

Intrinsically Disordered Proteins: Perspective on COVID-19 Infection and Drug Discovery

Rumiana Tenchov and Qiongqiong Angela Zhou*

Cite This: <https://doi.org/10.1021/acsinfectdis.2c00031>

Read Online

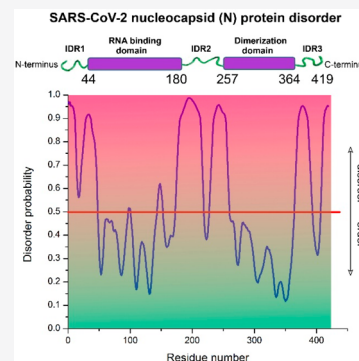
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Since the beginning of the COVID-19 pandemic caused by SARS-CoV-2, millions of patients have been diagnosed and many of them have died from the disease worldwide. The identification of novel therapeutic targets are of utmost significance for prevention and treatment of COVID-19. SARS-CoV-2 is a single-stranded RNA virus with a 30 kb genome packaged into a membrane-enveloped virion, transcribing several tens of proteins. The belief that the amino acid sequence of proteins determines their 3D structure which, in turn, determines their function has been a central principle of molecular biology for a long time. Recently, it has been increasingly realized, however, that there is a large group of proteins that lack a fixed or ordered 3D structure, yet they exhibit important biological activities—so-called intrinsically disordered proteins and protein regions (IDPs/IDRs). Disordered regions in viral proteins are generally associated with viral infectivity and pathogenicity because they endow the viral proteins the ability to easily and promiscuously bind to host proteins; therefore, the proteome of SARS-CoV-2 has been thoroughly examined for intrinsic disorder. It has been recognized that, in fact, the SARS-CoV-2 proteome exhibits significant levels of structural order, with only the nucleocapsid (N) structural protein and two of the nonstructural proteins being highly disordered. The spike (S) protein of SARS-CoV-2 exhibits significant levels of structural order, yet its predicted percentage of intrinsic disorder is still higher than that of the spike protein of SARS-CoV. Noteworthy, however, even though IDPs/IDRs are not common in the SARS-CoV-2 proteome, the existing ones play major roles in the functioning and virulence of the virus and are thus promising drug targets for rational antiviral drug design. Presented here is a COVID-19 perspective on the intrinsically disordered proteins, summarizing recent results on the SARS-CoV-2 proteome disorder features, their physiological and pathological relevance, and their prominence as prospective drug target sites.

KEYWORDS: COVID-19, intrinsically disordered protein, SARS-CoV-2, proteome, nucleocapsid protein, spike protein, drug design



1. INTRODUCTION

For a long time, one of the central principles of molecular biology has been the belief that the amino acid sequence of each protein determines its three-dimensional structure which, in turn, determines its function. Recently, it has been increasingly realized, however, that there is a large group of proteins and protein regions that lack a fixed or ordered 3D structure, yet they exhibit biological activities—so-called intrinsically disordered proteins and intrinsically disordered regions (IDPs/IDRs) (Figure 1).^{1–3} The highly dynamic disordered regions of these proteins have been linked to important phenomena such as enzyme catalysis and allosteric regulation and vital physiological functions such as cell signaling and transcription. They are also key players in the cellular liquid–liquid phase separation driving the formation of membraneless organelles, allowing the concentration of biomolecules to increase biochemical reaction efficiency and protection of nucleic acids or proteins to promote cell survival under stress.^{4,5} In viral proteins, disordered regions have been strongly correlated to the viral infectivity and pathogenicity because they provide the viral proteins with the ability to easily

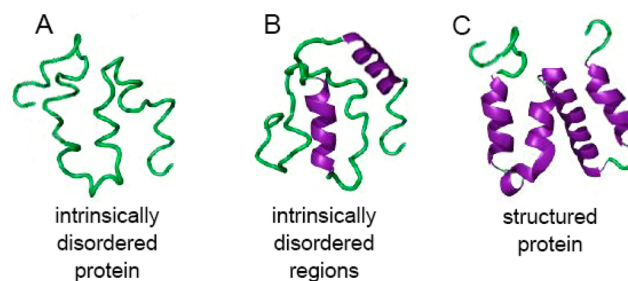


Figure 1. Schematic presentation of (A) intrinsically disordered proteins (IDPs), (B) intrinsically disordered regions (IDRs), and (C) structured proteins.

Received: January 13, 2022

and promiscuously bind to host proteins. With the COVID-19 disease caused by infection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) rapidly spreading around the world, the search for correlations between the viral protein disorder and the infection pathway has been intensified, hoping to find ways to mitigate the effects of the viral infection. Furthermore, the practice of rational drug design has largely ignored so far the presence of intrinsic disorder in target proteins. Understanding of the structure of these regions in the COVID-19 proteome would be a valuable asset for high-throughput screening in drug design and development for treatment of COVID-19.

In fact, the SARS-CoV-2 proteome has been found to exhibit significant levels of structural order. Except for the nucleocapsid (N) structural protein and two of the non-structural proteins (ORF6 and ORF9b), the majority of SARS-CoV-2 proteins are predominantly highly ordered proteins, with a few disordered regions.⁶ The spike (S) protein of SARS-CoV-2 exhibits significant levels of structural order, yet its predicted percentage of intrinsic disorder is still higher than that of the spike protein of SARS-CoV. Furthermore, the IDRs located in the SARS-CoV-2 proteins appear to be of high functional importance. Most of the SARS-CoV-2 proteins contain intrinsic disorder-based protein–protein interaction sites utilized for molecular recognition and interaction with certain partner proteins. The accumulated knowledge on the SARS-CoV-2 proteome is important for understanding the virulence of coronaviruses and to find ways to alleviate the effects of the viral infections.

Here, along with a concise review of the common knowledge on IDPs, we present a COVID-19 perspective on the intrinsically disordered proteins, summarizing and analyzing recent results on the SARS-CoV-2 proteome disorder features and their physiological and pathological relevance. As IDPs/IDRs typically undergo structural transitions upon attaching to their physiological associates, such knowledge generates an important base for better understanding the activity of these proteins, their interactions with host proteins, and their prominence as prospective drug target sites.

2. INTRINSICALLY DISORDERED PROTEINS

2.1. IDPs Lack a Fixed or Ordered 3D Structure. IDPs/IDRs do not exhibit a fixed three-dimensional structure. Instead, they fold dynamically into a series of conformations depending on the surrounding conditions.^{7–9} This allows them to have a wide range of binding associates and thus serve significant roles in critical biological processes such as cell signaling and transcription.^{2,10} It has been clear for quite some time now that IDPs/IDRs are functionally important and abundant in proteins implicated across the disease spectrum.¹¹

Whereas ordered proteins usually have a single, well-defined conformation representing a global free-energy minimum with the potential to bind small molecules with high affinity, the free-energy landscape of disordered proteins is characterized by a large number of local minima. These minima correspond to the many conformations within the structural ensemble populated by disordered proteins, which can transiently bind small molecules with weak affinity. Examples exist of disordered proteins interacting with other proteins or nucleic acids in which they undergo disorder-to-order transitions, resulting in low affinity complexes.^{12–15} Due to the difference in their structure, ordered and disordered proteins exhibit different hydration degrees. The hydration is significantly

higher for the IDPs in comparison to the similar size globular proteins.¹⁶ IDPs also exhibit a high propensity of binding to charged solute ions.^{17,18}

The IDPs are abundant in living systems. Eukaryotes generally have the highest amount of IDPs in their proteomes, estimated between 30 and 45%.^{19,20} More than half of eukaryotic proteins have long regions of disorder. Intrinsic disorder is particularly enriched in proteins implicated in cell signaling and transcription and are subject of tight control by the organisms.²¹ IDPs play key roles in many biological processes. They have numerous crucial functions that complement the functionality of ordered proteins. Highly dynamic disordered proteins have been related to functionally important processes such as allosteric regulation and enzyme catalysis. For many disordered proteins, the binding affinity to their receptors is regulated by post-translational modifications. The flexibility of disordered proteins facilitates the conformational requirements for binding enzymes and their receptors. IDPs are involved in regulation of transcription and translation, cellular signal transduction, protein phosphorylation, and the regulation of the self-assembly of large multiprotein complexes.

Many intrinsically disordered proteins undergo transitions to more ordered states or fold into stable secondary or tertiary structures upon binding to their targets—i.e., a coupled folding and binding route. Coupled folding and binding often produces a complex with high specificity and relatively low affinity, which is suitable for signal transduction proteins that must not only associate specifically to initiate the signaling process but also be capable of dissociation upon signaling completion. Another advantage of a system that undergoes coupled folding and binding is that the conformational flexibility expedites the post-translational modification of important transcription factors.

