N and O glycosylation of the SARS-CoV-2 spike protein

Miloslav Sanda[1,3,*], Lindsay Morrison[4], Radoslav Goldman[1,2,3]

[1] Georgetown University, Department of Oncology, 3800 Reservoir Rd NW, Washington, D.C. 20057

[2] Georgetown University, Department of Biochemistry and Molecular & Cellular Biology, 3800 Reservoir Rd NW, Washington, D.C. 20057

[3] Georgetown University, Clinical and Translational Glycoscience Research Center, 3800 Reservoir Rd NW, Washington, D.C. 20057

[4] Waters Corporation Inc., Beverly, MA 01915 USA

* To whom correspondence should be addressed. Tel: +1 202-6879868; Fax: +1 202-6871988; Email: ms2465@georgetown.edu

Abstract

Covid-19 pandemic outbreak is the reason of the current world health crisis. The development of effective antiviral compounds and vaccines requires detailed descriptive studies of the SARS-CoV-2 proteins. The SARS-CoV-2 spike (S) protein mediates virion binding to the human cells through its interaction with the ACE2 cell surface receptor and is one of the prime immunization targets. A functional virion is composed of three S1 and three S2 subunits created by furin cleavage of the spike protein at R682, a polybasic cleavage sites that differs from the SARS-CoV spike protein of 2002. We observe that the spike protein is O-glycosylated on a threonine (T678) near the furin cleavage site occupied by core-1 and core-2 structures. In addition, we have identified eight additional O-glycopeptides on the spike glycoprotein and we confirmed that the spike protein is heavily N-glycosylated. Our recently developed LC-MS/MS methodology allowed us to identify LacdiNAc structural motifs on all occupied N-glycopeptides and polyLacNAc structures on six glycopeptides of the spike protein. In conclusion, our study substantially expands the current knowledge of the spike protein's glycosylation and enables the investigation of the influence of the O-glycosylation on its proteolytic activation.

Keywords

COVID19, SARS-CoV-2 spike glycoprotein, N-glycopeptide, structure specific LC-MS/MS glycoprotein analysis, O-glycosylation

Introduction

The World Health Organization was informed of pneumonia cases of unknown etiology in Wuhan, Hubei Province, China on 31 December 2019 (1). A novel coronavirus was identified as the cause of the disease by further investigations (2). This new virus is related to the previously identified SARS-CoV (severe acute respiratory syndrome coronavirus) and has been named SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2). Symptoms of the coronavirus disease 2019 (COVID-19) are acute onset of fever, myalgia, dyspnea, cough and evidence of ground-glass lung opacities. We do not have currently an effective vaccine or treatment for the COVID-19 patients and continued research is urgently needed to address the challenges posed by the pandemic.

Transmembrane spike (S) glycoprotein of the SARS-CoV-2 interacts with the angiotensin-converting enzyme 2 (ACE2) presented on the surface of human cells and mediates viral entry (3–5). Both the viral spike and the human ACE2 (hACE2) are glycoproteins and their glycosylation affects their interactions or vaccine design. Covid 19 spike glycoprotein forms a trimeric structure on the surface of the virus envelope (6). Each spike protein consists of an S1 and an S2 subunit; the S1 subunit mediates binding of the virus to the ACE2 receptor while the S2 subunit enables fusion of the virion with the cell membrane and initiates viral entry. SARS-CoV-2 has 10 to 20 times higher affinity for the ACE2 receptor than the SARS-CoV (3) which may be, in part, related to glycosylation of the proteins. SARS-CoV-2 S glycoprotein carries 22 N-glycosylation sequons

(6) and at least 3 sites of mucin-type O-glycosylation were predicted (7) but were not yet observed experimentally. The latest analysis shows that 20 out of the 22 N-glycosylation sequons are occupied by complex, hybrid and oligomannosidic structures. Some of the sequons are predominantly occupied by oligomannose structures which could have influence on the trimeric structure. The studies also detected one O-glycopeptide occupied at sites, distinct from the predicted furin cleavage site at the S1/S2 boundary (6, 8, 9).

In this study, we report analysis of the site-specific glycoforms with focus on the resolution of structural motifs of the identified O- and N- glycopeptides. To this end, we used high-resolution LC-MS/MS with HCD fragmentation and modulated NCE (10) to study a recombinant SARS-CoV-2 S full-length protein expressed in human embryonic kidney (HEK 293) cells. Our analyses identified 9 occupied O-glycopeptides and 17 N-glycopeptides. We resolved, for the first time, LacdiNAc and polyLacNAc structural motifs associated with the N-glycopeptides and we identified novel O-glycopeptides including a glycopeptide near the furin cleavage site of the spike glycoprotein.

Materials and Methods

Materials

Recombinant SARS-CoV-2 spike (R683A, R685A, His-tag) protein expressed in HEK 293 cell line was obtained from Acrobiosystems (Newark, DE, USA). Trypsin Gold and Glu-C, Sequencing Grade were from Promega (Madison, WI), PNGase F, Neuraminidase, 1-3 and 1-4 betagalactosidase were from New England Biolabs (Ipswich, MA).

Glycopeptide preparation

Aliquots of the SARS-CoV-2 S protein were dissolved in sodium bicarbonate buffer to a final concentration of 1mg/ml. The protein solution was reduced with 5 mM DTT for 60 min at 60 °C, alkylated with 15 mM iodoacetamide for 30 min in the dark, and digested with Trypsin Gold (2.5 ng/µl) at 37°C in Barocycler NEP2320 (Pressure BioSciences, South Easton, MA) for 1 hour. GluC, PNGase F, neuraminidase and beta galactosidase digests of tryptic peptides were carried out as described previously (11, 12) with heat inactivation (99 °C for 10 min) prior to the addition of any enzyme.

