APPLICATION NOTE FOR AGILENT Q-TOFS

Top-down Localization of Multiple Phosphorylation Sites using the ExD AQ-251 Option

Rachel Franklin, Michael Hare, Joseph Meeuwsen, Phillip Zhu, Richard B. Cooley, Joseph Beckman

Introduction

Protein phosphorylation is the most common type of post-translational modification, estimated to exist on more than 40% of the human proteome.[1,2] Because of the labile nature of phosphorylation modifications, traditional tandem mass spectrometry (MS/MS) techniques using collision-induced dissociation (CID) often results in high false positive and false localization rates when multiple phosphorylations are present.[3,4] Furthermore, proteins that contain dense regions of serine and arginine residues (SR-rich) are poor candidates for enzymatic digestion due to their repeating sequences containing tryptic cleavage sites flanked by potential phosphorylation sites. Thus, alternative workflows and fragmentation strategies are required to address the limitations of conventional bottom up MS/MS for multiply-phosphorylated SR-rich proteins. This study establishes a top-down MS/MS method for localizing phosphorylation using electron capture dissociation (ECD) performed by the e-MSion ExD AQ-251 Option on an Agilent 6545XT AdvanceBio LC/Q-TOF. The method is applied to localize multiple phosphorylation sites on the SR-rich domain of the SARS-CoV-2 nucleocapsid protein.

Experimental

Materials

The SARS-CoV-2 nucleocapsid SR-rich domain is known to become highly phosphorylated in infected lung cells.[5] The first step involves a single phosphorylation event by an initial kinase which acts as a 'primer' that enables sequential phosphorylation by glycogen synthase kinase 3β (GSK-3β).[6] In this study, a single 'primer' phosphoserine was genetically encoded into the sequence at site Ser206 of the SARS-CoV-2 nucleocapsid using amber codon reassignment. An 80-amino acid construct of the SARS-CoV-2 nucleocapsid SR-rich domain containing a single phosphoserine was expressed in *E. Coli* and then purified through chromatography.[7] The synthetically primed protein was then reacted with the enzyme GSK-3β to produce an ensemble of multiply-phosphorylated proteoforms.

MS Sample preparation

Phosphorylated samples were desalted and buffer exchanged using Amicon Ultra 3 KDa MWCO centrifugal filter units (Millipore-Sigma). Filters were pre-rinsed with water and 0.1% formic acid before loading the sample. Then the samples were loaded and buffer-exchanged to 200 mM ammonium acetate according to the manufacturer's instructions. Samples were then diluted to a final concentration of 10 µM in 15% (v/v) acetonitrile with 0.1% formic acid.

Instrumentation

All samples were analyzed on an Agilent 6545XT AdvanceBio LC/Q-TOF configured with a Dual AJS electrospray source and equipped with the e-MSion ExD AQ-251 Option (Figure 1). Samples were introduced through direct infusion at a flow rate of 100 µL per minute. Instrument parameters used for analysis are listed in Table 1.

The autotune feature in the ExDControl software was used to optimize the ExD cell voltage profiles for transmission and for ECD fragmentation, both with the filament heating current set to 2.4 A. Even with the filament hot, the ExD cell voltage profile can be tuned to minimize ECD and maximize transmission. To tune for transmission, Agilent tuning mix was used. To tune for ECD, the peptide standard Substance P was used.

Acquisition

A targeted acquisition method was developed based on the MS1 spectrum of the SR-rich domain wherein individual observed charge states were isolated using a wide (9 *m/z* asymmetric) isolation window. The ExDControl software was configured to automatically apply ECD fragmentation during precursor isolation.

