

Next-Gen Protein Sequencing™ Enables Accurate AAV Serotype Profiling and Mixture Resolution for Gene Therapy

HIGHLIGHTS



High-resolution AAV serotype identification: NGPS accurately distinguishes AAV8 from AAV9 with single-molecule precision, providing insights beyond traditional nucleic acid or immunoassays.



Critical role of sample preparation: RapiGest SF, a gentle denaturant, greatly enhances AAV capsid protein digestion in the sample prep process, which is vital for effective downstream NGPS analysis.



Sensitivity for cross-contamination detection: Detecting AAV8 and AAV9 mixtures at 10% abundance, with low false discovery rates, makes NGPS a powerful tool for monitoring the purity of AAV and preventing cross-contamination in manufacturing environments.



Potential for gene therapy quality control: NGPS offers label-free AAV serotype verification, confirming capsid identity at a molecular level with potential for broader applications as the technology evolves.

SUMMARY

Adeno-associated viruses (AAVs) are widely used as delivery vectors in clinical gene therapy, with each serotype exhibiting unique capsid protein sequences that influence tissue tropism, transduction efficiency, and immune response. Ensuring the identity and purity of AAV preparations is essential during manufacturing and quality control, particularly when closely related serotypes are involved. In this study, we demonstrate the use of Next-Gen Protein Sequencing™ (NGPS™) to identify AAV8 and AAV9 capsid proteins by detecting three serotype-specific peptides capturing amino acid differences at position 105, 315, 539, and 540. NGPS enabled reliable identification of each serotype and unambiguous detection of each serotype from a mixture, with sensitivity sufficient to detect AAV8 and AAV9 at levels of 10% in a mixture based on distinct kinetic signatures. These results highlight the utility of NGPS as a high-resolution analytical tool for AAV serotype characterization and monitoring of contamination.

INTRODUCTION

Gene therapy is an emerging therapeutic strategy that involves altering a patient's genetic material to treat or reverse disease. One of the most widely used delivery vehicles for gene therapy is the adeno-associated virus (AAV), a non-pathogenic parvovirus that can infect both dividing and non-dividing cells. The FDA has approved two AAV gene therapies: Spark Therapeutics' *Luxturna* for retinal dystrophy in 2017, and Novartis' *Zolgensma* for spinal muscular atrophy in 2019, which was

supported by the AFM Téléthon (the French Muscular Dystrophy Association). There are currently 359 candidates using AAV technology to deliver genes to tissues for 133 diseases.¹ AAV vectors are favored for their safety, long-term gene expression potential, and ability to deliver therapeutic genes to a variety of tissues.^{2,3} Their growing role in clinical pipelines has increased the demand for reliable tools to ensure their identity, purity, and quality throughout development and manufacturing.

AAVs exist in multiple naturally occurring and engineered serotypes, each defined by specific capsid protein structures that determine tissue tropism. For example, AAV1 and AAV8 are known for efficient transduction of skeletal and cardiac muscle, while AAV2 and AAV9 are more commonly used for targeting liver and central nervous system tissues. Because the capsid sequence directly affects biodistribution, immunogenicity, and transduction efficiency,^{2,4} choosing the correct serotype is critical for therapeutic success. Consequently, accurate characterization of the AAV capsid is essential for preclinical development, clinical translation, regulatory compliance, and quality control.

While nucleic acid-based methods such as qPCR or next-generation DNA sequencing are commonly used to confirm genome integrity and vector identity, they provide limited information about the capsid proteins that ultimately determine serotype. These methods cannot distinguish between empty capsids, non-functional particles, or subtle protein-level differences between closely related serotypes. Protein-based assays such as SDS-PAGE and ELISA lack the resolution to detect minor sequence variants or low-level contamination. Although mass spectrometry offers high-resolution characterization, it requires specialized instrumentation and technical expertise, making it less accessible for routine or high-throughput serotype verification.

NGPS offers a novel approach for direct, single-molecule resolution analysis of AAV capsid proteins, providing detailed insights into their amino acid sequences and post-translational modifications (PTMs).⁴ This high level of specificity allows for accurate identification of serotype-defining sequence features, even among closely related variants. NGPS also offers the sensitivity needed to detect low-level contaminants that may be missed by traditional protein or nucleic acid-based methods.

