materials and methods

Identification and annotation of intrinsically disordered regions

The presence of IDRs in each SARS-CoV-2 protein was manually curated based on the most relevant literature data as well as with manual visual inspection of crystallographic and raw structural data. In addition, we combined our annotations with information retrieved from UniProt [14], (https://www.uniprot.org/) and ViralZone [15] (https://viralzone.expasy.org/). The intrinsically disordered regions (IDRs) were then annotated in DisProt, the database for manually curated intrinsically disordered proteins [16,17] (https://disprot.org/). Each SARS-CoV-2 protein described corresponds to a specific entry in DisProt: spike glycoprotein (DisProt: DP02772), nucleoprotein (DisProt: DP03212), ORF1ab (DisProt: DP02925), E protein (DisProt: DP03450), ORF7a protein (DisProt: DP03460) and ORF3a (DisProt: DP03003).

Identification and mapping of mutations on IDRs

The analysis of SARS-CoV-2 mutations, both missense and deletions, relies on variants of concern (VOC), that is, Alpha, Beta, Gamma, Delta and Omicron, and variants of interest (VOI), that is, Epsilon, Zeta, Eta, Theta, Iota, Kappa, Lambda and Mu, by using the outbreak.info resource (https://outbreak.info/). Mutations with at least a minimum prevalence of 80% were considered for the analysis and then mapped on the previously identified IDRs in the spike glycoprotein and Nucleoprotein of SARS-CoV-2.

The trimeric spike protein structure (PDB: 6ZGG [70]) was built using Chimera to display mutations specifically affecting viral immune escape and antibody evasion [71]. Disordered region 677–689, missing from the spike structure, was modelled on the chain A starting from the sequence using RANCH [72].

Acknowledgements

This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 778247 and No 952334, from the Italian Ministry of University and Research (MIUR), PRIN 2017 under grant agreement No 2017483NH8 and from the Swiss Federal Government through the State Secretariat for Education, Research and Innovation SERI. We are thankful to Dr. Daniel Kolakofsky and Dr. Alan Bridge

for helpful discussions. The Graphical Abstract was partially created with BioRender.com. Open Access Funding provided by Universita degli Studi di Padova within the CRUI-CARE Agreement.

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

PLM and FQ conceived the study. ES, FQ and PLM performed the data curation and analysed the data. PLM and SCET supervised the project. FQ, ES, MC, GM, SCET and PLM contributed to writing, critically reviewing and editing the manuscript.

Peer review

The peer review history for this article is available at https://publons.com/publon/10.1111/febs.16379.

Data availability statement

Data on intrinsically disordered regions (IDRs) can be found in DisProt (https://disprot.org/): spike glycoprotein (DisProt: DP02772), nucleoprotein (DisProt: DP03212), ORF1ab (DisProt: DP02925), E protein (DisProt: DP03450), ORF7a protein (DisProt: DP03460) and ORF3a (DisProt: DP03003). Data on SARS-CoV-2 variants are stored in the ViralZone resource https://viralzone.expasy.org/9556.

References

1 Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, et al. Thousands of proteins likely to have long disordered regions. *Pac Symp Biocomput Pac Symp Biocomput*. 1998;**437**:448.

2 van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev*. 2014;**114**:6589–631.

3 Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J Biomol Struct Dyn*. 2012;**30**:137–49.

4 Dyson HJ, Wright PE. How do intrinsically disordered viral proteins hijack the cell? *Biochemistry*. 2018;**57**:4045–6.

5 Mishra PM, Verma NC, Rao C, Uversky VN, Nandi CK. Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis. *Prog Mol Biol Transl Sci*. 2020;**174**: 1–78.

6 MacRaild CA, Richards JS. Anders RF & Norton RS (2016) antibody recognition of disordered antigens. *Structure*. 1993;**24**:148–57.

7 Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*. 2020;**367**:1260–3.

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

9 Guseva S, Perez LM, Camacho-Zarco A, Bessa LM, Salvi N, Malki A, et al. 1H, 13C and 15N Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. *Biomol NMR Assign*. 2021;**15**:255–60.

10 Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol*. 2020;**5**:1403–7.

11 Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution and epidemic spread of SARS-CoV-2 in Brazil. *Science*. 2020;**369**:1255–60.

12 Deng X, Gu W, Federman S, du Plessis L, Pybus OG, Faria NR, et al. Genomic surveillance reveals multiple introductions of SARS-CoV-2 into Northern California. *Science*. 2020;**369**:582–7.

13 Fenwick C, Croxatto A, Coste AT, Pojer F, André C, Pellaton C, et al. Changes in SARS-CoV-2 Spike versus nucleoprotein antibody responses impact the estimates of infections in population-based seroprevalence studies. *J Virol*. 2021;**95**:e01828–20.

14 UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;**49**:D480–9.

15 Hulo C, de Castro E, Masson P, Bougueleret L, Bairoch A, Xenarios I, et al. ViralZone: a knowledge resource to understand virus diversity. *Nucleic Acids Res*. 2011;**39**:D576–82.

16 Hatos A, Hajdu-Soltész B, Monzon AM, Palopoli N, Álvarez L, Aykac-Fas B, et al. DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res*. 2019;**25**:975.

