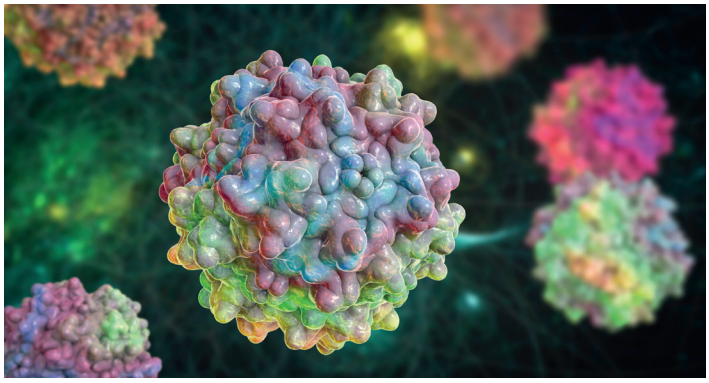
APPLICATION NOTE 000242

Adeno-associated virus capsid protein characterization and host cell protein profiling using micro-flow UHPLC-Orbitrap MS

Authors: Reiko Kiyonami[1], Jennifer Sutton[1], Katy Irvin[2], Ken Cook[3], Estee Toole[4], Min Du[5]

[1]Thermo Fisher Scientific, San Jose, CA;
[2]Thermo Fisher Scientific, Frederick, MD;
[3]Thermo Fisher Scientific, Hemel Hempstead, UK;
[4]Thermo Fisher Scientific, Florida, US;
[5]Thermo Fisher Scientific, Cambridge, MA

Keywords: Adeno-associated virus (AAV), capsid protein, viral protein, post-translational modification (PTM), host cell proteins (HCP), bottom-up analysis, peptide mapping, high resolution accurate mass (HRAM), Orbitrap technology, ultra-high-performance liquid chromatography (UHPLC), reversed-phase liquid chromatography (RPLC), micro-flow rate separation



Application benefits

- Excellent quality of MS and MS/MS data offered by Orbitrap™ mass spectrometer allows confident viral protein sequence confirmation, post-translational modification (PTM) identification and host cell protein identification
- Increased sensitivity enabled by micro-flow rate separation allows detection of low abundant PTMs on capsid proteins and low abundant host cell proteins
- Comparable robustness and quantitative reproducibility vs. analytical flow
- Ready-to-use instrument bottom-up method template for data acquisition
- Ready-to-use single informatics workflow for peptide mapping and HCP data processing and visualization

ThermoFisher SCIENTIFIC

Goals

- Develop a robust micro-flow LC-MS/MS method for sequence confirmation, site-specific PTM identification, and relative quantification of AAV viral proteins
- Develop a robust micro-flow LC-MS/MS method for AAV residual HCP impurities identification and relative quantification

Introduction

Recombinant adeno-associated viral (rAAV) vectors have emerged as the leading gene delivery vehicles for gene therapy due to their high-efficiency transduction and safety.[1,2] AAV capsid viral proteins (VPs) are important constituents of AAV product and play an important role for immunogenicity and tissue tropism in gene therapy.[3] Full characterization, including sequence and post-translational modification (PTM) identification of viral proteins is required to ensure the safety, quality, and efficacy of AAV products. In addition, AAV vector manufacturing process-related impurities, such as residual host cell proteins derived from the production cells, also need to be well characterized and controlled during the AAV vector production to mitigate immunogenicity and ensure product stability.[4] Bottom-up LC-MS/MS method is a well-established analytical approach for therapeutic protein sequence confirmation and site-specific PTM identification as well as therapeutic protein production related HCP analysis. Recently, it has been applied to AAV capsid protein characterization and corresponding HCP analysis[5-7] to support large scale rAAV vector production. Unlike therapeutic proteins, the AAV sample is more complex in structure and limited in sample volume with low titers due to small batch sizes, yielding very low amounts of capsid protein for subsequent LC-MS/MS analysis. Also, most AAVs are manufactured in human cell lines with a more complex proteome background, presenting more challenges for HCP analysis. In several AAV peptide mapping studies, nano-flow LC-MS/MS methods were chosen for increased sensitivity and to detect low abundant peptides and PTMs of AAV viral proteins.[5,6] However, nano-flow LC-MS/MS methods need to deliver peptide separation at nano-flow rates and generate ions without using any gases that are generally not supported by the high flow HPLC pump and standard ESI ion source.

In contrast, a micro-flow LC-MS/MS method can be carried out using standard UHPLC-MS platforms by simply switching from a larger i.d. column (e.g., 2.1 mm) to a smaller i.d. column (e.g., 1 mm). By using a smaller i.d. column with micro-flow rate separation, we can reduce the injection amount of AAV6 sample compared to high flow LC MS/MS, while maintaining comparable method robustness and reproducibility.

The Thermo Scientific™ Orbitrap Exploris™ 480 mass spectrometer provides excellent sensitivity over a large dynamic range by implementing a high-capacity transfer tube for maximum ion loading and an electrodynamic ion funnel that accommodates and transmits ions over a broader mass range with high efficiency.[8] 100% sequence coverage of AAV6 viral proteins was achieved with the developed micro-flow HPLC MS/MS method using an Orbitrap Exploris 480 mass spectrometer coupled with a Thermo Scientific™ Vanquish™ Horizon UHPLC system. In addition, close to 1,000 HCPs from a crude harvest AAV6 sample and 30 HCPs from the purified AAV6 sample were identified and relative quantified using the same UHPLC-Orbitrap MS platform. The analytical results are reported here.

Experimental

AAV sample

All AAV6 samples were generated from transient transfection in HEK293 cell.

For AAV viral protein peptide mapping analysis:
A purified and concentrated AAV6 sample was used. The titer of the AAV6 sample after concentration was 2.9E13 vg/mL determined using qPCR.

For AAV HCP analysis:
Sample 1: Crude harvest AAV6 sample. The titer was 1.0E11 vg/mL.
Sample 2: POROS AAVX affinity resin purified AAV6 sample (shorthand as purified sample in the experimental and results sessions). The titer was 2.0E12 vg/mL.

Sample preparation

Peptide mapping analysis:
50 µL each of the purified and concentrated AAV6 sample was buffer exchanged to water containing 5% formic acid and 5 mM TCEP. The buffer exchanged AAV6 sample was enzymatically digested in-solution using pepsin (1 mg/mL in the buffer adjusted to pH 2.0) at 70 ˚C over 1.5 h. The digestion was stopped by heating at 95 ˚C for 10 minutes. The digested sample was used for LC-MS/MS analysis.

HCP analysis:
250 µL each of the crude harvest and purified AAV6 samples containing spiked-in 200 ng intact protein (*Streptococcus* Protein AG_chimeric[9]) was buffer exchanged to 7 M Guanidine HCl 100 mM Tris, respectively. The buffer exchanged AAV samples were reduced with DDT (dithiothreitol) and alkylated with IAC (iodoacetic acid). The reduced and alkylated samples were further buffer exchanged to 50 mM Tris and enzymatically digested in-solution using trypsin at 37 ˚C over 2.0 h. The digestion was terminated by the addition of 10% formic acid. The digested sample was used directly for LC-MS/MS analysis.