The difference between ordered and disordered proteins starts at the level of their amino acid sequences. There are noticeable differences between ordered and disordered proteins in terms of their amino acid compositions, charge, flexibility, and hydrophobicity.²² Thus, the tendency of a protein to stay intrinsically disordered is programmed in its amino acid sequence.^{23,24} Indeed, IDPs are notably depleted in bulky hydrophobic (Ile, Leu, and Val) and aromatic amino acid residues (Trp, Tyr, Phe), which normally create the hydrophobic core of a folded globular protein and also have low content of the hydrophobic and uncharged Cys and Asn residues. These lessened residues are considered order-promoting amino acids. On the other hand, natively unfolded proteins are substantially enriched in polar, hydrophilic, disorder-promoting amino acids: Ala, Arg, Gly, Gln, Ser, Pro, Glu, and Lys. Many IDPs have a substantial excess of basic or acidic amino acids and are thus highly charged at neutral pH. The charge destabilizes a compact structure on such proteins. The relationship between amino acid composition and sequence and protein order or disorder has been carefully explored in an effort to develop predictors of intrinsic disorder. At present, there are a multitude of such predictors based on a variety of amino acid attributes.^{25–31}

IDPs are suggested to be relevant to the origin of life. Scientists had long hypothesized that the early genetic code was simpler than it is now. The number of protein-building amino acids have been supposedly 12 or 14 amino acids rather than 20.³² The last of the 20 amino acids to evolve was tryptophan, the largest amino acid and also the most structure-promoting one. It has been speculated that the earliest proteins

have been disordered and may have played a unique and crucial role in the origin of life. Even more, it has been suggested that the amino acids of the earliest proteins lacked aromatic rings, which are required to stabilize the active site of the structured proteins, thus the outcome is that the later-evolving amino acids may have been key to developing structure.³² The evolution of intrinsic disorder has been supposed to exhibit a wavy pattern, with disordered primordial proteins having mainly chaperone activities gradually substituted by highly ordered enzymes and protein intrinsic disorder reinvented at subsequent evolutionary steps along with development of more complex organisms,^{33–35} playing substantial roles in the evolution of multifunctionality.^{36–38} Moreover, the relationship between structural disorder and organism complexity, along with proteome size, has been discussed, suggesting that structural disorder may effectively increase the complexity of the species.^{39,40}

Interest in IDPs in protein science has been rapidly increasing in the years since 2000, as demonstrated by a search in CAS Content Collection.⁴¹ Currently, there are over 6000 publications on IDPs-related topics (Figure 2).

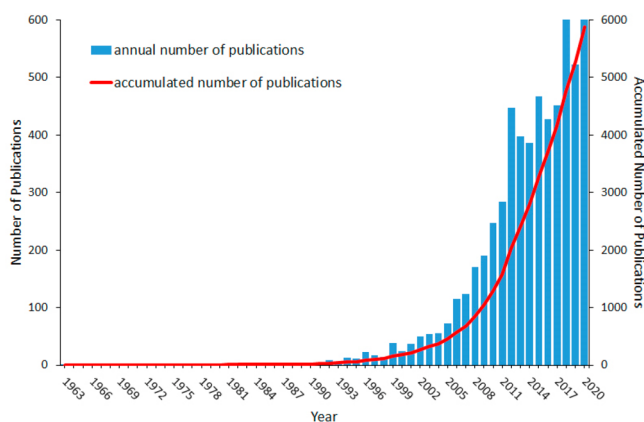


Figure 2. Annual number and accumulated number of IDP-related publications in CAS Content Collection.⁴¹

2.2. IDPs in Diseases. IDPs have numerous crucial biological functions that complement the functionality of ordered proteins. However, when malfunction occurs (e.g., misexpression, misprocessing, or misregulation), IDPs/IDRs tend to engage into some undesirable interactions and get involved in the development of various pathological states. As a matter of fact, many proteins which are associated with neurodegeneration, diabetes, cardiovascular disease, amyloidosis, and genetic diseases, as well as the majority of the human cancer-related proteins, are either IDPs or contain long IDRs.^{42–48} Recently, IDPs were suggested to be responsible for exothermic events observed in cerebrospinal fluid and brain proteome. These reversible exothermic transitions vary depending on neurodegenerative pathologies and possibly reflect processes of protein fibrillization and/or aggregation.^{49,50} Further, it has become clear that IDPs such as α -synuclein and tau protein are common among neurodegenerative diseases, especially Parkinson's disease. It has been recognized that α -synuclein can form amyloid fibrils, which are implicated with the pathogenesis of Parkinson's disease and other neurodegenerative conditions, collectively termed synucleinopathies.⁵¹ IDPs, such as α -synuclein, tau protein, p53, and BRCA1, are appropriate targets for drugs

modulating protein–protein interactions. From these and other examples, novel strategies for drug discovery based on IDPs have been developed.⁵²

The COVID-19 pandemic has been ongoing for almost 2 years. Although various treatments have been explored, efficient antiviral drugs are currently still in short supply. Targeting the IDPs/IDRs in the SARS-CoV-2 proteome could be an alternative strategy for rational antiviral drug design.

3. INTRINSICALLY DISORDERED PROTEINS IN SARS-CoV-2 PROTEOME AND THEIR ROLE IN COVID-19 INFECTION

Since the beginning of the COVID-19 pandemic caused by the SARS-CoV-2 virus, millions of patients have been diagnosed and many of them have died from the disease worldwide. The identification of novel therapeutic targets are of utmost significance for the prevention and treatment of COVID-19. The SARS-CoV-2 virus is a single-stranded RNA virus with a 30 kb genome packaged into a membrane-enveloped virion 80–90 nm in diameter, transcribing several tens of proteins (Figure 3).⁵³

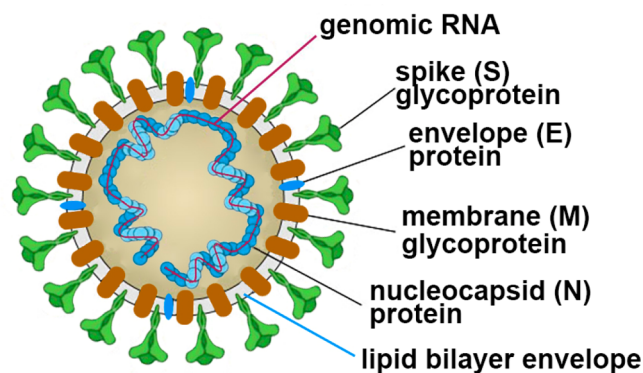


Figure 3. Schematic diagram of the coronavirus particle. The viral genomic RNA is assembled with the nucleocapsid (N) protein, which is enclosed by the lipid bilayer membrane. The viral surface proteins—spike (S), envelope (E), and membrane (M) proteins—are embedded in the lipid bilayer envelope.

SARS-CoV-2 forms a virion including its genomic RNA bundled in a particle comprising four structural proteins—spike (S) glycoprotein that binds to human angiotensin converting enzyme 2 (ACE2) receptor to mediate the entry of the virus into the host cell,⁵⁴ the membrane (M) protein facilitating viral assembly in the endoplasmic reticulum,^{55,56} the ion channel small envelope (E) protein,⁵⁷ and the nucleocapsid (N) protein,^{58–60} which assembles with viral RNA to form a ribonucleoprotein complex—the nucleocapsid.^{61,62} A recent extensive NMR exploration provided thorough knowledge on the structure of the SARS-CoV-2 proteins, which is required for understanding the basic principles of the viral life cycle and processes underlying viral infection and transmission.⁶³ Moreover, protocols for the large-scale production of more than 80% of all SARS-CoV-2 proteins are provided, highly valuable for further explorations.

As it is known that viral proteomes typically contain noticeable levels of intrinsically disordered proteins and viral proteins utilize intrinsic disorder during host cell invasion, it is important to examine the intrinsic disorder of the viral proteins associated with the SARS-CoV-2 infection.⁶⁴ Intrinsically

disordered proteins and intrinsically disordered regions play key roles in vital biological processes, including DNA/RNA and protein binding. The RNA–protein recognition often needs conformational changes in both RNA and protein, which is facilitated by the structural flexibility of disordered residues.⁶⁵ Knowledge related to coronavirus proteomes from the viewpoint of the intrinsic disorder propensities can provide a useful outlook for the viral pathogenicity.

