

Online Hydrophilic Interaction Chromatography (HILIC) Enhanced Top-Down Mass Spectrometry

Characterization of SARS-Cov-2 Spike Receptor Binding Domain

Jesse W. Wilson¹, Aivett Bilbao¹, Juan Wang², Yen-Chen Liao¹, Dusan Velickovic¹, Roza Wojcik³, Marta Passamonti⁴, Rui Zhao¹, Andrea Gargano⁴, Vincent R. Gerbasi², Ljiljana Pasa-Tolic¹, Scott Baker¹, Mowei Zhou¹

1. Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, 3335 Innovation Boulevard, Richland, Washington 99354, United States

2. Biological Sciences Division, Pacific Northwest National Laboratories, 902 Battelle Blvd, Richland, Washington 99354, United States

3. National Security Directorate, Pacific Northwest National Laboratories, 902 Battelle Blvd, Richland, Washington 99354, United States

4. Centre for Analytical Sciences Amsterdam, Amsterdam 1098 XH, The Netherlands; Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam 1098 XH, The Netherlands

Contact: mowei.zhou@pnnl.gov

Abstract:

SARS-CoV-2 cellular infection is mediated by the heavily glycosylated spike protein. Recombinant versions of the spike protein and the receptor binding domain (RBD) are necessary for seropositivity assays and can potentially serve as vaccines against viral infection. RBD plays key roles in the spike protein's structure and function, and thus comprehensive characterization of recombinant RBD is critically important for biopharmaceutical applications. Liquid-chromatography coupled to mass spectrometry (LCMS) has been widely used to characterize post-translational modifications in proteins including glycosylation. Most studies of RBDs were performed at the proteolytic peptide (bottom-up proteomics) or released glycan level because of the technical challenges in resolving highly heterogeneous glycans at the intact protein level. Herein, we evaluated several online separation techniques: 1. C2 reverse-phase liquid chromatography (RPLC), 2. capillary zone electrophoresis (CZE),

and 3. acrylamide-based monolithic hydrophilic interaction chromatography (HILIC) to separate intact recombinant RBDs with varying combinations of glycosylations (glycoforms) for top-down MS. Within the conditions we explored, the HILIC method was superior to RPLC and CZE at separating RBD glycoforms, which differ significantly in neutral glycan groups. In addition, our top-down analysis readily captured unexpected modifications (e.g., cysteinylolation, N-terminal sequence variation) and low abundance, heavily glycosylated proteoforms that may be missed by using glycopeptide data alone. The HILIC top-down MS platform holds great potential in resolving heterogeneous glycoproteins for facile comparison of biosimilars in quality control applications.

Introduction:

The heavily glycosylated spike protein on the surface of SARS-CoV-2 virion particles mediates internalization into human cells via interactions with the cellular surface protein angiotensin converting enzyme-2 (ACE-2).¹⁻⁵ Due to the direct interaction of the spike receptor-binding domain (RBD) and ACE-2, the RBD serves as a key target for neutralizing antibodies to prevent infection.^{1,6,7} Recombinant spike protein and RBD can potentially serve as vaccines,^{1,8,9} and RBDs are necessary for diagnostic purposes in immunoassays.^{7,9} RBD glycosylation has been demonstrated to play a role in ACE-2 binding and aids shielding of the spike protein from antibodies.^{2,4,10,11} In general, protein glycosylation is known to modulate the immune response,^{12,13} and recombinant protein glycosylation is often dependent on the expression platform employed.¹⁴⁻¹⁶ Thus, understanding the full glycosylation profile of the spike RBD is important for the development and quality control of novel therapeutics or vaccines,¹¹ where knowledge of the precise combination of all post translational modifications (PTMs) is necessary.

Individual glycosites can be occupied with many glycan structural variants, which result in different forms of the protein termed glycoforms. With glycoproteins, macroheterogeneity describes the glycan occupancy at a given glycosite, while microheterogeneity is the variation of glycan composition per glycosite. More recent glycoprotein observations have led to the idea of glycan meta-heterogeneity; a higher level of glycan regulation based on the variation in glycosylation across multiple sites.¹⁷ Conventional analysis of protein glycosylation relies on bottom-up mass spectrometry (MS) approaches to produce glycopeptides using enzymatic digestion,^{3,18-20} and release of glycans from the protein for comprehensive glycan profiling.²¹ Although these approaches lead to robust identification of glycosites and glycans (macro- and micro-heterogeneity, respectively), the overall connectivity between various intact glycoforms and other PTMs is lost (meta-heterogeneity). Top-down mass spectrometry experiments skip enzymatic digestion to analyze whole proteins where the relative abundance of exact proteoforms; proteins with varying glycosylation or PTMs, can be determined.²²

To better resolve heterogeneous samples (e.g. glycoproteins), liquid chromatography is often coupled to mass spectrometry (LCMS) to separate proteins based on their chemical properties before MS detection.²²⁻²⁴ Denaturing reversed-phase liquid chromatography (RPLC) is among the most widely used separation techniques with protein retention based on hydrophobic interactions.²⁵ Capillary zone electrophoresis (CZE) coupled to MS, in comparison, separates based on charge and size characteristics within a capillary using an applied electric field in the presence of a background electrolyte (BGE).²⁶ Ion exchange is another charge-based online separation method that has recently been developed to study intact therapeutic proteins such as antibodies.²⁷ Hydrophilic interaction chromatography (HILIC), which has the opposite selectivity to RPLC, separates molecules based on hydrophilicity with greater retention for hydrophilic molecules such as glycans.²⁸⁻³²

Two recent publications have used top-down mass spectrometry to study RBD glycosylation.^{33, 34} Both of these studies identified O-glycosylation at site T323 of the spike S1 protein and two N-glycosylation sites at N331 and N343, the same sites that were previously identified with bottom-up glycoproteomics.^{3, 19, 20} The study from Roberts et al.,³⁴ used denatured and native top-down analysis to determine the relative abundance of O-glycoforms of the RBD, while the study from Gstöttner et al.³³ combined bottom-up, and intact protein analysis with multiple glycosidase enzymatic steps to study the N- and O-linked glycosylation profile of the RBD. However, the modality of online separation for improving the intact protein analysis has not yet been thoroughly investigated in these reports. Herein we performed RPLC, CZE, and HILIC separations on several RBDs recombinantly expressed in HEK 293 cells from two different vendors (Sino Biological and RayBiotech). HILIC allowed the greatest separation of RBD glycoforms, which were different in their neutral glycans. Drastic differences in glycan composition were also detected between vendor sources when the same expression platform (from the same cell line) was used. Additionally, our results suggest the RBD exhibits more than 200 individual glycoforms that are assignable. Top-down MS also helped the discovery of unexpected PTMs on the

proteins that may affect the structure and function. When compared with glycopeptide data, HILIC top-down analysis better detected low abundance and/or heavily glycosylated proteoforms that may be missed due to more severe detection biases at the peptide level. We anticipate that online HILIC separation has great potential for defining the meta-heterogeneity of heterogeneous glycoproteins for biotherapeutic or biotechnology use.