Chromatography

For all experiments, chromatographic separations were carried out using a Thermo Scientific™ PepMap™ 100 C18 column (1.0 × 150 mm, 3 µm) on the Thermo Scientific™ Vanquish™ Horizon UHPLC system, consisting of the following modules:

- Thermo Scientific™ System Base Vanquish™ Horizon/Flex (P/N VF-S01-A-02)
- Thermo Scientific™ Vanquish™ Binary Pump H (P/N VH-P10-A)
- Thermo Scientific™ Vanquish™ Split Sampler HT (P/N VH-A10-A)
- Thermo Scientific™ Vanquish™ Column Compartment (P/N VH-C10-A)

Solvent A: Water with 0.1% formic acid

Solvent B: Acetonitrile with 0.1% formic acid.

Chromatographic separation details for the peptide mapping and HCP analysis are summarized in Tables 1 and 2.

Table 1. LC gradient condition for peptide mapping

Time	Flow (mL/min)	%A	%B
0	0.08	98	2
1	0.08	95	5
77	0.08	65	35
81	0.08	10	90
83	0.08	10	90
85.1	0.08	98	2
90	0.08	98	2

Table 2. LC gradient condition for HCP analysis

Time	Flow (mL/min)	%A	%B
0	0.08	98	2
5	0.08	98	2
6	0.08	95	5
106	0.08	65	35
113	0.08	10	90
117	0.08	10	90
117.1	0.08	98	2
125	0.08	98	2

The column temperature was set to 50 ˚C in Still-Air mode. For peptide mapping analysis, 5 µL of pepsin digested sample were injected on column. For HCP analysis, 60 µL of two trypsin digested samples were injected on column in triplicate.

Mass spectrometry

The Orbitrap Exploris 480 (P/N BRE725533) mass spectrometer was used for MS data collection. The instrument was operated with Thermo Scientific™ Xcalibur™ 4.2 SP1 software and controlled by Orbitrap Exploris Series 3.0 instrument control software. MS/MS data acquisition was carried out using a data-dependent MS/MS setup. The same MS settings were used for both peptide mapping and HCP experiments (Table 3).

Table 3. MS parameter settings for peptide mapping and HCP analysis

MS source setting	Value
Sheath gas	25
Aux gas	8
Sweep gas	0
Spray voltage (+V)	3,500
Capillary temp. (˚C)	250
Vaporizer temp. (˚C)	100
MS full MS/dd MS2 (top10) setting	**Value**
General	
Application mode	Peptide
Pressure mode	Standard
RF lens (%)	50
Full MS	
Scan range (*m/z*)	300–1,800
Resolution	60,000 at *m/z* 200
AGC target value (%)	300
Max inject time (ms)	100
dd-MS/MS (top 10)	
Resolution	15,000 at *m/z* 200
Isolation window (*m/z*)	2
AGC target value (%)	100
Max inject time (ms)	200
Fixed first mass (*m/z*)	75
Targeted mass exclusion	On
HCD collision energy (V)	28

Data processing

Both peptide mapping (sequence confirmation, site specific PTM identification, and relative quantification) and HCP (identification and relative quantification) analysis data processing and visualization were performed using Thermo Scientific™ BioPharma Finder™ 4.1 software. The software settings used for peptide mapping data analysis are listed in Table 4. In the Biopharma Finder 4.1 software, the HCP analysis capability, which allows host cell protein identification and quantitation, has been added to the peptide mapping analysis and combines product-specific peptide mapping results with host cell protein identification. This easy-to-use workflow starts with a peak list where the software identifies all the product specific peptides and PTMs then remaining unmatched peaks are searched using Comet[10] against larger host cell protein databases (e.g., CHO, Human, Mouse). The results are then merged. Automatic HCP quantitation is reported using the top 3 most abundant peptides to determine the average MS area for the protein.[11] The BioPharma Finder software settings used for HCP data analysis are listed in Table 5.

Table 4. BioPharma Finder 4.1 software parameter settings for peptide mapping data analysis

Component detection	Setting
Absolute MS signal threshold	2.00E+05
Typical chromatographic peak width (min)	0.17
Relative MS signal threshold (% base peak)	1.00
Relative analog threshold (% of highest peak)	1.00
Width of Gaussian filter (represented as 1/n of chromatographic peak width)	3.00
Minimum valley to be considered as two chromatographic peaks	80
Minimum MS peak width (Da)	1.20
Maximum MS peak width (Da)	4.20
Mass tolerance (ppm for high-res or Da for low-res)	6.00
Maximum retention time shift (min)	1.73
Maximum mass (Da)	30,000
Mass centroiding cutoff (% from base)	15
Identification	**Setting**
Maximum peptide mass	11,000
Mass accuracy (ppm)	5
Minimum confidence	0.8
Maximum number of modifications for a peptide	1
Unspecified modification	–
N-glycosylation	CHO
Protease	Pepsin
Protease specificity	High
Static modifications	**Setting**
N-term	Acetylation
Variable modifications	**Setting**
Side chain	Deamidation (N), Double oxidation, Glycation, H_2O loss, Hydroxylation, Mannosylation (S), NH_3 loss, Oxidation (MW), Phosphorylation (STY)

Table 5. BioPharma Finder 4.1 software parameter settings for HCP data analysis

Basic parameters	Setting
Protein database	Uniprot-proteome_ UP000005640_both swiss_prot_ TrEMBL_fasta with Streptococcus Protein AG_chimeric
Acquisition type	High-High (MS1 and MS2)
Precursor mass tolerance	10 ppm
Ions to search	a ion, b ion, y ion, NL ion
Mass range (MH+ peptide mass)	600–5,000
E-value cutoff	0.0005
Protease parameters	**Setting**
Protease termini	Fully digested
Max number internal miscleavages	1
Enable decoy search	Yes
Modifications	**Setting**
Static side chain	Carboxymethylation
Max # of variable modification per peptide	1
Variable side chain	–
Protein terminal modification	–
Advanced parameters	**Setting**
Enable methionine protein N-term clip	No

Results and discussion

Peptide mapping analysis

Although trypsin is the most commonly used enzyme for peptide mapping analysis, it is difficult to get complete viral protein sequence coverage using trypsin because of AAV capsid high structural stability and the suboptimal localization of trypsin cleavage sites in the AAV protein sequences. We used pepsin instead of trypsin for our peptide mapping analysis because a recent study[12] has shown that pepsin is thermally stable and active in acidic conditions and can improve characterization of commonly challenging protein sequence sections within viral protein sequences.

The base peak chromatogram of full MS obtained from the pepsin digest mixture is shown in Figure 1. The peptide mixture was well separated at a 80 µL/min flow rate using the PepMap 100 C18 column.