Glycopeptide analysis using DDA nano LC-MS/MS on the Orbitrap Fusion-Lumos

Digested proteins were separated using a 120-minute ACN gradient on a 250 mm x 75 µm C18 pepmap column at a flow rate of 0.3 µL/min as described previously (13). In brief, peptide and glycopeptide separation was achieved by a 5 min trapping/washing step using 99% solvent A (2% acetonitrile, 0.1% formic acid) at 10 µL/min followed by a 90 min acetonitrile gradient at a flow rate of 0.3 µL/min: 0-3 min 2% B (0.1% formic acid in ACN), 3-5 min 2-10% B; 5-60 min 10-45% B; 60-65 min 45-98% B; 65-70 min 98% B, 70-90 min equilibration by 2% B. Glycopeptides were analyzed using Orbitrap Fusion Lumos mass spectrometer with the electrospray ionization voltage at 3 kV and the capillary temperature at 275°C. MS1 scans were performed over m/z 400–1800 with the wide quadrupole isolation on a resolution of 120,000 (m/z 200), RF Lens at 40%, intensity threshold for MS2 set to 2.0e4, selected precursors for MS2 with charge state 3-8, and dynamic exclusion 30s. Data-dependent HCD tandem mass spectra were

collected with a resolution of 15,000 in the Orbitrap with fixed first mass 110 and 4 normalized collision energy 10, 20 and 35%. ETD and EThcD methods used calibrated charge dependent parameters and HCD supplemental activation was set to 15% NCE; we used the same chromatographic method and instrument settings for the ETD measurements as described above.

Glycopeptide analysis using cyclic ion mobility

LC-IM-MS/MS experiments were performed on a Waters Select Series cyclic ion mobility mass spectrometer with an ACQUITY M-Class solvent system. Tryptic peptides were separated using a 75 µm x 150 mm ACQUITY BEH C18 column with a 5 cm Symmetry C18 trap. Peptides were eluted over 60 minutes prior to electrospray ionization and analysis in positive mode. Glycoforms of the polybasic peptide were isolated in the quadrupole and fragmented in the trap region prior to ion mobility separations. Ion mobility methods entailing five passes of the cyclic device were previously optimized for HexNAcHex and HexNAcHexNeuAc oxonium ions and were used to separate and characterize the oxonium ion fragments of the targeted glycopeptides. Traveling wave parameters within the cyclic device were kept at default values, 375 m/s and 22V for the wave velocity and wave height, respectively. Calibration for collisional cross section was performed for both single pass and for each of the 5-pass methods using Major Mix.

Data analysis

Byonic software (protein metric) was used for the identification of summary formulas of glycans associated with the glycopeptides. Independent searches were performed on the data with different collision energy (CE) settings. All spectra of the identified glycopeptides were checked manually for the presence of structure-specific fragments. Analysis of the ion mobility data was performed

using DriftScope (v2.9) by manual extraction of the retention ranges associated with the glycopeptides.

Results and Discussion

N-glycopeptide analysis

We have identified 17 tryptic N-glycopeptides of the SARS-CoV-2 spike protein occupied by high mannose, hybrid and complex glycans. We have determined their site occupancy by PNGaseF deglycosylation in ^{18}O water as described (11) and we find majority of the sequons fully occupied (Table 1). We have found that two sequons are not glycosylated (N17, N603), that N234 is almost exclusively occupied by high mannose glycans. The remaining 17 sequons are dominated by complex glycans. In addition, we confirmed the presence of core fucosylation on 15 of the occupied sequons (Table 1).

Structural motifs of the N-glycans using modulated collision energy

We used our recently described workflows, using modulation of collision energy (CE) for selective fragmentation of the glycopeptides (10), to identify structural motifs of the N-glycosylated peptides of the SARS-CoV-2 S protein. We identified the LacdiNAc structural motif on all the occupied sequons of the SARS-CoV-2 S expressed in the HEK293 cells (Table 1). The low CE tandem mass spectrum (Figure 1) reveals structural features of an asymmetric LacdiNAc motif contained within a disialylated biantennary N-glycan. Presence of the m/z 366/407 ions distinguishes the LacNAc and LacdiNAc motifs; in addition, we observe the m/z 657/698 ions of

their sialylated counterparts. In addition to the fucosylated and/or sialylated LacdiNAc, we also identified polyLacNAc structures on 6 N-glycopeptides (Table 1) and we resolved extensive fucosylation of the core as well as the outer-arms of the N-glycopeptides as described previously (10). The presence of core fucose was confirmed on 15 sequons and we confirmed the presence of outer arm fucosylation (11, 14) on 7 sequons which is in contrast to the previously published data (8). This might be a result of slight differences in the HEK293 expression systems used or differences in the analytical methods. For example, our study analyzed a modified full-length protein not cleaved by convertases which could potentially cause some differences. It is, however, more likely that the energy optimized workflows improve the structural resolution. We do not achieve complete assignment of all linkages or quantification of the isobaric structures but the presence of these structural motifs, frequently associated with specific biological functions, is clearly established. The overall results show that 6 glycopeptides carry polyLacNac motifs, that all sequons occupied by complex glycans carry LacdiNAc to some degree, and that the LacdiNAc structures constitute majority (>50%) of the glycoforms on N165 and N1098. This may not necessarily reflect the N-glycoforms of a virion but the HEK293 expression system is commonly used for functional studies of the S glycoprotein or the production of vaccine candidates which means that resolution of the structures is highly relevant.

O-glycopeptide analysis

Previously published data describes one O-glycopeptide occupied at S323 and T325 (6, 8). We identified the same O-glycopeptides but, in addition, we have identified 8 O-glycopeptides occupied by core-1 and core-2 structures (Table 2). Occupancy of the sites varies from <1% to

57% and is very low for at least three of the glycopeptides. However, we detect approximately 13% occupancy with core-1 and core-2 structures at the T678 (Figure 2 and 3) located near the polybasic furin cleavage site between the S1 and S2 subunits which evolved in the SARS-CoV-2 S protein (7). This is relevant because O-glycans proximal to the convertase cleavage sites of protein substrates regulate proteolysis (15) and may regulate activation of the SARS-CoV-2 S protein. Supplemental Figure 1 documents identification of an O-glycopeptide following deglycosylation with PNGaseF. This is interesting because O-glycosylation in such a close proximity to N-glycosylation is rarely described; we do not know if this has any functional relevance but it shows that analysis of N-deglycosylated peptides for O-glycoforms may deserve attention. Retention times of the T687 O-glycoforms (Figure 4) follow the expected trends of structure dependent reverse phase chromatographic behavior of glycopeptides (16, 17).

Determination of sites of O-glycosylation using EThcD

Exoglycosidase digestion and EThcD fragmentation was used to determine exact sites occupied by glycans. We have identified 9 O-glycopeptides occupied by O-glycans (Table 2). Figure 3 shows a typical ETD/EThcD fragmentation spectrum of an O-glycopeptide in this case simplified by non-specific neuraminidase. The fragmentation shows that T687 is the major occupied site as the z6 carries a glycan but the z4 does not; thepeptide is occupied by core-1 (Figure 3A) as well as core-2 (Figure 3B) structures. In addition, we used a combination of neuraminidase, β1-3 and β1-4galoctosidasesto resolve the HexNAc attachment on the O-glycopeptide even in the HCD spectra (Supplemental Figure 3).