17 Quaglia F, Mészáros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. *Nucleic Acids Res*. 2022;**50**:D480–7.

18 Receveur-Bréchot V, Bourhis J-M, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. *Proteins Struct Funct Bioinform*. 2005;**62**:24–45.

19 Quaglia F, Hatos A, Piovesan D, Tosatto SCE. Exploring manually curated annotations of intrinsically disordered proteins with DisProt. *Currr Protoc Bioinform*. 2020;**72**:11.

20 Zumla A, Chan JFW, Azhar EI, Hui DSC, Yuen K-Y. Coronaviruses - drug discovery and therapeutic options. *Nat Rev Drug Discov*. 2016;**15**:327–47.

21 Dai L, Gao GF. Viral targets for vaccines against COVID-19. *Nat Rev Immunol*. 2021;**21**:73–82.

22 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**:1444–8.

23 Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*. 2020;**581**:215–20.

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

25 Örd M, Faustova I, Loog M. The sequence at Spike S1/S2 site enables cleavage by furin and phospho-regulation in SARS-CoV2 but not in SARS-CoV1 or MERS-CoV. *Sci Rep*. 2020;**10**:16944.

26 Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*. 2020;**369**:330–3.

27 Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol*. 2007;**15**:211–8.

28 Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci Rep*. 2020;**10**:14991.

29 Cubuk J, Alston JJ, Incicco JJ, Singh S, Stuchell-Brereton MD, Ward MD, et al. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun*. 2021;**12**:1936.

30 Perdikari TM, Murthy AC, Ryan VH, Watters S, Naik MT, Fawzi NL. SARS-CoV-2 nucleocapsid protein phase-separates with RNA and with human hnRNPs. *EMBO J*. 2020;**39**.

31 Savastano A, Ibáñez de Opakua A, Rankovic M, Zweckstetter M. Nucleocapsid protein of SARS-CoV-2 phase separates into RNA-rich polymerase-containing condensates. *Nat Commun*. 2020;**11**:6041.

32 Jia Z, Liu C, Chen Y, Jiang H, Wang Z, Yao J, et al. Crystal structures of the SARS-CoV-2 nucleocapsid protein C-terminal domain and development of nucleocapsid-targeting nanobodies. *FEBS J*. 2021;**16**:239.

33 Wu C, Qavi AJ, Hachim A, Kavian N, Cole AR, Moyle AB, et al. Characterization of SARS-CoV-2 nucleocapsid protein reveals multiple functional consequences of the C-terminal domain. *iScience*. 2021;**24**:102681.

34 Fontanet A, Autran B, Lina B, Kieny MP, Karim SSA, Sridhar D. SARS-CoV-2 variants and ending the COVID-19 pandemic. *The Lancet*. 2021;**397**:952–4.

35 Hung J-J, Hsieh M-T, Young M-J, Kao C-L, King C-C, Chang W. An external loop region of domain III of dengue virus type 2 envelope protein is involved in serotype-specific binding to mosquito but not mammalian cells. *J Virol*. 2004;**78**:378–88.

36 Fahimi H, Mohammadipour M, Haddad Kashani H, Parvini F, Sadeghizadeh M. Dengue viruses and promising envelope protein domain III-based vaccines. *Appl Microbiol Biotechnol*. 2018;**102**:2977–96.

37 Zolla-Pazner S, Cardozo T. Structure–function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. *Nat Rev Immunol*. 2010;**10**:527–35.

38 O'Connell RJ, Kim JH, Excler J-L. The HIV-1 gp120 V1V2 loop: structure, function and importance for vaccine development. *Expert Rev Vaccines*. 2014;**13**:1489–500.

39 Distefano M, Lanzarotti E, Fernández MF, Mangano A, Martí M, Aulicino P. Identification of novel molecular determinants of co-receptor usage in HIV-1 subtype F V3 envelope sequences. *Sci Rep*. 2020;**10**:12583.

40 Kern DM, Sorum B, Mali SS, Hoel CM, Sridharan S, Remis JP, et al. Cryo-EM structure of SARS-CoV-2 ORF3a in lipid nanodiscs. *Nat Struct Mol Biol*. 2021;**28**:573–82.

41 Siu K, Yuen K, Castano-Rodriguez C, Ye Z, Yeung M, Fung S, et al. Severe acute respiratory syndrome Coronavirus ORF3a protein activates the NLRP3 inflammasome by promoting TRAF3-dependent ubiquitination of ASC. *FASEB J*. 2019;**33**:8865–77.

42 Park SH, Siddiqi H, Castro DV, De Angelis AA, Oom AL, Stoneham CA, et al. Interactions of SARS-CoV-2 envelope protein with amilorides correlate with antiviral activity. *PLoS Pathog*. 2021;**17**:e1009519.

43 Zhou Z, Huang C, Zhou Z, Huang Z, Su L, Kang S, et al. Structural insight reveals SARS-CoV-2 ORF7a as an immunomodulating factor for human CD14+ monocytes. *iScience*. 2021;**24**:102187.