Disordered regions in viral proteins are generally associated with the viral infectivity and pathogenicity, thus it is of certain interest to evaluate the intrinsic disorder within the SARS-CoV-2 proteome. As a matter of fact, the SARS-CoV-2 proteome has been found to exhibit significant levels of structural order: except for the nucleocapsid (N) protein from the structural proteins (plus Nsp8 and ORF6 from the nonstructural proteins⁶), the majority of SARS-CoV-2 proteins are highly ordered proteins containing a very few intrinsically disordered protein regions.⁶ Noteworthy, however, even though IDPs/IDRs are not common in the SARS-CoV-2 proteome, the existing ones contribute significantly to the functioning and virulence of the virus and are thus promising drug targets for antiviral drug discovery.^{6,66,67}

3.1. Nucleocapsid (N) Protein Exhibits Large IDRs.

The RNA-binding nucleocapsid (N) protein stabilizes the genomic RNA inside the virus particle and regulates the viral genome transcription, replication, and packaging. It is composed of two structural domains: N-terminal RNA-binding domain (NTD, residues 45–181) and C-terminal dimerization domain (CTD, residues 248–365), bordered by three large intrinsically disordered regions (Figure 4). The primary role of the RNA-binding N protein in the coronavirus life cycle is to assemble with the genomic RNA into the viral RNA–protein

complex and to mediate its packaging into 80–90 nm virions.⁷¹

The N protein is highly disordered—its average percentage of predicted intrinsic disorder according to the various disorder prediction tools²⁷ has been estimated as 65%.⁶ The IDRs in SARS-CoV-2 nucleocapsid protein comprise three segments, including residues 1–44 (IDR1), 182–247 (IDR2), and 366–422 (IDR3) (Figure 4A).⁶⁶ Analysis of the N protein disorder propensity using the protein disorder prediction system (PrDOS)^{28,29} indicates a significant degree of disorder in these three domains (Figure 4B). The highly flexible intrinsic disordered linker region, which connects the NTD and CTD, is rich in serine and arginine residues, both of which are highly disorder-promoting.²⁴ The middle IDR is known to be responsible for its RNA-binding activity.⁶⁸ An IDR that borders the CTD (C-terminal tail peptide) plays a significant role in dimer–dimer association in human coronaviruses, whereas nucleotide binding is largely mediated by the central NTD core.⁶⁹ In fact, molecular recognition domains have been identified in all three IDRs of the N protein. The RNA–protein recognition requires conformational changes in both RNA and protein, which is enabled by the structural flexibility of disordered regions. These findings support the key function of the N protein in a ribonucleoprotein core formation via interaction with the genomic RNA, which is a crucial step for RNA encapsulation and virus life cycle. Thus, the IDRs appear as appropriate targets to inhibit the interaction of N protein with viral genomic RNA.⁷⁰

Many IDPs are known to undergo liquid–liquid phase separation into dense intracellular organelles with a multitude of important cellular functions.⁷¹ Liquid–liquid phase separation is a process by which biomolecules, such as proteins and/or nucleic acids, condense into a dense phase that often resembles liquid droplets.^{72,73} It typically occurs within the cell, forming compartments termed membraneless organelles.⁷³ Eukaryotes contain numerous such membraneless compartments, and they are highly dynamic structures that can rapidly form on demand, thus concentrating proteins and biochemical reactions at distinct locations as needed. Virtually all such membraneless compartments contain a large proportion of IDPs. Their lack of defined secondary or tertiary structure and high conformational flexibility match the dynamic behavior of the membraneless compartments. One type of such membraneless aggregates—the stress granules—are composed of proteins and RNAs and form when the cell is under stress and the translation initiation is limited.^{74–77} Here again, it would be noteworthy to examine whether and how stress enhances or diminishes the degree of intrinsic disorder of the IDPs involved in the stress granules.

It has been shown recently that the N protein of the SARS-CoV-2 virus undergoes liquid–liquid phase separation into stress granules through its N-terminal intrinsically disordered region IDR1.^{58,78,79} The condensation of the N protein into stress granules is possibly a way for SARS-CoV-2 to inhibit host cell innate immunity. A model has been developed in which phase separation of the SARS-CoV-2 N protein contributes both to suppression of the host immune response and to packaging genomic RNA during virion assembly.⁷⁸ Disruption of the N protein liquid–liquid phase separation process holds promise for antiviral intervention and offers new targets and strategies for the development of drugs to combat COVID-19.

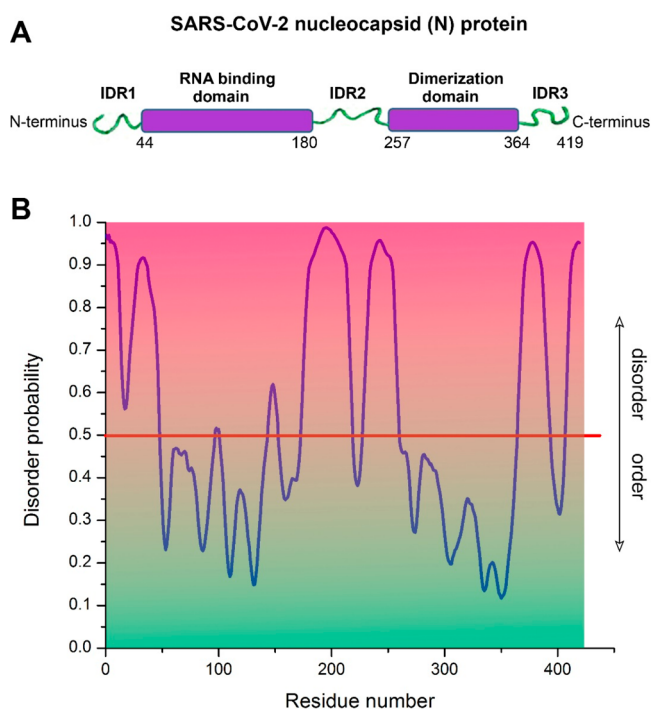


Figure 4. (A) Structure of the nucleocapsid (N) protein comprising three intrinsically disordered regions, IDR1, IDR2, and IDR3. (B) N protein disorder prediction using the protein disorder prediction system (PrDOS)^{28,29} (similar disorder predictions were obtained with other disorder prediction tools).

Recent NMR and crystal studies have confirmed the disordered nature of the N protein domains, including the NTD^{80,81} and CTD^{82,83} domains, whereas fluorescence recovery after photobleaching (FRAP) experiments confirmed the N protein liquid–liquid phase separation,⁷⁹ which concentrates components of the SARS-CoV-2 replication machinery, thus providing a means for enhanced viral transcription and replication.

A key to hold back the COVID-19 pandemic is to understand how SARS-CoV-2 manages to overcome host antiviral defense mechanisms. Stress granules, which are assembled throughout viral infection and function to sequester host and viral mRNAs and proteins, are part of the antiviral responses. It has been shown that the SARS-CoV-2 nucleocapsid (N) protein, an RNA binding protein essential for viral production, interacts with Ras-GTPase-activating protein SH3-domain-binding protein (G3BP) and disrupts stress granule assembly, both of which require the intrinsically disordered region 1 (IDR1) in the N protein. The N protein segregates into stress granules through liquid–liquid phase separation with G3BP and obstructs the interaction of G3BP with other stress-granule-related proteins. Furthermore, the N protein IDR domains important for phase separation with G3BP and stress granule disassembly are found to be essential for SARS-CoV-2 viral production. It has been suggested that N-protein-mediated stress granule disassembly is crucial for SARS-CoV-2 production.⁸⁴ It is thus implied that inhibition of the RNA-induced phase separation of the N protein provides a viable new strategy for the design of COVID-19 therapeutics.⁷⁹

3.2. Spike (S) Protein Exhibits Small but Critical IDRs.

The spike (S) protein is a large multidomain 1273 amino acid fusion protein creating the exterior of the CoV particles.⁸⁵ It protrudes from the virion and ornaments the viral surface like a crown. It is anchored in the viral membrane and mediates fusion of the viral membrane with the host cell membrane.⁸⁶ The spike protein is critical for the entry of the coronaviruses into the host, so it is an attractive antiviral target. It contains two distinct domains, S1 and S2, each of which consists of a number of functional subunits (Figure 5A).

The spike protein is cleaved by host proteases into the S1 and S2 subunits, which are responsible for receptor recognition and membrane fusion, respectively: subunit S1 activates viral infection by binding to host cell receptors, and S2 mediates the fusion of the virion and cellular membranes, thus promoting viral entry into the host cells.^{62,86} The spike protein binds to a specific surface receptor angiotensin converting enzyme 2 (ACE2) on the host cell plasma membrane via its N-terminal receptor-binding domain (RBD).⁸⁷ The S1 subunit consists of NTD and RBD; the S2 subunit contains a fusion peptide (FP), a heptad repeat 1 (HR1), a central helix (CH), a connector domain (CD), a heptad repeat 2 (HR2), a transmembrane domain (TM), and a cytoplasmic tail (CT) (Figure 5A).⁸⁸ Recent experimental studies confirmed that human ACE2 (hACE2) mediates SARS-CoV-2 S-protein-mediated entry into cells, establishing it as a functional entry receptor for this newly emerged coronavirus.⁵⁴

The site at the border between the S1 and S2 subunits—the S1/S2 protease cleavage site—is where spike protein is cleaved into S1 and S2 subunits during entry into the infected cells. The attachment of the spike protein with the host cell is activated by the host cell enzymes trypsin, cathepsin L, furin, and TMPRSS2. Comparison of the sequence of SARS-CoV-2 against other coronaviruses indicates that a unique amino acid

