

CHAPTER 1

Overview of Glycosylation Studies of SARS-CoV-2

LAUREN E. PEPI, ASIF SHAJAHAN, ANNE S. GLEINICH,
CHRISTIAN HEISS AND PARASTOO AZADI*

Complex Carbohydrate Research Center, University of Georgia, Athens, GA
30602, USA

*E-mail: azadi@ccrc.uga.edu

1.1 Introduction

Glycosylation occurs in approximately half of the expressed proteins in a cell. Glycosylation is the attachment of sugar residues to proteins as a post-translational modification (PTM).^{1,2} Enzymatic glycosylation is controlled by factors that differ greatly between different cell types and species. Elaborate glycosylation routes have been identified in a multitude of organisms; these routes lead to carbohydrate units, known as glycans, bound to proteins, known as glycoproteins, that are secreted by cells or components of the membranes, cytoplasm, and nucleus.² Asparagine (Asn)-linked glycans, referred to as *N*-linked glycans, and their structural and functional aspects, have been comprehensively studied.³ *N*-linked glycans are covalently attached to Asn by an *N*-glycosidic bond. All *N*-glycans contain the same common core sequence of Man α 1-3(Man α 1-6)Man β 1-4GlcNAc β 1-4GlcNAc β 1Asn-X-Ser/Thr, where X can be any amino acid apart from proline.⁴ The consensus sequence of Asn-X-Ser/Thr is necessary for *N*-glycosylation. There are three classes of *N*-glycans: high-mannose (oligomannose), complex, and hybrid. The high-mannose type

New Developments in Mass Spectrometry No. 15

Glycoprotein Analysis

Edited by Weston B. Struwe

© The Royal Society of Chemistry 2024

Published by the Royal Society of Chemistry, www.rsc.org

contains only *N*-acetylglucosamine (GlcNAc) residues and mannose (Man) residues. Typically, Man₅GlcNAc₂ through Man₉GlcNAc₂ are seen. Further processing, such as trimming of the mannose moieties and subsequent addition of carbohydrate residues (*e.g.* *N*-acetylglucosamine, galactose, fucose, and sialic acid) on the antenna(e), is referred to as maturation and produces hybrid- and complex-type *N*-glycans.⁴ Complex-type glycans are highly processed and can contain up to six branches elongated with Galβ1-4GlcNAc (LacNAc) repeats. Hybrid-type glycans are a mixture of complex and high-mannose types, with one branch typically resembling complex glycan structures and the other resembling high-mannose glycan structures. Most *N*-glycans are easily cleaved from the peptide backbone by *N*-glycosidase F protein (PNGase F), which breaks the bond between the core GlcNAc and Asn residue.⁴ *O*-GalNAc glycans, also referred to as mucin *O*-glycans, have a building block of *N*-acetylgalactosamine (GalNAc) attached to the hydroxyl group of serine or threonine.⁵ Their prediction is much more difficult, as no specific consensus sequence motifs have been identified. However, disordered regions caused by amino acid sequences rich in serine, threonine, and proline residues have been found to carry the majority of identified *O*-glycosylation sites in mapping studies.⁶ There are four core structures: core-1 Galβ1-3GalNAcαSer/Thr, core-2 GlcNAcβ1-6(Galβ1-3)GalNAcαSer/Thr, core-3 GlcNAcβ1-3GalNAcαSer/Thr, and core-4 GlcNAcβ1-6(GlcNAcβ1-3)GalNAcαSer/Thr.⁵ The high density of glycosylation sites on viral proteins that are exposed on the surface has often been described as a shield, ensuring immune evasion.⁷

Glycosylation on viral proteins and host receptors has been shown to facilitate viral attachment to host cells. Herein, we discuss the different implications of viral protein glycosylation and what tools are best utilized to characterize and locate glycosylation. This chapter highlights the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic and how glycosylation studies have assisted in vaccine and therapeutic development. Herein, we will focus on *N*-linked and mucin-type *O*-linked glycosylation as PTMs.

1.2 Viral Glycobiology

Glycoproteomic analysis has been an invaluable tool in various fields of research for decades. However, the global health crisis caused by the SARS-CoV-2 virus demanded an urgent development of therapeutics and vaccines. Without state-of-the-art glycoproteomic tools and established methods in glycoprotein analysis, this critical endeavor would not have been achievable within approximately 18 months.

SARS-CoV-2 is a member of the *Coronaviridae* family and belongs to the enveloped viral pathogens.⁸ In general, viruses hijack the molecular machinery of the host cell for reproduction and transmission of viral particles.⁹ The glycosylation pathways are also utilized by viruses to allow for the

secretion of viral proteins that are post-translationally modified to mimic the host 'self'.¹⁰ Therefore, the viral proteins either evade an immune response or elicit only a weakened response so that the viral replication is not significantly affected.¹¹ Apart from immune evasion, often heavily glycosylated viral glycoproteins are crucial for the functional integrity of viral building blocks, formation of viral particles, and infectivity.¹²

The prevalence of glycoforms post-translationally attached to the viral proteins is dependent on the host cell, *i.e.* cell type, cell differentiation, stage of the cell cycle, and other pathogenetic events, among other factors.¹³ The viral glycosylation pattern often mirrors that of the host due to the utilization of the host cell machinery. However, viral particles do not always follow the entire secretory pathway, and when released prematurely, a higher percentage of under-processed high-mannose-type glycans are evident within the glycosylation profile of the viral protein(s).¹⁰ These glycans can trigger immune responses through glycan-binding proteins (GBPs) or lectins of the eukaryotic immune system that serve as recognition receptors, *e.g.* the human C-type lectin dendritic cell-specific intercellular adhesion molecule-3 grabbing nonintegrin (DC-SIGN; CD209; CLEC4L) that is involved in interleukin signaling pathways.^{14,15}

It should be mentioned that certain large viruses, such as chloroviruses and mimiviruses, contribute to the glycosylation machinery in the host cell. For example, the chloroviruses encode information for carbohydrate-manipulating enzymes and influence the host's carbohydrate metabolism.¹⁶ Furthermore, mimiviruses contribute genes encoding glycosyltransferases.¹⁷

Other viruses interfere with the expression of host glycosyltransferases, changing the glycosylation profile as a result. One study showed that the expression of dormant genes encoding for fucosyltransferases (FUTs) in uninfected cells was activated after infection with herpes viruses. The increase in protein PTMs with sialyl-Lewis X (sLe^x) and Lewis Y (Le^y) antigens was hypothesized to contribute to the viral spread and immune evasion.¹⁸

The glycans on host cells can provide a point of attachment for the subsequent viral entry into the host cell *via* carbohydrate-binding domains or proteins that recognize fucose residues, proteoglycans, or sialic acids.^{9,19} A renowned example is the influenza virus and its ability to attach to sialylated glycoforms on the host cell through its hemagglutinin (HA) proteins on the viral surface. The human influenza viruses mainly utilize the α 2,6-linked sialic acid residues on human host cells, whereas the avian influenza viruses favor α 2,3-linked sialic acid residues.²⁰ However, interspecies transmission has been observed due to two factors: first, amino acid sequence mutations can alter the receptor-binding specificity of the HA, causing viral transmission between different species.²¹ Second, it has been shown that the human glycosylation profile in the upper respiratory tract mainly displays sialylated glycoforms with α 2,6-linkage, whereas α 2,3-linked sialic acid moieties are expressed in the lower respiratory tract.

Therefore, avian influenza viruses that reach the lower respiratory tract of humans can attach to cells and cause infection.^{20,22}

Proteoglycans and glycosaminoglycans (GAGs) are often a point for viral attachment at the host cell surface. The coronaviruses, including SARS-CoV-2, utilize glycosaminoglycans on the surface of cells and proteins. In particular, host cell-surface heparan sulfates (HS) are exploited to facilitate viral attachment while still allowing the interaction between the receptor-binding domain (RBD) on the viral spike protein and the human angiotensin-converting enzyme 2 (hACE2) receptor.²³ The interaction between SARS-CoV-2 and HS is described in more detail in the following sections.

1.3 SARS-CoV-2 and Its Structural Proteins

SARS-CoV-2 and its resultant coronavirus disease 2019 (COVID-19) have been the focus of many research efforts over the last 18 months due to the high transmissibility rate of viruses.^{24,25} SARS-CoV-2 predominantly attacks lung cells and, in some cases, can lead to pneumonia and acute respiratory distress syndrome (ARDS). Coronaviruses are enveloped, single-stranded RNA viruses. To date, seven coronaviruses related to human diseases have been identified. Of these, SARS-CoV-1, Middle East respiratory syndrome CoV (MERS-CoV), and SARS-CoV-2 can lead to upper respiratory infections and ARDS and are the most severe.²⁶ Although the SARS-CoV-2 virus is novel, it has similarities with SARS-CoV-1 and MERS-CoV. Herein, we discuss the structural proteins of SARS-CoV-2, their implication in viral infection, and their PTMs.