Materials and Methods:

Chemicals and Proteins

The glycoprotein standard alpha-1-acid glycoprotein (AGP, G9885) was purchased from Millipore Sigma (St. Louis MO). Spike RBD proteins expressed in HEK 293 cells were purchased from two sources: Sino Biological (Beijing, China) and RayBiotech (Atlanta, GA, USA). A wild type version of the RBD (SARS-CoV-2 spike protein amino acids 319-541) was purchased from both sources expressed with a C-terminal poly-histidine tag. Additionally, the N501Y (stronger ACE-2 binding)³⁵ mutant from Sino Biological and the N331Q mutant (removes the N331 glycosite) from RayBiotech were purchased for comparison. Ammonium acetate, tris(2-carboxyethyl)phosphine hydrochloride (TCEP), 1,4-Dithiothreitol (DTT), and iodoacetamide were purchased from Sigma (St. Louis, MO, USA). Peptide -N-Glycosidase F (PNGase F, P0704S) was acquired from New England Biolabs (Ipswich, MA, USA).

Sample preparation for intact protein analysis

AGP and the RBDs from RayBiotech came as lyophilized powders and were diluted to a starting concentration of 1 mg/mL with deionized water. The RBDs from Sino Biological were also lyophilized but were diluted to a starting concentration of 0.25 mg/mL following the manufacture's recommendations. Samples were then buffer exchanged into 100 mM ammonium acetate using Zeba 7 kDa desalting columns (Thermo Fisher) that were equilibrated with ammonium acetate. All samples were vialled at 0.25 mg/mL in 100mM ammonium acetate for MS analysis. Further sample preparation for PNGase F

treatment of RBDs as well as glycopeptide and released glycan analysis can be found in the Supplementary Methods.

Online Liquid chromatography for Intact protein mass spectrometry

Online RPLC and HILIC separations were performed on a Waters NanoAcquity UPLC with dual pump trapping mode. The nanoflow C2 column (100 μm ID capillary) was packed in-house. The HILIC column used was an acrylamide-based polymer monolith stationary phase polymerized in a 200 μm ID capillary prepared following the procedure described³⁶. Both C2 and HILIC utilized an online desalting C2 column prior to the analytical separation. Online CZE separation was performed using a CMP Scientific (Brooklyn, NY) EVE-001 capillary electrophoresis autosampler using a proprietary 100 cm PS2 coated capillary (cat: E-SC-PS2-360/150-50-100-B1) from CMP Scientific using denaturing BGE conditions. Further details for online separation can be found in the Supporting Information Methods.

Intact and top-down mass measurements were performed on a Thermo Fisher Orbitrap Eclipse tribrid, Thermo Exploris 480, or Thermo Q-Exactive HF mass spectrometers. The nanoelectrospray source was set to 1.8-2.0 kV, with the transfer tube at 305 °C, source fragmentation voltage set to 35 V for HILIC experiments and 15 V for C2 RPLC and CZE experiments. The RF lens was set at 70% for experiments. MS1 spectra were acquired with a mass range of 600–6000 m/z at 7500 resolution (at m/z of 400), AGC target of 8E5, maximum injection time of 200 ms, and 5 microscans. The Eclipse and Exploris were set to intact protein mode with low pressure. The Sino Biological WT RBD was run in triplicate per a separation method tested while the other RBDs were run in triplicate using HILIC separation.

Data analysis of intact mass spectra

All raw mass spectra used for comparing separation techniques were deconvolved to zero-charge spectra and output as a matrix of mass, abundance, and elution time slice using Protein Metrics

(Cupertino, CA) Intact Mass software (version 4.2)³⁷ with default settings. An R script was implemented (Version 4.0.2) to compare the list of deconvoluted masses from each separation method and to remove mass peaks not observed in triplicate using a mass tolerance of ± 2 Da. Further details of this peak filtering approach can be found in the Supporting Information methods. The R source code and deconvoluted peak data can be found in <https://github.com/EMSL-Computing/RBD-intact-peak-analysis>.

Results and Discussion:

HILIC provided the highest degree of separation for resolving the heterogeneity of RBD glycoforms by separating neutral glycans

To benchmark the separation methods for glycoform separation we used AGP (Orosomuroid 1) as a mammalian glycoprotein standard. Almost half of the mass of the ~40 kDa AGP is from extensive N-glycosylation (5x N-glycan sites) with a high degree of sialic acid incorporation.³⁸ Overall, for AGP, both CZE and HILIC demonstrated greater capacities to separate glycoforms than C2 RPLC (Figure S1). In contrast, C2 RPLC separation was mostly between the sequence variants of ORM1*F1 and ORM1*S (Figure S2). CZE separated AGP charge variants well based on the degree of sialylation per a glycoform (Figure S3) while HILIC separation demonstrated a trend of increasing glycosylation with retention time (mostly based on neutral glycans). These observations were consistent with previous studies.^{28, 39}

Variable amounts of sialic acid groups on the O-and the N-glycan sites have been reported on the RBD.^{20, 33, 34} Separation techniques to resolve these complex RBD glycoforms have not been systematically evaluated to the best of our knowledge. One might expect CZE to separate the RBDs well since there is some variance in charged sialic acid, while HILIC may be able to separate the more neutral N-glycans based on total glycan composition. Figure 1A-C displays the C2 RPLC, CZE, and monolithic HILIC separation for the Sino Biological WT RBD with total ion chromatograms (TICs), and extracted ion chromatograms (XICs) of selected RBD glycoforms that only differ in their glycan compositions. The

glycoform assignments were based on additional top-down, glycoproteomic, and released glycan data, which will be discussed in later sections. Minor separation was seen in both C2 RPLC and CZE for the selected high abundance RBD glycoforms (Figure 1A-B). In contrast, HILIC produced clear chromatographic separation of RBD glycoforms, with resolvable peaks even in the TIC (Figure 1C). The limited CZE separation may be due to the small degree of heterogeneity in sialic acid seen in recent reports of recombinant RBDs. The T323 O-glycan site is decorated with 1-2 sialic acids (the most abundant O-glycan has 2 sialic acids) with Core 1 or Core 2 O-glycan structures.^{33, 34} The identified N-glycans are complex and have between 0-3 sialic acids.^{20, 33} This combination of O- and N-glycan compositions leads to the majority of RBDs containing 2-4 sialic acids,³³ a narrower distribution than seen with proteins that separate well with CZE such as AGP (13-19 sialic acids).

As seen here, and previously, HILIC has a high capacity to separate neutrally charged glycans;^{28, 30} leading to a broader elution time profile for RBD glycoforms than observed with C2 RPLC or CZE (Figure 1C). From the monolithic HILIC elution profile alone, differences in N-glycan occupancy become readily apparent, which would otherwise be easily missed with CZE or C2 RPLC. Figures 1D-F display the intact mass distributions for the Sino Biological WT RBD at three different points along the HILIC elution (marked in Figure 1C). The first peak eluting at 19 min (Figure 1C) contains an RBD species weighing 29830.3 Da (Figure 1D), while the main elution peak at ~21 min weighs 31974.3 Da (Figure 1E), a mass difference of 2144 Da. This mass shift of 2144 Da readily corresponds to the addition of the N-glycan group with the composition H₃N₆F₃ (hexose (H), N-acetylhexosamine (N), fucose (F), sialic acid (S)), which was identified previously from glycoproteomics²⁰ and is observed in our released glycan and glycoproteomics data. Additionally, repeated peak spacing of 41 Da is observed in Figures 1D-E that can be attributed to the exchange of a H₁ for N₁. Late eluting species (Figure 1F) display abundant peak spacing of 146 Da that matches increasing amounts of fucose or sialic acid. Two fucose units weigh only

