Proteomics
Proteomics and Systems Biology

REVIEW

Site-specific glycosylation of SARS-CoV-2: Big challenges in mass spectrometry analysis

Diana Campos[1] | **Michael Girgis[2]** | **Miloslav Sanda[1,3]** (ID)

[1]Max-Planck-Institut fuer Herz- und Lungenforschung, Bad Nauheim, Germany

[2]Department of Bioengineering, Volgenau School of Engineering and Computing, George Mason University, Fairfax, Virginia, USA

[3]Clinical and Translational Glycoscience Research Center, Georgetown University, Washington, DC, USA

Correspondence
Miloslav Sanda, Max-Planck-Institut fuer Herz- und Lungenforschung, Ludwigstrasse 43, Bad Nauheim, 61231, Germany.
Email: Miloslav.Sanda@mpi-bn.mpg.de

Abstract

Glycosylation of viral proteins is required for the progeny formation and infectivity of virtually all viruses. It is increasingly clear that distinct glycans also play pivotal roles in the virus's ability to shield and evade the host's immune system. Recently, there has been a great advancement in structural identification and quantitation of viral glycosylation, especially spike proteins. Given the ongoing pandemic and the high demand for structure analysis of SARS-CoV-2 densely glycosylated spike protein, mass spectrometry methodologies have been employed to accurately determine glycosylation patterns. There are still many challenges in the determination of site-specific glycosylation of SARS-CoV-2 viral spike protein. This is compounded by some conflicting results regarding glycan site occupancy and glycan structural characterization. These are probably due to differences in the expression systems, form of expressed spike glycoprotein, MS methodologies, and analysis software. In this review, we recap the glycosylation of spike protein and compare among various studies. Also, we describe the most recent advancements in glycosylation analysis in greater detail and we explain some misinterpretation of previously observed data in recent publications. Our study provides a comprehensive view of the spike protein glycosylation and highlights the importance of consistent glycosylation determination.

KEYWORDS
glycoproteomics, N-glycosylation, O-glycosylation, SARS-CoV-2 glycoprotein

1 | INTRODUCTION

Viruses can be classified into two groups depending on whether they have a lipid bilayer membrane on their outer surface or not: enveloped viruses and nonenveloped viruses. A characteristic feature of enveloped viruses is a cell membrane-derived envelope modified with virally encoded proteins [1]. All attachment and fusion proteins of the enveloped viruses are usually modified by glycosylation and these glycosylated surface epitopes are key in the pathogen-host interplay.

The interactions between viral proteins and host cell proteome play a crucial role in the infection process [2, 3]. Viral envelope proteins, like cellular proteins, possess signal peptides directing them to the secretory pathway, and can be decorated with different kinds of post-translational modifications (PTMs). Viral protein PTMs, as well as host cell proteome, are the determining factors in the level of these interactions and the extent of host immune response [4, 5].

Glycosylation is one of the most important types of PTMs that could potentially impact protein structure, orientation, binding affinity, and metabolism [6]. It involves the covalent attachment of different types of glycans to specific sites on protein structures. Viral envelope proteins are often ornamented by glycans that can account for up to half of the molecular weight of these glycoproteins [7]. Despite the numerous types of glycosylation, N- and mucin-type O-linked glycosylation are the most widely exploited in viral research [8]. Prominent examples include the heavily glycosylated gp120 glycoprotein in HIV, the Ebola virus glycoprotein modified by a very high glycan content, and the HIV-1 glycoprotein gp160 that is glycosylated by the addition of multiple N-linked glycans [7, 9-11].

The N-linked glycosylation takes place on the nitrogen atom of the asparagine amide side chain, where this asparagine (N) residue is located within the sequence -N-X-T/S- (where X ≠ proline) [12–14]. The O-linked glycosylation is usually meant to attach the sugar moiety to the oxygen of the side chains of serine and/or threonine residues within the amino acid sequence [15–17]. Both types of site-specific glycosylation have been shown to affect viral glycoprotein secretion and function [8].

As the virus hijacks the host cellular machinery for replication, it subsequently uses the host glycosylation capabilities for the production of viral proteins. Thus, the viral surface antigens may encompass familiar host glycans, which can thereby change the ability of the host to recognize the virus and stimulate the immune response [18–20].

In addition, certain viruses are able to induce changes in the expression levels of host glycosyltransferases (enzymes that catalyze the glycosidic bond). For example, herpes viruses can induce the expression of host fucosyltransferases leading to the expression of sLex or Ley antigens [21, 22]. Another example is the shift of the total glycoprofile of HCV-infected hepatoma cells toward more fucosylated, sialylated, and a more complex N-glycan structures [23]. Furthermore, the changes in expression levels of host glycosylation enzymes may indicate occurrence of specific structural or functional modifications or highlight the demand for increased glycosylation capacity to support viral protein glycosylation.

Mass spectrometry (MS) techniques have been abundantly implemented for proteomic and glycoproteomic analysis to decipher pathogenic structures and explore mechanistic immune responses of the host [24–33].

In this review, we examined the most recent advancements in structural identification and quantitation of viral glycosylation. Given the ongoing pandemic and the high demand for protein and structure analysis, we mainly focused on an in-depth analysis of SARS-CoV-2. Also, we highlighted the most intricate challenges in the determination of site-specific glycosylation associated with viral spike protein.

2 | VIRAL GLYCOPROTEINS AND THEIR FUNCTIONS

Glycosylation of viral proteins is required for the progeny formation and infectivity of many viruses [31, 34]. Furthermore, distinct glycans play pivotal roles in various stages of the viral cycle [34, 35]. For example, glycans on viral entry proteins are greatly involved in the modulation of receptor binding and entry [36]. A prominent example is the influenza viruses that attach to glycans on cellular surface glycoproteins. Hemagglutinin (HA) and neuraminidase (NA) are the surface glycoproteins of influenza viruses, which interact with the terminal sialic acid (SA) of the host cell surface glycoproteins. NA can cleave the SA residues of host mucin to gain access to the epithelial cells, playing a secondary role in helping viruses to enter host cells [37]. On the other hand, NA can also cleave SA residues from glycoproteins of the enveloped virus itself and enhance infectivity by preventing aggregation of viral particles [38]. Host-cell dependent glycosylation of HA and NA has also proven to be critical. Glycosylation in the HA stalk region is important for protein folding and trafficking, and for pH stability [39, 40], while the extent of glycosylation near the HA receptor-binding site alters its affinity for SA-containing receptors [41, 42]. Glycosylation near the cleavage of HA site also modulates virus pathogenicity [43]. While the role of glycosylation of NA is less well understood, N-linked glycosylation is important for functional NA, and lack of NA glycosylation increases neurovirulence of the A/WSN/33 IAV strain in mice [44]. Besides, HA N-glycosylation affects T cell activation and cytokine production and thus promote immune evasion [45]. This and other types of immune evasion have been reported for different glycoproteins in different viruses. In fact, it is strongly believed that the high levels of glycosylation serve primarily as a protective shield against the host's immune system [46, 47].

While the innate immune system is constantly evolving a range of strategies to combat glycosylated epitopes of serious pathogens, mutations could lead to failures in the immune reactions. Alterations on viral glycoproteins can significantly impact viral characteristics, including the extent of protein glycosylation, which may jeopardize efficacy of existing vaccines [45, 48]. Moreover, antigen glycosylation complicates the development of vaccines and antibody-based therapies.

Glycans represent structural features that are not encoded in the gene sequence and yet play a crucial role in immune recognition affecting vaccine designs. Such vaccine candidates are often expressed in cell lines that do not recapitulate the glycosylation pattern on native pathogens, and potentially do not elicit biologically relevant immune responses. Therefore, it is important to identify common glycosylation patterns for translational applications. This has been particularly important in the study of SARS-CoV-2 spike (S) protein.

2.1 | SARS-CoV-2 spike protein glycosylation

SARS-CoV-2 initiates the infection cycle by binding to Angiotensin-Converting Enzyme II (ACE2) on the host epithelial cells in the respiratory tract. This leads to viral penetration, domination of the host cell's biological machinery, multiplication, and maturation of the virus [49–52]. Figure 1 demonstrates SARS-CoV-2 as represented by Ganji
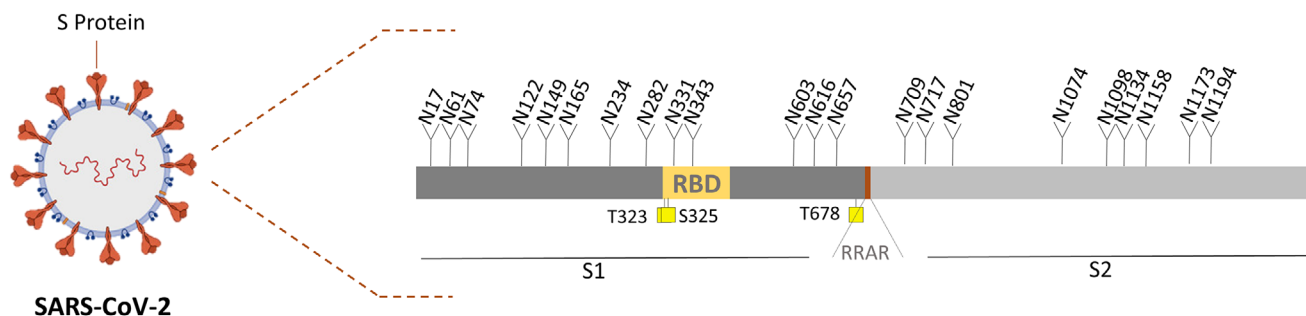
FIGURE 1 Anatomy of the SARS-CoV-2 particle showing its structural proteins (adapted from Ganji et al., [53]) with emphasis on the SARS-CoV-2 Spike (S) protein with its illustrative protein sequence components: S1 and S2 subunits, and Receptor binding domain (RBD). Also, the most commonly observed N- and O- glycosylation sites on S protein are depicted

et al., [53]. There are four structural proteins encoded by SARS-CoV-2 genome: the spike (S) glycoprotein, the membrane (M) protein, the envelope (E) protein, and the nucleocapsid (N) protein.