1.3.1 SARS-CoV-2 Spike Protein

1.3.1.1 *Functions of the Spike Protein in SARS-CoV-2 Infection*

The spike (S) protein is a trimeric class I fusion protein consisting of two subunits (S1 and S2).^{24,27–30} The S1 subunit facilitates viral attachment to the host cell through the RBD, and S2 allows for the fusion of the virus to human cellular membranes. SARS-CoV-2 utilizes the hACE2 as an entry receptor.^{24,31–36} SARS-CoV-2 has also been shown to utilize host cell-surface heparin (Hp) and HS as co-receptors to facilitate viral entry into the host cell.^{37–41} The receptor and co-receptor are independently engaged by the virus, but both are necessary to create an optimal entry point. Due to the key role played by the S protein in viral entry, it has been the focus of therapeutics. Understanding the S protein's post-translational modifications, including site-specific glycosylation, is important to correlate structural variation with immunogenicity for glycoprotein-based vaccine candidates.²⁷

The two subunits of the S protein are linked through transmembrane protease serine 2 (TMPRSS2) and furin cleavage sites. TMPRSS2 is an

acid-dependent protease, which carries out the cleavage of the S protein, followed by the fusion of the viral and host cell membranes. This cleavage occurs at two different positions in the S2 domain of the S protein.^{26,42} The first separates the RBD and fusion domains; the second exposes the fusion peptide. The acid-dependent cleavage occurs in the endosomes. The internal fusion peptides of SARS-CoV-2 and SARS-CoV-1 are identical, indicating a common mechanism of viral fusion and entry into the host cell. However, SARS-CoV-2 has 12 extra nucleotides upstream from the cleavage site, which form a similar sequence to canonical furin-like cleavage sites.^{26,43–45} It facilitates S protein priming and may increase the efficiency of SARS-CoV-2 spread compared to other coronaviruses.^{26,46,47} This could shed light on the increased infectivity rate of SARS-CoV-2 (200+ million humans infected and 1.6% mortality rate as of September 2021) compared to SARS-CoV-1 (8098 humans infected and 9% mortality rate) and MERS-CoV (2519 humans infected and 34.4% mortality rate).²⁶

1.3.1.2 Spike Protein Glycosylation

The S protein is heavily glycosylated and contains both *N*- and *O*-glycans. When the S protein is expressed in HEK293 cells, extensive *N*-glycosylation was located on 22 sites.^{27,29} The *N*-glycans present are highly heterogeneous and include high-mannose-type glycans, hybrid-type glycans, and highly processed complex-type glycans. Eight of the 22 sites were determined to be occupied predominantly with high-mannose-type structures by Watanabe *et al.* (N61, N122, N234, N603, N709, N717, N801, and N1074).²⁹ However, Shajahan *et al.*, Walls *et al.*, and Zhang *et al.* noted an abundance of complex-type glycans (Figure 1.1).^{30,48} Specifically, the preserved glycosites between SARS-CoV-1 and SARS-CoV-2 (N61, N122, N603, N709, N801, and N1074), which had previously been characterized as containing only high-mannose-type glycans, have since been identified as containing complex-type glycans.^{24,27,29,30} Sites N165, N331, and N343 express complex-type *N*-glycans containing both bi- and tri-antennary structures.^{27,29} Interestingly, sites N331 and N343, which are within the RBD region, showed predominantly high-mannose-type glycans when the S protein was expressed in individual subunits S1 and S2.^{7,27} Variable glycosylation of the S protein was also noted when comparing native S protein with the recombinant form used for vaccine design.⁴⁹ This variation likely results from distinctive cellular secretion pathways and variation between different cell types.

Shajahan *et al.* initially reported *O*-glycosylation at sites S323 and T325.^{24,27} This work utilized a liquid chromatography–tandem mass spectrometry (LC–MS/MS) workflow of the S protein expressed in HEK293 cells. However, since these initial glycosylation studies, there have been additional studies focusing on the *O*-glycosylation of the S protein. Sanda *et al.* reported the identification of a novel *O*-glycosylation site, T678, which is near the furin cleavage site of the S protein.⁵⁰ In addition to this site, Sanda *et al.* also

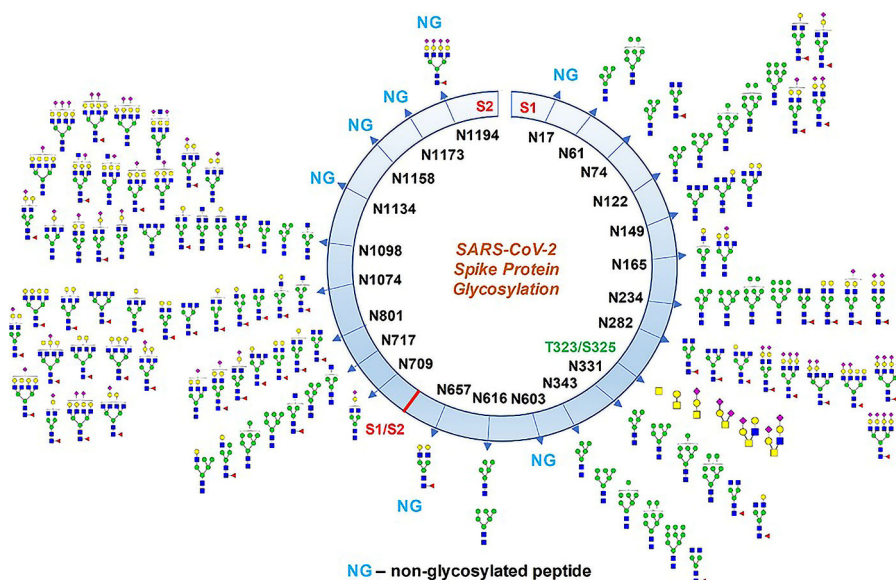


Figure 1.1 Glycosylation profile of the SARS-CoV-2 S protein. These data confirmed that 17 of the proposed 22 *N*-glycosylation sites were glycosylated. *O*-glycosylation was also determined to be present at site T323 or S325. Reproduced from ref. 27 with permission from Oxford University Press, Copyright 2020.

confirmed *O*-glycosylation at S323 and T325. Sites S323 and T325 contain core-1 and core-2 type *O*-glycans, and site T325 is noted as being the primarily occupied site of the two.^{27,29} The recently reported site T678 also contained core-1 and core-2 type *O*-glycans.⁵⁰ A large portion of the *O*-glycans identified in the SARS-CoV-2 S protein contain terminal sialic acid (NeuAc). A recent study from Tian *et al.* utilized PNGase F release in the presence of ¹⁸O water to confirm *N*-glycosylation sites and determine *O*-glycosylation sites.⁵¹ The ¹⁸O labeling in this study confirmed the 22 *N*-glycosylation sites previously reported and identified 17 novel *O*-glycosylation sites on the S protein extracted from virions. *O*-glycosylation has been noted for the recombinant S protein,^{27,29} but native virion S protein *O*-glycosylation had previously not been characterized. Tian *et al.* found 11 *O*-glycosylation sites on S1 and 6 at the N-terminal domain (NTD) of S2. These novel sites include S60, T124, S151, T236, T604/S605, T618, S659, T1076, T1077, S1097, and T1100, which are all located in close proximity to Asn residues.⁵¹ There were 35 Ser/Thr residues located within 3 amino acids of Asn residues, 11 of which were determined to be glycosylated, and 7 of these sites were located within 2 amino acids of an Asn residue. When a glycosylated Asn residue (N616) was mutated, it was noted that *O*-glycosylation on site T618 was abolished, suggesting that *O*-glycosylation occurs in an “O-follow-N” pattern, meaning that *O*-glycosylation is found near the glycosylated Asn residue in the N-sequence.⁵¹

1.3.1.3 Host Receptors and Their Glycosylation

S protein *N*-glycosylation sites N165 and N234 stabilize the RBD “up” conformation, which permits effective binding to hACE2.^{24,52} Mutation of these two glycosylation sites significantly reduces binding, as there is a conformational change to the “down” position. hACE2 is also glycosylated with seven *N*-glycosylation sites and one *O*-glycosylation site as shown in Figure 1.2.⁵² Site N90 has been indicated as playing an important role in SARS-CoV-2 and hACE2 binding. *In-silico* experiments show that mutating N90 leads to a stronger interaction of hACE2 with SARS-CoV-2, indicating that glycosylation at this site may impose steric hindrance for RBD binding.^{53,54} However, Devaux *et al.* have demonstrated that species with ACE2 sequences containing N90 along with K31, Y41, and K353 are more likely to be susceptible to SARS-CoV-2 infection.⁵⁵ This result is contradictory to the previously mentioned findings, which highlights the need for more research.

Independent of hACE2, cell-surface GAGs aid SARS-CoV-2 in host cell binding (Figure 1.3).^{24,37,41,56} GAGs bind to proteins that contain Cardin–Weintraub motifs (XB₂BXB₂ or XB₂BXB₂XB₂, where B is a basic residue and X

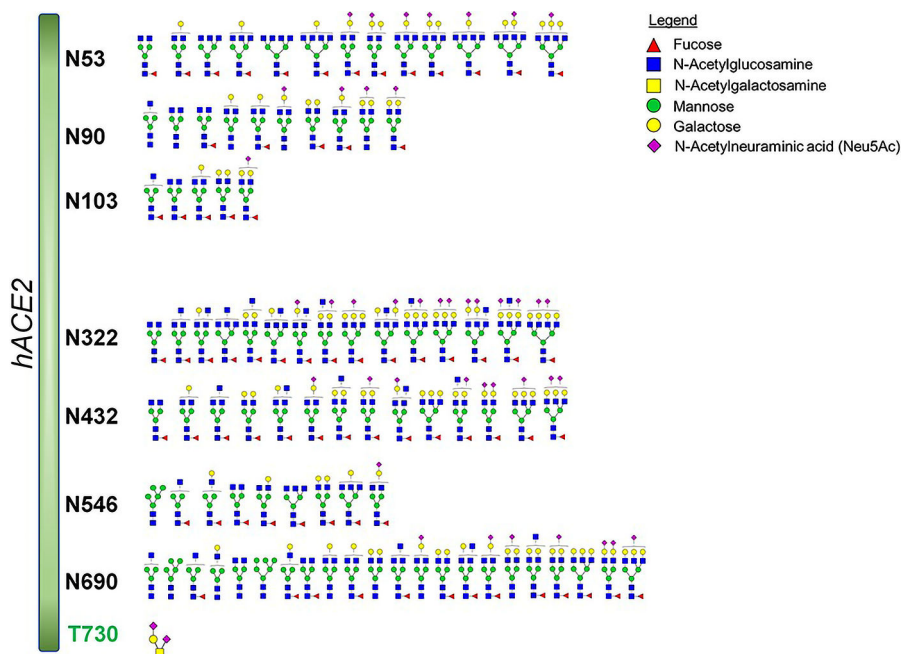


Figure 1.2 Glycosylation profile of human angiotensin-converting enzyme 2 (hACE2) receptor. This work outlined 7 *N*-glycosylation sites, each with 5 or more glycoforms present, and one *O*-glycosylation site. Mainly complex-type *N*-glycans were present. Reproduced from ref. 52 with permission from Oxford University Press, Copyright 2020.