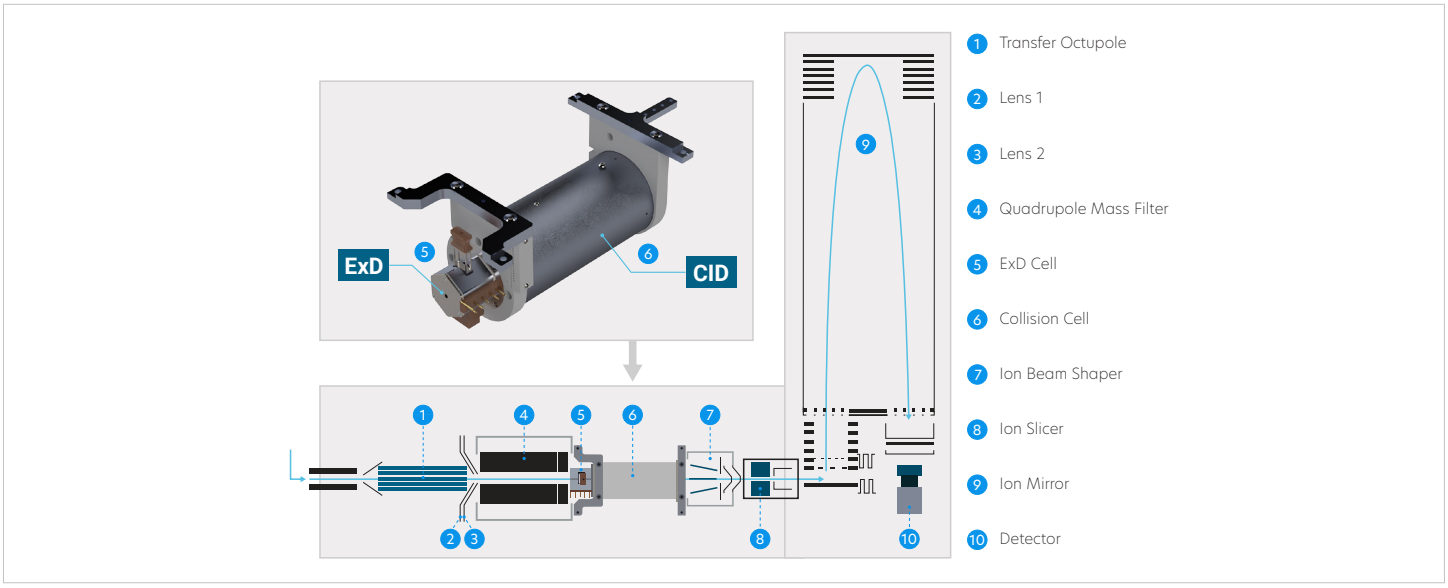

Figure 1. Schematic of the ExD cell in the 6545XT AdvanceBio LC/Q-TOF. The ExD cell is positioned directly upstream of the CID cell. ExD and CID techniques can be used together, independently of one another, or toggled off altogether. Both types of fragmentation occur without ion trapping or disruption to duty cycle.

Agilent 6545XT AdvanceBio LC/Q-TOF parameters	
Gas temperature	300 °C
Drying gas	5 L/min
Nebulizer	35 psi
Sheath temperature	275 °C
Sheath gas flow	12 L/min
Capillary voltage	3500 V
Fragmentor	250 V
Skimmer	150 V
Instrument Mode	2GHz (Extended Dynamic Range)
Slicer	High Resolution
Mass range - MS1	400 - 2,400 m/z
Mass range - MS2	120 – 2,400 m/z
Acquisition rate	1 spectrum/second
Isolation width	Wide
Collision gas pressure	24 psi

Table 1. Instrument parameters used for the analysis of the intact singly- and multiply-phosphorylated proteoforms

Data analysis

Intact masses were determined using the Max Entropy deconvolution algorithm in Agilent Mass Hunter BioConfirm software.

e-MSion's ExDViewer software (version 4.1.14) was used to analyze the top-down MS/MS data and produce figures of annotated spectra and sequence coverage maps. Agilent .d data files were loaded directly into ExDViewer as part of a Match-to-Target workflow. For the singly-phosphorylated 'primed' proteoform, the location of the phosphosite was known in advance due to site-specific genetic incorporation, so it was entered into the target sequence as an expected modification using the Unimod notation "(*Phospho*)". Knowledge about GSK-3's mechanism guided the search for subsequent phosphorylation sites, which are located on serine and threonine in a n+4 pattern.[6] Approximately 6 minutes of spectra were selected and averaged to perform peak-picking and matching on. The PeakPickerHiRes algorithm from OpenMS is used in ExDViewer to perform peak picking on profile data. Default peak-picking parameter values were used. The proprietary matching algorithm in ExDViewer compares *m/z* data to a theoretically generated spectrum, and generates a match score for each fragment ion based on the *m/z*, intensity, and noise baseline of each constituent peak in the cluster. Default matching parameter values were used, including a maximum mass error of 20 ppm, minimum signal-to-noise of 2.0, and minimum match score of 3.5. Only *b*, *c*, *y*, and *z*-type ions were considered.

