



OPEN ACCESS

EDITED BY

Yogan Khatri,
Cayman Chemical, Ann Arbor, United States

REVIEWED BY

Gustavo Martos,
Bureau International des Poids et mesures,
France
Bart Van Puyvelde,
Ghent University, Belgium

*CORRESPONDENCE

Kelvin H. Lee,
✉ KHL@udel.edu

[†]These authors have contributed equally to
this work

RECEIVED 18 February 2025

ACCEPTED 23 April 2025

PUBLISHED 02 May 2025

CITATION

Leibiger TM, Min L and Lee KH (2025) A
comparison of SWATH-MS methods for
measurement of residual host cell proteins in
adeno-associated virus preparations.
Front. Bioeng. Biotechnol. 13:1579098.
doi: 10.3389/fbioe.2025.1579098

COPYRIGHT

© 2025 Leibiger, Min and Lee. This is an open-
access article distributed under the terms of the
[Creative Commons Attribution License \(CC BY\)](#).
The use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in this
journal is cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

A comparison of SWATH-MS methods for measurement of residual host cell proteins in adeno-associated virus preparations

Thomas M. Leibiger[†], Lie Min[†] and Kelvin H. Lee^{*}

Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, United States

Introduction: Analysis of residual host cell proteins in adeno-associated virus (AAV) preparations is challenging due to low availability and high complexity of samples. One strategy to address these challenges is through development of improved liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods with greater sensitivity and reduced sample requirement.

Methods: In this work, we compare the performance of four sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) methods for identification and quantitation of residual HCPs in rAAV2, -5, -8, and -9 preparations produced with human embryonic kidney 293 (HEK293) cells and purified using immunoaffinity chromatography. Key SWATH-MS parameters including spectral library construction (data dependent vs. *in silico*), data processing software (DIA-NN vs. Skyline), and mass spectrometer instrument (Sciex TripleTOF 6600 vs. Sciex ZenoTOF 7600) were assessed. Method attributes including sample requirement and processing time, and method outputs including protein and precursor identifications, host cell protein quantitation comparisons across methods, and quantitation coefficients of variance (CV) were considered to help establish a SWATH-MS workflow well-suited for rAAV HCP analytics.

Results: A 78% increase in HCP identifications, 80% reduction in sample requirement, and 70% reduction in instrument runtime was achieved with an *in silico* spectral library, data processing in DIA-NN, and data collection with the Sciex ZenoTOF 7600 instrument (DIA-NN-7600 method) compared to a previously established method using a DDA-derived spectral library, data processing in Skyline, and data collection with the Sciex TripleTOF 6600 instrument (Skyline-DDA-6600 method). Additionally, the DIA-NN-7600 method shows median HCP quantitation CV below 10% for triplicate data acquisitions, and comparable quantitation to other methods for a panel of highly abundant residual HCPs previously identified in rAAV downstream processing.

Discussion: This work highlights a SWATH-MS method with data collection and processing specifically tailored for rAAV residual HCP analysis.

KEYWORDS

host cell proteins (HCPs), adeno-associated virus (AAV), mass spectrometry, SWATH-MS, DIA-NN, data independent acquisition (DIA), liquid chromatography-tandem mass spectrometry (LC-MS/MS)

1 Introduction

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) tools are crucially important for the in-depth understanding of biological systems, with proteomic applications spanning from the study of cellular mechanisms to drug discovery and biopharmaceutical process development. Various “bottom-up” LC-MS/MS approaches have been developed for detection, quantitation, and monitoring of peptides derived from proteolytic digestion of protein samples (Yates, 1998). Data dependent acquisition (DDA) LC-MS/MS allows for qualitative analysis of diverse protein mixtures using spectrum-based peptide matching to a reference protein database (Zhang et al., 2021). However, due to selection of only the most abundant precursor ions for fragmentation, DDA provides limited resolution on lower-abundances species (Zhang et al., 2013). Targeted acquisition methods such as selected reaction monitoring (SRM) using a triple quadrupole mass spectrometer or parallel reaction monitoring (PRM) with a Orbitrap or Q-TOF instrument can be used for quantitation of selected proteins (Lange et al., 2008; Peterson et al., 2012). For these approaches, pre-defined peptide ions are filtered and fragmented and a select set of fragment ions are measured (SRM), or a full MS/MS spectrum is acquired (PRM). While these targeted approaches can be suitable for providing highly accurate quantitation and monitoring of sample components, they are limited to analysis of several hundred peptides per sample due to the cycle time limitations for monitoring precursor-product ion transitions or collection of full MS/MS spectra for each selected peptide ion (Ronsein et al., 2015; Gotti et al., 2021). Data independent acquisition (DIA) LC-MS/MS allows for sensitive detection and quantitation of peptides by continuous collection of fragment ion spectra for all precursors (Ludwig et al., 2018). The most common DIA methods use sequential precursor isolation and fragmentation windows applied across a broad mass-to-charge (m/z) range allowing for full precursor ion coverage and MS/MS scanning (Venable et al., 2004; Gillet et al., 2012; Zhang et al., 2020). These DIA methods combine the broad coverage of DDA with the reproducible label-free protein quantitation associated with targeted acquisition methods (Ludwig et al., 2018; Gillet et al., 2012; Zhang et al., 2020).

DIA proteomic methods such as sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) are powerful tools within the field of bioprocessing for characterization of residual host cell protein (HCP) content which is a critical quality attribute for biologically-derived therapeutics (Jones et al., 2021; Walker et al., 2017; Carvalho et al., 2024). These tools have been widely applied for analysis of residual HCPs in monoclonal antibody (mAb) production systems

to understand underlying mechanisms of impurity persistence (Herman et al., 2023a; Herman et al., 2023b), design improved downstream polishing steps (Ito et al., 2024), and characterize HCP-specific impacts to drug stability and safety (Jones et al., 2021; Wang et al., 2009; Li et al., 2021; Chiu et al., 2017). However, the reporting of specific SWATH-MS methods for HCP characterization of other therapeutic modalities such as recombinant adeno-associated virus (rAAV) vectors is limited. rAAV is the most widely used viral vector for *in vivo* gene therapy applications with over 200 completed or ongoing clinical trials and 7 FDA approved products for the treatment of monogenic diseases (Clinicaltrials.gov, 2025; Clement and Grieger, 2016; Goswami et al., 2019). Despite residual HCP content being a critical quality attribute for rAAV products (Kontogiannis et al., 2024), directly applying previously established SWATH-MS methods for detection and monitoring of residual HCPs in rAAV bioprocesses presents unique challenges. Low rAAV titers necessitate larger-scale process development experiments to isolate enough purified material to meet LC-MS/MS mass injection targets (Wright, 2023). Additionally, rAAV harvest processes result in complex mixtures of both intracellular and secreted HCPs that may persist across one or more downstream unit operations (Vandenberghe et al., 2010). A SWATH-MS workflow suitable for rAAV residual HCP analytics combines broad protein identification and quantitation at a low coefficient of variance (CV) with low sample requirement.

DIA LC-MS/MS performance depends on both acquisition methods and software tools used for data processing. Gotti et al. benchmarked DIA acquisition methods and analysis software including both Skyline and DIA-NN using a panel of 48 human proteins spiked into an *E. coli* proteome background (Gotti et al., 2021). Further studies have used benchmark datasets to understand the impact of spectral library generation and DIA data processing tools on data outputs (Fröhlich et al., 2022; Lou et al., 2023). These evaluations capture the importance of data collection and processing parameters but are often performed using model reference datasets or through analysis of a select number of proteins spiked into a complex heterogeneous background mixture. In this work, SWATH-MS data acquisition and processing methods were explored in the specific context of residual HCP analysis for rAAV vectors with the goals of increasing HCP identifications and reducing sample requirement while yielding consistent protein quantitation and CV across multiple injections. Four rAAV serotypes (rAAV2, -5, -8, and -9) produced using a commercial HEK293 culture system and purified using POROS™ CaptureSelect™ AAVX affinity chromatography were analyzed for residual HCP content using SWATH-MS. The impacts of three key method parameters were evaluated—the spectral library construction, the DIA data processing software, and the mass spectrometer instrument. First, method performance using a spectral library constructed from project-specific DDA data was compared to a spectral library constructed *in silico* from a provided sequence database.