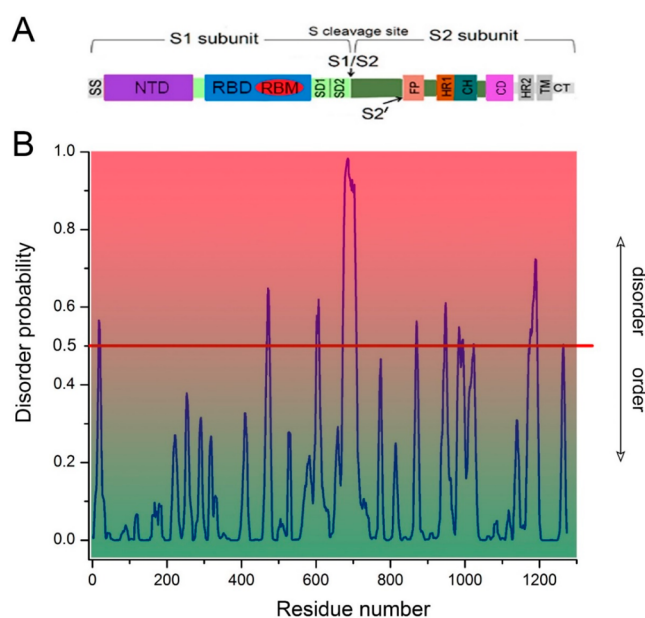


Figure 5. (A) Structure of the SARS-CoV-2 spike protein comprising S1 and S2 subunits;¹⁰² domain arrangement of spike protein: SS, signal sequence; NTD, N-terminal transactivation domain; RBD, receptor-binding domain; SD, subdomain; FP, fusion peptides; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; S1/S2 and S2', protease cleavage sites; TM, transmembrane domain; CT, cytoplasmic tail.⁷⁰ (B) Spike protein disorder prediction using the PONDR-VLXT^{97,98} (similar disorder predictions were obtained with other disorder prediction tools).

sequence pattern RRAR (Arg-Arg-Ala-Arg) is present at the S1/S2 junction of the spike protein, which is cleaved by the furin enzyme but is absent in other coronaviruses of the same clade, including SARS-CoV-1.^{70,89} Moreover, the structure reported for SARS-CoV-2 spike protein (PDB code: 6VSB) shows that the S1/S2 junction is in a disordered, solvent-exposed loop,⁹⁰ which is hypothesized to be responsible for the effective viral transmission.^{91,92}

Another cleavage event at an additional site inside the S2 subunit, the S2' site (Figure 5A), brings the viral and cellular membranes together, eventually creating a fusion pore that lets the viral genome reach the cell cytoplasm. ACE2 binding by the virus exposes the S2' site. The S2' site cleavage by transmembrane protease serine 2 (TMPRSS2) results in the release of the fusion peptide.⁸⁶

In the SARS-CoV-2 S protein, the first S1/S2 cleavage site is at residue R685^{89,93} (or R684 from a different study⁶⁴), whereas the second cleavage site generating the S2' subunit is located at residue R816⁹⁴ (or R815^{64,95}), bordering the FP located at residues 816–855.⁹⁶ An intrinsic disorder profile generated for the spike protein of the SARS-CoV-2 virus by the disorder predictor PONDR-VLXT^{97,98} (Figure 5B) indicates that both the S1/S2 cleavage site and the FP are located within IDRs. Considering the notion that the proteolytic digestion is considerably faster in unstructured relative to structured protein regions,^{45,99} this structural specificity of the SARS-CoV-2 spike protein might be of high functional importance (Figure 5B).

The S1 subunit of the S protein contains a RBD and an amino-terminal (N-terminal) domain. The RBD (residues 333–526¹⁰⁰) has a β -sheet core, bordered on either side by a short helix, and contains a receptor-binding motif (RBM,

residues 438–508¹⁰¹), which makes extensive contact with the ACE2 receptor, thus accounting for the interaction with ACE2. The tyrosine-rich RBM, which is stabilized by disulfide bonds, is also characterized by recognizable structural flexibility¹⁰¹—it does not possess a regular secondary structure except for the two small β -sheets.⁶² During SARS-CoV-2 virus infection, intrinsically disordered regions are detected at the interface of the spike protein and ACE2 receptor, providing a shape match to the complex. The key residues of the spike protein have strong binding affinity to ACE2, which can be a likely reason for the higher transmission rate of SARS-CoV-2.⁷⁰

Thus, the receptor binding and its membrane fusion are the initial and important steps in coronavirus infection and serve as primary targets for inhibiting the viral entry. Both exhibit regions of substantial intrinsic disorder, supposedly responsible for the viral cycle and pathogenicity.⁷⁰ Altogether, even though the spike protein of SARS-CoV-2 exhibits significant levels of structural order, its predicted percentage of intrinsic disorder is still higher than the spike protein of SARS-CoV (cf. 1.41 for SARS-CoV-2 vs 1.12 for SARS-CoV⁶), which may correlate with the higher infectivity and pathogenicity of SARS-CoV-2. Moreover, it was recently shown that the SARS-CoV-2 S glycoprotein exhibits a furin cleavage site at the borderline between the S1/S2 subunits, which sets this virus apart from other SARS-CoVs, supposedly enhancing its transmissibility due to the ubiquitous distribution of furin-like proteases in host cells.⁵⁴

A recent study made an advance in approaching anti-COVID-19 drug discovery by specifically focusing on targeting disordered protein regions in the virus proteome.⁶⁷ It was demonstrated how these IDRs can be targeted through molecular docking. As a result, 11 new drug candidates were identified that exhibited high binding and activity scores, as well as good antiviral properties.⁶⁷

3.3. M Protein Is Highly Ordered, Which May Contribute to the Greater Virulence of the SARS-CoV-2. The membrane (M) protein, one of the four structural proteins in the coronavirus virion, is a major transmembrane protein that is found in large numbers in the virion. It is a part of the protective proteinaceous layer responsible for the virus survival upon its transmission between the hosts. Estimates of the intrinsic disorder indicate that SARS-CoV-2 has one of the hardest protective outer shell among coronaviruses (Figure 6)—the percentage of intrinsic disorder of the M protein of only 6% (cf. 8% for SARS-CoV, 9% for MERS-CoV, 11% for HCoV-NL63, etc.).¹⁰³ Structural studies have confirmed the highly ordered organization of the M protein.¹⁰⁴ Thus, it might be anticipated for the SARS-CoV-2 virus to be highly resistant to antimicrobial substances in saliva and/or other bodily fluids, as well as in the environment, outside of the body. It is therefore expected to remain active for a longer time, which may account for greater contagiousness of SARS-CoV-2. Indeed, correlation has been confirmed between the virulence of various viruses and the percentage of intrinsic disorder of their M protein, with the less disordered viruses being more contagious.¹⁰⁵

A model that estimates the percentage of intrinsic disorder of viral proteins demonstrates that the degree of shell disorder in coronaviruses correlates with the levels of their fecal–oral and respiratory transmission.¹⁰⁶ The N protein is also important for the model, as it exhibits the greater disorder in the inner shell related to the mode of infection and virulence in other viruses.¹⁰³ In fact, it seems reasonable that the

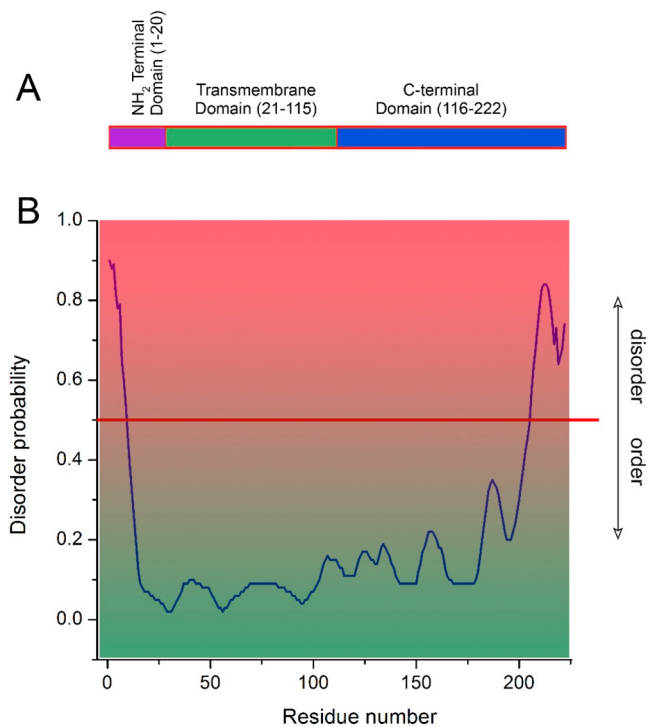


Figure 6. M protein (A) structure and (B) disorder prediction using the protein disorder prediction system (PrDOS)^{28,29}

performance of the virus depends on a delicate balance between the rigidity of its protective outer shell on one side, with the M protein as a major constituent, responsible for the virus survival, and the flexibility and adaptability of the inner core on the other side, especially the N nucleoprotein, upon RNA encapsulation throughout the virus life cycle, promoting binding to host proteins. Thus, the M protein orderliness and the N protein disorder seem to contribute to the SARS-CoV-2 high virulence and infectivity. It is also worth noting that, despite that the spike protein of SARS-CoV-2 exhibits significant levels of structural order, its predicted percentage of intrinsic disorder is still higher than that of the spike protein of SARS-CoV,⁶ which may also contribute to the higher infectivity and pathogenicity of SARS-CoV-2.