Like many viruses, SARS-CoV-2 launches its cellular invasion through its heavily glycosylated S protein [54–58]. Despite the relatively modest contribution of the glycans to the total molecular weight of the S trimer of the spike protein, they shield approximately 40% of the protein surface. The three-dimensional structure of the S protein shows that the protein surface is extensively shielded from antibody recognition by glycans, with the notable exception of the ACE2 receptor binding domain [59]. The S protein encompasses two protein subunits, S1 and S2. While the S1 subunit is the ACE2 receptor binding domain and modulates receptor recognition, S2 subunit orchestrates cellular adhesion and membrane fusion [60]. The viral S protein of SARS-Cov-2 has also been the ultimate target for vaccine production [61]. In fact, spike proteins are often used as immunogens for vaccines to generate neutralizing antibodies and are frequently targeted for inhibition by small molecules that might block host receptor binding and/or membrane fusion.

Glycosylation is a heterogenic process that depends on many factors, including age, underlying disease and ethnicity; therefore, assessing the glycosylation profile may be correlated to the observed differential susceptibilities among individuals to COVID-19 [62–65]. While overall shielding of the underlying protein surface does not appear to be highly sensitive to glycan microheterogeneity, it could impact the innate immune response by altering the ability of collectins and other lectins of the immune system to effectively bind to the S glycoprotein and neutralize the virus. Also, it may impact the adaptive immune response by altering the number of viable human leucocyte antigen (HLA) [66].

With reference to the sequence details of the S protein, so far, 22 highly occupied N-linked glycosylation sites have been identified, as well as a variable number of O-linked glycosylation sites at low stoichiometries [56, 67-70]. These SARS-CoV-2 S N-glycosites were reported occupied in most studies in recent glycoproteomic analyses [67–71].

A summary of the identified O-glycan sites is displayed in Table 1, with emphasis for the sites found on two critical areas of the genome: furin cleavage site and Receptor Binding Domain (RBD). SARS-CoV-2 has a polybasic cleavage site (RRAR) at the junction of S1 and S2, the two subunits of the spike. Three O-glycan sites (S673, T678 and S686) are predicted to flank the cleavage site [72]. The T678 occupancy was firstly identified by *Sanda* et al., in a recent publication [68]. The RBD in the spike protein is the most variable part of the coronavirus genome. O-glycan sites were also consistently found at RBD sequence, namely T323 and S325. It's also important to note that 16 out of 25 O-glycosites were located within three amino acids from the known N-glycosites. However, O-glycosylation was primarily found on peptides that were unoccupied by N-glycans. This suggests that although O-glycans constitute a minor component of the S protein, they may ensure maximum shielding of the minor fraction of peptides that are unoccupied by N-glycans [73]. In the future, it would be relevant to map O-glycosites on native viruses derived from specified respiratory cell subtypes.

2.2 | Strategies used for the analysis of viral spike glycoprotein

Various strategies have been recruited to analyze and confirm the nature, the structure and exact location of glycans on viral spike proteins. These strategies involve the sample preparation, the chromatographic separation, mass spectrometry techniques, and bioinformatics analysis.

The sample preparation usually involved the use of glycan releasing enzymes (such as PNGase F for N-glycans) as well as proteases (such as trypsin) with or without enrichment depending on the glycopeptide stoichiometry. Hydrophilic interaction liquid chromatography (HILIC), lectin affinity chromatography, and graphitized carbon chromatography were widely adopted for enrichment of glycopeptides. Figure 2 demonstrates the workflow for the sample preparation and data acquisition.

To simplify glycopeptide complex mixture, chromatographic separation can be implemented on reversed-phase C18 column, Hydrophilic interaction liquid chromatography (HILIC) [74, 75], and Porous graphitic carbon (PGC) columns [76, 77].

The separated glycans or glycopeptides are then analyzed by tandem MS/MS to determine glycosite location and glycan structure of

TABLE 1 A Summary of site-specific O-glycosylation analysis of SARS-Cov-2 Spike glycoproteins

| Reference[a]	Source of S protein	Expression system	Total sites	RBD sites	Furin cleavage sites*	Notes
Shajahan A. et al., 2020 [69]	S1+S2 (separately)	Hek 293	2	T323, S325	–	in gel digestion
Watanabe Y. et al., 2020 [70]	S protein 2P GSAS (682- 685)	Hek 293	2	T323, S325	–	
Sanda M. et al., 2021 [68]	S protein (R683A, R685A)	Hek 293	9	T323, S325	T678	
Gao C. et al., 2020 [92]	S protein (R683A, R685A)	Hek 293	5	T323	T678, S686	
Zhao P. et al., 2020 [58]	S protein 2P GGSG (682- 685)	Hek 293	27	T323, S325, S359, S366, S371, S373, T393, S399, S494, T547, T553, S555	–	O-protease OpeRATOR
Bagdonaite I. et al., 2021 [73]	S protein 2P AARA (682- 685)	Hek 293	12	T478	(T676), T678	in gel digestion and in solution workflow
Tian Y. et al., 2021 [96]	S1 subunity	Hek 293	14	T323, T523	T678	
Zhang Y. et al., 2021 [95]	S1 subunity	Hek 293	30	T323, S325	(T676), T678, (S680)	
Bagdonaite I. et al., 2021 [73]	S protein 2P AARA (682- 685)	Insect cells	15	T478	T678	in gel digestion
Bagdonaite I. et al., 2021[73]	Soluble RBD	Insect cells	6	T323, T333, T345, T415, T523		in gel digestion
Zhang Y. et al., 2021[95]	S protein	Insect cells	43	T323, S325, T333, S345, S477	(T676), T678	
Tian W. et al., 2021[101]	SARS-CoV-2 virions	Vero cells	17	T323		
Brun J. et al., 2021 [97]	SARS-CoV-2 virions (S1 subunity)	Calu-3	1	–	T678	

[a]Literature references are indicated as first author last name and year
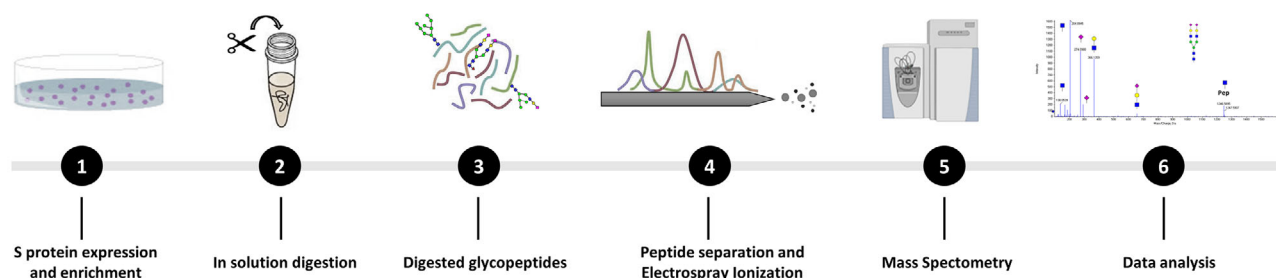
FIGURE 2 Schematic depiction of site-specific bottom up glycoproteomics workflow used to characterize SARS-CoV-2 S protein glycosylation. Different approaches to the workflow and other protocols used are described further in Section 2.2

both N- and O- glycosylation. Collision-based dissociation, such as collision induced dissociation (CID) and higher-energy C-Trap Dissociation (HCD) are the most commonly used fragmentation techniques in mass spectrometry. Electron-based dissociation, such as Electron Capture dissociation (ECD) and Electron Transfer Dissociation (ETD) can yield a gentle dissociation of the peptide backbone without neutral loss of the N-glycan moiety. Ultraviolet photodissociation (UVPD) which combines the features of both collision techniques. Moreover, combined EThcD and ETciD on Orbitrap mass spectrometers have also been implemented for a selective backbone fragmentation. Two

MS acquisitions have been adopted for this type of analysis: data-dependent acquisition (DDA) and data-independent acquisition (DIA). Characterization glycan composition, intact precursors or fragment ions, is done by a variety of software such as Byonic, pleco, GPQuest, GPSeeker, O-pair Search in MetaMorpheus, MSFragger-Glyco, and StrucGP.

In the following sections, we present the most intricate challenges in the assignments of site -specific glycosylations associated with SARS-CoV-2S spike protein. While the majority of site-specific N-glycosylation analysis were performed using the bottom-up
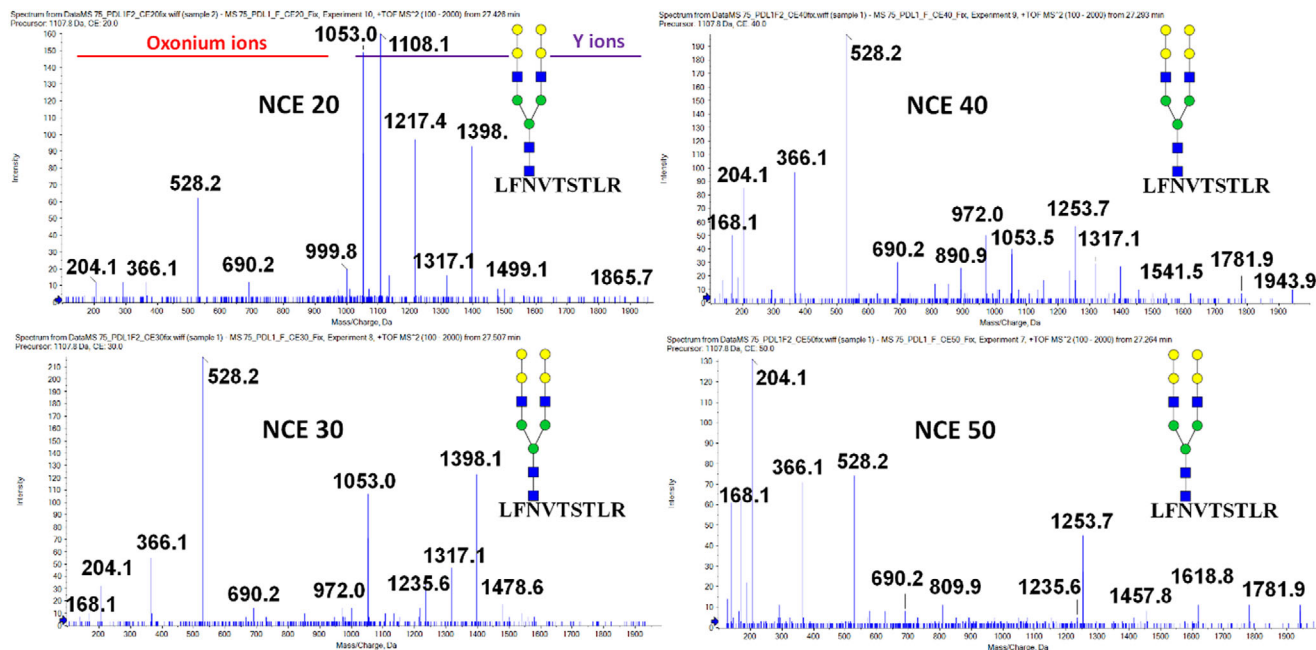
FIGURE 3 Comparison of HCD tandem mass spectra of the PD-L1 glycopeptide LFNVTSTLR occupied by a biantennary galili recorded at four different collision energies (NCE: 20, 30, 40, 50) (Sanda et al., [83])

approach, very few reports utilized the top-down approach in their O-glycosylation analysis [60, 78-80].