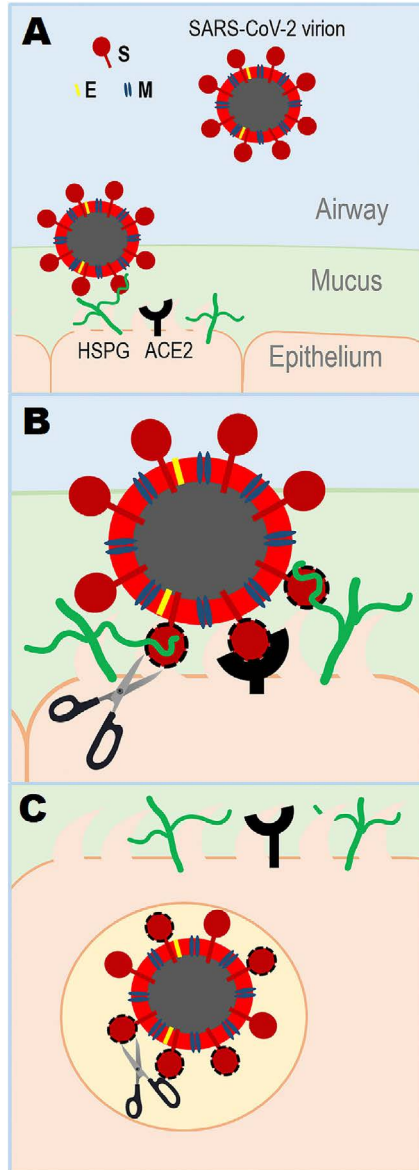


Figure 1.3 Proposed SARS-CoV-2 host cell entry model utilizing host cell-surface GAGs. (A) Virion binds to host cell heparan sulfate proteoglycan (HSPG). (B) Host cell-surface proteases initiate the proteolytic digestion of the S protein, which initiates viral–host cell membrane fusion due to conformational changes by host cell receptor binding of ACE2 and HSPG. (C) Virion enters the host cell where it can undergo further proteolytic processing by endosomal host cell protease. Reproduced from ref. 39 with permission from Elsevier, Copyright 2020.

is a hydrophobic residue), and such motifs were found at three sites on the S protein of SARS-CoV-2.⁵⁶ The RBD contains two GAG-binding motifs at sites Y453-S459 and P681-S686, and when the RBD is in the open conformation, these sites interact with cell-surface heparin/HS. SARS-CoV-1 is also known to utilize cell-surface GAGs during host cell entry. A study utilizing surface plasmon resonance (SPR) found that both monomeric and trimeric SARS-CoV-2 S proteins bound more tightly to Hp than SARS-CoV-1 or MERS-CoV S proteins.⁵⁶ The binding of cell-surface GAGs to S protein trimers enhances the binding affinity of ACE2. This indicates that cell-surface GAGs act as co-receptors for RBD-ACE2 interactions. When cell-surface Hp/HS was first removed using a mixture of heparin lyases (HSase) I, II, and III in multiple cell types, SARS-CoV-2 infection was prevented.^{24,37} This prevention was not seen for SARS-CoV-1, further indicating the tighter binding of the SARS-CoV-2 S protein to Hp/HS than SARS-CoV-1. The high affinity for the SARS-CoV-2 S protein binding to cell-surface GAGs suggests that treatment with GAGs may lead to competitive binding. This hypothesis was tested, and trisulfated (Tris) Hp, USP-Hp, and two fucoidan structures were able to compete with cell-surface GAGs for S protein binding.^{24,41} These findings suggest that treatment with GAGs and GAG-like structures may be a useful preventative measure for combating SARS-CoV-2 infection; however, further studies are needed.

1.3.1.4 *Advances in Vaccines and Therapeutics Utilizing the Spike Protein*

After the start of the COVID-19 global pandemic, researchers quickly noted the importance of the S protein in viral infection of human host cells. This made the S protein a logical target for vaccines and therapeutics. Some of the vaccine types being utilized or tested for SARS-CoV-2 include inactivated virus, protein subunit, viral vector, DNA and messenger RNA (mRNA), virus-like particle (VLP)/nanoparticle, and live-attenuated virus vaccines.⁵⁷ One of the main vaccine types utilized during this pandemic is based on mRNA technology. mRNA vaccines use a modified, harmless protein (or portion of the protein), which teaches human cells how to make this modified protein, triggering an immune response.⁵⁸ Some mRNA SARS-CoV-2 vaccines use a version of the S protein that is modified by two proline mutations.⁵⁷ These mutations lock the S protein in its prefusion conformation. The mRNA serves as a template for the expression of the locked S protein on the surface of human cells, which then produces the S protein, resulting in human cells with S protein expressed on their surface. This then allows the immune system to learn how to identify and fight the S protein. Other mRNA vaccines utilize the RBD region of the S protein.⁵⁷

Another vaccine type, viral vector, has also utilized the S protein of SARS-CoV-2. Viral vector vaccines use a harmless, unrelated virus as a vector to carry genetic instructions into the host cell.⁵⁸ In the case of SARS-CoV-2,

this is the S protein. Once the viral vector has entered the host, the S protein is expressed on the surface of host cells, triggering an immune response. Similar to the mRNA vaccines, some viral vector vaccines use a modified version of the S protein with two proline mutations. However, this version also has two additional mutations at the furin cleavage site. Other viral vector vaccines utilize the full-length S protein without the proline or furin cleavage-site modifications.⁵⁷

As many of the current vaccines used for SARS-CoV-2 apply a modified version of the S protein, it is likely that glycosylation differences will also be observed. Liu *et al.* recently reported a potential vaccine candidate, which utilizes the recombinant RBD domain of the S protein, prepared by means of a novel yeast expression system, *Pichia pastoris*.⁵⁹ This vaccine triggered RBD-directed antibody production in mice to produce anti-RBD-specific antibodies, and virus-neutralizing antibodies (NAbs) were maintained in mice for more than 6 months. The recombinant S protein used for this study showed a 25% increase in mass. However, when the recombinant S protein was treated with PNGase F, the mass matched the theoretical mass of the deglycosylated protein, indicating the presence of *N*-glycosylation.⁵⁹ The RBD region used in this study mainly consists of Man₅GlcNAc₂ and Gal₂GlcNAc₂Man₃GlcNAc₂.⁵⁹ Conversely, HEK293-derived RBD had previously been characterized and contained fucosylated di- and triantennary complex-type *N*-glycans, with a small amount of high-mannose-type glycans.⁶⁰ These studies highlight the differences in glycosylation that can occur when the S protein is expressed in different cell types.

1.3.2 Envelope, Nucleocapsid, and Membrane Proteins

1.3.2.1 Functions of the Envelope, Nucleocapsid, and Membrane Proteins in SARS-CoV-2 Infection

The membrane (M) protein is a glycoprotein composed of 222 amino acids and has an N-terminal ectodomain and a C-terminal endodomain. The M protein contains three N-terminal membrane-spanning domains, which are essential for viral particle assembly. The M protein, along with the envelope (E) protein, regulates intracellular trafficking and intracellular processing of the S protein.^{61,62}

The E protein is the smallest of the structural proteins and contains an N-terminal ectodomain and a C-terminal endodomain. The E protein has ion channel activity, which could be required for SARS-CoV-2 pathogenesis. The E protein also plays a vital role in virus assembly and release.³⁸

The nucleocapsid (N) protein has an N-terminal domain and a C-terminal domain, both of which can bind to RNA. The N protein helps in the packaging of the viral genome into the viral particles by interacting with the M protein and a component of the replicase complex, facilitating the binding to the replicase-transcriptase complex.⁶³

1.3.2.2 Post-translational Modifications of Envelope, Nucleocapsid, and Membrane Proteins

Based on *in-silico* experiments, both the M and E proteins are glycosylated. The E protein of SARS-CoV-2 contains two possible glycosylation sites (N48 and N66). These sites are both located C-terminally of the transmembrane site. However, due to the close proximity of site N48 to the membrane (steric hindrance and hydrophobicity), it is likely that the E protein is only glycosylated at site N66.^{38,64,65} *In vitro* experiments are still needed to confirm E protein glycosylation.

The M protein of SARS-CoV-2 contains 8 possible glycosylation sites (N5, N21, N41, N43, N117, N121, N203, and N216).^{38,61} Currently, the glycosylation of the M protein has not been reported outside of *in-silico* experiments. However, it has been found that the M protein of SARS-CoV-2 resembles a sugar transporter named SemiSWEET (sugars will eventually be exported transporter).⁶² SemiSWEETs are composed of a simple triple helix bundle, whereas SWEETs contain multiple triple helix bundles connected by an inversion linker helix.⁶⁶ SemiSWEETs catalyze the diffusion of sugars driven by their concentration gradients.⁶⁷ The presence of a sugar transporter could influence sucrose entry into the lysosome, endosome, and/or autophagosome, which could aid in the release of the virus into cells. This transporter may be an efficient mechanism to induce rapid viral proliferation and immune evasion.⁶²

The SARS-CoV-2 N protein is located in the nucleocapsid and does not go through the secretory pathway. Because of this, the N protein is not expected to be glycosylated. A recent study by Supekar *et al.* confirmed this and noted the phosphorylation present as a PTM (Figure 1.4).⁶³ Interestingly, the location of PTMs has been shown to be dependent on the source of the protein. The recombinant N protein on its own showed no signs

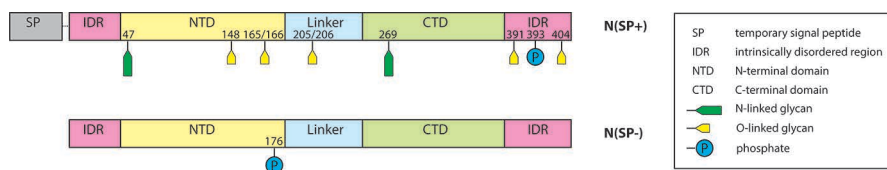


Figure 1.4 Domain structure of the N protein with a signal protein (SP+) or without a signal protein (SP-). The N protein containing a signal protein to be processed through the secretory pathway showed N- and O-glycosylations. However, the N protein lacking the signal peptide, which is more intuitive than the wild-type N protein, did not contain glycosylation. In both cases, phosphorylation was present; however, the location of phosphorylation changed from the intrinsically disordered region (IDR) with the signal peptide to the N-terminal domain (NTD) when the signal protein was absent. Reproduced from ref. 63 with permission from Oxford University Press, Copyright 2021.

of glycosylation, and phosphorylation was present at Ser176.⁶³ However, a commercial N protein preparation contained *N*- and *O*-linked glycosylation and phosphorylation at Thr393. It was determined that the commercial protein was designed with a proprietary N-terminal signal peptide sequence that allowed it to enter the secretory pathway of the HEK293 host cells. This led to distinctive glycosylation and phosphorylation of the recombinant N protein lacking the signal peptide sequence, which therefore remained in the cytosol.⁶³ The recombinant N protein lacking the signal peptide sequence more closely resembles the native N protein.