Results and Discussion

Singly-phosphorylated SR-rich domain

Genetic code expansion is a valuable tool for phosphoprotein characterization because of its ability to create pure, unambiguous site-specifically modified proteins. In this study, genetic code expansion was used to incorporate a single phosphoserine into the SR-rich domain of the SARS-CoV-2 nucleocapsid protein. The single phosphosite encoded acts as a 'primer' for subsequent phosphorylation by the kinase GSK-3β.[6] The 'primed' SR-rich domain was analyzed with top-down ECD MS/MS before and after reaction with GSK-3β (Figure 2).

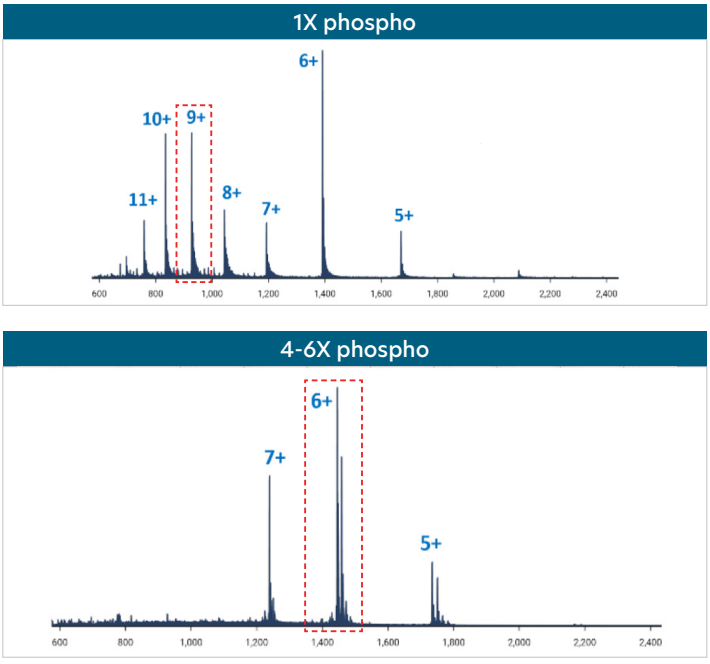

Figure 2. MS1 spectra showing the charge state distribution of the singly phosphorylated SR-rich domain (top) and the multiply phosphorylated SR-rich domain (bottom). The dashed red boxes indicate the charge state that was isolated and fragmented for phosphosite localization experiments