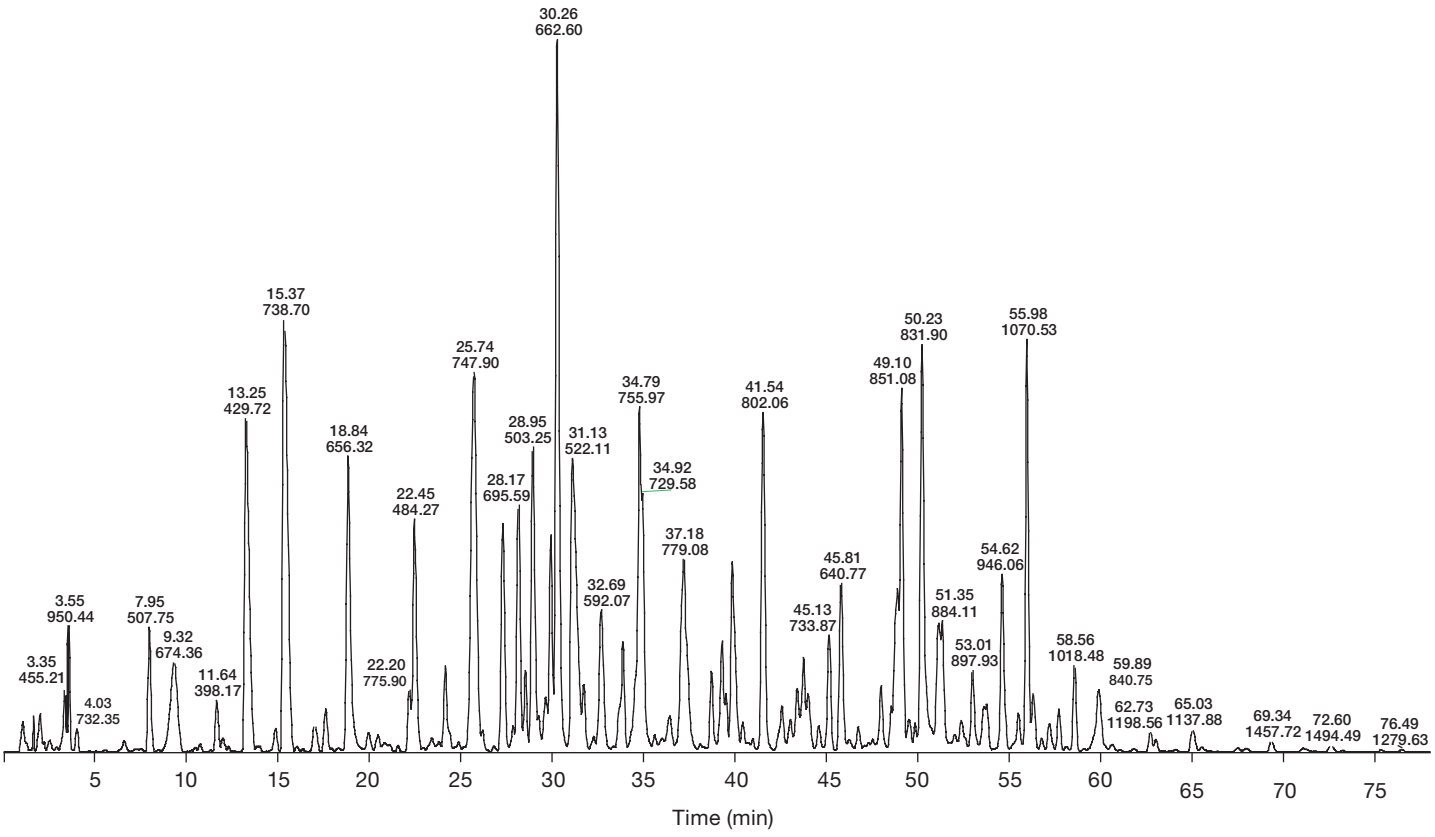


Figure 1. Base peak chromatogram of full scan MS from the pepsin digestion mixture of the concentrated AAV 6 sample

The raw data was processed using the peptide mapping workflow in the Biopharma Finder 4.1 software. 100% VP1 sequence coverage was achieved using high-resolution accurate mass MS and MS/MS data (Figure 2). Viral proteins share the same C-terminus sequence and only differ in the N-terminus sequences. It is therefore important to confirm the N-terminal sequence of each viral protein to identify all the viral proteins. The great spectral quality and high sensitivity of MS and MS/MS data obtained on the micro-flow UHPLC-Orbitrap Exploris 480 MS platform enabled confident identification of VP1, VP2, and VP3 N-terminus sequences (Figure 3) with or without acetylation.

Proteins	Number of MS peaks	MS peak area	Sequence coverage	Abundance (mol)
1:tr|O56137|O56137_9VIRU Capsid protein VP1 OS=Adeno-associated virus – 6 OX=68558	4,346	87.0%	100.0%	100.00%

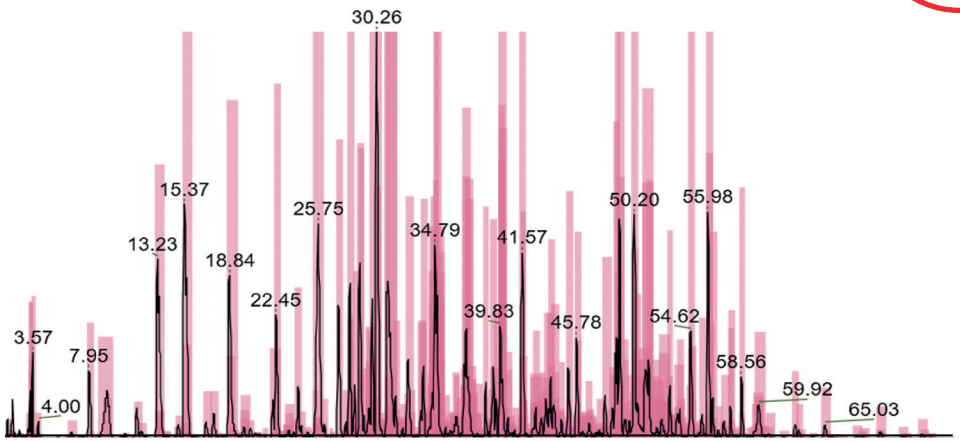


Figure 2. AAV6 VP1 sequence coverage from the peptide mapping analysis. The identified VP1 peptide peaks are highlighted with red. 100% sequence coverage was achieved in a single UHPLC-MS/MS run using pepsin enzyme.