In this application note, we demonstrate the advantage of NGPS to distinguish closely related AAV serotypes based on their capsid protein sequences. Specifically, we show that AAV8 and AAV9 – two serotypes with high sequence similarity – can be reliably differentiated through four key peptide-level differences. Using these distinguishing example peptides, we further demonstrate that NGPS can detect the low-level presence of AAV8 in an AAV9 background, and vice versa, at levels as low as 10%. These results highlight the power of NGPS for both confident serotype identification and monitoring cross-contamination in AAV production workflows.

MATERIALS AND METHODS

AAV SAMPLE PREPARATION

AAV8 and AAV9 protein samples were provided by Génethon, Évry, France, for technical evaluation.

LIBRARY PREPARATION

AAV samples were processed using Quantum-Si Library Preparation Kit – Lys-C V2 (catalog number 910-00012-02) following its protocol with minor modifications. Briefly, the samples were diluted to 90 µL of 0.25 µg/µL using Quantum-Si sample buffer. RapiGest SF surfactant (Waters catalog number 186002123) was reconstituted in sample buffer to a final concentration of 1%. 90 µL of each sample (0.25 µg/µL) was mixed with 10 µL of either sample buffer or 1% RapiGest

solution. The samples were reduced with 2 μ L of TCEP at 37°C for 30 minutes, followed by alkylation with 2 μ L CAA at room temperature for 30 minutes. Digestion was initiated by adding 2 μ L of rehydrated Lys-C (0.5 μ g/ μ L), and samples were incubated overnight (16–18 hours) at 37°C.

Following LysC digestion, samples underwent the diazo-transfer process. The samples were incubated with 6 μ L of K₂CO₃, 13 μ L of CuSO₄, and 2 μ L of ISA for 90 minutes at room temperature. The entire volume of each sample was then transferred to a beads tube for quenching and incubated at room temperature for 30 minutes with end-to-end mixing. Post-quenching, samples were filtered using the provided spin columns, and 6 μ L of acetic acid was added to adjust the pH. Linker conjugation was performed by adding 2 μ L each of EDTA, CTAB, and K-linker to 47 μ L of each sample. The reaction was incubated at 37°C overnight (16–18 hours).

NEXT-GEN PROTEIN SEQUENCING ON PLATINUM® PRO

After conjugation, the samples were taken out from 37°C and kept on ice until sequencing on Quantum-Si's Platinum Pro instrument, following the Quantum-Si Sequencing Kit V4 (catalog number 910-00038-04) protocol. All samples were loaded at 0.4 nM via full-chip experiments (which use both sides of the chip). The data was analyzed using *Primary Analysis* V2.14.0 and *Peptide Alignment* V2.15.0. The standard alignment score threshold of 5 was used for peptides of standard length (≥ 6 amino acids), while an alignment score threshold of 3.75 was used for short peptides (≤ 5 amino acids).

SDS-PAGE ANALYSIS

3 μ L of pre- and post-digestion samples were prepared in Novex™ Tris-Glycine SDS Sample Buffer 2X (Thermo Fisher Scientific catalog number LC2676), heated at 95°C for 5 minutes, and loaded onto a 4–20% Novex™ Tris-Glycine Mini Protein Gel (Thermo Fisher Scientific catalog

number XP04200BOX). The gel was run in Novex™ Tris-Glycine SDS Running Buffer (Thermo Fisher Scientific catalog number LC2675) at 200 V for 55 minutes. To visualize gel bands, the staining and destaining procedure was performed using SimplyBlue™ SafeStain (Thermo Fisher Scientific catalog number LC6060), following the manufacturer's microwave protocol.

RESULTS AND DISCUSSION

RAPIGEST ENABLES EFFECTIVE DIGESTION OF AAV SAMPLES

To evaluate digestion efficiency of AAV capsid proteins, 100 μ L of AAV8 and AAV9 samples at a concentration of 0.225 μ g/ μ L in sample buffer (from Quantum-Si Library Preparation Kit – Lys-C V2) were processed using its protocol with minor modifications (see *Materials and Methods*). Samples were reduced with TCEP, alkylated with CAA, and digested with Lys-C. SDS-PAGE analysis of pre- and post-digestion samples revealed no observable differences in banding patterns (Figure 1), indicating that the capsid proteins remained largely intact and resistant to Lys-C enzymatic cleavage.