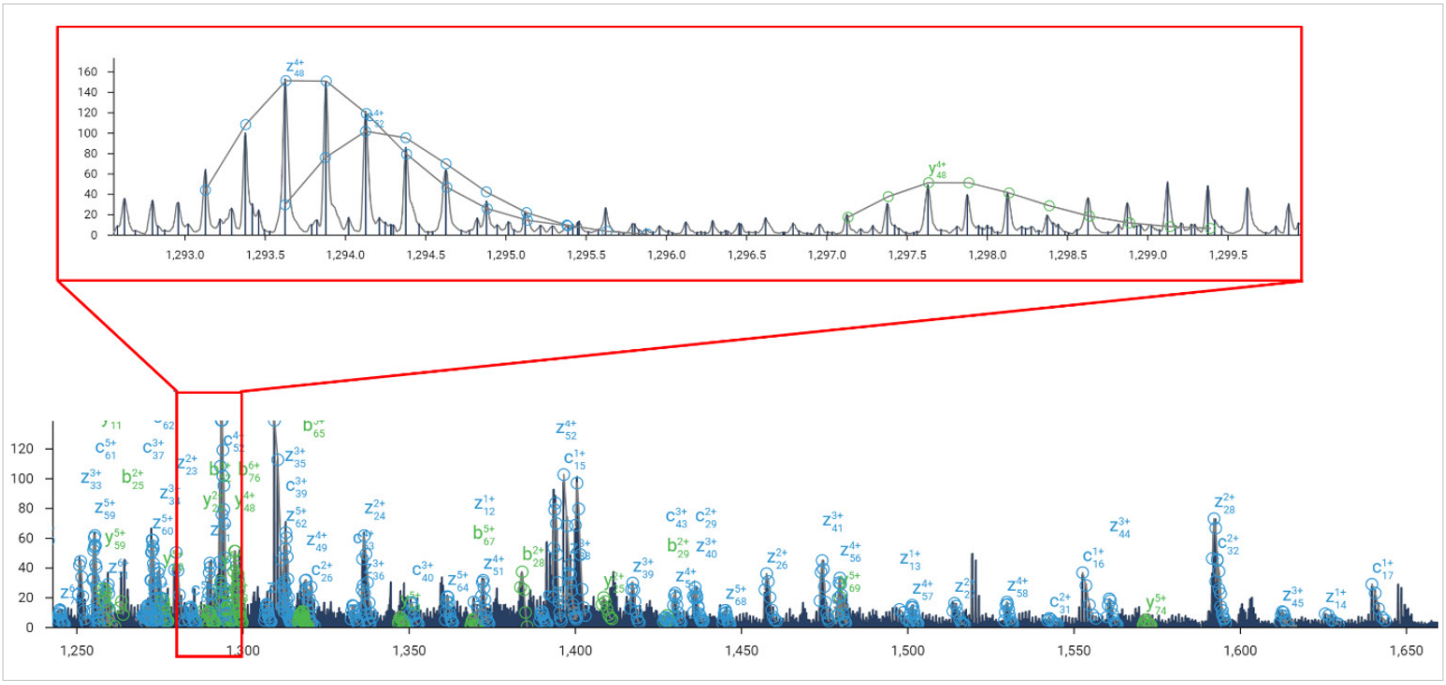

Figure 3. A representative mass-to-charge spectrum showing several of the ECD and CID fragment matches assigned by ExDViewer. The inset highlights the fragment (z_{48}^{4+}) which corresponds directly to the site of the genetically-encoded phosphorylation.

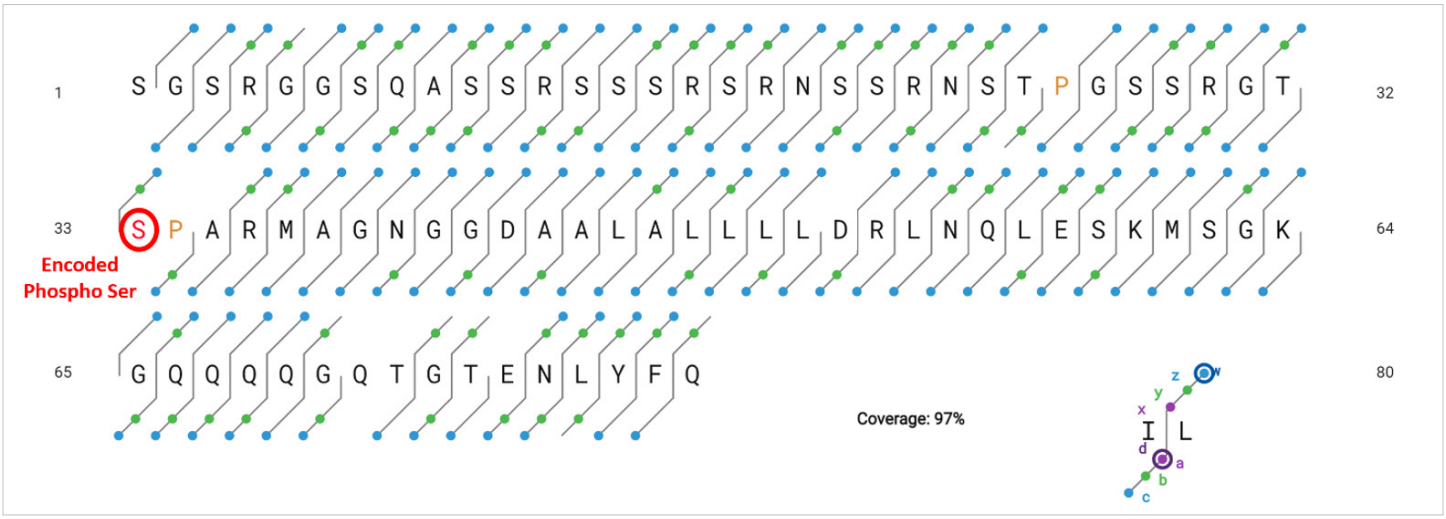

Figure 4. The sequence of the SR-rich domain with the site of genetically encoded phosphoserine indicated by a red circle. The blue dots in between residues represent ECD-derived *c*- and *z*-type ions, and the green dots represent CID-derived *b*- and *y*-type ions. *Note:* ECD does not cleave N-terminal to proline residues.

