

Quantitative proteomic analysis of residual host cell protein retention across adeno-associated virus affinity chromatography

Thomas M. Leibiger,¹ Lie Min,¹ and Kelvin H. Lee¹

¹University of Delaware, Department of Chemical and Biomolecular Engineering, Newark, DE 19713, USA

To better understand host cell protein (HCP) retention in adeno-associated virus (AAV) downstream processes, sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) was used to quantitatively profile residual HCPs for four AAV serotypes (AAV2, -5, -8, and -9) produced with HEK293 cells and purified using POROS CaptureSelect AAVX affinity chromatography. A broad range of residual HCPs were detected in affinity eluates after purification ($N_{total} = 2,746$), and HCP profiles showed universally present species ($N_{universal} = 1,117$) and species unique to one or more AAV serotype. SWATH-MS revealed that HCP persistence was dominated by high-abundance conserved species (HACS), which appeared across all serotype conditions studied. Due to the notable contribution of these species to overall residual HCP levels, physical and functional characteristics of HACS were examined to determine trends that coincide with persistence. Subnetwork interaction mapping and Gene Ontology function enrichment analysis revealed extensive physical interactions between these proteins and significant enrichment for biological processes, molecular functions, and reactome pathways related to protein folding, nucleic acid binding, and cellular stress. The abundant and conserved nature of these HCPs and their functions offers a new perspective for mechanistic evaluations of impurity retention for AAV downstream processes.

INTRODUCTION

Adeno-associated virus (AAV) is the dominant delivery platform for *in vivo* gene therapy due to low immunogenicity, a strong safety profile, and the ability of specific serotypes to target diverse tissue systems.¹ As of the end of 2023, there have been over 200 completed or in-progress clinical trials using recombinant AAV (rAAV) therapies and 6 approved products for the treatment of genetic disorders.² Despite the notable clinical and commercial success of AAV gene therapies, widespread availability of these treatments is limited by an inability to meet increasing demand for Good Manufacturing Practices-produced vectors.³ To combat supply challenges, AAV production systems using HEK293 cells have transitioned from adherent culture systems to larger-scale suspension formats, with increasing viral titers resulting from transfection and cell culture process parameter optimizations.^{4,5} Additionally, packaging and producer cell lines

are being developed to allow for improved culture scalability, reduced cost of goods, and elimination of process constraints associated with transient plasmid transfection.^{6,7}

As upstream platforms continue to evolve, downstream processes must adapt to ensure that highly pure AAV products can be isolated from larger and more complex feed streams. Aside from product isolation and final formulation, the AAV downstream process plays a key role in process- and product-related impurity removal, including residual host cell proteins (HCPs), exogenous DNA, fragmented and aggregated capsid species, and improperly packaged or empty AAV capsids.⁸ Reduction in DNA and HCP impurities has been correlated with improvements in *in vivo* transduction efficiency regardless of AAV serotype or tissue target,⁹ potentially allowing for reduced dosing requirements for highly pure products. Additionally, even with the use of a human expression system such as HEK293 cells, residual HCPs may still pose immunogenicity risks to patients due to the high dosing requirements of some therapies.¹⁰ However, the challenge of residual HCP clearance for AAV downstream bioprocesses is compounded by an increased impurity burden from cell lysis at harvest and the need for substantial volume reduction due to low product expression.

Several studies have explored HCP retention for AAV purification processes and examined co-purifying residual HCPs from AAV-producing HEK293 culture systems.^{11–14} Although these studies provide valuable insights, they do not quantitatively capture residual impurity profiles across different AAV serotypes from a controlled production and purification scheme. In this work, data-independent acquisition (DIA) liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods were applied to quantitatively profile residual HCP species after purification with POROS CaptureSelect AAVX affinity chromatography. Sequential window acquisition of all theoretical fragment ion mass spectra (SWATH-MS) was performed using an ion spectra library constructed from data-dependent acquisition (DDA) LC-MS/MS.

Received 6 August 2024; accepted 14 November 2024;
<https://doi.org/10.1016/j.omtm.2024.101383>.

Correspondence: Kelvin H. Lee, University of Delaware, Department of Chemical and Biomolecular Engineering, 590 Avenue 1743, Newark, DE 19713, USA.
E-mail: khl@udel.edu



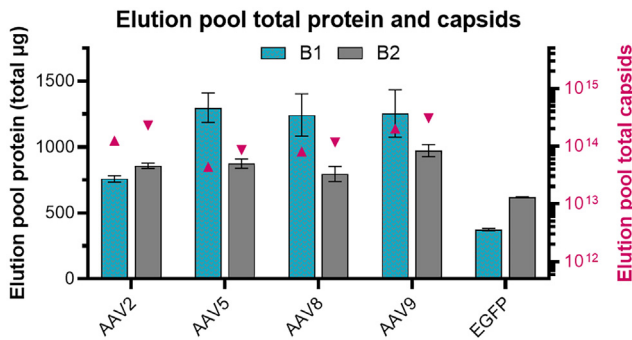


Figure 1. POROS CaptureSelect AAVX elution pool total protein and AAV capsids

POROS CaptureSelect AAVX affinity chromatography elution pool total protein amounts measured by Bradford and total AAV capsid content measured by AAVX Octet for AAV2, -5, -8, and -9, B1 and B2. Total protein (bars) is plotted on the left y axis on a normal scale. Elution pool total AAV capsid data (magenta triangles) is plotted on the right y axis on a log₁₀ scale. Error bars for protein concentration measurements correspond to SDs of technical replicates ($n = 3$).

The quantitative proteomic workflow was applied across four AAV serotypes (AAV2, -5, -8, and -9), with a model EGFP transgene produced from a suspension HEK293 culture system. Groups of high-abundance residual HCPs were identified that appeared across all AAV-containing samples. These high-abundance conserved HCP species, denoted herein as HACS, were characterized based on different abundance ranking criteria. Due to the notable contribution of HACS to residual impurity levels following affinity purification, physical and functional characteristics of these species were further examined. Molecular weight (MW) and isoelectric point (pI) profiles were compared to the full set of identified residual HCPs to determine whether HACS show physical property trends that might be related to downstream persistence. Physical subnetwork interaction mapping and Gene Ontology (GO) functional enrichment analyses were used to identify the participation of HACS in conserved protein complexes and to determine whether this group of HCPs show enrichment for specific biological processes, molecular functions, or reactome pathways. Overall, the characterizations and residual HCP trends identified in this work represent a step toward a more comprehensive understanding of process-related impurity retention and clearance for scalable AAV downstream processes.

RESULTS

AAV production, purification, and characterization

AAV-producing cultures for both batch 1 (B1) and batch 2 (B2) had peak viability between 24 and 48 h followed by declining viable cell density (VCD) and viability until harvest at 72 h post-transfection (Figure S1). The EGFP control condition, transfected only with the pEGFP gene of interest (GOI) plasmid, remained at >95% viability throughout the culture duration with an elevated VCD compared to AAV production conditions. Vector genome (VG) titers appeared to be serotype dependent, with AAV9 showing the highest lysate VG titer between 4 and 5×10^{11} VG/mL culture (Figure S2). All VG titers

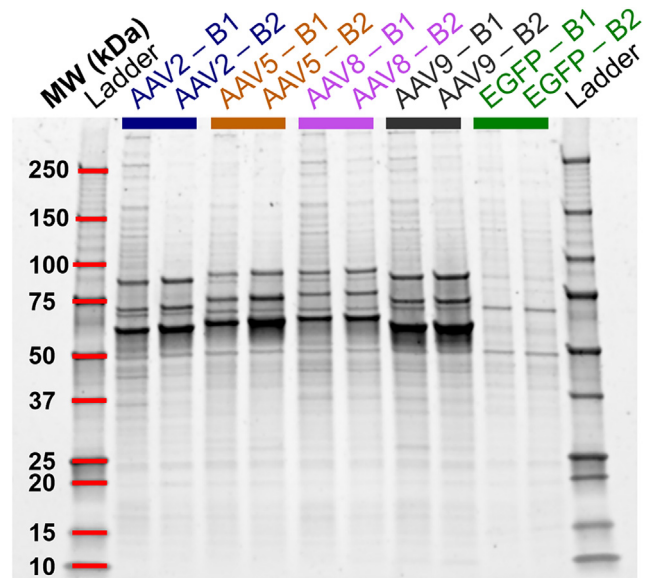


Figure 2. SYPRO Ruby-stained protein gel of AAV lots purified with POROS CaptureSelect AAVX resin

SYPRO Ruby-stained protein gel showing AAV2 (blue), AAV5 (orange), AAV8 (magenta), AAV9 (charcoal), and EGFP control (green) lysates after purification by POROS CaptureSelect AAVX affinity chromatography. Neutralized eluates were 4-fold concentrated and an equal volume of concentrated eluate was loaded per gel lane (2.5 µL). Viral capsid protein banding corresponding to VP1 (~80–81 kDa), VP2 (~65–66 kDa), and VP3 (~59–60 kDa) can be seen for each AAV-containing condition.