Structural analysis of the O-glycopeptides using beam type fragmentation

We chose cIMS (18) of oxonium ions to determine the structural features of O-glycopeptides of the SARS CoV-2 glycoprotein. We choose to use cIMS on the fragment to reduce influence of the peptide backbone on the structural resolution. We were able to confirm the presence of core-2 structures by the diHexNAc fragment m/z 407 in the HCD spectra using the Orbitrap Fusion Lumos (Figure 2A inset). The tandem mass spectra obtained from the cIMS instrument preserve large oxonium ions, such as the intact detached glycan m/z 1022 (Figure 5) which confirms that a hexasacharide occupies the O-glycopeptide AGC(cam)LIGAEHVNN(dea)SYEC(cam)DIPIGAGIC(cam)ASYQTQTNSPR but using beam type fragmentation we could not determine which serine or threonine is occupied. We cannot fully exclude the possibility of contribution from a second glycan at this peptide but neither the ETD not the HCD spectra show evidence of another occupied site besides the T678 of this peptide. We have also confirmed the presence of an extended core-1 structure associated with this glycopeptide by the fragments 528 and 819 observed in the spectra (Figure 5B).

Structural analysis of the O-glycopeptides using cIMS

We have used cIMS to separate isomeric oxonium ion fragments of the glycopeptides. We have used the m/z 657 ion to assign sialylation of the core-2 monosialylated structures. We used an optimized procedure based on a hemopexin glycopeptide standard, which we described previously (12), and we determined CCS of the fragment 657 (Figure 6) observed by fragmentation of the glycopeptide with sialyl-T antigen with linkage (α2-3) (CCS 234.9) and by fragmentation of an N-glycopeptide with sialyl-LacNAc with (α2-6) linkage (CCS 232.8) (data not shown). This is in

agreement with the previously published results on the linkages of the sialylated glycans (6, 8). We resolved two major IMS peaks in the cIMS of the fragment m/z 657 using a one pass method (Figure 7A). With 5 passes, the first peak was partially separated into two analytes with determined CCSs of 232.8 and 234.9 $Å^2$ and a second peak CCS 248.5 $Å^2$. This is reproducible for all 2HexNAc containing structures (Supplemental Figure 2). The CCS of the first peak fits exactly the previously observed CCS of sialylated α2-6 LacNAc while the CCS of the second peak fits the CCS of the sialylated α2-3 T-antigen. CCS 248.5 $Å^2$ of the third peak is in agreement with the previously described CCS of α2-3 LacNAc (19) (20). The peak corresponding to the 2-6 linked sialic acid is better visible under high collision energy (Figure 7B-D) due to different stability of the SA-Gal bond (21). We have determined a 7/3 ratio of the GlcNAc-Gal-2-3-SA and GalNAc-Gal-2-3SA in the mono-sialylated core-2 structure (Figure 7, panel A). We used the cIMS of the fragment m/z 731 (2HexNAc2Hex) to determine the ratio of the core-2 structure and the extended core-1 structure. We obtained 2 major peaks using 5 passes of the cIMS (data not shown); the first peak (drift time:69.70 ms; CCS: 238.8 $Å^2$) is consistent with a core-2 structure and the second peak (76.84 ms; CCS: 251.5 $Å^2$) with a linear core-1 extended structure with terminal GalNAc(1-3)Gal as described previously (22). The ratio of the core-1 with terminal GalNAc(1-3)Gal and the core-2 structure is 25/75.

Conclusion:

We have used our energy optimized LC-MS/MS and ion mobility MS/MS methods to resolve structural motifs of the N- and O-glycans of the SARS-CoV-2 S protein. We identified 17 N-glycopeptides, with many glycoforms including the LAcdiNAc and polyLacNAc structural

motives. This is important for functional studies and the use of the protein as an immunization target. In addition, we identified, for the first time, an O-glycopeptide adjacent to the polybasic furin cleavage site located between the S1/S2 subunits that carries core-1 and core-2 structures capped primarily with α2-3 sialic acid at the T678. The furin cleavage site is unique to the SARS-CoV-2 S protein compared to the SARS-CoV of 2002 and its cleavage is potentially regulated by the nearby O-glycans as described for other convertases. In addition, we identified 8 additional O-glycopeptides of variable occupancy and unknown functional significance. The study expands substantially the knowledge of the glycoforms of SARS-CoV-2 S expressed in the HEK293 cells and warrants further exploration of the impact of glycosylation on the S protein's function.

Funding:

Research reported in this publication was supported by the National Institutes of Health under awards S10OD023557, U01CA230692, and R01CA238455 to RG. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Compliance with Ethical Standards:

Conflict of Interest

The authors declare that they have no conflict of interest.

References:

1. WHO Situation report - 71, “Coronavirus disease 2019 (COVID-19)” (2020).

2. F. Wu, *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* (2020) https:/doi.org/10.1038/s41586-020-2008-3.

3. W. D., *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (80-.).* (2020) https:/doi.org/http://dx.doi.org/10.1126/science.aax0902.

4. J. S.A., *et al.*, CD209L (L-SIGN) is a receptor for severe acute respiratory syndrome coronavirus. *Proc. Natl. Acad. Sci. U. S. A.* (2004).

5. W. H. Li, *et al.*, Single-cell RNA-seq data analysis on the receptor ACE2 expression reveals the potential risk of different human organs vulnerable to 2019-nCoV infection. Frontiers of medicine. *Nature* (2003) https:/doi.org/10.1038/nature02145.

6. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science (80-.).* (2020) https:/doi.org/10.1126/science.abb9983.

7. K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, R. F. Garry, The proximal origin of SARS-CoV-2. *Nat. Med.* **26**, 450–452 (2020).

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

9. Y. Zhang, *et al.*, Site-specific N-glycosylation Characterization of Recombinant SARS-

CoV-2 Spike Proteins using High-Resolution Mass Spectrometry. *bioRxiv*, 2020.03.28.013276 (2020).

10. M. Sanda, J. Benicky, R. Goldman, Low Collision Energy Fragmentation in Structure-Specific Glycoproteomics Analysis. *Anal. Chem.* (2020) https:/doi.org/10.1021/acs.analchem.0c00519.