1 Da more than one sialic acid group, causing mass degeneracy that is difficult to disentangle due to

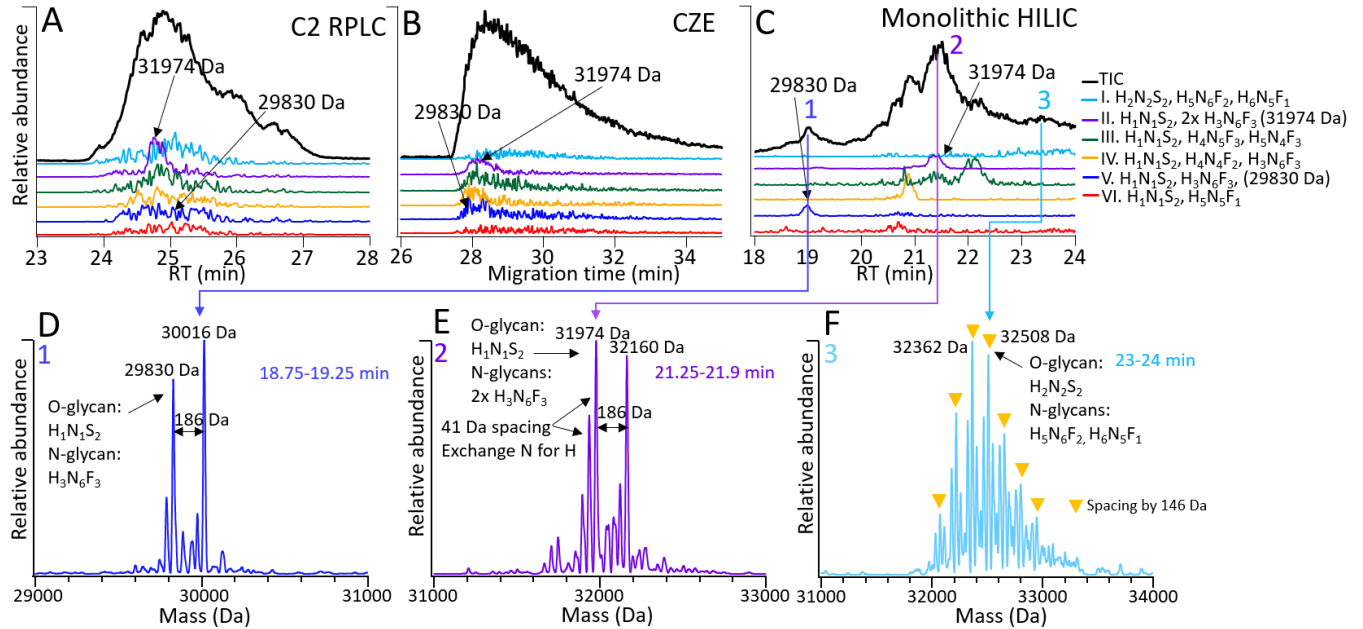


Figure 1. Sino Biological WT RBD intact glycoform separation comparison between C2 RPLC (A), CZE (B), monolithic HILIC (C). Each XIC corresponds to a defined glycoform of the RBD (charge state 16+ except for the lightest two glycoforms with one N-glycan are 15+). HILIC separated the RBD glycoforms the best by N-glycan occupancy and glycan composition. (D-F) Intact mass distributions for the RBD at given HILIC elution times matching 3 selected elution peaks marked in (C). (D) displays part of the RBD population with 1x N-glycan occupied on sites N331 or N343. The mass shift from the most abundant peaks from (D) to (E) (labeled elution peaks 1 and 2) correspond to 2144 Da adding the N-glycan $H_3N_6F_3$. The 186 Da spacing observed does not match a glycan mass but is an additional unknown modification. (F) Late eluting RBD species (labeled elution peak 3) display mass peak spacing of 146 Da suggesting increasing fucosylation and sialic acid. The addition of two fucoses is 1 Da heavier than adding one sialic acid. Labeled glycan compositions are based on matching top-down and released N-glycan data. Corresponding glycan key: hexose (H), N-acetylhexosamine (N), fucose (F), and sialic acid (S).

peak overlap and limited mass resolution. Overall, HILIC provides the broadest separation of RBD glycoforms based on the extent of glycosylation (increasing retention with the extent of glycosylation).

HILIC separation reduced spectral congestion and improved detection of low abundance glycoforms

A significant challenge in MS of heterogeneous macromolecules (such as glycoproteins) remains in the effective charge state determination prior to accurate mass determination by deconvolution. The complexity may arise from the lack of mass resolution, and high spectral baselines due to peak

coalescence. In addition to the advances of new algorithms⁴⁰ and instruments (e.g. charge detection MS),⁴¹⁻⁴³ online separations also play essential roles in reducing sample complexity. Figure 2 displays the deconvolved intact mass distributions with each tested RBD using HILIC separation. Each colored trace corresponds to the intact mass distribution at a given apex elution time from the HILIC separation. With each RBD the intact mass increases with retention time. Visual comparison of the intact mass profiles between RBDs immediately suggests the glycan compositions for the Sino Biological WT and N501Y RBDs (Figure 2A, D) are different from that of the RayBiotech WT RBD (Figure 2B). Importantly, the two Sino Biological RBDs (WT and N501Y) have very similar intact mass profiles that are only shifted by the N to Y amino acid substitution (49.07 Da). The RayBiotech WT RBD has a broader intact mass profile than the Sino Biological RBDs that are centered around a few high-abundance mass peaks. The N331Q mutant (removes N331 glycosite, Figure 2D) displayed a simpler intact mass profile with reduced glycosylation.

With each RBD, more than 200 peaks were detected due to the number of different glycoforms, creating a broad and congested mass distribution that is challenging to interpret. The peak spacing can correspond to adding certain glycan units (Figure 2B, peaks 31821 and 32024), but often the relation between mass peaks is unclear. Using the Ray Biotech WT RBD as an example that had the most complex glycosylation pattern among the 4 RBDs, Figure 3A displays the TIC for the HILIC separation with selected mass spectra of RBD species from the elution (Figure 3B). Summing the elution window together produced a heavily congested mass spectrum (Figure 3C). Some of the lighter RBD species for this sample (red and blue mass spectral traces in Figure 3B) are easily lost without the use of the HILIC glycoform separation due to the high baseline and overlapping charge state distributions observed. For instance, the 12+ RBD species weighing 29273.3 Da (retention time 16.5 min) overlaps with heavier 13+ RBD species weighing ~31570 Da that elute later (retention time 19.5 min). However, taking 30 second

windowed slices along the elution drastically reduces the spectral complexity before performing mass deconvolution, which eases detection of the lighter RBD species.

