For Research Use Only. Not for use in diagnostic procedures.

In-depth characterization of SARS-CoV-2 and mutant N501Y spike proteins

Haichuan Liu[1], Sibylle Heidelberger[2], Rachel Rowlinson[3], Raquel Faba[3] and Zoe Zhang[1]
[1]SCIEX, USA; [2]SCIEX, UK; [3]Peak Proteins, UK

INTRODUCTION

The spike protein of SARS-CoV-2 is the primary target for vaccine design, development of protein therapeutics, and scientific research surrounding COVID-19. However, it is challenging to achieve comprehensive characterization of SARS-CoV-2 spike protein due to its complex glycosylation patterns[1], ever-evolving mutations and deamidation that may affect the function and molecular aging of the spike protein.[2] Collision-based approaches such as collision-induced dissociation (CID) used in traditional peptide mapping workflows struggle with localizing glycosylations and differentiating amino acid isomers, such as aspartic (Asp) and isoaspartic (isoAsp) acid derived from deamidation. That leads to incomplete characterization of these important product quality attributes. By comparison, these challenges can be addressed using electron activated dissociation (EAD), as described previously.[3-4]

Here, a highly sensitive data-dependent acquisition (DDA) method comparing CID and EAD for obtaining complete characterization of SARS-CoV-2 spike protein and its N501 mutant was investigated. The advantages of EAD for improved confidence in identification of certain peptides and glycopeptide analysis will be highlighted.

MATERIALS AND METHODS

Sample preparation: The wild type (WT) and N501Y mutant forms of SARS-CoV-2 spike protein were expressed in a HEK293 system. Two protein samples (~50 µg in a concentration of ~0.8 mg/mL) were denatured using 8 M urea in 20 mM Tris HCl (pH = 7.9), reduced at 37ºC for 45 min by 20 mM dithiothreitol and alkylated in the dark at room temperature with 100 mM iodoacetamide. Subsequently, trypsin and chymotrypsin were added in a ratio of 1:10 (enzyme:protein) and the samples were incubated overnight at 37ºC. The digestion was quenched with 10% trifluoroacetic acid. 10–20 µL of the 2 digestion samples, which correspond to ~1-2 µg of proteins assuming 60% sample recovery, were injected for DDA using CID and 40 µL (~4 µg) were injected for EAD analyses.[5]

HPLC conditions: The peptides were separated with a linear 62 min gradient using an ACQUITY CSH C18 column (2.1 × 150 mm, 1.7 µm, 130 Å, Waters). A flow rate of 0.25 mL/min was used for all the separations mentioned above. The column was kept at 60ºC in the column oven of an ExionLC system (SCIEX). The mobile phases A and B consisted of 0.1% formic acid in water and 0.1% FA in acetonitrile, respectively.[5]

MS/MS conditions: LC-MS data were acquired in SCIEX OS software using the ZenoTOF 7600 system (SCIEX). The key TOF MS and DDA settings for peptide mapping with Zeno CID and Zeno EAD were described earlier.[5]

Data processing: CID and EAD DDA data were processed using the peptide mapping workflow templates in the Biologics Explorer software (SCIEX). The mass tolerances for MS and MS/MS were set to 10 ppm and 20 ppm, respectively. The instrument type was set to “CID (Glycopeptides)” and “EAD” for processing CID and EAD DDA data, respectively. Trypsin and chymotrypsin were chosen from the enzyme list. The carbamidomethyl at cysteine was set as the fixed modification. The variable modifications included deamidation at N and Q, Gln to pyro-Glu for N-terminal Q, oxidation at M and ammonia loss for N. For identification of N-linked glycopeptides, a large library containing 133 CHO N-glycan structures was chosen from the libraries available from the Biologics Explorer software. A library of 7 common O-glycan structures was used for subsequent identification of O-linked glycopeptides. It should be highlighted that identification of N- and O-linked glycopeptides can be performed in a stepwise manner, limiting false-positives despite a high search space, using an extended peptide mapping template within the Biologics Explorer software.[5]

RESULTS

The highly sensitive Zeno CID DDA method led to a near complete sequence coverage (>95%) for both WT and N501Y mutant of SARS-CoV-2 spike protein, thereby allowing in-depth characterization of mutations and modifications in the 2 proteins. The naturally occurring N501Y mutation was confidently confirmed based on the identification of peptides GFQPT**N**GVGYQPY and GFQPT**Y**GVGYQPY in the WT and N501Y samples, respectively, using CID or EAD (Figure 1).