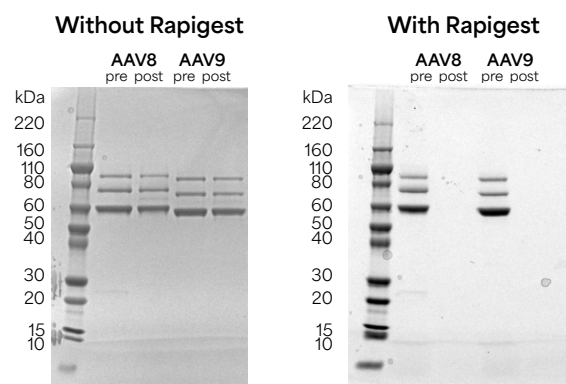


Figure 1. RapiGest improves the digestion efficiency of AAV capsid proteins. AAV8 and AAV9 samples were either left untreated or treated with RapiGest prior to Lys-C digestion and analyzed by SDS-PAGE analysis. RapiGest-treated samples exhibit complete digestion, as indicated by the absence of intact protein bands in the post-digestion lanes, demonstrating enhanced proteolytic efficiency.

Digestion of AAV protein samples with the Lys-C enzyme was significantly enhanced by the addition of RapiGest, a mild denaturant that makes proteins more amenable to solubilization and unfolding. RapiGest was added to a final concentration of 0.1% prior to the digestion workflow. The samples were subjected to TCEP reduction, CAA alkylation, and Lys-C digestion, followed by SDS-PAGE analysis. In contrast to the untreated condition, RapiGest-treated samples showed complete

disappearance of the intact protein bands post-digestion (Figure 1), confirming effective proteolysis. These findings suggest that AAV capsid proteins require mild denaturation for effective digestion. The inclusion of 0.1% RapiGest in the workflow facilitates the efficient digestion of AAV samples, making it a critical step for downstream NGPS applications.

NGPS ENABLES SEROTYPE DISCRIMINATION OF AAV8 AND AAV9 THROUGH RESIDUES 105, 315, 539, AND 540 OF CAPSID PROTEINS

AAV8 and AAV9 vary at several positions within their VP1 capsid protein sequences. Specifically, residues 105, 315, 539, and 540 act as four distinct, serotype-defining sequences that can be accurately identified using NGPS. At position 105, AAV8 encodes glutamine (Q), while AAV9 encodes lysine (K) (Figure 2A). The presence of lysine in AAV9 introduces a unique Lys-C cleavage site, generating a peptide beginning with the sequence EDT that is absent from AAV8 digests.

To determine if NGPS could identify this difference, Lys-C-digested samples were functionalized, sequenced on Platinum Pro, and data analyzed using the *Peptide Alignment* workflow against a combined AAV8/AAV9 reference file. NGPS confidently detected the EDT peptide in the AAV9 sample, with 147 alignments, a false discovery rate (FDR) of 1% (Figure 2B), and a distinct kinetic signature (Figure 2C). In contrast, the AAV8 sample produced only seven alignments for the same peptide with an FDR of 43% (Figure 2B). These results demonstrate that NGPS can distinguish AAV9 from AAV8 by detecting the unique EDT peptide resulting from cleavage at residue 105.