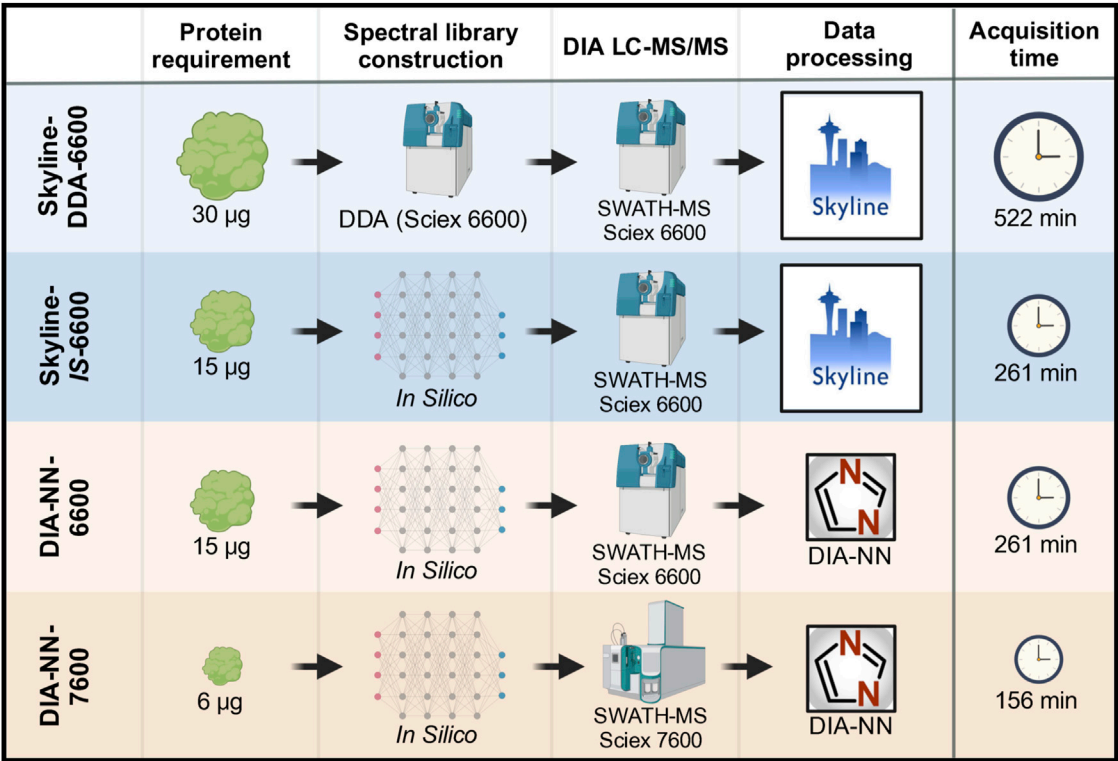


FIGURE 1
DIA LC-MS/MS data acquisition and analysis overview showing the four workflows tested. Protein requirement and acquisition time are shown for triplicate data acquisition of one sample.

Next, peptide and protein coverage and CV of triplicate data acquisitions were compared for data processing in two open-source software suites–Skyline and DIA-NN, the latter of which leverages advanced deep neural network (DNN) machine learning models for protein identification and quantitation. Lastly, the best performing SWATH-MS method run on the Sciex TripleTOF 6600 as determined by number of unique protein identifications across the sample set was transferred to the Sciex ZenoTOF 7600 instrument to examine instrument-specific impacts on HCP identification and quantitation. Overall, these evaluations resulted in a SWATH-MS method for rAAV HCP analysis with increased protein coverage, and reduced sample requirement and instrument runtime.

2 Results

A base-case DIA LC-MS/MS method that was previously established for analysis of residual HCPs in mAb downstream processing uses a spectral library constructed from project-specific DDA data, SWATH-MS data acquisition with a Sciex TripleTOF 6600 instrument, and data processing in Skyline (Skyline-DDA-6600) (Herman et al., 2023a; Herman et al., 2023b; Leibiger et al., 2024a). Protein identification and quantitation outputs for this base-case method were compared to a method using an *in silico* spectral library, a method using DIA-NN data processing software, and a method with data acquisition on a Sciex ZenoTOF 7600 mass spectrometer instrument (Figure 1).

2.1 Project-specific DDA vs. *in silico* spectral library (Sciex TripleTOF 6600)

DIA outputs depend heavily on the spectral library composition (Fröhlich et al., 2022; Lou et al., 2023; Zhang et al., 2023). For prior analyses, DDA data based-spectral libraries were built in Skyline with database search results from ProteinPilot software v5.1 (MacLean et al., 2010). Triplicate DDA data was acquired for each sample to feed into spectral library construction which contributed a substantial amount of sample and instrument runtime to the overall DIA workflow. Library-free mode in DIA-NN allows for *in silico* generation of spectral libraries for DIA applications, eliminating the need for DDA data collection and database search (Demichev et al., 2020). SWATH-MS data collected on the Sciex TripleTOF 6600 was processed in Skyline using both a project-specific DDA-derived library and an *in silico* spectral library built using DIA-NN to test library performance.

Use of the *in silico* spectral library in Skyline (Skyline-IS-6600) resulted in a slight increase in protein identifications across the sample set, which was measured to be statistically significant by a non-parametric Wilcoxon Signed-Rank test ($p < 0.05$) (Figure 2; Supplementary Figure S1). One out of the ten samples analyzed (AAV2 – B2) showed fewer HCP identifications for the *in silico* library (1,246) compared to the DDA library (1,286) causing the conserved protein identifications for the *in silico* library to be slightly below that of the DDA library (Supplementary Figure S1). HCP identifications for all other samples and total unique HCP

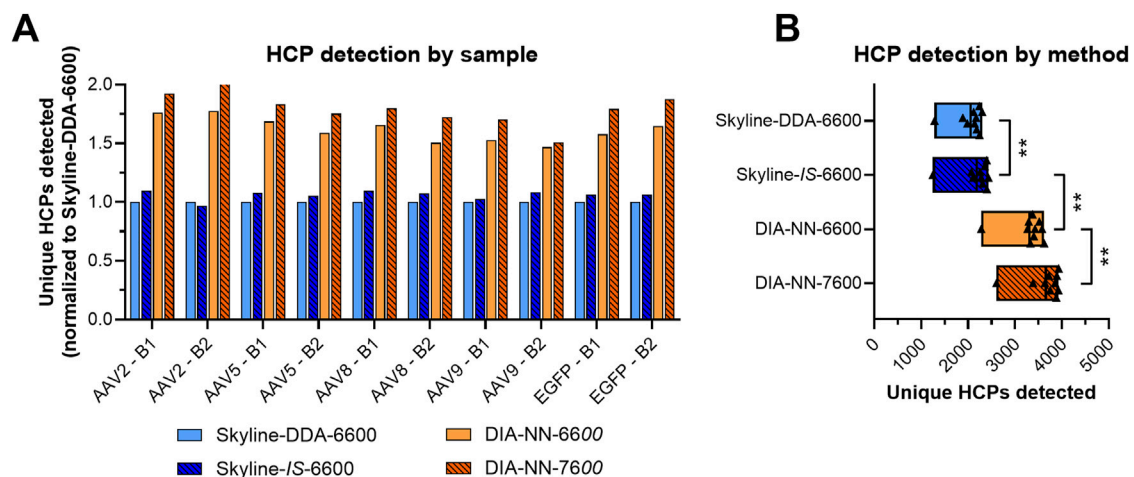


FIGURE 2

Total HCP identifications for the four different DIA LC-MS/MS workflows (A). Protein identification in triplicate injections was required for inclusion in each method group. Protein identifications averaged across the sample set ($N = 10$) for each method (B). Non-parametric Wilcoxon Signed-Rank statistical tests were applied to determine if the means of paired groups were significantly different. ** indicates statistical significance at $p < 0.01$.

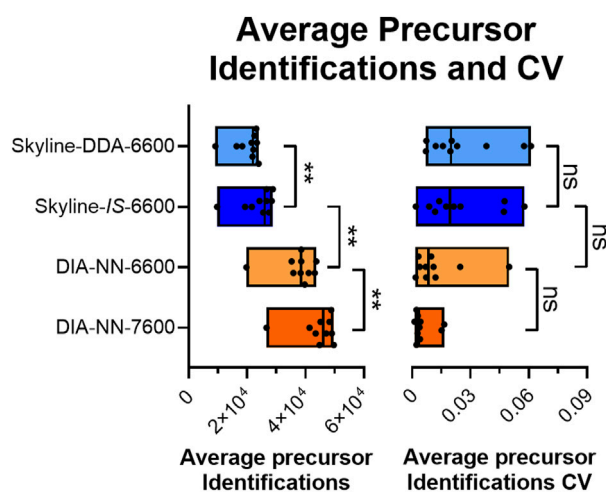


FIGURE 3

Average precursor identifications and precursor identification CV across triplicate injections for all four SWATH-MS methods across the sample set ($N = 10$). ** indicates statistical significance at $p < 0.01$. ns indicates that the differences were not statistically significant at a 95% confidence interval.

identifications were higher with the *in silico* spectral library (Figure 2; Supplementary Figure S1). Total precursor identifications trended similarly, with a slight increase for the Skyline-/S-6600 method compared to the base case Skyline-DDA-6600 method (Figure 3). There were 961 HCPs commonly identified and quantified in all samples for both DDA and *in silico* spectral library data processing in Skyline ($N = 10$). Normalized protein quantitation (ng of an individual HCP relative to the total HCP amount in μg in the sample) of these species showed excellent agreement between the two methods, with linear regression r^2 between 0.974 and 0.997 for all samples, and slopes ranging from 0.984–1.192 (Figure 4A; Supplementary Figure S2). There was no

statistically significant change to the protein quantitation CV across triplicate injections for the two spectral libraries, with median CV remaining below 10% in all cases (Figure 5). Peptide identification CV also remained consistent for both spectral libraries (Figure 3).