2.2.1 | Early reports of SARS-Cov-2 spike glycoprotein analysis

SARS COV-2 is considered the most frequently studied viral glycoprotein by mass spectrometry in the recent few years. Roughly, about three dozen of publications were focused on the analysis of N- and O-glycosylation of spike glycoprotein. In the first published report, Watanabe et al., analyzed N-glycosylation and partially O-glycosylation of HEK 293 overexpressed protein and utilized high-resolution mass spectrometry combined with HCD fragmentation [70]. While the glycopeptide tandem mass spectra of later reports were acquired using lower energy (NCE = 20 to 35) or stepped collision energy, the tandem MS data in first report were acquired using HCD set at 50%. Under NCE 50%, glycopeptides were extensively fragmented which results in very low abundances of peptide + Yn ions. Missing peptide + Yn ions limited glycan characterization with only information of intact mass and oxonium ions, which makes difficult to identify core structure [81]. In addition, Oxonium ions are fragmented to single units which reduce chance to correctly identify outer arm specific structures. NCE glycopeptide tandem mass spectra changes are visible in Figure 3 [82, 83]. To observe more complex information, it is possible to acquire data using 2 or 3 different collision energies while low collision energy is used for glycan structure elucidation and high collision energy for glycopeptide identification. Watanabe et al., described 22 occupied N-glycosites across the whole spike glycoprotein and two occupied O-glycosites located on RBD domain. In addition, they constructed

the three-dimensional structure based on the mapping of SARS-CoV-2 N-linked glycans. Many groups including our laboratory described glycosylation analysis of SARS-CoV-2 spike glycoprotein following this work [68].

2.2.2 | Glycosylation analysis using mass spectrometers with various instruments and different resolving power

SARS-CoV-2 S protein glycosylation was predominantly analyzed only by high-resolution mass spectrometers. In most cases, it was an orbitrap based analyzer working with a resolution of 60,000 to 120,000 or a Time Of Flight (TOF) based analyzer which was used in very few cases of the top-down approach [78, 84]. The use of high and ultrahigh resolving power allowed scientists to effectively analyze 22 glycosylation sites in the mixture of digested peptides and glycopeptides. Targeted multiple-reaction monitoring (MRM) and parallel reaction monitoring (PRM) methodology were used for detection and quantification of spike protein fragments in biological fluids such as saliva or extracted nasal swabs [85, 86]. Although these methodologies were only reported for nonglycosylated peptide analysis, we envision these will be implemented for spike protein glycosylation analysis of SARS-CoV-2 S protein in the near future.

2.2.3 | Data processing software

In most of the studies focused on SARS-CoV-2 S protein glycosylation, researchers used Byonic software (Protein Metrics, USA). Some of the

Reference	Source of S protein	expression system	digestion
Shajahan A. et al, 2020 [69]	S1+S2 (separately)	Hek 293	Try, Chy, Try+ Chy
Watanabe Y. et al, 2020 [70]	S protein 2P GSAS (682- 685)	Hek 293	Try, Chy, aLP
Wang D. et al., 2020 [93]	S protein 2P	Hek 293	Try, LysC+Chy, Asp-N+Chy
Sanda M. et al, 2020 [68]	S protein (R683A, R685A)	Hek 293	Try; Glu-C
Zhao P. et al, 2020 [58]	S protein 2P GGSG (682- 685)	Hek 293	Chy, aLP, Try+Gluc-C, Glu-C+AspN
Zhang S. et al, 2022 [94]	wild type S protein	Hek 293	Try, Chy
Wang D. et al., 2020 [93]	S1 subunity	Hek 293	Try, LysC+Chy, Asp-N+Chy
Tian Y. et al, 2021 [96]	S1 subunity	Hek 293	Try, Chy
Zhang Y. et al, 2021 [95]	S1 subunity	Hek 293	Try, Glu-C, Try+Glu-C
Brun J. et al, 2021 [97]	SARS-CoV-2 virions (S1 subunity)	Calu-3	Try, Chy, aLP
Tian Y. et al, 2021 [96]	SARS-CoV-2 virions	Vero cells	Try+Glu-C, Chy
Tian Y. et al, 2021 [96]	SARS-CoV-2 virions (D614G variant)	Vero cells	Try+Glu-C, Chy
Yao H. et al, 2020 [99]	SARS-CoV-2 virions	Vero cells	Try, Chy, aLP
Wang D. et al., 2020 [93]	S protein 2P	Insect cells	Try, LysC+Chy, Asp-N+Chy
Bangaru S. et al, 2020 [67]	S protein 2P QQAQ (682- 685)	Insect cells	Try+Chy, Chy, Try/elastase/Arg-C,
Zhang Y. et al, 2021 [95]	S protein 2P	Insect cells	Try, Glu-C, Try+Glu-C

(N-glycan site columns — S1: 17, 61, 74, 122, 149, 165, 234, 282, 331, 343, 603, 616, 657, 709, 717; S2: 801, 1074, 1098, 1134, 1158, 1173, 1194 — are displayed as a color-coded heatmap.)

FIGURE 4 Summary of site-specific N-glycosylation analysis of SARS-Cov-2. The N-glycan sites and glycan composition at each site are compared among different publications. The compositional analysis of the glycans is shown for each site displaying only the most abundant of each of the three types of N-glycans: unoccupied sites (gray), high-mannose (green), hybrid (orange), complex (blue), and paucimannose (yellow). When compositional analysis shows equal abundancy of 2 types of N-glycans, the two correspondent colors are displayed at the same site. Different sources of the S protein, expression systems, and proteases used for sample digestion are also shown for comparison

reports used pGlyco [58], and a few researchers used proprietary software. Interestingly, several reports classified glycans into numerous subgroups such us complex/hybrid and oligomannose without glycan structure-based separation using chromatographic techniques [87] or ion mobility [83].

Classifying glycans only based on the glycan composition utilizing precursor ions only could lead to misinterpretations of the glycoproteomics data. Similar problem could happen with the assumption that most of the glycans are composed of a single structure. For example, the core fucosylated glycan reported by several groups [58, 69, 70] without the contribution of the outer arm fucosylated form, might be an inaccurate interpretation of the fact. Incorporation of an orthogonal chromatography separation or ion mobility technique may be the ultimate solution to this issue. Several studies reported the occurrence of both forms in biological samples [88].

Zhao et al., used glycomics informed glycoproteomic analysis, which defines glycan space (used glycan data base) by glycomic analysis and could minimize glycoproteomics data misinterpretation. Furthermore, they disclosed the glycan informed glycoproteomics analysis protocol, where they determined searching glycan space by separated analysis of detached glycans using Peptide N-glycosidase (PNGase F) [58]. Also, Hackett et al., described software and glycospace dependent ID results using four different types of glycoproteomic software with default settings and fixed settings, even with fixed setting such as glycospace, [89] and so on. In addition, this methodology can help correct glycan assignment, but it is heavily dependent on the methodology used for glycan analysis. However, the results were different between studies (Figure 4).

2.2.4 | Levels of glycosylation (Glycan, glycopeptides, and glycoprotein)

Most of the studies on SARS-CoV-2 spike protein glycosylation were performed on the glycopeptide level. The advantage of glycopeptide analysis is its ability to link the sugar moiety to an exact glycosylation site, which is important in the case of analyzing spike glycoprotein and receptor binding domain as well as the development of effective vaccines. On the other hand, the ability of bottom up glycoproteomics to assign specific isomeric structure to a particular glycosite is very limited. Cho et al., used isomeric separation to compare detached glycans from SARS-CoV-2, SARS-CoV and Middle East respiratory Syndrome (MERS) spike glycoproteins [87]. Meanwhile, Gao et al., introduced various MALDI-MS/MS methods to analyze detached glycans. They showed a high level of LacdiNAc containing structures which were described previously in HEK 293 cell overexpressed glycoproteins [90–92]. Sanda et al., previously utilized ion mobility for separation of biantennary galactosylated complex glycan and LacdiNAc contains hybrid glycans isobaric glycopeptides [83], overexpressed PDL1 in the HEK 293 cells [91].

The approaches used for the analysis of O-glycosylations are more diverse, such as bottom up (Figure 4) or top down proteomics [78]. The rest of N-glycosylation analysis was performed using bottom up glycoproteomics strategy. Figure 4 demonstrates the various bottom-up methodologies used for analysis of N-glycosylation in all cases. The results were consistent despite the minor variation in a few aspects like using different proteolytic enzyme combination, multiple fragmentation settings, different processing softwares, using ion

mobility for separation of glycan isomers [68, 78] or using ETD fragmentation for occupied glycosite assignment [68, 92]. Top-down approach was exclusively used in the analysis of O-glycosylation of RBD domain since it has only two potential glycosites and allows effective TOP-down analysis of O-glycosylation.