For the singly-phosphorylated form, a range of charge states from 5-11^{+} was observed. The 9^{+} charge state was isolated and subjected to ECD fragmentation. Agilent's BioConfirm software was used to confirm the purity and expected mass of the singly phosphorylated SR-rich domain. Then, the .d file was loaded directly into ExDViewer for fragment ion matching & annotation (Figure 3).

An extensive series of mostly-ECD fragment ions directly confirmed 77/79 residues (97% sequence coverage), as well as the location of the one genetically encoded phosphoserine at the 33rd residue (Figure 4).

Multiply-phosphorylated SR-rich domain

Confident localization of modification sites in multiply-phosphorylated proteins is a formidable challenge for top down analysis by CID due to the labile nature of the phosphorylation modification.[4] Furthermore, SR-rich proteins present the additional challenge of non-covalent

interactions which can prevent dissociation after electron capture unless supplemental collisional activation is provided.[8,9] The ExD cell provides a unique advantage to the analysis of tightly-folded proteins like the SR-rich domain of the SARS-CoV-2 nucleocapsid protein because it enables collisional activation after ECD which helps dissociate ECD fragment ions without fragmenting labile phosphate groups.[10]

Top-down analysis of the SR-rich domain revealed several populations of multiply-phosphorylated protein after reacting with GSK-3. These populations ranged from 4-6X phosphorylated, with 5X phosphorylated protein being the most abundant form (Figure 5).

The charge states in the MS1 spectrum of the multiply-phosphorylated SR-rich domain ranged from 5-7^{+},