1.3.3 Mutations in Structural Protein Glycosylation

Since the onset of the COVID-19 pandemic, multiple variants of SARS-CoV-2 have emerged. Of these variants, four have been classified as variants of concern (VOCs). VOCs are designated by the World Health Organization (WHO) as having an increase in transmissibility or a detrimental change in COVID-19 epidemiology, or an increase in virulence or a change in clinical disease presentation, or a decrease in the effectiveness of public health and social measures or available diagnostics, vaccines, and therapeutics.⁶⁸ The Alpha (B.1.1.7) variant was first designated in mid-December 2020 in the United Kingdom. The Alpha variant is estimated to be 40–70% more transmissible than the wild-type SARS-CoV-2 virus.⁶⁹ This increase in transmissibility is likely due to the N501Y mutation, which increases the receptor-binding affinity of the S protein with hACE2. In addition to increased transmissibility, there is also an estimated 30–50% increase in mortality compared to the wild-type strain.⁶⁹ At the same time the Alpha variant was designated, the Beta variant (B.1.351) was designated in South Africa.⁶⁸ Unlike the wild-type strain, the Beta variant had a higher prevalence among young people with no underlying health conditions. In January 2021, the Gamma variant (P.1) was designated in Brazil.⁶⁸ Both the transmissibility and the mortality of this variant are much higher than those of the wild-type strain. Most recently, after an increase in infection rates was seen in India in Spring 2021, the Delta variant (B.1.617.2), which was previously being monitored as a variant of interest (VOI) since December 2020, has been deemed a VOC.⁶⁸ As of July 2021, the Delta variant had spread to 124 countries and has become the dominant strain of SARS-CoV-2 globally.

A few recent pre-prints have investigated the glycosylation change between the wild-type S protein and the D614G mutant.^{69,70} One study detected 21 *N*-glycosylation sites, 10 of which (N17, N61, N74, N331, N343, N657, N1074, N1158, N1173, and N1194) had little to no change in the distribution of both individual glycans and glycan types. Three of these unchanged sites are located in the stalk region/C-terminal portion of the S2 subunit. This indicates that the complex glycan shielding of the stalk region remains intact with the D614G mutant.⁶⁹ Sites N331 and N343 are

the only two *N*-glycosylation sites in the RBD; however, they are not in the binding motif of the RBD. There was no change noted in the glycosylation of these sites. This could show that they are not involved and do not affect RBD–ACE2 binding directly or indirectly; however, it could also indicate that glycosylation at these sites is important for RBD–ACE2 binding, as all VOCs maintain glycosylation at these sites. Eleven *N*-glycosylation sites (N122, N165, N234, N282, N603, N616, N709, N717, N801, N1098, and N1134) did have glycosylation changes. These sites are all located in the head region of the S protein. Almost all sites in the lower portion of the head showed significant changes, whereas only half of the sites in the top portion of the head did. Sites N122, N234, N603, N709, and N801 had an increase in high-mannose-type and a decrease in both hybrid- and complex-type glycans. Sites N165, N282, N616, N1098, and N1134 were increased in both high-mannose- and hybrid-type glycans and decreased in complex-type glycans. The glycans on site N282, which is located in the N-terminal domain and situated away from the RBD region, were less fucosylated in mutant S proteins compared to the wild-type (45–48% and 73–80%, respectively).⁷⁰ In these recent studies, the wild-type and mutant S proteins were grown and processed the same way, and therefore, variation in processing is likely not the cause of these changes. Based on the changes in glycosylation type paired with cryo-electron microscopy (cryo-EM) data, it has been proposed that the D614G substitution results in more open conformations or a higher percentage of RBD in the upward position that facilitates RBD–ACE2 interactions.⁶⁹ Additional studies are needed on how glycosylation of the different variants, especially the variants of concern, is changed.

1.4 Viral Protein Analysis

The structural determination of viral proteins is critical to elucidate the functional roles of these proteins in the viral life cycle. Earlier structural studies on the SARS-CoV-2 S protein were initiated with cryo-EM studies on its trimeric stabilized form. Subsequently, mass spectrometry was used for the detailed profiling of the S protein glycosylation. Later, the complementary data from cryo-EM and mass spectrometry were used for the three-dimensional (3D) modeling, visualizations, and molecular dynamics (MD) studies. This section also discusses the applications of biosensors in determining the structure-based binding interactions of the S protein with its receptor.

1.4.1 Cryo-electron Microscopy

The detailed structural characterization of the SARS-CoV-2 S protein provided a basis for the design and study of vaccines and inhibitors of

the virus. Cryo-EM provided the initial breakthroughs in understanding the structural proteins of SARS-CoV-2 in detail, particularly the S protein and its glycosylation. Cryo-EM bombards flash-frozen solutions of proteins or other biomolecules with electrons to produce microscopic images of individual biomolecules. It is used to reconstruct the 3D shape, or structure, of the molecule and thereby aid in uncovering protein function, changes in diseases, and drug targeting.⁷¹

For the expression of viral fusion glycoproteins, prefusion stabilization is performed to decrease the yield and likelihood of misfolded proteins due to the adaptation of the more stable postfusion structure.^{72,73} Prefusion-stabilized viral glycoprotein variants (S-2P) contain two consecutive proline substitutions in the S2 subunit in a turn between the central helix and the heptad repeat 1 (HR1).⁷³ Such a stabilization strategy, which was used for other betacoronavirus S proteins, was also used to determine high-resolution cryo-EM structures of the SARS-CoV-2 S protein.^{74,75} A prefusion-stabilized ectodomain trimer construct with an abrogated furin S1/S2 cleavage site, two consecutive proline-stabilizing mutations, and a C-terminal fold-on trimerization domain were designed in one study to enable single-particle cryo-EM studies.⁷⁴ Another group reported the expression of the ectodomain residues 1 through 1208 of the SARS-CoV-2 S protein based on the first reported genome sequence of SARS-CoV-2 by adding two stabilizing proline mutations in the C-terminal S2 fusion machinery (Figure 1.5).⁷⁵ Both studies report the expression of the recombinant prefusion-stabilized S ectodomain from FreeStyle 293 cells and purification by affinity chromatography and size-exclusion chromatography (SEC); therefore, there is no expected variation based on S protein expression systems between these two studies.^{73–75}

3D cryo-EM grids were prepared using the purified, fully glycosylated S protein. The cryo-EM reconstruction was reported to be obtained at a 2.8 Å resolution for a fully closed RBD state and 3.2 Å resolution for one-RBD-open asymmetric conformation.⁷⁴ In a parallel study, Wrapp *et al.* determined the asymmetric one-RBD-open conformation at 3.5 Å resolution.⁷⁵ The reconstructed cryo-EM structures of the SARS-CoV-2 S ectodomain trimer revealed that it is a 160 Å-long trimer with a triangular cross-section, resembling the closely related SARS-CoV-1 S structure.⁷⁴ Multiple conformational states of the SARS-CoV-2 S protein, based on the distinct organization of the RBD domain, were observed.^{74,75} The S protein trimer with one of the three RBDs rotated up (open state) in a receptor-accessible conformation is found as the predominant state.⁷⁵ Such variability in the conformation of RBD domains is also observed on SARS-CoV-1 S and MERS-CoV S trimers. The SARS-CoV-2 S1 subunit has a V-shaped architecture. The open conformation of the SARS-CoV-2 RBD domain is necessary for interacting with host receptor ACE2 and changes in the conformations that lead to the cleavage of the S2 site, membrane fusion, and viral entry.⁷⁴

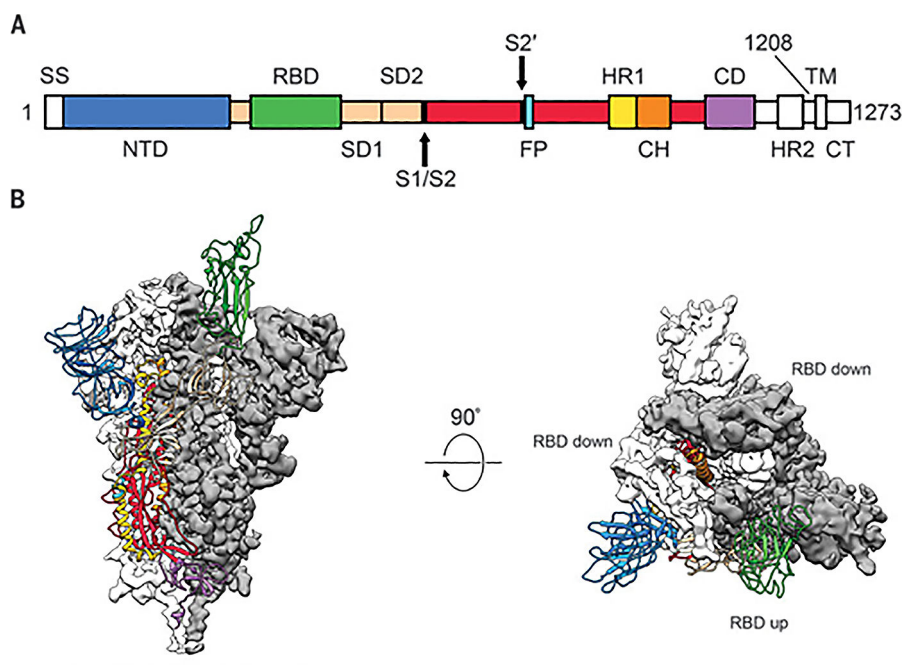


Figure 1.5 Structure of the SARS-CoV-2 S protein in the prefusion conformation. (A) The primary structure of the SARS-CoV-2 S protein, colored by domain. SS, signal sequence; S2', S2' protease cleavage site; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; CD, connector domain; HR2, heptad repeat 2; TM, transmembrane domain; CT, cytoplasmic tail. Protease cleavage sites are shown with arrows. (B) Side and top views of the prefusion structure of the SARS-CoV-2 S protein with a single RBD in the up conformation. Reproduced from ref. 75, <https://doi.org/10.1126/science.abb2507>, under the terms of the CC BY 4.0 license, <https://creativecommons.org/licenses/by/4.0/>.