2.2 Skyline vs. DIA-NN data processing (*in silico* spectral library, Sciex TripleTOF 6600) – HCP identifications

The DIA-NN software suite houses advanced data processing features which combine peptide- and spectrum-centric approaches with DNNs to discriminate between target and decoy precursors (Demichev et al., 2020). Multiple studies have evaluated the performance of DIA-NN using reference datasets and prepared mixtures of protein standards, broadly demonstrating the superior performance of this software particularly for coverage of low abundance species (Gotti et al., 2021; Fröhlich et al., 2022; Lou et al., 2023).

Data processing with DIA-NN significantly increased protein coverage compared to Skyline, with average protein identifications across the sample set increasing 51.3%, from 2,188 in Skyline (Skyline-/S-6600) to 3,310 in DIA-NN (DIA-NN-6600) (Figure 2). Precursor identifications were also significantly higher for the DIA-NN-6600 workflow (Figure 3). There was no observed impact on median HCP quantitation CV for triplicate data acquisition when processing data in DIA-NN compared to Skyline, with median CVs remaining below 10% for all samples (Figure 5). Precursor identification CV also remained consistent for both programs (Figure 3).

2.3 DIA-NN data processing with the Sciex ZenoTOF 7600

The DIA-NN-6600 method produced the greatest number of unique protein identifications for runs performed with the Sciex