11. P. Pompach, *et al.*, Site-specific glycoforms of haptoglobin in liver cirrhosis and hepatocellular carcinoma. *Mol. Cell. Proteomics* (2013) https:/doi.org/10.1074/mcp.M112.023259.

12. M. Sanda, *et al.*, Increased sialylation of site specific O-glycoforms of hemopexin in liver disease. *Clin. Proteomics* (2016) https:/doi.org/10.1186/s12014-016-9125-x.

13. J. Benicky, M. Sanda, Z. B. Kennedy, R. Goldman, N-Glycosylation is required for secretion of the precursor to brain-derived neurotrophic factor (proBDNF) carrying sulfated LacdiNAc structures. *J. Biol. Chem.* (2019) https:/doi.org/10.1074/jbc.RA119.009989.

14. W. Yuan, J. Benicky, R. Wei, R. Goldman, M. Sanda, Quantitative Analysis of Sex-Hormone-Binding Globulin Glycosylation in Liver Diseases by Liquid Chromatography-Mass Spectrometry Parallel Reaction Monitoring. *J. Proteome Res.* (2018) https:/doi.org/10.1021/acs.jproteome.8b00201.

15. K. T. B. G. Schjoldager, *et al.*, A systematic study of site-specific GalNAc-type O-glycosylation modulating proprotein convertase processing. *J. Biol. Chem.* (2011) https:/doi.org/10.1074/jbc.M111.287912.

16. P. Kozlik, R. Goldman, M. Sanda, Study of structure-dependent chromatographic behavior

of glycopeptides using reversed phase nanoLC. *Electrophoresis* (2017) https:/doi.org/10.1002/elps.201600547.

17. P. Kozlik, M. Sanda, R. Goldman, Nano reversed phase versus nano hydrophilic interaction liquid chromatography on a chip in the analysis of hemopexin glycopeptides. *J. Chromatogr. A* (2017) https:/doi.org/10.1016/j.chroma.2017.08.066.

18. K. Giles, *et al.*, A Cyclic Ion Mobility-Mass Spectrometry System. *Anal. Chem.* (2019) https:/doi.org/10.1021/acs.analchem.9b01838.

19. M. Guttman, K. K. Lee, Site-Specific Mapping of Sialic Acid Linkage Isomers by Ion Mobility Spectrometry. *Anal. Chem.* (2016) https:/doi.org/10.1021/acs.analchem.6b00265.

20. A. Barroso, *et al.*, Evaluation of ion mobility for the separation of glycoconjugate isomers due to different types of sialic acid linkage, at the intact glycoprotein, glycopeptide and glycan level. *J. Proteomics* (2018) https:/doi.org/10.1016/j.jprot.2017.11.020.

21. A. Depraz Depland, G. Renois-Predelus, B. Schindler, I. Compagnon, Identification of sialic acid linkage isomers in glycans using coupled InfraRed Multiple Photon Dissociation (IRMPD) spectroscopy and mass spectrometry. *Int. J. Mass Spectrom.* (2018) https:/doi.org/10.1016/j.ijms.2018.09.005.

22. C. Jin, D. J. Harvey, W. B. Struwe, N. G. Karlsson, Separation of isomeric o-glycans by ion mobility and liquid chromatography-mass spectrometry. *Anal. Chem.* (2019) https:/doi.org/10.1021/acs.analchem.9b01772.

Table 1. N-glycosylation of the SARS-CoV-2 S glycoprotein: A LacdiNAc; B PolyLacNAc; C Outer-arm fucosylation; D core fucosylation

Annotated Sequence	Occupied site	Site occupancy	identified structures	Identified structural motifs
[R].SSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTK.[R]	61,74		”51”	A,C,D
[K].TQSLLIVNNATNVVIK.[V]	122	100%	63	A,C,D
[K].VCEFQFCNDPFLGVYYHKNNK.[S]	149	100%	60	A,C,D
[R].VYSSANNCTFEYVSQPFLMDLEGK.[Q]	165	100%	63	A,C,D
[R].DLPQGFSALEPLVDLPIGINITR.[F]	234	>99%	45	A,C,D
[K].YNENGTITDAVDCALDPLSETK.[C]	282	100%	72	A,B,C,D
[R].FPNITNLCPFGE.[V]	331	>99%	11	A,C,D
[E].VFNATR.[F]	344	100%	48	A,C,D
[D].VNCTEVPVAIHADQLTPTWR.[V]	616	>99%	8	A
[R].AGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPR.[A]	657	96%	53	A,B,C,D
[K].DFGGFNFSQILPDPSKPSK.[R]	801	>99%	30	A,C,D
[K].NFTTAPAICHDGK.[A]	1074	98%	31	A,B,C,D
[R].EGVFVSNGTHWFVTQR.[N]	1098	>99%	78	A,B,C,D
[D].VVIGIVNNTVYDPLQPE.[L]	1134	92%	11	A
[K].NHTSPD.[V]	1158	“>99%”	4	A,B,C,D
[D].LGDISGINASVVNIQK.[E]	1173	57%	10	A,B,D
[K].NLNESLIDLQELGKYEQYIK.[W]	1194	98%	25	A,B,C,D

