

**Integrated metrology approaches for the
characterisation of recombinant adeno-associated
viral vectors: From molecular assays to mass
spectrometry**

Theodoros Kontogiannis

A thesis submitted for the degree of
Doctor of Philosophy in Biochemistry

University of Kent
School of Natural Sciences

2025

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent or any other University or Institute of learning.

Theodoros Kontogiannis

Date: 02/05/2025

Abstract

Characterisation of recombinant adeno-associated viral (rAAV) vector products for therapeutic use is essential to ensure the safety and efficacy of gene therapies. However, several current analytical techniques used to assess the critical quality attributes (CQAs) have limitations, such as low precision, limited resolution, and poor comparability between methods. To overcome these limitations and support the assessment of CQAs, a comprehensive measurement framework based on metrology is needed. This includes the development of advanced reference methods to enable comparability and traceability of measurements, support the optimisation of existing assays, and guide the development of reference materials. This thesis describes the development of higher-order analytical methods for measuring CQAs of rAAVs. Initially, rAAV preparations were characterised using standard and widely used techniques for quantification of genome and capsid titres, as well as assessment of the purity of rAAV samples and percentage of full particles. These current practice techniques, consisting of quantitative polymerase chain reaction (qPCR) for genome titre analysis, enzyme-linked immunosorbent assay (ELISA) for capsid titre analysis, and sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) for purity analysis, formed the basis for comparison with more advanced analytical methods developed in this study for comprehensive characterisation of both the rAAV capsid and genome. For genome characterisation, single-target digital polymerase chain reaction (dPCR) assays were initially developed for quantifying the genome titre and were compared to qPCR. However, while single-target assays provide information on genome titre, they do not offer insights into genome integrity. To address this limitation, multiplex dPCR assays were developed for a more comprehensive characterisation of rAAV preparations, enabling evaluation of genome integrity alongside quantification of genome titre across different regions of the rAAV genome. Genome integrity data obtained through multiplex dPCR in rAAV preparations were then compared to genome integrity data generated using nanopore sequencing. For impurity characterisation, residual DNA originating from the host cells and production plasmids was identified using nanopore sequencing, while mass spectrometry (MS) was employed to detect host cell proteins (HCPs). A total of 100 HCPs were identified, 24 of which have previously been reported in rAAV preparations in the literature. For capsid characterisation, an MS-based approach, traceable to the International System of Units (SI), was developed for absolute quantification of individual capsid proteins of rAAV9 (VP1, VP2, and VP3), determination of total capsid content, and measurement of capsid stoichiometry. The findings indicate that the measured capsid stoichiometry of the analysed rAAV9 sample deviates from the widely assumed 1:1:10 ratio of VP1:VP2:VP3. Additionally, MS analysis detected deamidations in both rAAV8 and rAAV9 capsids. Together, the findings of this study will contribute to improved measurement of rAAV CQAs, which can, in turn, inform process development, enhance understanding of how these attributes influence the overall quality of rAAV-based gene therapy products, and ultimately support improved product quality and batch-to-batch consistency.

Acknowledgements

First, I would like to express my deepest appreciation to Professor Mark Smales for his guidance and support throughout this project.

I want to extend my sincere thanks to Dr Carole Foy, Dr Julian Braybrook, and Dr Milena Quaglia for giving me the opportunity to work on this project and for their support along the way. It was a privilege to work in such a welcoming and encouraging environment.

To Dr Alexandra Whale, thank you for your patience, mentorship, and countless hours of teaching me about digital PCR.

Many thanks to Dr Christopher McElroy for teaching me about mass spectrometry and for always stepping in when an instrument failed. His good humour made even the most challenging days in the lab more manageable.

Thank you to Dr Rui Pereira for introducing me to next-generation sequencing, teaching me how to code, and providing helpful feedback and advice.

Finally, I want to thank my mum, dad, and brother for supporting me through every step of the PhD. Their belief in me made all the difference.

This work was supported by the Community for Analytical Measurement Science through a 2021 CAMS Fellowship Award funded by the Analytical Chemistry Trust Fund (<https://cams-uk.co.uk> | @CAMS_UK | www.linkedin.com/company/cams-uk/), the Department for Science, Innovation and Technology (DSIT) through the Chemical and Biological Metrology (CBM) programme, the National Measurement Laboratory (NML) at Laboratory of the Government Chemist (LGC), and University of Kent.