The conformational variation of the S1 subunit as the RBD undergoes a hinge-like movement was observed on the SARS-CoV-2 S protein. This likely contributes to the relatively poor local resolution of S1 compared with the more stable S2 subunit.⁷⁵ Such type of stochastic RBD movement was also observed on the closely related betacoronaviruses SARS-CoV-1 and MERS-CoV. This indicates that SARS-CoV-2 shares a similar mechanism with other betacoronavirus receptors, wherein the receptor binding to exposed RBDs triggers an unstable three-RBD up conformation, which subsequently leads to the shedding of S1 and the refolding of S2.^{72,73}

The SARS-CoV-2 S protein ectodomain model was built with the glycans at 44 of the 66 *N*-linked glycosylation sites per trimer and 16 out of 22 potential *N*-glycosylation sites by two independent cryo-EM studies.^{74,75} In general, the structure of the SARS-CoV-2 S protein is similar to that of the SARS-CoV-1 S protein, with one of the larger differences being the position

of the RBDs in their respective down conformations. The authors noted that the SARS-CoV-1 RBD in the down conformation packs tightly against the NTD of the neighboring protomer, whereas the SARS-CoV-2 RBD in the down conformation is angled closer to the central cavity of the trimer.⁷⁴ However, a high degree of structural homology was observed between the two proteins when the individual structural domains of the SARS-CoV-2 S protein were aligned to those of the SARS-CoV-1 S protein.^{74,75}

A recent cryo-EM structural analysis of the complex formed between a mutant S protein N501Y ectodomain and the ACE2 receptor ectodomain revealed that the overall structure at the binding site of the mutant is almost identical to that of the unmutated version. The only exception is the local rearrangements that result in the aromatic ring of Y501 being accommodated in a cavity that is sandwiched between Y41 and K353 of the ACE2 receptor. The same study also reported that the Y501 in the S protein and Y41 in the ACE2 receptor produce a perpendicular π -shaped π - π stacking interaction.⁷⁶

1.4.2 Mass Spectrometry

While the initial cryo-EM studies provided insights into the dense glycosylation of the SARS-CoV-2 S protein, cryo-EM yields low-resolution structures, and low-molecular-weight ligands may not be visualized. Because of these drawbacks, significant progress in the site-specific glycosylation analysis of the SARS-CoV-2 S protein occurred after the initial mass spectrometric studies.^{27,29,50} The glycosylation profiling on the S protein was conducted by multiple research groups on the S protein in trimeric stabilized and unstabilized forms, individual S1 and S2 subunits, and the S protein RBD domain.^{27,29,49,77} Moreover, various expression systems were used for the glycosylation study on the S protein, such as the most widely used HEK293-based system, insect cells, yeast cells, and, more recently, the native infectious virions themselves.^{27,29,35,49,77–79} A workflow of the analysis process of SARS-CoV-2 glycosylation is shown in Figure 1.6.⁴⁹

There was significant variability in the reporting of the site-specific N-glycosylation on the 22 sites of the SARS-CoV-2 S protein, as well as discrepancies in O-glycosylation,^{27,29,49,51,77–79} prompting warnings regarding the selection of appropriate expression systems for the development of inhibitors, antibodies, and vaccines.⁸⁰

For the site-specific analysis of the glycosylation of the SARS-CoV-2 S protein by mass spectrometry, the glycoprotein was digested by common proteases, such as trypsin, chymotrypsin, and α -lytic protease. In the majority of the studies, state-of-the-art nano-liquid chromatography (nLC) coupled with high-resolution mass spectrometers was used for glycopeptide separation and data acquisition. The tandem mass spectrometry (MS/MS) fragmentations were conducted predominantly through stepped collisional energy (SCE), higher-energy collisional dissociation (HCD), and

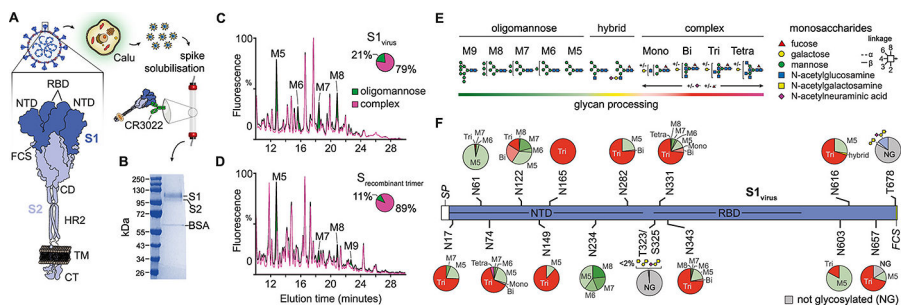


Figure 1.6 Analysis of the SARS-CoV-2 spike glycoprotein glycosylation by mass spectrometry. (A) The S protein is purified from SARS-CoV-2-infected Calu-3 cells by immunoaffinity purification using the antibody CR3022 against subunit S1. Spike S1, dark blue; S2, light blue; RBD, receptor-binding domain; NTD, N-terminal domain; FCS, furin cleavage site; CD, connecting domain; HR2, heptad repeat; TM, transmembrane domain; CT, cytoplasmic tail labeled. (B) The presence of S1 and S2 subunits of the virus-derived spike is shown by SDS-PAGE. (C and D) The distribution of oligomannose and complex-type glycans on the S1 virus and S recombinant trimer determined by the quantitative ultra-high-performance liquid chromatography (UHPLC) N-glycan analysis, respectively. (E) The process of N-glycan maturation depicted by color coding for the degree of glycan processing from oligomannose (green) to hybrid (yellow) to complex (purple). (F) Site-specific N- and O-glycosylation determined by quantitative glycoproteomics of subunit S1 of infectious virus. Pie charts show the degree of N-glycan processing depicted in (E). Reproduced from ref. 49 with permission from American Chemical Society, Copyright 2021.

HCD-product-triggered electron transfer dissociation (HCDpdEthcD) or HCDpd-collision-induced dissociation (CID) methods. As a rule, data were interpreted by software tools, such as Byonic (different versions) and Proteome Discoverer 2.2, for site-specific glycan determination, although extensive manual interpretation and validation of the mass spectrometric data were performed, especially with respect to the exact mass measurements and the presence of glycan oxonium ions and peptide/glycan fragments in the MS/MS spectra.^{27,29,35,49,51,77–79,81}

Subsequently, improvements in the tandem mass spectrometric method, instrumentation, and sample processing strategies have further empowered the glycosylation study of the SARS-CoV-2 S protein.^{77,82} A set of methods developed by Sanda *et al.* helped in the detection of LacdiNAc structural motifs on all occupied N-glycopeptides and polyLacNAc structures on six glycopeptides of the spike protein in addition to the detection of 9 O-glycosylation sites.^{50,77} The study includes novel methods, such as (1) peptide and glycopeptide separation on nLC by 5 min trapping/washing step using 99% solvent A (2% acetonitrile and 0.1% formic acid) at 10 $\mu\text{L min}^{-1}$, (2) using modulation of collision energy for selective fragmentation

of the glycopeptides to identify structural motifs of the *N*-glycosylated peptides, (3) cyclic ion mobility separation (cIMS)-based separation of glycans, and (4) beam-type tandem mass spectra fragmentation of glycans.⁵⁰ The extensive *O*-glycosylation characterization on the SARS-CoV-2 S protein is enabled by prior de-*N*-glycosylation of the S protein with PNGase F or Endoglycosidase H (Endo H) and treatment with exoglycosidases.^{50,51,81} Sequential deglycosylation, by Endo H and PNGase F in the presence of ¹⁸O H₂O after the tryptic digestion of the S protein, was also performed to determine the occupancy of *N*-linked glycans at each site.^{35,51}

1.4.3 Molecular Modeling

As described in detail in Section 1.4.2, the cryo-EM data of the SARS-CoV-2 S protein trimer, available in the Protein Data Bank (PDB), do not include the significant glycosylated portion of the protein structure due to the lack of flexibility.⁸¹ The *N*-glycosylation of the SARS-CoV-2 S protein was mapped into experimentally determined 3D structure using the cryo-EM structure of the trimeric S protein (PDB – 6VSB) by Watanabe *et al.* This combination of mass spectrometric and cryo-EM analysis studies revealed the distinct regions across the surface of the S protein that are occluded by *N*-linked glycans (Figure 1.7).²⁹

In another study, the defined *N*-glycans at each glycosylated sequon were mapped into the 3D structure of the S trimer generated by using a homology model of the S trimer (PDB – 6VSB). The authors generated three glycoform models denoted “Abundance,” “Oxford Class,” and “Processed” models, which represent glycosylation microheterogeneity and cross-validating glycomic and glycoproteomic analysis data.³⁵

Multiple all-atom MD simulations with explicit water were performed on three glycoform models of the SARS-CoV-2 S protein.³⁵ This study also determined the variants of the S protein reported at that time and measured the inter-residue distances between the most α -carbon-distal atoms of the *N*-glycan sites and S protein variant sites in 3D space. Variants with mutations on residues, which are closest to the *N*-glycosylation sites, were identified in this analysis. Furthermore, the authors studied the percentage of simulation time that each S protein residue is accessible to a probe that has the approximate size of an antibody variable domain. Such predicted antibody accessibility was visualized across the S protein sequence and mapped on the 3D surface. The study further concluded that a substantial number of sequence variants have mutations in regions of high calculated epitope accessibility and suggested that this is a potential selective pressure to avoid host immune response. The same study also showed the interactions of the glycosylated SARS-CoV-2 S trimer with the soluble glycosylated form of hACE2 by 3D structural modeling and MD simulations, demonstrating that the glycan at N74 of the S protein interacts with the glycan at N546 of hACE2.³⁵