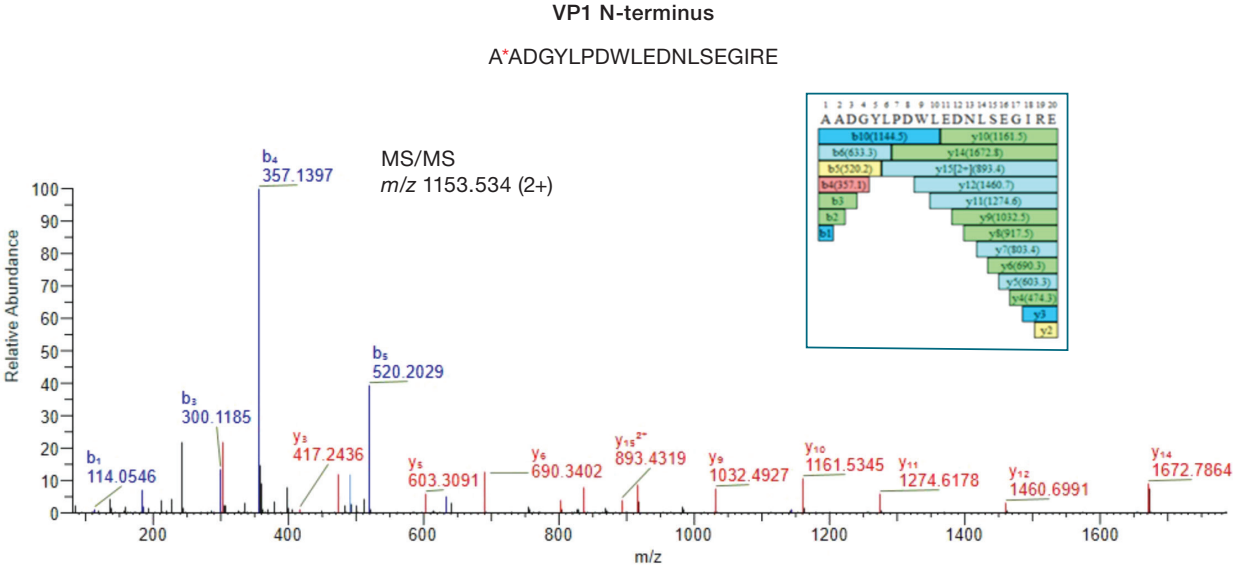


Figure 3. N-terminus sequences of AAV6 VP1, VP2, and VP3 proteins were confidently identified with high-quality MS/MS data.

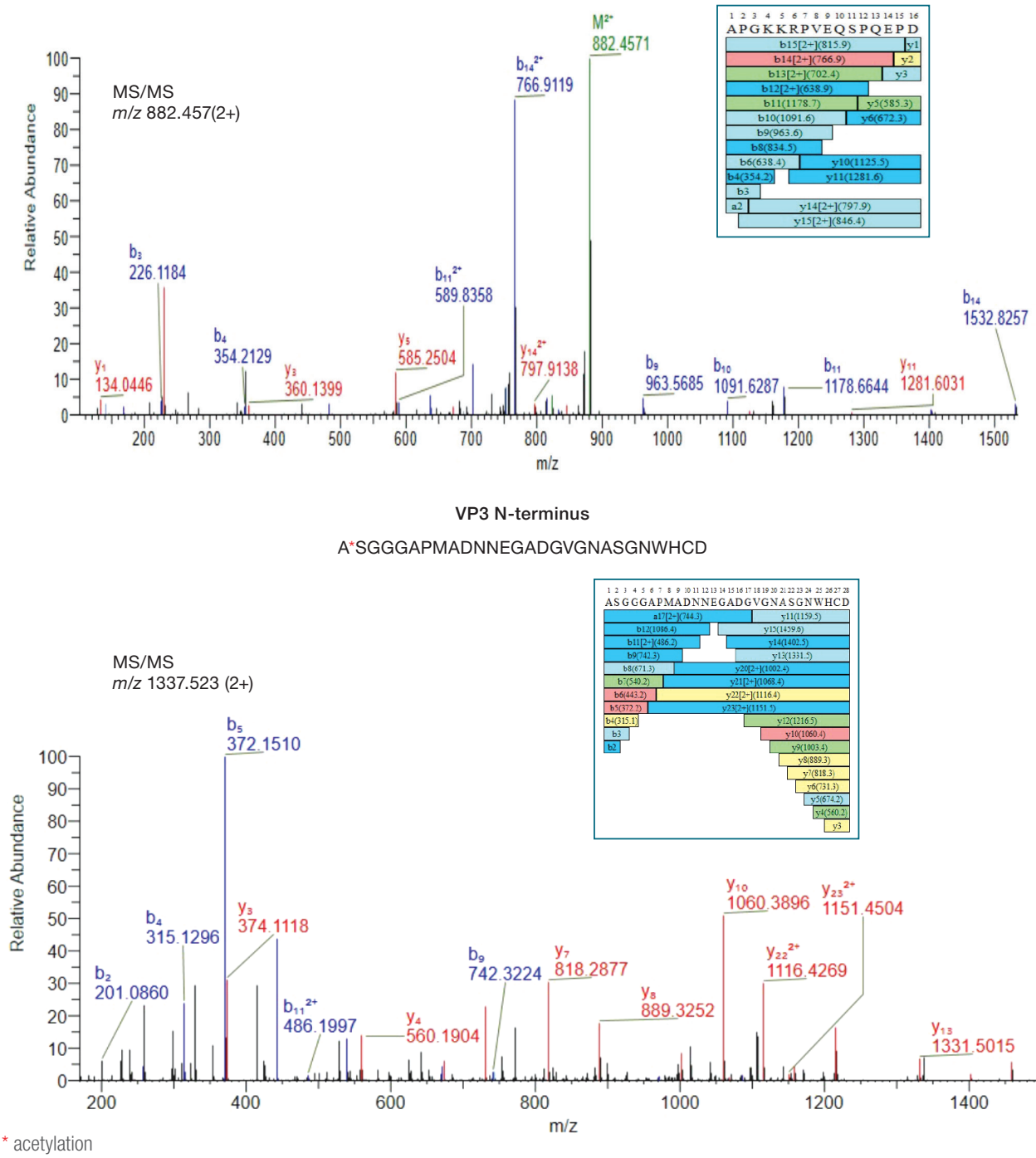


Figure 3 continued. N-terminus sequences of AAV6 VP1, VP2, and VP3 proteins were confidently identified with high-quality MS/MS data.

In addition to acetylation, other post-translational modifications, such as phosphorylation, deamidation, and oxidation of the viral proteins, were also identified from the high-quality MS and MS/MS data. Figure 4 shows the phosphorylation site on the peptide YYLNRTQNQSGSAQNKDLL confidently identified by the MS/MS data. The integrated peak area observed from the phosphorylated peptide was only 0.4% of the peak area observed relative to the non-modified.