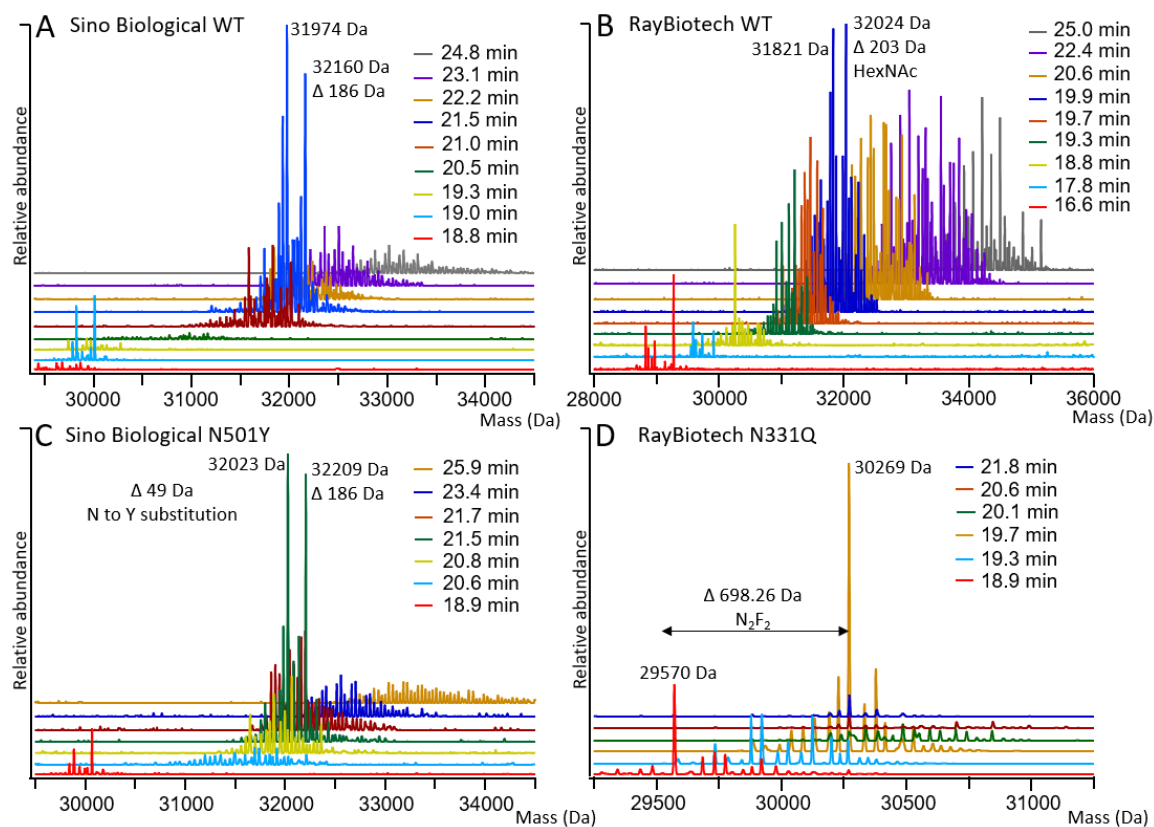


Figure 2. Deconvoluted intact mass analysis of tested RBDs separated with monolithic HILIC across the elution profile with apex times given for each elution slice. (A) Sino Biological WT RBD. Same raw data as shown in Figure 1. (B) RayBiotech WT RBD. (C) Sino Biological N501Y RBD. (D) RayBiotech N331Q RBD. HILIC separated slices are represented by the overlaid colors to combine the full mass distribution without peak filtering. The mass distributions of the WT RBDs are drastically different between the two vendors, while the N501Y mutant is simply shifted by the mass of the mutation in comparison to the WT RBD from the same vendor.

This spectral complexity inevitably resulted in variability and artifact peaks during deconvolution. We thus implemented a peak filtering approach to ensure the consistency of results (more details in the Supporting Methods section). In essence, elution time slices (18 total) for each separation method were deconvoluted separately. The resultant mass lists were then merged and filtered by removing mass peaks not observed in three technical replicates with a tolerance of ± 2 Da. A representative example of the deconvolution results after peak filtering for the Sino Biological WT is

plotted in Figure S4 with each separation method and compares the observed intact masses. This approach combined the intact mass distribution from triplicates and reduced the influence of noise in the deconvolution when comparing different samples or conditions. HILIC separation reproducibly detected the highest number of proteoforms (261 peaks), including low abundance species, in comparison to C2 RPLC (129 peaks) and CZE (177) (Figure S4C). It is noted that the filtering step kept ~10% of the total peaks, while ~70% and ~20% peaks showed up only in one and two replicates, respectively. The average median abundances of peaks observed in all triplicates were consistently higher than peaks observed in one or two replicates for all separation methods, suggesting higher variability in detection of low-abundance species (Figure S5). Using this peak filtering approach, we compared the total numbers of peaks after deconvolution using the HILIC separation across time slices (same as Figure 3B) vs. summing across the full elution window (same as Figure 3C) for all four RBDs (Figure S6A-D). Not surprisingly, separation increased the number of peaks by at least ~3 fold (Figure S6E) and showed the lowest median abundances of the detected peaks (Figure S6F). Overall, the windowed elution slices and peak filtering approach used here best utilizes the HILIC glycoform separation to detect the greatest number of proteoforms, while reducing the influence of noise from the deconvolution process from spectral congestion issues.

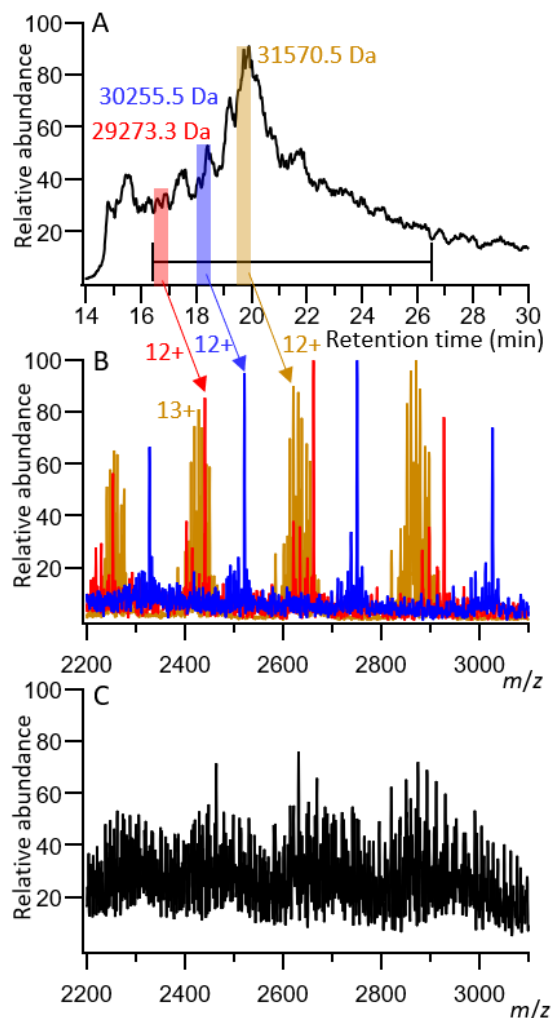


Figure 3. Example chromatogram (A) and overlaid mass spectra (B) from the RayBiotech WT HILIC elution. Using 30 second windowed slices (colored segments and traces, red: 16.5-17 min, blue: 18-18.5 min, and gold: 19.5-20 min) from the elution aids resolution of more RBD glycoforms due to the reduction in spectral congestion in comparison to summing the full elution window together (C) (16.5-26.5 min) where overlapping charge state distributions can be observed. For example, the 12+ RBD species weighing 29273.3 Da (most abundant) overlaps with heavier 13+ RBD species weighing ~31500 Da.