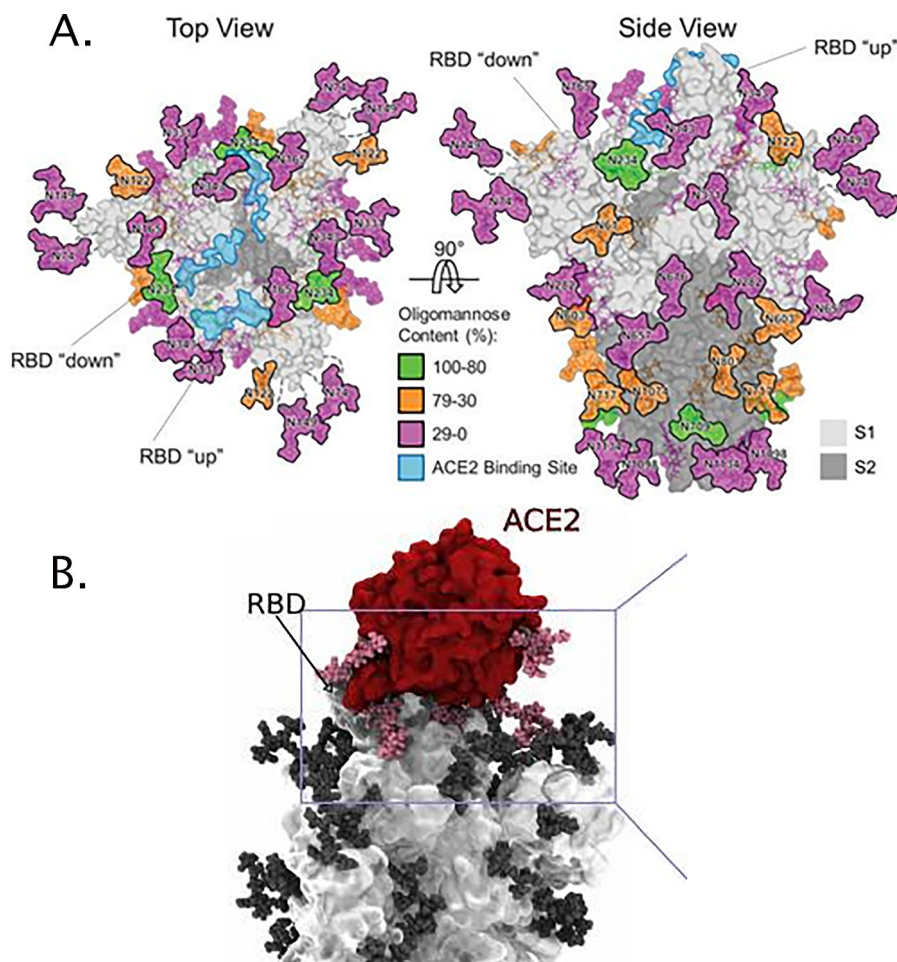


Figure 1.7 Mapping of SARS-CoV-2 S N-linked glycans by molecular modeling through cryo-EM and mass spectrometry data. (A) Representative N-glycans are mapped onto the prefusion with one RBD in the "up" conformation and the other two RBDs in the "down" conformation structure of the trimeric SARS-CoV-2 S glycoprotein (PDB ID 6VSB). (B) Interaction of glycosylated soluble human ACE2 and glycosylated SARS-CoV-2 S trimer immunogen visualized through MD simulation. Reproduced from ref. 35 with permission from Elsevier, Copyright 2021 and ref. 29, <https://doi.org/10.1126/science.abb9983>, under the terms of the CC BY 4.0 license, <https://creativecommons.org/licenses/by/4.0/>.

Several experimentally determined O-glycosylation sites were visualized in 3D on a fully N-glycosylated full-length S protein structure (PDB: model 6VSB) by Bagdonaite *et al.*, and such a model is the one with both N- and O-glycosites mapped on the SARS-CoV-2 S-protein. The authors used

Conformational Analysis Tools software to check for accessible *O*-glycosylation sites on the protein surface, and the solvent accessible surface (SAS) of Ser/Thr residues was estimated through the STRIDE algorithm. For the *O*-glycan visualization, the authors manually attached Gal β 3GalNAc α disaccharides to experimentally confirmed *O*-glycosylation sites.⁸¹

Molecular modeling and simulation studies complemented the structural and functional study of SARS-CoV-2 viral proteins through cryo-EM, mass spectrometry, and other experimental evidence. Casalino *et al.* studied the atomistic perspective on the roles of glycans and the structure and dynamics of the SARS-CoV-2 S protein. A full-length model of the glycosylated SARS-CoV-2 S protein in both open and closed states was built through structural and biological data and performed multiple microsecond-long, all-atom molecular dynamics simulations. This study highlighted an unexpected but vital structural role of *N*-glycans at sites N165 and N234 in modulating the conformational dynamics of the S protein RBD. The authors further confirmed this observation through biolayer interferometry experiments involving N165A and N234A mutations, which consequently reduced the ACE2 binding of the RBD due to the conformational shift of the RBD toward the “down” state. This study underscores the importance of glycosylation of the SARS-CoV-2 S protein by providing insights into the unseen structural and functional importance of glycosylation.⁸³ Cryo-EM studies only provided snapshots of the “up” and “down” states of the RBD. A recent study reported kinetically unbiased RBD-opening pathways through simulations of the fully glycosylated spike ectodomain. They combined the simulation data with ManifoldEM software analysis of cryo-electron microscopy data and biolayer interferometry experiments and reported that a gating role for the *N*-glycan at position N343 enables the RBD opening.⁸⁴

1.4.4 Surface Plasmon Resonance

Enzyme-linked immunosorbent assays (ELISAs) are typically used for the detection of antigens in a biological sample and can be designed as a high-throughput method. ELISA utilizes antibodies to detect a target antigen based on highly specific antibody–antigen interactions. The antigen of interest is immobilized to a solid surface and then complexed to a detection antibody conjugated with a fluorophore or enzyme for detection.⁸⁵ ELISA is a labor-intensive and expensive process and has a high likelihood of false-positive or false-negative results.⁸⁶ However, SPR sensing is a label-free and highly sensitive technique that is well suited for biomolecules, such as antibodies. SPR sensing has been used for the detection of antibodies in crude biofluids, particularly antibodies against SARS-CoV-1, and for the determination of kinetic parameters of the interaction. Thus, SPR-based binding and interaction studies were utilized for the study of (1) the interaction between selected SARS-CoV-2 proteins and putative host

cell-based interaction partners or receptors and (2) quantitative analysis of antibodies associated with SARS-CoV-2 in biofluids.

Very early studies on SARS-CoV-2 used SPR-based binding studies to quantify the kinetics of the interaction between the SARS-CoV-2 RBD and the host receptor ACE2.⁷⁵ ACE2 bound to the SARS-CoV-2 S ectodomain with an affinity of about 15 nM, and the authors highlighted that this equals a 10- to 20-fold higher affinity than the interaction of ACE2 with the SARS-CoV-1 S protein.⁷⁵

The interaction of the SARS-CoV-2 S protein with host cell-surface GAGs, which co-facilitate host cell entry, was determined by SPR-based binding assays. Liu *et al.* found that S protein monomers and trimers of SARS-CoV-2 bind to immobilized Hp with an equilibrium dissociation constant (K_D) in the nanomolar range (monomeric S protein K_D = 55 nM and trimeric S protein K_D = 64 nM), while a protein construct containing the RBD only interacted with the S protein of SARS-CoV-2 with a lower affinity (K_D = 1 μ M).²³ Furthermore, the interaction between Hp and the S protein did neither interfere with the S protein–ACE2 interaction nor with the proteolytic processing that the S protein undergoes. They tested an HS oligosaccharide library and determined in SPR-based competition assays that SARS-CoV-2 binds HS in a length-, sulfation-, and repetition-specific pattern with higher affinity than Hp. Their findings support the model of the presence of an additional HS/Hp-binding site outside the RBD that co-facilitates the interaction between the RBD of the S protein and the hACE2 through SPR techniques.²³

The neutralizing antibody CR3022 obtained from a convalescent SARS-CoV-1-infected patient was also reported to bind effectively to the SARS-CoV-2 S proteins.⁸⁷ SPR has also been used to study the binding interaction of various conformations of the SARS-CoV-2 S protein with the CR3022 antibody.⁸⁸

1.4.5 Biosensors

Biosensors, short for biological sensors, are devices composed of a transducer and a biological element, such as an antibody, enzyme, or nucleic acid. The biological response of the bioelement as it interacts with the analyte is converted into an electrical signal.^{89,90} A novel bioluminescence-based biosensor to probe the binding of the SARS-CoV-2 viral S protein to the hACE2 receptor based on nanoluciferase binary technology (Nano-BiT) was recently developed. This biosensor is composed of nanoluciferase, which is split into two complementary subunits: Large BiT and Small BiT. Each of these components is fused to the S1 domain of the SARS-CoV-2 S protein and the ACE2 ectodomain, respectively. Reassembly of functional nanoluciferase occurs during the ACE2-S1 interaction, which catalyzes a bioluminescent reaction that is detected in a highly sensitive manner. This bioreporter-based study was used to study the roles played by the

N-glycosylation on the RBD region of the SARS-CoV-2 S protein in viral entry, antigenicity, and immunogenicity.⁹¹ Abrogation of the *N*-glycosylation at the RBD region by PNGase F, Endo H, or Endo F enzymes resulted in a reduction in the binding of the RBD region with the ACE2 receptor.⁹¹

1.5 Concluding Remarks

Previous developments in glycoprotein analysis were vital to combating the SARS-CoV-2 global pandemic that emerged in late 2019. The increased transmissibility of SARS-CoV-2 compared to other human coronaviruses has led to over 200 million cases and over 6 million deaths worldwide.⁹² Viral glycosylation has been previously indicated as important in viral infectivity and transmission. Host receptor glycosylation and viral protein glycosylation can interact with each other to facilitate viral entry into the host. SARS-CoV-2 is composed of four structural proteins, three of which are glycosylated. The spike (S) protein is responsible for viral–host cell attachment and fusion. Through glycomic and glycoproteomic approaches, it was determined that the S protein is heavily glycosylated with both *N*- and *O*-glycans. It was also found that the human receptor hACE2 is glycosylated. Mutant S proteins lacking glycosylation at key sites for host binding were found to have a significant decrease in binding affinity, indicating the importance of glycosylation for viral fusion.

Herein, we have outlined and discussed the key factors of SARS-CoV-2 and its glycosylation, as well as the methods used to characterize its glycoproteins. Combining techniques, such as cryo-EM and mass spectrometry, has allowed scientists to determine structure–function relationships of the SARS-CoV-2 S protein. In addition, utilizing enzymatic techniques, including protease digestion, glycan removal, and ¹⁸O labeling, can give more comprehensive results compared to using one of these techniques alone. Thanks to established methods for glycoproteomic analysis, scientists were able to quickly sequence and map the SARS-CoV-2 S protein, which eventually led to the development of vaccines and therapeutics. More work is still needed to fully characterize SARS-CoV-2. Specifically, a comprehensive analysis of the wild-type S protein compared to the recently emerged variants of concern is needed to understand the increased transmissibility of these variants. In addition, more studies are required to map the *N*- and *O*-glycosylation of the E and M structural proteins to confirm the *in silico* results, which have been previously reported. One of the difficulties of mapping the glycosylation sites of these proteins is the difficulty in purifying them. Currently, only recombinant N and S proteins are available in the purified form, expressed in human cells, and E and M proteins are available in *E. coli* expression systems. Recent reviews by Bagdonaite *et al.*,⁸¹ Watanabe *et al.*,¹⁰ and Shajahan *et al.*²⁴ are recommended for further reading on viral glycosylation with a focus on glycoproteomics.