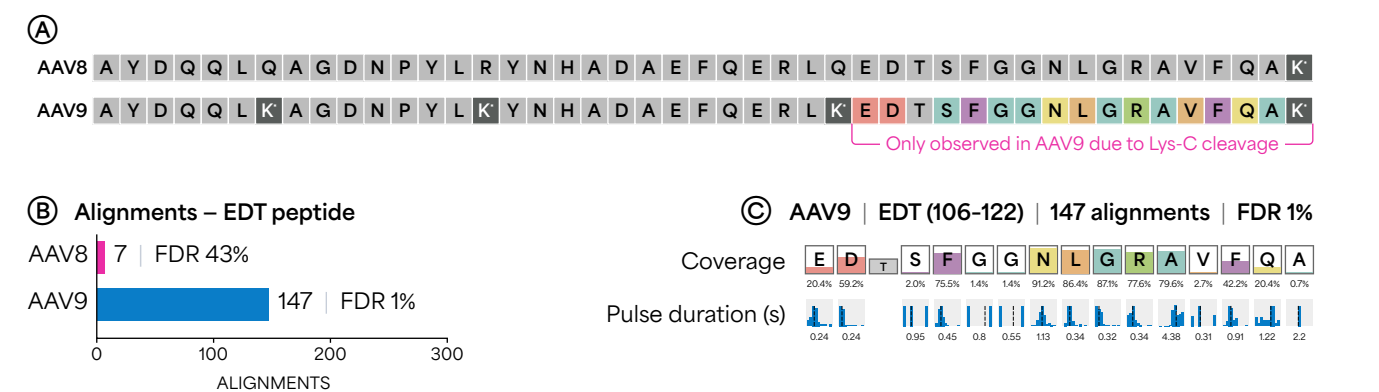


Figure 2. Identification of a serotype-specific peptide resulting from cleavage after residue 105 in AAV capsids. (A) Schematic showing residue 105 in the AAV capsid protein sequence: glutamine (Q) in AAV8 versus lysine (K) in AAV9. The presence of lysine in AAV9 enables Lys-C cleavage and generation of the AAV9 serotype-specific EDT peptide. (B) *Peptide Alignment* analysis reveals high alignment counts and low false discovery rate (FDR) for the EDT peptide in AAV9 samples, with no significant detection in AAV8. (C) Kinetic signature of the EDT peptide in AAV9 shows coverage and pulse duration values of expected residues across the peptide sequence.

The second key serotype-specific difference that NGPS can resolve is residue 315 of the AAV capsid protein. At this position, AAV8 encodes a serine (S), while AAV9 encodes an asparagine (N). This single-residue substitution gives rise to two distinct peptides following Lys-C digestion: RLS for AAV8 and RLN for AAV9 (Figure 3A). In the AAV8 sample, NGPS consistently identified 219 alignments for the RLS peptide with an FDR of 0%, indicating high confidence (Figure 3B).

In contrast, RLN was not confidently detected (1 alignment, FDR 100%). Conversely, the AAV9 sample revealed 201 confident alignments for RLN with an FDR of 1%, while RLS was not confidently detected (1 alignment, FDR 100%). Analysis of the kinetic signatures of these peptides confirmed the S315N substitution in AAV9 (Figure 3C). These results demonstrate that the distinct kinetic signatures of the RLS and RLN peptides can be used to confidently differentiate AAV8 and AAV9.