HILIC top-down analysis reported more low-abundance, heavily glycosylated proteoforms than what were predicted from glycopeptide data

From the intact mass measurements alone, initial fitting of the masses to previously reported O- and N-glycans proved very difficult. The measured masses were 206 Da heavier for the Sino Biological and 733 Da heavier for the RayBiotech RBDs than expected (after considering the known C-terminal

poly-histidine affinity tag), suggesting additional modifications. To better confirm the protein sequence, the N-glycans were removed with PNGase F, the protein was reduced, and then denatured to produce O-glycoforms for direct infusion top-down analysis, similar to the approach used by Roberts et al.³⁴ Removal of N-glycosylation significantly reduced spectral complexity and improved fragmentation for sequencing the protein backbones. After accounting for the expected sequence mass (including purification tags) and the O-glycoforms, the RayBiotech WT and N331Q RBDs were still 613.0 and 610.3 Da heavier than expected, respectively (Figure S7). Manual *de novo* sequencing of the N-terminus for the RayBiotech WT RBD by electron transfer dissociation (ETD) fragmentation (Figure S8),⁴⁴ determined that five amino acids with the sequence, KSMHM (weighing 614.77 Da), from the partially cleaved signaling peptide remained. Top-down with higher-energy collisional dissociation (HCD) for each RBD produced only C-terminal and glycan fragments that matched to the expected O-glycoform isolated but did not reveal the N-terminus (Figure S9), possibly due to the complication of O-glycosylation. Using the same approach, N-terminal ETD fragments for the Sino Biological WT (Figure S10) explained the extra 85.1 Da by an additional serine residue remained from the signaling peptide sequence, as was previously reported for this RBD.³⁴ We also identified potential cysteinylolation at C538 based on non-reducing intact mass analysis after N-glycan removal (Figure S11) and known information in recent reports.³³ Additionally, we detected an unknown 186 Da mass shift unique to the two Sino Biological RBDs (Figure S11C-D). This mass shift was present in the intact RBDs (without any treatment, Figure 1-2), but lost with TCEP reduction. It cannot be readily assigned to known N- or O-glycans but could be a non-covalent adduct protected by disulfides or an unknown covalent modification linked to a cysteine residue. The identity of this modification remains unclear but must be accounted for in the intact mass distribution.

Together, all the RBDs studied here had additional N-terminal residues added to the expected RBD sequence from the signaling peptide and cysteinylolation, consistent with other reports by Gstöttner et al.³³, and Roberts et al.³⁴ Notably, standard peptide mapping by bottom-up analysis produced 100%

and 92.1% sequence coverage for the RayBiotech WT and Sino Biological WT, respectively, (Figure S12 and S13) without considering the modified N-terminal sequence. Since the expected N-termini begin with arginine, additional preceding residues are easily missed due to the trypsin cleavage. Despite the relatively short sequence of the signaling peptides at N-termini, variable N-terminal cleavage points have been observed that leave additional residues, which could complicate RBD-based seropositivity assays if recombinant RBDs with variable N-termini are used.⁴⁵ Therefore, complete characterization of the sequence, especially with the power of top-down measurements can be important for quality control of RBDs.

After quantifying intact O-glycoforms and unexpected PTMs of all the four RBDs from the intact mass and top-down data, we examined bottom-up glycopeptide (Figure S14-S17) and released glycan data (Figure S18-S19, Table S1-S4) regarding their coverage of glycosylation. Overall, released N-glycans showed similar glycan profiles to glycopeptide data but captured few N-glycans with sialic acids, possibly due to the labile nature or detection bias of sialic acid groups.^{46, 47} Thus, we focused on combining the RBD intact masses after N-glycan removal (O-glycoforms) with the N-glycopeptide data for reconstruction of the mass distributions for each RBD by adapting the method reported by Yang et al.⁴⁸ Figure 4 plots the intact mass profile after peak filtering (top, black trace) with the corresponding reconstruction (bottom, red trace). Matching the reconstruction allowed us to attach assignments to at least half of the peaks in the filtered intact mass distributions including the selected glycoforms in Figure 1.

With all the RBDs investigated, the RayBiotech N331Q RBD (Figure 4D), which was the least heterogeneous sample, had the best fit between the reconstruction and the experimental data with a Pearson correlation of 0.83 and 52 assignable peaks in the reconstruction out of 107 experimental peaks matching within ± 2 Da. The RayBiotech WT RBD had a correlation of 0.76 and matched 98 of 150 peaks. Similarly, the Sino Biological RBDs had correlations of 0.74 and 0.67 with 210/261 and 160/202 mass

peaks matching the reconstruction for the WT and N501Y RBDs, respectively. Despite the overall agreements, a significant amount of higher mass peaks (>33 kDa and up to ~30% relative abundance) in the experiment were unaccounted for in the reconstruction for the RayBiotech WT RBD (Figure 4B). Underestimations of high mass species were also seen in the two Sino Biological RBDs (the tails in the mass distribution >33 kDa in Figure 5A, C), but those species were at much lower abundance (up to ~5%). This discrepancy could be attributed to the detection biases in the glycopeptide analysis.

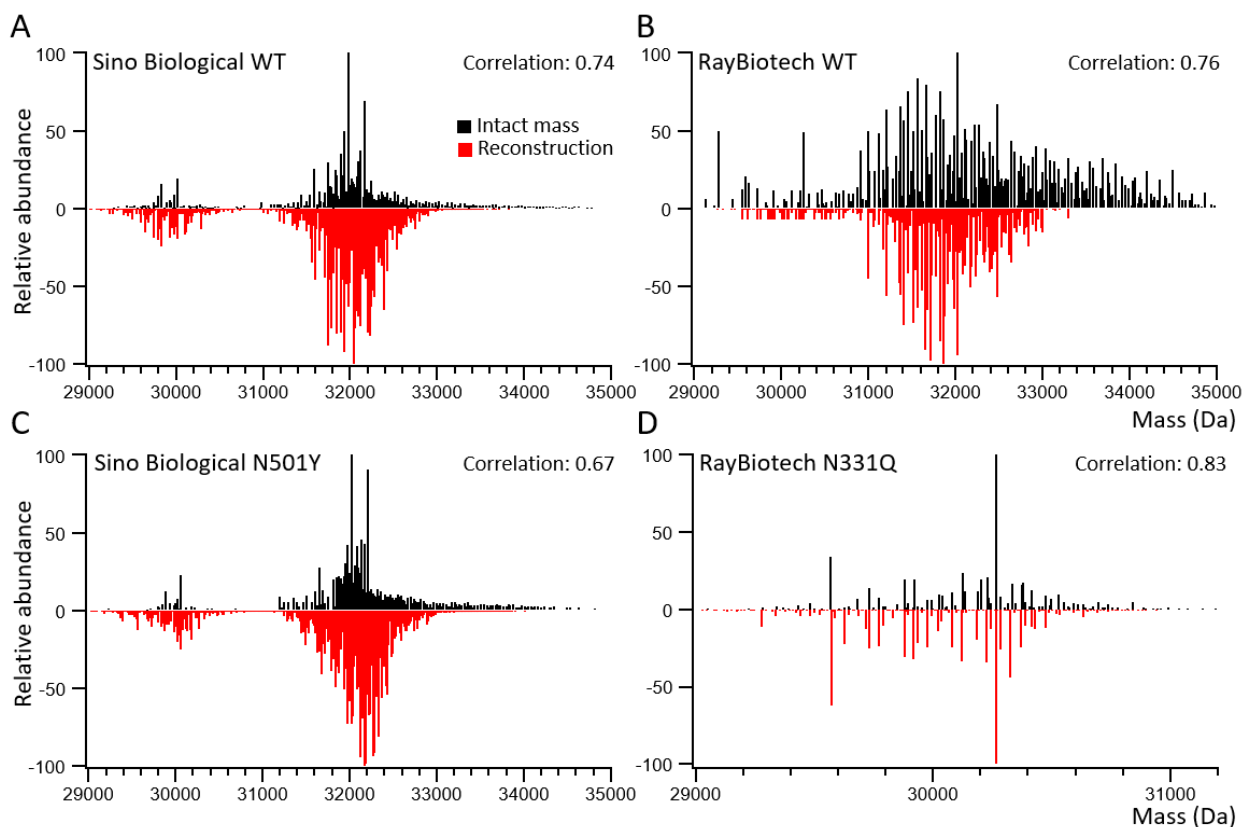


Figure 4. Comparison of HILIC separated deconvolved RBD mass distributions after peak filtering (top, black trace) with the reconstructed mass distribution from top-down and glycopeptide N-glycan analysis (bottom, red trace) for each RBD (A-D). The Pearson correlation between the intact mass spectrum and the reconstruction is given for every RBD. Note: For the Sino Biological RBDs only glycosite N331 was observed to be sometimes unoccupied. The relative abundances of unoccupied glycosites were estimated to 8% based on the best fit to the intact mass distribution as described in Supporting Method.