In the group of the nonstructural SARS-CoV-2 proteins, several moderately disordered proteins—Nsp8, ORF6, and ORF9b—have been identified, as well.⁶ The C-terminal region of the nonstructural protein Nsp2 has been predicted to be disordered.¹⁰⁷ Because it interacts with host proteins, which regulate translation initiation and endosome vesicle sorting, compounds that block these interactions could be valuable candidates for drug development. Although the other proteins exhibit lower disorder content, nearly all contain at least one IDR. Moreover, intrinsic disorder has been reported at the cleavage sites of replicase protein 1ab of SARS-CoV-2.⁶

Generally, there is a large variation in the distribution of disordered residue fractions in the viral proteomes, with certain viral species highly enriched in intrinsic disorder.⁴⁰ Examples of viral proteins with a high degree of intrinsic disorder (>80%) according to the DisProt database¹⁰⁸ are shown in Table 1. Indeed, the high content of intrinsic disorder predicted in viruses agrees with a recent study which showed that viral proteins were significantly enriched in polar residues and depleted in hydrophobic residues compared with that of archaea and bacteria,¹⁰⁹ which correlates with their disorder-

Table 1. Examples of Viral Proteins with >80% Intrinsic Disorder According to the DisProt Database¹⁰⁸

protein name	organism
antitermination protein N	<i>Escherichia</i> phage lambda
protein E7	human papillomavirus type 16
antitoxin phd	<i>Escherichia</i> phage P1
phosphoprotein	human respiratory syncytial virus A
nuclear export protein	influenza A virus
latent membrane protein 2A	Epstein–Barr virus (strain GD1)
protein Tat	human immunodeficiency virus 1
early E1A protein	human adenovirus A serotype 2
transcriptional repressor arc	<i>Salmonella</i> phage P22

edness.^{110,111} The high intrinsic disorder in viral proteins is supposedly linked to important functional implications, helping the viruses to highjack various pathways of the host cells and/or to accommodate to their hostile habitats.^{110,111}

The roles of IDPs/IDRs in certain viral proteins upon viral infections have been explored but are not yet completely understood, including membrane-binding protein λ N of bacteriophage, horde virus protein TGBp1, influenza virus nonstructural protein 2, basic protein δ Ag of hepatitis B virus, and human adenovirus type 5.¹

4. OUTLINE OF THE PROSPECTIVE DRUG TARGET SITES

As discussed above, even though IDPs/IDRs are not common in the SARS-CoV-2 proteome, the existing ones contribute significantly to the functioning and virulence of the virus and are thus promising drug targets for antiviral drug discovery.^{6,66,67} Therefore, it might be productive to focus on targeting those intrinsically disordered protein regions that lack a stable structure instead of following the common drug discovery approach along the conventional protein structure–function paradigm, thus designing drugs to bind to fixed 3D structures. Moreover, such an approach already has proven valuable in designing new SARS-CoV-2 drug candidates targeting IDRs.⁶⁷

- The nucleocapsid (N) protein plays a key function for the ribonucleoprotein core formation via interaction with the genomic RNA, a critical step in the virus life cycle. Thus, the three IDRs in the N protein appear as appropriate targets to inhibit its interaction with the viral genomic RNA.⁷⁰
- The N protein segregates into stress granules through liquid–liquid phase separation with Ras-GTPase-activating protein SH3-domain-binding protein (G3BP) and obstructs the interaction of G3BP with other stress-granule-related proteins. The process requires active involvement of the intrinsically disordered region 1 (IDR1) of the N protein. Disruption of N protein liquid–liquid phase separation process holds promise for antiviral intervention and offers new targets and strategies for the development of drugs to combat COVID-19.
- Because of the vital function of the spike (S) protein for the entry of coronaviruses into the host, it is a reasonable target for inhibition by neutralizing antibodies, and characterization of the spike protein structure provides strategic information for rational vaccine design. The receptor binding and its membrane fusion are the initial and important steps in the coronavirus infection, both of

which involve specific intrinsically disordered regions of the protein as key players and serve as primary targets for inhibiting the viral entry. Neutralizing antibodies targeting the SARS-CoV-2 S trimer have exhibited protection from viral infection in animal models and are currently being assessed as therapeutics in humans.⁶² Such antibodies include human monoclonal antibodies isolated from COVID-19 recovering donors and single-domain nanobodies. Although some neutralizing antibodies target the NTD or the S2 domain, most of them bind to the RBD, producing steric hindrance and thus blocking ACE2 attachment.⁸⁴ Recent biophysical modeling suggested that the SARS-CoV-2 spike protein may provide an adaptable allosteric mechanism utilized to fine-tune the response to antibody binding, which may be useful for therapeutic intervention by targeting specific hotspots of allosteric interactions.¹¹²

- Research consortia have been established with the goal to bring together scientists with different expertise to advance our understanding of COVID-19 and speed-up drug discovery.^{113,114} A COVID-19 NMR research consortium was set up in 2020, aiming to support the search for antiviral drugs using an NMR-based screening and establishing protocols for large-scale production of all druggable SARS-CoV-2 proteins and RNAs for rational drug design and fast mapping of compound binding sites.⁶³

5. CONCLUSIONS AND OUTLOOK

The notion of the abundance of IDPs and IDRs in proteomes is altering protein science. The concept of protein intrinsic disorder provides answers for many scientific problems that cannot be easily explained based on the classic structure–function paradigm. Functions of IDPs/IDRs can be controlled by multiple processes, such as post-translational modifications, interaction with various other molecules, etc. The multilevel structural and functional complexity of IDPs/IDRs makes them particularly sensitive to the environment. Furthermore, intrinsic disorder may be related to the emergent behavior of several systems characterized by the presence of specific patterns and can be used to explain the biochemical reaction–diffusion processes. Understanding how structured and disordered proteins work in concert is crucial for understanding protein functions.

Appearance of novel viruses and associated epidemics around the globe are currently a major concern. Knowledge on the structures and functions of the viral proteins is of utmost significance for identification of novel therapeutic targets for prevention and treatment of the viral diseases.⁶³ In this Perspective, we summarized information available on SARS-CoV-2 proteome with regards to the occurrence of intrinsic disorder in SARS-CoV-2 proteins. It has been recognized that the SARS-CoV-2 proteome exhibits substantial levels of structural order. From the SARS-CoV-2 proteome, only the nucleocapsid (N) protein is highly disordered, with an average percentage of predicted intrinsic disorder of 65%. The spike (S) protein of SARS-CoV-2 exhibits significant levels of structural order, yet its predicted percentage of intrinsic disorder is still higher than that of the spike protein of SARS-CoV. Furthermore, although the other structural proteins of SARS-CoV-2 apart from the N protein are characterized by a low degree of disorder, their existing IDRs contribute

significantly to the functioning and virulence of the virus and are thus promising drug targets for antiviral drug design. As these IDRs typically undergo structural transitions upon attaching to their physiological associates, such knowledge generates an important base for better understanding the activity of these proteins and their interactions with host proteins in different physiological conditions. The association of viral and host proteins needs to be explored to find means suitable for limiting replication, maturation, and ultimately pathogenesis of the viruses. Although structural biology techniques can be utilized in drug development, the practice of rational drug design has essentially ignored so far the presence of intrinsic disorder in target proteins. Understanding of the structure of these regions in the COVID-19 proteome would be a valuable asset for high-throughput screening in drug development for the disease.

AUTHOR INFORMATION

Corresponding Author

Qiongqiong Angela Zhou – CAS, a division of the American Chemical Society, Columbus, Ohio 43210, United States;

orcid.org/0000-0001-6711-369X; Email: qzhou@cas.org

Author

Rumiana Tenchov – CAS, a division of the American Chemical Society, Columbus, Ohio 43210, United States;

orcid.org/0000-0003-4698-6832

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsinfectdis.2c00031>

Notes

The authors declare no competing financial interest.