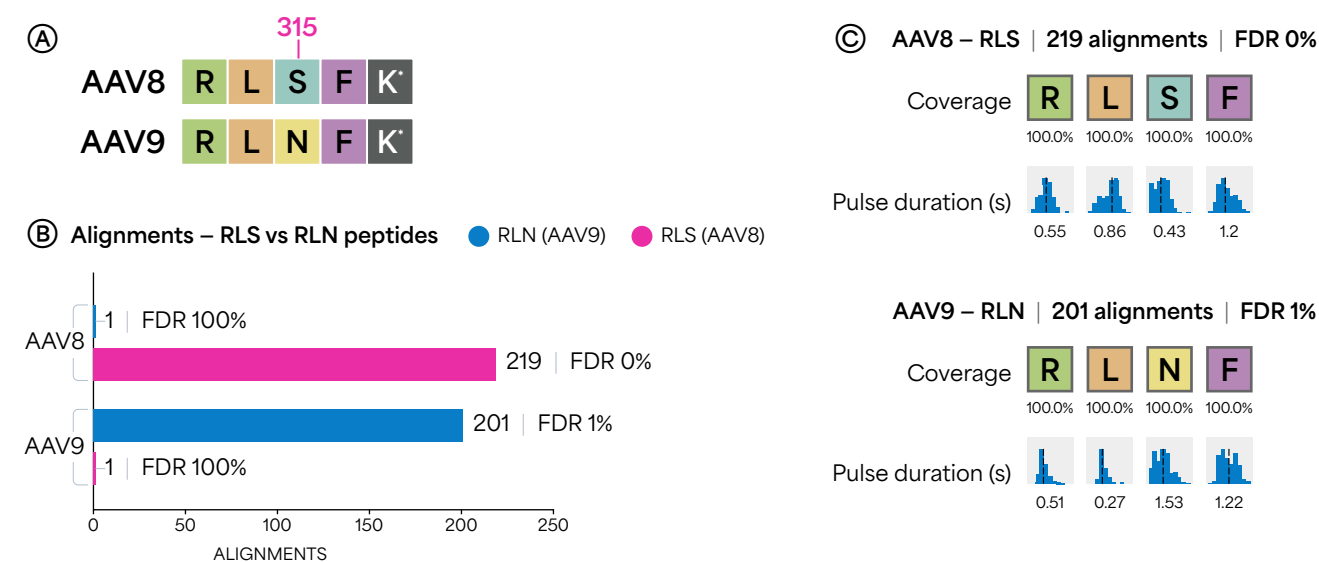


Figure 3. Serotype-specific differentiation based on residue 315 in AAV capsid proteins. (A) Schematic highlighting residue 315: serine (S) in AAV8 and asparagine (N) in AAV9, resulting in RLS and RLN peptide variants, respectively. (B) *Peptide Alignment* analysis reveals high alignment counts and low FDR for RLS in AAV8 and RLN in AAV9, with no significant detection of the other serotype. (C) Kinetic signatures of the RLS and RLN peptides in AAV8 and AAV9, respectively, showing coverage and pulse duration values of expected residues across the peptide sequences.

The final set of serotype-specific substitutions occurs at residues 539 and 540, where AAV8 encodes serine (S) and asparagine (N), while AAV9 encodes leucine (L) and serine (S), respectively. Lys-C digestion produces the peptides DDE in AAV8 and EGE in AAV9 (Figure 4A). While these peptides also span additional substitutions not detected by NGPS, the current analysis focuses on the two positions that can be confidently resolved. In the AAV8 sample, NGPS identified 331 alignments for the DDE peptide with an FDR of 1%, and only 4 alignments for the EGE peptide with an FDR of 75% (Figure 4B).

Meanwhile, the AAV9 sample yielded 136 alignments for the EGE peptide with an FDR of 1%, and only 2 alignments for the DDE peptide with an FDR of 100%. Analysis of the kinetic signatures of these peptides confirmed the presence of the S539L and N540S substitutions (Figure 4C). This result, along with the EDT peptide at residue 105 and the RLS/RLN peptide at residue 315, offers orthogonal evidence that supports serotype identification and discrimination through NGPS.