While the CID MS/MS spectra (Figure 1A and 1C) of these 2 peptides were dominated by a *y2* ion due to preferential cleavage at proline near the C-terminus, a more even distribution of sequence ions (*b/y* and *c/z*) was observed in the EAD spectra (Figure 1B and 1D), thereby allowing for improved confidence in peptide identification.

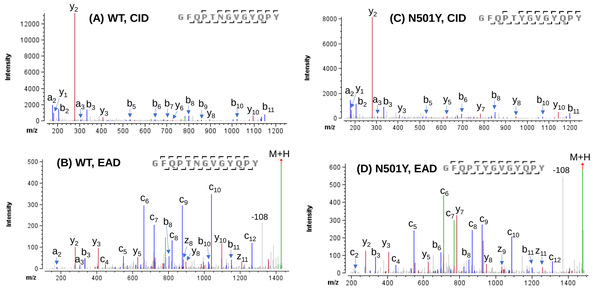
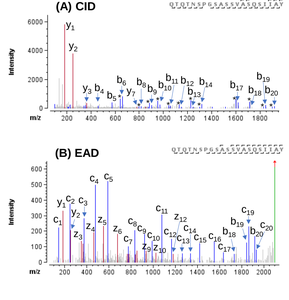


Figure 1. CID and EAD MS/MS spectra of WT peptide and its N501Y mutant. The N501Y mutation was confidently confirmed by the high-quality CID or EAD data of 2 peptides. While the CID spectra of WT and N501Y peptides (A and B) are dominated by the presence of a *y2* ion resulting from the preferential cleavage at proline near the C-terminus, EAD generated a more even distribution of sequence ions (*b/y* and *c/z*) for the two peptides (C and D). The neutral loss of 108 amu of the remaining precursor in the EAD spectra corresponds to a side chain loss from the tyrosine residue. Note: Spectra were deconvoluted and deisotoped. Not all fragments were labelled for spectral clarity.

The sequences of both WT and N501Y forms analyzed in this study were engineered to increase their resistance to proteases and the stability between S1 and S2 subunits. Specifically, the sequence RRAR in a furin recognition site was replaced by GSAS and 2 amino acid residues at positions 986 and 987 were replaced by 2 proline residues. These sequence changes were confidently confirmed by CID and EAD data. Displayed in Figure 3 are CID and EAD MS/MS spectra of the peptide with RRAR to GSAS mutation.



Both MS/MS spectra were rich in sequence ions, enabling confident peptide identification and confirmation of the sequence change. However, compared to the CID spectrum (Figure 2A), where y_1 and y_2 were the dominant fragment ions, the EAD spectrum (Figure 2B) was featured with more sequence ions (*c/z* and *b/y*) in a more even distribution. Additionally, while the *b*-H_2O ions were prevalent in the CID spectrum (labelled with “*” in Figure 2A) of this serine-rich peptide (6 Ser in total), these fragments were not detected in the EAD spectrum (Figure 2B). These unique characteristics of EAD improved confidence in characterization of mutations of single or multiple amino acid residues in the SARS-CoV-2 spike protein.

Figure 2. CID and EAD MS/MS spectra of the peptide with engineered site (RRAR to GSAS). Peaks labeled “*” in the CID spectrum correspond to the H_2O-loss products of the corresponding *b* ions (A). Both CID (A) and EAD (B) provided confident peptide identification and confirmation of the sequence change. As compared to CID (A), which generated abundant *b*-H_2O ions from this serine-rich peptide, EAD (B) produced more extensive sequence ions (*c/z* and *b/y*) and minimal H_2O-loss products, thereby improving the confidence in sequence analysis. Spectra were deconvoluted and deisotoped.