Table 2. O-glycosylation of the SARS-CoV-2 S glycoprotein

Peptide	Modification	%
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(2)Hex(2)NeuAc(2)	5.03
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(2)Hex(1)NeuAc(1)	0.49
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(2)Hex(2)	1.5
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(2)Hex(2)NeuAc(1)	2.91
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(1)Hex(1)NeuAc(2)	0.79
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(1)Hex(1)NeuAc(1)	0.72
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(1)Hex(1)	<0.1
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]; 1xHexNAc(1)	1.34
[E].CDIPIGAGICASYQTQTNSPR.[A]	2xCarbamidomethyl [C1; C10]	87.22
[R].TQLPPAYTNSFTR.[G]		99.73
[R].TQLPPAYTNSFTR.[G]	1xDeamidated [N9]	0.18
[R].TQLPPAYTNSFTR.[G]	1xHexNAc(1)	0.05
[R].TQLPPAYTNSFTR.[G]	1xHexNAc(1)Hex(1)	0.02
[R].TQLPPAYTNSFTR.[G]	1xHexNAc(3)	0.01
[R].TQLPPAYTNSFTR.[G]	1xHexNAc(1)Hex(1)NeuAc(2)	0.01
[K].YFKNHTSPDVD.[L]	1xDeamidated [N4]; 1xHexNAc(1)Hex(1)NeuAc(2)	7.18
[K].YFKNHTSPDVD.[L]	1xDeamidated [N4]	70.33
[K].YFKNHTSPDVD.[L]	1xDeamidated [N4]; 1xHexNAc(2)Hex(2)NeuAc(2)	22.49
[R].VQPTESIVR.[F]	1xHexNAc(2)Hex(2)NeuAc(1)	7.13
[R].VQPTESIVR.[F]	1xHexNAc(1)Hex(1)NeuAc(2)	30.56
[R].VQPTESIVR.[F]	1xHexNAc(1)Hex(1)NeuAc(1)	5.65
[R].VQPTESIVR.[F]	1xHexNAc(1)Hex(1)	13.93
[R].VQPTESIVR.[F]		42.75
[R].VYSTGSNVFQTR.[A]	1xHexNAc(1)	0.023
[R].VYSTGSNVFQTR.[A]		99.67
[E].VPVAIHADQLTPTWR.[V]	1xHexNAc(1)Hex(1)NeuAc(2)	0.013
[E].VPVAIHADQLTPTWR.[V]		99.98
[D].STECSNLLLQYGSFCTQLNR.[A]	2xCarbamidomethyl [C4; C15]; 1xHexNAc(1)Hex(1); 1xHexNAc(2)Hex(3)Fuc(1)	87.23
[D].STECSNLLLQYGSFCTQLNR.[A]	2xCarbamidomethyl [C4; C15]; 1xHexNAc(2) [S/T]; 1xHexNAc(2)Hex(1)NeuAc(2)	12.77
[K].LPDDFTGCVIAWNSNNLD.[S]	1xCarbamidomethyl [C8]	82.04
[K].LPDDFTGCVIAWNSNNLD.[S]	1xCarbamidomethyl [C8]; 2xHexNAc(1)Hex(2)Fuc(1)	14.37
[K].LPDDFTGCVIAWNSNNLD.[S]	1xCarbamidomethyl [C8]; 1xHexNAc(3)Hex(2)Fuc(1)NeuAc(1)	3.59
[D].YSVLYNSASFSTFK.[C]	1xDeamidated [N6]	3.35
[D].YSVLYNSASFSTFK.[C]	1xDeamidated [N6]; 1xHexNAc(1)Hex(1)Fuc(1)NeuAc(1)	0.037
[D].YSVLYNSASFSTFK.[C]		96.61

Figure 1. HCD fragmentation of the N165 glycopeptide carrying an asymmetric biantennary glycan with sialylated LacdiNAc structural motif.

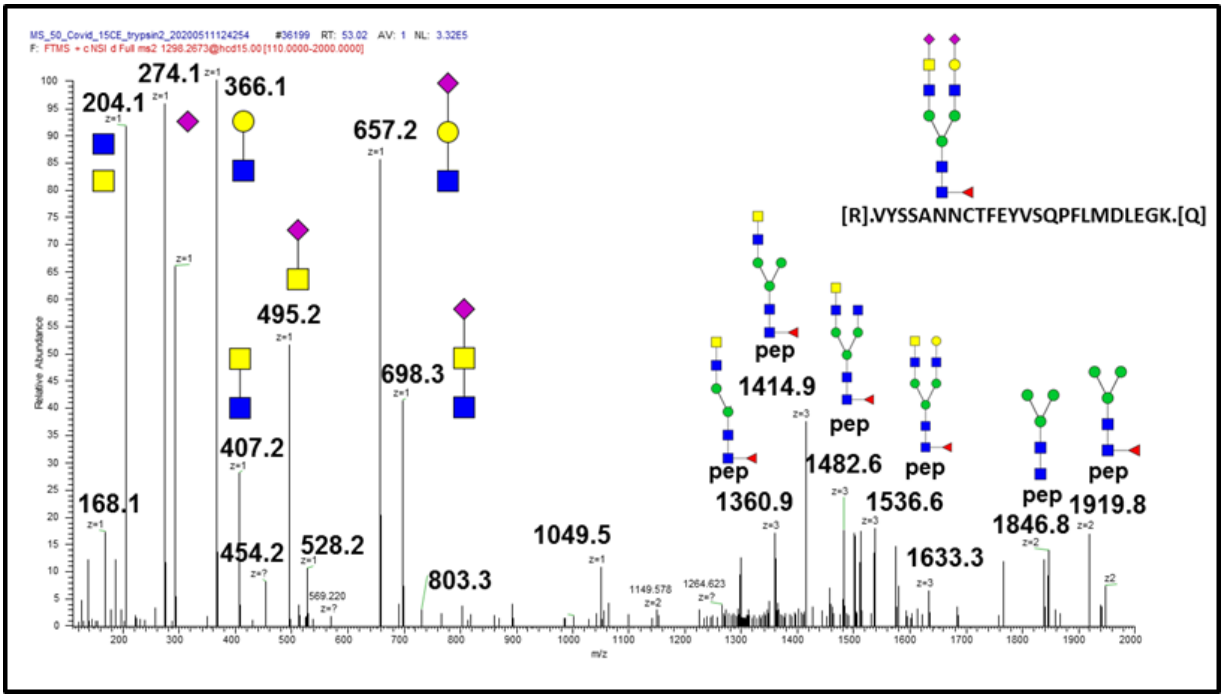

Figure 2. HCD tandem mass spectra of the SARS-CoV-2 S protein O-glycosylated on T678 with the following structures: (A) extended core-1 and core-2 structures; (B) disialylated core-1 structure. Inset: oxonium ions in the HCD fragmentation spectrum confirm the presence of core-2 structures.

