Use of mass spectrometry to assist production of SARS-CoV-2 spike protein

Peptide mapping analysis of COVID-19 spike protein using the SCIEX X500B QTOF System and Protein Metrics Inc. software enables the confirmation of induced protein variants

Raquel Faba-Rodriguez[1], Rachel Rowlinson[1], Sibylle Heidelberger[2], Chris Nortcliffe[2], Stephen Lock[2]
[1]Peak Proteins, UK; [2]SCIEX, UK

In this work, mass spectrometry has been used in assessing purity during protein production experiments. Both wild type (WT) and mutant forms of SARS-CoV-2 spike protein were expressed in a HEK293 system and were confirmed by analysis with the SCIEX X500B System. Rapid and efficient characterization was performed by using automated data analysis with Byos software (Protein Metrics Inc.).

Proteins and other biotherapeutics have a broad range of applications, however, their expression and purification often require monitoring to obtain a product with the desired structure and quality attributes. Understanding the quality and heterogeneity of any protein sample is key to making data-driven decisions. During protein production, mass spectrometry can rapidly determine whether specific sequences have been inserted or deleted, or it can identify key amino acid point mutations that have lasting effects on biological activity. Production of the specific COVID-19 spike protein is critical in the current climate as this protein forms the basis of many of the vaccines for SARS-CoV-2. Due to the high degree of glycosylation, it is important that standards are accurate and correct when used for comparing antigen quality when developing vaccines and antibody tests.

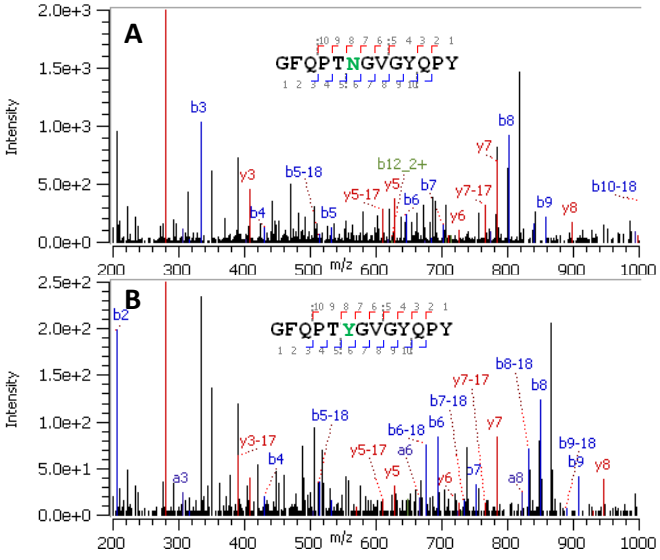

Figure 1. MS/MS spectra of WT and mutant peptide of COVID-19 spike protein. A) WT N501 and B) mutant N501Y. The *b* and *y* ions across both peptides confirm the site of the point mutation with high confidence. In particular the *y*8 and *b*6 fragments in each sequence confirm the alteration of asparagine (N) to tryptophan (Y).

In this technical note, liquid chromatography coupled to mass spectrometry (LC-MS) analysis of proteolytic digests of SARS-CoV-2 spike protein provided excellent sequence coverage for this highly complex protein and confirmed several sites of mutation, including the variant N501Y— a common mutation found in the UK, Brazilian and South African variants. In addition, various N- and O-linked glycopeptides were successfully identified across the protein using the Byos software (Protein Metrics Inc.).

Key features of the SCIEX solution for protein production

- Rapid identification and confirmation of mutation sites with high-quality, high-resolution QTOF MS and MS/MS data using automated information-dependent acquisition (IDA)
- Characterization of post-translational modifications, such as N- and O-linked glycosylations, in the same injection, without the need for compound-specific tuning
- Ease-of-use in instrument control of the X500B System and data acquisition using SCIEX OS Software, suitable for every user-level
- Streamlined data analysis with direct compatibility of raw data with state-of-the-art software suite from Protein Metrics Inc.