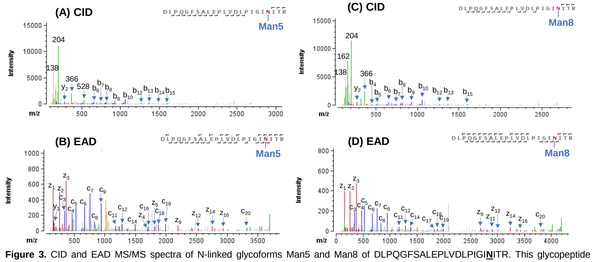
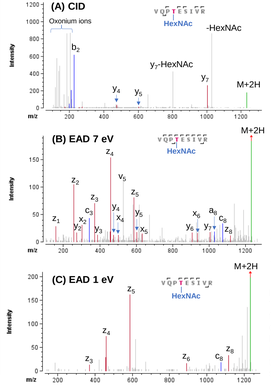


Figure 3. CID and EAD MS/MS spectra of N-linked glycoforms Man5 and Man8 of DLPQGFSALEPLVDLPIGINITR. This glycopeptide contains an N-glycosylation site at Asn234. The CID spectra of glycoforms Man5 and Man8 (A and C) were dominated by the oxonium ions (such as *m/z* 204) from the glycans. Although the detection of a series of low-abundance *b*-ions in CID allowed confident identification of these 2 glycopeptides, no fragments containing the glycan were observed and hence site-specific information about glycosylation was lost. By comparison, EAD of these 2 glycopeptides (B and D) generated more sequence ions (*c/z*) for improved confidence in identification and produced fragments with the glycan (such as, c_{20} and z_4) for unambiguous localization of the glycosylation. Spectra were deconvoluted and deisotoped for clarity.

The SARS-CoV-2 spike protein contains a variety of glycan structures, including predominantly high-mannose and complex N-glycan types as well as trace level of O-glycosylation.[1] In this study, 21 N-linked glycosylation sites were identified from CID and EAD data. In agreement with published results,[1] most of the N-linked glycosylation sites carried high-mannose and complex glycans (example in Figures 3). The CID MS/MS spectra (Figure 3A/3C) were enriched with abundant oxonium ions (such as, *m/z* 204) generated from the glycan moieties. Although the detection of a series of low-abundance *b/y*-ions in CID enabled confident sequence identification, no fragments containing the glycans were observed. Thus, site-specific information about glycosylation was missing from CID data.



By comparison, EAD (Figure 3B/3D) generated more complete sequence ions (*c/z*) for improved confidence in glycopeptide identification and preserved the intact glycans on the fragments, thereby allowing precise localization of glycosylation. This result confirms the capability of EAD for in-depth characterization of glycopeptides.

O-glycosylation was found to be present at very low levels. The main O-glycosylation site was identified to be T323 with both CID and EAD data (Figure 4). The occupancy of O-linked glycan at T323 was confirmed based on a series of EAD fragments with or without the glycan HexNAc (Figure 4B and 4C). Specifically, the accurate site information about this glycan was obtained from a non-glycosylated z_5 and a HexNAc-containing z_6 ion detected in the EAD data acquired using an electron kinetic energy (KE) of 1 eV (Figure 4C).
Adversely, the CID spectrum was dominated by oxonium ions and fragments with dissociated glycans, impeding localization of the glycan.

Figure 4. CID and EAD MS/MS spectra of O-linked glycopeptide VQPTESIVR. The CID spectrum (A) was featured with abundant oxonium ions and no fragments containing the O-glycan. By contrast, EAD with KE of 7 eV (B) and 1 eV (C) preserved the O-glycan in the fragments, allowing for accurate localization of this glycan to the Thr instead of Ser residue. While EAD with KE=7 eV (B) provided a more complete fragment coverage of this glycopeptide, the EAD spectrum obtained using KE=1 eV (C) was simpler allowing for simplified interpretation.