It is known that glycosylation can reduce the ionization efficiency of glycopeptides relative to un-glycosylated peptides, which complicates quantitative glycan analysis without derivitization.^{49, 50}

Additionally, sialic acid is known to substantially shift the retention time of glycopeptides where more heavily sialylated peptides elute later in reverse-phase chromatography. Sialic acid also can undergo more substantial in-source neutral fragmentation than other glycan groups.⁵¹ These missing glycoforms in the reconstruction likely have even larger N-glycans with more sialic acid and/or fucose present that are not readily detected from the glycopeptide or released glycan analysis. For instance, the highest matched reconstruction and experimental mass peak for the RayBiotech WT RBD weighs 33296 Da (O-glycan: H₂N₂S₂, combined N-glycans: H₁₀N₁₀F₂S₃). Additional higher mass peaks of 33587 and 33878 Da are observed in intact mass data that are each spaced by sialic acid (291 Da), and thus supporting this possibility of suppressed signal in glycoproteomics data. In native or denaturing intact mode MS, ionization efficiency is largely driven by the protein backbone, which alleviates the ionization bias based on the extent of glycosylation.⁵²

We also noted that both Sino Biological RBDs showed more heterogenous glycoform distributions (~3 times more peaks) in the reconstructions than in the experimental intact mass distribution. Herein, we collected dual enzyme digestion data to separately define the glycosylation on N331 and N343 for the two Sino Biological RBDs. We suspect the N-glycans on these two glycosites were likely correlated and not randomly combined as assumed in the reconstruction (i.e., connectivity between the micro-heterogeneity). In our RayBiotech WT RBD data, the N331 and N343 glycans were defined on a single tryptic peptide, maintaining the native connectivity between the two sites, and therefore yielding a more similar reconstruction to the experimental spectrum. While the connectivity and variation of glycosylation across multiple sites (i.e. meta-heterogeneity) requires further confirmation, our preliminary analysis showed that a top-down framework with HILIC glycoform separation can be highly beneficial for defining the combinations of glycosylation with reduced detection biases.

Conclusions:

Here, we demonstrated that online monolithic HILIC provided improved detection of RBD intact glycoforms compared to C2 RPLC and CZE, attributable to HILIC's superior separation of neutral glycans.^{28, 30} While all separation methods helped reduce spectral congestion, the heterogeneity of the glycoproteins still caused remarkable variations in detection of glycoforms even among technical replicates. The peak filtering approach used here maintained reproducible RBD proteoforms while limiting the influence of noise with these heterogeneous and difficult to study proteins. The intact mass profile of the detected glycoforms provides a rapid assessment of the integrity of the RBDs, and readily revealed unexpected mass shifts from PTMs and sequence variations. When combined with top-down, glycopeptide, and released glycan analyses, up to 75% of the intact masses could be assigned based on the computationally reconstructed mass profiles integrating all the data. Interestingly, our reconstruction based on glycopeptide data generally showed fewer low abundance, and/or heavily glycosylated species than the top-down data, especially for the more heterogeneous RBDs. Such discrepancies were consistently reported in several recent studies of RBD/spike glycosylation,^{33, 43} and with other glycoprotein analysis,⁴⁷ to differing extents. Given the known experimental biases in glycopeptide analysis, incorporating top-down data will be highly valuable for more complete characterization of glycosylation.

While it remains technically challenging to directly assign individual glycoforms by online top-down fragmentation, the continuing advances in MS instrumentation methods such as proton transfer charge reduction^{53, 54} and charge detection mass spectrometry⁴³ will likely allow more comprehensive analysis of heterogeneous glycoproteins in the near future. New online separation modalities such as the HILIC method described here will also be indispensable for reducing sample complexity prior to MS analysis. Currently, HILIC separation of intact glycoproteins is less common in comparison to the use of HILIC for glycopeptide or released glycan separation, however HILIC has shown great promise in the separation of glycoproteins that vary in glycan occupancy and the amount of neutral glycans.^{28, 30, 55}

Given the importance of protein glycosylation in modulating immune responses^{12, 13} and many other biological processes, we envision HILIC coupled to MS will have great potential in the characterization of heterogeneous glycoprotein products.

Acknowledgements

A portion of this research was performed on a project award 10.46936/intm.proj.2020.51671/60000250 from the Environmental Molecular Sciences Laboratory (EMSL), a DOE Office of Science User Facility sponsored by the Biological and Environmental Research program under Contract No. DE-AC05-76RL01830. This research was also supported by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. A portion of this work was supported by Congressionally Directed Medical Research program award #PR201356 to VRG. M.P. acknowledges the European Research Council (ERC), Project 694151, STAMP, for funding.

We would like to thank David S. Roberts and Ying Ge for helpful discussions and for data comparison; Thomas Fillmore and Ronald Moore for liquid chromatography and instrumental tuning conditions; Matthew Monroe for data deposition.

Conflict of interest

The authors declare no competing financial interest.

Data Availability Statements

Raw MS data for online RBD glycoform separations, glycopeptide analysis, released glycan, and top-down experiments after N-glycan removal are available at MassIVE with accession MSV000088601 (FTP password: HEK3874).