Methods

Sample preparation: Coomassie stained bands were subjected to 10 minute washes with 180 µL 50% acetonitrile/water, 180 µL 100 mM ammonium bicarbonate and then with 180 µL of 50 mM ammonium bicarbonate/50% acetonitrile (10 minutes each). To each gel band, 75 µL 10 of mM dithiothreitol (DTT)/100 mM ammonium bicarbonate were added and the band was incubated (65°C, 45 minutes). The reduction solution was discarded and the samples were alkylated by adding 75 µL of 50 mM iodoacetamide/ 100 mM ammonium bicarbonate and incubated (room temperature in the dark, 20 minutes). The alkylation solutions were discarded and the bands were then washed with 180 µL of 50 mM ammonium bicarbonate, followed by 180 µL of 50 mM ammonium bicarbonate/50% acetonitrile (*v*:*v*), before a final wash with 180 µL acetonitrile for 10 minutes. The bands were dried down in vacuum desiccator for 10 minutes. A mixture of 14 µL of 12.5 ng/µL trypsin solution and 14 µL of 12.5 ng/µL chymotrypsin solution (made up in 50 mM ammonium bicarbonate pH 8.0) was added to each sample and the bands were incubated (overnight at room temperature). The peptides were extracted by adding 20 µL 0.1% trifluoroacetic acid (TFA)/60% acetonitrile (*v*:*v*) to each band and placing the samples on the shaker (10 minutes). This extraction was repeated and the supernatant from both extractions was pooled and dried down in the vacuum desiccator (10 minutes). Samples were reconstituted to 25 µL with 0.1% TFA ready for analysis on the X500B System.

Chromatography: A volume of 10 µL of the digest was injected onto a Phenomenex Luna Omega PS-C18 column (2.1 × 150 mm, 1.6 µm particle size, 100 Å) using an ExionLC™ AD System. The aqueous mobile phase (A) consisted of water with 0.1% formic acid while the organic phase (B) consisted of acetonitrile with 0.1% formic acid. Temperature was set to 40°C on the column and a gradient profile was used (Table 1) with 0.3 µL/min.

Table 1. Chromatographic gradient.

Time	%A	%B
2.0	*95*	*5.0*
7.5	*55*	*45*
8.0	*5.0*	*95*
9.0	*5.0*	*95*
9.1	*95*	*5.0*

Mass spectrometry: Mass spectrometry data were acquired in positive ionization mode using the SCIEX X500B System fitted with an Turbo V™ Ion Source. An IDA approach was used. MS and MS/MS conditions are listed in Table 2.

Table 2. Mass spectrometry parameters.

Parameter	MS	MS/MS
Scan mode	*TOF-MS*	*IDA dependent*
Gas 1	*30 psi*	
Gas 2	*40 psi*	
Curtain gas	*35 psi*	
Source temperature	*450°C*	
Ion spray voltage	*5500 V*	
Declustering potential	*100 V*	
Collision energy	*10 V*	*Rolling*
CAD gas	*7*	
Maximum candidate ion	*10*	
Intensity threshold	*100 cps*	
Charge states	*2 to 7*	
Exclusion time	*3 s after 1 occurrence*	
Start mass	*300 m/z*	*100 m/z*
Stop mass	*1,800 m/z*	*1,250 m/z*
Accumulation time	*0.15 s*	*0.05 s*
Time bins to sum	*12*	*16*

Data processing: Data processing was performed using Byos software (Protein Metrics Inc.) Byos allowed an extensive list of 59 N-glycans and 78 O-glycans to be screened during the data processing to identify all glycosylated peptides.

Sequence coverage

With the emergence of the novel disease, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), production of protein constructs became a key driver for both academic and industrial research. The spike protein of SARS-CoV-2 is very important in receptor recognition and interaction with the cell membrane. It consists of two subunits, the S1 (aa14-685) receptor binding domain which binds to the host receptor for angiotensin-converting enzyme 2, and the S2 (aa686-1273) domain which is involved in viral cell membrane fusion.[1] The protein is highly glycosylated with 22 identified N-glycosylation sites along with O-glycosylation on T323.[2]

Double digestion with trypsin and chymotrypsin was performed to provide great sequence coverage and confirm that inserted mutations were present and correct. The sequence coverage of spike protein based on MS/MS data was 87.86%.