2.2.5 | N-glycosylation depends on different source of material

The N-glycosylation of SARS-CoV-2 S proteins has been investigated extensively in the literature. Several studies used recombinant S proteins produced in different expression systems [58, 68-70, 93-95]. This is mostly due to the difficulties inherent to the study of the wild type viral proteins. Human embryonic kidney cell line HEK 293 was the most frequently used expression system to produce the recombinant S protein (Figure 4). Some studies compared the recombinant S protein expressed in HEK cells with protein overexpressed in insect cells [93, 95]. As expected, N-glycosylation from different expression systems yielded variable results. Figure 4 describes the different expression systems, source of protein, used proteases and references for the various N-glycosylation analysis. The blue color indicates a complex glycan structure, green refers to a high mannose glycan, orange to hybrid glycans, and yellow to paucimannose glycans.

A total of 3 research groups focused on analysis of S1 subunit encompassing the RBD domain which was overexpressed in HEK 293 cell line [93, 95, 96], while a single study used S1 subunit isolated from virion cultivated on Calu-3 cells [97]. The three studies of S1 subunit overexpressed in HEK 293 cells showed only complex glycans as a major structural component on all sites. On the other hand, Zhang et al., reported N603 mostly occupied by high mannose glycan [95].

Interestingly, a total of 11 out of 16 studies demonstrated that N234 was occupied by oligomannose moiety, while the remaining groups concluded that N234 is decorated with a complex glycan structure which has a direct impact on receptor binding [93–96]. To note that three of the studies reporting complex glycan structures at N234 used only the S1 subunit expressed in HEK 293 cells, instead of the full size (S1+S2 subunits) spike protein construct. In addition, Brun et al., showed that almost one third of the glycosites (including N234) is occupied by oligomannose structures in the S1 subunit, isolated from virion grown on Calu-3 cells. Three research groups used insect-cell lines for the production of SARS-CoV-2 spike glycoprotein [67, 93, 95]. It is known that insect cells produce predominantly short version of N-gycans also known as paucimannose glycans. These glycan type was identified on almost all N-glycosylation sites on these studies. Wang et al., and Bangarus et al., displayed only a few sites (including N234) occupied by oligomannose glycans [67, 93].

Zhang et al., used insect cells to overexpress their spike glycoprotein, but they identified mostly oligomannose structures in all sites. Similar N-glycosylation pattern was detected using the same overexpression system in the case of other proteins such as prostate-specific membrane protein (PSAM) [98]. Following studies described analysis of

spike glycoprotein isolated from virions [96, 97, 99]. The glycosylation profiles in vero cells showed more oligomannose structures compared to of HEK 293 cells overexpressed protein (including N234 site) with Yao et al., being an exception. Yao et al., demonstrated complex glycan structures on all sites and except three positions (N234 and N607 which were partially occupied by oligomannose and complex glycans and 122 which was occupied predominantly by paucimannose glycans) [99]. One of interesting issues is the identification of unoccupied sites. Two reports have shown unoccupied sites. Shajahan et al., reported four unoccupied N-glycosites (N17, N603, N1134, N1158, and N1173) and Sanda et al., reported one unoccupied site N603. These sites were clearly identified as unoccupied sites and, in some cases, confirmed by PNGaseF deglycosylation. Zhang et al., showed site N1173 with very low occupancy. The other 13 reports showed all sites occupied or nonidentified, but did not reported unoccupied sites. The number of identified sites vary depending on the protein source and the digestion system. In our case, work was mainly focused on the analysis of O-glycosylation near to the furin cleavage site and digestion system was chosen for the cleavage of N657 glycosite and predicted O-glycosites S673 and T678, to reduce potential false-positive O-glycan identification due to incomplete N deglycosylation. Using Try/GluC combination, we were not able to identify site N17 (very short peptide) and sites N709 and N717 (potential double occupied glycopeptide). This pattern is consistent with other reports of Tian et al., used the same digesting enzyme combination. Other reports described using a digestion mixture of trypsin/chymotrypsin with minor changes and were able to describe glycosylation on missing sites in our report.

Determination of the expression system independent of the glycosylation events can help understand protein sequence and structure. To the best of our knowledge, there is no analysis of glycosylation of spike glycoprotein isolated from real patient samples.

Detailed knowledge of S protein glycosylation it's not only important for vaccine development, but also to understand its role on receptor binding. It was recently shown, that N-glycans of different types can have different effects on interaction with receptors [84]. In this study, we showed how S1 protein expressed in different expression models (and hence different major N-glycosylation types) including baculovirus-insect, Chinese hamster ovarian (CHO) cells and two variants of HEK 293 cells, imposed different binding affinities to ACE2.

2.2.6 | O-Glycosylation on different proteoforms and different used techniques

One of the important aspects of spike glycoproteins analysis was the form (partial sequence or mutation) of glycoprotein for overexpression. As mentioned above, Wang et al., [93] Tian et al., [96] and Zhang et al., [95] overexpressed subunit S1 containing receptor binding domain (RBD). Several groups described analysis of protein which has mutated furin cleavage sites such as R683A, R685A [68, 92]; R683S, R685S [70] and R683G, A684S, and R685G [58]. Spike protein is overexpressed as one intact unit. Many studies described glycosylation analysis of
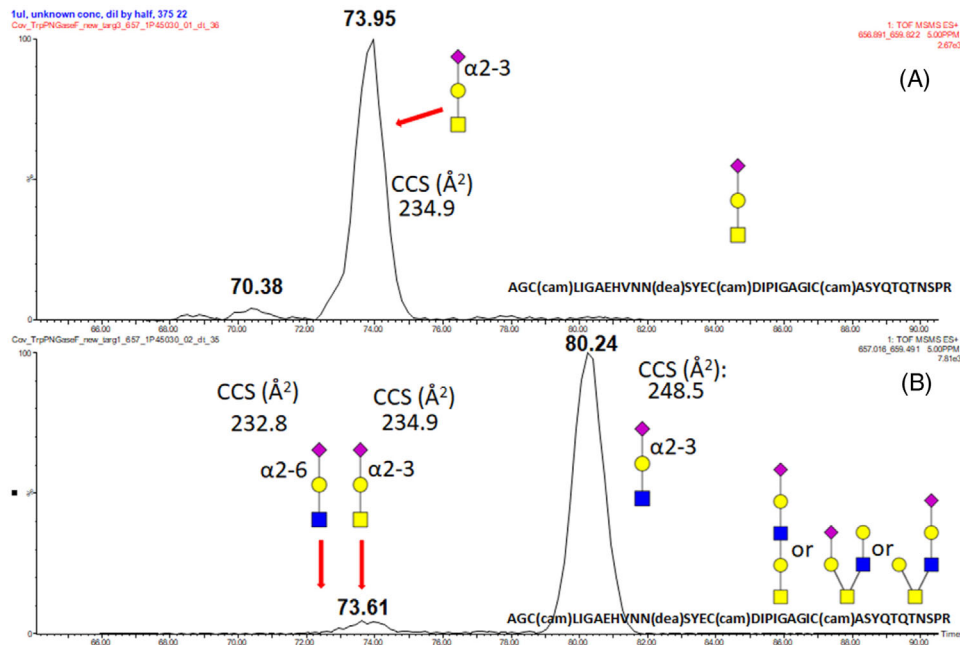
FIGURE 5 An example of separation of O-glycostructures using of ion mobility technique. (Sanda et al., [68])

protein overexpressed as trimeric complex similar to real viral spike complex [94]. Table 1 epitomizes the research studies performed on the O-glycosylation sites.

epth of O-glycoproteomics analysis is shown in column "Total sites" which is number of occupied sites identified in spike glycoprotein. Contrary to N-glycosylation, there is no known sequence motif for O-glycosylation. Therefore, the number of identified sites greatly depends on the used methodology, enrichment, and the digestion system. Overall, O-glycosylation degree (site occupancy) of overexpressed protein is relatively low compared to N-glycosylation. Many different groups (9 out of 11 reports) identified 1 or 2 occupied O-glycosites on the receptor binding domain S323 and S325. These glycosites could have an influence on the binding of spike protein to ACE2 receptor [52, 84]. This information is listed in column "RBD" sites in Table 1. Another physiologically important O-glycosylation region was predicted to be near the furin cleavage site. This polybasic furin cleavage sequence is specific for human SARS-CoV-2 spike glycoprotein [52]. There are three predicted O-glycosites near this site: S673, T678, and S686. A total of eight reports described T678 occupied sites which is only two amino acid from the furin cleavage site. Also, Gao et al., showed occupied S686 in his report. There was no described identification of occupied S673. In addition, Zhang et al., disclosed occupancy of sites T676 and S680 which were not predicted to be occupied [100]. Mutations of T676 and T678 were used in recent study that confirmed influence of O-glycosylation on furin cleavage efficiency [100]. Some publications used EthcD fragmentation to efficiently assign O-glycosites [68, 73]. The rest of the publication mostly used beam type fragmentation and deduced glycosite occupancy mostly on the beam type fragmentation. Major glycans occupied O-glycosites were core 1 and core 2 sialylated structures. Some of the reports identified fucosylated O-glycans. Sanda et al., used cyclic ion mobility on the oxonium

ions fragments to analyze sialic acid linkages. They found that O-glycan structures that were, commonly described as core 2, could be in the case of spike glycoprotein a mixture of extended core 1 and core 2 with different sialic acid linkages (Figure 5).

Two studies used top-Ddown proteomics to analyze O-glycosylation of receptor binding domain only[78, 79]. The O-glycopeptide enrichment, in the case of Zhang Y et al., can significantly increase the number of identified glycosites [94]. However, the physiological significance is questionable due to the very low site occupancy described in other reports. Occupancy of S/T of N-glycomotives is very rare but two reports showed partial O-glycosylation occupancy of peptides after N-glycan cleavage [68, 101].

2.2.7 | Source of data variability between laboratories, instrumentation, and methodologies

It is important to note that the identified glycoforms of the overexpressed spike protein greatly depend on the expression system used (Figure 4) and may not necessarily reflect the virion protein glycoforms isolated from infected cells of SARS-CoV-2 patients. A number of different protein forms and expression systems have been used to deduce structure and glycosylation of SARS-CoV-2 S protein. Many of those were explored as candidates in vaccine research. To the best of our knowledge, only three studies have been carried out on SARS-CoV-2 virion preparations [96, 97, 99]. The rest of these publication described glycosylation of overexpressed protein in human embryonic kidney cells or insect-cell expression systems. As expected, the glycosylation composition strongly influenced by the expression system as well as the number of passages [90].