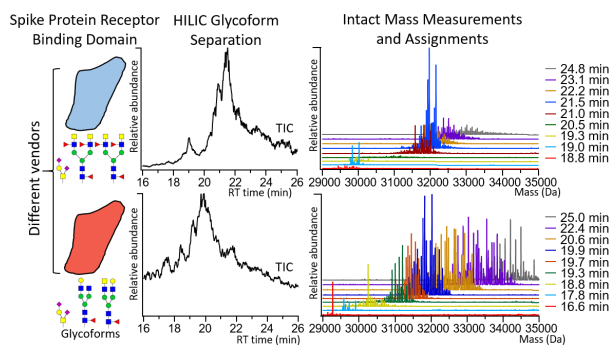
References

1. Du, L.; He, Y.; Zhou, Y.; Liu, S.; Zheng, B.-J.; Jiang, S., The spike protein of SARS-CoV — a target for vaccine and therapeutic development. *Nature Reviews Microbiology* **2009**, *7* (3), 226-236.
2. Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E., Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **2020**, *6* (10), 1722-1734.
3. Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M., Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369* (6501), 330-333.
4. Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J., Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Scientific Reports* **2020**, *10* (1), 14991.
5. Choi, Y. K.; Cao, Y.; Frank, M.; Woo, H.; Park, S.-J.; Yeom, M. S.; Croll, T. I.; Seok, C.; Im, W., Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane. *Journal of Chemical Theory and Computation* **2021**, *17* (4), 2479-2487.
6. Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **2020**, *581* (7807), 215-220.
7. Premkumar, L.; Segovia-Chumbez, B.; Jadi, R.; Martinez, D. R.; Raut, R.; Markmann, A. J.; Cornaby, C.; Bartelt, L.; Weiss, S.; Park, Y.; Edwards, C. E.; Weimer, E.; Scherer, E. M.; Rouphael, N.; Edupuganti, S.; Weiskopf, D.; Tse, L. V.; Hou, Y. J.; Margolis, D.; Sette, A.; Collins, M. H.; Schmitz, J.; Baric, R. S.; de Silva, A. M., The receptor-binding domain of the viral spike protein is an immunodominant and highly specific target of antibodies in SARS-CoV-2 patients. *Science Immunology* **2020**, *5* (48), eabc8413.
8. Qi, X.; Ke, B.; Feng, Q.; Yang, D.; Lian, Q.; Li, Z.; Lu, L.; Ke, C.; Liu, Z.; Liao, G., Construction and immunogenic studies of a mFc fusion receptor binding domain (RBD) of spike protein as a subunit vaccine against SARS-CoV-2 infection. *Chemical Communications* **2020**, *56* (61), 8683-8686.
9. Houlihan, C. F.; Beale, R., The complexities of SARS-CoV-2 serology. *The Lancet Infectious Diseases* **2020**, *20* (12), 1350-1351.
10. Bouwman, K. M.; Tomris, I.; Turner, H. L.; van der Woude, R.; Shamorkina, T. M.; Bosman, G. P.; Rockx, B.; Herfst, S.; Snijder, J.; Haagmans, B. L.; Ward, A. B.; Boons, G.-J.; de Vries, R. P., Multimerization- and glycosylation-dependent receptor binding of SARS-CoV-2 spike proteins. *PLOS Pathogens* **2021**, *17* (2), e1009282.
11. Zhao, X.; Chen, H.; Wang, H., Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Frontiers in Molecular Biosciences* **2021**, *8* (53).
12. Li, J.; Hsu, H.-C.; Mountz, J. D.; Allen, J. G., Unmasking Fucosylation: from Cell Adhesion to Immune System Regulation and Diseases. *Cell Chemical Biology* **2018**, *25* (5), 499-512.
13. Rabinovich, G. A.; van Kooyk, Y.; Cobb, B. A., Glycobiology of immune responses. *Annals of the New York Academy of Sciences* **2012**, *1253* (1), 1-15.
14. Gupta, S. K.; Shukla, P., Glycosylation control technologies for recombinant therapeutic proteins. *Applied Microbiology and Biotechnology* **2018**, *102* (24), 10457-10468.
15. Aghamohseni, H.; Ohadi, K.; Spearman, M.; Krahn, N.; Moo-Young, M.; Scharer, J. M.; Butler, M.; Budman, H. M., Effects of nutrient levels and average culture pH on the glycosylation pattern of camelid-humanized monoclonal antibody. *Journal of Biotechnology* **2014**, *186*, 98-109.
16. Kotidis, P.; Kontoravdi, C., Harnessing the potential of artificial neural networks for predicting protein glycosylation. *Metabolic Engineering Communications* **2020**, *10*, e00131.

17. Čaval, T.; Heck, A. J. R.; Reiding, K. R., Meta-heterogeneity: Evaluating and Describing the Diversity in Glycosylation Between Sites on the Same Glycoprotein. *Molecular & Cellular Proteomics* **2021**, *20*, 100010.
18. Camperi, J.; Pichon, V.; Delaunay, N., Separation methods hyphenated to mass spectrometry for the characterization of the protein glycosylation at the intact level. *Journal of Pharmaceutical and Biomedical Analysis* **2020**, *178*, 112921.
19. Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P., Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **2020**, *30* (12), 981-988.
20. Antonopoulos, A.; Broome, S.; Sharov, V.; Ziegenfuss, C.; Easton, R. L.; Panico, M.; Dell, A.; Morris, H. R.; Haslam, S. M., Site-specific characterization of SARS-CoV-2 spike glycoprotein receptor-binding domain. *Glycobiology* **2020**, *31* (3), 181-187.
21. Yang, X.; Bartlett, M. G., Glycan analysis for protein therapeutics. *Journal of Chromatography B* **2019**, *1120*, 29-40.
22. Donnelly, D. P.; Rawlins, C. M.; DeHart, C. J.; Fornelli, L.; Schachner, L. F.; Lin, Z.; Lippens, J. L.; Aluri, K. C.; Sarin, R.; Chen, B.; Lantz, C.; Jung, W.; Johnson, K. R.; Koller, A.; Wolff, J. J.; Campuzano, I. D. G.; Auclair, J. R.; Ivanov, A. R.; Whitelegge, J. P.; Paša-Tolić, L.; Chamot-Rooke, J.; Danis, P. O.; Smith, L. M.; Tsybin, Y. O.; Loo, J. A.; Ge, Y.; Kelleher, N. L.; Agar, J. N., Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nature Methods* **2019**, *16* (7), 587-594.
23. Füssl, F.; Strasser, L.; Cari, S.; Bones, J., Native LC-MS for capturing quality attributes of biopharmaceuticals on the intact protein level. *Current Opinion in Biotechnology* **2021**, *71*, 32-40.
24. Campuzano, I. D. G.; Sandoval, W., Denaturing and Native Mass Spectrometric Analytics for Biotherapeutic Drug Discovery Research: Historical, Current, and Future Personal Perspectives. *Journal of the American Society for Mass Spectrometry* **2021**.
25. D'Atri, V.; Murisier, A.; Fekete, S.; Veuthey, J.-L.; Guillarme, D., Current and future trends in reversed-phase liquid chromatography-mass spectrometry of therapeutic proteins. *TrAC Trends in Analytical Chemistry* **2020**, *130*, 115962.
26. Gomes, F. P.; Yates III, J. R., Recent trends of capillary electrophoresis-mass spectrometry in proteomics research. *Mass Spectrometry Reviews* **2019**, *38* (6), 445-460.
27. Shi, R. L.; Xiao, G.; Dillon, T. M.; Ricci, M. S.; Bondarenko, P. V., Characterization of therapeutic proteins by cation exchange chromatography-mass spectrometry and top-down analysis. *MAbs* **2020**, *12* (1), 1739825-1739825.
28. Domínguez-Vega, E.; Tengattini, S.; Peintner, C.; van Angeren, J.; Temporini, C.; Haselberg, R.; Massolini, G.; Somsen, G. W., High-resolution glycoform profiling of intact therapeutic proteins by hydrophilic interaction chromatography-mass spectrometry. *Talanta* **2018**, *184*, 375-381.
29. Gargano, A. F. G.; Roca, L. S.; Fellers, R. T.; Bocxe, M.; Domínguez-Vega, E.; Somsen, G. W., Capillary HILIC-MS: A New Tool for Sensitive Top-Down Proteomics. *Analytical Chemistry* **2018**, *90* (11), 6601-6609.
30. Gargano, A. F. G.; Schouten, O.; van Schaick, G.; Roca, L. S.; van den Berg-Verleg, J. H.; Haselberg, R.; Akeroyd, M.; Abello, N.; Somsen, G. W., Profiling of a high mannose-type N-glycosylated lipase using hydrophilic interaction chromatography-mass spectrometry. *Analytica Chimica Acta* **2020**, *1109*, 69-77.
31. Gargano, A. F. G.; Shaw, J. B.; Zhou, M.; Wilkins, C. S.; Fillmore, T. L.; Moore, R. J.; Somsen, G. W.; Paša-Tolić, L., Increasing the Separation Capacity of Intact Histone Proteoforms Chromatography Coupling Online Weak Cation Exchange-HILIC to Reversed Phase LC UVPD-HRMS. *Journal of Proteome Research* **2018**, *17* (11), 3791-3800.
32. Kohler, I.; Verhoeven, M.; Haselberg, R.; Gargano, A., Hydrophilic interaction chromatography – mass spectrometry for metabolomics and proteomics: state-of-the-art and current trends. *Microchemical Journal* **2021**, 106986.