Acknowledgements

This work received funding from the US National Institutes of Health (R24GM137782, S10OD018530), The National Science Foundation Materials Innovation Platform funded through Cooperative Agreement DMR-1933525, and the US Department of Energy, Office of Science, Basic Energy Sciences (under Award DE-SC0015662 to DOE—Center for Plant and Microbial Complex Carbohydrates at the Complex Carbohydrate Research Center, USA).

References

1. R. G. Spiro, *Annu. Rev. Biochem.*, 1970, **39**, 599–638.
2. R. G. Spiro, *Glycobiology*, 2002, **12**, 43R–56R.
3. R. D. Cummings and J. M. Pierce, *Chem. Biol.*, 2014, **21**, 1–15.
4. P. Stanley, N. Taniguchi and M. Aebi, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, Cold Spring Harbor, NY, 2015, pp. 99–111.
5. I. Brockhausen and P. Stanley, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, Cold Spring Harbor, NY, 2015, pp. 113–123.
6. C. Steentoft, S. Y. Vakhrushev, H. J. Joshi, Y. Kong, M. B. Vester-Christensen, K. T. Schjoldager, K. Lavrsen, S. Dabelsteen, N. B. Pedersen, L. Marcos-Silva, R. Gupta, E. P. Bennett, U. Mandel, S. Brunak, H. H. Wandall, S. B. Levery and H. Clausen, *EMBO J.*, 2013, **32**, 1478–1488.
7. Y. Watanabe, Z. T. Berndsen, J. Raghvani, G. E. Seabright, J. D. Allen, O. G. Pybus, J. S. McLellan, I. A. Wilson, T. A. Bowden, A. B. Ward and M. Crispin, *Nat. Commun.*, 2020, **11**, 2688.
8. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, *Nat. Microbiol.*, 2020, **5**, 536–544.
9. C. M. Szymanski, R. L. Schnaar and M. Aebi, in *Essentials of Glycobiology*, ed. A. Varki, R. D. Cummings, J. D. Esko, P. Stanley, G. W. Hart, M. Aebi, A. G. Darvill, T. Kinoshita, N. H. Packer, J. H. Prestegard, R. L. Schnaar and P. H. Seeberger, Cold Spring Harbor, NY, 2015, pp. 527–538.
10. Y. Watanabe, T. A. Bowden, I. A. Wilson and M. Crispin, *Biochim. Biophys. Acta, Gen. Subj.*, 2019, **1863**, 1480–1497.
11. D. J. Vigerust and V. L. Shepherd, *Trends Microbiol.*, 2007, **15**, 211–218.
12. A. Varki, *Glycobiology*, 2017, **27**, 3–49.
13. J. M. Keller, P. G. Spear and B. Roizman, *Proc. Natl. Acad. Sci. U. S. A.*, 1970, **65**, 865–871.
14. G. A. Rabinovich, Y. van Kooyk and B. A. Cobb, *Ann. N. Y. Acad. Sci.*, 2012, **1253**, 1–15.

15. D. A. Mitchell, Q. Zhang, L. Voorhaar, D. M. Haddleton, S. Herath, A. S. Gleinich, H. S. Randeva, M. Crispin, H. Lehnert, R. Wallis, S. Patterson and C. R. Becer, *Chem. Sci.*, 2017, **8**, 6974–6980.
16. J. L. Van Etten, I. Agarkova, D. D. Dunigan, M. Tonetti, C. De Castro and G. A. Duncan, *Viruses*, 2017, **9**, 88.
17. F. Piacente, M. Gaglianone, M. E. Laugier and M. G. Tonetti, *Int. J. Mol. Sci.*, 2015, **16**, 29315–29328.
18. K. Nyström, A. Grahn, M. Lindh, M. Brytting, U. Mandel, G. Larson and S. Olofsson, *Glycobiology*, 2007, **17**, 355–366.
19. J. E. Jones, V. Le Sage and S. S. Lakdawala, *Nat. Rev. Microbiol.*, 2021, **19**, 272–282.
20. G. M. Air, *Curr. Opin. Virol.*, 2014, **7**, 128–133.
21. M. Matrosovich, A. Tuzikov, N. Bovin, A. Gambaryan, A. Klimov, M. R. Castrucci, I. Donatelli and Y. Kawaoka, *J. Virol.*, 2000, **74**, 8502–8512.
22. K. Shinya, M. Ebina, S. Yamada, M. Ono, N. Kasai and Y. Kawaoka, *Nature*, 2006, **440**, 435–436.
23. L. Liu, P. Chopra, X. Li, K. M. Bouwman, S. M. Tompkins, M. A. Wolfert, R. P. de Vries and G. J. Boons, *ACS Cent. Sci.*, 2021, **7**, 1009–1018.
24. A. Shajahan, L. E. Pepi, D. S. Rouhani, C. Heiss and P. Azadi, *Anal. Bioanal. Chem.*, 2021, **413**, 7179–7193.
25. H. Ge, X. Wang, X. Yuan, G. Xiao, C. Wang, T. Deng, Q. Yuan and X. Xiao, *Eur. J. Clin. Microbiol. Infect. Dis.*, 2020, **39**, 1011–1019.
26. A. A. Rabaan, S. H. Al-Ahmed, S. Haque, R. Sah, R. Tiwari, Y. S. Malik, K. Dhama, M. I. Yatoo, D. K. Bonilla-Aldana and A. J. Rodriguez-Morales, *Infez. Med.*, 2020, **28**, 174–184.
27. A. Shajahan, N. T. Supekar, A. S. Gleinich and P. Azadi, *Glycobiology*, 2020, **30**, 981–988.
28. J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach and F. Li, *Nature*, 2020, **581**, 221–224.
29. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan and M. Crispin, *Science*, 2020, **369**, 330–333.
30. Y. Zhang, W. Zhao, Y. Mao, Y. Chen, S. Wang, Y. Zhong, T. Su, M. Gong, D. Du and X. Lu, *Mol. Cell. Proteomics*, 2021, 100058.
31. J. D. Allen, Y. Watanabe, H. Chawla, M. L. Newby and M. Crispin, *J. Mol. Biol.*, 2021, **433**, 166762.
32. W. Chen, Z. Hui, X. Ren, Y. Luo, J. Shu, H. Yu and Z. Li, *bioRxiv*, 2020, DOI: 10.1101/2020.12.01.406025.
33. Y. Hou, C. Peng, M. Yu, Y. Li, Z. Han, F. Li, L.-F. Wang and Z. Shi, *Arch. Virol.*, 2010, **155**, 1563–1569.
34. R. Yan, Y. Zhang, Y. Li, L. Xia, Y. Guo and Q. Zhou, *Science*, 2020, **367**, 1444–1448.
35. P. Zhao, J. L. Praissman, O. C. Grant, Y. Cai, T. Xiao, K. E. Rosenbalm, K. Aoki, B. P. Kellman, R. Bridger, D. H. Barouch, M. A. Brindley, N. E. Lewis, M. Tiemeyer, B. Chen, R. J. Woods and L. Wells, *Cell Host Microbe*, 2020, **28**, 586–601.e586.

36. X. Zhao, D. Chen, R. Szabla, M. Zheng, G. Li, P. Du, S. Zheng, X. Li, C. Song, R. Li, J.-T. Guo, M. Junop, H. Zeng and H. Lin, *J. Virol.*, 2020, **94**, e00940-00920.
37. T. M. Clausen, D. R. Sandoval, C. B. Spliid, J. Pihl, H. R. Perrett, C. D. Painter, A. Narayanan, S. A. Majowicz, E. M. Kwong, R. N. McVicar, B. E. Thacker, C. A. Glass, Z. Yang, J. L. Torres, G. J. Golden, P. L. Bartels, R. N. Porell, A. F. Garretson, L. Laubach, J. Feldman, X. Yin, Y. Pu, B. M. Hauser, T. M. Caradonna, B. P. Kellman, C. Martino, P. L. S. M. Gordts, S. K. Chanda, A. G. Schmidt, K. Godula, S. L. Leibel, J. Jose, K. D. Corbett, A. B. Ward, A. F. Carlin and J. D. Esko, *Cell*, 2020, **183**, 1043–1057.e1015.
38. G. Duarte, M. J. García-Murria, B. Grau, J. M. Acosta-Cáceres, L. Martínez-Gil and I. Mingarro, *Open Biol.*, 2020, **10**, 200209.
39. S. Y. Kim, W. Jin, A. Sood, D. W. Montgomery, O. C. Grant, M. M. Fuster, L. Fu, J. S. Dordick, R. J. Woods and F. Zhang, *Antiviral Res.*, 2020, **181**, 104873.
40. J. Liu, J. Li, K. Arnold, R. Pawlinski and N. S. Key, *Res. Pract. Thromb. Haemostasis*, 2020, **4**, 518–523.
41. R. Tandon, J. S. Sharp, F. Zhang, V. H. Pomin, N. M. Ashpole, D. Mitra, M. G. McCandless, W. Jin, H. Liu and P. Sharma, *J. Virol.*, 2021, **95**, DOI: 10.1128/JVI.01987-20.
42. S. Belouzard, V. C. Chu and G. R. Whittaker, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5871–5876.
43. N. G. Seidah and A. Prat, *Nat. Rev. Drug Discovery*, 2012, **11**, 367–383.
44. E. Braun and D. Sauter, *Clin. Transl. Immunol.*, 2019, **8**, e1073.
45. G. Izaguirre, *Viruses*, 2019, **11**, 837.
46. B. Coutard, C. Valle, X. de Lamballerie, B. Canard, N. G. Seidah and E. Decroly, *Antiviral Res.*, 2020, **176**, 104742.
47. J. K. Millet and G. R. Whittaker, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 15214–15219.
48. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veasley, *Cell*, 2020, **181**, 281–292.e286.
49. J. Brun, S. Vasiljevic, B. Gangadharan, M. Hensen, A. V. Chandran, M. L. Hill, J. L. Kiappes, R. A. Dwek, D. S. Alonzi, W. B. Struwe and N. Zitzmann, *ACS Cent. Sci.*, 2021, **7**, 586–593.
50. M. Sanda, L. Morrison and R. Goldman, *Anal. Chem.*, 2021, **93**, 2003–2009.
51. W. Tian, D. Li, N. Zhang, G. Bai, K. Yuan, H. Xiao, F. Gao, Y. Chen, C. C. L. Wong and G. F. Gao, *Cell Res.*, 2021, **31**, 1123–1125.
52. A. Shajahan, S. Archer-Hartmann, N. T. Supekar, A. S. Gleinich, C. Heiss and P. Azadi, *Glycobiology*, 2021, **31**, 410–424.
53. K. K. Chan, D. Dorosky, P. Sharma, S. A. Abbasi, J. M. Dye, D. M. Kranz, A. S. Herbert and E. Procko, *Science*, 2020, **369**, 1261–1265.