Table of Contents

Declaration.....	i
Abstract	ii
Acknowledgements	iii
Table of Contents.....	iv
List of Abbreviations.....	ix
List of Figures	xiv
List of Tables.....	xviii
Chapter 1: Introduction	1
1.1 Introduction to Gene Therapy	1
1.2 Viral Vectors: An Overview.....	2
1.3 AAV Biology.....	3
1.4 AAV as a Viral Vector	6
1.4.1 AAV Serotypes.....	6
1.4.2 Advantages & Disadvantages of rAAV Vectors.....	7
1.4.3 rAAV Production.....	8
1.4.4 Adherent and Suspension Cell Lines	11
1.5 Purification of rAAVs; Downstream Processing	11
1.5.1 Types of Impurities.....	12
1.5.2 Methods of Downstream Processing.....	12
1.6 Critical Quality Attributes of rAAVs	14
1.7 Metrology in Biotherapeutic Viral Vector Characterisation.....	16
1.7.1 Viral Vectors and Measurement Challenges	16
1.7.2 Viral Vector Reference Materials	16
1.7.3 rAAV Vector Reference Materials.....	18
1.8 Physical Titre	20
1.8.1 qPCR	20
1.8.2 dPCR	21
1.8.3 rAAV Genome Titration with qPCR and dPCR	23
1.8.4 ELISA for Capsid Titre Measurement	27
1.8.5 Optical Density	28

1.8.6	Infectious Titre	28
1.9	Identity Assays	29
1.9.1	Genetic Identity.....	29
1.9.2	Genome Integrity	30
1.9.3	Assessing the Methylation Status of rAAV Genomes	34
1.9.4	rAAV Capsid Identity and Stoichiometry.....	34
1.9.5	rAAV Capsid Post-translational Modification Mapping and Identification	41
1.10	Analytics and Viral Vector Purity.....	45
1.10.1	Analysis of Empty, Full, and Partial Capsids	46
1.10.2	Analysis of DNA Impurities.....	51
1.10.3	Host Cell Protein Analysis.....	52
1.11	Aims of the Work Described in this Thesis	56
Chapter 2: Generation and Characterisation of Model Recombinant AAV Materials		58
2.1	Introduction	58
2.2	Aims of this Chapter.....	60
2.3	Materials & Methods	60
2.3.1	Plasmids Used for Production of rAAV2.....	60
2.3.2	Preparation of Luria-Bertani Media and Agar.....	61
2.3.3	Amplification of Plasmids and Sequence Validation	61
2.3.4	PacI-digest of pAAV.CMV.PI.EGFP.WPRE.bGH Plasmid	62
2.3.5	Production of rAAV2 in HEK293 cells	62
2.3.6	Extraction and Purification of rAAV2 Particles.....	63
2.3.7	Model rAAV Sample Origin	64
2.3.8	ELISA.....	68
2.3.9	BCA Assay	69
2.3.10	SDS-PAGE Analysis.....	70
2.3.11	DNA Extraction Using the Commercial Takara Kit.....	71
2.3.12	qPCR Analysis of rAAV Samples	72
2.4	Results	73
2.4.1	Sequence Confirmation of Plasmids	73
2.4.2	Generation of DNA Templates for Assay Development	73
2.4.3	Measurement of rAAV2 and rAAV9 Sample Capsid Content.....	75
2.4.4	Measurement of rAAV Genome Titre	76

2.4.5	Calculation of the Percentage of Full rAAV Particles	80
2.4.6	Total Protein Content Estimation of rAAV8.....	81
2.4.7	Analysis of Capsid Proteins and Protein Impurities in rAAV8 Samples.....	81
2.5	Discussion.....	83
Chapter 3: Characterisation of rAAV Genome Titre and Integrity		86
3.1	Introduction	86
3.2	Aims of this Chapter.....	88
3.3	Materials & Methods	88
3.3.1	Assay Design.....	88
3.3.2	Uniplex dPCR Assays	90
3.3.3	Multiplex dPCR Assays	91
3.3.4	Pre-treatment with DNase I and Preparation of Dilutions	92
3.3.5	Linear Range Assessment.....	93
3.3.6	MspI Digest	93
3.3.7	Preparation of dPCR.....	94
3.3.8	Data Processing and Analysis.....	96
3.3.9	Calculation of target DNA concentration	98
3.3.10	Statistical analysis	99
3.3.11	Measurement of genome integrity.....	99
3.4	Results.....	100
3.4.1	Temperature Gradient.....	100
3.4.2	dPCR Platform Comparison.....	101
3.4.3	Evaluation of Linear Range of dPCR Assays	103
3.4.4	Comparison of dPCR and qPCR	108
3.4.5	Comparison of Pre-Extracted rAAV DNA and In-Partition Lysed rAAV Particles 110	
3.4.6	Quantification of rAAV Samples with Uniplex Assays	112
3.4.7	Development of 4-plex and 5-plex dPCR Assays.....	115
3.4.8	Comparison of 4-plex Assay Between QIAcuity and Absolute Q.....	121
3.4.9	4-plex Optimisation for rAAV Genome Characterisation	123
3.4.10	Genome Integrity Analysis	127
3.5	Discussion.....	128
Chapter 4: Characterisation of rAAV Identity and Purity		136
4.1	Introduction	136