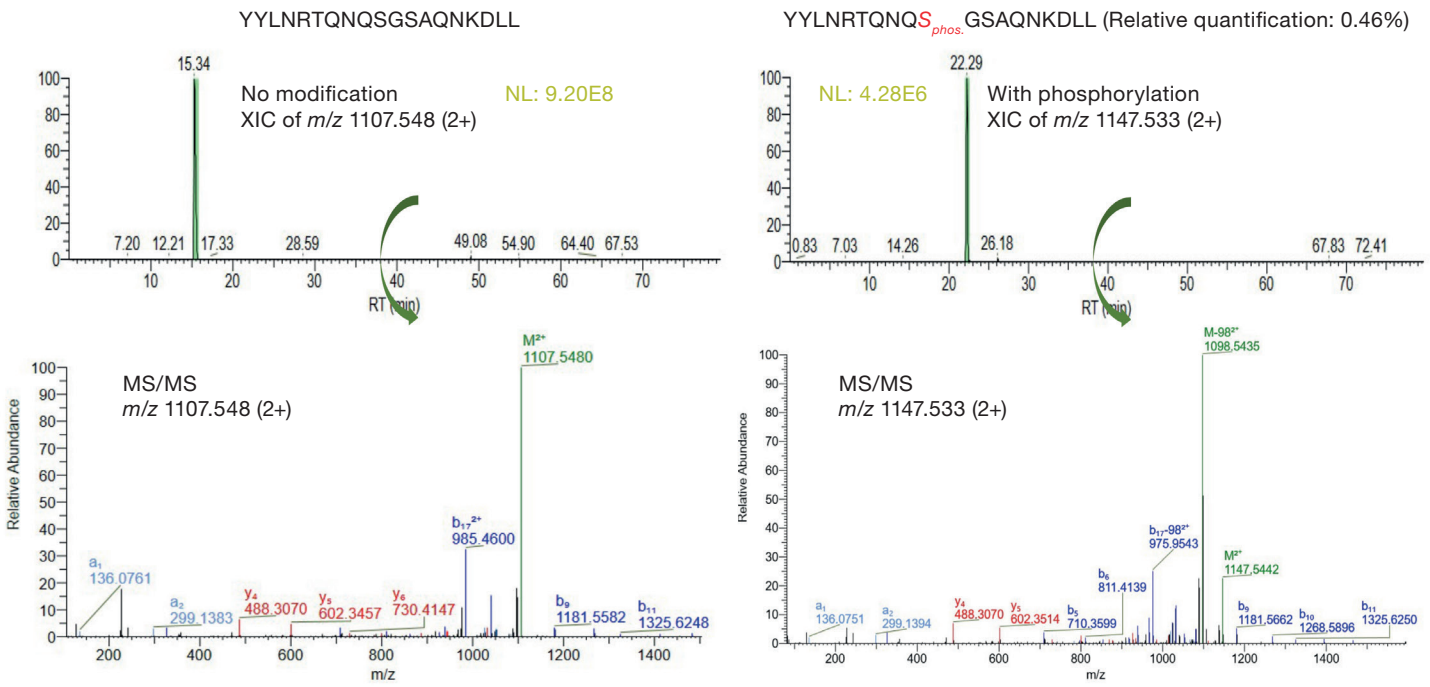


Figure 4. Confident phosphorylation identification with high-quality MS/MS data and relative quantification of the phosphorylated peptide vs. non-phosphorylated peptide using the integrated peak areas of the precursor ions

Figure 5 shows the MS/MS identification results of a peptide without and with oxidation. The low abundant oxidized peptide (0.6%) was confidently identified with high-quality MS/MS data.

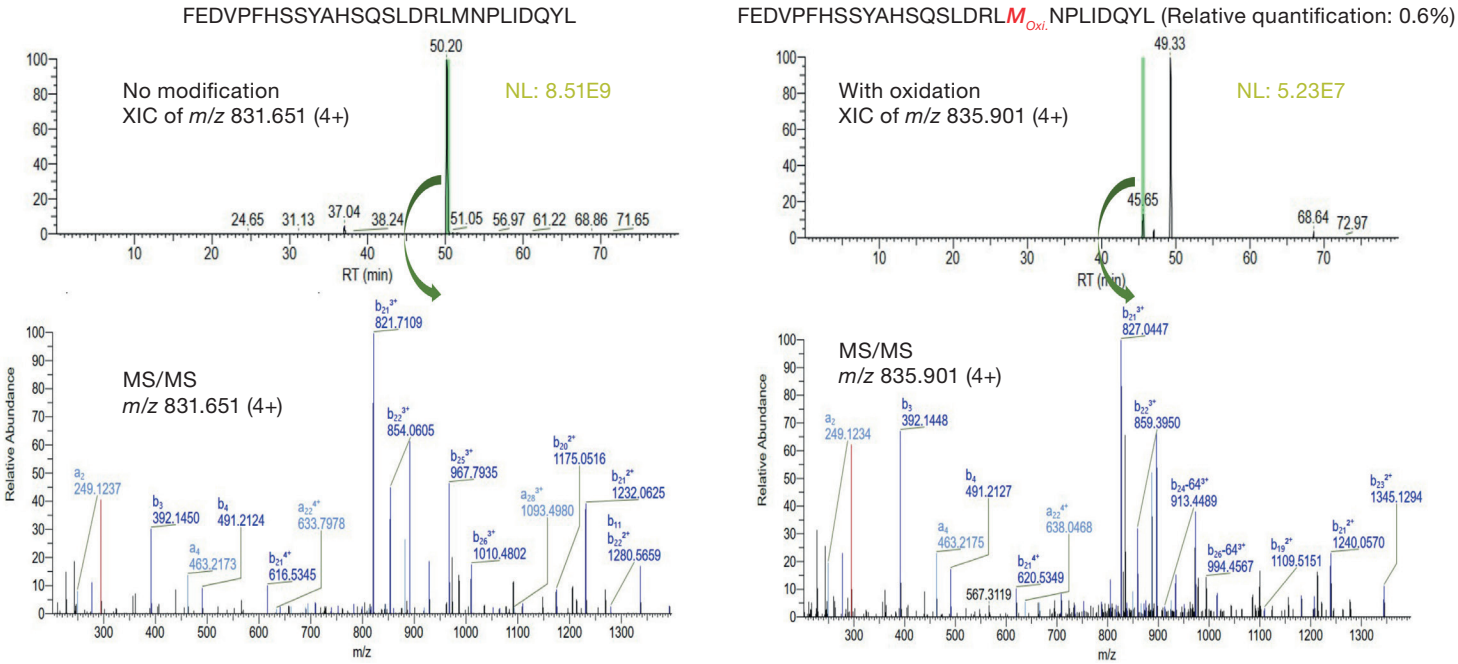


Figure 5. Confident low abundant oxidation identification with high-quality MS/MS data and relative quantification of the oxidized peptide vs. non-oxidized peptide using the integrated peak areas of the precursor ions

Another example of low abundant PTM, deamidation identification is shown in Figure 6. The relative abundance of the deamidated peptide (YAKSAN$V_{deami.}$DF) was calculated as 0.33%.

HCP identification and relative quantification

The trypsin digests of the crude AAV6 sample and the purified AAV6 sample were each analyzed in triplicate. Base peak chromatograms from the triplicate runs of purified AAV6 sample are shown in Figure 7. Excellent retention time reproducibility and separation efficiency were achieved using the micro-flow separation.