(A)

(B)
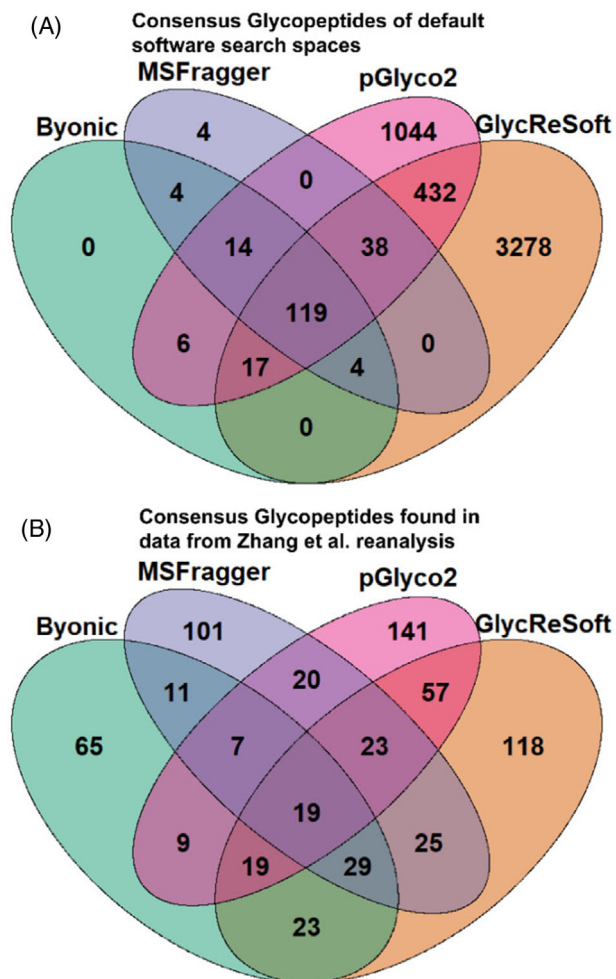
FIGURE 6 Comparison of software used for N-glycoproteomics data processing from Hackett et al., [89]. (A) Comparison of glycoproteomics software, and (B) data analysis from Zhang et. al., [105]

In addition, the type of digestion, LC-MS system as well as the software used for identification and quantification could be a source of interlaboratory variation when describing glycosylation of SARS-CoV-2 spike glycoprotein. A recent interlaboratory study was focused on examining the variations of data processing and evaluation. This study disclosed significant differences in observation of the number of glycoforms as well as composition with various software, evaluation criteria, and experience of researchers [102]. Figure 4 epitomizes the N-glycosylation analysis with information about the source of analyzed material, digestion system used, and glycan identification. As mentioned above Hackett et al., compared four different types of software and the results varied significantly (Figure 6) [89].In their previous work he used Watanabe et al., data [70] and his software tool named GlycReSoft [89]. They observed similar quantities and glycan distribution as their original work but identified many sulfated glycans especially on the site N1074 (6 out of major 30) and penta-antennary or polyLacNAc containing glycopeptides. This indicates that the results strongly depend on the glycan composition of database used for the

study. Figure 6 represents the various types of software and their corresponding outcomes. The N-linked glycans from the native spikes glycans are very similar to that of the recombinant glycoprotein glycans [96].

Differences in data are not only influenced by experimental variation, but also by data misinterpretation as well. For example, bioinformatics mass data analysis by Krishnan et al., has mixed nonidentified N-glycosites in some cases as unoccupied [103]. Furthermore, Segreto et al., misinterpred O-glycosylation data obtained by EthcD as data obtained by HCD fragmentation with modulated collision energy [104].

3 | CONCLUDING REMARKS AND OUTLOOK

In this review, we collected and compared most of results on glycosylation of SARS-CoV-2. We showed the discrepancies found on both glycan structural characterization and site occupancy among the different studies. These could be related to the differences in the expression systems, conditions, and/or form of expressed spike glycoprotein. Moreover, the complex molecular structures and glycan heterogeneity of the spike protein binding domain are still an unresolved mystery due to the challenges in conventional Mass Spectrometry techniques and/or analysis. Furthermore, we demonstrated the most recent advancements in N-glycosylation as well as O-glycosylation analysis in greater detail. Also, we highlighted the most recent discrepancies in the current literature of O-glycosylation in the SARS-CoV-2 spike glycoprotein and explained some misinterpretation of previously results in recent publications. Data analysis of O-glycosylation continues to be a major source of variation. A unified technique and a reliable software may help standardize the workflow of this complex type of analysis.

ORCID
Miloslav Sanda https://orcid.org/0000-0002-7735-3635

REFERENCES
1. Buchmann, J. P., & Holmes, E. C. (2015). Cell walls and the convergent evolution of the viral envelope. *Microbiology and Molecular Biology Reviews*, 79(4), 403–418. https://doi.org/10.1128/MMBR.00017-15
2. Long, J. S., Mistry, B., Haslam, S. M., & Barclay, W. S. (2019). Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology*, 17(2), 67–81. https://doi.org/10.1038/s41579-018-0115-z
3. Taubenberger, J. K., & Kash, J. C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell host & microbe*, 7(6), 440–451. https://doi.org/10.1016/j.chom.2010.05.009
4. Braun, L., Brenier-Pinchart, M.-P., Yogavel, M., Curt-Varesano, A., Curt-Bertini, R.-L., Hussain, T., Kieffer-Jaquinod, S., Coute, Y., Pelloux,

H., Tardieux, I., Sharma, A., Belrhali, H., Bougdour, A., & Hakimi, M.-A. (2013). A Toxoplasma dense granule protein, GRA24, modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation. *Journal of Experimental Medicine*, *210*(10), 2071–2086. https://doi.org/10.1084/jem.20130103

5. Duyen, H. T. L., Ngoc, T. V., Ha, D. T., Hang, V. T. T., Kieu, N. T. T., Young, P. R., Farrar, J. J., Simmons, C. P., Wolbers, M., & Wills, B. A. (2011). Kinetics of plasma viremia and soluble nonstructural protein 1 concentrations in dengue: Differential effects according to serotype and immune status. *Journal of Infectious Diseases*, *203*(9), 1292–1300. https://doi.org/10.1093/infdis/jir014

6. Prabakaran, S., Lippens, G., Steen, H., & Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *4*(6), 565–583. https://doi.org/10.1002/wsbm.1185

7. Depetris, R. S., Julien, J.-P., Khayat, R., Lee, J. H., Pejchal, R., Katpally, U., Cocco, N., Kachare, M., Massi, E., David, K. B., Cupo, A., Marozsan, A. J., Olson, W. C., Ward, A. B., Wilson, I. A., Sanders, R. W., & Moore, J. P. (2012). Partial enzymatic deglycosylation preserves the structure of cleaved recombinant HIV-1 envelope glycoprotein trimers. *Journal of Biological Chemistry*, *287*(29), 24239–24254. https://doi.org/10.1074/jbc.M112.371898

8. Bagdonaite, I., & Wandall, H. H. (2018). Global aspects of viral glycosylation. *Glycobiology*, *28*(7), 443–467. https://doi.org/10.1093/glycob/cwy021

9. Jeffers, S. A., Sanders, D. A., & Sanchez, A. (2002). Covalent modifications of the ebola virus glycoprotein. *Journal of Virology*, *76*(24), 12463–12472. https://doi.org/10.1128/jvi.76.24.12463-12472.2002

10. Pabst, M., Chang, M., Stadlmann, J., & Altmann, F. (2012). Glycan profiles of the 27 N-glycosylation sites of the HIV envelope protein CN54gp140. *Biological Chemistry*, *393*(8), 719–730. https://doi.org/10.1515/hsz-2012-0148

11. Yang, W., Shah, P., Toghi Eshghi, S., Yang, S., Sun, S., Ao, M., Rubin, A., Jackson, J. B, & Zhang, H. (2014). Glycoform analysis of recombinant and human immunodeficiency virus envelope protein gp120 via higher energy collisional dissociation and spectral-aligning strategy. *Analytical Chemistry*, *86*(14), 6959–6967. https://doi.org/10.1021/ac500876p

12. Knauer, R., & Lehle, L. (1999). The oligosaccharyltransferase complex from yeast. *Biochimica et Biophysica Acta*, *1426*(2), 259–273. https://doi.org/10.1016/s0304-4165(98)00128-7

13. Mohanty, S., P Chaudhary, B., & Zoetewey, D. (2020). Structural insight into the mechanism of N-Linked glycosylation by oligosaccharyltransferase. *Biomolecules*, *10*(4), 624. https://doi.org/10.3390/biom10040624

14. Welply, J. K., Shenbagamurthi, P., Lennarz, W. J., & Naider, F. (1983). Substrate recognition by oligosaccharyltransferase. Studies on glycosylation of modified Asn-X-Thr/Ser tripeptides. *Journal of Biological Chemistry*, *258*(19), 11856–11863.