33. Gstöttner, C.; Zhang, T.; Resemann, A.; Ruben, S.; Pengelley, S.; Suckau, D.; Welsink, T.; Wuhrer, M.; Domínguez-Vega, E., Structural and Functional Characterization of SARS-CoV-2 RBD Domains Produced in Mammalian Cells. *Analytical Chemistry* **2021**, *93* (17), 6839-6847.
34. Roberts, D. S.; Mann, M.; Melby, J. A.; Larson, E. J.; Zhu, Y.; Brasier, A. R.; Jin, S.; Ge, Y., 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* **2021**.
35. Tian, F.; Tong, B.; Sun, L.; Shi, S.; Zheng, B.; Wang, Z.; Dong, X.; Zheng, P., N501Y mutation of spike protein in SARS-CoV-2 strengthens its binding to receptor ACE2. *eLife* **2021**, *10*, e69091.
36. Passamonti, M.; de Roos, C.; Schoenmakers, P. J.; Gargano, A. F. G., Poly(acrylamide-co-N,N'-methylenebisacrylamide) Monoliths for High-Peak-Capacity Hydrophilic-Interaction Chromatography–High-Resolution Mass Spectrometry of Intact Proteins at Low Trifluoroacetic Acid Content. *Analytical Chemistry* **2021**.
37. Bern, M.; Caval, T.; Kil, Y. J.; Tang, W.; Becker, C.; Carlson, E.; Kletter, D.; Sen, K. I.; Galy, N.; Hagemans, D.; Franc, V.; Heck, A. J. R., Parsimonious Charge Deconvolution for Native Mass Spectrometry. *Journal of Proteome Research* **2018**, *17* (3), 1216-1226.
38. Baerenfaenger, M.; Meyer, B., Intact Human Alpha-Acid Glycoprotein Analyzed by ESI-qTOF-MS: Simultaneous Determination of the Glycan Composition of Multiple Glycosylation Sites. *Journal of Proteome Research* **2018**, *17* (11), 3693-3703.
39. Ongay, S.; Neusüß, C., Isoform differentiation of intact AGP from human serum by capillary electrophoresis–mass spectrometry. *Analytical and Bioanalytical Chemistry* **2010**, *398* (2), 845-855.
40. Rolland, A. D.; Prell, J. S., Approaches to Heterogeneity in Native Mass Spectrometry. *Chemical Reviews* **2021**.
41. Wörner, T. P.; Snijder, J.; Bennett, A.; Agbandje-McKenna, M.; Makarov, A. A.; Heck, A. J. R., Resolving heterogeneous macromolecular assemblies by Orbitrap-based single-particle charge detection mass spectrometry. *Nature Methods* **2020**, *17* (4), 395-398.
42. Kafader, J. O.; Melani, R. D.; Durbin, K. R.; Ikwuagwu, B.; Early, B. P.; Fellers, R. T.; Beu, S. C.; Zabrouskov, V.; Makarov, A. A.; Maze, J. T.; Shinholt, D. L.; Yip, P. F.; Tullman-Ercek, D.; Senko, M. W.; Compton, P. D.; Kelleher, N. L., Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes. *Nature Methods* **2020**, *17* (4), 391-394.
43. Miller, L. M.; Barnes, L. F.; Raab, S. A.; Draper, B. E.; El-Baba, T. J.; Lutomski, C. A.; Robinson, C. V.; Clemmer, D. E.; Jarrold, M. F., Heterogeneity of Glycan Processing on Trimeric SARS-CoV-2 Spike Protein Revealed by Charge Detection Mass Spectrometry. *Journal of the American Chemical Society* **2021**, *143* (10), 3959-3966.
44. Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolić, N.; Paša-Tolić, L.; Smith, R. D.; Payne, S. H.; Kim, S., Informed-Proteomics: open-source software package for top-down proteomics. *Nature Methods* **2017**, *14* (9), 909-914.
45. D'Ippolito, R. A.; Drew, M. R.; Mehalko, J.; Snead, K.; Wall, V.; Putman, Z.; Esposito, D.; DeHart, C. J., Refining the N-Termini of the SARS-CoV-2 Spike Protein and Its Discrete Receptor-Binding Domain. *Journal of Proteome Research* **2021**, *20* (9), 4427-4434.
46. Nie, H.; Li, Y.; Sun, X.-L., Recent advances in sialic acid-focused glycomics. *Journal of Proteomics* **2012**, *75* (11), 3098-3112.
47. Čaval, T.; Buettner, A.; Habegger, M.; Reusch, D.; Heck, A. J. R., Discrepancies between High-Resolution Native and Glycopeptide-Centric Mass Spectrometric Approaches: A Case Study into the Glycosylation of Erythropoietin Variants. *Journal of the American Society for Mass Spectrometry* **2021**, *32* (8), 2099-2104.

48. Yang, Y.; Liu, F.; Franc, V.; Halim, L. A.; Schellekens, H.; Heck, A. J. R., Hybrid mass spectrometry approaches in glycoprotein analysis and their usage in scoring biosimilarity. *Nature Communications* **2016**, *7* (1), 13397.
49. Stavenhagen, K.; Hinneburg, H.; Thaysen-Andersen, M.; Hartmann, L.; Silva, D. V.; Fuchser, J.; Kaspar, S.; Rapp, E.; Seeberger, P. H.; Kolarich, D., Quantitative mapping of glycoprotein micro-heterogeneity and macro-heterogeneity: an evaluation of mass spectrometry signal strengths using synthetic peptides and glycopeptides. *Journal of Mass Spectrometry* **2013**, *48* (6), 627-639.
50. Delafield, D. G.; Li, L., Recent Advances in Analytical Approaches for Glycan and Glycopeptide Quantitation. *Molecular & Cellular Proteomics* **2021**, *20*.
51. Medzihradzky, K. F.; Kaasik, K.; Chalkley, R. J., Characterizing Sialic Acid Variants at the Glycopeptide Level. *Analytical Chemistry* **2015**, *87* (5), 3064-3071.
52. Yang, Y.; Wang, G.; Song, T.; Lebrilla, C. B.; Heck, A. J. R., Resolving the micro-heterogeneity and structural integrity of monoclonal antibodies by hybrid mass spectrometric approaches. *MAbs* **2017**, *9* (4), 638-645.
53. Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J. B.; Fellers, R. T.; Zabrouskov, V.; Syka, J. E. P.; Kelleher, N. L.; Fornelli, L., Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Analytical Chemistry* **2019**, *91* (24), 15732-15739.
54. Stephenson, J. L.; McLuckey, S. A., Ion/Ion Proton Transfer Reactions for Protein Mixture Analysis. *Analytical Chemistry* **1996**, *68* (22), 4026-4032.
55. Camperi, J.; Combès, A.; Fournier, T.; Pichon, V.; Delaunay, N., Analysis of the human chorionic gonadotropin protein at the intact level by HILIC-MS and comparison with RPLC-MS. *Analytical and Bioanalytical Chemistry* **2020**, *412* (18), 4423-4432.



TOC Graphic