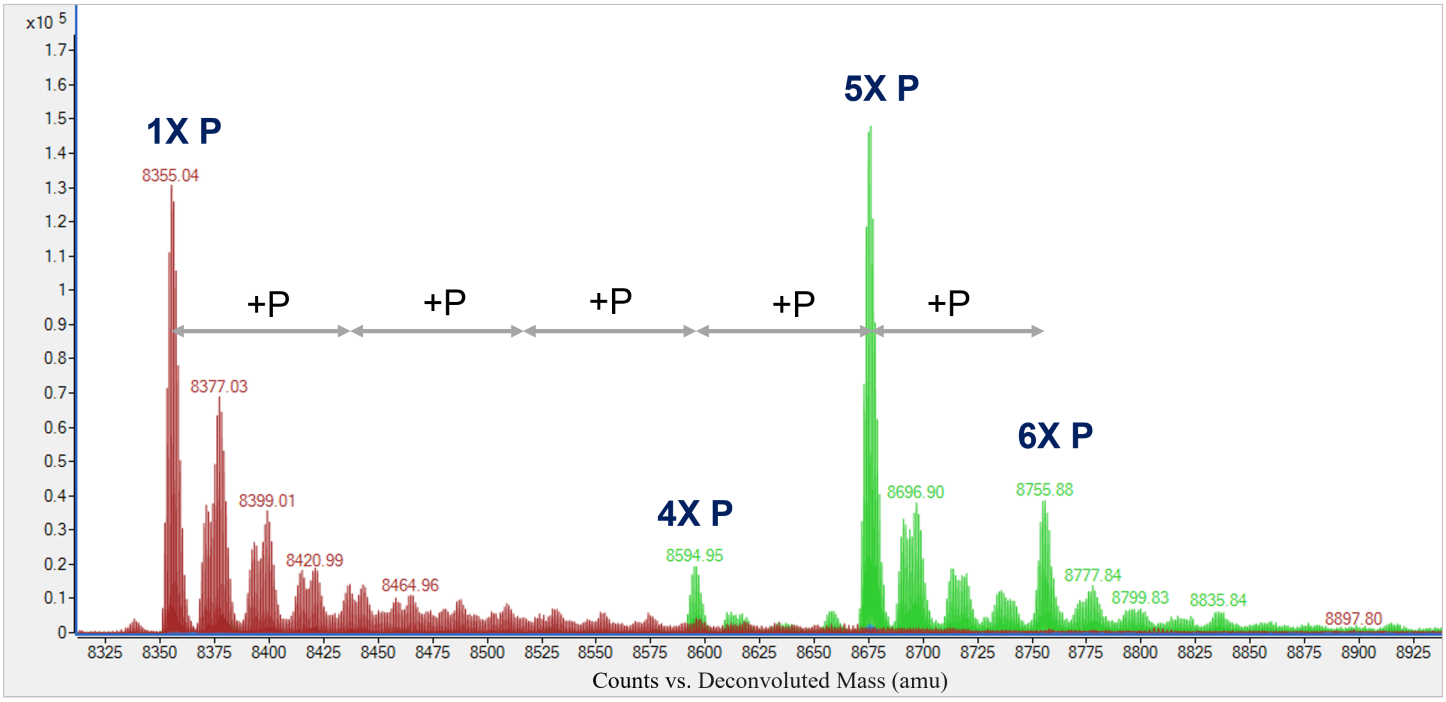

Figure 5. The deconvoluted mass spectra of the singly-phosphorylated (red) and the multiply-phosphorylated (green) SR-rich domain. Deconvolution was performed using the Agilent BioConfirm software, and the spectra were exported using the 'copy to clipboard' function. Unlabeled higher mass peaks are salt adducts.

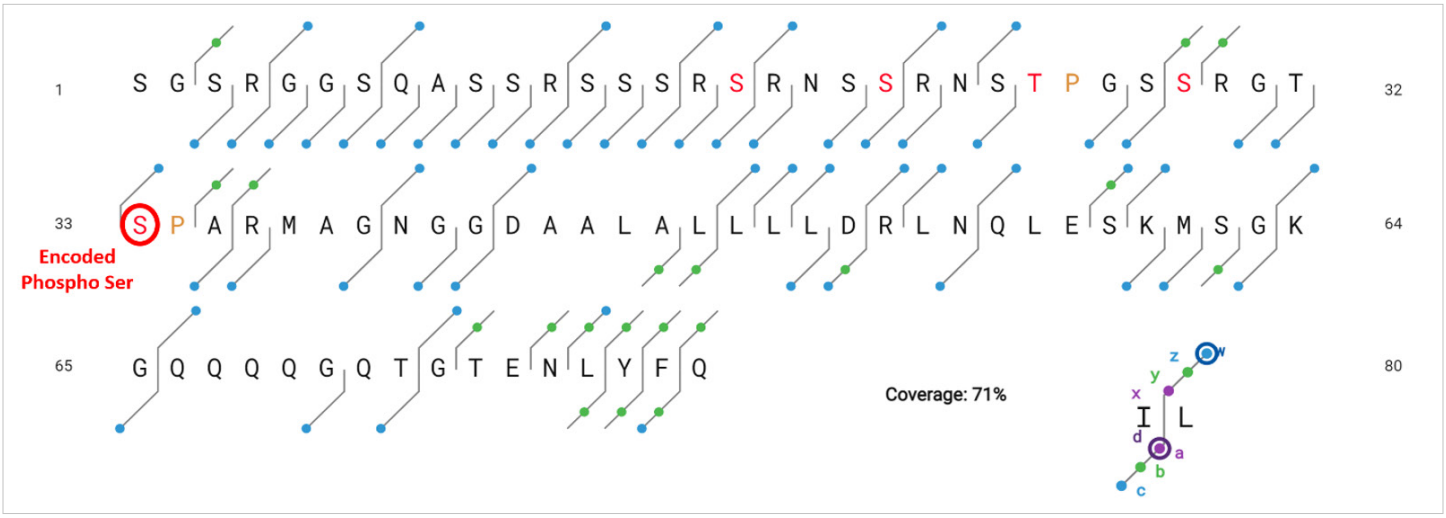

Figure 6. The sequence coverage map for the 5X phosphorylated SR-rich domain. The location of the genetically encoded phosphoserine is indicated by a red circle, while the remaining 4 sites of phosphorylation are indicated in red text.

compared to the wider range of 5-11^{+} for the singly-phosphorylated SR-rich domain. The 6^{+} charge state of the 5X phosphorylated form was isolated and subjected to ECD and CID fragmentation (Figure 6).