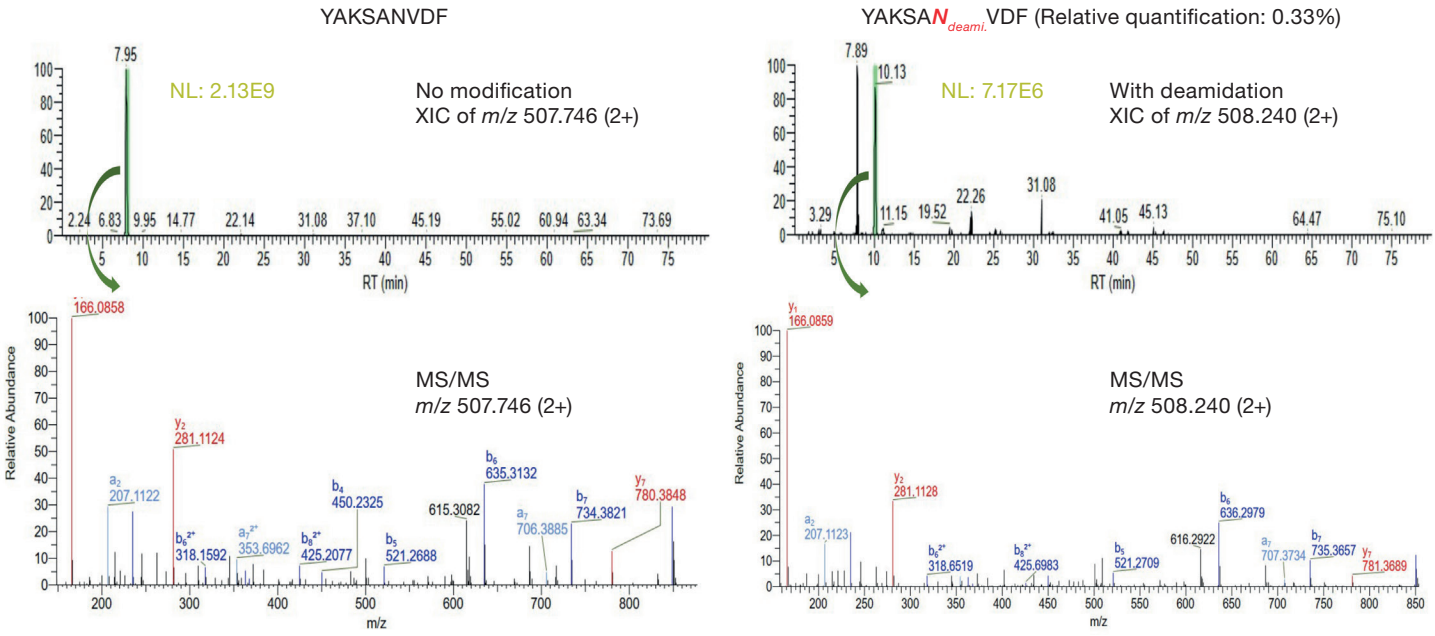


Figure 6. High-quality MS/MS data enables confident identification of low abundant deamidated peptide and relative quantification of the deamidated peptide vs. non-deamidated peptide using the integrated peak areas of the precursor ions.

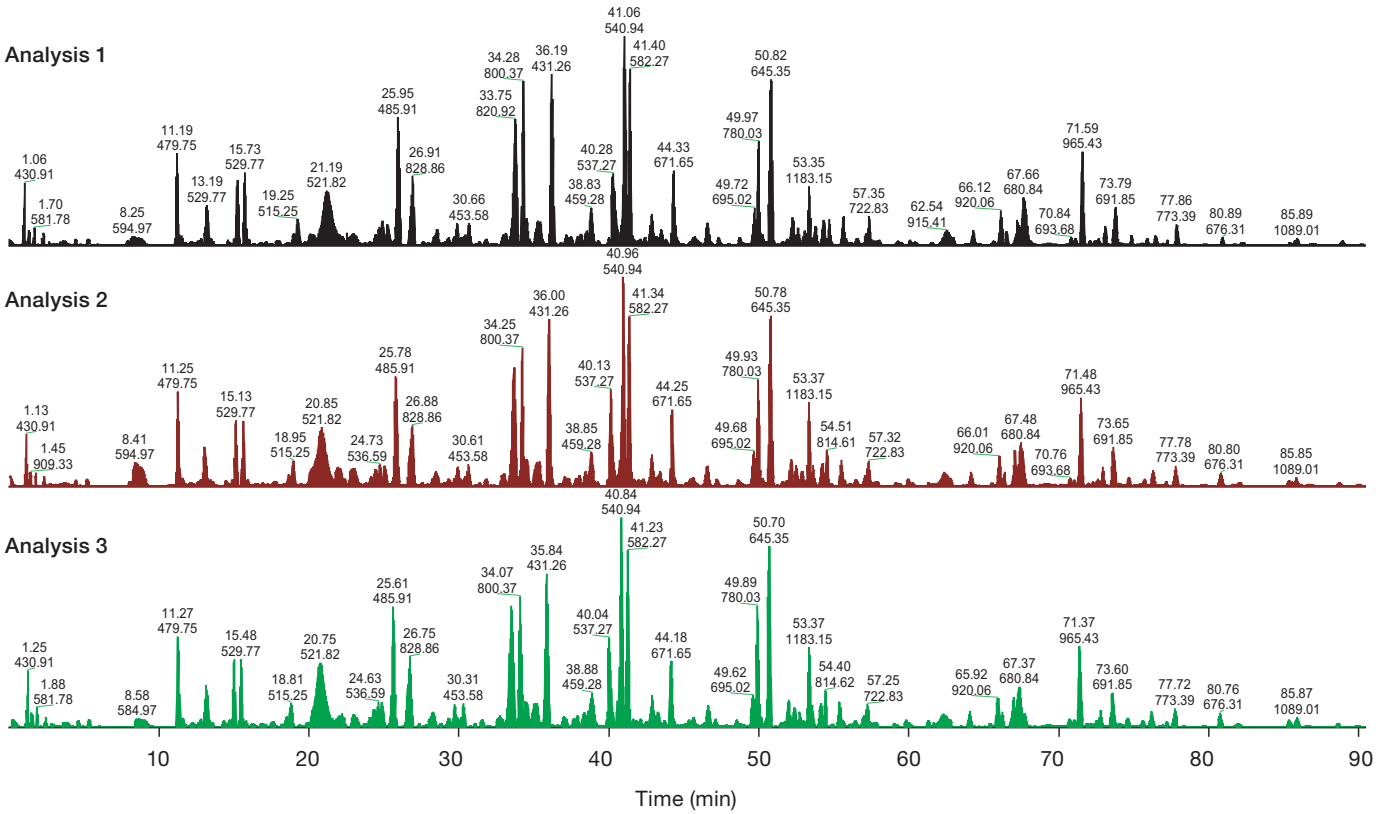


Figure 7. Base peak chromatograms of purified AAV6 sample for HCP analysis

The six raw files were processed using the peptide mapping workflow via the host cell protein analysis feature in Biopharma Finder 4.1 software. The Human-UniProt-proteome database appended with the sequence for *Streptococcus* protein was used for HCP identification.

In the crude harvest AAV6 sample, 817 HCPs were identified with at least three unique peptides. With the high separation efficiency and great sensitivity offered by this micro-UHPLC-MS/MS approach, the MS/MS data were obtained over a wide dynamic range. As shown in Figure 8, although the peaks of detected peptides from the HCP – myristoylated alanine-rich C-kinase substrate (highlighted in purple) were buried in the background because of their low intensity signals, the Orbitrap Exploris 480 mass spectrometer still triggered MS/MS acquisition and generated high-quality MS/MS spectral data for confident peptide identification. Using the spiked-in 200 ng intact *Streptococcus* Protein AG_chimeric as reference, the estimated concentration of this HCP was calculated to be 3.6 pmol/mL in the crude harvest AAV6 sample using the average area of top 3 peptides.[11] In addition, the integrated peak area values for these low abundant peptides demonstrated good reproducibility. The percentage coefficient of variance of the average peak area (top 3 peptides) across the triplicate runs was 11%.