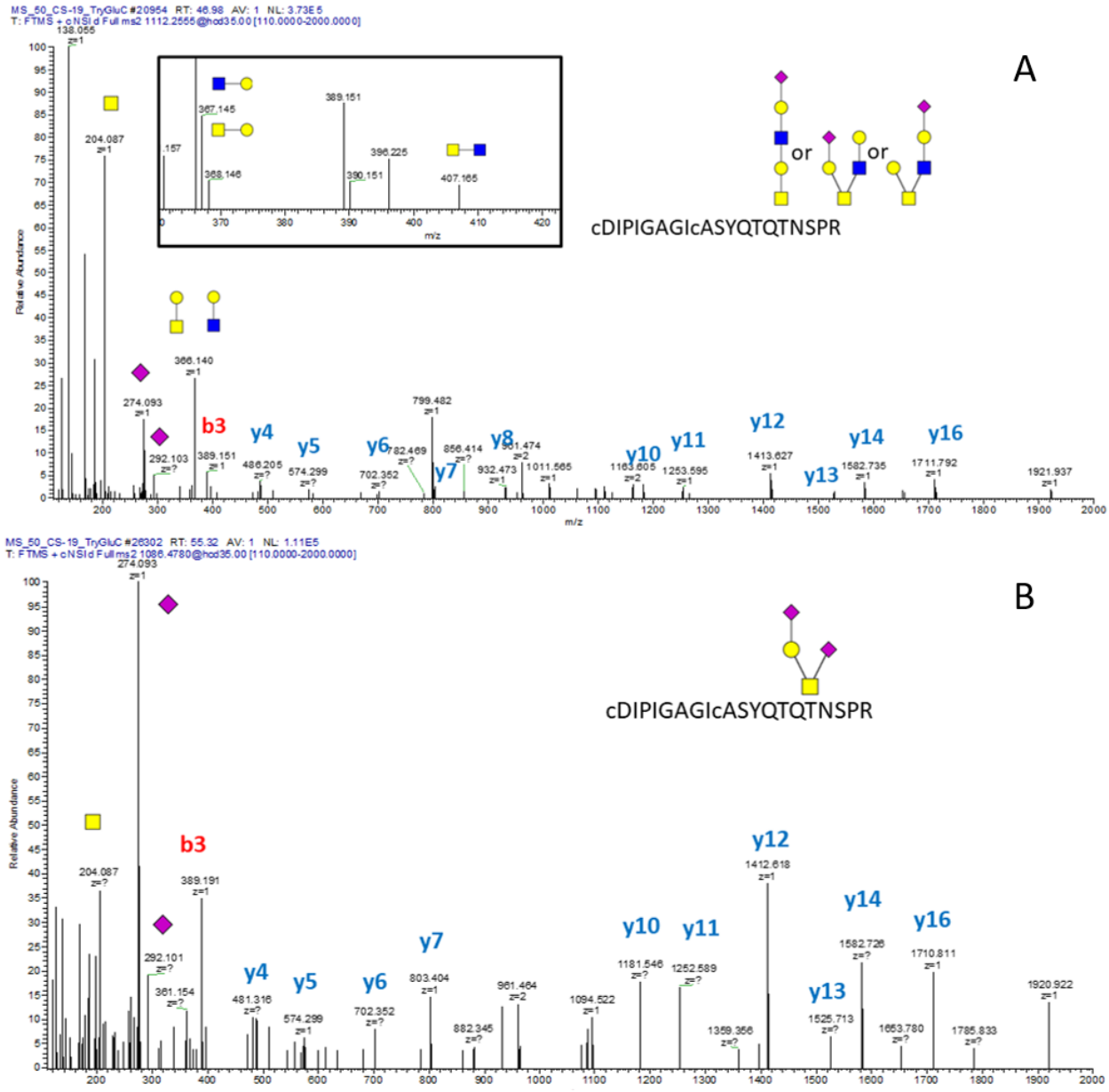

Figure 3. EThcD tandem mass spectra of a tryptic/GluC glycopeptide treated with PNGaseF and non-specific neuraminidase confirms occupancy of the T678 by core-1 (B) and core-2 (A) structures.

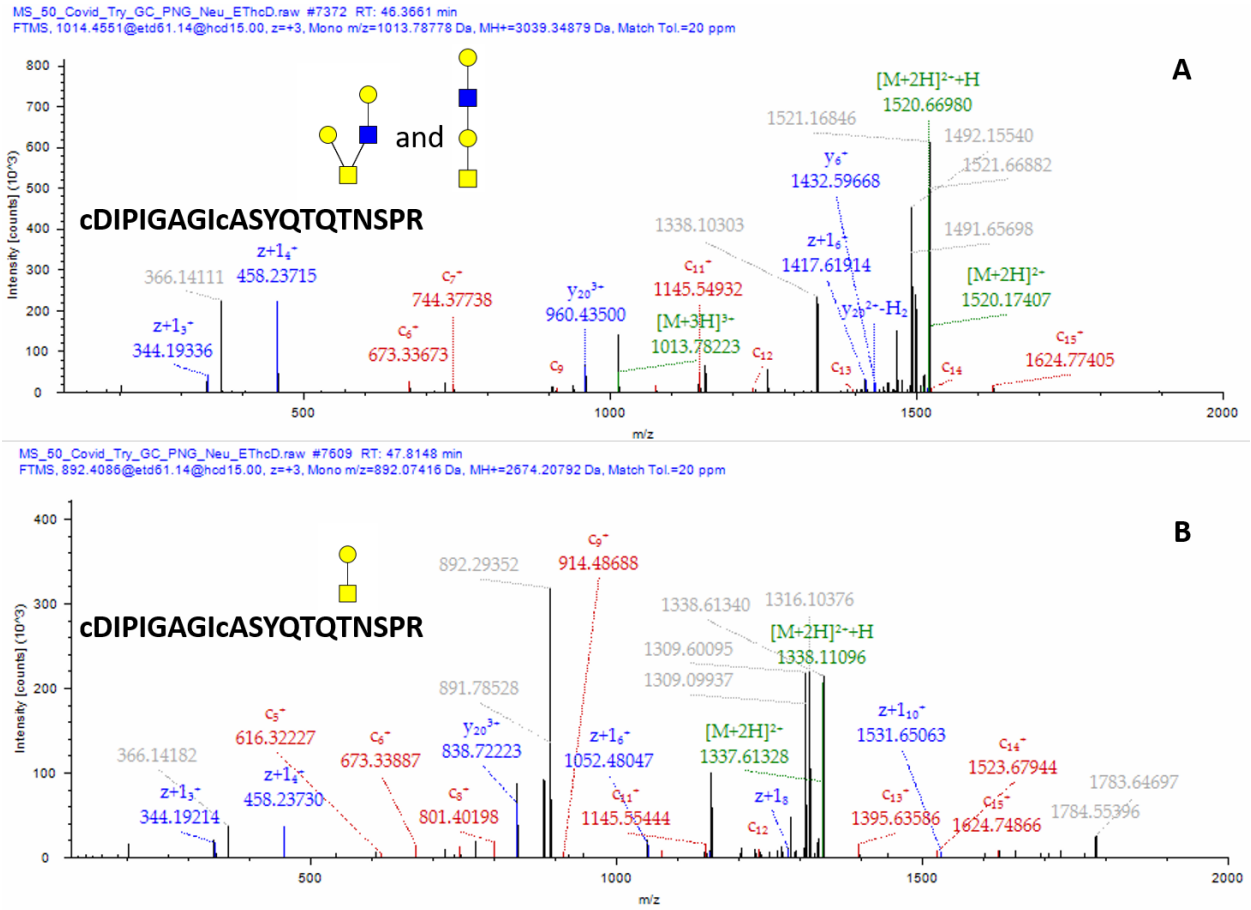

Figure 4. The CDIPIGAGICASYQTQTNSPR O-glycopeptides of SARS-CoV-2 S protein with the expected glycoform-dependent RT shifts visible in the XIC chromatograms.