Site-specific mutations and insertions

SARS-CoV-2 spike protein contains an internal furin recognition site with the sequence RRAR, which results in an S1/S2 protease cleavage site. When expressing the protein, this is undesirable as cleavage increases heterogeneity of the final product and results in multiple chains. Because of this, the site was replaced by the sequence GSAS. Residues 986 and 987 were also replaced with prolines to stabilize this prefusion complex.[3] From a protein production perspective, confirmation of these mutation sites is critical to ensure the product has the desired three dimensional structure. Figure 2 shows the MS and MS/MS data of the two peptides containing these regions discovered using Byos software. Both peptides have very good fragment coverages, confirming the presence of the altered amino acids.

The third engineered mutation was the addition of a C-terminal foldon, a trimerization domain that has been shown in previous coronaviruses to encourage formation of the active multimeric complex.[3] This is especially important for structural studies such as cryogenic electron microscopy and X-ray crystallography to study the conformations of these biomolecules. The MS and MS/MS of the longest, identified peptide in this addition is shown in Figure 3 along with the other peptides that were found to support addition of this domain.

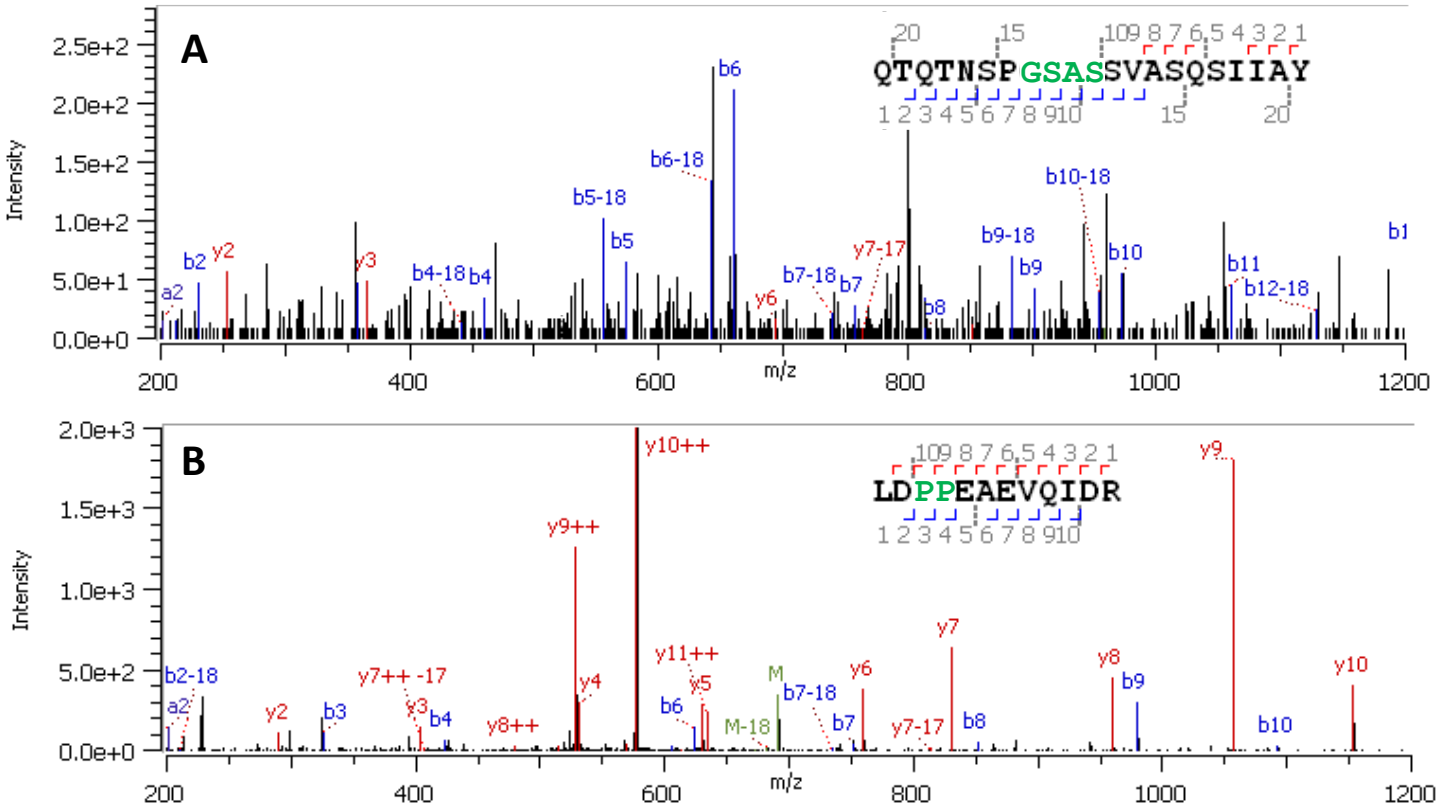

Figure 2. MS/MS spectra of introduced mutations. A) Exchange of aa682-685 RRAR to GSAS. The GSAS mutation was introduced at the S1/S2 domain boundary cleavage site to prevent cleavage in S1 and S2 subunits by furin protease. Image A shows the MS/MS and coverage of this modified peptide confirming the new sequence. B) The two proline mutations K986P and V987P stabilize the spike proteins in the prefusion conformation. The MS/MS spectrum confirms the peptide with this new sequence with high quality *b*- and *y*-ions.