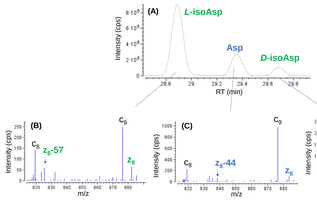


Figure 5. Differentiation of isoAsp and Asp isomers of deamidated peptide GFQPT**N**GVGYQPY using EAD. Three isomeric species were observed in this deamidated peptide. The detection of a signature *z*ϕ–57 fragment allowed confident assignment of two isoAsp isomers (B and D), while the absence of this fragment and the detection of a *z*ϕ–44 ion enabled assignment of the Asp isomer (C). The presence of 2 isoAsp isomers can be explained by racemization from the naturally occurring L- form to its D- counterpart, as reported previously.[6]

Deamidation is an important modification that affects protein stabilities, structures and functions. The effects of deamidation on molecular aging of the SARS-CoV-2 spike protein were studied recently.[2] It is critical to characterize deamidation and its isomers for in-depth understanding of its effect on the function of the spike protein.
As reported previously, it is challenging to differentiate deamidation isomers such as, Asp and isoAsp, based on their elution patterns or using CID.[3,4] However, EAD is capable of generating signature fragments (*z*–57 and *z*–44 for the isoAsp and Asp isomers, respectively) allowing differentiation of these 2 isomers.[3,4] Illustrated in Figure 5 is an example of isomer differentiation for a deamidated peptide of the SARS-CoV-2 spike protein (GFQPT**N**GVGYQPY), which contains the N501Y mutation site using EAD. The 3 isomeric deamidated species observed in the extracted ion chromatogram (XIC) (Figure 5A) can be differentiated based on the detection of a signature z_ϕ–57 fragment for the isoAsp isomer (Figure 5B and 5D) and a z_ϕ–44 fragment for the Asp isomer (Figure 5C). The presence of 2 isoAsp isomers can be attributed to racemization, which converted the naturally occurring L- form to its D- counterpart in lower abundance, as reported previously.[6]

CONCLUSIONS

- Comprehensive characterization of SARS-CoV-2 spike protein and its mutant was achieved using DDA with CID or EAD
- Nearly complete sequence coverages (>95%) were obtained for the wild type and N501Y mutant of the spike protein in a single injection using the highly sensitive CID method
- The mutations in the SARS-CoV-2 spike proteins were confidently confirmed using CID or EAD, with EAD providing better coverage and more even distribution of sequence ions for higher confidence in identification
- EAD provided additional layer of confidence in identification of N- and O-linked glycopeptides as well as localization of glycosylation

REFERENCES

1. Yasunori Watanabe *et al.* (2020) Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**(6501):330-333.
2. Ramiro Lorenzo et al. (2021) Deamidation drives molecular aging of the SARS-CoV-2 spike protein receptor-binding motif. J Biol. Chem. 297(4):101175.
3. Differentiation of aspartic and isoaspartic acid using electron activated dissociation (EAD). SCIEX technical note, RUO-MKT-02-12550-B.
4. An evaluation of single injection platform method for advanced characterization of protein therapeutics using electron activation dissociation (EAD). SCIEX technical note, RUO-MKT-02-13965-A.
5. Comprehensive characterization of SARS-CoV-2 spike protein and its mutant by LC-MS. SCIEX technical note, RUO-MKT-02-14590-A.
6. Marine Morvan and Ivan Miksik. (2021) Recent advances in chiral analysis of proteins and peptides. Separations 8(8): 112.

TRADEMARKS/LICENSING

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to www.sciex.com/diagnostics. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries (see www.sciex.com/trademarks).

© 2022 DH Tech. Dev. Pte. Ltd. RUO-MKT-10-14639-A

To receive a copy of this poster:
- Scan the code with your phone camera
- Complete the form