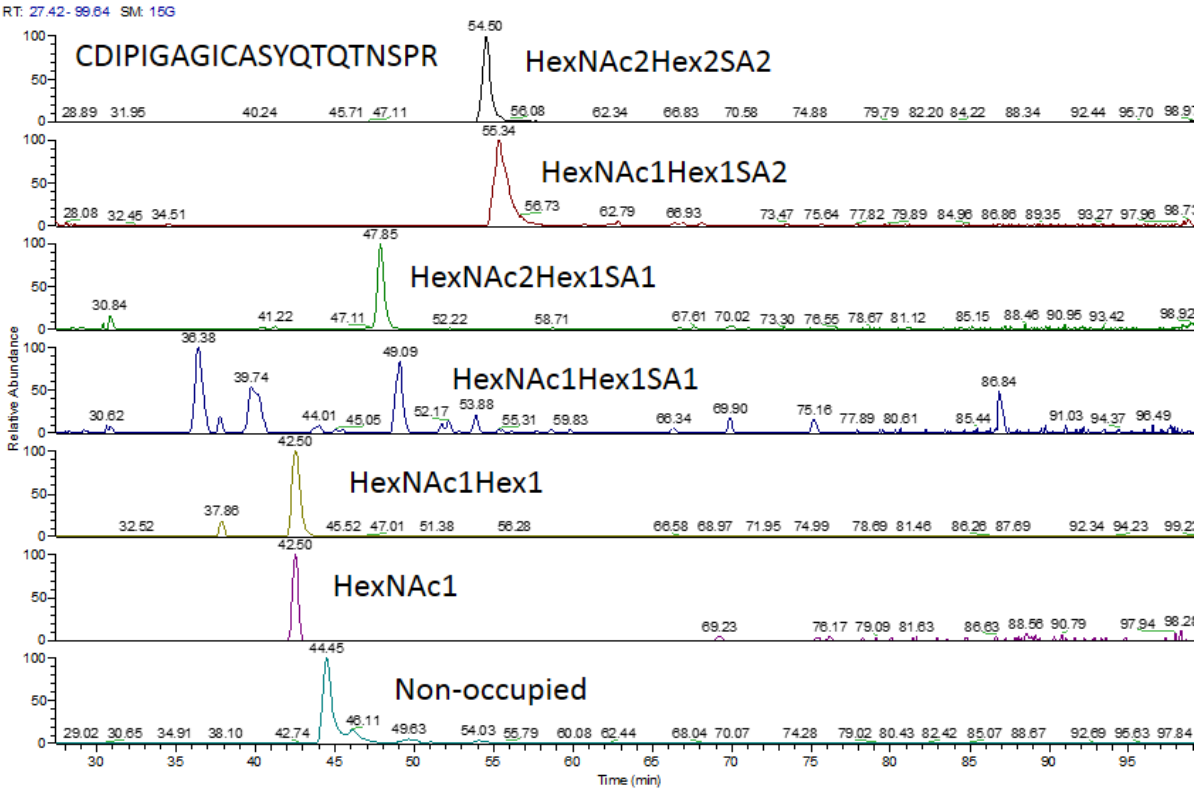

Figure 5. Beam type tandem mass spectra of the AGC(cam)LIGAEHVNN(dea)SYEC(cam)DIPIGAGIC(cam)ASYQTQTNSPR (HexNAc2Hex2SA1) O-glycopeptide with assigned extended core-1 and core-2 structures. The structures are characterized by the following fragments: (**A**) oxonium ions 366 and 657, generated from both core -1 and core-2 structures; (**B**) oxonium ion 407 specific for the core-2 and ions 528 and 819 specific to the extended core-1 structure; and (**C**) oxonium ion 1022 correspoding to the detached intact glycans.

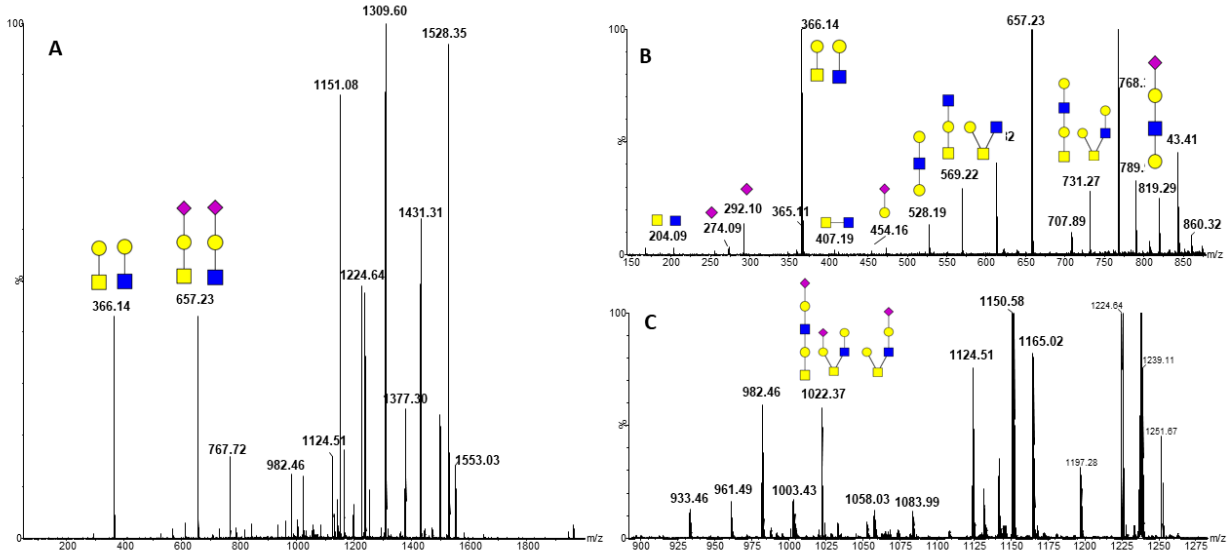

Figure 6. cIMS of the fragment m/z 657 with measured CCS assignments produced by fragmentation of the AGC(cam)LIGAEHVNN(dea)SYEC(cam)DIPIGAGIC(cam)ASYQTQTNSPR (HexNAcHexSA) (A) and (HexNAc2Hex2SA) (B) O-glycopeptides produced by tryptic digests and PNGaseF deglycosylation of the SARS-CoV-2 S glycoprotein.