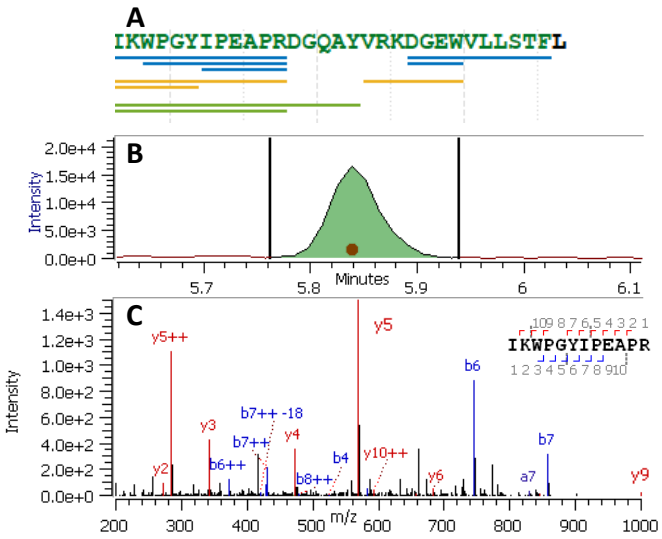

Figure 3: Data from C-terminal insertion of the trimerization sequence domain. A) Sequence map with identified peptides indicated as bars. B) XIC of precursor IKWPGYIPEAPRDGQAY and C) corresponding MS/MS spectrum.

The final mutation was at the location of N501Y which was found as a natural mutation in several variants of the spike protein.[5] This mutation has been correlated with an increase in infectivity and is therefore of particular interest in research. Figure 1 shows the comparison of the MS/MS of the WT version of this peptide compared to the N501Y variant. High quality MS/MS ensures identification critical *b*- and *y*-ions, providing excellent coverage of the single site of mutation.

N- and O-linked glycosylations

Many coronaviruses that have infected humans in recent history have been heavily glycosylated and SARS-CoV-2 is no exception. Although the focus of the experiments presented here were on rapid analysis for determination of mutation sites, 21 of the 22 previously identified N-linked glycan locations[4] were confirmed with this rapid analysis. Various glycans were identified on different sites from high-mannose species such as M5-M9 as well as more complex processed glycans. Byos software from Protein Metrics Inc. was used to identify these glycans from a selected list of 59 O-linked and 78 N-linked glycans. Figure 4 shows an example of the presence of several O-glycans on a peptide with two potential sites. MS/MS data suggests that the glycan is attached to the threonine residue (T323) rather than the serine (S325) with a delta mod. score of greater than 10. This is supported by previous work that found strong evidence for T323 being the predominantly occupied site.[2] Figure 5 shows a particular peptide, FSNVTWF, which was identified with several different N-linked glycosylations, with each providing excellent MS and MS/MS quality.