REFERENCES

- (1) Mishra, P. M.; Verma, N. C.; Rao, C.; Uversky, V. N.; Nandi, C. K. Intrinsically disordered proteins of viruses: Involvement in the mechanism of cell regulation and pathogenesis. *Progress in molecular biology and translational science* **2020**, *174*, 1–78.
- (2) Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999**, *293*, 321–331.
- (3) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graph Model* **2001**, *19*, 26–59.
- (4) Uversky, V. N. Protein intrinsic disorder-based liquid-liquid phase transitions in biological systems: Complex coacervates and membrane-less organelles. *Adv. Colloid Interface Sci.* **2017**, *239*, 97–114.
- (5) Brocca, S.; Grandori, R.; Longhi, S.; Uversky, V. Liquid–Liquid Phase Separation by Intrinsically Disordered Protein Regions of Viruses: Roles in Viral Life Cycle and Control of Virus–Host Interactions. *International Journal of Molecular Sciences* **2020**, *21*, 9045.
- (6) Giri, R.; Bhardwaj, T.; Shegane, M.; Gehi, B. R.; Kumar, P.; Gadhave, K.; Oldfield, C. J.; Uversky, V. N. Understanding COVID-19 via comparative analysis of dark proteomes of SARS-CoV-2, human SARS and bat SARS-like coronaviruses. *Cellular and molecular life sciences: CMLS* **2021**, *78*, 1655–1688.
- (7) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **2015**, *16*, 18–29.
- (8) Uversky, V. N. Intrinsically Disordered Proteins. In *Brenner's Encyclopedia of Genetics*, 2nd ed.; Maloy, S., Hughes, K., Eds.; Academic Press: San Diego, CA, 2013; pp 124–126.
- (9) Dobrev, V. S.; Fred, L. M.; Gerhart, K. P.; Metallo, S. J. Characterization of the Binding of Small Molecules to Intrinsically Disordered Proteins. In *Methods in Enzymology*; Rhoades, E., Ed.; Academic Press, 2018; Chapter 21, Vol. 611, pp 677–702.
- (10) *Intrinsically Disordered Proteins*, 1st ed.; Academic Press, 2018; Vol. 611, p 754.
- (11) Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically disordered proteins in human diseases: Introducing the D(2) concept. *Annual Review of Biophysics* **2008**, *37*, 215–246.
- (12) Heller, G. T.; Aprile, F. A.; Vendruscolo, M. Methods of probing the interactions between small molecules and disordered proteins. *Cellular and molecular life sciences: CMLS* **2017**, *74*, 3225–3243.
- (13) Hsu, C. C.; Buehler, M. J.; Tarakanova, A. The Order-Disorder Continuum: Linking Predictions of Protein Structure and Disorder through Molecular Simulation. *Sci. Rep.* **2020**, *10*, 2068.
- (14) Liu, H.; Guo, X.; Han, J.; Luo, R.; Chen, H.-F. Order-disorder transition of intrinsically disordered kinase inducible transactivation domain of CREB. *J. Chem. Phys.* **2018**, *148*, 225101.
- (15) Moritsugu, K.; Terada, T.; Kidera, A. Disorder-to-Order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling. *J. Am. Chem. Soc.* **2012**, *134*, 7094–7101.
- (16) Gallat, F. X.; Laganowsky, A.; Wood, K.; Gabel, F.; van Eijck, L.; Wuttke, J.; Moulin, M.; Härtlein, M.; Eisenberg, D.; Colletier, J. P.; Zaccai, G.; Weik, M. Dynamical coupling of intrinsically disordered proteins and their hydration water: comparison with folded soluble and membrane proteins. *Biophysical journal* **2012**, *103*, 129–136.
- (17) Wohl, S.; Jakubowski, M.; Zheng, W. Salt-Dependent Conformational Changes of Intrinsically Disordered Proteins. *J. Phys. Chem. Lett.* **2021**, *12*, 6684–6691.
- (18) Wicky, B. I. M.; Shammass, S. L.; Clarke, J. Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 9882–9887.
- (19) Basile, W.; Salvatore, M.; Bassot, C.; Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS computational biology* **2019**, *15*, e1007186–e1007186.
- (20) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics* **2019**, DOI: 10.3389/fphy.2019.00010.
- (21) Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science* **2008**, *322*, 1365–1368.
- (22) Uversky, V. N. Intrinsic Disorder, Protein-Protein Interactions, and Disease. *Adv. Protein Chem. Struct. Biol.* **2018**, *110*, 85–121.
- (23) Uversky, V. N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically disordered proteins* **2013**, *1*, e24684–e24684.
- (24) Campen, A.; Williams, R. M.; Brown, C. J.; Meng, J.; Uversky, V. N.; Dunker, A. K. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett.* **2008**, *15*, 956–63.
- (25) Nielsen, J. T.; Mulder, F. A. A. Quality and bias of protein disorder predictors. *Sci. Rep.* **2019**, *9*, 5137.
- (26) Hu, G.; Katuwawala, A.; Wang, K.; Wu, Z.; Ghadermarzi, S.; Gao, J.; Kurgan, L. fDPnn: Accurate intrinsic disorder prediction with putative propensities of disorder functions. *Nat. Commun.* **2021**, *12*, 4438.
- (27) List of disorder prediction software, https://en.wikipedia.org/wiki/List_of_disorder_prediction_software (accessed 2021-10-11).
- (28) Protein DisOrder prediction system, <https://prd05.hgc.jp/cgi-bin/top.cgi> (accessed 2021-10-07).
- (29) Ishida, T.; Kinoshita, K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic acids research* **2007**, *35*, W460–W464.