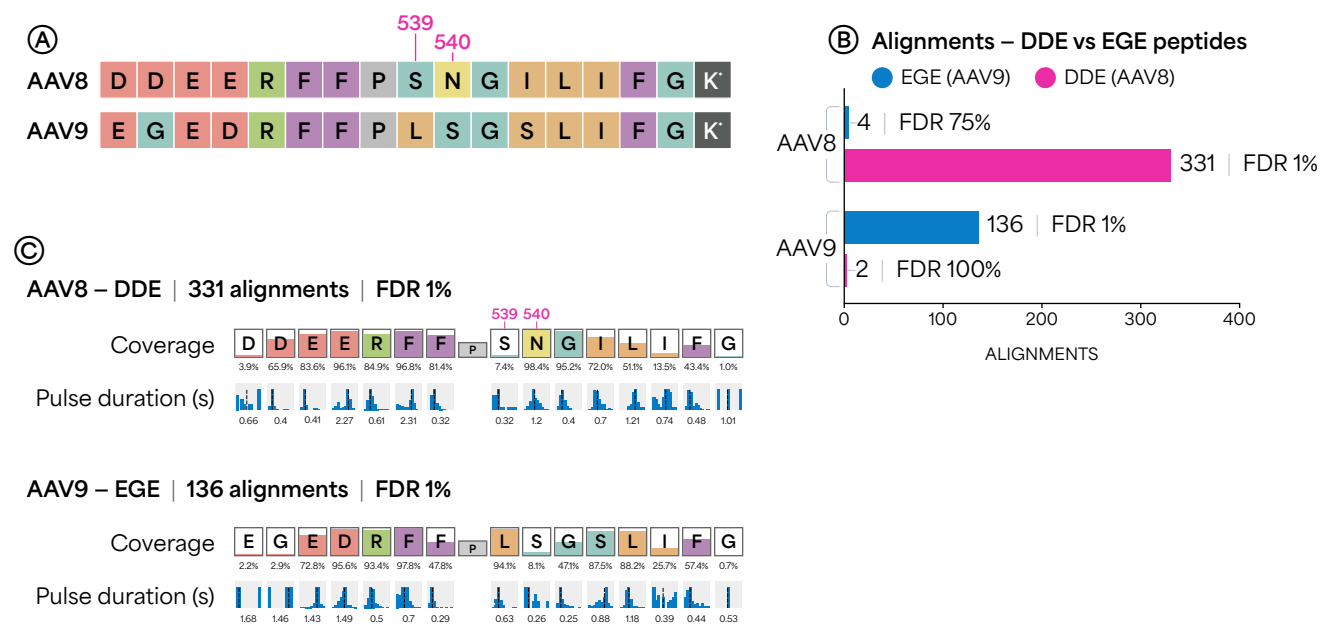


Figure 4. Serotype-specific differentiation based on residues 539 and 540 in AAV capsid proteins. (A) Schematic highlighting residues 539 and 540: serine/asparagine (SN) in AAV8 and leucine/serine (LS) in AAV9, resulting in DDE and EGE peptide variants, respectively. (B) *Peptide Alignment* analysis reveals high alignment counts and low FDR for DDE in AAV8 and EGE in AAV9, with no significant detection of the other serotype. (C) Kinetic signatures of the DDE and EGE peptides in AAV8 and AAV9, respectively, showing coverage and pulse duration values of expected residues across the peptide sequences.

NGPS ENABLES DETECTION OF AAV8 AND AAV9 MIXED AT 10% ABUNDANCE LEVEL

Cross-contamination between AAV serotypes during production or purification can compromise the safety and efficacy of gene therapy products.⁶ Detecting such low-level contaminants is essential for ensuring product purity and regulatory compliance. To evaluate the sensitivity of NGPS for detecting AAV cross-contamination, we designed a dilution series consisting of four samples: pure AAV8, pure AAV9, and AAV8:AAV9 mixtures at 10:1 and 1:10 ratios. NGPS data were analyzed using the *Peptide Alignment* workflow, focusing on three serotype-specific peptide signatures: EDT (derived from residue 105), RLS/RLN (derived from residue 315), and DDE/EGE (derived from residues 539 and 540).

For the EDT peptide (Q105K substitution), we observed a clear trend in detection across the AAV8:AAV9 mixtures (Figure 5A). In the pure AAV9 sample, NGPS detected 147 alignments with an FDR of 1%, reflecting confident identification of the AAV9-specific signal. The AAV8:AAV9 1:10 mixture also showed strong detection, with 237 alignments and an FDR of 2%. At lower abundance, the AAV8:AAV9 10:1 mixture yielded 25 alignments with an FDR of 4%, indicating that NGPS can detect AAV9 at 10% abundance in an AAV8 background. By contrast, the pure AAV8 sample produced negligible detection of EDT peptide, with only 7 alignments at an FDR of 43%.