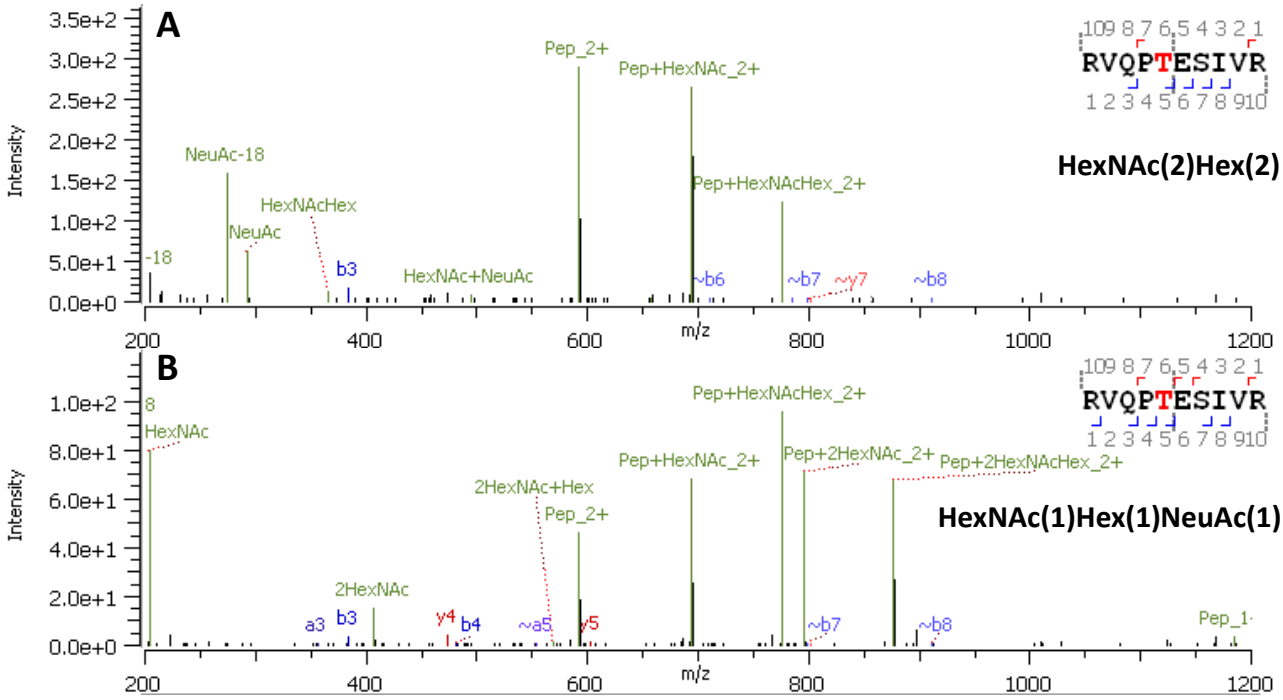

Figure 4. MS/MS spectra of O-linked glycosylated peptide RVQPTESIVR in COVID-19 spike protein. A) HexNAc(2)Hex(2) B) HexNAc(1) Hex(1)NeuAc(1). The suggested site of modification was T323.