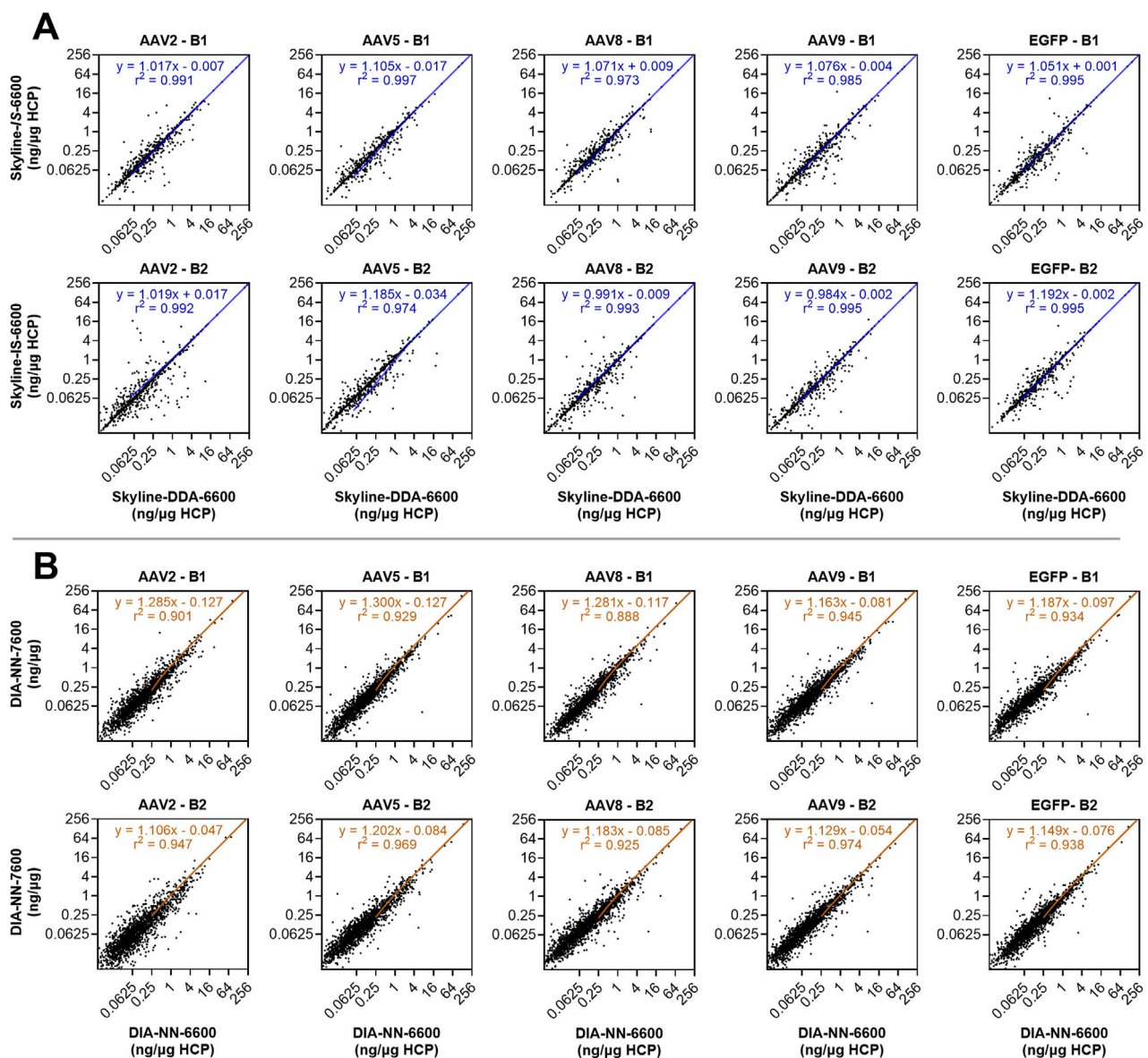


FIGURE 4

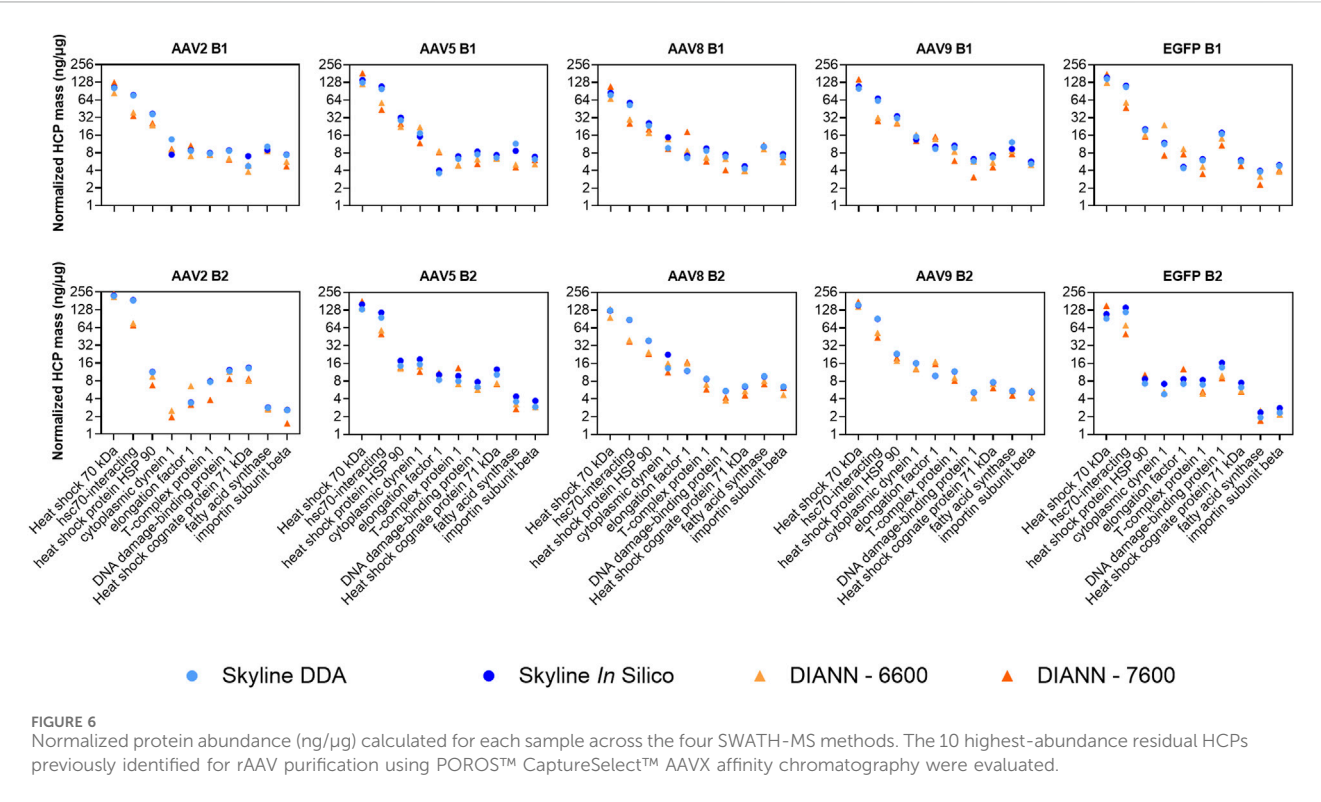
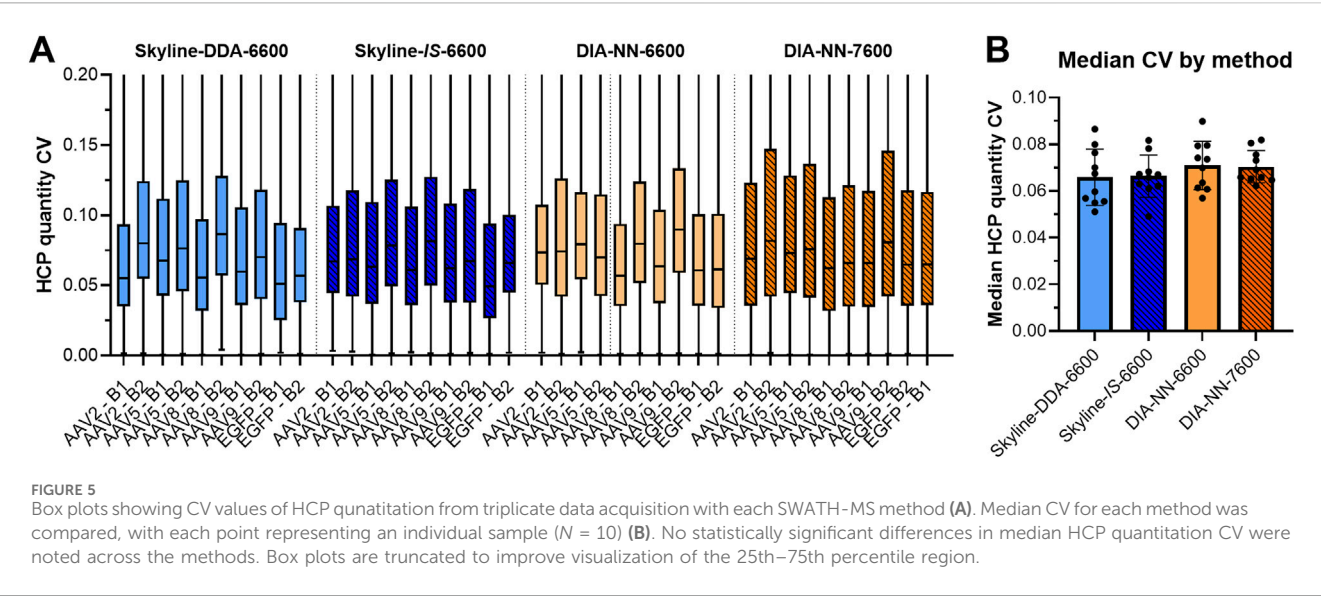
Individual HCP quantitation (ng/μg) for the 961 HCPs commonly quantified with both the DDA and *in silico* spectral libraries in Skyline across all samples ($N = 10$) (A). Individual HCP quantitation (ng/μg) for the 1,758 HCPs commonly quantified with both Sciex instruments using DIA-NN data processing and the *in silico* spectral library across all samples ($N = 10$) (B). Linear regression trendlines of the non-transformed data are shown on each subplot, with equations of fit and goodness of fit r^2 .

TripleTOF 6600 system. This method was transferred to the Sciex ZenoTOF 7600 instrument (DIA-NN-7600 method), which yielded a further increase in both protein and precursor identifications (Wilcoxon Signed-rank test, $p < 0.05$). Median HCP quantitation CV was once again unimpacted and remained below 10% for all samples. There were 1,758 HCPs commonly identified and quantified across triplicate injections for all samples using both the DIA-NN-6600 and DIA-NN-7600 methods ($N = 10$). The normalized protein abundances (ng/μg) of these species showed r^2 of 0.939 for the combined datasets, with a range of 0.888–0.974 across the sample set (Figure 4B; Supplementary Figure S3). However, linear regression slopes were all >1.1 (range 1.106–1.300) indicating that samples re-analyzed with the Sciex ZenoTOF 7600 instrument yielded higher

normalized protein abundance outputs despite showing good quantitation linearity with the DIA-NN-6600 data. This variability can be attributed to the separate sample handling required for analysis with the Sciex ZenoTOF 7600 mass spectrometer which included re-measurement of protein concentration, re-digestion, and separate LC-MS/MS data acquisition.

2.4 HCP quantitative consistency between DIA LC-MS/MS methods

Direct quantitative comparisons of individual HCPs between Skyline and DIA-NN were not performed across the full datasets due



to differences in the algorithms governing peak detection, peak integration, and false discovery rate (FDR) calculation. Protein identification outputs for the two software suites were often split across two or more different isoforms that share considerable sequence homology making direct, comprehensive, comparisons challenging. To address this challenge, quantitative outputs for a select set of residual HCPs were manually inspected across the two data processing software. A set of highly abundant conserved HCPs were previously detected within all intermediately purified rAAV2, -5, -8, and -9 preparations analyzed from our production and purification scheme (Leibiger et al., 2024a). The 10 most abundant HCPs across this highly conserved group were compared for the four analysis workflows to determine normalized quantitation (ng/μg) trending of these species (Figure 6). Consistent trends in quantitation were observed across all four methods. Quantitation outputs for these 10 HCPs across the two software suites using the *in silico* spectral library were also compared comprehensively for all samples and showed good linearity between software with r^2 of 0.892 and 0.793 for DIA-NN-6600 and DIA-NN-7600 compared to Skyline-IS-6600, respectively

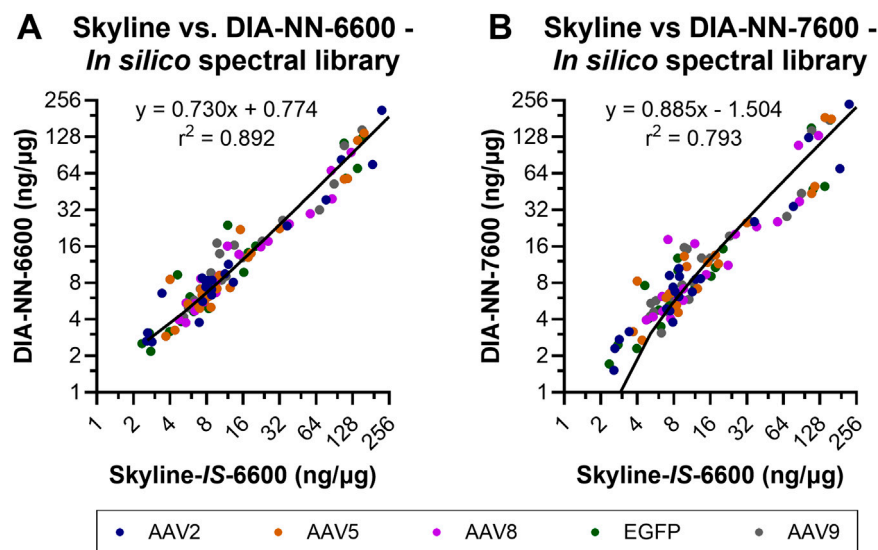


FIGURE 7

Normalized protein abundance (ng/ μ g) comparisons between Skyline and DIA-NN for the 10 highest-abundance residual HCPs previously identified (Leibiger et al., 2024a). Skyline-IS-6600 was compared to DIA-NN-6600 (A) and DIA-NN-7600 (B). Data for all samples is plotted together, with color coding designations of rAAV serotypes or EGFP control material. Linear regression was performed across the full datasets. Linear regression trendlines of the non-transformed data are shown on each subplot, with equations of fit and goodness of fit r^2 .