15. Darebna, P., Novak, P., Kucera, R., Topolcan, O., Sanda, M., Goldman, R., & Pompach, P. (2017). Changes in the expression of N- and O-glycopeptides in patients with colorectal cancer and hepatocellular carcinoma quantified by full-MS scan FT-ICR and multiple reaction monitoring. *Journal of Proteomics*, *153*, 44–52. https://doi.org/10.1016/j.jprot.2016.09.004

16. Sanda, M., Pompach, P., Benicky, J., & Goldman, R. (2013). LC-MS3 quantification of O-glycopeptides in human serum. *Electrophoresis*, *34*(16), 2342–4349. https://doi.org/10.1002/elps.201200658

17. Yang, S., Höti, N., Yang, W., Liu, Y., Chen, L., Li, S., & Zhang, H. (2017). Simultaneous analyses of N-linked and O-linked glycans of ovarian cancer cells using solid-phase chemoenzymatic method. *Clinical Proteomics*, *14*, 3. https://doi.org/10.1186/s12014-017-9137-1

18. Crispin, M., & Doores, K. J. (2015). Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Current Opinion in Virology*, *11*, 63–69. https://doi.org/10.1016/j.coviro.2015.02.002

19. Marth, J. D., & Grewal, P. K. (2008). Mammalian glycosylation in immunity. *Nature Reviews Immunology*, *8*(11), 874–887. https://doi.org/10.1038/nri2417

20. Tate, M., Job, E., Deng, Y.-M., Gunalan, V., Maurer-Stroh, S., & Reading, P. (2014). Playing hide and seek: How glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses*, *6*(3), 1294–1316. https://doi.org/10.3390/v6031294

21. Nordén, R., Halim, A., Nyström, K., Bennett, E. P., Mandel, U., Olofsson, S., Nilsson, J., & Larson, G. (2015). O-linked glycosylation of the mucin domain of the herpes simplex virus type 1-specific glycoprotein gC-1 is temporally regulated in a seed-and-spread manner. *Journal of Biological Chemistry*, *290*(8), 5078–5091. https://doi.org/10.1074/jbc.M114.616409

22. Nordén, R., Nyström, K., Adamiak, B., Halim, A., Nilsson, J., Larson, G., Trybala, E., & Olofsson, S. (2013). Involvement of viral glycoprotein gC-1 in expression of the selectin ligand sialyl-Lewis X induced after infection with herpes simplex virus type 1. *APMIS*, *121*(4), 280–289. https://doi.org/10.1111/j.1600-0463.2012.02967.x

23. Xiang, T., Yang, G., Liu, X., Zhou, Y., Fu, Z., Lu, F., Gu, J., Taniguchi, N., Tan, Z., Chen, X., Xie, Y., Guan, F., & Zhang, X.-L. (2017). Alteration of N-glycan expression profile and glycan pattern of glycoproteins in human hepatoma cells after HCV infection. *Biochimica et Biophysica Acta General Subjects*, *1861*(5 Pt A), 1036–1045. https://doi.org/10.1016/j.bbagen.2017.02.014

24. Amiri-Dashatan, N., Koushki, M., & Rezaei-Tavirani, M. (2022). Mass spectrometry-based proteomics research to fight COVID-19: An expert review on hopes and challenges. *Omics*, *26*(1), 19–34. https://doi.org/10.1089/omi.2021.0182

25. Bittremieux, W., Adams, C., Laukens, K., Dorrestein, P. C., & Bandeira, N. (2021). Open science resources for the mass spectrometry-based analysis of SARS-CoV-2. *Journal of Proteome Research*, *20*(3), 1464–1475. https://doi.org/10.1021/acs.jproteome.0c00929

26. Mahmud, I., & Garrett, T. J. (2020). Mass spectrometry techniques in emerging pathogens studies: COVID-19 perspectives. *Journal of the American Society for Mass Spectrometry*, *31*(10), 2013–2024. https://doi.org/10.1021/jasms.0c00238

27. Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., Meyer, J. G., Quan, Q., Muehlbauer, L. K., Trujillo, E. A., He, Y., Chopra, A., Chieng, H. C., Tiwari, A., Judson, M. A., Paulson, B., Brademan, D. R., Zhu, Y., Serrano, L. R., … Jaitovich, A. (2021). Large-scale multi-omic analysis of COVID-19 severity. *Cell Systems*, *12*(1), 23–40. e7 e27. https://doi.org/10.1016/j.cels.2020.10.003

28. Sorelle, J. A., Patel, K., Filkins, L., & Park, J. Y. (2020). Mass spectrometry for COVID-19. *Clinical Chemistry*, *66*(11), 1367–1368. https://doi.org/10.1093/clinchem/hvaa222

29. Terracciano, R., Preianò, M., Fregola, A., Pelaia, C., Montalcini, T., & Savino, R. (2021). Mapping the SARS-CoV-2-host protein-protein interactome by affinity purification mass spectrometry and proximity-dependent biotin labeling: A rational and straightforward route to discover host-directed Anti-SARS-CoV-2 therapeutics. *International Journal of Molecular Sciences*, *22*(2), 532. https://doi.org/10.3390/ijms22020532

30. Van Puyvelde, B., & Dhaenens, M. (2021). Add mass spectrometry to the pandemic toolbox. https://doi.org/10.7554/eLife.75471

31. Witkowska, D. (2020). Mass spectrometry and structural biology techniques in the studies on the coronavirus-receptor interaction. *Molecules (Basel, Switzerland)*, *25*(18), 4133. https://doi.org/10.3390/molecules25184133

32. Xing, Y., Zhao, B., Yin, L., Guo, M., Shi, H., Zhu, Z., Zhang, L., He, J., Ling, Y., Gao, M., Lu, H., Mao, E., & Zhang, L. (2021). Vitamin C supplementation is necessary for patients with coronavirus disease: An

ultra-high-performance liquid chromatography-tandem mass spectrometry finding. *Journal of Pharmaceutical and Biomedical Analysis*, *196*, 113927. https://doi.org/10.1016/j.jpba.2021.113927

33. Yuan, Z.-C., & Hu, B. (2021). Mass spectrometry-based human breath analysis: Towards COVID-19 diagnosis and research. *Journal of Analysis and Testing*, *5*(4), 287–297. https://doi.org/10.1007/s41664-021-00194-9

34. Orlova, O. V., Drutsa, V. L., Spirin, P. V., Ivanov, A. V., Prasolov, V. S., Rubtsov, P. M., Kochetkov, S. N., & Beljelarskaya, S. N. (2015). The role of HCV e2 protein glycosylation in functioning of virus envelope proteins in insect and mammalian cells. *Acta Naturae*, *7*(1), 87–97.

35. Lennemann, N. J., Rhein, B. A., Ndungo, E., Chandran, K., Qiu, X., & Maury, W. (2014). Comprehensive functional analysis of N-linked glycans on Ebola virus GP1. *mBio*, *5*(1), e00862–00813. https://doi.org/10.1128/mBio.00862-13

36. Stone, J. A., Nicola, A. V., Baum, L. G., & Aguilar, H. C. (2016). Multiple novel functions of henipavirus O-glycans: The first O-glycan functions identified in the paramyxovirus family. *PLOS Pathogens*, *12*(2), e1005445. https://doi.org/10.1371/journal.ppat.1005445

37. Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., & Klenk, H.-D. (2004). Neuraminidase is important for the initiation of influenza virus infection in human airway epithelium. *Journal of Virology*, *78*(22), 12665–12667. https://doi.org/10.1128/JVI.78.22.12665-12667.2004

38. Palese, P., & Compans, R. W. (1976). Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA): mechanism of action. *Journal of General Virology*, *33*(1), 159–163. https://doi.org/10.1099/0022-1317-33-1-159

39. Roberts, P. C., Garten, W., & Klenk, H. D. (1993). Role of conserved glycosylation sites in maturation and transport of influenza A virus hemagglutinin. *Journal of Virology*, *67*(6), 3048–3060. https://doi.org/10.1128/JVI.67.6.3048-3060.1993

40. Tsuchiya, E., Sugawara, K., Hongo, S., Matsuzaki, Y., Muraki, Y., & Nakamura, K. (2002). Role of overlapping glycosylation sequons in antigenic properties, intracellular transport and biological activities of influenza A/H2N2 virus haemagglutinin. *Journal of General Virology*, *83*(Pt 12), 3067–3074. https://doi.org/10.1099/0022-1317-83-12-3067

41. Alymova, I. V., York, I. A., Air, G. M., Cipollo, J. F., Gulati, S., Baranovich, T., Kumar, A., Zeng, H., Gansebom, S., & Mccullers, J. A. (2016). Glycosylation changes in the globular head of H3N2 influenza hemagglutinin modulate receptor binding without affecting virus virulence. *Science Reports*, *6*, 36216. https://doi.org/10.1038/srep36216

42. Ohuchi, M., Ohuchi, R., Feldmann, A., & Klenk, H. D. (1997). Regulation of receptor binding affinity of influenza virus hemagglutinin by its carbohydrate moiety. *Journal of Virology*, *71*(11), 8377–8384. https://doi.org/10.1128/JVI.71.11.8377-8384.1997

43. Ohuchi, M., Orlich, M., Ohuchi, R., Simpson, B. E. J., Garten, W., Klenk, H.-D., & Rott, R. (1989). Mutations at the cleavage site of the hemagglutinin alter the pathogenicity of influenza virus a/chick/penn/83 (H5N2). *Virology*, *168*(2), 274–280. https://doi.org/10.1016/0042-6822(89)90267-5

44. Wagner, R., Wolff, T., Herwig, A., Pleschka, S., & Klenk, H.-D. (2000). Interdependence of hemagglutinin glycosylation and neuraminidase as regulators of influenza virus growth: a study by reverse genetics. *Journal of Virology*, *74*(14), 6316–6323. https://doi.org/10.1128/jvi.74.14.6316-6323.2000

45. Hütter, J., Rödig, J. V., Höper, D., Seeberger, P. H., Reichl, U., Rapp, E., & Lepenies, B. (2013). Toward animal cell culture-based influenza vaccine design: Viral hemagglutinin N- glycosylation markedly impacts immunogenicity. *Journal of Immunology*, *190*(1), 220–230. https://doi.org/10.4049/jimmunol.1201060

46. Sommerstein, R., Flatz, L., Remy, M. M., Malinge, P., Magistrelli, G., Fischer, N., Sahin, M., Bergthaler, A., Igonet, S., Ter Meulen, J., Rigo, D., Meda, P., Rabah, N., Coutard, B., Bowden, T. A., Lambert, P.-H., Siegrist, C.-A., & Pinschewer, D. D. (2015). Arenavirus glycan shield promotes neutralizing antibody evasion and protracted infection. *PLOS Pathogens*, *11*(11), e1005276. https://doi.org/10.1371/journal.ppat.1005276

47. Walls, A. C., Tortorici, M. A., Frenz, B., Snijder, J., Li, W., Rey, F. A., Dimaio, F., Bosch, B.-J., & Veesler, D. (2016). Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nature Structural & Molecular Biology*, *23*(10), 899–905. https://doi.org/10.1038/nsmb.3293

48. Altman, M. O., Angel, M., Košík, I., Trovão, N. S., Zost, S. J., Gibbs, J. S., Casalino, L., Amaro, R. E., Hensley, S. E., Nelson, M. I., & Yewdell, J. W. (2019). Human influenza A Virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit. *mBio*, *10*(2), e00204–00219. https://doi.org/10.1128/mBio.00204-19

49. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T. S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell*, *181*(2), 271–280. e278. https://doi.org/10.1016/j.cell.2020.02.052

50. Liya, G., Yuguang, W., Jian, L., Huaiping, Y., Xue, H., Jianwei, H., Jiaju, M., Youran, L., Chen, M., & Yiqing, J. (2020). Studies on viral pneumonia related to novel coronavirus SARS-CoV-2, SARS-CoV, and MERS-CoV: A literature review. *Apmis*, *128*(6), 423–432. https://doi.org/10.1111/apm.13047

51. South, A. M., Diz, D. I., & Chappell, M. C. (2020). COVID-19, ACE2, and the cardiovascular consequences. *American Journal of Physiology. Heart and Circulatory Physiology*, *318*(5), H1084–H1090. https://doi.org/10.1152/ajpheart.00217.2020

52. Walls, A. C., Park, Y.-J., Tortorici, M. A, Wall, A., Mcguire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell*, *181*(2), 281–292. e286. https://doi.org/10.1016/j.cell.2020.02.058

53. Ganji, R., & Reddy, P. H. (2020). Impact of COVID-19 on mitochondrial-based immunity in aging and age-related diseases. *Frontiers in Aging Neuroscience*, *12*, 614650. https://doi.org/10.3389/fnagi.2020.614650

54. Poh, C. M., Carissimo, G., Wang, B., Amrun, S. N., Lee, C Y.-P., Chee, R. S.-L., Fong, S.-W., Yeo, N. K.-W., Lee, W.-H., Torres-Ruesta, A., Leo, Y.-S., Chen, M. I.-C., Tan, S.-Y., Chai L Y. A., Kalimuddin, S., Kheng, S. S G., Thien, S.-Y., Young, B. E., Lye, D. C., … Ng, L. F. P. (2020). Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nature Communication*, *11*(1), 2806. https://doi.org/10.1038/s41467-020-16638-2

55. Romeo, A., Iacovelli, F., & Falconi, M. (2020). Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors. *Virus Research*, *286*, 198068. https://doi.org/10.1016/j.virusres.2020.198068

56. Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C.-L., Abiona, O., Graham, B. S., & Mclellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*, *367*(6483), 1260–1263. https://doi.org/10.1126/science.abb2507

57. Zhang, B.-Z., Hu, Y.-F., Chen, L.-L., Yau, T., Tong, Y.-G., Hu, J.-C., Cai, J.-P., Chan, K.-H., Dou, Y., Deng, J., Wang, X.-L., Hung, I. F.-N., To, K. K.-W., Yuen, K. Y., & Huang, J.-D. (2020). Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell Research*, *30*(8), 702–704. https://doi.org/10.1038/s41422-020-0366-x

58. Zhao, P., Praissman, J. L., Grant, O. C., Cai, Y., Xiao, T., Rosenbalm, K. E., Aoki, K., Kellman, B. P., Bridger, R., Barouch, D. H., Brindley, M. A., Lewis, N. E., Tiemeyer, M., Chen, B., Woods, R. J., & Wells, L. (2020). Virus-receptor interactions of glycosylated SARS-CoV-2 spike and human ACE2 receptor. *Cell Host & Microbe*, *28*(4), 586–601. e586. https://doi.org/10.1016/j.chom.2020.08.004

59. Grant, O. C., Montgomery, D., Ito, K., & Woods, R. J. (2020). Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Science Reports*, 10(1), 14991. https://doi.org/10.1038/s41598-020-71748-7

60. Gstöttner, C., Zhang, T., Resemann, A., Ruben, S., Pengelley, S., Suckau, D., Welsink, T., Wuhrer, M., & Domínguez-Vega, E. (2021). Structural and functional characterization of SARS-CoV-2 RBD domains produced in mammalian cells. *Analytical Chemistry*, 93(17), 6839–6847. https://doi.org/10.1021/acs.analchem.1c00893

61. Pinto, D., Park, Y. J., Beltramello, M., Walls, A. C., Tortorici, M. A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., Peter, A., Guarino, B., Spreafico, R., Cameroni, E., Case, J. B., Chen, R. E., Havenar-Daughton, C., Snell, G., Telenti, A., … Corti, D. (2020). Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *bioRxiv*, 2020.2004.2007.023903. https://doi.org/10.1101/2020.04.07.023903

62. Gebrehiwot, A. G., Melka, D. S., Kassaye, Y. M., Rehan, I. F., Rangappa, S., Hinou, H., Kamiyama, T., & Nishimura, S.-I. (2018). Healthy human serum N-glycan profiling reveals the influence of ethnic variation on the identified cancer-relevant glycan biomarkers. *Plos One*, 13(12), e0209515. https://doi.org/10.1371/journal.pone.0209515

63. Krištić, J., Vučković, F., Menni, C., Klarić, L., Keser, T., Beceheli, I., Pučić-Baković, M., Novokmet, M., Mangino, M., Thaqi, K., Rudan, P., Novokmet, N., Šarac, J., Missoni, S., Kolčić, I., Polašek, O., Rudan, I., Campbell, H., Hayward, C., … Lauc, G. (2014). Glycans are a novel biomarker of chronological and biological ages. *Journals of Gerontology. Series A*, *Biological Sciences and Medical Sciences*, 69(7), 779–789. https://doi.org/10.1093/gerona/glt190

64. Pavić, T., Dilber, D., Kifer, D., Selak, N., Keser, T., Ljubičić, Đ., Vukić Dugac, A., Lauc, G., Rumora, L., & Gornik, O. (2018). N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *Journal of Translational Medicine*, 16(1), 323. https://doi.org/10.1186/s12967-018-1695-0

65. Rudman, N., Gornik, O., & Lauc, G. (2019). Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Letters*, 593(13), 1598–1615. https://doi.org/10.1002/1873-3468.13495

66. Khatri, K., Klein, J. A., White, M. R., Grant, O. C., Leymarie, N., Woods, R. J., Hartshorn, K. L., & Zaia, J. (2016). Integrated omics and computational glycobiology reveal structural basis for influenza A virus glycan microheterogeneity and host interactions. *Molecular & Cellular Proteomics*, 15(6), 1895–1912. https://doi.org/10.1074/mcp.M116.058016

67. Bangaru, S., Ozorowski, G., Turner, H. L., Antanasijevic, A., Huang, D., Wang, X., Torres, J. L., Diedrich, J. K., Tian, J. H., Portnoff, A. D., Patel, N., Massare, M. J., Yates, J. R. 3rd, Nemazee, D., Paulson, J. C., Glenn, G., Smith, G., & Ward, A. B. (2020). Structural analysis of full-length SARS-CoV-2 spike protein from an advanced vaccine candidate. *Science*, 370(6520), 1089-1094. https://doi.org/10.1101/2020.08.06.234674

68. Sanda, M., Morrison, L., & Goldman, R. (2021). N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Analytical Chemistry*, 93(4), 2003–2009. https://doi.org/10.1021/acs.analchem.0c03173

69. Shajahan, A., Supekar, N. T., Gleinich, A. S., & Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology*, 30(12), 981–988. https://doi.org/10.1093/glycob/cwaa042

70. Watanabe, Y., Allen, J. D., Wrapp, D., Mclellan, J. S., & Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. *Science*, 369(6501), 330–333. https://doi.org/10.1126/science.abb9983

71. Shajahan, A., Pepi, L. E., Rouhani, D. S., Heiss, C., & Azadi, P. (2021). Glycosylation of SARS-CoV-2: Structural and functional insights. *Analytical and bioanalytical chemistry*, 413(29), 7179–7193. https://doi.org/10.1007/s00216-021-03499-x

72. Andersen, K. G., Rambaut, A., Lipkin, W. I, Holmes, E. C., & Garry, R. F. (2020). The proximal origin of SARS-CoV-2. *Nature Medicine*, 26(4), 450–452. https://doi.org/10.1038/s41591-020-0820-9

73. Bagdonaite, I., Thompson, A. J., Wang, X., Søgaard, M., Fougeroux, C., Frank, M., Diedrich, J. K., Yates, J. R., Salanti, A., Vakhrushev, S. Y., Paulson, J. C., & Wandall, H. H. (2021). Site-specific O-glycosylation analysis of SARS-CoV-2 spike protein produced in insect and human cells. *Viruses*, 13(4), 551. https://doi.org/10.3390/v13040551

74. Kozlik, P., Goldman, R., & Sanda, M. (2018). Hydrophilic interaction liquid chromatography in the separation of glycopeptides and their isomers. *Analytical and bioanalytical chemistry*, 410(20), 5001–5008. https://doi.org/10.1007/s00216-018-1150-3

75. Pedrali, A., Tengattini, S., Marrubini, G., Bavaro, T., Hemström, P., Massolini, G., Terreni, M., & Temporini, C. (2014). Characterization of intact neo-glycoproteins by hydrophilic interaction liquid chromatography. *Molecules (Basel, Switzerland)*, 19(7), 9070–9088. https://doi.org/10.3390/molecules19079070

76. Mechref, Y., Hu, Y., Desantos-Garcia, J. L., Hussein, A., & Tang, H. (2013). Quantitative glycomics strategies. *Molecular & Cellular Proteomics*, 12(4), 874–884. https://doi.org/10.1074/mcp.R112.026310

77. Young, C., Condina, M. R., Briggs, M. T., Moh, E. S. X., Kaur, G., Oehler, M. K., & Hoffmann, P. (2021). In-house packed porous graphitic carbon columns for liquid chromatography-mass spectrometry analysis of N-glycans. *Frontiers in Chemistry*, 9, 653959. https://doi.org/10.3389/fchem.2021.653959

78. Roberts, D. S., Mann, M., Li, B. H., Kim, D., Brasier, A. R., Jin, S., & Ge, Y. (2022). Distinct core glycan and O-glycoform utilization of SARS-CoV-2 omicron variant spike protein RBD revealed by top-down mass spectrometry. *bioRxiv*, 2022.2002.2009.479776. https://doi.org/10.1101/2022.02.09.479776