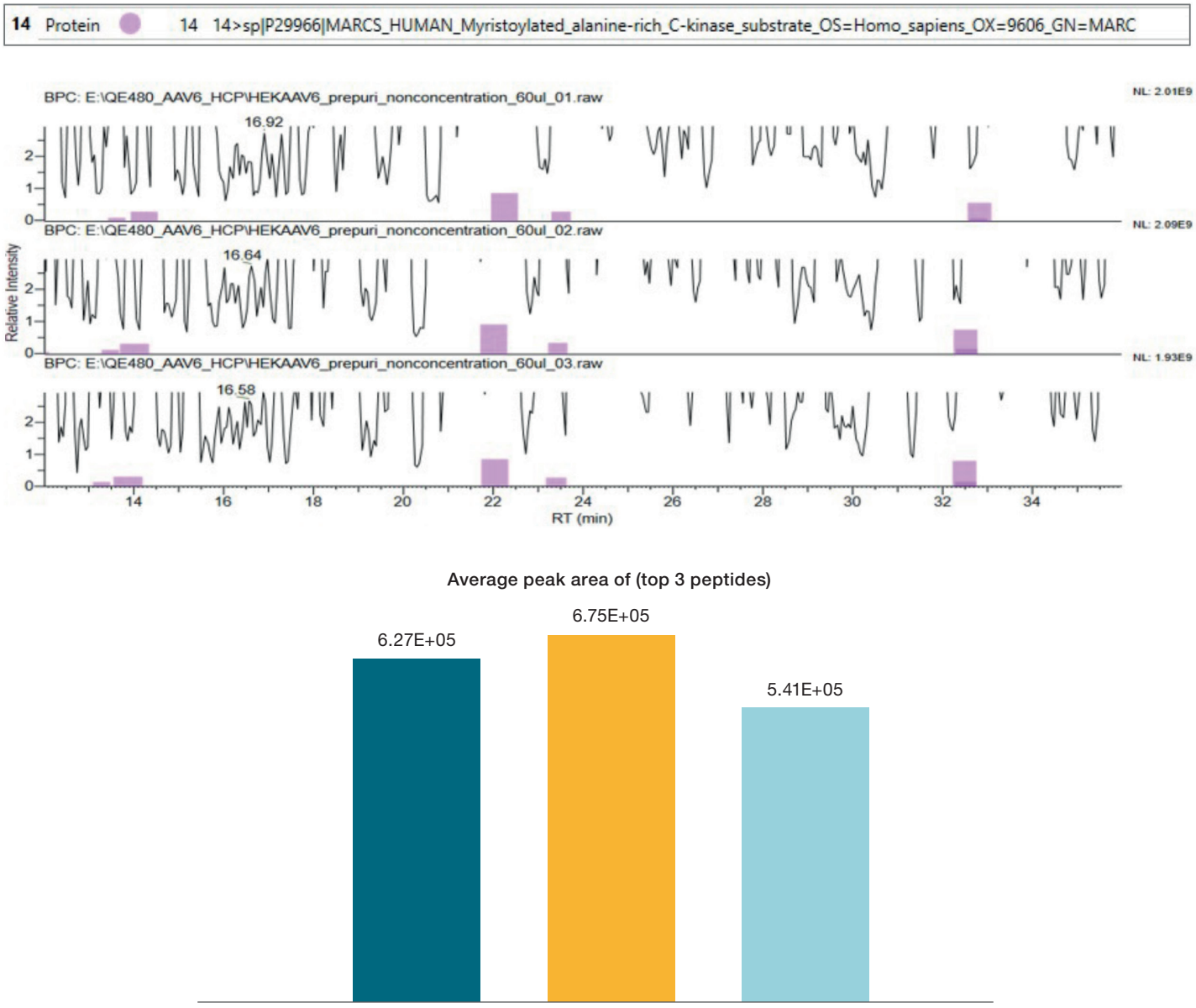


Figure 8. Example of low abundant HCP (estimated concentration: 3.6 pmol/mL) identification and quantification from triplicate runs of the crude harvest AAV6 sample. The purple bars show the peptide peaks detected for this low abundant HCP.

The number of HCP decreased significantly after the purification and only 30 HCPs that have at least three unique peptides were identified from the purified AAV6 sample (Figure 9). Figure 10 shows the relative quantification results for several HCPs identified in both crude harvest and purified AAV6 samples. Overall, all HCPs showed lower concentrations in the purified AAV sample, demonstrating that the POROS AAVX affinity resin was able to remove the HCPs efficiently from the AAV6 harvest.

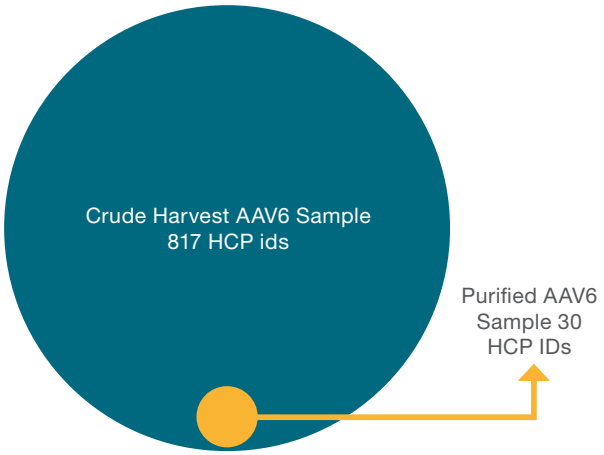


Figure 9. The numbers of identified HCPs with at least three unique peptides from the harvest vs. purified AAV6 samples using the Human-UniProt-proteome database