(Figure 7). Slopes of best fit were 0.730 and 0.885 for DIA-NN-6600 and DIA-NN-7600 compared to Skyline-IS-6600, respectively, indicating that DIA-NN gives relatively lower normalized HCP abundance compared to Skyline for these HCPs of interest. The lower normalized HCP abundance in DIA-NN can be attributed to dynamic background noise correction, a feature not included in Skyline. Correction factor constants were determined to directly compare protein quantitation based on the mean variability between methods, calculated at 1.250 and 1.365 for DIA-NN-6600 and DIA-NN-7600 vs. Skyline-IS-6600, respectively. After applying these correction constants, the median relative difference in protein quantitation between methods was 14.47% for DIA-NN-6600 and 22.57% for DIA-NN-7600 compared to Skyline-IS-6600 (26.38% and 32.89%, respectively, without application of correction constants). Despite quantitation agreement for most HCPs, some method-specific trends were also observed. For instance, the normalized abundance (ng/ μ g) of hsc-70 interacting protein was measured to be approximately 2-fold higher in Skyline compared to DIA-NN (Figure 6).

3 Discussion

Process-related impurity retention mechanisms and product impacts have been rigorously explored for mAbs (Herman et al., 2023a; Herman et al., 2023b; Levy et al., 2014; Hogwood et al., 2014), but there have been relatively few studies examining impurity retention for viral vector products (Bracewell et al., 2021; Rumachik et al., 2020; Soni et al., 2024). These studies are challenging to perform for rAAV vectors due to the complex nature and limited availability of samples. The intracellular expression of most rAAV serotypes requires cell lysis at harvest

for optimal product recovery which introduces a complex background of cellular impurities into the product stream. Additionally, compared to a high titer mAb bioprocess which can produce 8 g/L, a typical rAAV upstream process may only generate ~6 mg of packaged vector per liter of cell culture ($\sim 1 \times 10^{12}$ VG/mL culture) making it difficult to meet LC-MS/MS protein mass injection targets with small scale studies that must also support other product-specific analytical workflows (Leibiger et al., 2024a; Szkodny and Lee, 2022). Scaling up current rAAV production systems to better support product and process analytics presents additional complications. rAAV cell culture processes commonly use transient three plasmid transfection of HEK293 cells which is limited by poor mixing of DNA complexes at larger scales and is associated with high-cost consumables such as transfection reagents and plasmid DNA preparations (Lyle et al., 2024). To address these challenges and allow for more rigorous characterizations of residual HCP profiles across rAAV purification, we report a SWATH-MS workflow leveraging *in silico* spectral library construction and DNN data processing with reduced sample requirement and increased sensitivity for low abundance HCPs.

3.1 *In silico* spectral library reduces sample requirements and improves protein coverage

Project-specific DDA-based spectral libraries contain high-confidence peptide spectra identifications derived from peptide sequences, precursor and fragment ion data, and retention times. In contrast, *in silico* spectral libraries are composed of computationally-derived peptide fragmentation patterns predicted from a provided protein sequence database (Tiway et al., 2019).

This approach is based on deep learning algorithms and has been used to build hybrid targeted libraries with deeper protein coverage for specific protein families (Lou et al., 2020; Lou et al., 2021), and comprehensive proteome-wide libraries (Demichev et al., 2020; Gessulat et al., 2019). Using an *in silico* spectral library for SWATH-MS of rAAV samples addresses two key challenges in data generation—low sample availability and high sample complexity. Replacing the project-specific DDA spectral library with an *in silico* library constructed from the NCBI:Hu_RefSeqGRCh38 database eliminated the need for DDA data acquisition, reducing sample requirement by 50% while increasing sample throughput and lowering analysis costs. Data processed using the *in silico* library also yielded an increase in identified precursors and proteins across the sample set without impacting individual HCP quantitation trends or CV (Figures 2, 3, 4A). The increase in protein identifications with the *in silico* spectral library is driven by fragmentation pattern prediction of low intensity precursor ions captured in the computationally derived library which may go undetected in DDA data collection. As observed here and previously shown by Rice et al., peptide identifications captured in the DDA but not the *in silico* spectral libraries can occur which may be influenced by charge states of longer ionized peptides with more acidic or basic residues (Rice and Belani, 2022). We observed an increase in mean peptide length for species uniquely detected with the DDA spectral library which aligns with previously reported findings (Rice and Belani, 2022). Using the *in silico* spectral library allows for smaller-scale process development studies or testing of multiple purification conditions from a given material lot which can save time, reduce development costs, and increase sample throughput for rAAV HCP characterization without compromising data quality.

3.2 Deep neural network data processing in DIA-NN substantially boosts protein coverage

DIA-NN is a software suite specifically developed for processing DIA proteomics data by leveraging DNNs for differentiation of target and decoy precursor ion patterns using both peptide-centric and spectrum-centric approaches (Demichev et al., 2020). DIA-NN data processing substantially improved HCP identifications for the rAAV samples analyzed, with total detected HCPs increasing 38.8% (2,830 to 3,928, $N = 8$) and universally detected HCPs increasing 88.6% (1,095 to 2,065, $N = 8$) compared to data processed using Skyline (Supplementary Figure S1). The observed variation in data outputs between the two software can be attributed to differences in their computational frameworks. In Skyline, peak detection and integration is primarily driven by direct extraction of chromatographic elements (MacLean et al., 2010). This process includes extraction of m/z intensity and retention time, resampling (i.e., linear interpolation of raw chromatograms), peak detection by local minima, local maxima, and inflection points, peptide modification peak grouping, peptide identification, and peak integration by area under the curve (MacLean et al., 2010; Pino et al., 2020). Peptide identification is based on a static model that uses coefficient-based weighting of peak group features including log intensity, co-elution count,

identified count, library intensity correlation, shape score, weighted co-elution, and delta retention time from predicted (Pino et al., 2020). DIA-NN uses a DNN-based machine learning approach for peak detection and integration instead of a relying on static coefficient-based weighting of peak group features (Demichev et al., 2020). Peak group features used for peptide identification are dynamically weighted in DIA-NN based on training data and iterative dataset-specific refinement (Demichev et al., 2020). DIA-NN also employs forward propagating DNNs to discriminate between target and decoy precursors, selecting the best representative precursor-fragment pair for each spectrum while predicting and correcting for interference of co-fragmenting precursors. Because of these features, DIA-NN is particularly well suited for processing proteomics data from complex samples with many co-eluting peptides, making it a preferable software suite for DIA data processing of rAAV samples.

3.3 Sciex ZenoTOF 7600 improves HCP detection sensitivity with reduced sample load

Re-digestion and re-collection of SWATH-MS data using the Sciex ZenoTOF 7600 system yielded an additional 10.6% increase in average HCP identifications while reducing sample load by 60% (2 μ g injection compared to 5 μ g injection with the Sciex TripleTOF 6600 system) (Figure 2). The reduced sample loading with the Sciex ZenoTOF 7600 instrument further addresses sample availability constraints for residual HCP analysis of rAAV samples. This finding is consistent with previously reported SWATH-MS comparisons between the instruments which showed increased protein coverage for the Sciex ZenoTOF 7600 compared to the Sciex TripleTOF 6600 when analyzing a K562 human cell line digest standard at varying sample loading amounts (Wang et al., 2022). Differences in instrument performance result from hardware and software improvements for the newer Sciex ZenoTOF 7600 model, as specified by the manufacturer. The Sciex ZenoTOF 7600 has a faster scan rate leading to enhanced ion picking and a more sensitive ion detector with greater dynamic range compared to the Sciex TripleTOF 6600. The Sciex ZenoTOF 7600's Zeno trap device can be used to further increase ion utilization and boost signal-to-noise ratio by capturing and releasing ions to the TOF analyzer in synchronized pulses (Wang et al., 2022). The use of this device has been reported to increase precursor and protein identifications potentially allowing for further method sensitivity and sample requirement improvements (Wang et al., 2022).

3.4 Conclusion

The base-case Skyline-DDA-6600 workflow requires triplicate data acquisition for DDA library construction with a 5 μ g protein injection target, and triplicate DIA data acquisition with another 5 μ g protein injection using a Sciex TripleTOF 6600 instrument. Transitioning to *in silico* spectral library generation, DIA-NN data processing, and data acquisition using a Sciex ZenoTOF 7600 instrument provided notable advantages in sample

requirement, protein coverage, and instrument runtime. The DIA-NN-7600 method gave a 77.0% average increase in HCP identifications across the full sample set (range 50.7%–102.6%, $N = 10$), an 80% reduction in sample requirement (30 μg –6 μg), and a 70% reduction in instrument runtime (87 min DDA and SWATH-MS to 52 min SWATH-MS only, per injection) compared to the Skyline-DDA-6600 method. These method improvements enhance the ability to study residual HCPs in rAAV bioprocessing, ultimately allowing for more comprehensive process development and vector characterization.