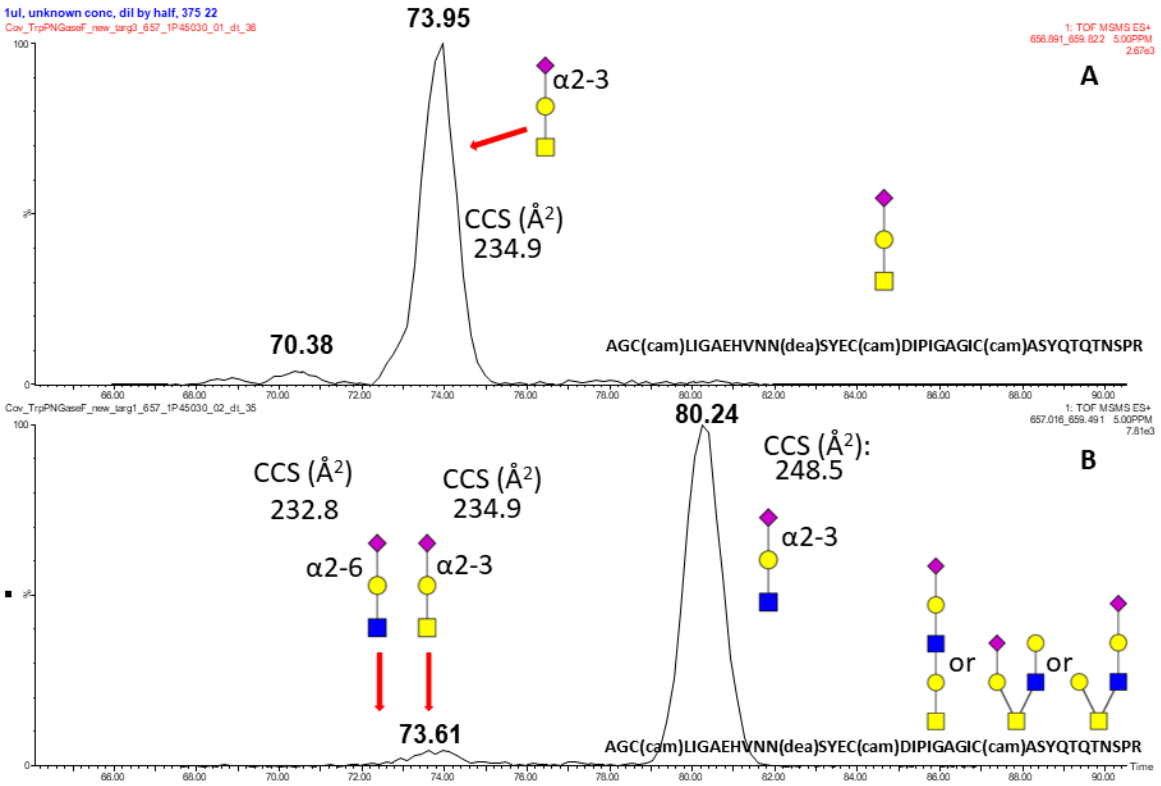
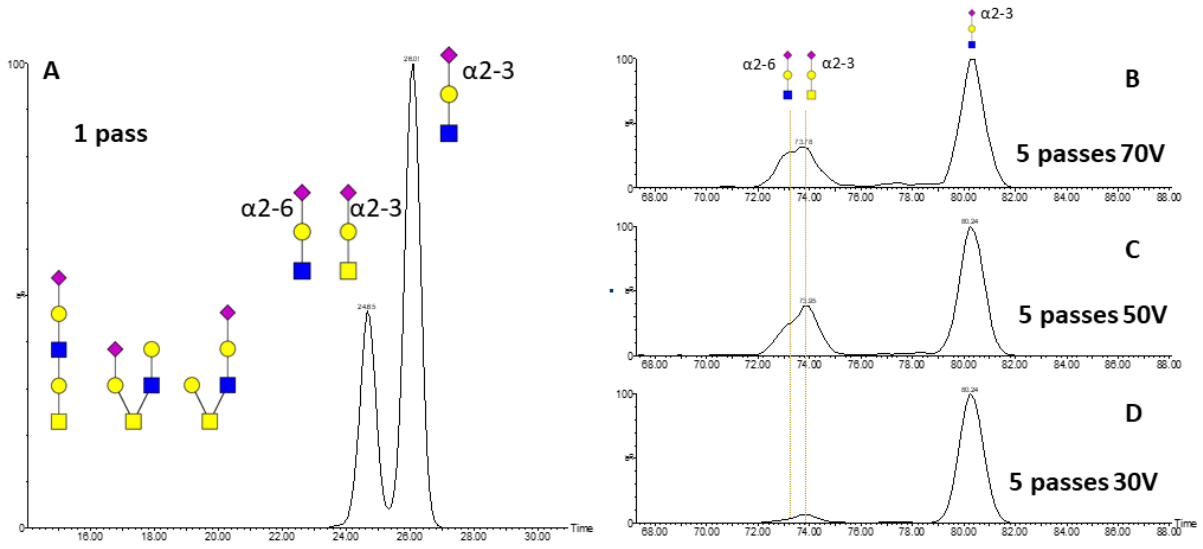
Figure 7. cIMS of the m/z 657 fragment produced by fragmentation of the AGC(cam)LIGAEHVNN(dea)SYEC(cam)DIPIGAGIC(cam)ASYQTQTNSPR O-Glycopeptide produced by tryptic digest and PNGaseF deglycosylation of the SARS-CoV-2 S glycoprotein using the following settings: (**A)** one pass cIMS does not resolve GalNAcGal(2-3)SA and GlcNAcGal(2-6)SA; (**B**). 5 passes cIMS with 70V CE; (**C**) 5 passes cIMS with 50V CE; and (**D**) 5 passes cIMS with 30V CE. The ion mobilograms (**B,C,D**) show that multiple passes improve resolution of isobaric (GalNAcGal(2-3)SA and GlcNAcGal(2-6)SA) structures and reveals differences in the stability of the sialic acid linkages.