were greater than 1×10^{11} VG/mL culture, with some variability between biological replicate cultures observed, as shown in Figure S2. VG titer measurement in the supernatant compared to the lysate varied by serotype. AAV8 was the most secreted, with approximately 66% (range 63%–68%) of total VG titer appearing in the supernatant. AAV2 was the least secreted, with approximately 1% or less of total VG titer appearing in the supernatant. Affinity purification was performed with no observed process deviations, and UV absorbance measured at 280 nm baselined below 10 mAU in all cases during the wash block prior to elution (Figure S3). Capsid yields were lower for AAV2 (43% and 65% for B1 and B2, respectively) and AAV5 (82% and 63% for B1 and B2, respectively) compared to AAV8 (89% and 95% for B1 and B2, respectively) and AAV9 (92% and 97% for B1 and B2, respectively). Smaller but distinct elution peaks can be seen on the EGFP control chromatograms, indicating the presence of non-specific binding and elution of HCP impurities. Elution pool total protein and total capsid content show slight variability between biological duplicates, with lower elution pool protein measured for EGFP control lysate conditions compared to AAV-containing conditions after purification (Figure 1). Following purification, protein gel banding consistent with viral capsid proteins VP1, VP2, and VP3 was seen for all AAV-containing samples (Figure 2). A wide distribution of residual HCPs can also be visualized on the gel with a notable presence of cellular impurities within the 25- to 75-kDa range. Intact

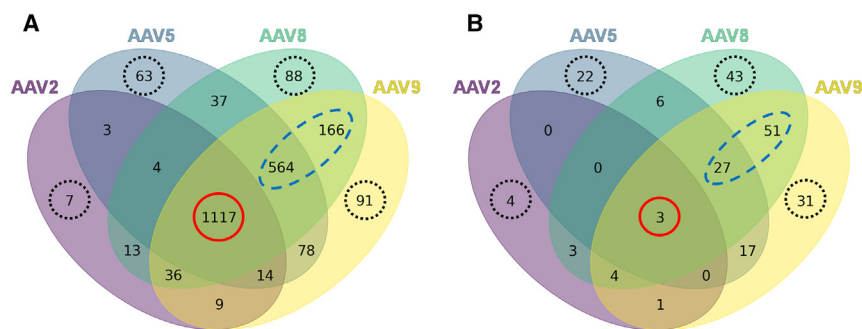


Figure 3. Residual host cell protein identifications by serotype condition after POROS CaptureSelect AAVX purification

All identified residual host cell proteins (HCPs) (A) and HCPs present in AAV-containing conditions but absent in the EGFP control lysate conditions (B) after POROS CaptureSelect AAVX affinity purification. Protein detection in all three triplicate LC-MS/MS injections for both biological duplicates was required for inclusion in each serotype group. For control-corrected conditions, HCPs detected in at least one EGFP control biological duplicate were removed. HCPs detected across all AAV serotypes are shown in red circles. A notable overlap in residual protein species was observed between AAV8 and AAV9 samples, shown in blue dashed circles. Proteins that appeared unique to each AAV serotype are shown in black dotted circles.

protein LC-MS was performed after AAV purification to confirm serotype identities prior to sample digestion and LC-MS/MS analysis. VP1, VP2, and VP3 were detected, with close matching to theoretical values for all serotypes (Table S1).^{15,16} Deconvoluted mass spectra for VP3 species are shown in Figure S4 for reference. Negative staining transmission electron microscopy (nsTEM) imaging further confirmed the presence of AAV capsids following affinity purification and shows residual process-related impurities (Figure S5). Purified capsids appeared as distinct spheres in the nsTEM images, but larger aggregates were also observed for all serotypes. These aggregates appeared predominantly in areas surrounding what appear to be degraded or improperly formed AAV capsids carried through affinity purification.

Qualitative residual HCP characterization

SWATH-MS data were collected and processed in triplicate for each sample. Detection in all three triplicate data acquisitions was required for an HCP to be considered present in any given sample. Overall, 2,746 unique protein species derived from the producing HEK293 cell line were identified in at least one sample ($n = 10$), with 1,117 species being detected for all AAV-containing samples ($n_{AAV} = 8$) following POROS CaptureSelect AAVX affinity purification (Figure 3A). Subsets of retained cellular impurities were detected for AAV-containing samples that were absent for EGFP control material after processing (Figure 3B). Interestingly, within this group of control-corrected residual HCPs, universal species that were detected for all four serotypes and serotype-specific species were observed. A greater overlap in conserved HCPs was seen between AAV8 and AAV9 after EGFP control correction ($N = 78$), and AAV2 showed a lower number of shared HCPs with other serotypes. Of the 1,117 HCPs detected in all AAV-containing samples, 1,091 (97.7%) were measured in both EGFP control biological duplicates, and 1,114 (99.7%) were measured in at least one EGFP control biological duplicate, indicating that most cellular impurities were carried through the process in a measurable capacity regardless of the presence of AAV capsids. HCPs that appeared unique to one specific AAV serotype were measured in the EGFP control lysate at a lower consistency (Figure 3).

Quantitative residual HCP characterization

HCP protein amounts (ng) calculated from SWATH-MS were averaged across triplicate injections for each sample. Median coefficients of variation (CVs) for HCP amounts (ng) calculated from triplicate injections ranged from approximately 5% to 7.5% across the sample set ($n = 10$) (Figure S6). All SWATH-MS outputs including raw data showing average protein amounts (ng) and CV determined from yeast alcohol dehydrogenase (ADH) reference injections and protein abundances (ng/ μ g total HCP) are included in Table S2. Despite the broad distribution of residual HCPs identified with LC-MS/MS, quantitative proteomic profiling using SWATH-MS revealed that cellular impurity persistence is driven primarily by groups of high-abundance HCPs (Figure 4). Total HCP identifications ranged from 1,286 to 2,303 unique proteins across the purified AAV samples, but the 25 highest-abundance species within each sample, ranging from 1.1% to 1.9% of total HCPs identified for each respective sample, comprised a median of 47.7% of total HCP amount (ng) across the samples (range 38.0%–61.9%). High-abundance species were present for biological duplicate samples (Figure S7) and were largely conserved across serotype conditions (Figure S8). High-abundance species were compared across serotypes to identify HACS. There were 39 HCPs measured to be within the 100 most-abundant species across all AAV-containing conditions, which are referred to as the top-100 \cap ($N = 39$) subgroup of HCPs. There were 251 HCPs measured to be within the 500 most-abundant species in all AAV-containing samples, which are referred to as the top-500 \cap ($N = 251$) subgroup of HCPs. The top-500 \cap ($N = 251$) subgroup was used for abundance-ranking assessments. A number of these 251 HCPs have been reported for their persistence across AAV downstream processes, including heat shock 70-kDa protein 1B (HSPA1B),¹² nucleolin,^{11,17,18} nucleophosmin,^{18,19} Y-box-binding protein 1,¹² *ruvB*-like 2,¹² alpha enolase,¹² and various heterogeneous nuclear ribonucleoproteins.¹⁸

Abundance ranking of residual HCPs

High-abundance or difficult-to-remove residual HCPs can pose risks to downstream bioprocesses by contributing substantially to intermediate and final impurity profiles, impacting outright impurity levels

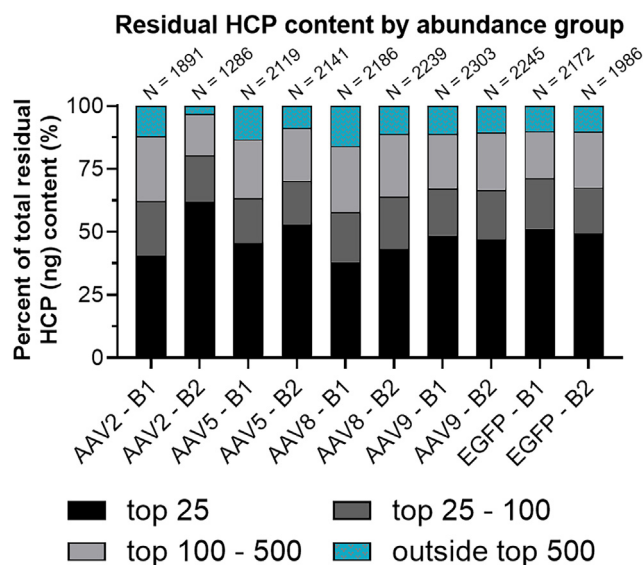


Figure 4. Percentage of total residual HCP content by abundance group
Residual HCP profiling for the top 25, 25–100, and 100–500 most-prevalent species in each sample by normalized abundance (ng/ μ g HCP). Low-abundance HCPs outside of the top 500 are shown in turquoise. Total identified HCPs in each sample are shown at the top of each bar.

and the ability of a process to match process-related impurity specifications that may be on file with regulatory authorities. To identify the most persistent cellular impurities, three separate abundance-ranking criteria were applied across the top-500 ($N = 251$) HCP subgroup. Within the top-500 ($N = 251$) subgroup, comparisons between HCP species showed a disproportionate abundance distribution, with HSPA1B, hsc70-interacting protein (ST13), and heat shock protein HSP 90-beta (HSP90AB1) appearing at notably high amounts compared to other HCPs (Figure 5). The 25 highest-abundance species across all AAV-containing conditions are summarized in Table 1, with HCPs listed in order of highest median abundance ranking. To avoid bias in ranking designation by median HCP abundance, mean abundance ranking and weighted abundance ranking were also performed for comparison, with trending across the ranking systems shown in Figure 5C. Ranking systems show agreement based on linear regression R^2 of 0.966, 0.932, and 0.970 for median vs. mean abundance rank, median vs. weighted abundance rank, and mean vs. weighted abundance rank, respectively. The agreement between ranking systems indicates that consistent HCP abundance trends are observed even when applying different ranking metrics, giving confidence to HACS designations. Full-abundance ranking outputs are displayed in Table S3.

Physical and functional HCP characterization

Physical subnetwork interaction mapping of top-100 ($N = 39$) HCPs showed notable protein-protein interactions, indicating high-confidence participation of these species in conserved protein complexes (Figure 6). GO functional enrichment analyses for this subgroup revealed enrichment for species involved in specific bio-

logical processes, molecular functions, and reactome pathways, namely those related to protein folding, RNA binding, and cellular response to stress (Figure S9). These functional enrichments were statistically significant (false discovery rate [FDR] < 0.05) against enrichment backgrounds of varying specificity, including the human genome, all identified HCPs ($N_{total} = 2,746$), and all conserved HCPs in AAV-containing samples ($N_{universal} = 1,117$). Trends of lower mean and median pI and higher mean and median MW across the top-100 ($N = 39$) group of HCPs compared to all identified HCPs were also observed (Figure 7). These trends were also seen when comparing subgroups of varying abundance within each AAV serotype condition (Figures S10 and S11).

DISCUSSION

Based on MW calculation, a typical AAV2 (~ 3.74 MDa) production process with a VG titer of 1×10^{12} /mL yields approximately 6 mg of packaged vector per liter of cell culture. A high-titer monoclonal antibody process by comparison can produce 8 g/L,²⁰ an over 1,000-fold higher product concentration. Because AAV2 remains cell associated,²¹ recovery requires cell lysis, which introduces increased impurity burden compared to bioprocesses with secreted products readily harvested from the culture supernatant. This combination of relatively low product expression and elevated impurity burden necessitates rigorous downstream process development to yield highly pure vectors. Characterization of impurity retention trends can contribute to improved process understanding, and it ultimately allows for the implementation of targeted strategies to minimize impurity production in the upstream process and maximize removal in the downstream process. Previously, Dong et al. identified a few co-purifying cellular proteins retained across CsCl gradient ultracentrifugation,¹¹ and Strobel et al. performed CsCl and iodixanol gradient ultracentrifugation fractionation and characterized several prominent retained cellular proteins via SDS-PAGE band excision and LC-MS/MS.¹⁸ However, impurity profiles and mechanisms driving HCP retention for intermediately purified AAV products of variable serotype produced from suspension HEK293 culture and purified using a scalable affinity chromatography method are poorly understood. In this work, we sought to characterize residual HCP profiles after POROS CaptureSelect AAVX affinity chromatography for four AAV serotypes (AAV2, -5, -8, and -9) with the goal of identifying quantitative trends for individual HCP species and assessing the physical and functional properties of HACS to postulate mechanisms of retention and allow for further experimentation focused on targeted HCP reduction and removal strategies.

Detection of highly abundant conserved residual HCPs

Across all four vector serotypes studied, a small subset of residual HCPs dominated overall HCP content, with a median of 47.7% of residual HCP amount (ng) being derived from only 25 HCP species. Further analysis of abundant impurities revealed that these HCPs were highly conserved across biological replicates and serotype conditions, meaning processes that specifically target downregulation or removal of these highly abundant residual HCPs can substantially lower total residual HCP levels. This could be

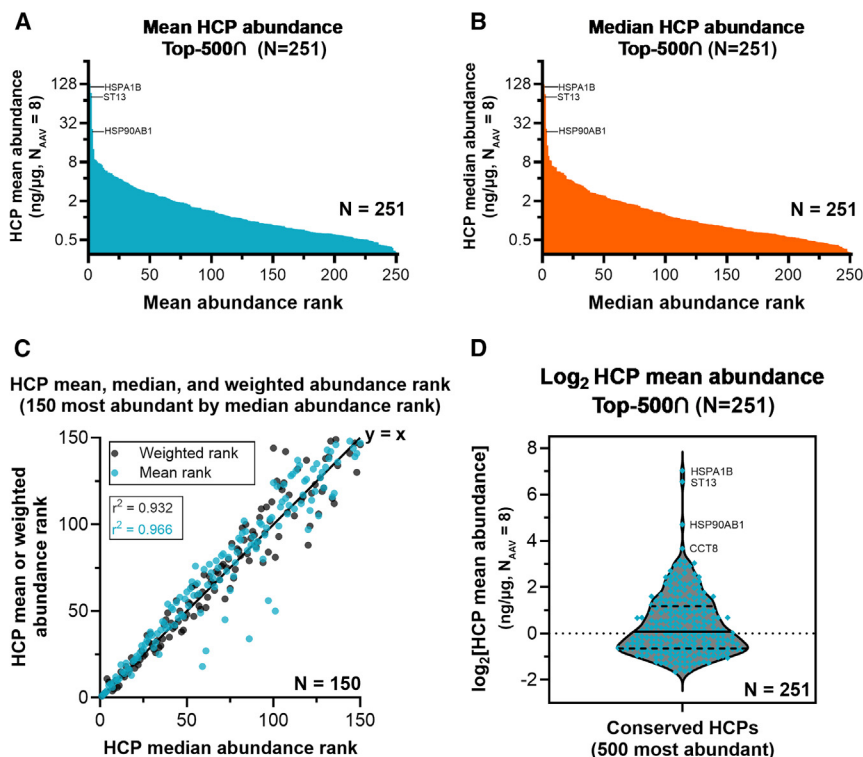


Figure 5. HCP abundance profiling by mean, median, and weighted ranking

Individual HCP abundance profiling for the 251 species that appear within the 500 most-abundant subgroup across all AAV-containing samples (top-500). These species were ranked by mean abundance (ng/μg HCP) across all samples (A) and median abundance (ng/μg HCP) (B), shown in log₂ space. An additional ranking system based on weighted abundance scoring was devised, which was compared to the mean and median abundance ranking for the 150 most-abundant HCPs by median rank (C). Several HCPs appeared in particularly high abundance, which can be visualized more clearly from the log₂ transformed HCP mean abundance violin plot (D).

Residual HCP physical property trends

Across all AAV serotypes studied and for the EGFP control condition, POROS CaptureSelect AAVX affinity purification more readily cleared smaller HCPs that are neutral or positively charged at operating pH. Across the human proteome, the distribution of pI is known to be bimodal, with a minima around physiological pH (~7.4) due to protein instability at a pH near that of the pI.²⁷ The profile of residual HCPs identified herein adheres to this distribu-

tion; however, the higher-abundance species skewed toward proteins with reduced pI (between 4 and 6) that carry a negative charge at affinity-binding conditions of pH 7.5 (Figures 7 and S10). Additionally, the overall distribution of residual HCP MW shifts upward in correspondence to protein abundance, suggesting that the affinity chromatography process removes a higher proportion of smaller proteins, while larger species have greater propensities for retention (Figures 7 and S11). These trends were observed equally across all AAV-containing samples and EGFP control samples, suggesting that this protein MW and charge trending with residual HCP abundance is capsid independent. It is possible that HCP MW and charge properties influence the tendency of non-specific binding to the AAVX affinity ligand or drive the formation of HCP-rich assemblies that can more easily persist across the chromatography process. However, additional experimentation is needed to rigorously demonstrate these mechanisms.

achieved through HCP knockout,²² upstream process development,²³ affinity chromatography wash selection,²⁴ or design of further downstream unit operations. The observation that 97.7% of HCPs detected in all AAV-containing conditions were also seen in both biological replicates of the EGFP control condition suggests a product-independent retention mechanism such as non-specific interaction with the affinity resin. Visible elution peaks (UV absorbance at 280 nm) for affinity chromatograms of EGFP control material containing no AAV capsids (Figure S3C), along with notable protein banding via SDS-PAGE for purified EGFP control material, indicate binding of HCPs to the resin and subsequent elution at low pH (Figure 2).

Despite most residual impurities appearing to be retained in a vector-independent manner, the presence of specific HCPs in AAV-containing conditions that are absent in EGFP control conditions suggests one or more mechanisms of vector-mediated retention. This retention may be driven by protein-protein interactions with viral capsids or upregulation of certain cellular proteins in the presence of AAV production. The notable overlap in residual HCPs retained for AAV8 and AAV9 conditions that were absent in all other AAV and EGFP control conditions may be driven by serotype-specific properties. AAV8 and AAV9 capsids share 93% aligned amino acid sequence identity and have been previously shown to share substantial binding similarities to serum proteins compared to other AAV serotypes,^{25,26} which could relate to propensity for specific capsid interactions with cellular proteins.

Larger aggregate species can be seen in nsTEM images for all four AAV serotypes following affinity purification (Figure S5). Visual inspection of images suggests that degraded or improperly assembled capsids may function as agglomeration sites for aggregates, which could contain residual HCPs and DNA. The persistence of these larger aggregate species through purification may be driven by physical association with AAV capsids or by non-specific interactions with the affinity resin. Interestingly, the three highest-abundance residual HCPs identified in this work—HSPA1B, ST13, and HSP90AB1—are involved in binding to misfolded or aggregated protein assemblies.^{28–30}

Table 1. The 25 most abundant residual HCPs by median normalized abundance (ng/μg total HCP) ranking aggregated across all AAV-containing samples ($n_{AAV} = 8$)

Median abundance rank	Protein name (<i>Homo sapiens</i>)	Protein accession	Median HCP abundance, ng/μg total HCP
1	heat shock 70-kDa protein 1B*	NP_005337.2	126.89
2	hsc70-interacting protein isoform 1*	NP_003923.2	89.59
3	heat shock protein HSP 90-beta isoform a*	NP_031381.2	26.21
4	cytoplasmic dynein 1 heavy chain 1	NP_001367.2	14.37
5	fatty acid synthase isoform X1	XP_011521840.1	9.95
6	elongation factor 1-alpha 1*	NP_001393.1	8.52
7	T-complex protein 1 subunit theta isoform 1*	NP_006576.2	8.38
8	heat shock protein HSP 90-alpha isoform 2*	NP_005339.3	6.98
9	heat shock protein 75-kDa, mitochondrial isoform 1 precursor*	NP_057376.2	6.84
10	heat shock cognate 71-kDa protein isoform X1*	XP_011541100.1	6.68
11	DNA damage-binding protein 1*	NP_001914.3	6.63
12	importin subunit beta-1 isoform 1	NP_002256.2	5.74
13	splicing factor 3B subunit 3*	NP_036558.3	5.73
14	far upstream element-binding protein 2 isoform 1*	NP_003676.2	5.71
15	T-complex protein 1 subunit epsilon isoform a*	NP_036205.1	5.52
16	clathrin heavy chain 1 isoform 1	NP_004850.1	5.45
17	protein kinase C iota type	NP_002731.4	5.44
18	chromobox protein homolog 3 isoform X1*	XP_005249668.1	5.08
19	nucleolin*	NP_005372.2	4.87
20	endoplasmic reticulum chaperone BiP precursor*	NP_005338.1	4.36
21	vimentin	NP_003371.2	4.32
22	cleavage stimulation factor subunit 2 isoform 2*	NP_001316.1	4.26
23	26S proteasome non-ATPase regulatory subunit 2 isoform 1*	NP_002799.3	4.26
24	T-complex protein 1 subunit zeta isoform a*	NP_001753.1	4.18
25	heat shock-related 70-kDa protein 2*	NP_068814.2	4.04

Species presence within the 500 most-abundant HCPs by normalized abundance (top-500 \cap) across all AAV sample conditions was required for ranking eligibility.

*These proteins have primary functions related to either protein folding, response to unfolded proteins, or nucleic acid binding.

Residual HCP functional enrichment analysis

Within the top-100 \cap ($N = 39$) subgroup identified in this study, there is a functional enrichment for proteins involved in protein folding, chaperone-mediated protein folding, and protein refolding biological processes. These enrichments are significant (FDR <0.05) not only against an enrichment background of the human genome but also against the group of HCPs identified in all AAV-containing samples ($N_{universal} = 1,117$). Among the 25 highest-abundance HCPs summarized in Table 1, 19 species have primary functions related to protein folding, unfolded protein response (UPR), or nucleic acid binding. Subnetwork mapping using STRING shows substantial physical interactions between the top-100 \cap ($N = 39$) group, with 34 of 39 species sharing at least one physical interaction network edge with another species. This demonstrates a high prevalence of known protein-protein interactions between these species and participation in conserved protein complexes, which may influence overall retention properties (Figure 6). GO enrichment in nucleic acid binding was also observed, which may relate to previous studies describing extensive interactions between host cell nucleic acid-binding proteins and the AAV genome throughout the virus life cycle.^{12,19,31}

HCP retention mechanisms

Various HCP retention mechanisms have been broadly described for recombinant protein bioprocessing. Direct association of HCPs with the target protein has been demonstrated,^{32–34} which can be caused by electrostatic interactions, hydrophobic effects, and hydrogen bonding properties. Chromatin and histone-rich protein complex formation³⁵ and high-MW aggregate formation driven by proteins involved in the UPR have been shown to contribute to HCP persistence.³⁶ Recently, Panikulam et al. has also reported HCP network-driven retention whereby an increased propensity for impurity co-elution is mediated by HCP-HCP interactions from species participating in conserved protein networks.³⁷ Based on the results of the present study, it appears that HCP retention for AAV purification processes using POROS CaptureSelect AAVX affinity chromatography is a complex process that may be influenced by protein interactions with the resin, physical properties of individual proteins, and functional characteristics of HCPs such as involvement in protein or nucleic acid binding and participation in conserved protein networks.

Problematic residual HCPs

The retention of particular “problematic” HCPs has been shown to impact the stability of therapeutic proteins through protease-mediated degradation,³⁸ and can pose additional risks due to impacts to formulation components,²² unwanted biological activity, or drug modifications.³⁹ Although immunogenicity risks for human HCPs may be lower compared to non-human-derived proteins, the safety and AAV product quality impacts of individual HCPs derived from HEK293 cells remain unclear.⁴⁰ Several residual HCPs that have been denoted “high risk” in the context of mammalian bioprocessing for therapeutic protein production were detected within the top-500 \cap ($N = 251$) subgroup, including alpha-enolase (drug quality, modification), peptidyl-prolyl *cis-trans* isomerase A (drug

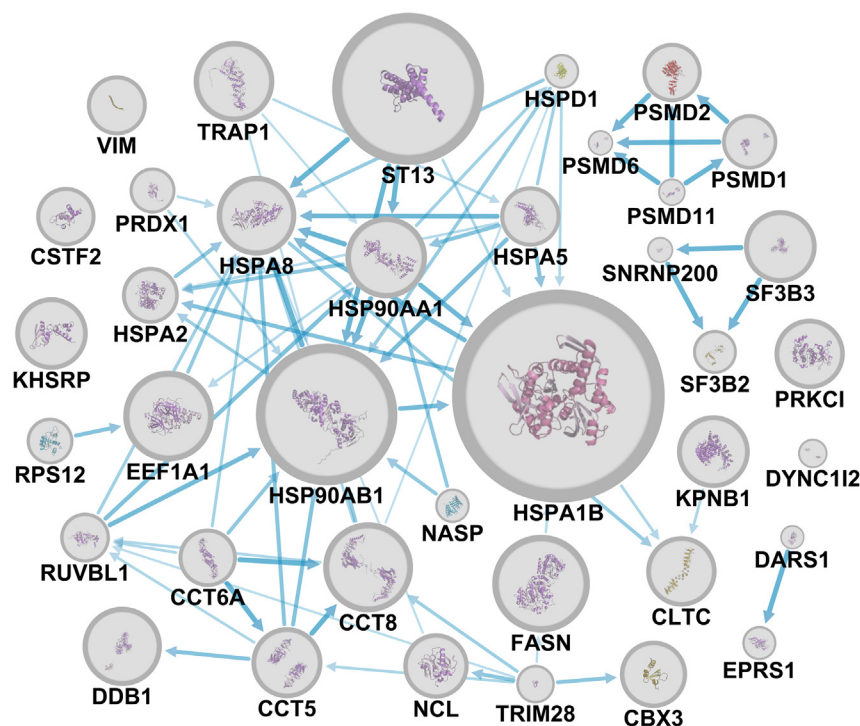


Figure 6. Physical subnetwork interaction mapping for high-abundance conserved HCP species

HCP species that appeared within the 100 highest-abundance subgroup for all AAV-containing samples are shown (top-100 Ω). Blue arrows denote physical subnetwork interactions between proteins, with thicker, darker lines corresponding to higher-confidence scores. Protein nodes are sized proportional to the \log_2 of the protein's median normalized abundance (ng/ μ g HCP) across AAV-containing conditions ($n_{AAV} = 8$).

quality, aggregation), endoplasmic reticulum chaperone BiP precursor (drug quality, aggregation), pyruvate kinase (immunogenicity), and peroxiredoxin-1 (immunogenicity).³⁹ Due to the high-abundance and potentially problematic nature of these HCPs, they may be of particular interest in AAV downstream processes for monitoring or targeted removal strategies.

Several LC-MS/MS characterizations have previously been reported for the assessment of residual HCPs for various AAV serotypes, which have examined impurities found in highly pure commercially acquired vector preparations produced from HEK293 cells.^{41,42} Consistent with our findings, these studies detected universally present residual HCPs across different AAV serotypes and serotype-specific impurities. Among the 33 HCPs identified by Hu et al. in 5 different serotype (AAV1, -2, -5, -6, and -9) samples purchased from Charles River Laboratories,⁴¹ 30 were identified in at least 1 sample for our study, and 21 of 33 were identified in all AAV-containing samples across the 4 serotypes studied ($n_{AAV} = 8$). Similarly, Smith et al. reported the 10 highest abundance HCPs measured in purified AAV2 material acquired from Patheon Viral Vector Services.⁴² Comparatively, 7 of 10 of these HCPs were detected in this work, with 5 of 10 being detected in all AAV-containing samples across the 4 serotypes studied ($n_{AAV} = 8$). Variability in residual HCP profiles between studies are likely influenced by many factors, including differences in production and purification methods used.

Overall, the results of this study represent a step toward a more complete understanding of HCP persistence in scalable AAV down-

stream processing. The finding that HCP levels after AAV affinity purification are dominated by a subset of highly abundant cellular proteins that are largely conserved across serotypes allows manufacturers to specifically target downregulation of these cellular proteins in the upstream process or remove these species through the improved design of affinity chromatography processes or subsequent downstream unit operations. It has been shown that complete removal of the 25 highest-abundance HCPs can correspond with a 50% reduction in total residual HCP levels following primary product capture. Furthermore, through comparisons to EGFP control conditions as well as physical and functional assessment of HACS, underlying mechanisms of retention are postulated that warrant further investigation, including non-specific resin interactions, aggregate formation, and network-mediated retention. Although evidence for these mechanisms was observed, additional experimentation is needed to explicitly demonstrate their involvement in HCP retention. Further probing of these mechanisms and their possible roles in HCP persistence for chromatographic AAV purification processes can lead to targeted removal strategies and ultimately contribute to the delivery of highly pure vectors to patients. Additionally, a more thorough understanding of AAV bioprocess-related impurity retention and downstream clearance could be achieved by comprehensive proteomic analysis of chromatography load material and flowthrough fractions, by the evaluation of chromatography wash additives, and through the study of HCP removal across further downstream polishing stages.

MATERIALS AND METHODS

AAV production

EXPI293 cells (Thermo Fisher) in exponential growth phase were exchanged into fresh EXPI293 Expression Medium (Thermo Fisher) at 2.5×10^6 viable cells/mL. Cells were seeded into 8 individual 1-L shake flasks (Corning) at a 200-mL working volume. All plasmids used for transfection were acquired from Addgene: pAdDeltaF6 (pAdH, Addgene, catalog no. 112867), pAAV-GFP (pEGFP, Addgene, catalog no. 32395),⁴³ pAAV2/2 (pRep2/Cap2, Addgene, catalog no. 104963), pAAV2/5 (pRep2/Cap5, Addgene, catalog no. 104964),

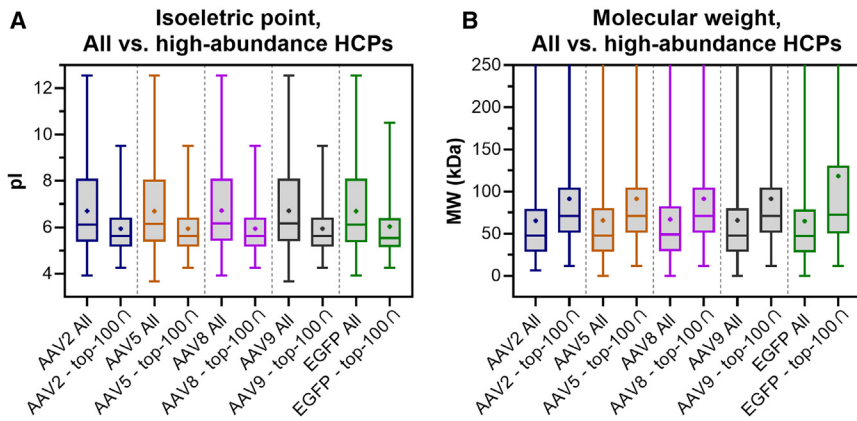


Figure 7. Isoelectric point and molecular weight boxplots for all identified and highly abundant HCPs within each serotype group

pI (A) and molecular weight (MW) (B) distributions for all identified HCPs across biological duplicates of each serotype condition (“All”) and HCPs that ranked within the top 100 highest abundance for both biological duplicates of each serotype condition (top-100). Lines within the boxes indicate medians, and plus symbols indicate means. Boxes span 25th–75th percentiles and bars show data ranges. Upper ends for data ranges of MW are cut off, and full data ranges are shown in [Figure S11](#).

pAAV2/8 (pRep2/Cap8, Addgene, catalog no. 112864), and pAAV2/9n (pRep2/Cap9, Addgene, catalog no. 112865) ([Table S4](#)). Transfections were carried out for each flask using the following parameters—total DNA delivery: 1.57 μg total DNA/ 10^6 cells, complexation volume: 10 mL (5% culture volume), complexation medium: Opti-Plex Complexation Buffer (Thermo Fisher), plasmid ratio: 1.5:1:1 M plasmid ratio (AAV5, -8, and -9) or 5:1:0.31 M plasmid ratio (AAV2) of pRep2/CapX:pAdH:pEGFP, transfection reagent: FectoVIR-AAV (Polyplus), DNA to transfection reagent ratio: 1.35 μg DNA/ μL FectoVIR-AAV, complexation time: 30 min. After the complexation hold time, complexes were added to cells, and flasks were placed into a Multitron HT (Infors, 25 mm throw) incubator at 135 rpm, 37°C, 80% relative humidity, and 5% CO₂. All cultures were harvested at 72 h post-transfection. Flasks, 2 \times 1 L, were transfected for each AAV serotype (2, 5, 8, and 9). Culture samples were taken from each flask at 24, 48, and 72 h post-transfection and were measured for cell growth trends using a Vi-Cell XR (Beckman Coulter) cell viability analyzer. The transfection procedure described above was performed two separate times with different plasmid and transfection reagent lots, and using EXPI293 cells grown from different vial thaws, for a total of 16 \times 1-L shake flasks ($n = 4$ flasks for AAV2, -5, -8, and -9). All rAAV vectors produced in this study were packaged with EGFP transgenes.

EGFP control material generation

In parallel with each biological replicate production batch, a 1-L flask was transfected as described previously, but only with delivery of the pEGFP plasmid. These flasks ($n = 2$) were used to generate “EGFP control” lysate material from cells receiving GOI plasmid and transfection reagent, but with no AAV production or viral helper genes delivered. DNA mass delivery of the pEGFP plasmid matching the 1:1:1.5 (pRep2/CapX:pAdH:pEGFP) molar plasmid ratio condition was used. The EGFP control flasks were cultured for 72 h, harvested, and purified with POROS CaptureSelect AAVX affinity chromatography in parallel with the AAV production cultures. Prior to purification, EGFP control lysates were 3-fold diluted with 1 \times PBS to control for an increased cell density at harvest compared to AAV-producing conditions. Methods used to produce and process the EGFP control material were intended

to most closely match the AAV-producing conditions without expression of AAV production or viral helper elements, thereby creating control material lots that could be taken through the purification and proteomic analysis workflows for comparison to AAV-production conditions.

Harvest and quantitative real-time PCR

Cells were harvested by centrifuging at 1,000 relative centrifugal force (RCF) with a 5920 R benchtop centrifuge (Eppendorf) for 10 min. Cell pellets were resuspended in 25 mL mammalian lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 2 mM MgCl₂, pH 8.5) and were subjected to 3 freeze-thaw cycles with a dry ice and ethanol slurry. Lysate mixtures were treated with Benzonase Nuclease (Sigma-Aldrich) at 25 U/mL and were incubated at 37°C for 60 min. Cell debris was removed by centrifuging at 3,428 RCF using a 5920 R benchtop centrifuge (Eppendorf) followed by 0.2 μm vacuum filtration (Fisher Scientific). Supernatants were treated with 2 M MgCl₂ to a concentration of 2 mM, followed by 25 U/mL Benzonase Nuclease (Sigma-Aldrich) and incubation at 37°C for 60 min. An additional exogenous DNA digestion step was performed for both lysates and supernatants by treating samples with DNase I (New England Biolabs) and incubating at 37°C for 60 min (2.5 μL sample, 2.5 μL DNase I, 2.5 μL DNase buffer, 17.5 μL molecular biology water). Prior to qPCR, capsids were digested by adding 2.5 μL of 20 mg/mL Proteinase K (Thermo Fisher) and incubated at 56°C for 90 min followed by Proteinase K inactivation at 95°C for 30 min. Samples (lysates and supernatants) were measured for VG titer relative to a PvuII (New England Biolabs) linearized pEGFP plasmid standard curve ([Figure S12](#)) diluted in molecular biology water. Linearized plasmid was desalted using a QIAquick PCR Purification Kit (Qiagen) and measured for concentration using UV absorbance (A₂₆₀) on a DS-11 FX+ (DeNovix) instrument. A 6-point standard curve (10⁹–10⁴ copies per reaction) was created by serial dilutions of the linearized plasmid. Samples were 10-fold diluted in molecular biology water after Proteinase K inactivation, and qPCR was performed on a CFX384 Touch Real-Time PCR system (Bio-Rad). TaqMan Fast Advanced Master Mix (Thermo Fisher) along with 900 nM primers/250 nM probe (IDT) targeting the EGFP

transgene were used (Table S5). After qPCR measurement, duplicate lots for each AAV serotype were pooled together to give approximately 50 mL lysate for each biological duplicate of the 4 AAV serotypes, for a total of 10 lysate material lots ($n = 2$ biological replicate lots for AAV2, -5, -8, and -9 and EGFP control).

POROS CaptureSelect AAVX affinity purification

Due to the low supernatant VG titer measured for AAV2, purifications were performed only from lysate material to allow for more consistent comparisons of residual impurity profiles. Lysate samples were purified using a “base-case” affinity chromatography process run with an AKTA Pure (Cytiva) fast protein liquid chromatography system. A TRICORN 5/50 (Cytiva) column was packed with 1 mL of fresh POROS CaptureSelect AAVX (Thermo Fisher) affinity resin for each run. 12 mL of each lysate lot was diluted with 12 mL equilibration buffer (20 mM Tris-HCl, 0.1 M NaCl, pH 7.5) and was sterile filtered using a 0.2- μ m vacuum filter (Fisher Scientific) immediately prior to loading. The following protocol was used for each run: column equilibration with 10 column volumes (ColV) of equilibration buffer, product loading (20 mL 1:1 diluted lysate), washing with 12 ColV equilibration buffer, and elution with 15 ColV elution buffer (0.1 M glycine-HCl, pH 2.6). All product contact steps (loading, washing, and elution) were performed at 2-min residence time (0.5 mL/min). Eluates were collected in 50-mL conical tubes (Fisher Scientific) containing 1.5 mL (10 vol/vol %) neutralization buffer (1 M Tris-HCl, pH 8.7).

Bradford, Octet AAVX, SDS-PAGE, intact protein LC-MS, and nsTEM

Elution pools were measured for protein concentration using Coomassie Plus (Bradford) Assay Reagent (Thermo Fisher) with an albumin standard (BSA) (Thermo Fisher) standard curve. Samples and standards were measured in triplicate with 40 μ L sample or BSA added to 260 μ L Coomassie Plus Reagent for each well. Six-point BSA standard curves ranged from 250 to 3.125 μ g/mL and were fit with second-order polynomials with $R^2 > 0.99$ in all cases. Flat bottom 96-well polystyrene plates (CELLTREAT) were used along with a SpectraMax M5 plate reader (Molecular Devices). Absorbance was measured at 595 nm, and all standard and sample wells were blank corrected by subtracting the triplicate averaged absorbance of 40 μ L elution buffer + 10% v/v neutralization buffer added to 260 μ L Coomassie Plus Reagent.

Elution pools and 1:1 diluted chromatography load materials were measured for capsid titer using Octet AAVX biosensors (Sartorius) on an Octet R8 (Sartorius) instrument, as described previously.⁴⁴ The following method parameters were used: 160 s baseline with Octet Sample Diluent (Sartorius), 600 s sample time, 5 probe regeneration cycles of 5 s each with regeneration buffer (10 mM glycine-HCl, pH 1.7), 30°C plate temperature, and 1,000 rpm agitation. Each well of a 96-well F-bottom microplate (Greiner) was loaded with 200 μ L sample or buffer. Probe hydration was performed with a 5-min incubation in Octet Sample Diluent prior to starting sample reads. Freshly thawed serotype-specific internal AAV reference stan-

dards serially diluted in Octet Sample Diluent were used for each plate, ranging from 1×10^{10} to 5×10^{12} capsids/mL. Octet Analysis Studio software version 12.2 was used for data analysis with dose response-4-parameter logistic regression (weighted Y) standard curves. Chromatography capsid yields were calculated by dividing the averaged elution capsid recovery ($n = 2$) by the averaged load capsid amount ($n = 2$).

Neutralized elution pools, 16.5 mL, were concentrated to a target volume of 4 mL using AMICON Ultra-15 10 kDa spin filters (Sigma-Aldrich). Prior to use, spin filter units were incubated with 15 mL 0.1% Pluronic F68 (Thermo Fisher) in $1 \times$ PBS for 5 min followed by sequential flushing with 0.01% Pluronic F68 in $1 \times$ PBS and then 0.001% Pluronic F68 + 200 mM NaCl in $1 \times$ PBS. After concentration, 5 μ L of each sample was 1:1 diluted with $2 \times$ Laemmli sample buffer (Bio-Rad) and heated at 90°C for 10 min. Then, 8 μ L diluted sample was added to each lane of a pre-cast polyacrylamide gel (4%–15% Mini-PROTEAN TGX Precast Protein Gel, 15-well, 15 μ L, Bio-Rad). The gel was run at 150 V for 50 min in Tris/glycine/SDS running buffer (Bio-Rad), fixed for 40 min in 50% methanol +7% acetic acid, stained for 3 h with SYPRO Ruby Protein Stain (Thermo Fisher), and washed for 1 h in 10% methanol +7% acetic acid. Imaging was performed using an Azure 600 imaging system (Azure Biosystems) set to the SYPRO Ruby channel (472 nm excitation/684 nm emission).

Intact protein LC-MS was performed using a Waters BioAccord with ACQUITY Premier system equipped with a ZORBAX RRHD StableBond C18 column (Agilent, 300 Å, 2.1×100 mm, 1.8 μ m). Prior to injection of each purified sample, 1.5×10^{11} VG were treated with 10% v/v acetic acid for 15 min followed by 5 min of centrifugation at 16,260 RCF.^{16,45} Additional method parameters for LC-MS data collection are listed in Table S6.

For nsTEM imaging, 400-mesh copper grids with a carbon film (Electron Microscopy Sciences) were glow discharged using an easiGlow discharge unit (Pelco) to render the films hydrophilic. The freshly glow discharged grids were floated on a drop of sample for several seconds, followed by washing with 4 drops of water and staining with 2% aqueous uranyl acetate. Excess stain was blotted with filter paper, and grids were allowed to dry prior to imaging. Images were collected using a Zeiss Libra 120 TEM instrument.

Trypsin digestion and sample cleanup

For each sample, a volume corresponding to 50 μ g protein was buffer exchanged into 100 mM triethylammonium bicarbonate at a 100- μ L volume target, and protein concentration measurement was performed using Coomassie Plus (Bradford), as described above. Samples containing 50 μ g protein were subjected to trypsin digestion, as described previously.^{46,47} Briefly, reduction and alkylation were achieved with 2.5 μ L of 100 mM Tris(2-carboxyethyl)phosphine (Thermo Fisher) incubated at 60°C for 1 h and with 5 μ L 150 mM iodoacetamide (Sigma-Aldrich) incubated in the dark for 30 min, respectively. Enzymatic digestion was carried out at 37°C for 16 h

with sequencing grade trypsin (Promega) at an enzyme-to-substrate mass ratio of 1:50. Digestions were then acidified with 4 μ L 20% formic acid (FA, Fisher Scientific). For samples with a final protein measurement less than 50 μ g, digestion was performed with reagent amounts scaled accordingly. Digested samples were subjected to OMIX C18 tip (Agilent) cleanup per manufacturer's instructions, with substitution of FA for heptafluorobutyric acid. Tips were conditioned with 50% acetonitrile (ACN, Fisher Scientific) and equilibrated with 1% FA. Washes were performed with 0.1% FA in water (Fisher Scientific) and elution with 50% ACN, 0.1% FA. Tips were regenerated, and the cleanup procedure was repeated twice for each sample using the same respective tip. Eluates were dried with a SpeedVac vacuum concentrator (Thermo Fisher).

LC-MS/MS data acquisition

Dried samples were resuspended in 2% ACN, 0.1% FA. Samples for triplicate injection were spiked with pre-digested ADH (Waters) at 5 fmol/ μ L, except for AAV5 B2, AAV8 B2, and AAV9 B2, which were spiked with 7.1, 7.0, and 5.4 fmol/ μ L, respectively. Samples equivalent to 5 μ g digested materials were injected for each LC-MS/MS analysis. Retention time standards (Biognosys) were prepared according to the manufacturer's recommendation, and 0.25 μ L was added with each injection. LC-MS/MS analysis was performed on a TripleTOF 6600 (Sciex) equipped with Eksigent nano 425 LC operating in microLC flow mode. LC separation was performed on a ChromXP C18CL column (3 mm, 120 \AA , 150 \times 0.3 mm) (Sciex) with mobile phase A (0.1% FA in water) and mobile phase B (0.1% FA in ACN) at a flow rate of 5 μ L/min. A program of 3%-25% mobile phase B over 68 min, 25%-35% mobile phase B over 5 min, 35%-80% mobile phase B over 2 min, and 80% mobile phase B for 3 min was used to elute peptides.

DDA was performed in positive ion mode with an MS1 full scan over a mass range of 400–1250 m/z , with a scan time of 250 ms, followed by MS/MS over a mass range of 100–1,500 m/z , with a scan time of 50 ms. The top 30 precursor ions were selected for fragmentation. Triplicate DDA data were acquired for all samples except "EGFP-B2," which contained enough protein for only duplicate injections. DIA SWATH-MS experiments were performed with an MS1 full scan followed by 64 MS/MS acquisitions with variable window sizes.^{46,48,49} Triplicate SWATH-MS data were acquired for all samples. Examples of total ion counts (intensity for all ions eluted vs. time) for each sample are shown in [Figure S13](#) for reference.

LC-MS/MS protein identification and quantitation

Database searches were performed as previously described,^{46,47} with modifications as follows: ProteinPilot version 5.1 (Sciex) was used to process and submit DDA data for searches against a local copy of the NCBI:Hu_RefSeqGRCh38 database supplemented with ADH, AAV replication and capsid proteins, retention time calibration standards, and common contaminants using the search engine Paragon Algorithm (Sciex). Search parameters were specified to include cysteine modifications by iodoacetamide, trypsin digestion, and a detected protein threshold at 10%. Identifications were based

on protein matches with an FDR of 1% and at least 2 peptide detections at 95% confidence limit.

For spectral library construction, triplicate DDA datasets for all project-specific samples were combined for a comprehensive database search using ProteinPilot software (version 5.1). The resulting group file from the ProteinPilot search was imported to Skyline (version 20.2.0.343, MacCoss Lab, University of Washington) to build a consolidated spectral library using BiblioSpec with peptides identified at 95% confidence limit or above.⁵⁰ In total, the spectral library consisted of 84,563 peptide precursors from 72,609 peptides mapped to 3,432 proteins. SWATH-MS data were extracted using a Skyline command-line interface with 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, resolving power of 36,000, and retention time tolerance of 4 min. Peaks were automatically picked and integrated with the mProphet algorithm based on the target decoy approach.⁵¹ A detection q value was assigned for each peak. Peak integration was manually checked for ADH, which contains the following peptides: ANELLINVK, SISIVGSYVGNR and VVGLSTLPEIYK. Peak areas were exported to the MSstats version 4.8.0 (Olga Vitek Lab, Northeastern University) input format.⁵² In MSstats, peak areas were \log_2 transformed, normalized with global standard (ADH) normalization, and the top three features of each protein were summarized using Tukey median polish to obtain the protein areas.

Triplicate SWATH-MS data for each sample were extracted and processed as a combined unit. Peptide detection at $q < 0.01$ in all three replicates was required for protein identification in each sample. Tukey's median polish was used for peak area summation, followed by protein peak area integration in MStats, then normalization performed against peak areas for ADH. Relative HCP amounts (ng) were calculated from normalized peak areas and the known mass of ADH spiked into each sample, with the assumption that all proteins generate a consistent concentration-dependent response.

SWATH-MS data normalization

Protein amounts (ng) calculated for each HCP using SWATH-MS were normalized across all samples. First, non-HCP species were removed from the analysis, including capsid proteins from AAV2, -5, -8, and -9 (YP_680426.1, Q9YIJ1, YP_077180.1, and Q6JC40), Rep78 protein (YP_680423.1), EGFP (C5MKY7), ADH (P00330), and modified trypsin (AC_000). The remaining cellular proteins were normalized by dividing the average calculated relative abundance across triplicate injections for each individual HCP by the total average relative abundance of all HCPs detected in the sample. These normalized outputs were multiplied by 1,000 to give individual ng/ μ g total HCP. These normalized values (ng/ μ g total HCP) were used for all further analyses.

HCP abundance ranking

HCP species were ranked across all AAV-containing samples using three separate methods to designate abundance order: median HCP

abundance, mean HCP abundance, and rank-weighted abundance. For median HCP abundance ranking, species were ordered based on the largest median normalized HCP abundance (ng/ μ g) across AAV-containing samples ($n_{AAV} = 8$). For mean HCP abundance ranking, species were ordered by the largest average normalized HCP abundance (ng/ μ g) across AAV-containing samples ($n_{AAV} = 8$). For rank-weighted abundance, HCP species for each AAV-containing sample were assigned a value ranging from 100 to 1 corresponding to the highest normalized abundance (ng/ μ g) species (assigned 100) to the 100th highest abundance species (assigned 1), and HCPs were sorted from highest total ranking summed across all samples ($n_{AAV} = 8$). To be eligible for abundance ranking, HCPs were required to be present within the 500 most-abundant species across all AAV-containing samples ($n_{AAV} = 8$). To further examine conserved high-abundance residual HCPs across samples, groups of the 50, 100, and 500 most-abundant residual HCPs were compared across AAV serotypes and to the EGFP control condition. To be considered within each abundance subgroup, the triplicate-averaged normalized HCP abundance (ng/ μ g) was required to be within the threshold (top 50, 100, or 500) for both biological duplicates. HCPs that were within the designated subgroups of for all AAV-containing samples were classified as the top-50 \cap ($N = 18$), top-100 \cap ($N = 39$), and top-500 \cap ($N = 251$) subgroups.

GO and functional enrichment analysis

The physical subnetwork interactions between the top-100 \cap ($N = 39$) subgroup were mapped using StringApp (version 2.0.3) in Cytoscape (version 3.10.1).⁵³ The default confidence score cutoff criteria of 0.4 was used for physical interaction mapping. GO functional enrichment analysis was performed using the *Homo sapiens* genome as a background. GO Molecular Function, GO Biological Process, and Reactome Pathway were assessed for enrichment, with an FDR <0.05 considered significant. Additional GO functional enrichment analyses were performed for the top-100 \cap ($N = 39$) subgroup against enrichment backgrounds of HCPs identified in at least one sample ($N_{total} = 2,746$) and HCPs identified in all AAV-containing samples ($N_{universal} = 1,117$).

DATA AND CODE AVAILABILITY

All data generated or analyzed as part of this study are available upon reasonable request.

ACKNOWLEDGMENTS

This work was supported in part by financial assistance award 70NANB17H002 from the US Department of Commerce, National Institute of Standards and Technology. Microscopy access was supported by grants from NIH-NIGMS (National Institute of General Medical Sciences) (P20 GM103446), the NIGMS (P20 GM139760) and the State of Delaware. The authors thank Shannon Modla in the Delaware Biotechnology Institute's Bio-Imaging Center for assistance with nsTEM imaging. The authors are grateful to Sartorius for providing the Octet AAVX biosensors. The graphical abstract figure was created using BioRender.

AUTHOR CONTRIBUTIONS

T.M.L. conceptualized the research and designed and performed the experiments. L.M. generated the LC-MS/MS data. T.M.L. and L.M. analyzed the data, created the figures, and wrote the original draft. K.H.L. edited the draft, secured the funding, and provided supervision.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.omtm.2024.101383>.

REFERENCES

- Clement, N., and Grieger, J.C. (2016). Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Molecular Therapy-Methods & Clinical Development* 3, Review. <https://doi.org/10.1038/mtm.2016.2>.
- Clinicaltrials.gov (accessed November 8, 2023).
- Dobrowsky, T., Gianni, D., Pieracci, J., and Suh, J. (2021). AAV manufacturing for clinical use: Insights on current challenges from the upstream process. *Current Opinion in Biomedical Engineering* 20, 100353. <https://doi.org/10.1016/j.cobme.2021.100353>.
- Zhao, H., Lee, K.J., Daris, M., Lin, Y., Wolfe, T., Sheng, J., Plewa, C., Wang, S., and Meisen, W.H. (2020). Creation of a High-Yield AAV Vector Production Platform in Suspension Cells Using a Design-of-Experiment Approach. *Mol. Ther. Methods Clin. Dev.* 18, 312–320. <https://doi.org/10.1016/j.omtm.2020.06.004>.
- Blessing, D., Vachey, G., Pythoud, C., Rey, M., Padrun, V., Wurm, F.M., Schneider, B.L., and Déglon, N. (2019). Scalable Production of AAV Vectors in Orbitally Shaken HEK293 Cells. *Mol. Ther. Methods Clin. Dev.* 13, 14–26. Article. <https://doi.org/10.1016/j.omtm.2018.11.004>.
- Lee, Z., Lu, M., Irfanullah, E., Soukup, M., and Hu, W.S. (2022). Construction of an rAAV Producer Cell Line through Synthetic Biology. *ACS Synth. Biol.* 11, 3285–3295. <https://doi.org/10.1021/acssynbio.2c00207>.
- Merten, O.W. (2024). Development of Stable Packaging and Producer Cell Lines for the Production of AAV Vectors. *Microorganisms* 12, 384. <https://doi.org/10.3390/microorganisms12020384>.
- Srivastava, A., Mallela, K.M.G., Deorkar, N., and Brophy, G. (2021). Manufacturing challenges and rational formulation development for AAV viral vectors. *J. Pharmaceut. Sci.* 110, 2609–2624. Review. <https://doi.org/10.1016/j.xphs.2021.03.024>.
- Ayuso, E., Mingozzi, F., Montane, J., Leon, X., Anguela, X.M., Haurigot, V., Edmonson, S.A., Africa, L., Zhou, S., High, K.A., et al. (2010). High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene Ther.* 17, 503–510. Article. <https://doi.org/10.1038/gt.2009.157>.
- Hebben, M. (2018). Downstream bioprocessing of AAV vectors: industrial challenges & regulatory requirements. *Cell Gene Ther. Insights* 4, 131–146.
- Dong, B., Duan, X., Chow, H.Y., Chen, L., Lu, H., Wu, W., Hauck, B., Wright, F., Kapranov, P., and Xiao, W. (2014). Proteomics Analysis of Co-Purifying Cellular Proteins Associated with rAAV Vectors. *PLoS One* 9, e86453. <https://doi.org/10.1371/journal.pone.0086453>.
- Satkunanathan, S., Wheeler, J., Thorpe, R., and Zhao, Y. (2014). Establishment of a Novel Cell Line for the Enhanced Production of Recombinant Adeno-Associated Virus Vectors for Gene Therapy. *Hum. Gene Ther.* 25, 929–941. Article. <https://doi.org/10.1089/hum.2014.041>.
- Rumachik, N.G., Malaker, S.A., Poweleit, N., Maynard, L.H., Adams, C.M., Leib, R.D., Cirolia, G., Thomas, D., Starnes, S., Holt, K., 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. <https://doi.org/10.1016/j.omtm.2020.05.018>.
- Zhang, S., Xiao, H., and Li, N. (2024). Analysis of Host Cell Proteins in AAV Products with ProteoMiner Protein Enrichment Technology. *Anal. Chem.* 96, 1890–1897. <https://doi.org/10.1021/acs.analchem.3c03884>.
- Lam, A.K., Zhang, J., Frabutt, D., Mulcrone, P.L., Li, L., Zeng, L., Herzog, R.W., and Xiao, W. (2022). Fast and high-throughput LC-MS characterization, and peptide mapping of engineered AAV capsids using LC-MS/MS. *Mol. Ther. Methods Clin. Dev.* 27, 185–194. <https://doi.org/10.1016/j.omtm.2022.09.008>.
- Jin, X., Liu, L., Nass, S., O'Riordan, C., Pastor, E., and Zhang, X.K. (2017). Direct Liquid Chromatography/Mass Spectrometry Analysis for Complete Characterization of

- Recombinant Adeno-Associated Virus Capsid Proteins. *Hum. Gene Ther. Methods* 28, 255–267. <https://doi.org/10.1089/hgtb.2016.178>.
17. Qiu, J., and Brown, K.E. (1999). A 110-kDa nuclear shuttle protein, nucleolin, specifically binds to adeno-associated virus type 2 (AAV-2) capsid. *Virology* 257, 373–382. Article. <https://doi.org/10.1006/viro.1999.9664>.
 18. Strobel, B., Miller, F.D., Rist, W., and Lamla, T. (2015). Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications. *Hum. Gene Ther. Methods* 26, 147–157. <https://doi.org/10.1089/hgtb.2015.051>.
 19. Satkunathan, S., Thorpe, R., and Zhao, Y. (2017). The function of DNA binding protein nucleophosmin in AAV replication. *Virology* 510, 46–54. Article. <https://doi.org/10.1016/j.virol.2017.07.007>.
 20. Szkodny, A.C., and Lee, K.H. (2022). Biopharmaceutical Manufacturing: Historical Perspectives and Future Directions. *Annu. Rev. Chem. Biomol. Eng.* 13, 141–165. <https://doi.org/10.1146/annurev-chembioeng-092220-125832>.
 21. Vandenberghe, L.H., Xiao, R., Lock, M., Lin, J., 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, 1251–1257. <https://doi.org/10.1089/hum.2010.107>.
 22. 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, 1006–1015. Article. <https://doi.org/10.1002/bit.26237>.
 23. Goey, C.H., Bell, D., and Kontoravdi, C. (2018). Mild hypothermic culture conditions affect residual host cell protein composition post-Protein A chromatography. *mAbs* 10, 476–487. <https://doi.org/10.1080/19420862.2018.1433977>.
 24. Shukla, A.A., and Hinckley, P. (2008). Host Cell Protein Clearance During Protein A Chromatography: Development of an Improved Column Wash Step. *Biotechnol. Prog.* 24, 1115–1121. Article. <https://doi.org/10.1002/btpr.50>.
 25. Mietzsch, M., Jose, A., Chipman, P., Bhattacharya, N., Daneshparvar, N., McKenna, R., and Agbandje-McKenna, M. (2021). Completion of the AAV Structural Atlas: Serotype Capsid Structures Reveals Clade-Specific Features. *Viruses* 13, 101. <https://doi.org/10.3390/v13010101>.
 26. Denard, J., Rouillon, J., Leger, T., Garcia, C., Lambert, M.P., Griffith, G., Jenny, C., Camadro, J.M., Garcia, L., and Svinartchouk, F. (2018). AAV-8 and AAV-9 Vectors Cooperate with Serum Proteins Differently Than MV-1 and AAV-6. *Mol. Ther. Methods Clin. Dev.* 10, 291–302. <https://doi.org/10.1016/j.omtm.2018.08.001>.
 27. Kozlowski, L.P. (2016). IPC - Isoelectric Point Calculator. *Biol. Direct* 11, 55. <https://doi.org/10.1186/s13062-016-0159-9>.
 28. Hohfeld, J., Minami, Y., and Hartl, F.U. (1995). HIP, A NOVEL COCHAPERONE INVOLVED IN THE EUKARYOTIC HSC70/HSP40 REACTION CYCLE. *Cell* 83, 589–598.
 29. Young, J.C. (2010). Mechanisms of the Hsp70 chaperone system. *Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire* 88, 291–300. <https://doi.org/10.1139/o09-175>.
 30. Chen, B., Piel, W.H., Gui, L., Bruford, E., and Monteiro, A. (2005). The HSP90 family of genes in the human genome: Insights into their divergence and evolution. *Genomics* 86, 627–637. <https://doi.org/10.1016/j.ygeno.2005.08.012>.
 31. Bevington, J.M., Needham, P.G., Verrill, K.C., Collaco, R.F., Basur, V., and Trempe, J.P. (2007). Adeno-associated virus interactions with B23/Nucleophosmin: Identification of sub-nucleolar virion regions. *Virology* 357, 102–113. <https://doi.org/10.1016/j.virol.2006.07.050>.
 32. 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, 904–912. Article. <https://doi.org/10.1002/bit.25158>.
 33. Sisodiya, V.N., Lequieu, J., Rodriguez, M., McDonald, P., and Lazzareschi, K.P. (2012). Studying host cell protein interactions with monoclonal antibodies using high throughput protein A chromatography. *Biotechnol. J.* 7, 1233–1241. <https://doi.org/10.1002/biot.201100479>.
 34. Tarrant, R.D.R., Velez-Suberbie, M.L., Tait, A.S., Smales, C.M., and Bracewell, D.G. (2012). Host cell protein adsorption characteristics during protein A chromatography. *Biotechnol. Prog.* 28, 1037–1044. <https://doi.org/10.1002/btpr.1581>.
 35. Gagnon, P., Nian, R., Lee, J., Tan, L., Latiff, S.M.A., Lim, C.L., Chuah, C., Bi, X., Yang, Y., Zhang, W., and Gan, H.T. (2014). Nonspecific interactions of chromatin with immunoglobulin G and protein A, and their impact on purification performance. *J. Chromatogr. A* 1340, 68–78. <https://doi.org/10.1016/j.chroma.2014.03.010>.
 36. Oh, Y.H., Becker, M.L., Mendola, K.M., Choe, L.H., Min, L., Lee, K.H., Yigzaw, Y., Seay, A., Bill, J., Li, X., et al. (2023). Characterization and implications of host-cell protein aggregates in biopharmaceutical processing. *Biotechnol. Bioeng.* 120, 1068–1080. <https://doi.org/10.1002/bit.28325>.
 37. Panikulam, S., Hanke, A., Kroener, F., Karle, A., Anderka, O., Villiger, T.K., and Lebesgue, N. (2024). Host cell protein networks as a novel co-elution mechanism during protein A chromatography. *Biotechnol. Bioeng.* 121, 1716–1728. <https://doi.org/10.1002/bit.28678>.
 38. Yang, B., Li, W., Zhao, H., Wang, A., Lei, Y., Xie, Q., and Xiong, S. (2019). Discovery and characterization of CHO host cell protease-induced fragmentation of a recombinant monoclonal antibody during production process development. *J. Chromatogr., B: Anal. Technol. Biomed. Life Sci.* 1112, 1–10. <https://doi.org/10.1016/j.jchromb.2019.02.020>.
 39. Jones, M., Palackal, N., Wang, F., Gaza-Bulsecu, G., Hurkmans, K., Zhao, Y., Chitikila, C., Clavier, S., Liu, S., Menesale, E., et al. (2021). "High-risk" host cell proteins (HCPs): A multi-company collaborative view. *Biotechnol. Bioeng.* 118, 2870–2885. <https://doi.org/10.1002/bit.27808>.
 40. 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. <https://doi.org/10.1016/j.copbio.2021.06.026>.
 41. Hu, Y., Hu, M., Ye, X., Wu, Z., Kang, J., Wong, C., Palackal, N., Qiu, H., and Li, N. (2024). A simple and sensitive differential digestion method to analyze adeno-associated virus residual host cell proteins by LC-MS. *J. Pharm. Biomed. Anal.* 242, 116009. <https://doi.org/10.1016/j.jpba.2024.116009>.
 42. Smith, J., Strasser, L., Guapo, F., Milian, S.G., Snyder, R.O., 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. <https://doi.org/10.1016/j.ejpb.2023.06.019>.
 43. Gray, J.T., and Zolotukhin, S. (2011). Design and Construction of Functional AAV Vectors. *Methods Mol. Biol.* 807, 25–46. https://doi.org/10.1007/978-1-61779-370-7_2.
 44. Leibiger, T.M., Remmler, L.A., Green, E.A., and Lee, K.H. (2024). Biolayer interferometry for adeno-associated virus capsid titer measurement and applications to upstream and downstream process development. *Mol. Ther.* 32, 101306. <https://doi.org/10.1016/j.omtm.2024.101306>.
 45. Zhang, X., Jin, X., Liu, L., Zhang, Z., Koza, S., Yu, Y.Q., and Chen, W. (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. <https://doi.org/10.1089/hum.2021.046>.
 46. 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, 2221–2238. <https://doi.org/10.1002/bit.28128>.
 47. Valente, K.N., Lenhoff, A.M., and Lee, K.H. (2015). Expression of Difficult-to-Remove Host Cell Protein Impurities During Extended Chinese Hamster Ovary Cell Culture and Their Impact on Continuous Bioprocessing. *Biotechnol. Bioeng.* 112, 1232–1242. <https://doi.org/10.1002/bit.25515>.
 48. Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (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, O111.016717. <https://doi.org/10.1074/mcp.O111.016717>.
 49. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., and Aebersold, R. (2018). Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* 14, e8126. Review. <https://doi.org/10.15252/msb.20178126>.

50. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. <https://doi.org/10.1093/bioinformatics/btq054>.
51. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.Y., Hengartner, M.O., and Aebersold, R. (2011). mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* 8, 430–435. <https://doi.org/10.1038/nmeth.1584>.
52. Choi, M., Chang, C.Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014). MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524–2526. <https://doi.org/10.1093/bioinformatics/btu305>.
53. Doncheva, N.T., Morris, J.H., Gorodkin, J., and Jensen, L.J. (2019). Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* 18, 623–632. <https://doi.org/10.1021/acs.jproteome.8b00702>.