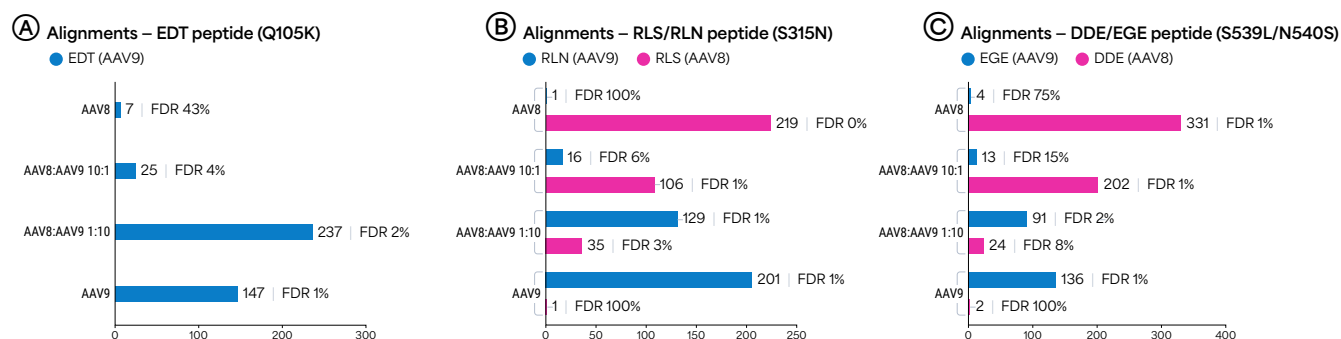


Figure 5. NGPS detection of AAV cross-contamination using serotype-specific peptides. Detection performance across AAV8:AAV9 dilution series for (A) EDT peptide (AAV9-specific), (B) RLS and RLN peptides (AAV8- and AAV9-specific, respectively), and (C) DDE and EGE peptides (AAV8- and AAV9-specific, respectively). Bar graphs show number of alignments with corresponding false discovery rates (FDR) indicated besides each bar. All peptides successfully detected their respective serotypes even at 10% abundance levels with good statistical confidence, demonstrating the sensitivity of NGPS for identifying low-level AAV cross-contamination.

For the S315N substitution, the RLS peptide showed strong selectivity for AAV8, with pure AAV8 samples resulting in 219 alignments at an FDR of 0% (Figure 5B). Notably, even when AAV8 made up only 10% of the AAV8:AAV9 1:10 mixture, 35 alignments were still detected at a good FDR of 3%, demonstrating NGPS's sensitivity for identifying low-level AAV8 contamination. Conversely, the RLN peptide showed opposite selectivity, with pure AAV9 samples generating 201 alignments at an FDR of 1% (Figure 5B). When AAV9 was only 10% of the AAV8:AAV9 10:1 mixture, 16 alignments were detected at a 6% FDR, confirming that NGPS reliably detects low levels of contamination.

Finally, for the S539L/N540S substitutions, the DDE peptide exhibited strong AAV8 selectivity, with pure AAV8 samples producing 331 alignments at an FDR of 1% (Figure 5C). Even at 10% abundance in the AAV8:AAV9 1:10 mixture, DDE remained detectable with 24 alignments at an FDR of 8%, demonstrating consistent sensitivity for low-level AAV8 contamination across multiple peptides. Conversely, the EGE peptide demonstrated clear AAV9 specificity, generating 136 alignments with an FDR of 1% in pure AAV9 samples (Figure 5C). Notably, even when AAV9 comprised only 10% of the AAV8:AAV9 10:1 mixture, EGE was detected with 13 alignments at a good FDR of 15%.

The three peptide pairs (EDT, RLS/RLN, and DDE/EGE) show that NGPS can effectively detect AAV cross-contamination at 10% abundance levels with high confidence. Conducting validation experiments with multiple replicates can help establish reliable criteria for limits of detection. Additionally, there is potential to create a combined metric from each peptide into an overall detection score.

CONCLUSION

This application note highlights NGPS's ability to effectively and precisely differentiate between AAV8 and AAV9 capsid proteins by analyzing amino acid differences at specific positions. We confidently distinguished the two serotypes and detected their presence as low as 10%. This demonstrates NGPS's capability for high-resolution, sequence-specific analysis of AAV samples without relying on traditional methods. Identifying contaminant serotypes showcases the platform's reliability for quality control in AAV-based gene therapy production. As Quantum-Si's NGPS technology evolves, it may reveal additional serotype-specific peptides, further enhancing AAV characterization and contamination detection in gene therapy vector development and production.

REFERENCES

1. <https://www.airfinity.com/articles/the-aav-gene-therapy-regulatory-landscape>
2. Nonnenmacher M, Weber T. Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther.* 2012;19(6):649-658. doi:10.1038/gt.2012.6
3. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov.* 2019;18(5):358-378. doi:10.1038/s41573-019-0012-9
4. Huang LY, Halder S, Agbandje-McKenna M. Parvovirus glycan interactions. *Curr Opin Virol.* 2014;7:108-118. doi:10.1016/j.coviro.2014.05.007
5. Reed BD, Meyer MJ, Abramzon V, et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science.* 2022;378(6616):186-192. doi:10.1126/science.abo7651
6. Grimm D, Kay MA, Kleinschmidt JA. Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. *Mol Ther.* 2003;7(6):839-850. doi:10.1016/s1525-0016(03)00095-9