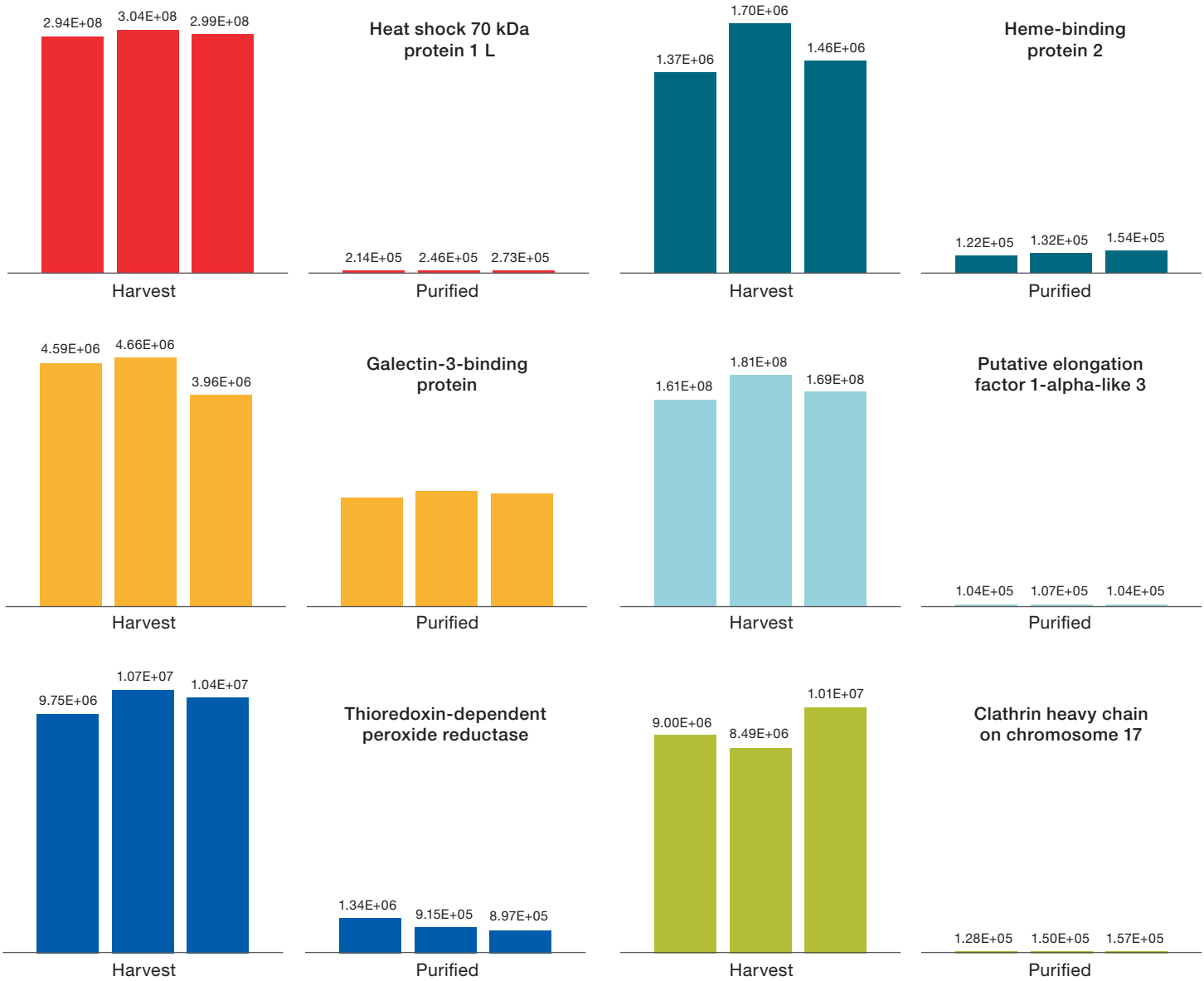


Figure 10. Quantitative trend of representative HCPs between the crude harvest and purified AAV6 samples using the average peak area of the top 3 peptides

thermo scientific

Conclusion

- A robust micro-flow LC-MS/MS method was developed using the Orbitrap Exploris 480 mass spectrometer equipped with the Vanquish Horizon UHPLC system and applied successfully for peptide mapping of AAV6 viral proteins and HCP profiling of pre-purified and purified AAV6 samples.
- 100% AAV6 viral protein sequence coverage was achieved from a single LC-MS/MS run using pepsin for AAV6 sample digestion.
- High sensitivity and high-quality MS and MS/MS data offered by the developed method allowed confident low abundant site-specific PTM identification and relative quantification of the viral proteins.
- POROS AAVX affinity resin offered effective removal of the HCPs from the crude harvest AAV6 sample during the purification process.
- A single software, Biopharma Finder 4.1, allowed data processing for both peptide mapping analysis and HCP analysis.

References

1. Samulski, R.J.; Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes, *Annu. Rev. Virol.* **2014**, *1*(1), 427–451.
2. Naso, M.F.; Tomkowicz, B.; Perry, W.L.; Stroh, W.R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy, *BioDrugs* **2017**, *31*, 317–334.
3. Popa-Wagner, R.; Porwal, M.; Kann, M.; Reuss, M.; Weimer, M.; Florin, L.; Kleinschmidt, J. A. Impact of VP1-Specific Protein Sequence Motifs on Adeno-Associated Virus Type 2 Intracellular Trafficking and Nuclear Entry. *Journal of Virology,* **2012**, *86*(17), 9163–9174.
4. Wright. J. F. Product-Related Impurities in Clinical-Grade Recombinant AAV Vectors: Characterization and Risk Assessment, *Biomedicines* **2014**, *2*, 80-97.
5. Jin, X.; Liu, L.; Nass, S.; O'Riordan, C.; Pastor, E.; Kate Zhang, X. Direct Liquid Chromatography/Mass Spectrometry Analysis for Complete Characterization of Recombinant Adeno-Associated Virus Capsid Proteins. *Human Gene Therapy Methods* **2017**, *28*(5), 255-267.
6. Rumachik, N.G.; Malaker, S.A.; Paulk, N.K. VectorMOD: Method for Bottom-Up Proteomic Characterization of rAAV Capsid Post-Translational Modifications and Vector Impurities. *Frontiers in Immunology*, Methods article, published: 01 April 2021. https://doi.org/10.3389/fimmu.2021.657795
7. Johnson, S.; Wheeler, J. X.; Thorpe, R.; Collins, M.; Takeuchi, Y; Zhao, Y. Mass Spectrometry Analysis Reveals Differences in the Host Cell Protein Species Found in Pseudotyped Lentiviral Vectors. *Biologicals* **2018** Mar; *52*, 59–66.
8. Orbitrap Exploris 480 mass spectrometer. https://planetorbitrap.com/orbitrap-exploris-480
9. Viner, R.; Cardasis, H.L.; Vitkovske, V.; Lin, S.; Ling, R.V.; Taujenis, L.; Gohil, V.; Opperman, K.; Rogers, R.; Capkauske, E.; Bargaila, K.; Alminaite, A.; Siurkus, J. Development of an All-Recombinant Intact Protein Standard for LC & MS Application Development & System Suitability Testing. *Thermo Scientific poster*, PO21707-EN 0417S. https://assets.thermofisher.com/TFS-Assets/CMD/posters/PO-21707-LC-MS-Inact-Protein-Standard-HPLC2017-PO21707-EN.pdf
10. Eng, J.K.; Jahan, T.A.; Hoopmann, M.R. Comet: An Open-source MS/MS Sequence Database Search Tool. *Proteomics* **2013**, *13*(1), 22-4.
11. Silva, J.C.; Gorenstein, M.V.; Li, G.Z.; Vissers, J.P.C.; Geromanos, S.J. Absolute Quantification of Proteins by LCMS^{E}. *Mol Cell Proteomics* **2006**, *5*(1), 144-56.
12. Toole, E.N.; Dufresne, C.P.; Ray, S.; Schwann, A.; Cook, K.; Ivanov; A.R. Rapid Highly Efficient Digestion and Peptide Mapping of Adeno-Associated Viruses. *Anal. Chem.* **2021**, *93*(30), 10403–10410. https://doi.org/10.1021/acs.analchem.1c02117

Find out more at **thermofisher.com**

ThermoFisher SCIENTIFIC