44 Schubert K, Karousis ED, Jomaa A, Scaiola A, Echeverria B, Gurzeler L-A, et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat Struct Mol Biol*. 2020;**27**:959–66.

45 Shin D, Mukherjee R, Grewe D, Bojkova D, Baek K, Bhattacharya A, et al. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*. 2020;**587**:657–62.

46 Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, et al. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science*. 2020;**368**:779–82.

47 Wang Q, Wu J, Wang H, Gao Y, Liu Q, Mu A, et al. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*. 2020;**182**:417–428.e13.

48 Lin S, Chen H, Ye F, Chen Z, Yang F, Zheng Y, et al. Crystal structure of SARS-CoV-2 nsp10/nsp16 2'-O-methylase and its implication on antiviral drug design. *Signal Transduct Target Ther*. 2020;**5**:131.
49 Viana R, Moyo S, Amoako DG, Tegally H, Scheepers C, Althaus CL, et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022;**41586**:3832.
50 Planas D, Saunders N, Maes P, Guivel-Benhassine F, Planchais C, Buchrieser J, et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature*. 2021;**38**:389.
51 Cele S, Jackson L, Khoury DS, Khan K, Moyo-Gwete T, Tegally H, et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*. 2021;**38**:824.
52 Kupferschmidt K. Where did "weird" Omicron come from? *Science*. 2021;**374**:1179.
53 Batra M, Tian R, Zhang C, Clarence E, Sacher CS, Miranda JN, et al. Role of IgG against N-protein of SARS-CoV2 in COVID19 clinical outcomes. *Sci Rep*. 2021;**11**:3455.
54 Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature*. 2020;**584**:450–6.
55 Errico JM, Zhao H, Chen RE, Liu Z, Case JB, Ma M, et al. Structural mechanism of SARS-CoV-2 neutralization by two murine antibodies targeting the RBD. *Cell Rep*. 2021;**37**:109881.
56 Lok S-M. An NTD supersite of attack. *Cell Host Microbe*. 2021;**29**:744–6.
57 Corey L, Beyrer C, Cohen MS, Michael NL, Bedford T, Rolland M. SARS-CoV-2 Variants in Patients with Immunosuppression. *N Engl J Med*. 2021;**385**:562–6.
58 Goh GK-M, Dunker AK, Uversky VN. Correlating Flavivirus virulence and levels of intrinsic disorder in shell proteins: protective roles vs. immune evasion. *Mol Biosyst*. 2016;**12**:1881–91.
59 Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature*. 2021;**593**:130–5.
60 Li Q, Nie J, Wu J, Zhang L, Ding R, Wang H, et al. SARS-CoV-2 501Y.V2 variants lack higher infectivity but do have immune escape. *Cell*. 2021;**184**:2362–2371.e9.
61 Zeng C, Evans JP, Faraone JN, Qu P, Zheng Y-M, Saif L, et al. Neutralization of SARS-CoV-2 Variants of Concern Harboring Q677H. *MBio*. 2021;**12**:e0251021.
62 Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, 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.
63 Cheng MH, Zhang S, Porritt RA, Noval Rivas M, Paschold L, Willscher E, et al. Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation. *Proc Natl Acad Sci USA*. 2020;**117**:25254–62.
64 Burbelo PD, Riedo FX, Morishima C, Rawlings S, Smith D, Das S, et al. Sensitivity in Detection of Antibodies to Nucleocapsid and Spike Proteins of Severe Acute Respiratory Syndrome Coronavirus 2 in Patients With Coronavirus Disease 2019. *J Infect Dis*. 2020;**222**:206–13.
65 He Y, Zhou Y, Wu H, Kou Z, Liu S, Jiang S. Mapping of Antigenic Sites on the Nucleocapsid Protein of the Severe Acute Respiratory Syndrome Coronavirus. *J Clin Microbiol*. 2004;**42**:5309–14.
66 Lin Y-S, Hsu W-L, Hwang J-K, Li W-H. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol*. 2007;**24**:1005–11.
67 Brown CJ, Takayama S, Campen AM, Vise P, Marshall TW, Oldfield CJ, et al. Evolutionary rate heterogeneity in proteins with long disordered regions. *J Mol Evol*. 2002;**55**:104–10.
68 Wei C, Shan K-J, Wang W, Zhang S, Huan Q, Qian W. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *J Genet Genomics Yi Chuan Xue Bao*. 2021;**48**:1111–21.
69 Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*. 2021;**596**:276–80.
70 Wrobel AG, Benton DJ, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol*. 2020;**27**:763–7.
71 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;**25**:1605–12.
72 Tria G, Mertens HDT, Kachala M, Svergun DI. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*. 2015;**2**:207–17.
73 Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science*. 2005;**309**:1864–8.
74 Pinto D, Park Y-J, Beltramello M, Walls AC, Tortorici MA, Bianchi S, et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*. 2020;**583**:290–5.
75 Jaworski JP. Neutralizing monoclonal antibodies for COVID-19 treatment and prevention. *Biomed J*. 2021;**44**:7–17.
76 Hurt AC, Wheatley AK. Neutralizing Antibody Therapeutics for COVID-19. *Viruses*. 2021;**13**:628.