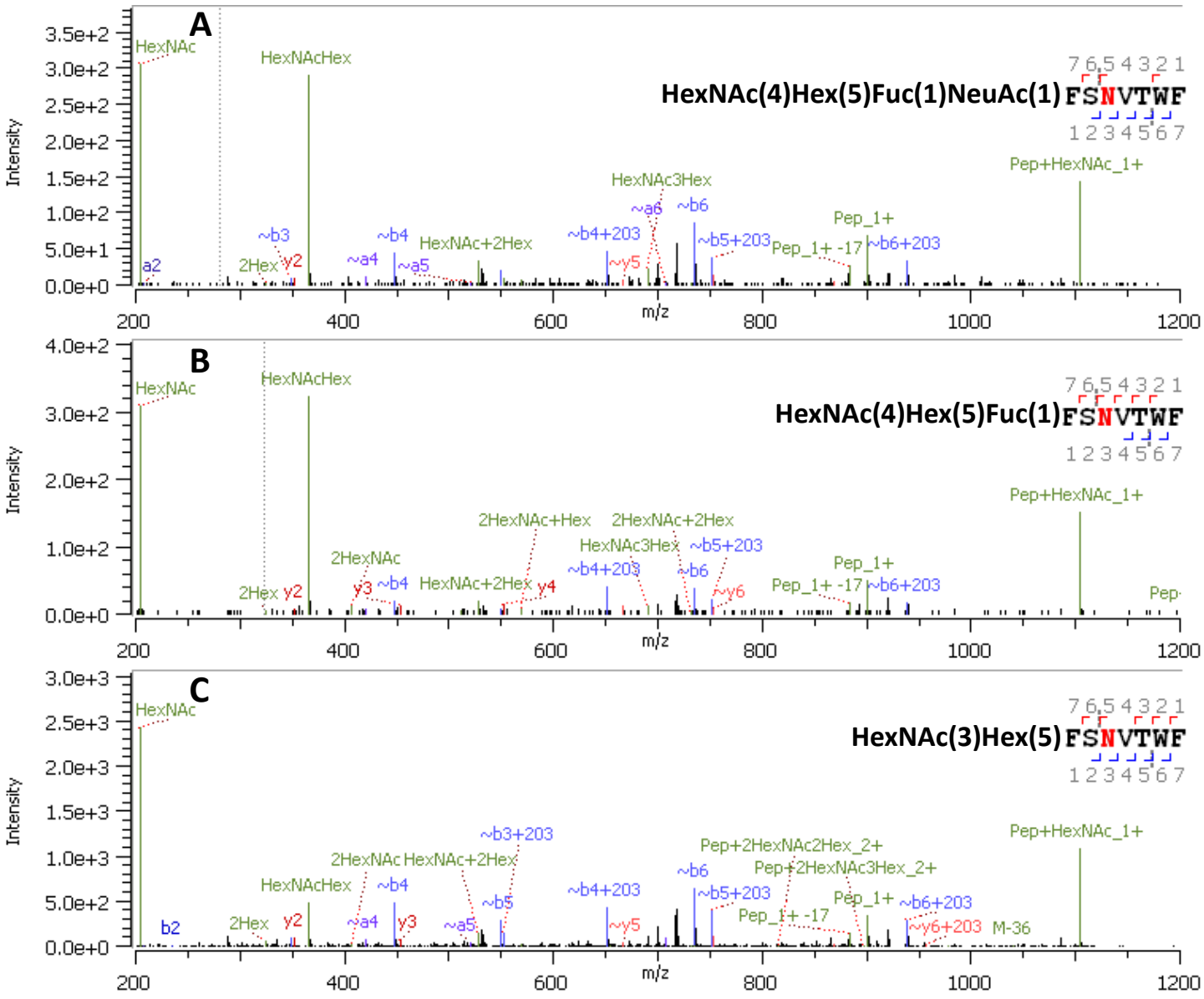

Figure 5. MS/MS spectra of N-linked glycosylated peptide FSNVTWF in COVID-19 spike protein. A) HexNac(4)Hex(5)Fuc(1)NeuAc(1), B) HexNac(4)Hex(5)Fuc(1) C) HexNAc(3) Hex(5).

Conclusions

- A straightforward characterization of the SARS-CoV-2 spike protein using an automated IDA strategy for peptide mapping on the SCIEX X500B System with SCIEX OS Software was shown
- The sequence coverage of the protein obtained in less than 10 min was greater than 87%, allowing for integrity confirmation in parallel with the confirmation of mutation sites via high-quality data, including N501Y variant mutation
- Various N-linked glycosylation sites were identified and several different glycoforms could be confirmed using high-quality MS and MS/MS data
- The O-linked glycosylated peptide was identified and MS/MS data indicated localization at T323

References

1. Huang, Y., Yang, C., Xu, Xf. et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin* **41,** 1141–1149 (2020).
2. Shajahan,A. ,Supekar N T., Gleinich A S., Acadi, P.,Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **30,** 12 981-988 (2020).
3. Walls, A.C. Park, Y. Tortorici, M.A. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**(2): 281-292 (2020).
4. Wrapp, D., Wang, N., Corbett, K.S., et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367,** 6483 1260-1263 (2020).
5. Leung, K., Shum, M.H., Leung, G.M., et al. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill*. **26,** 1 (2021).

Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices