Molecular Therapy
Methods & Clinical Development
Original Article

Development of LC-MS methods for AAV capsid protein quantification and host cell protein profiling

Theodoros Kontogiannis,[1,2] Christopher McElroy,[2] Milena Quaglia,[3] Carole Foy,[2] Alexandra S. Whale,[2] Julian Braybrook,[2] and C. Mark Smales[1,4]

[1]School of Natural Sciences, University of Kent, Canterbury, Kent CT2 7NJ, UK; [2]National Measurement Laboratory at the Laboratory of the Government Chemist (LGC), Guildford, Surrey GU2 7XY, UK; [3]Reading Scientific Services, Reading Science Centre, Whiteknights Campus, Pepper Lane, Reading, Berkshire RG6 6LA, UK; [4]National Institute for Bioprocessing Research and Training (NIBRT), Foster Avenue, Mount Merrion, Blackrock, A94 X099 County Dublin, Ireland

Accurate quantification and characterization of recombinant adeno-associated virus (rAAV) capsid proteins are critical for evaluating product quality and safety, ensuring batch consistency, and informing process development of their manufacture. The capsid consists of three proteins derived from the same gene, and while the mean capsid stoichiometry is nominally 1:1:10 (VP1:VP2:VP3), capsids with different stoichiometries exist. Recent studies show that variations in the capsid stoichiometry can impact vector infectivity. Here, a mass spectrometry (MS)-based method was developed to quantify VP1, VP2, and VP3 in rAAV9 capsids and determine stoichiometry. Additionally, the methodology delivers precise measurement of total capsid content and provides a greater depth of information than traditional ELISA capsid titer measurements. The method could be further refined as a reference method to standardize measurements and assign values to reference materials. Host cell proteins consistent with other findings reported in the literature were also identified and reported. The consistent detection of these host cell proteins across different studies highlights their potential relevance to gene therapy products and the importance of their monitoring. Our report exhibits the utility of MS for precise rAAV characterization and presents the first approach to using MS for the standardized measurement of rAAV across different drug products.

INTRODUCTION

Recombinant adeno-associated virus (rAAV) vectors are used extensively in gene therapy, but their intricate composition makes characterization challenging.[1,2] AAVs are non-enveloped, icosahedral viruses that contain a single-stranded DNA genome enclosed within a protein capsid. AAVs can demonstrate heterogeneity between and within production batches when manufacturing rAAVs, such as differences in capsid stoichiometry, post-translational modifications (PTMs), and the presence of host cell proteins (HCPs) that remain after purification.[3,4] The rAAV protein capsid consists of three different proteins, and the accurate quantification of these capsid proteins, VP1, VP2, and VP3, is critical for evaluating product quality and safety as well as ensuring batch consistency. Therefore, work toward the establishment of reference methods for the quantification of the VP1, VP2, and VP3 capsid proteins is an important step required for standardized capsid characterization, improved measurement consistency, and to ensure the safety of rAAV-based gene therapies.[4–6]

It is generally considered that an individual rAAV capsid consists of a collective 60 copies of the 3 viral proteins (VP1, VP2, VP3) in a nominal ratio of approximately 1:1:10, regardless of the serotype.[7,8] However, recent studies revealed that this stoichiometry is not fixed but rather depends on the relative expression levels of each VP protein.[9,10] Understanding capsid stoichiometry is crucial, as variations in VP ratios can affect the infectivity of the viral vectors, potentially resulting in less effective or even ineffective therapeutic batches.[11–13]

Traditionally, VP stoichiometry has been assessed using electrophoresis techniques such as sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). While SDS-PAGE can provide estimates of VP stoichiometry and serotype identity, it is limited by its low resolution and lack of quantitative precision or accuracy.[14,15] Methods such as capillary electrophoresis and liquid chromatography (LC) coupled with UV or fluorescence detection are increasingly used in the industry to enhance accuracy and precision.[16,17]

Mass spectrometry (MS) is a well-established, highly accurate analytical technique for confirming protein sequence and is capable of identifying and locating PTMs.[18–22] Additionally, the growing use of MS-based methods for identifying HCPs in biopharmaceuticals has deepened our understanding of manufacturing processes and established workflows, offering the opportunity to apply MS for rAAV

Received 13 May 2025; accepted 11 August 2025;
https://doi.org/10.1016/j.omtm.2025.101562.

Correspondence: Theodoros Kontogiannis, School of Natural Sciences, University of Kent, Canterbury, Kent CT2 7NJ, UK.
E-mail: tk437@kent.ac.uk

characterization as well.[19,23–25] Moreover, MS is widely recognized as a highly accurate and selective method for protein quantification. Its precision and reliability have established it as one of the primary tools for developing reference methods and materials for biologics.

Here, we describe a tailored MS-based approach for accurate quantification of rAAV serotype 9 (rAAV9) capsid proteins and for determining overall capsid stoichiometry. The method has the potential to be developed further as a reference method to standardize measurements from field methods and characterize reference materials. Data were traceable to the International System of Units (SI) via amino acid analysis of the peptides utilized as standards. As an essential part of the development of the method, the amino acid sequence of the rAAV9 and the presence of potential PTMs were first confirmed by both intact mass analysis (LC-MS) and peptide mapping (tandem mass spectrometry [LC-MS/MS]. Moreover, an LC-MS/MS-based approach was also used to identify HCPs in an rAAV8 sample.

Our findings reveal that the measured stoichiometry of the rAAV9 sample deviates from the widely assumed 1:1:10 ratio of VP1:VP2:VP3. The MS methods described here provide a robust and detailed framework for analyzing rAAV vectors, offering enhanced accuracy and insight into their stoichiometry. These findings are particularly significant, as variations in stoichiometry can directly impact the infectivity of rAAV vectors, potentially compromising the efficacy of the viral preparations.[11–13] Identifying and controlling these variations is therefore important for ensuring the reliability and effectiveness of rAAV-based therapies.

RESULTS

Measurement of the intact masses of the rAAV capsid proteins with LC-MS for serotype identification

Initially, intact mass analysis of the rAAV VP proteins was undertaken. The theoretical and observed masses by this analysis are reported in Table S1, while the deconvoluted spectra and the detection of the peaks corresponding to the capsid proteins of rAAV8 and rAAV9 are shown in Figure 1. In the intact mass analysis of rAAV9 samples, three masses were identified at 81,290 Da, 66,210 Da, and 59,730 Da. The 81,290-Da mass corresponds to the VP1 protein of rAAV9, encompassing amino acids 2–736, with an acetylated N-terminal alanine and deletion of the initial methionine. The 66,210-Da mass corresponds to the VP2 protein, spanning residues 139–736, with the first threonine removed. Lastly, the 59,730-Da mass is associated with the VP3 protein, covering amino acids 204–736, with acetylation of the N-terminal alanine and removal of both the first threonine and methionine residues. These modifications were previously reported in the literature and confirmed by peptide mapping in the next section.[25,26]

Similarly, in the intact mass analysis of rAAV8, three distinct masses were observed at 81,670 Da, 66,520 Da, and 59,800 Da. The 81,670-Da mass corresponds to the VP1 protein of rAAV8, covering the amino acid sequence from 2–738, with an acetylated N-terminal alanine and the removal of the initial methionine. The 66,520-Da mass aligns with the VP2 protein, spanning the sequence from amino acids 139–738, with the first threonine removed. The 59,800-Da mass corresponds to the VP3 protein, covering amino acids 205–738, with acetylation of the N-terminal alanine and the removal of both the first threonine and methionine residues.

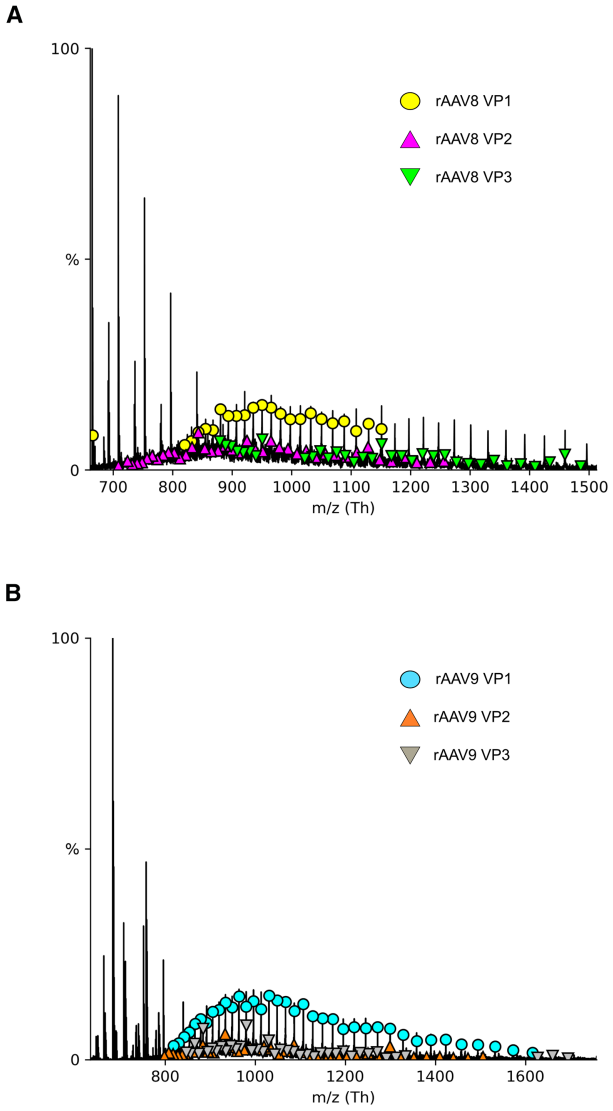
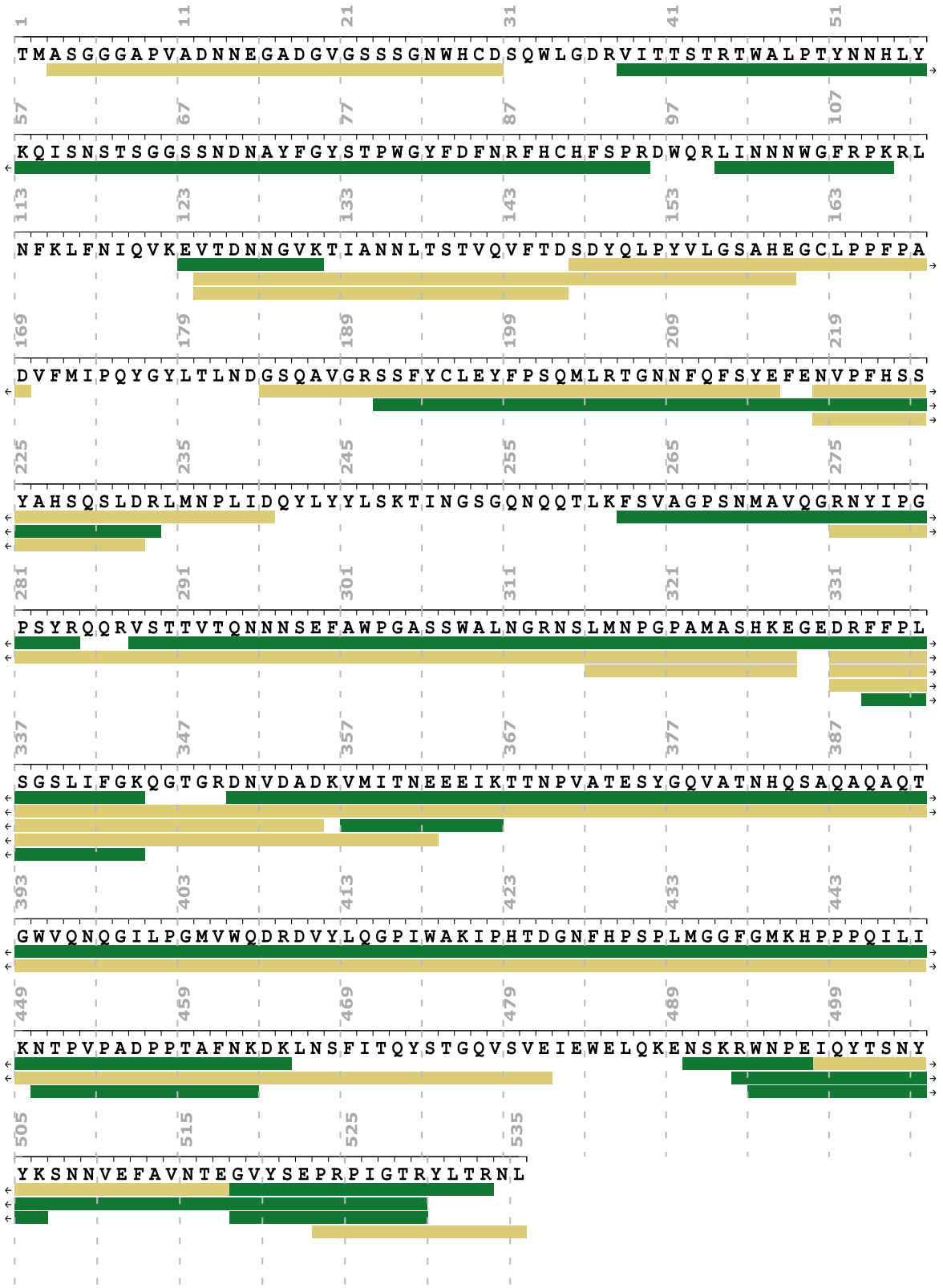


Figure 1. Mass spectrometry analysis and identification of the capsid proteins VP1, VP2, and VP3 in rAAV8 and rAAV9 samples

The deconvoluted spectra, generated using UniDec, are presented in (A) for rAAV8 and (B) for rAAV9. The spectra show distinct peaks corresponding to the capsid proteins VP1, VP2, and VP3. For rAAV8 (A), VP1 is represented by green circles, VP2 by pink triangles, and VP3 by yellow inverted triangles. For rAAV9 (B), VP1 is indicated by light blue circles, VP2 by orange triangles, and VP3 by gray inverted triangles. The *m/z* values of the peaks are allocated to the respective capsid proteins, with VP3 being the most abundant species in both rAAV8 and rAAV9 spectra. The image was created with UniDec.



(legend on next page)

MS peptide mapping and sequence coverage analysis of the rAAV VP proteins

After intact analysis, peptide mapping and MS/MS analysis of the rAAV VP proteins were conducted. In undertaking such analyses, achieving high sequence coverage is crucial for confirming the sequence of the VP protein and serotype and detecting the presence of any mutations, truncations, and other sequence modifications. Digestions with Glu-C and trypsin were conducted as separate experiments on rAAV9 and rAAV8 samples, and the sequence coverage was estimated based on peptides identified from both digests.[27,28]

When considering the specific rAAV8 capsid proteins, sequence coverages of 78.2%, 76.7%, and 81.9% were achieved for VP1 (Figure S1), VP2 (Figure S2), and VP3 (Figure S3), respectively. For the rAAV9, sequence coverages of 77.7%, 81.5%, and 88.6% were achieved for VP1 (Figure S4), VP2 (Figure S5), and VP3 (Figure 2), respectively. The amino acid sequences of the VPs of rAAV8 and rAAV9 determined from these analyses corroborate the results from the intact MS analysis, both revealing the deletion of the initial methionine in VP1 and acetylation of the N-terminal alanine. Similarly, in VP3, the first two residues were deleted, and the N-terminal alanine was acetylated in agreement with others' findings.[26]

Detection of PTMs by MS analysis: Deamidation and oxidation

Next, the mapping of PTMs and their location in the VP proteins was conducted. In particular, the mapping of deamidations and oxidations in AAV capsid proteins as critical quality attributes is important for assessing and ensuring protein stability, which directly impacts vector performance and consistency during manufacturing.[30] Furthermore, when selecting peptides for quantification purposes, it is important to select peptides that are not prone to deamidation and oxidation as these modifications alter the peptide mass, potentially complicating quantification and reducing the utility of a peptide as a standardized, quantitative marker. By avoiding peptides prone to these modifications, a more accurate and reliable quantification is ensured over the duration of the study.

A list of the deamidations detected in rAAV8 and rAAV9 is presented in Table 1. The corresponding b and y ion spectra for the identification of deamidations in rAAV9 and rAAV8 are shown in Figures S6 and S7, respectively. The ability to detect a 1-Da mass difference, indicative of deamidation, is necessary for such studies, and as such, the high-resolution capability of the Orbitrap MS was considered ideal. Interestingly, the deamidations observed in the rAAV8 samples reported here at N57, N514, and N540 have previously been reported by Giles et al., suggesting that these residues are inherently prone to deamidation, likely due to the accessibility, specific amino acid sequence, and three-dimensional protein structure surrounding these three asparagine residues.[30] When rAAV9 capsid VPs were analyzed, deamidation was observed at N254, N303 or N304, and N512. The N512 deamidation site in rAAV9 has also been reported previously.[31] A comparison of the deamidation sites between the rAAV8 and rAAV9 samples is reported in Table 1. Importantly, no oxidations were detected in either of the rAAV8 or rAAV9 samples.

Identification of peptides from each capsid protein suitable for determination of rAAV9 VP capsid stoichiometry

To quantify the VP1, VP2, and VP3 capsid proteins of rAAV9, the protein sequences (UniProt: Q6JC40) were retrieved from UniProt. Peptides unique to each capsid protein were then selected *in silico* using Skyline (University of Washington) and PeptideCutter (ExPASy) (Swiss Institute of Bioinformatics) with Glu-C as the specified endoproteinase.[32,33] In the workflow, Glu-C was selected because it produces these unique peptides after digesting the capsid proteins, whereas other proteases do not generate unique peptides. These unique peptides act as signature peptides for the quantification of the entire capsid protein since they are present only in the capsid protein they derive from. Ideally, multiple peptides would be used for the quantification of each capsid protein. However, this is challenging for the rAAV capsid proteins as they share a significant part of their sequence with one another (Figure 3).[34] Peptides that are unique to VP1 can be found in the N-terminal region of VP1, while peptides unique to VP2 and VP3 are found at the N-terminal regions of VP2 and VP3, respectively. For example, the VP3 unique N-terminal peptide only belongs to VP3 as its generation by Glu-C is not possible from VP1 or VP2 due to the location of Glu-C cleavage sites, which would result in a different peptide sequence (Figure 3).

A peptide located at the C-terminal of the capsid proteins (present in all three capsid proteins) was also selected to quantify the total capsid content (Figure 3). This peptide is referred to as VPc. Among the peptides of the capsid proteins of rAAV9 identified previously, the unique peptides for VP1 (referred to as VP1u), VP3 (referred to as VP3u), and the common capsid peptide (referred to as VPc) were successfully detected and selected for the quantification method. This nomenclature is used throughout the paper to clearly distinguish between protein-specific and shared peptides. The VP1u, VP3u, and VPc endogenous peptides were identified based on their b- and y-ion spectra (Figure S8) after LC-MS/MS analysis of the digested rAAV9. However, the unique peptide for VP2 was not detected. Various approaches were explored, including the use of different columns, modifications to the chromatography conditions, and use of different Glu-C digest conditions and alternative proteases such as AspN and Arg-C instead of Glu-C, but none were successful in identifying VP2u. Since VP2u was not identified during the study, an alternative approach was developed to estimate the

Figure 2. Protein sequence coverage map of rAAV9 VP3 protein

Peptide map showing sequence coverage of VP3 (88.6%) after digestion with Glu-C and trypsin and subsequent MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C digestion in yellow. The image was created using the Peptide Stack Visualization tool (Aarhus University Visualization Group).[29]

Table 1. Deamidation sites detected in trypsin-generated peptides from the capsid proteins of rAAV8 and rAAV9 samples

Serotype and deamidated sequences (trypsin digest)	Amino acid position
rAAV8	
YLGPFnGLDK	N57
YHLnGR	N514
FFPSnGILIFGK	N540
rAAV9	
TWALPTYNnHLYK	N254
LINnNWGFRPK	N303 or N304
VSTTVTQNNNSEFAWPGASSWALnGR	N512

All deamidations observed in both rAAV8 and rAAV9 were partial, as the corresponding non-deamidated peptides were also detected. An exception was observed at N57, where the non-deamidated form was not detected. n corresponds to deamidated asparagine (N) residue.

stoichiometry of the capsid proteins. By quantifying VPc, the total capsid protein content was determined. The quantity of VP2 was then estimated by subtracting the measured amounts of VP1u and VP3u from the total VPc value. This method enabled the estimation of the stoichiometry of VP1, VP2, and VP3 in the rAAV9 capsid. These signature peptides were used for the measurement of the rAAV9 stoichiometry.

In the quantification workflow, the rAAV9 capsid proteins were first subjected to reduction, alkylation, and proteolytic digestion using Glu-C to generate the unique peptides of interest: VP1u, VP3u, and VPc. Quantification of these peptides was performed using stable isotope dilution (SID). A known amount of isotopically labeled peptide mix containing standards for VP1u, VP3u, and VPc was spiked into the digested rAAV9 sample. Standard curves were prepared by mixing varying concentrations of the corresponding synthesized natural peptides with a constant amount of the isotopically labeled peptide mix (same as the amount spiked into the sample). The unknown concentrations of the endogenous rAAV9 peptides were determined by interpolating the measured peak area ratios of the quantifier transitions (endogenous peptides to their isotopically labeled counterparts) into the corresponding standard curve. Figure 4 illustrates the different steps of the rAAV9 capsid protein quantification workflow.

To ensure assay specificity, the selected peptide sequences were screened against the human proteome using the NCBI Protein BLAST tool to confirm that they do not match any potential HCPs.[34–36] To further validate the specificity of the method, an HEK293 cell lysate (Novus Biologicals, catalog no. NBP2-25048) was digested with Glu-C. The digest was then analyzed using the LC-MS/MS multiple reaction monitoring (MRM) method, which monitors the transitions corresponding to the VP1u, VP3u, and VPc peptides. No corresponding signals above the background level were detected, confirming that the assays are specific to the target peptides (Figure S9).

To enhance the sensitivity and the signal-to-noise ratio of the quantification method, the collision energy (CE) was individually optimized for each peptide (Figure S10).[32] Based on this analysis (Figure S10), the CEs that yielded the highest intensity for each transition were selected. Furthermore, digestion efficiency and the release of the VP1u, VP3u, and VPc signature peptides from the capsid were verified by monitoring their release over time and confirming that peptide levels reached a plateau, suggesting complete recovery of each signature peptide from the digest (Figure S11).

Amino acid analysis and purity assessment of natural peptides

Amino acid analysis was initially undertaken to determine the exact concentration of the synthesized natural peptides used for the generation of the standard curve (Figure S12).[37,38] The measured concentrations (mg/mL) of each peptide stock are shown in Table S2, while the composition of controls, sample blends, calibration blends, and blanks used in the analysis are shown in Table S3. Additionally, the selected transitions for the MRM experiment for each amino acid targeted in the analysis are shown in Table S4. The amount of each amino acid released from the peptide of interest was determined using Equation S1.

To ensure accurate quantification, it was also important to consider the purity of the synthesized natural peptides, as peptide synthesis can produce small fragments that could affect the analysis.[34,35] The purity of each synthesized natural peptide (VP1u, VP3u, and VPc) was assessed using intact MS with an Orbitrap Q-Exactive Plus instrument (Figure S13). It is important to note that the measured purity of the peptides, especially that of VP3u, was substantially different from what was ordered (95% purity) (Table S5).

MS quantification of the VP1u, VP3u, and VPc peptides from rAAV9 samples

For quantitative analysis, experiments were performed in triplicate, with each experimental replicate involving independently generated standard curves. The standard curves were plotted as the mean ratio of peak areas of the quantifier transition (three injections per standard curve) of each synthesized natural peptide to the isotopically labeled counterpart against the known concentration (nM) of the synthesized natural peptide (Figure S14).

The mean quantities of VP1u, VP2, VP3u, and VPc across the three experiments in the stock, along with their uncertainties, are reported in Table 2. The measured values in the stock were VP1u at 69.7 (36.7) nM, VP2 at 385.3 (12.8) nM, VP3u at 512.9 (117.9) nM, and VPc at 967.9 (135.7) nM. The mean stoichiometry of the rAAV9 sample calculated from the three experiments was 1:5.5:7.3 (VP1:VP2:VP3), which deviates from the widely assumed ratio of 1:1:10. The inter-experimental variability (coefficient of variation [CV]) for VP2 was calculated based on the inter-experimental standard deviation (SD) and the mean concentration of VP2. The inter-experimental SD for VP2 was calculated using the error propagation formula and considering the covariance between peptides using Equations S2 and S3, and it was only 3.3%. The combined variability

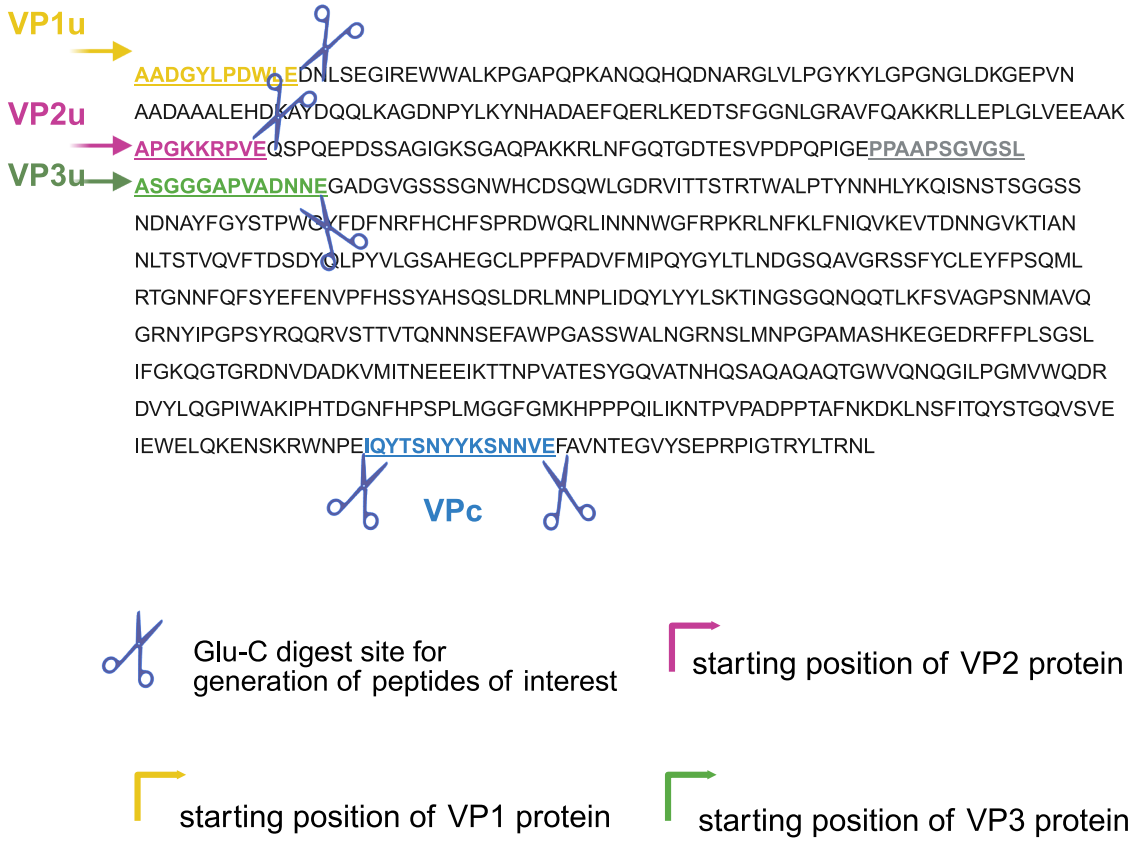


Figure 3. Schematic representation of unique peptide generation through Glu-C digestion

The sequence of the rAAV9 capsid protein is shown with yellow arrows indicating the starting position of VP1, pink arrows marking the starting position of VP2, and green arrows marking the starting position of VP3. Scissors denote Glu-C digestion sites, resulting in unique peptides for VP1u (yellow), VP2u (pink), and VP3u (green), as well as a common peptide VPc (shown in blue). The VP3u peptide shown in green is unique to VP3. In VP1 or VP2, Glu-C digestion at the same region generates longer peptides that include additional upstream amino acids (highlighted in gray), distinguishing them from the VP3-derived peptide. Created in BioRender (https://BioRender.com/uxi92qn).

for VP1u, VP3u, and VPc was calculated by including the CV from three primary sources: the amino acid analysis, the purity assessment of each synthesized natural peptide, and the variability observed across the three experiments using Equation S4. Among these, the variability across the three experiments is by far the largest contributor (99.9% of total variability for VP1u, 99.5% of total variability for VP3u, and 99.9% of total variability for VPc). VP3u and VPc demonstrated good agreement between experiments, with a CV of 23.0% and 14.0%, respectively. In contrast, VP1u exhibited higher variability, with a CV of 52.6%, as expected, given the lower quantity in the sample. Interestingly, the capsid stoichiometry estimated using the MS-based method was not consistent with that determined by SDS-PAGE densitometry (1:0.7:3.4) (Figure S15), potentially due to factors outlined in the discussion section.

The method was also used to measure the total capsid content of the sample. The VPc peptide is present in all three rAAV9 capsid proteins and reflects the total capsid concentration. Assuming there are 60 copies of the VP proteins in a capsid, the MS-based concentration of the rAAV9 sample was calculated as 9.71×10^9 (1.36×10^9) capsids/μL. Enzyme-linked immunosorbent assay (ELISA)-based analysis of the rAAV9 sample gave a capsid concentration of 1.6×10^{10} (1.04×10^9) capsids/μL (Figure S16). This corresponds to a 39.3% (6.2) relative difference between the mean values of ELISA and the MS-based approach described here.

Identification of host cell proteins in the rAAV8 samples by MS analysis

The rAAV8 and rAAV9 samples were analyzed for the presence of HCPs (human proteins from HEK293 production cells). One hundred HCPs were identified in the trypsin-digested rAAV8 sample (Table S6), some of which have been previously reported by others.[23,39]

These HCPs are involved primarily in DNA packaging, gene regulation, chromatin structure, translation, signal recognition, RNA processing, and protein folding. All identified proteins, the number of unique peptides matched, −10logP scores, and coverage (%) are reported in Table S6. Twenty-four of the HCPs identified in this study have been previously reported in the literature (Table 3),

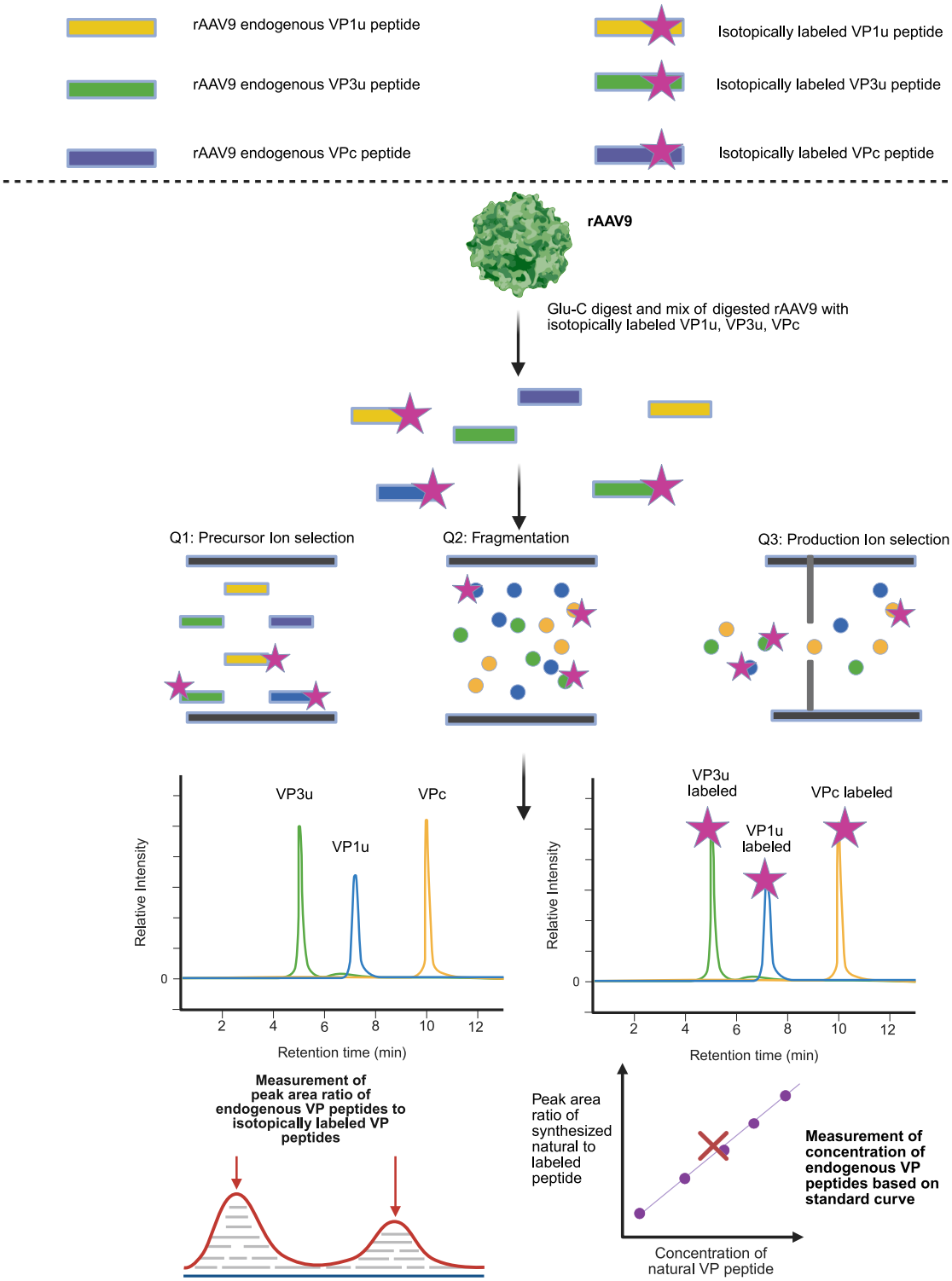


Figure 4. Multi-step workflow for quantification of VP1u, VP3u, and VPc

rAAV9 particles are denatured and digested with Glu-C to generate peptides corresponding to VP1u, VP3u, and VPc. These peptides are mixed with isotopically labeled standards for VP1u, VP3u, and VPc and analyzed using an MRM method. In Q1, precursor ions are selected based on their *m/z*; in Q2, the ions are fragmented; and in Q3,

(legend continued on next page)

highlighting their frequent occurrence in rAAV preparations and suggesting potential roles in rAAV biology or production processes. For the rAAV9 sample, no HCPs were detected, likely due to a more rigorous purification process. As shown in Figure S17, the chromatogram of the rAAV9 sample is much "cleaner" compared to that of the rAAV8 sample, where multiple peaks are observed, corresponding to the identified HCPs. Furthermore, the presence of multiple impurities in the rAAV8 preparation is demonstrated by SDS-PAGE (Figure S18).

DISCUSSION

In this study, we present the development of an MS-based method for the precise quantification of VP1, VP2, and VP3 proteins, as well as the total protein quantity of rAAV9 and the measurement of the capsid stoichiometry. Currently, there is no standardized reference measurement method for quantifying rAAV capsid proteins. However, by implementing MS-based approaches in accordance with International Organization for Standardization 17511:2020 guidelines, we have developed a method that ensures precision and higher-order reference measurement for capsid protein quantification.[42] To support the accuracy of this method, well-defined rAAV standards suitable for absolute quantification are necessary.

In our approach, the quantities of capsid proteins are measured by quantifying their signature peptides in SI units, which ensures traceability and comparability across different laboratories. Furthermore, by measuring the concentration of the synthesized natural peptides through amino acid analysis, the method ensures traceability to known standards. Moreover, the use of isotopically labeled peptides enables precise quantification by accounting for variability in sample preparation, differences in ionization efficiency, and differential recovery of the signature peptides during LC, thereby enhancing the reproducibility and robustness of the method. Although the signature peptides VP1u, VP3u, and VPc may have different physicochemical properties, the use of isotopically labeled counterparts introduced into the same matrix compensates for these differences. Quantification is based on the peak area ratio between the endogenous and labeled peptides, so any variation in LC recovery or ionization efficiency affects both forms equally, ensuring that the ratio remains consistent and reliable.

Active participation in inter-laboratory studies that compare values assigned to viral vector reference materials, alongside adherence to established written standards, is crucial for maintaining measurement accuracy over time. Moreover, the experimental variability of this method suggests that it is suitable for identifying differences in VP1, VP2, and VP3 quantities and overall capsid content between rAAV9 production batches, provided these differences are greater than the variability between experiments. These findings underscore the need for method standardization to enable reliable comparative analyses, particularly when detecting subtle variations between production batches.

A limitation of the SID-MS method is that it quantifies each capsid protein regardless of whether it is incorporated into an assembled capsid or present as a free-floating protein. As a result, the measured stoichiometry reflects the overall abundance of VP1, VP2, and VP3 in the preparation. However, since previous studies have shown that capsid stoichiometry is directly influenced by the expression levels of these proteins, it is reasonable to assume that the VP1:VP2:VP3 ratio in the total protein pool closely reflects the ratio observed in the assembled capsids.[9–11] Additionally, this method is specific to rAAV9; the peptides used cannot be applied to quantify other serotypes. rAAV9 was chosen due to its extensive use and well-documented advantages, including broad tissue tropism, efficient transduction of neuronal, non-neuronal, and astrocytic cells, as well as its ability to cross the blood-brain barrier.[43,44] Given these advantages and its extensive application, rAAV9 was chosen as the primary focus for developing a stoichiometry VP protein measurement method, whereby such an approach could then be extended to additional serotypes in the future. The VP1u peptide is present in other serotypes such as AAV1, AAV2, AAV3, AAV6, AAV8, and AAV10, providing a potential foundation for broader application.

The results reveal a measured stoichiometry of 1:5.5:7.3 (VP1:VP2:VP3). It is important to note that this method measures the average stoichiometry of the production batch while acknowledging that individual capsids within the batch may exhibit heterogeneous stoichiometries.[10] The observed ratio of 1:5.5:7.3 deviates from the widely assumed AAV capsid stoichiometry of 1:1:10, showing a substantially higher proportion of VP2. The observed low VP1:VP2 ratio (1:5.5:7.3) may have functional implications. Bosma et al. demonstrated that a low VP1:VP2 ratio (<0.5) is associated with reduced rAAV potency.[11] Moreover, similar deviations from the assumed 1:1:10 ratio have been reported by another study, which observed between 0 and 2 copies of VP1, 8–11 copies of VP2, and 48–51 copies of VP3 in AAV1 capsids.[9] Additionally, although a study by Tiambeng et al. reported VP1:VP2:VP3 ratios close to the expected 1:1:10 in some AAV samples, it also found substantially lower levels of VP1 and VP2 in others, with ratios as low as 1:1:35 and 1:1:49.[45] In the same study, rAAV8 particles produced using two different methods showed capsid protein ratios of 4.9:11.6:43.5 (VP1:VP2:VP3) for process A and 4.1:4.8:51.1 for process B. These findings highlight the impact of production conditions, particularly on VP2 levels, while VP1 and VP3 proportions remained relatively stable, illustrating that there is no standard capsid stoichiometry.[45] Overall, studies have suggested that the stoichiometry of the capsids directly

product ions are selected and then detected. The peak areas of the quantifier transitions for endogenous peptides and their isotopically labeled counterparts are measured, and the ratios are calculated. A standard curve, prepared with the synthesized natural VP1u, VP3u, and VPc peptides of known concentrations mixed with isotopically labeled standards, is used to determine the concentrations of endogenous VP1u, VP3u, and VPc in the rAAV9 sample. Created in BioRender (https://BioRender.com/tenyadv).

Table 2. Measured mean stoichiometry and concentrations of VP1u, VP3u, and VPc peptides in rAAV9 stock

Sample	Mean VP1: VP2:VP3 ratio	VP1u, nM	VP2, nM	VP3u, nM	VPc, nM
rAAV9	1:5.5:7.3	69.7 (36.7)	385.3 (12.8)	512.9 (117.9)	967.9 (135.7)

Data are shown as mean (uncertainty), where uncertainty reflects the propagated variability from three experimental replicates, amino acid analysis and purity assessment of each synthesized natural peptide. For VP2, uncertainty represents inter-experimental variability, including covariance.

depends on the expression levels of each individual capsid protein, which may differ based on the production method used.[9–11,45] Therefore, the low VP1:VP2 ratio observed may be attributed to high expression levels of VP2 in the host cells used for the production of rAAV9.

The discrepancy in capsid stoichiometry measurements between the MS-based method and SDS-PAGE may be attributed to limitations of both approaches. SDS-PAGE is a semi-quantitative technique, and a key limitation is that staining intensity does not necessarily correlate directly with protein abundance. This is due to differences in dye-binding efficiencies and variability in protein-stain interactions.[46] Consequently, differential staining of VP1, VP2, and VP3 may contribute to the observed discrepancy. Moreover, silver staining has a limited linear dynamic range, which limits its accuracy in determining capsid stoichiometry.[47] Thus, silver staining intensity, particularly when bands of very different intensity are present as in the case of VP3 compared to VP1 and VP2, does not correlate well with protein quantity across a large dynamic range.[47] Additionally, the MS-based method has limitations that may bias the measured ratio. For example, VP2 is estimated indirectly by subtracting the quantities of VP1u and VP3u from the total capsid content (VPc). Therefore, since the method also quantifies free capsid proteins, the presence of free VPc peptides in the sample could artificially inflate the calculated VP2 levels and contribute to the discrepancy observed between MS and SDS-PAGE.

In contrast, there was good agreement between our developed MS-based approach for quantifying the total capsid concentration and ELISA. The antibody used in the ELISA assay, ADK9, specifically binds to a conformational epitope on fully assembled rAAV9 capsids.[48,49] This epitope includes the contact residues 456–459, 492, 494–498, 584, and 587–589. The binding of the antibody also occludes residues corresponding to the VPc peptide (701–708, 712), indicating that the binding site of the antibody and the location of the VPc peptide are in close proximity. The observed 39.3% relative difference between the mean values of our MS-based approach and ELISA can be attributed to the fundamental differences between these two methods. Our MS-based method directly quantifies specific ions, providing precise measurements of the signature peptides for each capsid protein. In contrast, ELISA relies on absorbance measurements, which are susceptible to interference from substances absorbing at the same wavelength or non-specific antibody binding. Furthermore, while ELISA predominantly detects fully assembled capsids via specific antibody interactions, the MS-based method quantifies both fully assembled capsids, partially assembled capsids, and any free proteins or peptides, including potential degradation products in the sample. Based on this broader detection range, one might expect the MS-based method to yield higher values than ELISA. However, the observed difference can be attributed to several factors. A key source of variability lies in the large dilution factor (2×10^4) applied to the rAAV9 sample for ELISA to fit within the standard curve. The serial dilutions may introduce errors or reduce precision due to pipetting errors and/or mixing variability.[50,51] These differences underscore the importance of considering method-specific biases and assumptions when comparing results from ELISA and MS-based quantification.

Interestingly, a VP3 variant initiating from a downstream start codon has been reported in previous studies.[14,45] However, this variant was not detected in either the rAAV8 or rAAV9 samples in this study. For rAAV9, this is due to the absence of a suitable downstream start codon required for production of the variant. In contrast, rAAV8 does contain a downstream start codon (M212) that could potentially lead to the production of such variant. However, we found no evidence for the presence of this variant in our rAAV8 sample. This may be due to its low abundance in our sample, possibly below the detection limit of our LC-MS method.[52] Specifically, the number of copies of each capsid protein incorporated into the rAAV particle correlates with their relative expression levels, which can vary considerably depending on the production process.[11,45,52] It is therefore plausible that in our production system, the expression of the VP3 variant was lower than in previous studies, and not detectable with the sensitivity of our current analytical approach.

Finally, our study revealed deamidation sites in both rAAV8 and rAAV9 vectors, which we suggest should be monitored closely to evaluate their impact on vector stability and functionality. Our study identified 100 distinct HCPs in rAAV8 samples, showing the complexity of HCPs in rAAV production. Of these, 24 have been previously reported in the literature, highlighting their relevance. Their presence may be explained by various mechanisms, including interactions with capsid proteins, encapsulation of HCPs within the capsid, association with the packaged genome, or indirect binding via other HCPs that interact directly with components of the rAAV vector.[40,53] Consistent detection of these HCPs across different production batches and serotypes is crucial for understanding their potential impact on gene therapy products, including their influence on product stability, safety, immunogenicity, and therapeutic efficacy. Further analysis of rAAV HCPs will help optimize processes to remove HCPs, especially those that may pose risks to patients.

In summary, we have developed an MS-based method that enhances the precision and accuracy of quantifying rAAV9 capsid proteins and determining the capsid stoichiometry. This approach overcomes

Table 3. Host cell proteins identified in this study that have also been reported in previous studies

Protein name	UniProt accession no.	Coverage, %	−10logP score	Identified in the literature
40S ribosomal protein S3	P23396	77	162	Leibiger et al.[40]
40S ribosomal protein S7	P62081	52	120.3	Leibiger et al.[40]
40S Ribosomal protein S12	P25398	15	68.8	Leibiger et al.[40]
40S Ribosomal protein S18	P62269	46	122.7	Leibiger et al.[40]
40S ribosomal protein S20	P60866	47	110.1	Leibiger et al.[40]
60S ribosomal protein L29	P47914	14	93.5	Rumachik et al.[23]
Serine/arginine-rich splicing factor 1	Q07955	49	149.5	Dong et al.[41]
Splicing factor proline- and glutamine-rich	P23246	18	138.6	Leibiger et al.[40]
Heterogeneous nuclear ribonucleoprotein A1	P09651	40	143	Leibiger et al.[40]
Heterogeneous nuclear ribonucleoprotein A/B	Q99729	24	131.9	Leibiger et al.[40]
Heterogeneous nuclear ribonucleoprotein U	Q00839	20	146.2	Leibiger et al.[40]
Histone H2B type F	Q5QNW6	51	128.8	Leibiger et al.[40]
Histone H4	P62805	50	100.6	Leibiger et al.[40]
Nucleolin	P19338	32	189.5	Leibiger et al.,[40]; Dong et al.[41]
Heat shock cognate protein 71 kDa	P11142	49	185	Rumachik et al.,[23]; Leibiger et al.[40]
Heat shock protein 70 kDa 1A	P0DMV8	54	188.8	Rumachik et al.[23]
Heat shock protein 70 kDa 1B	P0DMV9	54	188.8	Leibiger et al.[40]
Putative elongation factor 1-alpha-like 3	Q5VTE0	51	179.8	Rumachik et al.[23]
Elongation factor 1-alpha 1	P68104	51	179.8	Leibiger et al.[40]
Chromobox protein homolog 1	P83916	54	158	Leibiger et al.[40]
Chromobox protein homolog 3	Q13185	65	154.8	Leibiger et al.[40]
Polyadenylate-binding protein 1	P11940	19	146.6	Leibiger et al.[40]
Y-box-binding protein 1	P67809	48	161.5	Leibiger et al.[40]
Ubiquitin 40S ribosomal protein S27a	P62979	20	88.9	Leibiger et al.[40]

the limitations of traditional methods such as SDS-PAGE and ELISA, providing a more detailed and reliable analysis of rAAV vectors. Our method shows potential as a reference for quantifying VP1, VP2, and VP3. It is capable of detecting changes in VP1, VP2, and VP3 levels between production batches as long as these differences surpass the variability observed between experiments. Future research and inter-laboratory studies are essential to further validate these findings, refine the methodology, and explore its application to other AAV serotypes, ultimately advancing the field of gene therapy and improving product quality and consistency.

MATERIALS AND METHODS

Model rAAV sample origin

An rAAV9 sample suitable for pre-clinical studies was commercially produced and purchased from VectorBuilder by Professor Giovanna Mallucci (Altos Labs, Cambridge Institute of Science) and was generously gifted to us for this study. rAAV8 was a generous gift from Professor Paul Dalby (Department of Biochemical Engineering, University College London). The production and purification of the rAAV8 sample was carried out by Dr. Yiwen Li (Department of Biochemical Engineering, University College London). The rAAV8 sample was produced by transfecting HEK293T cells with pAAV2/8 (Addgene plasmid 112864, which expresses the Rep and capsid proteins), the pAAV-CAG-GFP (Addgene plasmid 37825, which carries the EGFP genome), and the pAdDelta F6 (Addgene plasmid 112867, which encodes for the helper virus proteins). The subsequent rAAV8 particles were harvested from the cells and purified by affinity chromatography using the commercial POROS GoPure AAVX pre-packed column (0.5 × 5 cm, 1 mL) (Thermo Fisher Scientific, catalog no. A36652). The rAAV8 particles were eluted in 100 mM citric acid (pH 3) and neutralized in Tris buffer (pH 8) for a final pH of 7.

LC-MS analysis of intact VPs

For analysis, 5 μL rAAV9 (0.5 μg) or 5 μL rAAV8 (0.26 μg) was injected into an Avantor ACE 3 C4-300 column (100 × 2.1 mm inner diameter) (Avantor, catalog no. ACE-213-1002) maintained at 50°C. The mobile phase consisted of 0.5% formic acid (Fisher Scientific, catalog no. 10785711) in ultra-pure water with a resistance of 18.2 MΩ · cm (Elga Purelab Ultra) (solvent A) and acetonitrile (Romil, catalog no. H050) (solvent B). The flow rate was set at 0.25 mL/min, with a curve setting of 5. The gradient started from 5% solvent B for 1 min, which was then increased in a linear gradient to 60% solvent B at 61 min, 95% B at 62 min, and was maintained at 95% B up to

70 min. Subsequently, at 71 min, the gradient was returned to 5% solvent B, where it remained until the end of the analysis at 90 min.

LC-MS analysis was performed using a Vanquish UHPLC System (Thermo Fisher Scientific) and the Q-Exactive Plus Orbitrap instrument equipped with an HESI-II probe source (Thermo Fisher Scientific) in positive electrospray ionization (ESI) mode. The settings were as follows: a source temperature of 320°C, spray voltage of 3.5 kV, S-lens radiofrequency (RF) level of 100, sheath gas flow rate of 25 a.u., and auxiliary gas flow rate of 5 a.u. The maximum injection time was set to 200 ms, with an automatic gain control (AGC) target value of 3e−6 and 1 microscan. Data were acquired in 70K resolution mode over a mass range of 300–4,000 *m/z*. Intact protein mode was enabled for this experiment.

For the identification of VP1, VP2, VP3 in the rAAV8 sample, the following UniDec (University of Oxford) settings were implemented: *m/z* range: 290–4,040, charge range: 50–100, mass range: 59,700–85,000 Da, sample mass: every 10 Da, peak detection range: 500 Da, and peak detection threshold: 0.01. Similarly, for the identification of VP1, VP2, and VP3 in the rAAV9 sample, the following UniDec settings were implemented: *m/z* range: 290–4,040, charge range: 30–85, mass range: 59,700–82,000 Da, sample mass: every 10 Da, peak detection range: 1,600 Da, peak detection threshold: 0.01.

Preparation and digestion of rAAV8 and rAAV9 samples with Glu-C

A total of 5 μL rAAV9 or 5 μL rAAV8 was mixed with 5 μL of a solution comprising 8 M guanidine hydrochloride (gHCl) and 80 mM dithiothreitol (DTT) in 200 mM ammonium bicarbonate (AMBIC) for denaturation of the capsid proteins.[54] This resulted in a final concentration of 4 M gHCl and 40 mM DTT. The sample was then heated at 80°C for 20 min in a thermoshaker (Microtherm, Camlab), with shaking at 50 rpm. Following heat treatment, the sample was allowed to cool to room temperature for 10 min. Subsequently, 5 μL 45 mM iodoacetamide (IAA) was added, achieving a final IAA concentration of 15 mM. To dilute the gHCl concentration to 0.15 M, 245 μL of 50 mM ammonium acetate was added, followed by 50 μL endoproteinase Glu-C (V8 Protease) (Merck, catalog no. 10791156001) solution from a 1-μg/μL stock of Glu-C. The enzymatic digestion was then carried out at 37°C for 17 h. Post-digestion, the sample volume was reduced to 65 μL using a vacuum centrifuge (Concentrator Plus, Eppendorf) at the V-HV setting at 60°C.

Preparation and digestion of rAAV8 and rAAV9 samples with trypsin

rAAV9 and rAAV8 samples were prepared and denatured following the same procedure as described above. We added 5 μL 80 mM DTT to quench the alkylation reaction, and the gHCl concentration was diluted to 0.5 M by adding 60 μL 10 mM AMBIC. Finally, 10 μL trypsin solution (from a 1-μg/μL stock of trypsin) (Promega, catalog no. V5111) was added, and enzymatic digestion was carried out at 37°C for 17 h.

LC-MS/MS analysis for protein sequencing of VPs and HCP identification

We injected 5 μL of the digested rAAV9 or rAAV8 sample into an Avantor ACE 3 C18-300 (100 × 2.1 mm i.d.) column (Avantor, catalog no. ACE-213-1002) maintained at 40°C. In this experiment, 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) were used as mobile phases. A 20% acetonitrile solution was used as the needle wash, while 70% acetonitrile in ultra-pure water was used as the rear seal wash. The gradient used for analysis started from 3% solvent B with a flow rate of 0.2 mL/min for the first 3 min. After this, there was a linear increase to 45% solvent B at 41 min, followed by an increase to 100% B at 44 min, which was then held until 48 min. Finally, there was a return to 3% solvent B using a linear gradient at 52 min, which was then held to the end of the run at 65 min.

The LC-MS/MS analysis was conducted using the Vanquish UHPLC System and the Q-Exactive Plus Orbitrap instrument with an HESI-II probe source in positive ESI mode with the following settings: a 320°C source temperature, 2.5 kV spray voltage, 50 S-lens RF, 25 a. u. sheath gas flow rate, 10 a.u. auxiliary gas flow rate. For the first step of the analysis (full MS), the following settings were applied: 100-ms maximum injection time and 3e−6 AGC target value. Data were acquired in 70K resolution mode over a range of 200–1,800 *m/z*. For the subsequent data-dependent MS/MS, the following settings were applied: maximum injection time 50 ms AGC target 1e−5; the top 5 ions were analyzed, while data were acquired at a resolution of 17.5K and an isolation window of 1.5 *m/z*.

The data were analyzed with PEAKS version 7.5 (Bioinformatics Solutions) using the following settings: precursor mass: 20 ppm using monoisotopic mass, fragment ion: 0.5 Da, maximum missed cleavages per peptide: 1, maximum allowed variable PTM per peptide: 2; non-specific cleavages were not allowed.

For rAAV8 and rAAV9 capsid protein peptide mapping, all peptides were identified using the PEAKS version 7.5 software (Bioinformatics Solutions) with a false discovery rate (FDR) set at 5% and *de novo* average local confidence score ≥80%, and they were validated through three independent in-solution digest experiments with each enzyme (trypsin and Glu-C), followed by MS/MS analyses. Sequence coverage of the rAAV9 and rAAV8 capsid proteins was achieved using both trypsin and Glu-C enzymes for digestion.

For HCP peptide mapping, an FDR of 0.1% was applied for each peptide, and each protein identification required at least one unique peptide. Additionally, the −10logP score for each identified protein was set to ≥40. HCP identifications were further validated through three independent in-solution digestion experiments. The Human Proteome Database was downloaded from UniProt (UniProt: UP000005640 AND reviewed:true, 20,420 entries, downloaded on January 12, 2023).

Preparation of standard curves and sample quantification

The three peptides (VP1u, VP3u, and VPc; for sequences, see Table S7) were obtained as isotopically labeled peptides (30 nmol

each) and as natural peptides (synthesized) from Thermo Fisher Scientific (ordered as 2 mg each, 95% purity). Initially, each labeled peptide was dissolved in 2 mL 5% acetonitrile in water and subsequently diluted gravimetrically in ultra-pure water. A master mix containing the three isotopically labeled peptides (VP1u, VP3u, and VPc) was then prepared gravimetrically at a nominal ratio of 1:10:11 (VP1: VP3:VPc), corresponding to final concentrations of 5 nM for VP1u, 50 nM for VP3u, and 55 nM for VPc.

The synthesized natural peptide VP1u was dissolved in 2 mL 5% acetonitrile and 0.1% formic acid in water. The synthesized natural peptide VP3u was dissolved in 2 mL 0.1% formic acid in water. The synthesized natural peptide VPc was initially dissolved in 75% DMSO and 0.1% formic acid in water. After the VPc peptide was fully dissolved, it was serially diluted gravimetrically in ultra-pure water to achieve a final DMSO concentration (at the highest concentration level of the standard curve) of 0.0375% to ensure that the presence of DMSO was sufficiently low that it would not influence the accurate quantification of peptides by affecting the tuning or chromatography. The VP1u and VP3u natural peptide stocks were also diluted gravimetrically to reach the appropriate concentrations (nM) for the generation of the calibration curve.

Separate calibration curves were prepared gravimetrically for each signature peptide. Each point of the calibration curves for VP1u, VP3u, and VPc consisted of a 1:1 (volume ratio) mixture of synthesized natural peptide VP1 and a master mix comprising isotopically labeled VP1u, VP3u, and VPc peptides. The concentration of the isotopically labeled peptides remained constant across all points of the calibration curves (5 nM for VP1u, 50 nM for VP3u, and 55 nM for VPc), while the concentration of each synthesized natural VP peptide decreased. Dilutions for each standard curve point were prepared gravimetrically. The resulting standard curves were generated by gravimetrically preparing 1:1 mixtures of specific concentrations of the synthesized natural peptide with the isotopically labeled peptide master mix, ensuring a gradually increasing ratio of peak areas between each natural peptide and the isotopically labeled counterpart.

Finally, the digested rAAV9 sample was also gravimetrically combined with the isotopically labeled peptide master mix in a 1:1 volume ratio. Specifically, the Glu-C-digested rAAV9 sample was first dried completely using the same vacuum centrifuge under the same conditions. The dried sample was then resuspended in 50 μL 0.1% formic acid in water, and 50 μL master mix containing isotopically labeled peptides in 0.1% formic acid in water was added.

Gravimetric preparations

Each sample weighing was carried out using a Mettler Toledo XP205 analytical balance, which was calibrated daily using a certified weight set (serial no. 27252). The set included a 0.1-g weight from Mettler and 0.5-, 1-, 5-, 10-, 20-, 50-, 100-, and 200-g weights from Troemner.

Analysis of digested rAAV9 by triple quadrupole MS

Within each experiment, the standard curve was injected in triplicate. Between each replicate, four blanks consisting of 0.1% formic acid in ultra-pure water were injected. The sample was injected after the blanks, followed by another set of four blanks. In total, nine sample injections were performed per experiment, and the mean peak area ratio for each peptide was calculated across the nine injections.

In general, samples were analyzed using an ACQUITY UPLC M-Class system (Waters Corporation) coupled to a Xevo TQ-XS triple quadrupole mass spectrometer with the MRM method analyzing three transitions for each peptide—one quantifier (transition 1) and two qualifiers (transitions 2 and 3) (Table S8).[55,56]

Separation was performed on an ACQUITY UPLC HSS T3 C18 column (1.8 μm, 1.0 × 150 mm) (Waters Corporation, catalog no. 176001130). The column temperature was maintained at 40°C. The mobile phase A consisted of ultra-pure water containing 0.1% formic acid, while mobile phase B was acetonitrile with 0.1% formic acid.

The LC method started with 2% acetonitrile for the first 2 min. This was followed by a linear increase to 45% acetonitrile from 2 to 10 min, then a further increase to 80% from 10 to 12 min. The elution was continued with 98% acetonitrile from 12 to 12.5 min, maintained at 98% until 13.5 min, returned to 2% acetonitrile from 13.5 to 14 min, and then held at this level until 30 min. Each step was programmed with a curve setting of 6. Needle wash procedures included a weak wash with 5% acetonitrile in water, a strong wash consisting of 50% water/25% acetonitrile/25% isopropanol (v/v), and seal washing with a 1:1 mixture of water and acetonitrile for 5 min/sample. Sample injections were made using a 10-μL M-Class PEEKsil needle equipped with a 20-μL sample loop and a 100-μL sample syringe. A 5-μL volume of each sample was injected for analysis.

Peptides were detected in positive ESI mode under the following source conditions: source voltage of 3.0 kV, cone voltage of 35 V, source temperature of 150°C, and desolvation temperature of 300°C. The gas flow rates were set at 600 L/h for desolvation, 150 L/h for the cone, and a nebulizer pressure of 7 bar. Argon was utilized as the collision gas at a steady flow rate of 100 μL/min. For the data acquisition method, the peak width was set to 20, and 12 points per peak were required.

The peak areas of the quantifier transition for each peptide were measured through peak integration using the TargetLynx Application Manager within the MassLynx version 4.2 software (Waters Corporation). Peaks were detected at the exact retention times corresponding to the elution of each peptide, with the isotopically labeled standards serving as controls to confirm retention time alignment. Standard peak detection parameters were applied, including a peak-to-peak noise amplitude of 10, a balance setting of 10, and a splitting setting of 50.

ELISA

ELISA for analysis of the capsid titer was performed on the rAAV9 sample using the commercial AAV9 Titration ELISA kit (PROGEN) following the manufacturer's recommended protocols. The rAAV9 sample was analyzed in triplicate.

DATA AND CODE AVAILABILITY

The data that support the results of this study can be obtained from the corresponding author upon reasonable request.

ACKNOWLEDGMENTS

This work was financially supported by the Community for Analytical Measurement Science (CAMS) through a 2021 CAMS Fellowship Award funded by the Analytical Chemistry Trust Fund, the Department for Science, Innovation, and Technology through the Chemical and Biological Metrology programme, the National Measurement Laboratory at the LGC and the University of Kent. The graphical abstract and figures were created with BioRender.com. We would like to thank Dr. Luise Luckau, Dr. Giles Drinkwater, and Dr. Simon Cowen for their guidance and support throughout the project. We would like to extend our sincere thanks to Professor Paul Dalby and Dr. Yiwen Li for generously providing the rAAV8 sample, and Professor Giovanna Mallucci for providing the rAAV9 sample.

AUTHOR CONTRIBUTIONS

All authors contributed to the writing and editing of the manuscript.

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.2025.101562.

REFERENCES

1. Naso, M.F., Tomkowicz, B., Perry, W.L., and Strohl, W.R. (2017). Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. BioDrugs *31*, 317–334. https://doi.org/10.1007/s40259-017-0234-5.

2. van der Loo, J.C.M., and Wright, J.F. (2016). Progress and challenges in viral vector manufacturing. Hum. Mol. Genet. *25*, R42–R52. https://doi.org/10.1093/hmg/ddv451.

3. Tustian, A.D., and Bak, H. (2021). Assessment of quality attributes for adeno-associated viral vectors. Biotechnol. Bioeng. *118*, 4186–4203. https://doi.org/10.1002/bit.27905.

4. Kontogiannis, T., Braybrook, J., McElroy, C., Foy, C., Whale, A.S., Quaglia, M., and Smales, C.M. (2024). Characterization of AAV vectors: A review of analytical techniques and critical quality attributes. Mol. Ther. Methods Clin. Dev. *32*, 101309. https://doi.org/10.1016/j.omtm.2024.101309.

5. Moullier, P., and Snyder, R.O. (2012). Recombinant Adeno-Associated Viral Vector Reference Standards. Methods Enzymol. *507*, 297–311. https://doi.org/10.1016/B978-0-12-386509-0.00015-6.

6. Wright, J.F. (2021). Quality Control Testing, Characterization and Critical Quality Attributes of Adeno-Associated Virus Vectors Used for Human Gene Therapy. Biotechnol. J. *16*, e2000022. https://doi.org/10.1002/biot.202000022.

7. Bennett, A., Patel, S., Mietzsch, M., Jose, A., Lins-Austin, B., Yu, J.C., Bothner, B., McKenna, R., and Agbandje-McKenna, M. (2017). Thermal Stability as a Determinant of AAV Serotype Identity. Mol. Ther. Methods Clin. Dev. *6*, 171–182. https://doi.org/10.1016/j.omtm.2017.07.003.

8. Rose, J.A., Maizel, J.V., Inman, J.K., and Shatkin, A.J. (1971). Structural Proteins of Adenovirus-Associated Viruses. J. Virol. *8*, 766–770.

9. Snijder, J., van deWaterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A.J.R. (2014). Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry. J. Am. Chem. Soc. *136*, 7295–7299. https://doi.org/10.1021/ja502616y.

10. Wörner, T.P., Bennett, A., Habka, S., Snijder, J., Friese, O., Powers, T., Agbandje-McKenna, M., and Heck, A.J.R. (2021). Adeno-associated virus capsid assembly is divergent and stochastic. Nat. Commun. *12*, 1642. https://doi.org/10.1038/s41467-021-21935-5.

11. Bosma, B., du Plessis, F., Ehlert, E., Nijmeijer, B., de Haan, M., Petry, H., and Lubelski, J. (2018). Optimization of viral protein ratios for production of rAAV serotype 5 in the baculovirus system. Gene Ther. *25*, 415–424. https://doi.org/10.1038/s41434-018-0034-7.

12. Erickson, S.B., Pham, Q., Cao, X., Glicksman, J., Kelemen, R.E., Shahraeini, S.S., Bodkin, S., Kiyam, Z., and Chatterjee, A. (2024). Precise Manipulation of the Site and Stoichiometry of Capsid Modification Enables Optimization of Functional Adeno-Associated Virus Conjugates. Bioconjug. Chem. *35*, 64–71. https://doi.org/10.1021/acs.bioconjchem.3c00411.

13. Onishi, T., Nonaka, M., Maruno, T., Yamaguchi, Y., Fukuhara, M., Torisu, T., Maeda, M., Abbatiello, S., Haris, A., Richardson, K., et al. (2023). Enhancement of recombinant adeno-associated virus activity by improved stoichiometry and homogeneity of capsid protein assembly. Mol. Ther. Methods Clin. Dev. *31*, 101142. https://doi.org/10.1016/j.omtm.2023.101142.

14. Oyama, H., Ishii, K., Maruno, T., Torisu, T., and Uchiyama, S. (2021). Characterization of Adeno-Associated Virus Capsid Proteins with Two Types of VP3-Related Components by Capillary Gel Electrophoresis and Mass Spectrometry. Hum. Gene Ther. *32*, 1403–1416. https://doi.org/10.1089/hum.2021.009.

15. Su, Q., Sena-Esteves, M., and Gao, G. (2020). Analysis of Recombinant Adeno-Associated Virus (rAAV) Purity Using Silver-Stained SDS-PAGE. Cold Spring Harb. Protoc. *2020*, 095679. https://doi.org/10.1101/pdb.prot095679.

16. Fekete, S., Aebischer, M.K., Imiołek, M., Graf, T., Ruppert, R., Lauber, M., D'Atri, V., and Guillarme, D. (2023). Chromatographic strategies for the analytical characterization of adeno-associated virus vector-based gene therapy products. TrAC Trends Anal. Chem. *164*, 117088. https://doi.org/10.1016/j.trac.2023.117088.

17. Fernandes, R.P., Escandell, J.M., Guerreiro, A.C.L., Moura, F., Faria, T.Q., Carvalho, S.B., Silva, R.J.S., Gomes-Alves, P., and Peixoto, C. (2022). Assessing Multi-Attribute Characterization of Enveloped and Non-Enveloped Viral Particles by Capillary Electrophoresis. Viruses *14*, 2539. https://doi.org/10.3390/v14112539.

18. Hao, Z., Zhang, T., Xuan, Y., Wang, H., Qian, J., Lin, S., Chen, J., Horn, D.M., Argoti, D., Beck, A., et al. (2015). Intact Antibody Characterization Using Orbitrap Mass Spectrometry. In ACS Symposium Series, *1202*, J.E. Schiel, D.L. Davis, and O.V. Borisov, eds., pp. 289–315. https://doi.org/10.1021/bk-2015-1202.ch010.

19. Mary, B., Maurya, S., Arumugam, S., Kumar, V., and Jayandharan, G.R. (2019). Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes. FEBS J. *286*, 4964–4981. https://doi.org/10.1111/febs.15013.

20. Murray, S., Nilsson, C.L., Hare, J.T., Emmett, M.R., Korostelev, A., Ongley, H., Marshall, A.G., and Chapman, M.S. (2006). Characterization of the capsid protein glycosylation of adeno-associated virus type 2 by high-resolution mass spectrometry. J. Virol. *80*, 6171–6176. https://doi.org/10.1128/JVI.02417-05.

21. Rathore, D., Faustino, A., Schiel, J., Pang, E., Boyne, M., and Rogstad, S. (2018). The role of mass spectrometry in the characterization of biologic protein products. Expert Rev. Proteomics *15*, 431–449. https://doi.org/10.1080/14789450.2018.1469982.

22. Serrano, M.A.C., Furman, R., Chen, G., and Tao, L. (2023). Mass spectrometry in gene therapy: Challenges and opportunities for AAV analysis. Drug Discov. Today *28*, 103442. https://doi.org/10.1016/j.drudis.2022.103442.

23. Rumachik, N.G., Malaker, S.A., Poweleit, N., Maynard, L.H., Adams, C.M., Leib, R.D., Cirolia, G., Thomas, D., Stamnes, 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.

24. Campbell, J.J., Almond, N., Bae, Y.-K., Bhuller, R., Briones, A., Cho, S.-J., Cleveland, M.H., Cleveland, T.E., Galaway, F., He, H.-J., et al. (2024). Standards and Metrology

for Viral Vectors as Molecular Tools: Outcomes from a CCQM Workshop. Biologics *4*, 187–201. https://doi.org/10.3390/biologics4020013.

25. Beaumal, C., Guapo, F., Smith, J., Carillo, S., and Bones, J. (2025). Combination of hydrophilic interaction liquid chromatography and top-down mass spectrometry for characterisation of adeno-associated virus capsid proteins. Anal. Bioanal. Chem. *417*, 3405–3417. https://doi.org/10.1007/s00216-025-05874-4.
26. 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.
27. 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. Anal. *207*, 114427. https://doi.org/10.1016/j.jpba.2021.114427.
28. 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.
29. Nielsen, S.D.-H., Liang, N., Rathish, H., Kim, B.J., Lueangsakulthai, J., Koh, J., Qu, Y., Schulz, H.-J., and Dallas, D.C. (2024). Bioactive milk peptides: An updated comprehensive overview and database. Crit. Rev. Food Sci. Nutr. *64*, 11510–11529. https://doi.org/10.1080/10408398.2023.2240396.
30. Giles, A.R., Sims, J.J., Turner, K.B., Govindasamy, L., Alvira, M.R., Lock, M., and Wilson, J.M. (2018). Deamidation of Amino Acids on the Surface of Adeno-Associated Virus Capsids Leads to Charge Heterogeneity and Altered Vector Function. Mol. Ther. *26*, 2848–2862. https://doi.org/10.1016/j.ymthe.2018.09.013.
31. Zhou, Y., and Wang, Y. (2023). Direct deamidation analysis of intact adeno-associated virus serotype 9 capsid proteins using reversed-phase liquid chromatography. Anal. Biochem. *668*, 115099. https://doi.org/10.1016/j.ab.2023.115099.
32. MacLean, B., Tomazela, D.M., Abbatiello, S.E., Zhang, S., Whiteaker, J.R., Paulovich, A.G., Carr, S.A., and MacCoss, M.J. (2010). Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry. Anal. Chem. *82*, 10116–10124. https://doi.org/10.1021/ac102179j.
33. Prakash, A., Tomazela, D.M., Frewen, B., Maclean, B., Merrihew, G., Peterman, S., and Maccoss, M.J. (2009). Expediting the development of targeted SRM assays: Using data from shotgun proteomics to automate method development. J. Proteome Res. *8*, 2733–2739. https://doi.org/10.1021/pr801028b.
34. Kiyonami, R., and Domon, B. (2010). Selected Reaction Monitoring Applied to Quantitative Proteomics. In Methods in Molecular Biology, *658*, P.R. Cutillas and J.F. Timms, eds. (Humana Press), pp. 155–166. https://doi.org/10.1007/978-1-60761-780-8_9.
35. Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008). Selected reaction monitoring for quantitative proteomics: A tutorial. Mol. Syst. Biol. *4*, 222. https://doi.org/10.1038/msb.2008.61.
36. Sayers, E.W., Bolton, E.E., Brister, J.R., Canese, K., Chan, J., Comeau, D.C., Connor, R., Funk, K., Kelly, C., Kim, S., et al. (2022). Database resources of the national center for biotechnology information. Nucleic Acids Res. *50*, D20–D26. https://doi.org/10.1093/nar/gkab1112.
37. Torma, A.F., Groves, K., Biesenbruch, S., Mussell, C., Reid, A., Ellison, S., Cramer, R., and Quaglia, M. (2017). A candidate liquid chromatography mass spectrometry reference method for the quantification of the cardiac marker 1-32 B-type natriuretic peptide. Clin. Chem. Lab. Med. *55*, 1397–1406. https://doi.org/10.1515/cclm-2016-1054.
38. Zhang, L., Illes-Toth, E., Cryar, A., Drinkwater, G., Di Vagno, L., Pons, M.-L., Mateyka, J., McCullough, B., Achtar, E., Clarkson, C., et al. (2024). A candidate reference measurement procedure for the quantification of α-synuclein in cerebrospinal fluid using an SI traceable primary calibrator and multiple reaction monitoring. Analyst *149*, 4842–4850. https://doi.org/10.1039/D4AN00634H.
39. Aloor, A., Zhang, J., Gashash, E.A., Parameswaran, A., Chrzanowski, M., Ma, C., Diao, Y., Wang, P.G., and Xiao, W. (2018). Site-Specific N-Glycosylation on the AAV8 Capsid Protein. Viruses *10*, E644. https://doi.org/10.3390/v10110644.
40. Leibiger, T.M., Min, L., and Lee, K.H. (2024). Quantitative proteomic analysis of residual host cell protein retention across adeno-associated virus affinity chromatography. Mol. Ther. Methods Clin. Dev. *32*, 101383. https://doi.org/10.1016/j.omtm.2024.101383.
41. 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.
42. International Organization for Standardization (ISO) (2020). In vitro diagnostic medical devices—Requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. www.iso.org/standard/69984.html.
43. Issa, S.S., Shaimardanova, A.A., Solovyeva, V.V., and Rizvanov, A.A. (2023). Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. Cells *12*, 785. https://doi.org/10.3390/cells12050785.
44. Liu, D., Zhu, M., Zhang, Y., and Diao, Y. (2021). Crossing the blood-brain barrier with AAV vectors. Metab. Brain Dis. *36*, 45–52. https://doi.org/10.1007/s11011-020-00630-2.
45. Tiambeng, T.N., Yan, Y., Patel, S.K., Cotham, V.C., Wang, S., and Li, N. (2025). Characterization of adeno-associated virus capsid proteins using denaturing size-exclusion chromatography coupled with mass spectrometry. J. Pharm. Biomed. Anal. *253*, 116524. https://doi.org/10.1016/j.jpba.2024.116524.
46. Gauci, V.J., Wright, E.P., and Coorssen, J.R. (2011). Quantitative proteomics: Assessing the spectrum of in-gel protein detection methods. J. Chem. Biol. *4*, 3–29. https://doi.org/10.1007/s12154-010-0043-5.
47. Grove, H., Færgestad, E.M., Hollung, K., and Martens, H. (2009). Improved dynamic range of protein quantification in silver-stained gels by modelling gel images over time. Electrophoresis *30*, 1856–1862. https://doi.org/10.1002/elps.200800568.
48. Adachi, K., Enoki, T., Kawano, Y., Veraz, M., and Nakai, H. (2014). Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. Nat. Commun. *5*, 3075. https://doi.org/10.1038/ncomms4075.
49. Emmanuel, S.N., Smith, J.K., Hsi, J., Tseng, Y.-S., Kaplan, M., Mietzsch, M., Chipman, P., Asokan, A., McKenna, R., and Agbandje-McKenna, M. (2022). Structurally Mapping Antigenic Epitopes of Adeno-associated Virus 9: Development of Antibody Escape Variants. J. Virol. *96*, e01251-21. https://doi.org/10.1128/JVI.01251-21.
50. Verch, T., Roselle, C., and Shank-Retzlaff, M. (2016). Reduction of Dilution Error in Elisas Using an Internal Standard. Bioanalysis *8*, 1451–1464. https://doi.org/10.4155/bio-2016-0053.
51. Higgins, K.M., Davidian, M., Chew, G., and Burge, H. (1998). The Effect of Serial Dilution Error on Calibration Inference in Immunoassay. Biometrics *54*, 19–32. https://doi.org/10.2307/2533992.
52. Smith, J., Guapo, F., Strasser, L., Millán-Martín, S., Milian, S.G., Snyder, R.O., and Bones, J. (2023). Development of a Rapid Adeno-Associated Virus (AAV) Identity Testing Platform through Comprehensive Intact Mass Analysis of Full-Length AAV Capsid Proteins. J. Proteome Res. *23*, 161–174. https://doi.org/10.1021/acs.jproteome.3c00513.
53. 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.
54. Wingfield, P.T. (2001). Use of protein folding reagents. Curr. Protoc. Protein Sci. Appendix 3, Appendix 3A. https://doi.org/10.1002/0471140864.psa03as00.
55. Dawson, P.H., French, J.B., Buckley, J.A., Douglas, D.J., and Simmons, D. (1982). The use of triple quadrupoles for sequential mass spectrometry: 2—A detailed case study. Org. Mass Spectrom. *17*, 212–219. https://doi.org/10.1002/oms.1210170504.
56. Michaud, S.A., Pětrošová, H., Sinclair, N.J., Kinnear, A.L., Jackson, A.M., McGuire, J.C., Hardie, D.B., Bhowmick, P., Ganguly, M., Flenniken, A.M., et al. (2024). Multiple reaction monitoring assays for large-scale quantitation of proteins from 20 mouse organs and tissues. Commun. Biol. *7*, 6. https://doi.org/10.1038/s42003-023-05687-0.

OMTM, Volume *33*

Supplemental information

Development of LC-MS methods for AAV capsid protein quantification and host cell protein profiling

Theodoros Kontogiannis, Christopher McElroy, Milena Quaglia, Carole Foy, Alexandra S. Whale, Julian Braybrook, and C. Mark Smales

Peptide Release Monitoring

The completion of the digestion and release of the VP1u, VP3u, and VPc peptides from the capsid was assessed. The digested rAAV9 sample was spiked with the isotopically labeled peptide master mix post-digestion, as previously described. Aliquots were collected at 1, 3, 5, 7, and 23 h and analyzed using the developed multiple reaction monitoring (MRM) method on the triple quadrupole mass spectrometer, with three replicate injections per time point. For each time point, the ratio of the peak areas of the endogenous peptides (VP1u, VP3u, and VPc) to their corresponding isotopically labeled standards was calculated. The ratio of the peak areas to their isotopically labeled counterparts reaches a plateau after 3 hours of digestion for VP1u and VP3u and 1 hour for VPc, suggesting the completeness of the process (Figure S11).

LC-MS Analysis for Purity Assessment of VP1u, VP3u, VPc Peptides

For purity analysis, 1 µg of each synthesized natural peptide (VP1u, VP3u, VPc) was injected into an Avantor ACE 3 C18-300 (100 × 2.1 mm id) column (Avantor, Cat No. ACE-213-1002) maintained at 40 °C. In this experiment, 0.1% formic acid in water (solvent A) and 0.1% formic acid in acetonitrile (solvent B) were used as mobile phases. A 20% acetonitrile solution was used as the needle wash, while 70% acetonitrile in ultra-pure water was used as the rear seal wash. The LC method used for analysis started from 3% solvent B with a flow rate of 0.2 ml/min for the first 3 min. After this, there was a linear increase to 45% solvent B at 41 min, followed by an increase to 100% B at 44 min, which was then held until 48 min. Finally, there was a return to 3% solvent B using a linear gradient at 52 min, which was then held to the end of the run at 65 min.

MS analysis was performed using a Q-Exactive Plus Orbitrap instrument equipped with a HESI-II probe source (Thermo Fisher Scientific) in positive ESI mode. The settings were as follows: a source temperature of 320 °C, spray voltage of 3.5 kV, S-lens RF level of 100, sheath gas flow rate of 25 a.u., and auxiliary gas

flow rate of 5 a.u. The maximum injection time was set to 100 ms, with an AGC target value of 3e6 and 1 microscan. Data were acquired in 70K resolution mode over a mass range of 200–1,800 m/z.

Identification of the VP1u, VP3u, and VPc peptides in each chromatogram was carried out by searching for their specific m/z ratios (646.29 for VP1u, 600.76 for VP3u, and 861.90 for VPc) using the Extracted Ion Chromatogram (XIC) function in Xcalibur™ software (Thermo Fisher Scientific). The purity percentages of each synthesized natural peptide were then determined by manual peak integration using Freestyle™ software (Thermo Fisher Scientific) (Table S7).

The analysis revealed that the peptides were not 95% pure as ordered, as confirmed by manual peak integration of the chromatograms shown in Figure S11 (Panels B, D, and F) using Freestyle software (Thermo Fisher Scientific). The measured purities are detailed in Table S7.

Amino Acid Analysis

Amino acid analysis of the peptides VP1, VP3, and VPc was carried out based on a previously reported approach [1,2]. SI-traceability was ensured by using certified reference materials of natural amino acids obtained from the National Metrology Institute of Japan (NMIJ) (NMIJ; Tsukuba, Japan). This method analyzes the volatile amino acids present in the peptide sequences (VP1u, VP3u, VPc), including glycine, leucine, alanine, valine, proline, lysine, and isoleucine using amino acid-certified reference materials (CRM). Arginine and phenylalanine, though volatile, were not included in the data analysis as they are absent from these peptides. Glycine (CRM 6022-a, 99.9 (0.2) % purity), L-leucine (CRM 6012-a, 99.9 (0.2) % purity), L-alanine (CRM 6011-a, 99.9 (0.2) % purity), L-valine (CRM 6015-a, 99.8 (0.2) % purity), L-lysine monohydrochloride (CRM 6018-a, 99.8 (0.2) % purity), L-isoleucine (CRM 6012-a, 99.7 (0.2) % purity), L-proline (CRM 6016-a, 99.9 (0.2) % purity). The labeled amino acids (Cambridge Isotopes Laboratory, Andover, MA, USA) used were: L-glycine ($^{13}C_2{}^{15}N$, 98% purity), L-leucine ($^{13}C_6{}^{15}N$, 98% purity), L-alanine

($^{13}C_3{}^{15}N$, 98% purity), L-valine ($^{13}C_5{}^{15}N$, 98% purity), L-lysine/2HCl ($^{13}C_6$, $^{15}N_2$, 98% purity), L-isoleucine ($^{13}C_6$, 98% purity), L-proline ($^{13}C_5{}^{15}N$, 98% purity).

Prior to preparing the blends for amino acid analysis, the VP1u synthesized natural peptide was diluted 1:20 in ultra-pure water, whereas VP3u and VPc were each diluted 1:10 in ultra-pure water. In summary, three sample blends were gravimetrically prepared, each containing isotopically labeled amino acids along with the target peptide (VP1, VP3, or VPc). Alongside, three calibration blends were made, combining natural amino acids and labeled standards in equimolar amounts to the sample blends. Additional preparations included reagent blanks, a natural amino acid standard solution without the labeled counterpart, a labeled amino acid standard solution without the natural amino acids, and a recovery vial (Table S5). The labeled and natural amino acid stocks were prepared in 0.1 M hydrochloric acid (diluted in ultra-pure water from stock 6M) (Sigma-Aldrich, Cat No. 13-1686).

The samples were freeze-dried in a Genevac EZ-2 plus vacuum centrifuge (SP Scientific), followed by microwave-assisted hydrolysis at 190 °C for 4 h in an Ethos EZ Microwave Digestion System (Analytix) with 6 M hydrochloric acid (Sigma-Aldrich, Cat No. 13-1686). Following hydrolysis, the samples were freeze-dried, reconstituted in 80 µL of N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA) with 1% tert-butyldimethylchlorosilane (TBDMSCl) (Sigma-Aldrich, Cat No. 375934) for derivatization, and incubated at 85 °C for 90 min in a thermoshaker (Microtherm, Camlab) with shaking at 50 rpm. The samples were then analyzed using gas chromatography-mass spectrometry (GC-MS) on an Agilent 7890B gas chromatography system equipped with a Zebron ZB-5HT Inferno GC column (30 m × 0.25 mm i.d., 0.25 mm film thickness) and a Z-guard column (2 m, 0.53 mm i.d.) (Phenomenex, Cat No. 7HG-G015-02-GGA), coupled to an Agilent GC-MS system (Agilent) equipped with a CTC Combi-PAL Autosampler system (CTC Analytics AG). The sample blends were injected five times, bracketed by the calibration blends. A GC-

MS/MS method with multiple reaction monitoring (MRM) was used to analyze the amino acids in the samples. The transitions monitored are shown in Table S6.

The analysis was conducted in a split mode of 1:5, with a split flow set to 12.2 mL/min and inlet temperature maintained at 300 °C. A volume of 1 µL was injected for each sample. The temperature program began with a hold at 130 °C for 2 minutes, followed by a ramp to 200 °C at 25 °C/min, then an increase to 320 °C at 120 °C/min, where the temperature was held for 6 min.

Using the MRM method, the peak area ratio of each analyzed amino acid (either natural in the calibration blend or endogenous in the sample blend) to its corresponding isotopically labeled counterpart was measured across five injections. This ratio was calculated for each amino acid in all three sample blends and calibration blends. Subsequently, the amount of each amino acid released from the peptide of interest (W_x) in each sample blend was determined using Equation S1. In this equation, W_Z represents the concentration of the pure reference material in the standard solution (µg/g). The terms m_y and m_x refer to the masses (g) of the labeled standard and the sample that are combined to prepare the sample blend. The terms m_z and m_{yc} correspond to the masses (g) of the unlabeled and labeled reference standards, respectively, used to prepare the calibration blend. R_B and R_{BC} represent the observed signal ratios (average from 5 injections) between the unlabeled and labeled peptides in the sample and in the calibration blend, respectively.

$$W_x = W_z \times \frac{m_y}{m_x} \times \frac{m_z}{m_{yc}} \times \frac{R_B}{R_{BC}} \qquad \text{(Equation S1)}$$

To determine the nmol/g concentration of each amino acid, the calculated amount (W_x) was first divided by the molecular weight of the corresponding amino acid and then multiplied by 1000 and divided by the number of times that amino acid appears in the peptide of interest. To estimate the stock concentration (nmol/g) of each peptide, the average value across the three sample blends and all relevant amino acid

measurements were calculated and multiplied by the corresponding dilution factor. The standard deviation for each peptide was determined from all amino acid measurements across the three blends. Finally, the resulting concentration was converted to mg/mL based on the molecular weight and measured purity of each peptide. The combined standard deviation was calculated using the error propagation formula, incorporating the standard deviations from both the amino acid analysis and the peptide purity assessment.

The amino acid analysis results for all synthesized natural peptides were consistent, although slight variations were observed for VP3u. These variations can be explained by the lower purity of VP3u, as impurities may contain amino acids present in the peptide sequence, leading to inflated measurements.[3] This highlights the importance of accounting for peptide purity in quantitative analyses.

Collision Energy Optimization

Choosing a collision energy (CE) that leads to high-intensity transitions is important for improving the sensitivity of the method. The intensity of each transition peak for VP1u, VP3u, and VPc across 12 different CEs was plotted. Based on this analysis (Figure S10), the CEs that yielded the highest intensity for each transition were selected. For VP1u, a CE of 14 was optimal for all three transitions (intensity of 1.5×10^6 for T1, 1.7×10^6 for T2, 8.4×10^5 for T3). For VP3u, CE 14 was used for T1 (5.1×10^6), while CE 16 was selected for T2 and T3 (4.7×10^6 for T2 and 3.6×10^5 for T3). For VPc, CE 30 was chosen for T1 (1.1×10^6), CE 32 for T2 (1.2×10^6), and CE 28 for T3 (7.6×10^5) (Table S2, Figure S10).

Equations for Calculation of Variability

The inter-experimental variability for VP2 was calculated using an error propagation formula that considers the correlation between peptides within the capsid. The peptides show a strong linear relationship, meaning that when the concentration of one peptide increases, the concentrations of VP3u and VP1u also tend to increase. This correlation was taken into account to ensure an accurate estimation

of variability. Therefore, the sample covariances for the VP3u-VPc pair, VPc-VP1u pair, and VP1u-VP3u pair were first calculated using the mean concentration values from each experiment. The sample covariance was calculated in Microsoft Excel 365 using the COVARIANCE.S(array1, array2) function. For each peptide pair, array1 and array2 correspond to the mean values of the respective peptides across the three experiments. The arrays were adjusted accordingly for each peptide pair being evaluated. In general, the sample covariance between two variables, x and y, is calculated using Equation S2. In this equation, xi and yi refer to the measured values for each experiment for peptide 1 and peptide 2, respectively. The terms $\bar{x}$ and $\bar{y}$ represent the mean values of the x and y datasets, and n is the number of observations.

$$\mathrm{Cov}(X, Y) = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{n-1} \qquad \text{(Equation S2)}$$

The inter-experimental variability for VP1u, VP3u, and VPc was calculated as the standard deviation of the mean values obtained across the three experiments. Therefore, the inter-experimental variability for VP2 was calculated using Equation S3, where $\sigma_{\{VPx\}}$ represents the inter-experimental standard deviation of each peptide, and Cov represents the sample covariance for each peptide pair.

$$\sigma_{\{VP2\}} = \sqrt{\sigma_{\{VP1u\}} + \sigma_{\{VP3u\}} + \sigma_{\{VPc\}} - 2 \times \mathrm{Cov}(VP3u, VPc) - 2 \times \mathrm{Cov}(VP1u, VPc) + 2 \times \mathrm{Cov}(VP1u, VP3u)}$$

(Equation S3)

The inter-experimental variability for VP2 was calculated by dividing the measured standard deviation by the mean calculated concentration. The combined variability (combined CV) for the measurement of VP1u, VP3u, and VPc was calculated by estimating the CVs associated with amino acid analysis (CV_{AAA}) peptide purity assessment (CV_{purity}), and the variability observed across the three

experiments (CV_{exp}), and then applying an error propagation formula to combine these sources of variability (Equation S4).

$$CV_{combined} = \sqrt{CV^2{}_{AAA} + CV^2{}_{purity} + CV^2{}_{exp}} \qquad \text{(Equation S4)}$$

SDS-PAGE Analysis

A volume of 19.5 µL (total 1 µg) of rAAV8 sample was mixed with 7.5 µL of NuPAGE LDS Sample Buffer (4 ×) (Thermo Fisher Scientific, Cat No. NP0007) and 3 µL of NuPAGE Reducing Agent (10 ×) (Thermo Fisher Scientific, NP0009), resulting in a final volume of 30 µL. The mixture was thoroughly pipetted to ensure proper mixing, then incubated at 70 °C for 10 min in a Bio-Rad C1000 Touch thermal cycler, followed by an additional incubation at 95 °C for 10 min to fully denature the capsids. NuPAGE 10% Bis-Tris, 1.0 mm mini 10-well precast gels (Thermo Fisher Scientific, Cat No. NP0301BOX) were used for the analysis. A total of 25 µL of the prepared rAAV8 sample was loaded into its respective well, and 5 µL of the PageRuler™ Plus Prestained Protein Ladder (10–250 kDa) (Thermo Fisher Scientific, Cat No. 26619) was added to the first well. Electrophoresis was performed at 180 V for 70 min. After electrophoresis, the gel was fixed by shaking it for 10 minutes in a solution of 40 mL deionised water (Elga Purelab Ultra), 50 mL methanol (Romil, Cat No. H411), and 10 mL acetic acid (Fisher Scientific, Cat No. 10041250). The fixing solution was replaced with a staining solution prepared using the commercial Colloidal Blue Staining Kit (Thermo Fisher Scientific, Cat No. LC6025). The staining solution was prepared by mixing 55 mL of deionised water, 20 mL of methanol, and 20 mL of Stain A. The gel was shaken in the staining solution for 10 min at room temperature, after which 5 mL of Stain B was added. The gel was incubated with gentle shaking in the staining solution for 4 h. Finally, the staining solution was carefully poured off, and the gel was transferred to a container with 200 mL of deionised water. The gel was gently shaken for 10 h to remove any excess stain.

For SDS-PAGE analysis of the rAAV9 sample, 5.6×10^{10} rAAV9 capsids were prepared in 20 µL of reducing Laemmli sample buffer and boiled for 5 minutes. The entire sample was loaded onto a 4% stacking / 10% resolving SDS-PAGE gel and electrophoresed alongside a molecular weight marker (Bio-Rad Precision Plus Protein™ Dual Color Standard, Cat No. 1610374) at 4 A until the dye front reached the bottom of the gel. Following electrophoresis, the gel was fixed overnight in 40% ethanol and 10% acetic acid, then silver stained using the PlusOne™ Silver Staining Kit (GE Healthcare). Gels were scanned using an Azure 300 imaging system, and the relative band intensities of VP1, VP2, and VP3 were quantified using ImageJ software.

Table S1. Theoretical and measured masses (Da) of VP1, VP2, VP3 for rAAV8, rAAV9.

AAV Serotype	Capsid Protein	Theoretical Mass (Da)	Experimental Mass (Da)	Parts per million (ppm)
rAAV8	VP1	81,667	81,670	37
	VP2	66,519	66,520	15
	VP3	59,805	59,800	84
rAAV9	VP1	81,291	81,290	12
	VP2	66,210	66,210	0
	VP3	59,733	59,730	50

Table S2. Mean concentrations (mg/mL and total mg) of synthesized natural peptide stock solutions, as determined by amino acid analysis and corrected for peptide purity. The values in parentheses represent the combined SD, encompassing the measurements of amino acid analysis and purity assessment.

Natural Peptide	Concentration of Peptide (mg/mL) in Stock	Total mg of Peptide in Stock
VP1u	1.124 (0.029)	2.25 (0.058)
VP3u	0.096 (0.002)	0.38 (0.008)
VPc	0.449 (0.003)	1.79 (0.012)

Table S3. Composition of controls, sample blends, calibration blends, and blanks used for amino acid analysis.

Component	Unlabeled Amino Acids	Peptide VP Hydrolysate	Labeled Amino Acids
Reagent Blank	-	-	-
Unlabeled Amino Acid Control	X	-	-
Labeled Amino Acid Control	-	-	X
Calibration Blends 1 - 3	X	-	X
Sample Blends 1 - 3	-	X	X
Sample Blank	-	X	-

Table S4. MRM transitions used for each amino acid in the analyzed samples, listing the transitions for both unlabeled and labeled amino acids.

Amino Acid	Quantifier Transition (m/z) Unlabeled MTBSTFA Derivative	Quantifier Transition (m/z) Labeled MTBSTFA Derivative
Alanine	260/232	264/235
Glycine	246/218	249/220
Valine	288/260	294/265
Leucine	302/274	308/279
Isoleucine	302/274	308/279
Proline	286/258	292/263
Lysine	431/300	439/307

Table S5. Purity assessment of VP1u, VP3u, and VPc peptides by LC-MS. The values in parentheses represent the SD from three replicate injections.

Natural Peptide	Mean Purity %
VP1u	79.7 (1.1) %
VP3u	22.1 (0.5) %
VPc	86.9 (0.1) %

Table S6. Full list of HCPs identified in the rAAV8 sample. The table can be found in the Supplemental spreadsheet.

Table S7. Sequences of the peptides (VP1u, VP3u, VPc), including the modified amino acids in the synthesized isotopically labeled peptides. Ac. = acetylation.

Peptide	Endogenous Sequence	Modified amino acid in isotopically labeled peptide	Mass Shift (Da)
VP1u	(Ac.)AADGYLPDWLE	Leucine (L)	+7
VP3u	(Ac.)ASGGGAPVADNNE	Proline (P)	+6
VPc	IQYTSNYYKSNNVE	Lysine (K)	+8

Table S8. Selected transitions and corresponding optimal collision energies for the MRM experiment.

T1= quantifier transition, T2 and T3 are the qualifier transitions.

Peptide	m/z	T1 m/z	T2 m/z	T3 m/z	T1 CE	T2 CE	T3 CE
VP1u	646.29	633.19	659.21	772.29	14	14	14
VP1u Labeled	652.81	633.20	672.25	785.32	14	14	14
VP3u	600.76	443.11	758.28	829.31	14	16	16
VP3u Labeled	606.77	443.11	770.29	841.33	14	16	16
VPc	861.90	1016.45	1217.55	870.35	30	32	28
VPc Labeled	869.42	1024.47	1225.54	877.43	30	32	28

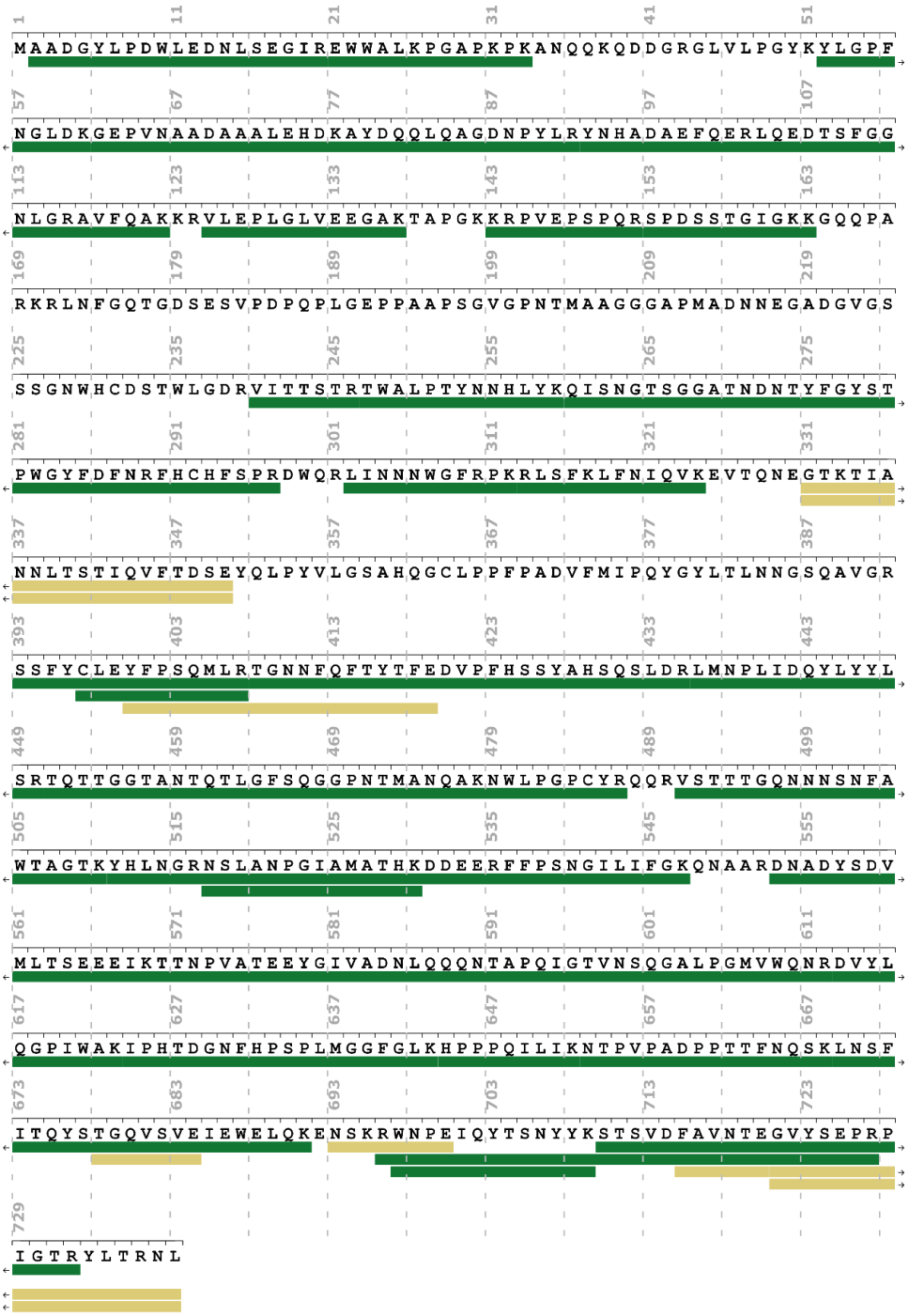


Figure S1. Protein sequence coverage map of rAAV8 VP1 protein. The peptide map shows sequence coverage of VP1 (78.2%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualization tool (Aarhus University Visualization Group).[4]

Figure S2. **Protein sequence coverage map of rAAV8 VP2 protein.** The peptide map shows sequence coverage of VP2 (76.7%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualization tool (Aarhus University Visualization Group).[4]

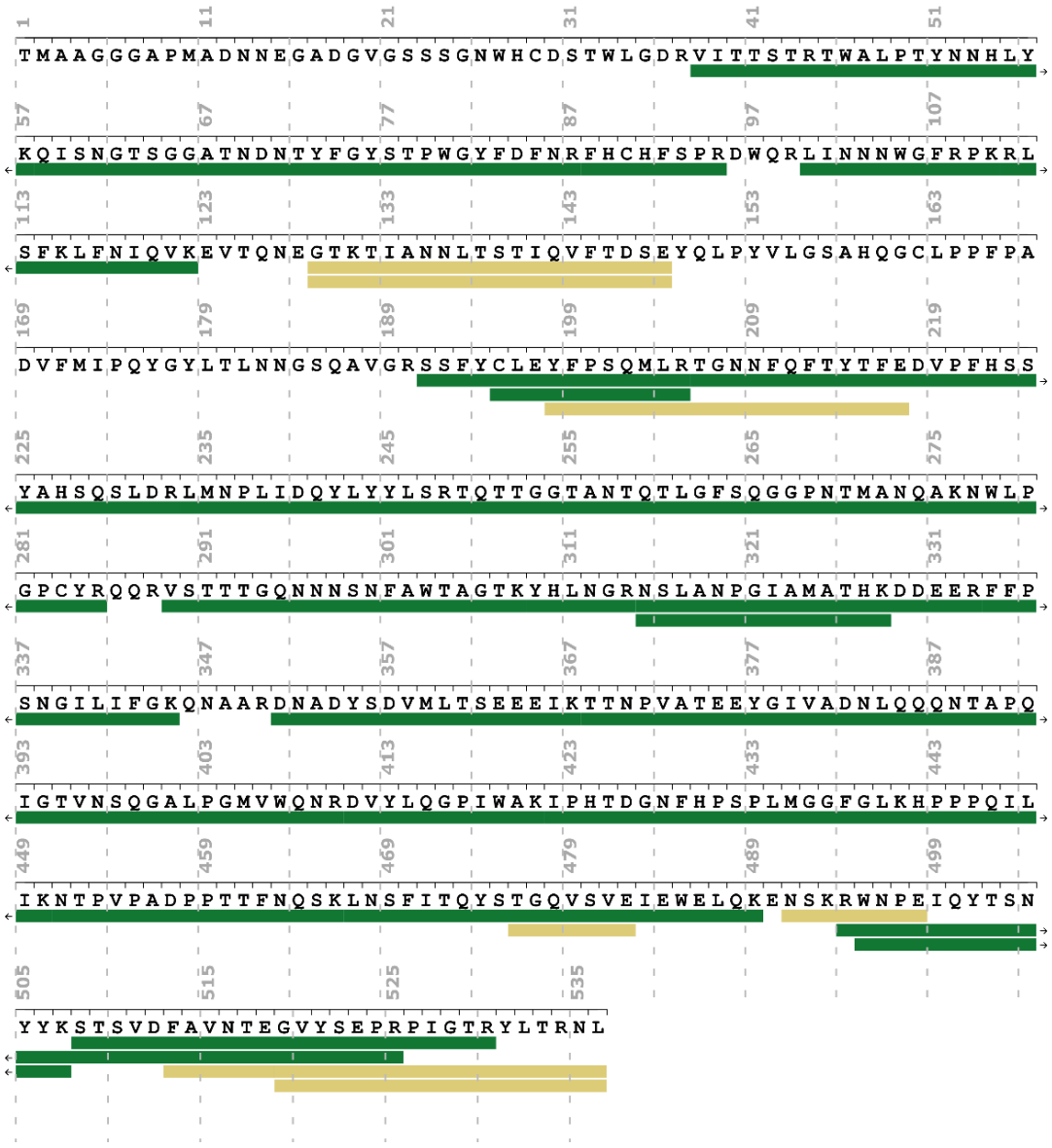


Figure S3. **Protein sequence coverage map of rAAV8 VP3 protein.** The peptide map shows sequence coverage of VP3 (81.9%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualization tool (Aarhus University Visualization Group).[4]

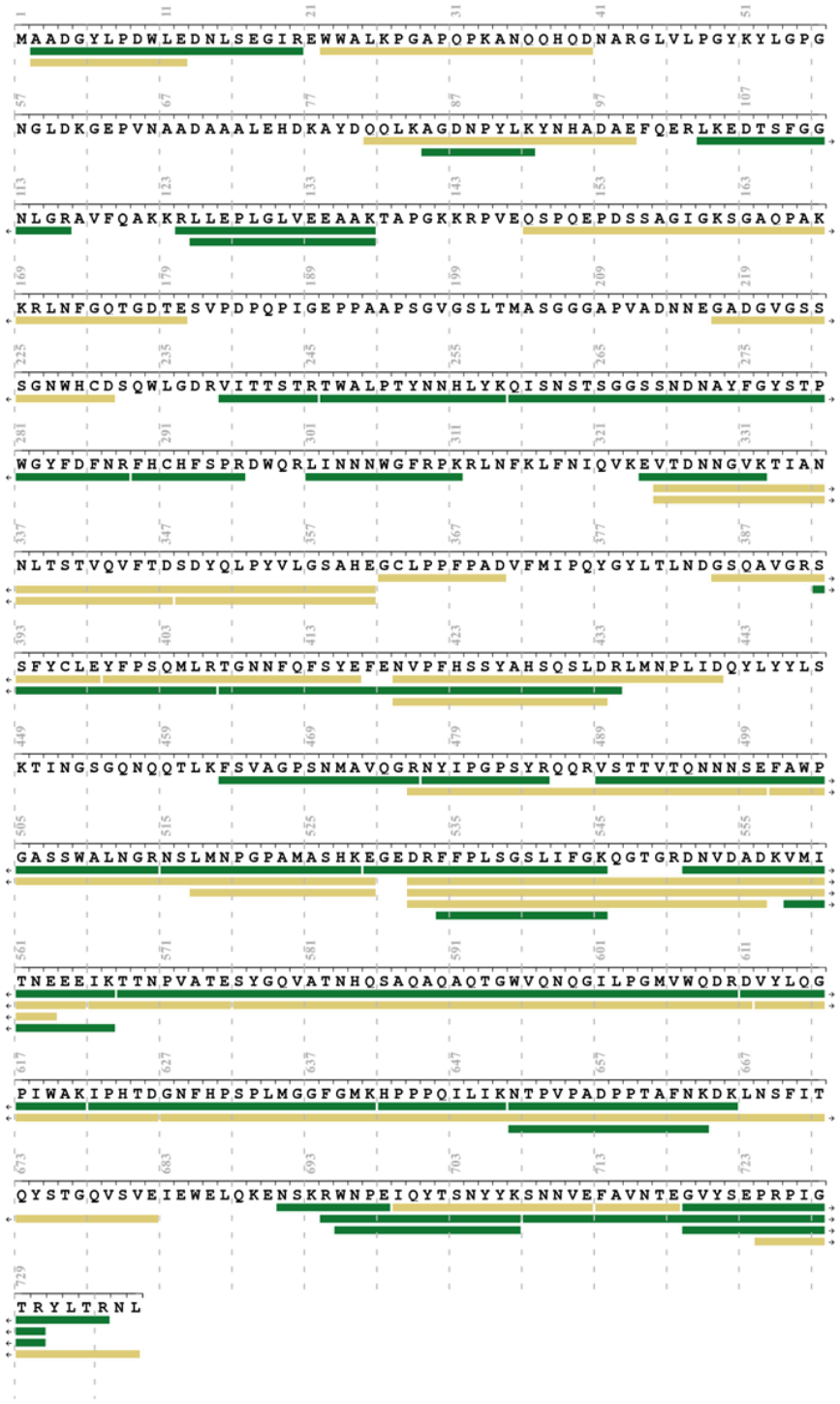


Figure S4. **Protein sequence coverage map of rAAV9 VP1 protein.** The peptide map shows sequence coverage of VP1 (77.7%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualization tool (Aarhus University Visualization Group).[4]

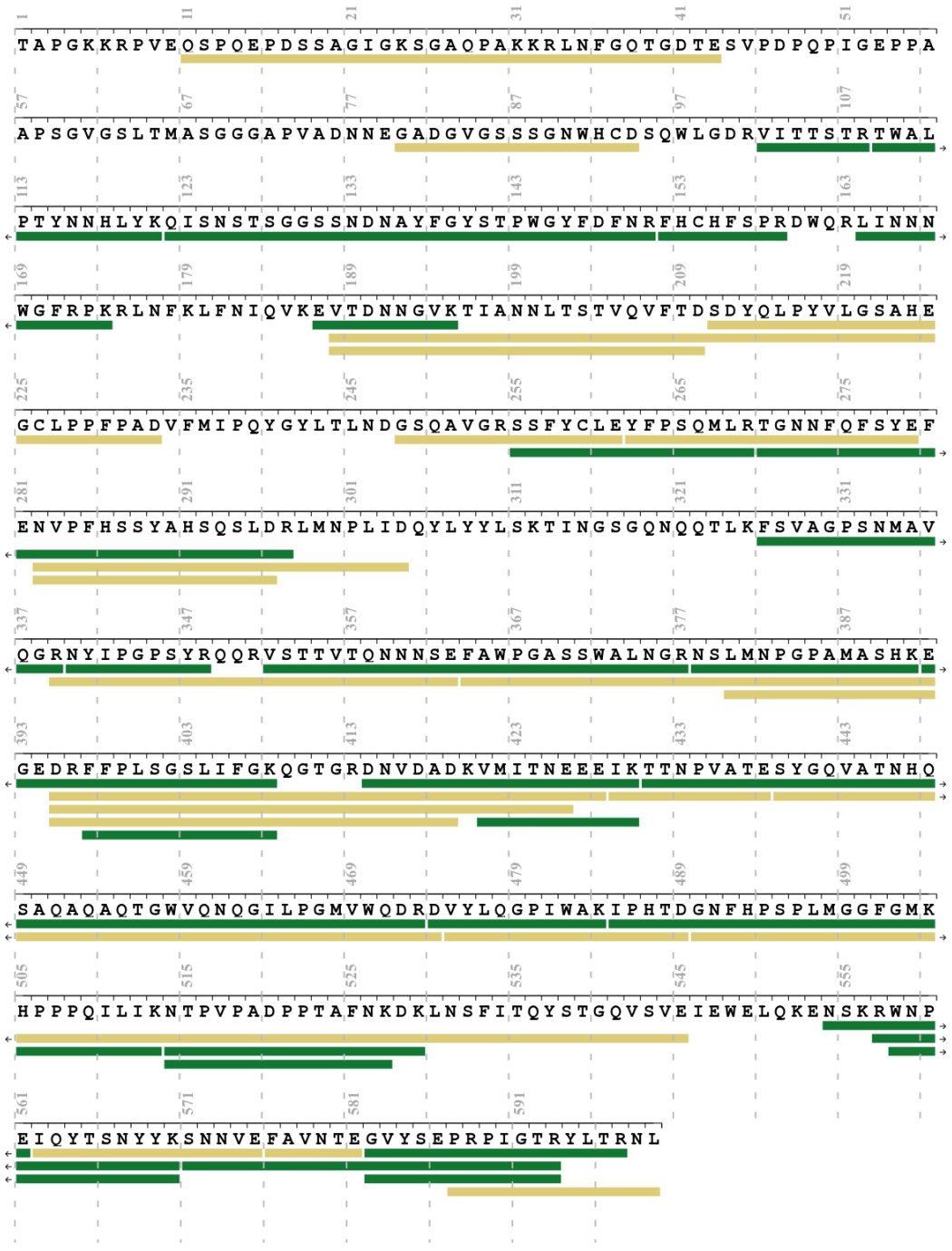
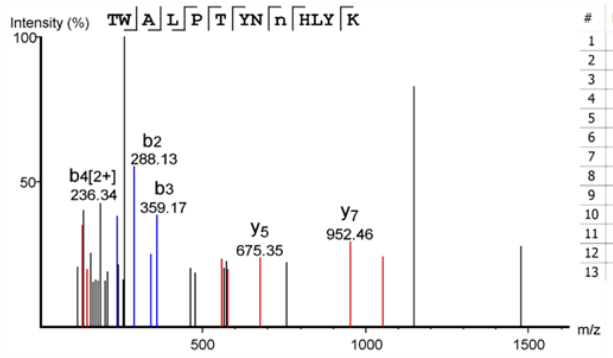


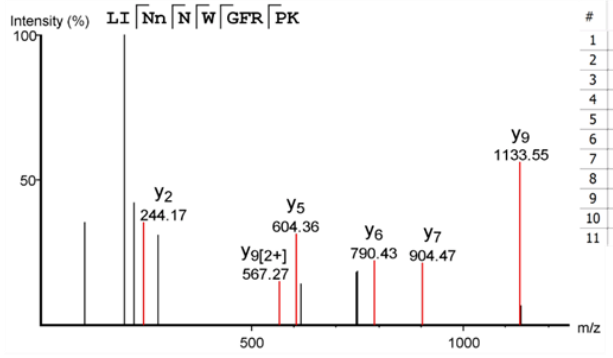
Figure S5. **Protein sequence coverage map of rAAV9 VP2 protein.** The peptide map shows sequence coverage of VP2 (81.5%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualization tool (Aarhus University Visualization Group).[4]

A



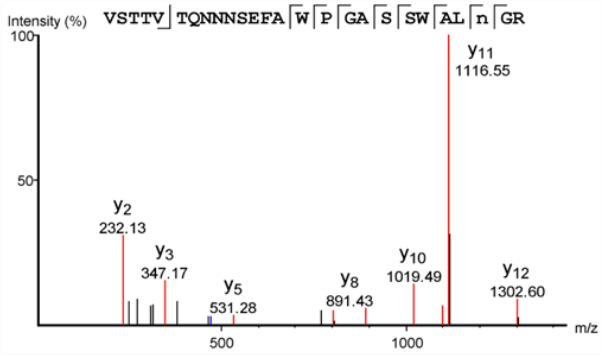
#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	102.06	84.04	85.03	51.53	T				
2	288.13	270.12	271.11	144.57	W	1520.75	1502.74	1503.73	760.88
3	359.17	341.22	342.14	180.09	A	1334.67	1316.66	1317.65	667.84
4	472.26	454.25	455.23	236.34	L	1263.64	1245.63	1246.61	632.32
5	569.31	551.30	552.28	285.15	P	1150.55	1132.54	1133.53	575.78
6	670.36	652.35	653.33	335.68	T	1053.50	1035.49	1036.47	527.25
7	833.42	815.41	816.39	417.21	Y	952.46	934.44	935.43	476.73
8	947.46	929.45	930.44	474.23	N	789.39	771.38	772.36	395.19
9	1062.49	1044.48	1045.46	531.74	N(+.98)	675.35	657.34	658.32	338.17
10	1199.55	1181.54	1182.52	600.27	H	560.31	542.31	543.29	280.66
11	1312.63	1294.62	1295.61	656.82	L	423.26	405.25	406.23	212.13
12	1475.70	1457.69	1458.67	738.35	Y	310.18	292.17	293.15	155.59
13					K	147.11	129.07	130.09	74.06

B



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	114.09	96.08	97.06	57.55	L				
2	227.18	209.17	210.15	114.09	I	1246.63	1228.62	1229.61	623.82
3	341.22	323.21	324.19	171.11	N	1133.55	1115.54	1116.52	567.27
4	456.25	438.24	439.22	228.62	N(+.98)	1019.51	1001.50	1002.48	510.25
5	570.29	552.28	553.26	285.64	N	904.47	886.47	887.45	452.74
6	756.37	738.36	739.34	378.68	W	790.43	772.43	773.41	395.72
7	813.39	795.38	796.36	407.19	G	604.36	586.35	587.33	302.68
8	960.46	942.45	943.43	480.73	F	547.34	529.32	530.31	274.17
9	1116.56	1098.55	1099.53	558.78	R	400.27	382.26	383.24	200.63
10	1213.61	1195.60	1196.58	607.31	P	244.17	226.15	227.14	122.58
11					K	147.11	129.10	130.09	74.06

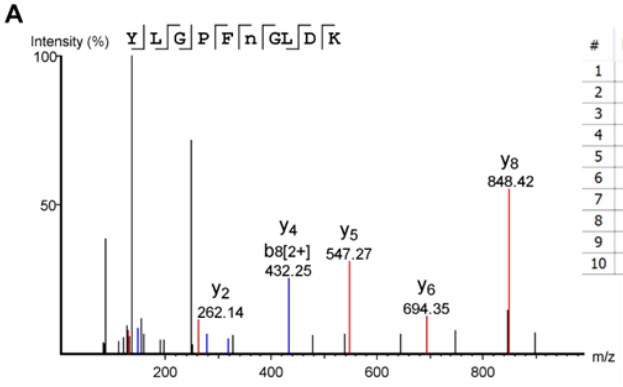
C



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	100.08	82.07	83.05	50.54	V				
2	187.11	169.10	170.08	94.05	S	2696.23	2678.22	2679.21	1348.62
3	288.16	270.15	271.13	144.58	T	2609.20	2591.19	2592.17	1305.10
4	389.20	371.19	372.18	195.10	T	2508.15	2490.14	2491.13	1254.58
5	488.27	470.26	471.25	244.64	V	2407.11	2389.10	2390.08	1204.05
6	589.32	571.31	572.29	295.16	T	2308.04	2290.03	2291.01	1154.52
7	717.38	699.37	700.35	359.19	Q	2206.99	2188.98	2189.96	1103.99
8	831.42	813.41	814.39	416.21	N	2078.93	2060.92	2061.90	1039.97
9	945.46	927.45	928.44	473.23	N	1964.89	1946.88	1947.86	982.94
10	1059.51	1041.50	1042.48	530.25	N	1850.85	1832.84	1833.82	925.92
11	1146.54	1128.53	1129.51	573.77	S	1736.80	1718.79	1719.78	868.90
12	1275.58	1257.57	1258.55	638.29	E	1649.77	1631.76	1632.74	825.39
13	1422.65	1404.64	1405.62	711.83	F	1520.73	1502.72	1503.70	760.86
14	1493.69	1475.68	1476.66	747.34	A	1373.66	1355.65	1356.63	687.33
15	1679.77	1661.76	1662.74	840.38	W	1302.60	1284.61	1285.60	651.81
16	1776.82	1758.81	1759.79	888.91	P	1116.55	1098.51	1099.52	558.77
17	1833.84	1815.83	1816.81	917.42	G	1019.49	1001.48	1002.46	510.25
18	1904.88	1886.87	1887.85	952.94	A	962.47	944.46	945.44	481.73
19	1991.91	1973.90	1974.88	996.45	S	891.43	873.42	874.40	446.22
20	2078.94	2060.93	2061.91	1039.97	S	804.40	786.39	787.37	402.70
21	2265.02	2247.01	2247.99	1133.01	W	717.37	699.36	700.34	359.18
22	2336.06	2318.05	2319.03	1168.53	A	531.28	513.28	514.26	266.14
23	2449.14	2431.13	2432.12	1225.07	L	460.25	442.24	443.22	230.63
24	2564.17	2546.16	2547.14	1282.58	N(+.98)	347.17	329.16	330.14	174.08
25	2621.19	2603.18	2604.16	1311.10	G	232.13	214.13	215.11	116.57
26					R	175.12	157.11	158.09	88.06

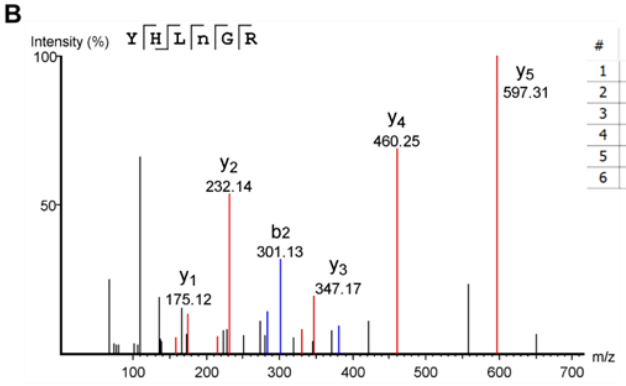
Figure S6. Identification of deamidations in rAAV9 based on MS/MS spectra. (A-C) shows the b and y ions generated from MS/MS fragmentation for identifying (A) the deamidation site at N254, (B) the deamidation at N303 (or N304), (C) the deamidation at N512. Peaks corresponding to b ions are shown in blue, and peaks corresponding to y ions are shown in red. The peptide sequence is displayed at the top of the figure, with b ions indicated by lines extending to the left below the corresponding amino acid and y ions indicated by lines extending to the right above the corresponding amino acid. The inserted tables show measured m/z values of all identified b (blue) and y (red) ions. Images are output of PEAKS v7.5.

A



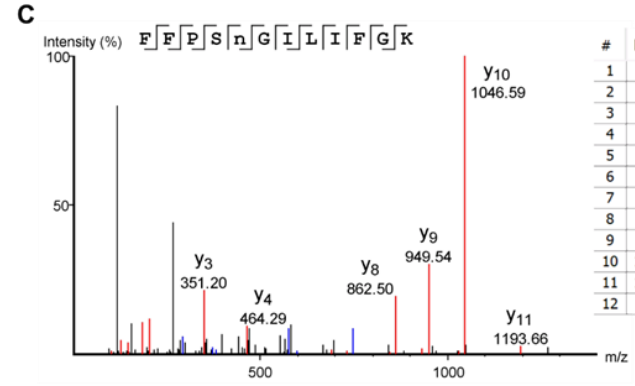
#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	164.07	146.06	147.11	82.54	Y				
2	277.16	259.14	260.13	139.08	L	961.50	943.49	944.47	481.25
3	334.18	316.17	317.60	167.59	G	848.42	830.40	831.39	424.71
4	431.23	413.22	414.20	216.11	P	791.39	773.38	774.37	396.20
5	578.30	560.29	561.27	289.65	F	694.35	676.33	677.31	347.67
6	693.32	675.31	676.30	347.16	N(+.98)	547.27	529.26	530.25	274.14
7	750.35	732.34	733.32	375.67	G	432.25	414.23	415.22	216.62
8	863.43	845.42	846.40	432.25	L	375.22	357.21	358.20	188.11
9	978.46	960.45	961.43	489.73	D	262.14	244.13	245.11	131.57
10					K	147.11	129.10	130.07	74.06

B



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	164.07	146.06	147.04	82.54	Y				
2	301.13	283.12	284.10	151.07	H	597.31	579.30	580.28	299.16
3	414.21	396.20	397.19	207.61	L	460.25	442.24	443.22	230.63
4	529.24	511.23	512.21	265.12	N(+.98)	347.17	329.16	330.14	174.08
5	586.26	568.25	569.24	293.63	G	232.14	214.13	215.11	116.57
6					R	175.12	157.11	158.09	88.06

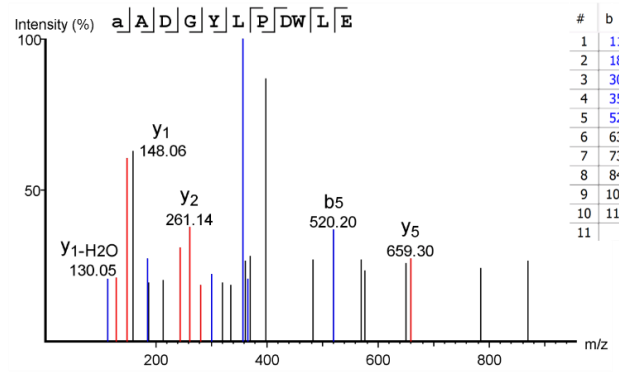
C



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	148.08	130.09	131.05	74.54	F				
2	295.14	277.13	278.12	148.07	F	1193.66	1175.65	1176.63	597.33
3	392.20	374.25	375.17	196.60	P	1046.59	1028.57	1029.56	523.79
4	479.23	461.22	462.20	240.11	S	949.54	931.52	932.51	475.27
5	594.26	576.25	577.37	297.63	N(+.98)	862.50	844.46	845.48	431.75
6	651.28	633.27	634.25	326.14	G	747.48	729.47	730.38	374.25
7	764.36	746.35	747.48	383.16	I	690.46	672.44	673.43	345.73
8	877.45	859.44	860.42	439.22	L	577.37	559.36	560.34	289.19
9	990.53	972.52	973.50	495.77	I	464.29	446.28	447.26	232.64
10	1137.60	1119.59	1120.57	569.30	F	351.20	333.19	334.18	176.10
11	1194.62	1176.61	1177.59	597.83	G	204.13	186.12	187.11	103.06
12					K	147.11	129.10	130.09	74.06

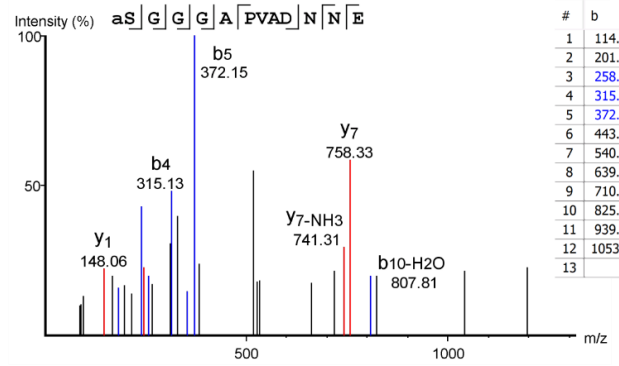
Figure S7. Identification of deamidations in rAAV8 based on MS/MS spectra. (A-C) shows the b and y ions generated from MS/MS fragmentation for identifying (A) the deamidation site at N57, (B) the deamidation at N514, (C) the deamidation at N540. Peaks corresponding to b ions are shown in blue, and peaks corresponding to y ions are shown in red. The peptide sequence is displayed at the top of the figure, with b ions indicated by lines extending to the left below the corresponding amino acid and y ions indicated by lines extending to the right above the corresponding amino acid. The inserted tables show measured m/z values of all identified b (blue) and y (red) ions. Images are output of PEAKS v7.5.

A



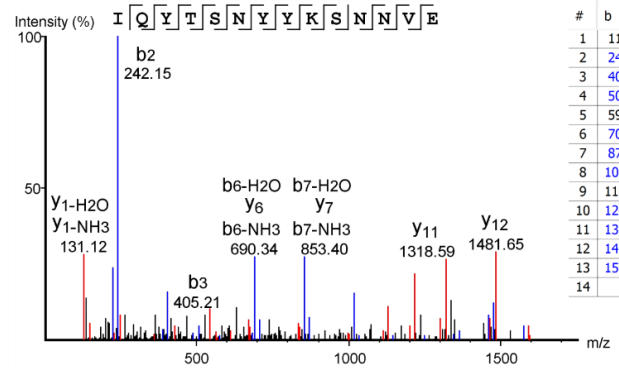
#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	114.05	96.04	97.03	57.53	A(+42.01)				
2	185.09	167.08	168.07	93.05	A	1178.54	1160.53	1161.51	589.77
3	300.12	282.11	283.09	150.56	D	1107.50	1089.49	1090.47	554.25
4	357.14	339.13	340.11	179.07	G	992.47	974.46	975.45	496.74
5	520.20	502.19	503.18	260.60	Y	935.45	917.44	918.42	468.23
6	633.29	615.28	616.26	317.14	L	772.39	754.38	755.36	386.69
7	730.34	712.33	713.31	365.79	P	659.30	641.29	642.28	330.15
8	845.37	827.36	828.34	423.18	D	562.25	544.24	545.22	281.60
9	1031.45	1013.44	1014.42	516.22	W	447.22	429.21	430.20	224.11
10	1144.53	1126.52	1127.50	572.77	L	261.14	243.13	244.09	131.07
11					E	148.06	130.05	131.03	74.53

B



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	114.06	96.04	97.03	57.53	A(+42.01)				
2	201.09	183.08	184.06	101.04	S	1087.46	1069.45	1070.44	544.23
3	258.11	240.10	241.08	129.55	G	1000.43	982.42	983.41	500.72
4	315.13	297.12	298.10	158.07	G	943.41	925.40	926.38	472.21
5	372.15	354.14	355.12	186.58	G	886.39	868.38	869.36	443.70
6	443.19	425.18	426.16	222.09	A	829.37	811.36	812.34	415.18
7	540.24	522.23	523.21	270.62	P	758.33	740.32	741.31	379.67
8	639.31	621.30	622.28	320.16	V	661.28	643.27	644.25	331.14
9	710.35	692.34	693.32	355.67	A	562.21	544.20	545.18	281.61
10	825.37	807.81	808.35	413.19	D	491.17	473.16	474.15	246.09
11	939.42	921.41	922.39	470.21	N	376.15	358.14	359.12	188.57
12	1053.46	1035.45	1036.43	527.23	N	262.10	244.09	245.08	131.55
13					E	148.06	130.05	131.03	74.53

C



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	114.09	96.08	97.06	57.55	I				
2	242.15	224.14	225.12	121.58	Q	1609.71	1591.70	1592.69	805.36
3	405.21	387.20	388.18	203.11	Y	1481.65	1463.63	1464.64	741.33
4	506.26	488.25	489.23	253.63	T	1318.59	1300.59	1301.56	659.80
5	593.29	575.28	576.27	297.15	S	1217.54	1199.55	1200.52	609.76
6	707.34	689.33	690.34	354.17	N	1130.52	1112.50	1113.49	566.22
7	870.41	852.39	853.40	435.70	Y	1016.47	998.44	999.44	508.73
8	1033.47	1015.45	1016.47	517.23	Y	853.40	835.37	836.37	427.20
9	1161.56	1143.55	1144.53	581.28	K	690.34	672.31	673.31	345.67
10	1248.59	1230.58	1231.56	624.80	S	562.25	544.24	545.23	281.62
11	1362.64	1344.61	1345.63	682.31	N	475.21	457.20	458.19	238.11
12	1476.68	1458.67	1459.64	738.84	N	361.17	343.16	344.14	181.09
13	1575.74	1557.73	1558.72	788.37	V	247.13	229.09	230.10	124.06
14					E	148.06	130.05	131.12	74.53

Figure S8. Identification of VP1u, VP3u, and VPc peptides based on their MS/MS spectra. (A-C) shows the b and y ions generated from MS/MS fragmentation for identifying (A) VP1u. (B) VP3u, and (C) VPc. Peaks corresponding to b ions are shown in blue, and peaks corresponding to y ions are shown in red. The peptide sequence is displayed at the top of the figure, with b ions indicated by lines extending to the left below the corresponding amino acid and y ions indicated by lines extending to the right above the corresponding amino acid. The inserted tables show measured m/z values of all identified b (blue) and y (red) ions. Images are output of PEAKS v7.5.

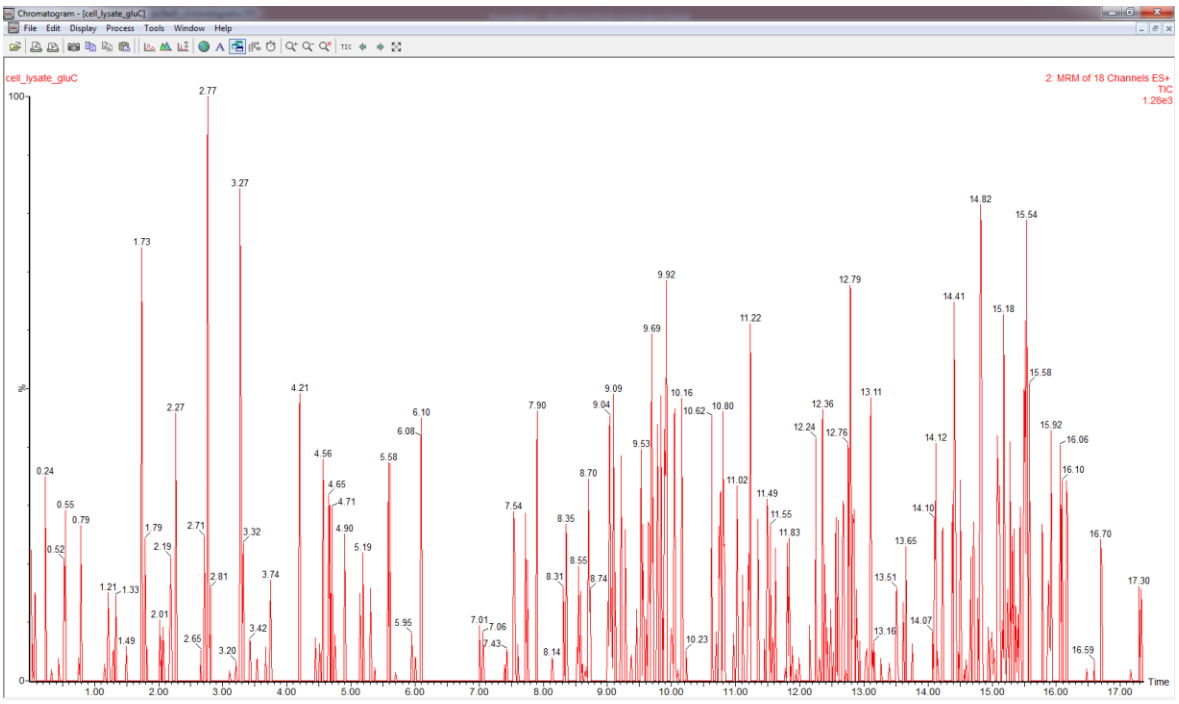


Figure S9. MRM chromatogram of the transitions for signature peptides in a Glu-C digested HEK293 cell lysate. Transitions for both endogenous and isotopically labeled VP1u, VP3u, and VPc peptides (18 transitions in total) were monitored. No specific signals corresponding to the target transitions were detected. All monitored transitions remained within background levels, with a maximum signal intensity of 1.28×10^3.

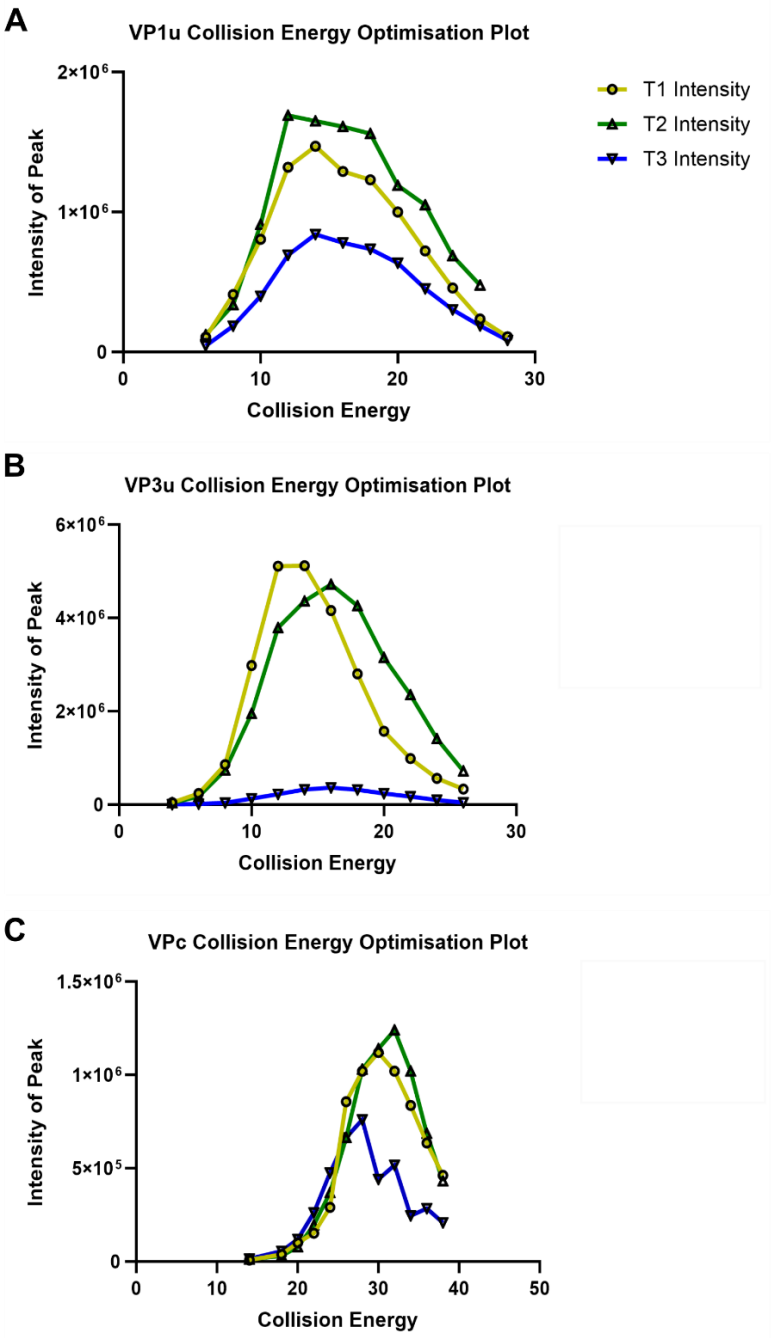


Figure S10. Collision energy (CE) optimization for transitions of rAAV capsid peptides VP1u, VP3u, and VPc. Panels (A), (B), and (C) correspond to the CE optimization plots for VP1u, VP3u, and VPc, respectively. Each plot depicts the intensity trends of three transitions (T1, T2, T3) across 12 tested CEs. The yellow line represents the optimization for T1, the green line for T2, and the blue line for T3. The optimal CE for each transition, yielding the highest intensity, is indicated by the peak of the respective curve.

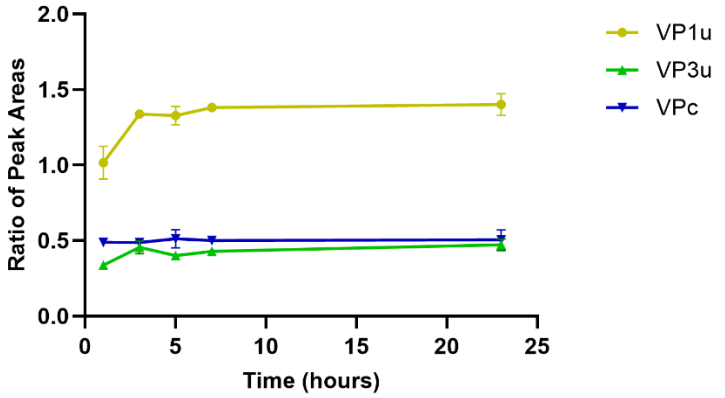


Figure S11. Release of VP1u, VP3u, and VPc peptides from rAAV9 during Glu-C digestion over time. The ratios of the peak areas of the quantifier transitions for each endogenous rAAV VP1u, VP3u, and VPc peptide to a fixed amount of their corresponding isotopically labeled standards were measured at 0, 1, 3, 5, 7, and 23 h post-digestion. The yellow line corresponds to VP1u, the green line to VP3u, and the blue line to VPc. Error bars indicate the SD of three replicate injections. For data points where error bars are not visible, the SD was smaller than the size of the symbol.

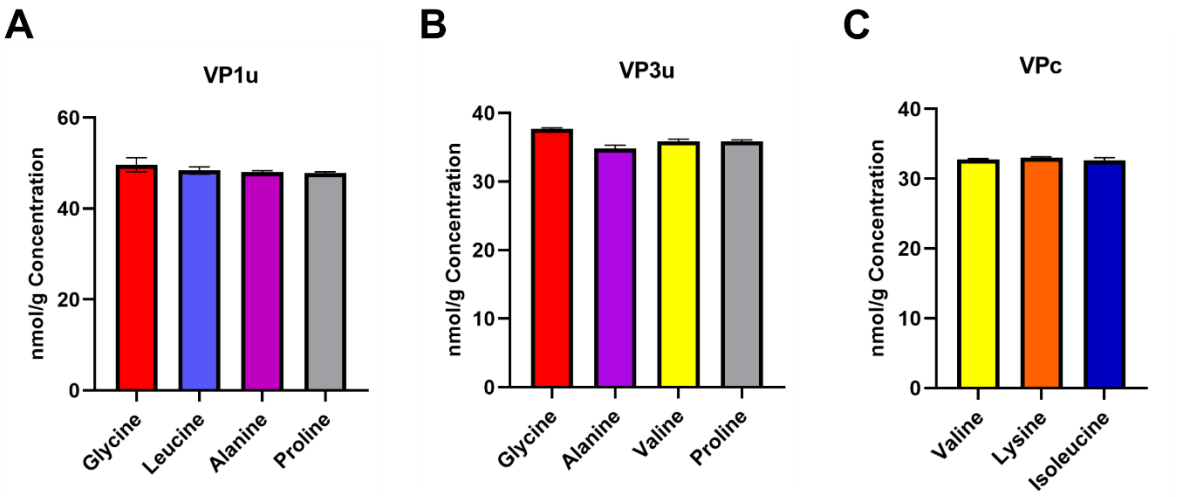


Figure S12. Amino acid analysis of VP1u, VP3u, and VPc synthesized natural peptides with GC-MS/MS MRM. (A) Amino acid analysis of VP1u, showing nmol/g concentrations of glycine (red), leucine (light blue), alanine (purple), and proline (grey). (B) Amino acid analysis of VP3u showing nmol/g concentrations of glycine (red), alanine (purple), valine (yellow), and proline (grey). (C) Amino acid analysis of VPc, showing nmol/g concentrations of valine (yellow), lysine (orange), and isoleucine (dark blue). Error bars correspond to SD measured across three replicate sample blends.

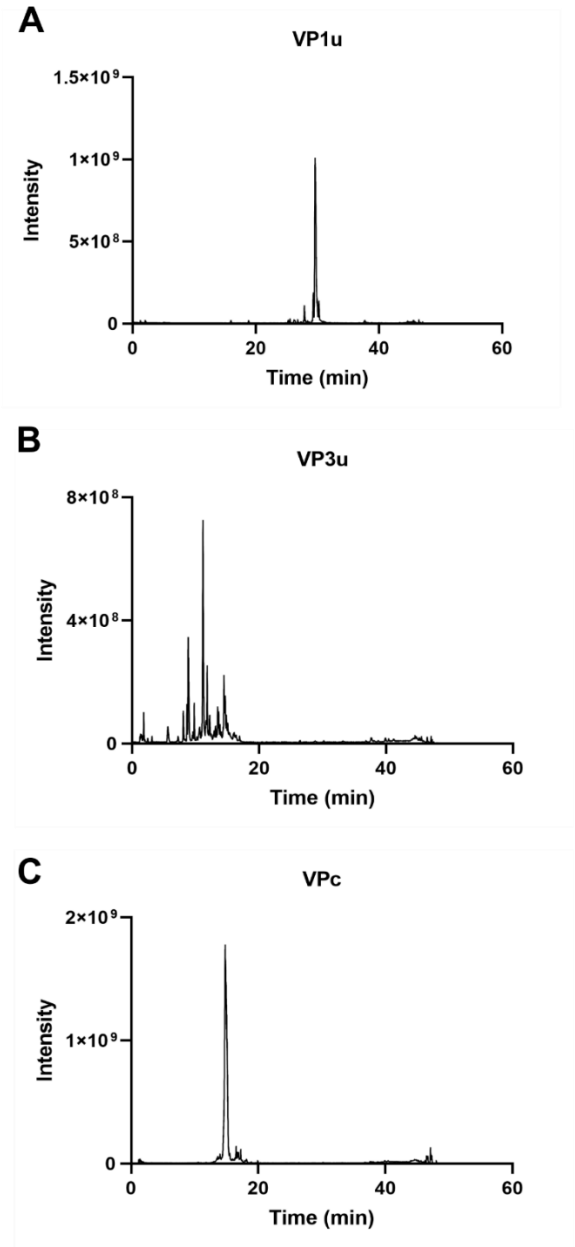


Figure S13. Intact LC-MS analysis of VP1u, VP3u, and VPc synthesized natural peptides. The highest-intensity peak in each chromatogram corresponds to the synthesized natural peptide of interest. (A) Chromatogram of VP1u natural peptide. (B) Chromatogram of VP3u natural peptide. (C) Chromatogram of VPc natural peptide.

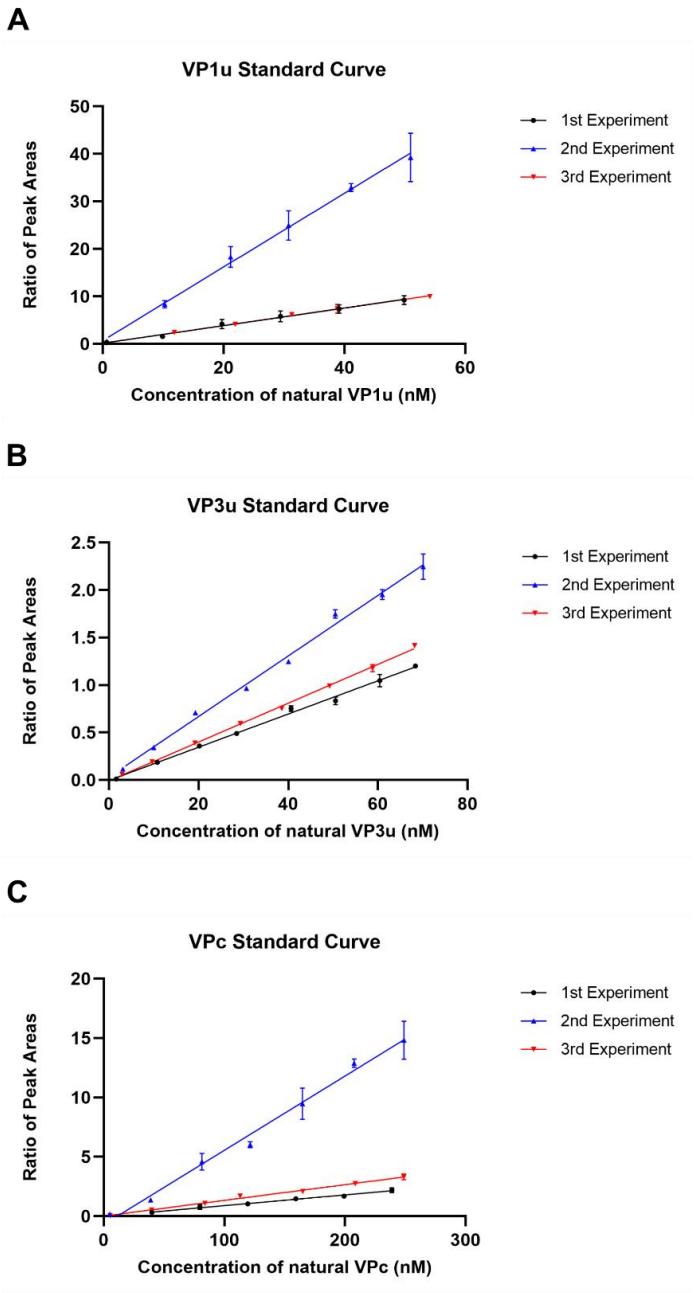


Figure S14. Standard curves for quantification of VP1u, VP3u, and VPc peptides. Panels (A), (B), and (C) display the standard curves for VP1u, VP3u, and VPc, respectively. The y-axis represents the ratio of peak areas between the quantifier transitions of the synthesized natural and labeled peptides, while the x-axis indicates the concentration of the synthesized natural peptide across the standard curve. Each plot includes three standard curves corresponding to experimental replicates, with black, blue, and red lines representing the first, second, and third experiments, respectively. All standard curves have $R^2 > 0.99$. Error bars indicate the SD of three replicate injections of the standard curve. For data points where error bars are not visible, the SD was smaller than the size of the symbol.

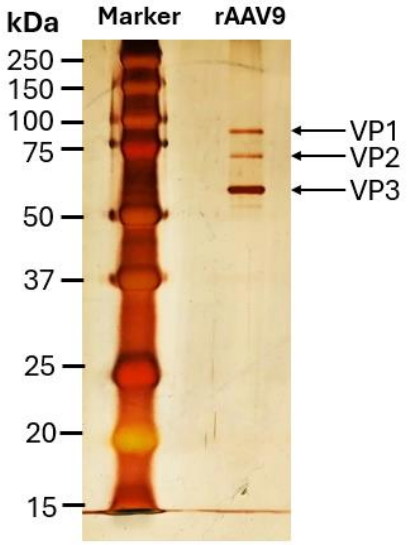


Figure S15. SDS-PAGE analysis of rAAV9 capsid proteins on a 10% polyacrylamide gel. The position of each VP capsid protein (VP1, VP2, VP3) on the gel is indicated. Bands were visualised by silver staining and subsequent densitometry analysis of the VP bands was undertaken using ImageJ. Densitometry analysis showed a ratio of VP1:VP2:VP3 staining intensity of 1:0.7:3.4.

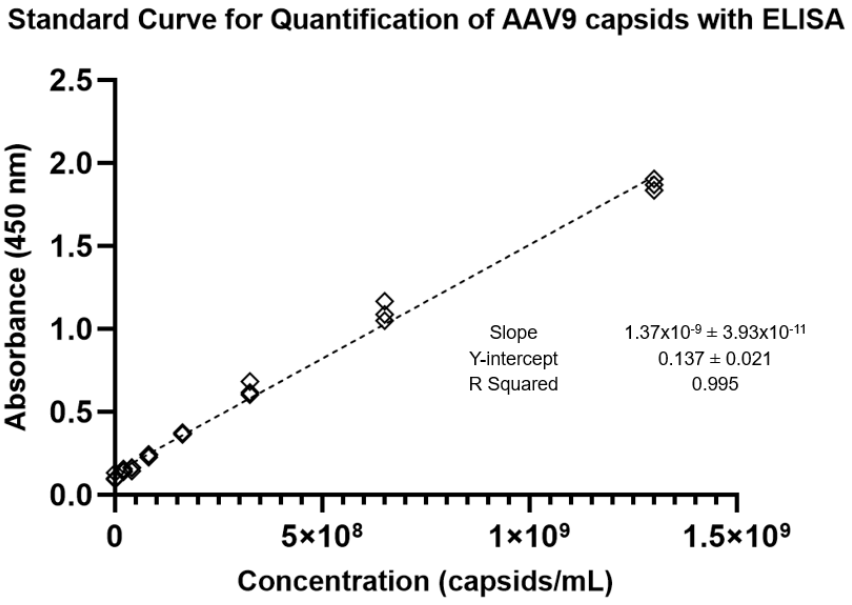


Figure S16. **ELISA standard curve for rAAV9 quantification.** Absorbance values (y-axis) were plotted against known concentrations of rAAV9 capsids obtained from PROGEN (x-axis). All three technical replicates for each level of the standard curve are plotted. Slope, y-intercept, and R^2 values are shown.

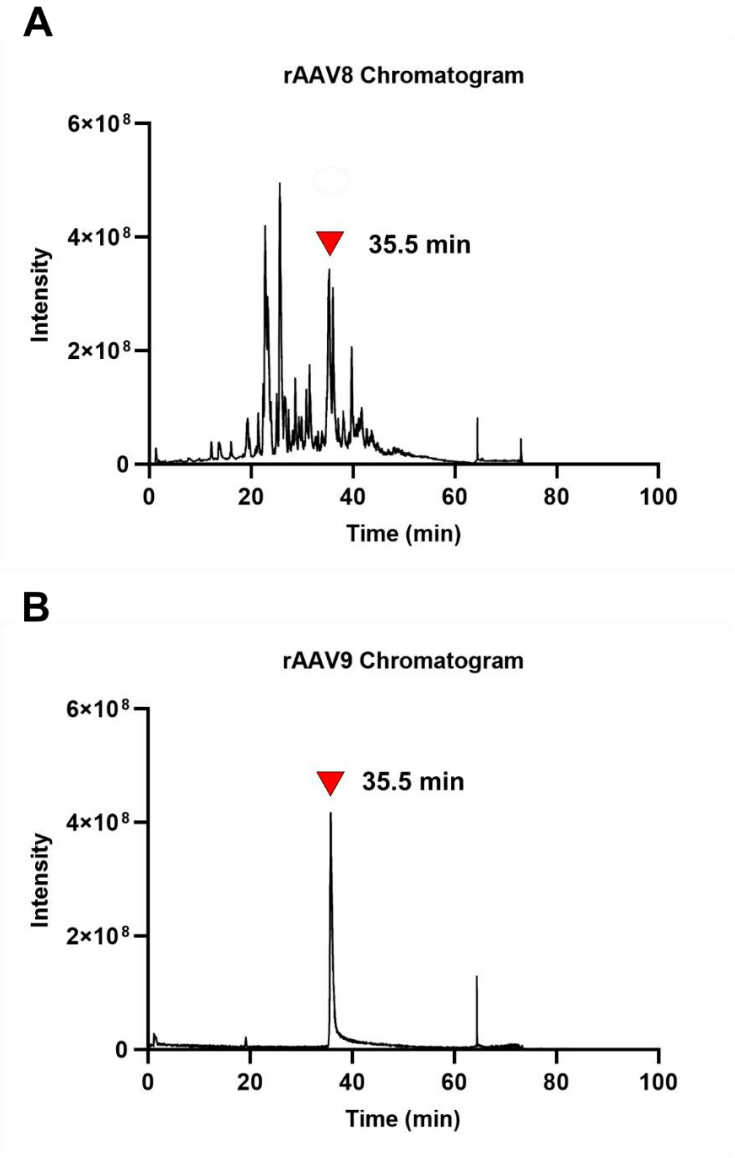


Figure S17. Chromatograms of rAAV8 and rAAV9 samples analyzed with LC-MS. (A) Chromatogram of rAAV8. (B) Chromatogram of rAAV9. The red arrow indicates the retention time at which the capsid proteins VP1, VP2, and VP3 of rAAV8 (A) and rAAV9 (B) were detected. The column flushing step begins at 62 minutes with a transition to 95% solvent B, explaining the appearance of a peak at approximately 65 minutes in both samples.

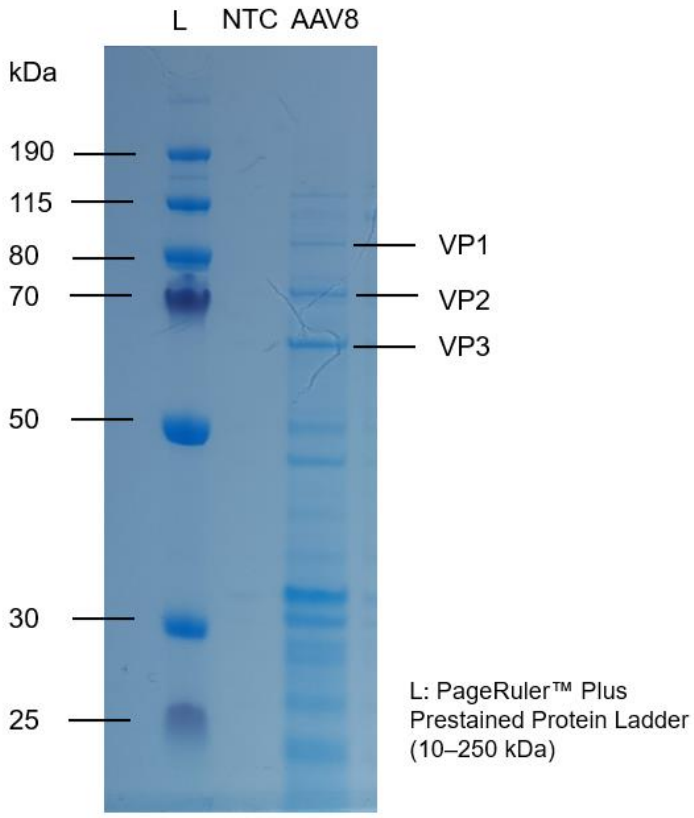


Figure S18. SDS-PAGE gel analysis of rAAV8 preparation. Bands are observed at molecular weights corresponding to the expected sizes of rAAV8 capsid proteins VP1 (81,667 Da), VP2 (66,519 Da), and VP3 (59,805 Da) (as indicated on the figure). Additional bands are present, indicating protein impurities within the sample.

References

1. Torma, A. F., Groves, K., Biesenbruch, S., Mussell, C., Reid, A., Ellison, S., Cramer, R., & Quaglia, M. (2017). A candidate liquid chromatography mass spectrometry reference method for the quantification of the cardiac marker 1-32 B-type natriuretic peptide. *Clin Chem Lab Med (CCLM)*, *55*(9). https://doi.org/10.1515/cclm-2016-1054

2. Zhang, L., Illes-Toth, E., Cryar, A., Drinkwater, G., Di Vagno, L., Pons, M.-L., Mateyka, J., McCullough, B., Achtar, E., Clarkson, C., et al. (2024). A candidate reference measurement procedure for the quantification of α-synuclein in cerebrospinal fluid using an SI traceable primary calibrator and multiple reaction monitoring. *Analyst*, *149*(19), 4842–4850. https://doi.org/10.1039/D4AN00634H

3. Melanson, J. E., Thibeault, M.-P., Stocks, B. B., Leek, D. M., McRae, G., & Meija, J. (2018). Purity assignment for peptide certified reference materials by combining qNMR and LC-MS/MS amino acid analysis results: Application to angiotensin II. *Anal Bioanal Chem*, *410*(26), 6719–6731. https://doi.org/10.1007/s00216-018-1272-7

4. Nielsen, S. D.-H., Liang, N., Rathish, H., Kim, B. J., Lueangsakulthai, J., Koh, J., Qu, Y., Schulz, H.-J., & Dallas, D. C. (2024). Bioactive milk peptides: An updated comprehensive overview and database. *Crit Rev Food Sci Nutr*, *64*(31), 11510–11529. https://doi.org/10.1080/10408398.2023.2240396