4 Materials and methods

4.1 rAAV production

rAAV vectors were produced by transient three plasmid transfection of suspension HEK293 cells. Plasmid DNA was complexed with FectoVIR-AAV[®] chemical transfection reagent, and DNA complexes were added to HEK293 cells at a density of 2.5×10^6 cells/mL dropwise while gently swirling flasks. The cell line present in this study was obtained from Thermo Fisher Scientific. EXPI293[®] (Thermo Fisher Scientific) cells were exchanged into fresh EXPI293[®] Expression Media (Thermo Fisher Scientific) immediately prior to transfection. Detailed transfection parameters are summarized in [Supplementary Table S1](#). Plasmid information and acknowledgements are listed in [Supplementary Table S2](#). For each rAAV serotype (rAAV2, -5, -8, and -9) 2×1 L shake flasks were transfected for each biological replicate, giving a total of 4×1 L shake flasks per serotype. After harvest, the 2×1 L flasks of each biological replicate were pooled together to give a total of eight rAAV-containing harvest lots. An additional 1 L flask for each repeated transfection lot was generated with only the pEGFP plasmid delivered to cells to produce ‘control’ EGFP material containing no AAV capsids. For each biological replicate production, a total of 9×1 L shake flasks were transfected (two for each rAAV serotype, and one for control EGFP lot). Biological replicate transfections were performed with independent plasmid lots, cell vial thaws, and transfection reagent lots to account for process variability. After pooling replicate flasks, 10 material lots in total were produced, $N = 2$ for rAAV2, -5, -8, -9, and EGFP.

4.2 rAAV harvest and quantitative real-time PCR

Culture harvest treatment and quantitative real-time PCR (real-time qPCR) were performed as described previously ([Leibiger et al., 2024a; Leibiger et al., 2024b](#)). Cultures were harvested by centrifuging for 10 min at $1,000 \times g$ using a 5920 R centrifuge (Eppendorf). Lysates were generated by three freeze/thaw cycles of resuspended cell pellets in Mammalian Lysis Buffer (50 mM Tris-HCl, 150 mM NaCl, 2 mM MgCl_2 , pH 8.5) followed by 25 U/mL Benzonase[®] Nuclease (Sigma Aldrich) treatment at 37°C for 60 min. Cell debris was removed by centrifuging harvest material at $3,428 \times g$ followed by filtration with 0.22 μm bottle top filter units (Fisher Scientific). Prior to real-time qPCR, harvest lots were treated with DNase (New England Biolabs) by adding 2.5 μL of sample to a

mixture of 2.5 μL DNase I, 2.5 μL DNase Buffer, and 17.5 μL molecular biology water and incubating at 37°C for 60 min. Capsid digestion was performed by adding 2.5 μL of 20 mg/mL Proteinase K (Thermo Fisher Scientific) and incubating at 56°C for 90 min. TaqMan[™] Fast Advanced Master Mix (Thermo Fisher Scientific) was used for real-time qPCR along with a primer (900 nM) and probe (250 nM) set (IDT) targeting a region within the EGFP transgene as described previously ([Leibiger et al., 2024a; Leibiger et al., 2024b](#)).

4.3 rAAV purification

Duplicate flasks for each rAAV serotype were pooled to give approximately 50 mL of clarified lysate per rAAV-containing lot. Purification was performed using POROS[™] CaptureSelect[™] AAVX affinity resin and an AKTA Pure[™] (Cytiva) fast protein liquid chromatography system. To account for serotype-dependent vector secretion into the culture supernatant, only purified rAAV isolated from culture lysates was analyzed in this work. For each affinity purification, 10 mL of clarified lysate was 1:1 diluted with Equilibration Buffer (20 mM Tris-HCl, 0.1 M NaCl, pH 7.5) and loaded onto a 1 mL freshly packed column using TRICORN 5/50 (Cytiva) column hardware. Column equilibration was performed with 10 column volumes (ColV) of Equilibration Buffer, followed by loading, washing with 12 ColV of Equilibration Buffer, and elution with 15 ColV of Elution Buffer (0.1 M Glycine-HCl, pH 2.6). Elution pools were collected in 50 mL conical tubes with 1.5 mL (10 v/v%) Neutralization Buffer (1 M Tris-HCl, pH 8.7). Loading, washing, and elution were performed at 2-min residence time (0.5 mL/min).

4.4 LC-MS/MS sample preparation

Coomassie Plus (Bradford) Assay Reagent (Thermo Fisher Scientific) was used to measure protein concentration of affinity chromatography elution pools as described previously ([Leibiger et al., 2024a](#)). Protein digestion and desalting were performed as described previously ([Herman et al., 2023a; Hamaker et al., 2022; Oh et al., 2023](#)). Sample volumes of 100 μL containing 50 μg of measured protein were reduced with 2.5 μL 100 mM TCEP (Thermo Fisher Scientific) and denatured for 1 h at 60°C, followed by alkylation with 5 μL 150 mM iodoacetamide (Sigma Aldrich) in the dark at room temperature for 30 min. Trypsin (Promega) was then added at an enzyme to substrate mass ratio of 1:50 and digestion proceeded at 37°C for 16 h. Trypsin has been shown to have a limited ability to digest AAV viral capsid proteins, and is therefore recommended for protein digestion of rAAV samples to increase relative HCP signal intensity compared to viral capsid proteins ([Smith et al., 2023; Guapo et al., 2022](#)). Digestion was stopped by adding 4 μL 20% formic acid (Thermo Fisher Scientific). For samples with a final protein amount of less than 50 μg , digestion was performed with reagent amounts scaled accordingly. Samples were desalted with Omix C18 tips (Agilent), dried with a Speed Vac (Thermo Fisher Scientific) and redissolved in 45 μL 2% acetonitrile with 0.1% formic acid. Samples were spiked with pre-digested ADH

(Waters) to a concentration of 5 fmol/ μ L and retention time calibrants (iRT, Biognosys, Schlieren, Switzerland). Samples equivalent to 5 μ g and 2 μ g of digested proteins were analyzed in technical triplicate for LC-MS/MS on the Sciex TripleTOF 6600 and Sciex ZenoTOF 7600, respectively.

4.5 LC-MS/MS data acquisition on Sciex TripleTOF 6600

LC-MS/MS data acquisition was done as described previously with an Eksigent Nano 425 LC (Sciex) coupled to a Sciex TripleTOF 6600 mass spectrometer (Sciex) and a dual spray source (Herman et al., 2023a; Hamaker et al., 2022; Oh et al., 2023). Samples were injected into a ChromXP C18CL Sciex column (3 mm, 120 $\text{\AA}$, 150 mm $\times$ 0.3 mm). Separation was performed using water with 0.1% formic acid (mobile phase A) and acetonitrile with 0.1% formic acid (mobile phase B) at a flow rate of 5 μ L/min. A separation gradient of 3%–25% mobile phase B over 68 min, 25%–35% mobile phase B over 5 min, and then 35%–80% mobile phase B over 2 min was used followed by column regeneration and re-equilibration. The column temperature was 30°C. SWATH-MS was performed with a survey scan in the mass range of 400–1,200 m/z followed by 64 variable-size isolation windows for MS/MS with accumulation time of 35 ms. Ion source gas one and two were set at 30 and 35 psi, respectively. DDA was performed with a survey scan over a mass range of 400–1,250 m/z and the top 30 precursor ions were selected for fragmentation and MS/MS detection over a mass range of 100–1,500 m/z.

4.6 LC-MS/MS data acquisition on sciex ZenoTOF 7600

LC-MS/MS analysis was done on an Acquity UPLC M-class system (Waters) coupled to a Sciex ZenoTOF 7600 mass spectrometer (Sciex) with an OptiFlow source. Samples were injected into a C18 microtrap column (Phenomenex, 10 $\times$ 0.3 mm) and washed with 0.1% formic acid for 2 min at 8 μ L/min, then eluted to a C18 column (Phenomenex, Kinetex 2.6 mm XB C18, 100 $\text{\AA}$, 150 $\times$ 0.3 mm). Mobile phase A and B consisted of 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively, and a flow rate of 5 μ L/min was used. Elution was performed with a program of 3% B for 1 min, 3%–32% B over 44 min, 32%–80% B over 1 min followed by column regeneration and re-equilibration. The column temperature was 30°C. SWATH-MS was performed with a survey scan in the mass range of 400–1,200 m/z followed by 32 variable-size isolation windows for MS/MS with accumulation time of 25 ms. Ion source gas one and two were set at 30 and 60 psi, respectively.

4.7 Spectral library construction

The DDA data based-spectral library was built previously (Leibiger et al., 2024a). Project specific triplicate DDA datasets

for all samples analyzed were pooled for a combined database search using Paragon Algorithm in Protein Pilot software (v5.1) (MacLean et al., 2010). A local copy of NCBI:Hu_RefSeqGRCh38 database was used, supplemented with ADH, retention time calibrants, and common contaminants. The resulting group file from Protein Pilot search was imported to Skyline (v20.2.0.343, MacCoss Lab, University of Washington) for building a consolidated spectral library using BiblioSpec with peptides identified at 95% confidence score or higher. In total, the spectral library consisted of 84,563 peptide precursors mapped to 3,432 proteins. An *in silico* library was built in DIA-NN (v1.8.1) with the local copy of NCBI:Hu_RefSeqGRCh38 database supplemented with ADH, retention time calibrants, and common contaminants (Demichiev et al., 2020).