54. K. Suryamohan, D. Diwanji, E. W. Stawiski, R. Gupta, S. Miersch, J. Liu, C. Chen, Y.-P. Jiang, F. A. Fellouse and J. F. Sathirapongsasuti, *Commun. Biol.*, 2021, **4**, 1–11.
55. C. A. Devaux, L. Pinault, I. O. Osman and D. Raoult, *Front. Public Health*, 2020, **8**, 608765.
56. P. S. Kwon, H. Oh, S.-J. Kwon, W. Jin, F. Zhang, K. Fraser, J. J. Hong, R. J. Linhardt and J. S. Dordick, *Cell Discovery*, 2020, **6**, 1–4.
57. L. Dai and G. F. Gao, *Nat. Rev. Immunol.*, 2021, **21**, 73–82.
58. A. S. Gleinich, L. E. Pepi, A. Shajahan, C. Heiss and P. Azadi, *Am. Pharm. Rev.*, 2021, **24**, 14–21.
59. B. Liu, Y. Yin, Y. Liu, T. Wang, P. Sun, Y. Ou, X. Gong, X. Hou, J. Zhang and H. Ren, *Engineering*, 2021, **13**, 107–115.
60. C. Gstöttner, T. Zhang, A. Resemann, S. Ruben, S. Pengelley, D. Suckau, T. Welsink, M. Wuhler and E. Domínguez-Vega, *Anal. Chem.*, 2021, **93**, 6839–6847.
61. A. A. Dawood, *Vacunas*, 2021, **22**, 1–9.
62. S. Thomas, *Pathog. Dis.*, 2020, **5**, 342.
63. N. T. Supekar, A. Shajahan, A. S. Gleinich, D. S. Rouhani, C. Heiss, D. G. Chapla, K. W. Moremen and P. Azadi, *Glycobiology*, 2021, **31**, 1080–1092.
64. I. M. Nilsson and G. von Heijne, *J. Biol. Chem.*, 1993, **268**, 5798–5801.
65. M. Orzáez, J. Salgado, A. Giménez-Giner, E. Pérez-Payá and I. Mingarro, *J. Mol. Biol.*, 2004, **335**, 631–640.
66. L. Feng and W. B. Frommer, *Trends Biochem. Sci.*, 2015, **40**, 480–486.
67. B. Jia, X. F. Zhu, Z. J. Pu, Y. X. Duan, L. J. Hao, J. Zhang, L.-Q. Chen, C. O. Jeon and Y. H. Xuan, *Front. Plant Sci.*, 2017, **8**, 2178.
68. World Health Organization, Tracking SARS-CoV-2 variants, World Health Organisation, <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/>, Accessed 30 Jul 2024.
69. N. L. Washington, K. Gangavarapu, M. Zeller, A. Bolze, E. T. Cirulli, K. M. Schiabor Barrett, B. B. Larsen, C. Anderson, S. White, T. Cassens, S. Jacobs, G. Levan, J. Nguyen, J. M. Ramirez, C. Rivera-Garcia, E. Sandoval, X. Wang, D. Wong, E. Spencer, R. Robles-Sikisaka, E. Kurzban, L. D. Hughes, X. Deng, C. Wang, V. Servellita, H. Valentine, P. De Hoff, P. Seaver, S. Sathe, K. Gietzen, B. Sickler, J. Antico, K. Hoon, J. Liu, A. Harding, O. Bakhtar, T. Basler, B. Austin, D. MacCannell, M. Isaksson, P. G. Febbo, D. Becker, M. Laurent, E. McDonald, G. W. Yeo, R. Knight, L. C. Laurent, E. de Feo, M. Worobey, C. Y. Chiu, M. A. Suchard, J. T. Lu, W. Lee and K. G. Andersen, *Cell*, 2021, **184**, 2587–2594.e2587.
70. C.-W. Kuo, T.-J. Yang, Y.-C. Chien, P.-Y. Yu, S.-T. D. Hsu and K.-H. Khoo, *Glycobiology*, 2022, **32**, 60–72.
71. E. Callaway, *Nature*, 2020, **578**, 201.
72. J. Pallesen, N. Wang, K. S. Corbett, D. Wrapp, R. N. Kirchdoerfer, H. L. Turner, C. A. Cottrell, M. M. Becker, L. Wang, W. Shi, W. P. Kong, E. L. Andres, A. N. Kettenbach, M. R. Denison, J. D. Chappell, B. S. Graham,

- A. B. Ward and J. S. McLellan, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E7348–E7357.
73. C. L. Hsieh, J. A. Goldsmith, J. M. Schaub, A. M. DiVenere, H. C. Kuo, K. Javanmardi, K. C. Le, D. Wrapp, A. G. Lee, Y. Liu, C. W. Chou, P. O. Byrne, C. K. Hjorth, N. V. Johnson, J. Ludes-Meyers, A. W. Nguyen, J. Park, N. Wang, D. Amengor, J. J. Lavinder, G. C. Ippolito, J. A. Maynard, I. J. Finkelstein and J. S. McLellan, *Science*, 2020, **369**, 1501–1505.
74. A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire and D. Veasley, *Cell*, 2020, **181**, 281–292 e286.
75. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham and J. S. McLellan, *Science*, 2020, **367**, 1260–1263.
76. X. Zhu, D. Mannar, S. S. Srivastava, A. M. Berezuk, J. P. Demers, J. W. Saville, K. Leopold, W. Li, D. S. Dimitrov, K. S. Tuttle, S. Zhou, S. Chittori and S. Subramaniam, *PLoS Biol.*, 2021, **19**, e3001237.
77. A. Antonopoulos, S. Broome, V. Sharov, C. Ziegenfuss, R. L. Easton, M. Panico, A. Dell, H. R. Morris and S. M. Haslam, *Glycobiology*, 2021, **31**, 181–187.
78. Y. Zhang, W. Zhao, Y. Mao, Y. Chen, S. Wang, Y. Zhong, T. Su, M. Gong, D. Du, X. Lu, J. Cheng and H. Yang, *Mol. Cell. Proteomics*, 2020, 100058.
79. Argentinian AntiCovid consortium, *Sci. Rep.*, 2020, **10**, 21779.
80. Y. Wang, Z. Wu, W. Hu, P. Hao and S. Yang, *ACS Omega*, 2021, **6**, 15988–15999.
81. I. Bagdonaite, A. J. Thompson, X. Wang, M. Sogaard, C. Fougereux, M. Frank, J. K. Diedrich, J. R. Yates, A. Salanti, S. Y. Vakhrushev, J. C. Paulson and H. H. Wandall, *Viruses*, 2021, **13**, 551.
82. D. Wang, J. Baudys, J. L. Bundy, M. Solano, T. Keppel and J. R. Barr, *Anal. Chem.*, 2020, **92**, 14730–14739.
83. L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda and R. E. Amaro, *ACS Cent. Sci.*, 2020, **6**, 1722–1734.
84. T. Sztain, S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, E. Seitz, R. S. McCool, F. L. Kearns, F. Acosta-Reyes, S. Maji, G. Mashayekhi, J. A. McCammon, A. Ourmazd, J. Frank, J. S. McLellan, L. T. Chong and R. E. Amaro, *Nat. Chem.*, 2021, **13**, 963–968.
85. M. N. Alhajj, M. Zubair and A. Farhana, Enzyme linked immunosorbent assay, *Stat. Pearls*, 2023.
86. S. Sakamoto, W. Putalun, S. Vimolmangkang, W. Phoolcharoen, Y. Shoyama, H. Tanaka and S. Morimoto, *J. Nat. Med.*, 2018, **72**, 32–42.
87. K. Rattanapisit, B. Shanmugaraj, S. Manopwisedjaroen, P. B. Purwono, K. Siri Wattananon, N. Khorattanakulchai, O. Hanittinan, W. Boonyayothin, A. Thitithanyanont, D. R. Smith and W. Phoolcharoen, *Sci. Rep.*, 2020, **10**, 17698.
88. R. Henderson, R. J. Edwards, K. Mansouri, K. Janowska, V. Stalls, S. M. C. Gobeil, M. Kopp, D. Li, R. Parks, A. L. Hsu, M. J. Borgnia, B. F. Haynes and P. Acharya, *Nat. Struct. Mol. Biol.*, 2020, **27**, 925–933.

89. A. Turner, I. Karube and G. S. Wilson, in *Biosensors: Fundamentals and applications*, Oxford university press, 1987.
90. A. P. Turner, *Chem. Soc. Rev.*, 2013, **42**, 3184–3196.
91. T. Azad, R. Singaravelu, Z. Taha, T. R. Jamieson, S. Boulton, M. J. F. Crupi, N. T. Martin, E. E. F. Brown, J. Poutou, M. Ghahremani, A. Pelin, K. Nouri, R. Rezaei, C. B. Marshall, M. Enomoto, R. Arulanandam, N. Alluqmani, R. Samson, A. C. Gingras, D. W. Cameron, P. A. Greer, C. S. Ilkow, J. S. Diallo and J. C. Bell, *Mol. Ther.*, 2021, **29**, 1984–2000.
92. World Health Organization, Coronavirus disease (COVID-19), [https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-\(covid-19\)](https://www.who.int/news-room/fact-sheets/detail/coronavirus-disease-(covid-19)), [Accessed 30 Jul 2024].