79. Roberts, D. S., Mann, M., Melby, J. A., Larson, E. J., Zhu, Y., Brasier, A. R., Jin, S., & Ge, Y. (2021). Structural O-glycoform heterogeneity of the SARS-CoV-2 spike protein receptor-binding domain revealed by top-down mass spectrometry. *Journal of the American Chemical Society*, 143(31), 12014–12024. https://doi.org/10.1021/jacs.1c02713

80. Wilson, J. W., Bilbao, A., Wang, J., Liao, Y.-C., Velickovic, D., Wojcik, R., Passamonti, M., Zhao, R., Gargano, A. F. G., Gerbasi, V. R., PasÄ-Tolić, L., Baker, S. E., & Zhou, M. (2022). Online hydrophilic interaction chromatography (HILIC) enhanced top-down mass spectrometry characterization of the SARS-CoV-2 spike receptor-binding domain. *Analytical Chemistry*, 94(15), 5909–5917. https://doi.org/10.1021/acs.analchem.2c00139

81. Klein, J. A., & Zaia, J. (2020). Assignment of coronavirus spike protein site-specific glycosylation using GlycReSoft. *bioRxiv*, 2020.2005.2031.125302. https://doi.org/10.1101/2020.05.31.125302

82. Hinneburg, H., Stavenhagen, K., Schweiger-Hufnagel, U., Pengelley, S., Jabs, W., Seeberger, P. H., Silva, D. V., Wuhrer, M., & Kolarich, D. (2016). The art of destruction: optimizing collision energies in quadrupole-time of flight (Q-TOF) instruments for glycopeptide-based glycoproteomics. *Journal of the American Society for Mass Spectrometry*, 27(3), 507–519. https://doi.org/10.1007/s13361-015-1308-6

83. Sanda, M., Benicky, J., & Goldman, R. (2020). Low collision energy fragmentation in structure-specific glycoproteomics analysis. *Analytical Chemistry*, 92(12), 8262–8267. https://doi.org/10.1021/acs.analchem.0c00519

84. Huang, C., Tan, Z., Zhao, K., Zou, W., Wang, H., Gao, H., Sun, S., Bu, D., Chai, W., & Li, Y. (2021). The effect of N-glycosylation of SARS-CoV-2 spike protein on the virus interaction with the host cell ACE2 receptor. *iScience*, 24(11), 103272. https://doi.org/10.1016/j.isci.2021.103272

85. Bezstarosti, K., Lamers, M. M., Doff, W. A. S., Wever, P. C., Thai, K. T. D., Van Kampen, J. J. A., Haagmans, B. L., & Demmers, J. A. A. (2021). Targeted proteomics as a tool to detect SARS-CoV-2 proteins

in clinical specimens. *PLOS One*, *16*(11), e0259165. https://doi.org/10.1371/journal.pone.0259165

86. Hober, A., Tran-Minh, K. H., Foley, D., Mcdonald, T., Vissers, J. P., Pattison, R., Ferries, S., Hermansson, S., Betner, I., Uhlén, M., Razavi, M., Yip, R., Pope, M. E., Pearson, T. W., Andersson, L. N., Bartlett, A., Calton, L., Alm, J. J., Engstrand, L., & Edfors, F. (2021). Rapid and sensitive detection of SARS-CoV-2 infection using quantitative peptide enrichment LC-MS analysis. *Elife*, *10*, e70843. https://doi.org/10.7554/eLife.70843

87. Cho, B. G., Gautam, S., Peng, W., Huang, Y., Goli, M., & Mechref, Y. (2021). Direct comparison of N-Glycans and their isomers derived from spike glycoprotein 1 of MERS-CoV, SARS-CoV-1, and SARS-CoV-2. *Journal of Proteome Research*, *20*(9), 4357–4365. https://doi.org/10.1021/acs.jproteome.1c00323

88. Yuan, W., Wei, R., Goldman, R., & Sanda, M. (2019). Optimized fragmentation for quantitative analysis of fucosylated N-glycoproteins by LC-MS-MRM. *Analytical Chemistry*, *91*(14), 9206–9212. https://doi.org/10.1021/acs.analchem.9b01983

89. Hackett, W. E., & Zaia, J. (2021). The Need for community standards to enable accurate comparison of glycoproteomics algorithm performance. *Molecules (Basel, Switzerland)*, *26*(16), 4757. https://doi.org/10.3390/molecules26164757

90. Benicky, J., Sanda, M., Brnakova Kennedy, Z., & Goldman, R. (2019). N-Glycosylation is required for secretion of the precursor to brain-derived neurotrophic factor (proBDNF) carrying sulfated LacdiNAc structures. *Journal of Biological Chemistry*, *294*(45), 16816–16830. https://doi.org/10.1074/jbc.RA119.009989

91. Benicky, J., Sanda, M., Brnakova Kennedy, Z., Grant, O. C., Woods, R. J., Zwart, A., & Goldman, R. (2021). PD-L1 glycosylation and its impact on binding to clinical antibodies. *Journal of Proteome Research*, *20*(1), 485–497. https://doi.org/10.1021/acs.jproteome.0c00521

92. Gao, C., Zeng, J., Jia, N., Stavenhagen, K., Matsumoto, Y., Zhang, H., Li, J., Hume, A. J., Mühlberger, E., van Die, I., Kwan, J., Tantisira, K., Emili, A., & Cummings, R. D. (2020). SARS-CoV-2 spike protein interacts with multiple innate immune receptors. *bioRxiv*, 2020.2007.2029.227462. https://doi.org/10.1101/2020.07.29.227462

93. Wang, D., Baudys, J., Bundy, J. L., Solano, M., Keppel, T., & Barr, J. R. (2020). Comprehensive analysis of the glycan complement of SARS-CoV-2 spike proteins using signature ions-triggered electron-transfer/higher-energy collisional dissociation (EThcD) mass spectrometry. *Analytical Chemistry*, *92*(21), 14730–14739. https://doi.org/10.1021/acs.analchem.0c03301

94. Zhang, S., Go, E. P., Ding, H., Anang, S., Kappes, J. C., Desaire, H., & Sodroski, J. G. (2022). Analysis of glycosylation and disulfide bonding of wild-type SARS-CoV-2 spike glycoprotein. *Journal of Virology*, *96*(3), e0162621. https://doi.org/10.1128/JVI.01626-21

95. Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Wang, S., Zhong, Y., Su, T., Gong, M., Du, D., Lu, X., Cheng, J., & Yang, H. (2021). Site-specific N-glycosylation characterization of recombinant SARS-CoV-2 spike proteins. *Molecular & Cellular Proteomics*, *20*, 100058. https://doi.org/10.1074/mcp.RA120.002295

96. Tian, Y., Parsons, L. M., Jankowska, E., & Cipollo, J. F. (2021). Site-specific glycosylation patterns of the SARS-CoV-2 spike protein derived from recombinant protein and viral WA1 and D614G strains. *Frontiers in Chemistry*, *9*, 767448. https://doi.org/10.3389/fchem.2021.767448

97. Brun, J., Vasiljevic, S. ž., Gangadharan, B., Hensen, M., Chandran, A. V., Hill, M. L., Kiappes, J. L., Dwek, R. A., Alonzi, D. S., Struwe, W. B., & Zitzmann, N. (2021). Assessing antigen structural integrity through glycosylation analysis of the SARS-CoV-2 viral spike. *ACS Central Science*, *7*(4), 586–593. https://doi.org/10.1021/acscentsci.1c00058

98. Yuan, W., Liu, B., Sanda, M., Wei, R., Benicky, J., Novakova, Z., Barinka, C., & Goldman, R. (2022). Glycoforms of human prostate-specific membrane antigen (PSMA) in human cells and prostate tissue. *Prostate*, *82*(1), 132–144. https://doi.org/10.1002/pros.24254

99. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z., Cheng, L., Shi, D., Lu, X., Lei, J., Crispin, M., Shi, Y., Li, L., & Li, S. (2020). Molecular architecture of the SARS-CoV-2 virus. *Cell*, *183*(3), 730–738. e713. https://doi.org/10.1016/j.cell.2020.09.018

100. Zhang, L., Mann, M., Syed, Z. A., Reynolds, H. M., Tian, E., Samara, N. L., Zeldin, D. C., Tabak, L. A., & Ten Hagen, K. G. (2021). Furin cleavage of the SARS-CoV-2 spike is modulated by O-glycosylation. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(47), e2109905118. https://doi.org/10.1073/pnas.2109905118

101. Tian, W., Li, D., Zhang, N., Bai, G., Yuan, K., Xiao, H., Gao, F., Chen, Y., Wong, C. C. L., & Gao, G. F. (2021). O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. *Cell Research*, *31*(10), 1123–1125. https://doi.org/10.1038/s41422-021-00545-2

102. Kawahara, R., Chernykh, A., Alagesan, K., Bern, M., Cao, W., Chalkley, R. J., Cheng, K., Choo, M. S., Edwards, N., Goldman, R., Hoffmann, M., Hu, Y., Huang, Y., Kim, J. Y., Kletter, D., Liquet, B., Liu, M., Mechref, Y., Meng, B., … Thaysen-Andersen, M. (2021). Community evaluation of glycoproteomics informatics solutions reveals high-performance search strategies for serum glycopeptide analysis. *Nature Methods*, *18*(11), 1304–1316. https://doi.org/10.1038/s41592-021-01309-x

103. Krishnan, S., & Krishnan, G. P. (2021). N-glycosylation network construction and analysis to modify glycans on the spike (S) glycoprotein of SARS-CoV-2. *Frontiers in Bioinformatics*, *1*, https://doi.org/10.3389/fbinf.2021.667012

104. Segreto, R., Deigin, Y., Mccairn, K., Sousa, A., Sirotkin, D., Sirotkin, K., Couey, J. J., Jones, A., & Zhang, D. (2021). Should we discount the laboratory origin of COVID-19? *Environmental Chemistry Letters*, 1–15. https://doi.org/10.1007/s10311-021-01211-0

105. Zhou, D., Tian, X., Qi, R., Peng, C., & Zhang, W. (2021). Identification of 22 N-glycosites on spike glycoprotein of SARS-CoV-2 and accessible surface glycopeptide motifs: Implications for vaccination and antibody therapeutics. *Glycobiology*, *31*(1), 69–80. https://doi.org/10.1093/glycob/cwaa052