4.8 LC-MS/MS data processing

For the Skyline-DDA-6600 workflow, SWATH-MS data acquired on Sciex TripleTOF 6600 was processed in Skyline (v22.2.0.351) as described previously (Herman et al., 2023a; Hamaker et al., 2022; Oh et al., 2023). Triplicate SWATH-MS datasets were processed together using the Skyline command line interface with the DDA data based spectral library and the following settings: max 1 missed cleavage allowed, variable carbamidomethyl modification of cysteine, the six most intense b- or y-ions at charge 1+ or 2+, from ion 3 to last ion, ion match tolerance 0.05 m/z, five to six fragments picked from library, resolving power of 36,000, retention time tolerance of 4 min. Peaks were selected and scored with mProphet algorithm based on a target decoy approach (Reiter et al., 2011). Peaks detected at a q value above 0.01 were removed from further analysis. Peak areas were exported to MSstats (Choi et al., 2014), in which peak areas were log₂-transformed, normalized with global standard (ADH) normalization, and the top three features of each protein were summarized using Turkey median polish to obtain protein areas. ADH peak area was calculated based on the following peptides–ANELLINVK, SISIVGSYVGNR, and VVGLSTLPEIYK. Protein areas were normalized to ADH peak area and protein amounts in ng were estimated based on the assumption that all proteins of a given amount of mol generate an equal response.

For the Skyline-IS-6600 workflow, the SWATH-MS datasets acquired on TripleTOF 6600 were processed in Skyline the same as the Skyline-6600-DDA library workflow except the *in silico* library was used instead of the DDA-derived spectral library. Peak areas from Skyline were exported to MSstats and integrated for protein areas to calculate protein amounts in ng.

For the DIA-NN-6600 and DIA-NN-7600 workflows, SWATH-MS data acquired on Sciex TripleTOF 6600 and Sciex ZenoTOF 7600, respectively, were processed with DIA-NN (v. 1.8.1) using the *in silico* library with default settings, except for protein interference which was set at isoform IDs. False discovery rate at protein and peptide levels were set to 1%, and match-between-runs was selected. The main report from DIA-NN was imported to MSstats and integrated for protein peak areas to calculate protein amounts in ng.

4.9 LC-MS/MS data normalization

HCP amounts in ng were calculated based on ADH response factors and protein molecular weights. Protein quantities obtained for all four methods were normalized for each sample as described previously (Leibiger et al., 2024a). Non-HCP species including AAV capsid proteins, AAV replication proteins, assembly activating protein, EGFP, ADH, and modified trypsin were removed from the analyses. The amounts of remaining HCP species were normalized by dividing the triplicate-averaged individual HCP amounts (ng) by the total HCP amount (ng) of each sample. Fractional HCP content outputs were then multiplied by 1,000 to give HCP mass in ng of individual HCP per μg total HCP. These normalized outputs (ng/ μg) were used for all subsequent analyses.

4.10 Statistical analysis

Statistical analyses were performed in Graphpad (v. 10.2.3). Due to the small sample size, all statistical comparisons between DIA methods were performed using non-parametric Wilcoxon Signed-Rank tests with p values <0.05 considered significant.

Data availability statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the jPOST partner repository with the dataset identifier PXD056294 (TripleTOF 6600 data) and PXD060891 (ZenoTOF 7600 data). All other data generated and discussed in this work is available upon reasonable request.

Author contributions

TL: Conceptualization, Formal Analysis, Investigation, Methodology, Writing – original draft. LM: Conceptualization,

Formal Analysis, Investigation, Methodology, Writing – original draft. KL: Funding acquisition, Supervision, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was funded in part by NIST #70NANB21H085; #70NANB21H086; and #70NANB20H037.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2025.1579098/full#supplementary-material>

References

- Bracewell, D. G., Smith, V., Delahaye, M., and Smales, C. M. (2021). Analytics of host cell proteins (HCPs): lessons from biopharmaceutical mAb analysis for Gene therapy products. *Curr. Opin. Biotechnol.* 71, 98–104. doi:10.1016/j.copbio.2021.06.026
- Carvalho, S., Profit, L., Krishnan, S., Gomes, R., Alexandre, B., Clavier, S., et al. (2024). SWATH-MS as a strategy for CHO host cell protein identification and quantification supporting the characterization of mAb purification platforms. *J. Biotechnol.* 384, 1–11. Article. doi:10.1016/j.jbiotec.2024.02.001
- Chiu, J., Valente, K. N., Levy, N. E., Min, L., Lenhoff, A. M., and Lee, K. H. (2017). Knockout of a difficult-to-remove CHO host cell protein, lipoprotein lipase, for improved polysorbate stability in monoclonal antibody formulations. *Biotechnol. Bioeng.* 114 (5), 1006–1015. doi:10.1002/bit.26237
- Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., et al. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30 (17), 2524–2526. doi:10.1093/bioinformatics/btu305
- Clement, N., and Grieger, J. C. (2016). Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol. Ther. Methods Clin. Dev.* 3, 16002. doi:10.1038/mtm.2016.2
- Clinicaltrials.gov (2025). *Clinicaltrials.gov*. (Accessed April 2, 2025).
- Demichev, V., Messner, C., Vernardis, S., Lilley, K., and Ralser, M. (2020). DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* 17, 41–44. doi:10.1038/s41592-019-0638-x
- Fröhlich, K., Brombacher, E., Fahrner, M., Vogele, D., Kook, L., Pinter, N., et al. (2022). Benchmarking of analysis strategies for data-independent acquisition proteomics using a large-scale dataset comprising inter-patient heterogeneity. *Nat. Commun.* 13, 2622. doi:10.1038/s41467-022-30094-0
- Gessulat, S., Schmidt, T., Zolg, D., Samaras, P., Schnatbaum, K., Zerweck, J., et al. (2019). Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* 16 (6), 509–518. doi:10.1038/s41592-019-0426-7
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11 (6), O111.016717. doi:10.1074/mcp.O111.016717
- Goswami, R., Subramanian, G., Silayeva, L., Newkirk, I., Doctor, D., Chawla, K., et al. (2019). Gene therapy leaves a vicious cycle. *Front. Oncol.* 9, 297. doi:10.3389/fonc.2019.00297