Fragmentation of the SR-rich domain became more difficult with increasing levels of phosphorylation; 10V of supplemental collision energy was necessary to localize all 5 phosphosites. The total sequence coverage obtained from *b*, *c*, *y*, and *z*-type ion assignments was 71% for the 5X phosphorylated SR-rich domain (Figure 6).

Conclusions

The ExD AQ-251 Option enables ECD in an Agilent 6545XT AdvanceBio LC/Q-TOF, unlocking new workflows such as top-down phosphosite mapping for difficult proteins like the SR-rich domain of the SARS-CoV-2 nucleocapsid. Using ExDViewer to perform straightforward analysis of the results, here we showed that ECD enabled extensive amino acid sequence coverage and the ability to localize multiple modifications otherwise not possible by CID alone. Furthermore, the ability to combine ECD with a small amount of CID provided a clear advantage for the analysis of multiply phosphorylated SR-rich proteins; the supplemental collisional activation facilitated detection of ECD fragment ions by disrupting the non-covalent interactions that can otherwise hold the protein together after electron capture.[9,10] Altogether, the ExD Option facilitated a powerful new approach for the mapping of critical labile modifications, offering a complementary technique to conventional bottom-up MS/MS.

References

1. Cohen, P. (2002) The origins of protein phosphorylation. *Nat Cell Bio* 4, E127–E130 https://doi.org/10.1038/ncb0502-e127
2. Gnad, F., Gunawardena, J., Mann,M.(2011) PHOSIDA 2011: the posttranslational modification database, *Nucleic Acids Research*, Volume 39, D253–D260, https://doi.org/10.1093/nar/gkq1159
3. Hoopmann, M., Kusebauch, U., Palmblad, M. et al. (2020) Insights from the First Phosphopeptide Challenge of the MS Resource Pillar of the HUPO Human Proteome Project. *J. Prot.* Res 19:12 4754-4765 DOI: 10.1021/acs.jproteome.0c00648
4. Potel, C., Lemeer, S., and Heck, A.J.R. (2019) Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal Chem* 91:1 126-141 DOI: 10.1021/acs.analchem.8b04746
5. Bouhaddou, M., Memon, D., Meyer, B., et al. (2020) The Global Phosphorylation Landscape of SARS-CoV-2 Infection. *Cell*. 182(3) 685-712.e19, ISSN 0092-8674
6. Wu, C., Yeh, S., Tsay, Y., et al. (2009) Glycogen Synthase Kinase-3 Regulates the Phosphorylation of Severe Acute Respiratory Syndrome Coronavirus Nucleocapsid Protein and Viral Replication. *Mech. Sig. Trans.* 284:8 https://doi.org/10.1074/jbc.M805747200
7. Zhu, P., Gafken, P.R., Mehl, R.A., Cooley, R.B. (2019) A Highly Versatile Expression System for the Production of Multiply Phosphorylated Proteins *ACS Chem Bio.* 14 (7), 1564-1572 DOI: 10.1021/acschembio.9b00307
8. Lopez-Clavijo, A.F., Duque-Daza, C.A., Creese, A.J., Cooper, H.J. (2015) Electron capture dissociation mass spectrometry of phosphopeptides: Arginine and phosphoserine *Int. J. of Mass Spectrom*, ISSN 1387-3806, https://doi.org/10.1016/j.ijms.2015.07.024.
9. Kim D, Pai PJ, Creese AJ, Jones AW, Russell DH, Cooper HJ. (2015) Probing the electron capture dissociation mass spectrometry of phosphopeptides with traveling wave ion mobility spectrometry and molecular dynamics simulations. *J Am Soc Mass Spectrom*. 1004-13. doi: 10.1007/s13361-015-1094-1
10. Beckman, J., Voinov, V., Hare, M., et al. (2021) Improved Protein and PTM Characterization with a Practical Electron-Based Fragmentation on Q-TOF Instruments. *J Am Soc Mass Spec* DOI: 10.1021/jasms.0c00482

SALES

e-MSion, Inc.
2121 NE Jack London Street
Suite 140
Corvallis, Oregon 97330
Sales@e-MSion.com

https://e-msion.com/products/agilent/