- (30) Prediction of intrinsically unstructured proteins, <https://iupred2a.elte.hu/> (accessed 2021-10-24).
- (31) Necci, M.; Piovesan, D.; Hoque, M. T.; Walsh, I.; Iqbal, S.; Ventrusco, M.; Sormanni, P.; Wang, C.; Raimondi, D.; Sharma, R.; Zhou, Y.; Litfin, T.; Galzitskaya, O. V.; Lobanov, M. Y.; Vranken, W.; Wallner, B.; Mirabello, C.; Malhis, N.; Dosztányi, Z.; Erdős, G.; Mészáros, B.; Gao, J.; Wang, K.; Hu, G.; Wu, Z.; Sharma, A.; Hanson, J.; Paliwal, K.; Callebaut, I.; Bitard-Feildel, T.; Orlando, G.; Peng, Z.; Xu, J.; Wang, S.; Jones, D. T.; Cozzetto, D.; Meng, F.; Yan, J.; Gsponer, J.; Cheng, J.; Wu, T.; Kurgan, L.; Promponas, V. J.; Tamana, S.; Marino-Buslje, C.; Martínez-Pérez, E.; Chasapi, A.; Ouzounis, C.; Dunker, A. K.; Kajava, A. V.; Leclercq, J. Y.; Aykac-Fas, B.; Lambroughi, M.; Maiani, E.; Papaleo, E.; Chemes, L. B.; Álvarez, L.; González-Foutel, N. S.; Iglesias, V.; Pujols, J.; Ventura, S.; Palopoli, N.; Benítez, G. I.; Parisi, G.; Bassot, C.; Elofsson, A.; Govindarajan, S.; Lamb, J.; Salvatore, M.; Hatos, A.; Monzon, A. M.; Bevilacqua, M.; Mičetić, I.; Minervini, G.; Paladin, L.; Quaglia, F.; Leonardi, E.; Davey, N.; Horvath, T.; Kovacs, O. P.; Murvai, N.; Pancsa, R.; Schad, E.; Szabo, B.; Tantos, A.; Macedo-Ribeiro, S.; Manso, J. A.; Pereira, P. J. B.; Davidović, R.; Veljkovic, N.; Hajdu-Soltész, B.; Pajkos, M.; Szaniszló, T.; Guharoy, M.; Lazar, T.; Macossay-Castillo, M.; Tompa, P.; Tosatto, S. C. E.; Predictors, C.; DisProt, C. Critical assessment of protein intrinsic disorder prediction. *Nat. Methods* **2021**, *18*, 472–481.
- (32) Katsnelson, A. Did Disordered Proteins Help Launch Life on Earth? *ACS Central Science* **2020**, *6*, 1854–1857.
- (33) Uversky, V. N. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* **2013**, *22*, 693–724.
- (34) Uversky, V. N. Dancing Protein Clouds: The Strange Biology and Chaotic Physics of Intrinsically Disordered Proteins. *J. Biol. Chem.* **2016**, *291*, 6681–8.
- (35) Kulkarni, P.; Uversky, V. N. Intrinsically Disordered Proteins: The Dark Horse of the Dark Proteome. *Proteomics* **2018**, *18*, 1800061.
- (36) Dunker, A. K.; Bondos, S. E.; Huang, F.; Oldfield, C. J. Intrinsically disordered proteins and multicellular organisms. *Semin Cell Dev Biol.* **2015**, *37*, 44–55.
- (37) Niklas, K. J.; Dunker, A. K.; Yruela, I. The evolutionary origins of cell type diversification and the role of intrinsically disordered proteins. *J. Exp. Bot.* **2018**, *69*, 1437–1446.
- (38) Kulkarni, P.; Behal, A.; Mohanty, A.; Salgia, R.; Nedelcu, A. M.; Uversky, V. N. Co-opting disorder into order: Intrinsically disordered proteins and the early evolution of complex multicellularity. *Int. J. Biol. Macromol.* **2022**, *201*, 29–36.
- (39) Schad, E.; Tompa, P.; Hegyi, H. The relationship between proteome size, structural disorder and organism complexity. *Genome Biol.* **2011**, *12*, R120.
- (40) Xue, B.; Dunker, A. K.; Uversky, V. N. 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.
- (41) CAS Content Collection; <https://www.cas.org/about/cas-content> (accessed 2021-06-09).
- (42) Kulkarni, P.; Uversky, V. N. Intrinsically Disordered Proteins in Chronic Diseases. *Biomolecules* **2019**, *9*, 147.
- (43) Du, Z.; Uversky, V. N. A Comprehensive Survey of the Roles of Highly Disordered Proteins in Type 2 Diabetes. *International Journal of Molecular Sciences* **2017**, *18*, 2010.
- (44) Uversky, V. N. Intrinsic disorder in proteins associated with neurodegenerative diseases. *Frontiers in Bioscience-Landmark* **2009**, *14*, 5188–5238.
- (45) Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradovic, Z.; Dunker, A. K. Intrinsic disorder in cell-signaling and cancer-associated proteins. *J. Mol. Biol.* **2002**, *323*, 573–584.
- (46) Cheng, Y. G.; LeGall, T.; Oldfield, C. J.; Dunker, A. K.; Uversky, V. N. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry* **2006**, *45*, 10448–10460.
- (47) Uversky, V. N. Amyloidogenesis of natively unfolded proteins. *Current Alzheimer Research* **2008**, *5*, 260–287.
- (48) Midic, U.; Oldfield, C. J.; Dunker, A. K.; Obradovic, Z.; Uversky, V. N. Unfoldomics of Human Genetic Diseases: Illustrative Examples of Ordered and Intrinsically Disordered Members of the Human Diseaseome. *Protein and Peptide Letters* **2009**, *16*, 1533–1547.
- (49) Antonova, B.; Naydenov, E.; Koynova, R.; Tumangelova-Yuzeir, K.; Tenchov, B. Exothermic transitions in the heat capacity profiles of human cerebrospinal fluid. *Eur. Biophys. J.* **2020**, *49*, 231–238.
- (50) Tenchov, B.; Abarova, S.; Koynova, R.; Traikov, L.; Tancheva, L. Low-temperature exothermic transitions in brain proteome of mice, effect of scopolamine. *Thermochim. Acta* **2017**, *650*, 26–32.
- (51) Stefanis, L. α -Synuclein in Parkinson's disease. *Cold Spring Harbor perspectives in medicine* **2012**, *2*, a009399–a009399.
- (52) Wang, J.; Cao, Z.; Zhao, L.; Li, S. Novel strategies for drug discovery based on Intrinsically Disordered Proteins (IDPs). *International journal of molecular sciences* **2011**, *12*, 3205–3219.
- (53) Kim, D.; Lee, J. Y.; Yang, J. S.; Kim, J. W.; Kim, V. N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921.e10.
- (54) 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*, 281–292.e6.
- (55) Alharbi, S. N.; Alrefaei, A. F. Comparison of the SARS-CoV-2 (2019-nCoV) M protein with its counterparts of SARS-CoV and MERS-CoV species. *Journal of King Saud University - Science* **2021**, *33*, 101355.
- (56) Thomas, S. The Structure of the Membrane Protein of SARS-CoV-2 Resembles the Sugar Transporter SemiSWEET. *Pathog Immun* **2020**, *5*, 342–363.
- (57) Aldaais, E. A.; Yegnaswamy, S.; Albahrani, F.; Alsowaikeet, F.; Alramadan, S. Sequence and structural analysis of COVID-19 E and M proteins with MERS virus E and M proteins-A comparative study. *Biochemistry and biophysics reports* **2021**, *26*, 101023–101023.
- (58) Cubuk, J.; Alston, J. J.; Incicco, J. J.; Singh, S.; Stuchell-Breterton, M. D.; Ward, M. D.; Zimmerman, M. I.; Vithani, N.; Griffith, D.; Wagoner, J. A.; Bowman, G. R.; Hall, K. B.; Soranno, A.; Holehouse, A. S. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **2021**, *12*, 1936.
- (59) Schiavina, M.; Pontoriero, L.; Uversky, V. N.; Felli, I. C.; Pierattelli, R. The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1–248 residue construct: sequence-specific resonance assignments through NMR. *Biomolecular NMR Assignments* **2021**, *15*, 219–227.
- (60) Bai, Z.; Cao, Y.; Liu, W.; Li, J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses* **2021**, *13*, 1115.
- (61) Masters, P. S. The Molecular Biology of Coronaviruses. *Advances in Virus Research*; Academic Press, 2006; Vol. 66, pp 193–292.
- (62) Yang, H.; Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology* **2021**, *19*, 685–700.
- (63) Altincekic, N.; Korn, S. M.; Qureshi, N. S.; Dujardin, M.; Ninot-Pedrosa, M.; Abele, R.; Abi Saad, M. J.; Alfano, C.; Almeida, F. C. L.; Alshamleh, I.; de Amorim, G. C.; Anderson, T. K.; Anobom, C. D.; Anorma, C.; Bains, J. K.; Bax, A.; Blackledge, M.; Blechar, J.; Böckmann, A.; Brigandat, L.; Bula, A.; Bütikofer, M.; Camacho-Zarco, A. R.; Carlomagno, T.; Caruso, I. P.; Ceylan, B.; Chaikuad, A.; Chu, F.; Cole, L.; Crosby, M. G.; de Jesus, V.; Dhamotharan, K.; Felli, I. C.; Ferner, J.; Fleischmann, Y.; Fogeron, M.-L.; Fourkoti, N. K.; Fuks, C.; Fürtig, B.; Gallo, A.; Gande, S. L.; Gerež, J. A.; Ghosh, D.; Gomes-Neto, F.; Gorbatyuk, O.; Guseva, S.; Hacker, C.; Häfner, S.; Hao, B.; Hargittay, B.; Henzler-Wildman, K.; Hoch, J. C.; Hohmann, K. F.; Hutchison, M. T.; Jaudzems, K.; Jović, K.; Kaderli, J.; Kalniņš, G.; Kaņepe, I.; Kirchdoerfer, R. N.; Kirkpatrick, J.; Knapp, S.; Krishnathas, R.; Kutz, F.; zur Lage, S.; Lambert, R.; Lang, A.; Laurents, D.; Lecoq, L.; Linhard, V.; Löh, F.; Malki, A.; Bessa, L. M.