- Gotti, C., Roux-Dalvai, F., Joly-Beauparlant, C., Mangnier, L., Leclercq, M., and Droit, A. (2021). Extensive and accurate benchmarking of DIA acquisition methods and software tools using a complex proteomic standard. *J. Proteome Res.* 20 (10), 4801–4814. doi:10.1021/acs.jproteome.1c00490
- Guapo, F., Strasser, L., Millán-Martín, S., Anderson, I., and Bones, J. (2022). Fast and efficient digestion of adeno associated virus (AAV) capsid proteins for liquid chromatography mass spectrometry (LC-MS) based peptide mapping and post translational modification analysis (PTMs). *J. Pharm. Biomed. Analysis* 207, 114427. doi:10.1016/j.jpba.2021.114427
- Hamaker, N. K., Min, L., and Lee, K. H. (2022). Comprehensive assessment of host cell protein expression after extended culture and bioreactor production of CHO cell lines. *Biotechnol. Bioeng.* 119 (8), 2221–2238. doi:10.1002/bit.28128
- Herman, C. E., Min, L., Choe, L. H., Maurer, R. W., Xu, X. K., Ghose, S., et al. (2023a). Analytical characterization of host-cell-protein-rich aggregates in monoclonal antibody solutions. *Biotechnol. Prog.* 39 (4), e3343. doi:10.1002/btpr.3343
- Herman, C. E., Min, L., Choe, L. H., Maurer, R. W., Xu, X. K., Ghose, S., et al. (2023b). Behavior of host-cell-protein-rich aggregates in antibody capture and polishing chromatography. *J. Chromatogr. A* 1702, 464081. doi:10.1016/j.chroma.2023.464081
- Hogwood, C., Bracewell, D., and Smales, C. (2014). Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Curr. Opin. Biotechnol.* 30, 153–160. doi:10.1016/j.copbio.2014.06.017
- Ito, T., Lutz, H., Tan, L., Wang, B., Tan, J., Patel, M., et al. (2024). Host cell proteins in monoclonal antibody processing: control, detection, and removal. *Biotechnol. Prog.* 40 (4), e3448. doi:10.1002/btpr.3448
- Jones, M., Palackal, N., Wang, F. Q., Gaza-Bulsecos, G., Hurkmans, K., Zhao, Y. W., et al. (2021). “High-risk” host cell proteins (HCPs): a multi-company collaborative view. *Biotechnol. Bioeng.* 118 (8), 2870–2885. doi:10.1002/bit.27808
- Kontogiannis, T., Braybrook, J., McElroy, C., Foy, C., Whale, A., Quaglia, M., et al. (2024). Characterization of AAV vectors: a review of analytical techniques and critical quality attributes. *Mol. Ther. Methods Clin. Dev.* 32 (3), 101309. doi:10.1016/j.omtm.2024.101309
- Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 4, 222. doi:10.1038/msb.2008.61
- Leibiger, T., Min, L., and Lee, K. (2024a). Quantitative proteomic analysis of residual host cell protein retention across adeno-associated virus affinity chromatography. *Mol. Ther. Methods Clin. Dev.* 32 (4), 101383. doi:10.1016/j.omtm.2024.101383
- Leibiger, T., Remmler, L., Green, E., and Lee, K. (2024b). Biolayer interferometry for adeno-associated virus capsid titer measurement and applications to upstream and downstream process development. *Mol. Ther.* 32, 101306. doi:10.1016/j.omtm.2024.101306
- Levy, N. E., Valente, K. N., Choe, L. H., Lee, K. H., and Lenhoff, A. M. (2014). Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing. *Biotechnol. Bioeng.* 111 (5), 904–912. doi:10.1002/bit.25158
- Li, X., An, Y., Liao, J., Xiao, L., Swanson, M., Martinez-Fonts, K., et al. (2021). Identification and characterization of a residual host cell protein hexosaminidase B associated with N-glycan degradation during the stability study of a therapeutic recombinant monoclonal antibody product. *Biotechnol. Prog.* 37, e3128. doi:10.1002/btpr.3128
- Lou, R., Cao, Y., Li, S., Lang, X., Li, Y., Zhang, Y., et al. (2023). Benchmarking commonly used software suites and analysis workflows for DIA proteomics and phosphoproteomics. *Nat. Commun.* 14, 94. doi:10.1038/s41467-022-35740-1
- Lou, R., Liu, W., Li, R., Li, S., He, X., and Shui, W. (2021). DeepPhospho accelerates DIA phosphoproteome profiling through *in silico* library generation. *Nat. Commun.* 12 (1), 6685. doi:10.1038/s41467-021-26979-1
- Lou, R., Tang, P., Ding, K., Li, S., Tian, C., Li, Y., et al. (2020). Hybrid spectral library combining DIA-MS data and a targeted virtual library substantially deepens the proteome coverage. *ISCIENCE* 23 (3), 100903. doi:10.1016/j.isci.2020.100903
- Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 14, e8126. doi:10.15252/msb.20178126
- Lyle, A., Stamatis, C., Linke, T., Hulley, M., Schmelzer, A., Turner, R., et al. (2024). Process economics evaluation and optimization of adeno-associated virus downstream processing. *Biotechnol. Bioeng.* 121, 2435–2448. doi:10.1002/bit.28402
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26 (7), 966–968. doi:10.1093/bioinformatics/btq054
- Oh, Y. H., Becker, M. L., Mendola, K. M., Choe, L. H., Min, L., Lee, K. H., et al. (2023). Characterization and implications of host-cell protein aggregates in biopharmaceutical processing. *Biotechnol. Bioeng.* 120 (4), 1068–1080. doi:10.1002/bit.28325
- Peterson, A., Russell, J., Bailey, D., Westphall, M., and Coon, J. (2012). Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* 11 (11), 1475–1488. Article. doi:10.1074/mcp.O112.020131
- Pino, L., Searle, B., Bollinger, J., Nunn, B., MacLean, B., and MacCoss, M. (2020). The Skyline ecosystem: informatics for quantitative mass spectrometry proteomics. *MASS Spectrom. Rev.* 39 (3), 229–244. doi:10.1002/mas.21540
- Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M. Y., et al. (2011). mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* 8 (5), 430–435. doi:10.1038/nmeth.1584
- Rice, S., and Belani, C. (2022). Optimizing data-independent acquisition (DIA) spectral library workflows for plasma proteomics studies. *Proteomics* 22 (17), e2200125. doi:10.1002/pmic.202200125
- Ronsein, G., Pamir, N., von Haller, P., Kim, D., Oda, M., Jarvik, G., et al. (2015). Parallel reaction monitoring (PRM) and selected reaction monitoring (SRM) exhibit comparable linearity, dynamic range and precision for targeted quantitative HDL proteomics. *J. Proteomics* 113, 388–399. doi:10.1016/j.jprot.2014.10.017
- Rumachik, N. G., Malaker, S. A., Poweleit, N., Maynard, L. H., Adams, C. M., Leib, R. D., et al. (2020). Methods matter: standard production platforms for recombinant AAV produce chemically and functionally distinct vectors. *Mol. Ther. Methods Clin. Dev.* 18, 98–118. doi:10.1016/j.omtm.2020.05.018
- Smith, J., Strasser, L., Guapo, F., Milian, S., Snyder, R., and Bones, J. (2023). SP3-based host cell protein monitoring in AAV-based gene therapy products using LC-MS/MS. *Eur. J. Pharm. Biopharm.* 189, 276–280. doi:10.1016/j.ejpb.2023.06.019
- Soni, H., Lako, I., Placidi, M., and Cramer, S. (2024). Implications of AAV affinity column reuse and vector stability on product quality attributes. *Biotechnol. Bioeng.* 121, 2449–2465. doi:10.1002/bit.28500
- Szkodny, A. C., and Lee, K. H. (2022). Biopharmaceutical manufacturing: historical perspectives and future directions. *Annu. Rev. Chem. Biomol. Eng.* 13, 141–165. doi:10.1146/annurev-chembioeng-092220-125832
- Tiway, S., Levy, R., Gutenbrunner, P., Soto, F., Palaniappan, K., Deming, L., et al. (2019). High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* 16 (6), 519–525. doi:10.1038/s41592-019-0427-6
- Vandenbergh, L. H., Xiao, R., Lock, M., Lin, J. P., Korn, M., and Wilson, J. M. (2010). Efficient serotype-dependent release of functional vector into the culture medium during adeno-associated virus manufacturing. *Hum. Gene Ther.* 21 (10), 1251–1257. doi:10.1089/hum.2010.107
- Venable, J., Dong, M., Wohlschlegel, J., Dillin, A., and Yates, J. (2004). Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* 1 (1), 39–45. Article. doi:10.1038/NMETH705
- Walker, D., Yang, F., Carver, J., Joe, K., Michels, D., and Yu, X. (2017). A modular and adaptive mass spectrometry-based platform for support of bioprocess development toward optimal host cell protein clearance. *MABS* 9 (4), 654–663. doi:10.1080/19420862.2017.1303023
- Wang, X., Hunter, A. K., and Mozier, N. M. (2009). Host cell proteins in biologics development: identification, quantitation and risk assessment. *Biotechnol. Bioeng.* 103 (3), 446–458. doi:10.1002/bit.22304
- Wang, Z., Müller, M., Batruch, I., Chelur, A., Textoris-Taube, K., Schwecke, T., et al. (2022). High-throughput proteomics of nanogram-scale samples with Zeno SWATH MS. *Elife* 11, e83947. doi:10.7554/eLife.83947
- Wright, J. (2023). AAV vector production: troublesome host innate responses in another setting. *Mol. Ther. Methods Clin. Dev.* 30, 332. doi:10.1016/j.omtm.2023.08.001
- Yates, J. (1998). Mass spectrometry and the age of the proteome. *J. MASS Spectrom.* 33 (1), 1–19. doi:10.1002/(SICI)1096-9888(199801)33:1<1::AID-JMS624>3.0.CO;2-9
- Zhang, F., Ge, W., Huang, L., Li, D., Liu, L., Dong, Z., et al. (2023). A comparative analysis of data analysis tools for data-independent acquisition mass spectrometry. *Mol. Cell. Proteomics* 22, 100623. doi:10.1016/j.mcpro.2023.100623
- Zhang, F., Ge, W., Ruan, G., Cai, X., and Guo, T. (2020). Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* 20 (17–18), e1900276. doi:10.1002/pmic.201900276
- Zhang, X., Jin, X., Liu, L., Zhang, Z., Koza, S., Yu, Y., et al. (2021). Optimized reversed-phase liquid chromatography/mass spectrometry methods for intact protein analysis and peptide mapping of adeno-associated virus proteins. *Hum. Gene Ther.* 32, 1501–1511. doi:10.1089/hum.2021.046
- Zhang, Y., Fonslow, B., Shan, B., Baek, M., and Yates, J. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chem. Rev.* 113 (4), 2343–2394. doi:10.1021/cr3003533