- Martin, R. W.; Matzel, T.; Maurin, D.; McNutt, S. W.; Mebus-Antunes, N. C.; Meier, B. H.; Meiser, N.; Mompeán, M.; Monaca, E.; Montserret, R.; Mariño Perez, L.; Moser, C.; Muhle-Goll, C.; Neves-Martins, T. C.; Ni, X.; Norton-Baker, B.; Pierattelli, R.; Pontoriero, L.; Pustovalova, Y.; Ohlenschläger, O.; Orts, J.; Da Poian, A. T.; Pypers, D. J.; Richter, C.; Riek, R.; Rienstra, C. M.; Robertson, A.; Pinheiro, A. S.; Sabbatella, R.; Salvi, N.; Saxena, K.; Schulte, L.; Schiavina, M.; Schwalbe, H.; Silber, M.; Almeida, M. d. S.; Sprague-Piercy, M. A.; Spyroulias, G. A.; Sreeramulu, S.; Tants, J.-N.; Tars, K.; Torres, F.; Tóws, S.; Treviño, M. Á.; Trucks, S.; Tsika, A. C.; Varga, K.; Wang, Y.; Weber, M. E.; Weigand, J. E.; Wiedemann, C.; Wirmmer-Bartoschek, J.; Wirtz Martin, M. A.; Zehnder, J.; Hengesbach, M.; Schlundt, A. Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Frontiers in Molecular Biosciences* **2021**, *8*, 653148.
- (64) Elrashdy, F.; Redwan, E. M.; Uversky, V. N. Intrinsic disorder perspective of an interplay between the renin-angiotensin-aldosterone system and SARS-CoV-2. *Infect Genet Evol* **2020**, *85*, 104510.
- (65) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197–208.
- (66) Chang, C. K.; Hou, M. H.; Chang, C. F.; Hsiao, C. D.; Huang, T. H. The SARS coronavirus nucleocapsid protein-forms and functions. *Antiviral Res.* **2014**, *103*, 39–50.
- (67) Mudide, A.; Alterovitz, G. SARS-CoV-2 Drug Discovery based on Intrinsically Disordered Regions. *Pac Symp. Biocomput* **2021**, *26*, 131–142.
- (68) Nelson, G. W.; Stohlman, S. A. Localization of the RNA-binding domain of mouse hepatitis virus nucleocapsid protein. *Journal of General Virology* **1993**, *74*, 1975–1979.
- (69) Redzic, J. S.; Lee, E.; Born, A.; Issaian, A.; Henen, M. A.; Nichols, P. J.; Blue, A.; Hansen, K. C.; D'Alessandro, A.; Vögeli, B.; Eisenmesser, E. Z. The Inherent Dynamics and Interaction Sites of the SARS-CoV-2 Nucleocapsid N-Terminal Region. *J. Mol. Biol.* **2021**, *433*, 167108.
- (70) Yesudhas, D.; Srivastava, A.; Gromiha, M. M. COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics. *Infection* **2021**, *49*, 199–213.
- (71) Darling, A. L.; Liu, Y.; Oldfield, C. J.; Uversky, V. N. Intrinsically Disordered Proteome of Human Membrane-Less Organelles. *Proteomics* **2018**, *18*, e1700193.
- (72) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324*, 1729–1732.
- (73) Shin, Y.; Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, eaaf4382.
- (74) Wolozin, B.; Ivanov, P. Stress granules and neurodegeneration. *Nat. Rev. Neurosci.* **2019**, *20*, 649–666.
- (75) Campos-Melo, D.; Hawley, Z. C. E.; Droppelmann, C. A.; Strong, M. J. The Integral Role of RNA in Stress Granule Formation and Function. *Frontiers in Cell and Developmental Biology* **2021**, DOI: 10.3389/fcell.2021.621779.
- (76) Protter, D. S. W.; Parker, R. Principles and Properties of Stress Granules. *Trends in Cell Biology* **2016**, *26*, 668–679.
- (77) Wheeler, J. R.; Matheny, T.; Jain, S.; Abrisch, R.; Parker, R. Distinct stages in stress granule assembly and disassembly. *eLife* **2016**, *5*, e18413.
- (78) Lu, S.; Ye, Q.; Singh, D.; Cao, Y.; Diedrich, J. K.; Yates, J. R.; Villa, E.; Cleveland, D. W.; Corbett, K. D. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **2021**, *12*, 502.
- (79) 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.
- (80) Dinesh, D. C.; Chalupska, D.; Silhan, J.; Koutna, E.; Nencka, R.; Veverka, V.; Boura, E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog* **2020**, *16*, e1009100.
- (81) Kang, S.; Yang, M.; Hong, Z.; Zhang, L.; Huang, Z.; Chen, X.; He, S.; Zhou, Z.; Zhou, Z.; Chen, Q.; Yan, Y.; Zhang, C.; Shan, H.; Chen, S. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin B* **2020**, *10*, 1228–1238.
- (82) Ye, Q.; West, A. M. V.; Silletti, S.; Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* **2020**, *29*, 1890–1901.
- (83) Zhou, R.; Zeng, R.; von Brunn, A.; Lei, J. Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Molecular Biomedicine* **2020**, *1*, 2.
- (84) Luo, L.; Li, Z.; Zhao, T.; Ju, X.; Ma, P.; Jin, B.; Zhou, Y.; He, S.; Huang, J.; Xu, X.; Zou, Y.; Li, P.; Liang, A.; Liu, J.; Chi, T.; Huang, X.; Ding, Q.; Jin, Z.; Huang, C.; Zhang, Y. SARS-CoV-2 nucleocapsid protein phase separates with G3BPs to disassemble stress granules and facilitate viral production. *Sci. Bull. (Beijing)* **2021**, *66*, 1194–1204.
- (85) Graham, R. L.; Baric, R. S. Recombination, Reservoirs, and the Modular Spike: Mechanisms of Coronavirus Cross-Species Transmission. *Journal of Virology* **2010**, *84*, 3134–3146.
- (86) Belouard, S.; Millet, J. K.; Licitra, B. N.; Whittaker, G. R. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* **2012**, *4*, 1011–33.
- (87) Li, F.; Li, W.; Farzan, M.; Harrison, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. *Science* **2005**, *309*, 1864–1868.
- (88) Wang, M. Y.; Zhao, R.; Gao, L. J.; Gao, X. F.; Wang, D. P.; Cao, J. M. SARS-CoV-2: Structure, Biology, and Structure-Based Therapeutics Development. *Front Cell Infect Microbiol* **2020**, *10*, 587269.
- (89) 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.
- (90) 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*, 1260–1263.
- (91) 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.
- (92) Hasan, A.; Paray, B. A.; Hussain, A.; Qadir, F. A.; Attar, F.; Aziz, F. M.; Sharifi, M.; Derakhshankhah, H.; Rasti, B.; Mehrabi, M.; Shahpasand, K.; Saboury, A. A.; Falahati, M. A review on the cleavage priming of the spike protein on coronavirus by angiotensin-converting enzyme-2 and furin. *J. Biomol Struct Dyn* **2021**, *39*, 3025–3033.
- (93) Pillay, T. S. Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *Journal of Clinical Pathology* **2020**, *73*, 366–369.
- (94) Verma, J.; Subbarao, N. A comparative study of human betacoronavirus spike proteins: structure, function and therapeutics. *Arch. Virol.* **2021**, *166*, 697–714.
- (95) Bestle, D.; Heindl, M. R.; Limburg, H.; Van Lam van, T.; Pilgram, O.; Moulton, H.; Stein, D. A.; Harges, K.; Eickmann, M.; Dolnik, O.; Rohde, C.; Klenk, H.-D.; Garten, W.; Steinmetzer, T.; Böttcher-Friebertshäuser, E. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance* **2020**, *3*, e202000786.
- (96) Koppiseti, R. K.; Fulcher, Y. G.; Van Doren, S. R. Fusion Peptide of SARS-CoV-2 Spike Rearranges into a Wedge Inserted in Bilayered Micelles. *J. Am. Chem. Soc.* **2021**, *143*, 13205–13211.
- (97) Romero, P.; Obradovic, Z.; Li, X.; Garner, E. C.; Brown, C. J.; Dunker, A. K. Sequence complexity of disordered protein. *Proteins* **2001**, *42*, 38–48.
- (98) PONDR - Predictor of Natural Disordered Regions; <http://www.pondr.com/> (accessed 2021-10-16).

- (99) Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zamai, M.; Zambonin, M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **1986**, *25*, 1847–1851.
- (100) 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*, 215–220.
- (101) 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. *Cellular & Molecular Immunology* **2020**, *17*, 621–630.
- (102) Kalathiya, U.; Padariya, M.; Mayordomo, M.; Lisowska, M.; Nicholson, J.; Singh, A.; Baginski, M.; Fahraeus, R.; Carragher, N.; Ball, K.; Haas, J.; Daniels, A.; Hupp, T. R.; Alfaro, J. A. Highly Conserved Homotrimer Cavity Formed by the SARS-CoV-2 Spike Glycoprotein: A Novel Binding Site. *J. Clin. Med.* **2020**, *9*, 1473.
- (103) Goh, G. K.-M.; Dunker, A. K.; Foster, J. A.; Uversky, V. N. Shell disorder analysis predicts greater resilience of the SARS-CoV-2 (COVID-19) outside the body and in body fluids. *Microbial pathogenesis* **2020**, *144*, 104177–104177.
- (104) Neuman, B. W.; Kiss, G.; Kunding, A. H.; Bhella, D.; Baksh, M. F.; Connelly, S.; Droese, B.; Klaus, J. P.; Makino, S.; Sawicki, S. G.; Siddell, S. G.; Stamou, D. G.; Wilson, I. A.; Kuhn, P.; Buchmeier, M. J. A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* **2011**, *174*, 11–22.
- (105) Goh, G. K.-M.; Dunker, A. K.; Foster, J. A.; Uversky, V. N. A Novel Strategy for the Development of Vaccines for SARS-CoV-2 (COVID-19) and Other Viruses Using AI and Viral Shell Disorder. *J. Proteome Res.* **2020**, *19*, 4355–4363.
- (106) Goh, G. K.-M.; Dunker, A. K.; Uversky, V. N. Understanding Viral Transmission Behavior via Protein Intrinsic Disorder Prediction: Coronaviruses. *Journal of Pathogens* **2012**, *2012*, 738590.
- (107) Mompeán, M.; Treviño, M.; Laurents, D. V. Partial structure, dampened mobility, and modest impact of a His tag in the SARS-CoV-2 Nsp2 C-terminal region. *Eur. Biophys. J.* **2021**, *50*, 1129–1137.
- (108) DisProt, the database of intrinsically disordered proteins; <https://www.disprot.org/> (accessed 2022-02-03).
- (109) Berezovsky, I. N. The diversity of physical forces and mechanisms in intermolecular interactions. *Phys. Biol.* **2011**, *8*, 035002.
- (110) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.; Dunker, A. K.; Uversky, V. N. Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept. Lett.* **2010**, *17*, 932–951.
- (111) Xue, B.; Williams, R. W.; Oldfield, C. J.; Goh, G. K.-M.; Dunker, A. K.; Uversky, V. N. Do Viral Proteins Possess Unique Features? *Flexible Viruses* **2011**, 1–34.
- (112) Verkhivker, G. M.; Di Paola, L. Integrated Biophysical Modeling of the SARS-CoV-2 Spike Protein Binding and Allosteric Interactions with Antibodies. *J. Phys. Chem. B* **2021**, *125*, 4596–4619.
- (113) Pietzner, M.; Wheeler, E.; Carrasco-Zanini, J.; Raffler, J.; Kerrison, N. D.; Oerton, E.; Auyeung, V. P. W.; Luan, J. a.; Finan, C.; Casas, J. P.; Ostroff, R.; Williams, S. A.; Kastenmüller, G.; Ralser, M.; Gamazon, E. R.; Wareham, N. J.; Hingorani, A. D.; Langenberg, C. Genetic architecture of host proteins involved in SARS-CoV-2 infection. *Nat. Commun.* **2020**, *11*, 6397.
- (114) Genetic architecture of host proteins involved in SARS-CoV-2 infection; <https://omicscience.org/apps/covidpgwas/> (accessed 2022-01-10).