4.2	Aims of this Chapter.....	138
4.3	Materials & Methods	138
4.3.1	Samples	138
4.3.2	DNase I reaction.....	138
4.3.3	rAAV9 DNA Extraction using the Heat Method	141
4.3.4	rAAV9 DNA Extraction using the PureLink Method	141
4.3.5	Library Preparation	142
4.3.6	Nanopore Sequencing.....	144
4.3.7	Overview of Sequencing Runs	145
4.3.8	Data Processing and Analysis	146
4.3.9	Samples for dPCR	147
4.3.10	Preparation of Samples for dPCR.....	147
4.3.11	Preparation of dPCR Reactions	148
4.3.12	Digital PCR Data Processing and Analysis	148
4.3.13	Measurement of Genome Integrity by dPCR.....	148
4.4	Results.....	148
4.4.1	Library Quality Control.....	148
4.4.2	Yield, Quality and Read Length Metrics.....	150
4.4.3	Sequencing Depth and Read Length Distribution.....	151
4.4.4	Variant Calling Analysis	156
4.4.5	Chimeric Backbone/rAAV Molecule Analysis.....	159
4.4.6	Genome Integrity Assessment from the Nanopore Sequencing Data.....	159
4.4.7	Genome Integrity Analysis using dPCR	160
4.4.8	Comparison of Genome Integrity as Determined by dPCR or Nanopore Sequencing	162
4.4.9	rAAV genome Purity Assessment using the Nanopore Sequencing Data.....	163
4.4.10	Investigating the PureLink Extraction Blank Yield and Read Identification Data	166
4.5	Discussion.....	168
Chapter 5: Characterisation of rAAV Capsid Proteins		176
5.1	Introduction	176
5.2	Aims of this Chapter.....	178
5.3	Materials & Methods	179
5.3.1	LC-MS Analysis of Intact VPs	179

5.3.2	Preparation of solutions.....	179
5.3.3	Preparation and Digestion of rAAV8 and rAAV9 Samples with Glu-C.....	180
5.3.4	Preparation and Digestion of rAAV8 and rAAV9 Samples with Trypsin	181
5.3.5	LC-MS/MS Analysis for Protein Sequencing.....	181
5.3.6	Identification of Peptides from each Capsid Protein Suitable for Determination of rAAV9 VP Capsid Stoichiometry	182
5.3.7	Gravimetical Preparations	184
5.3.8	Preparation of Isotopically Labelled Peptide Mastermix	184
5.3.9	Preparation of Natural Peptides	185
5.3.10	LC-MS analysis for Purity Assessment of VP1u, VP3u, VPc Peptides	185
5.3.11	Amino Acid Analysis	186
5.3.12	Preparation of Standard Curves	189
5.3.13	Analysis of Digested rAAV9 by Triple Quadrupole MS	190
5.3.14	Development of Multiple Reaction Monitoring (MRM) Method	191
5.3.15	LC-MS/MS Quantification Method Workflow Overview.....	192
5.3.16	Equations for Calculation of Variability.....	195
5.4	Results	196
5.4.1	Measurement of the Intact Masses of the rAAV Capsid Proteins with LC-MS for Serotype Identification.....	196
5.4.2	Mass Spectrometry Peptide Mapping and Sequence Coverage Analysis of the rAAV VP Proteins.....	200
5.4.3	Detection of PTMs by MS Analysis; Deamidation and Oxidation	205
5.4.4	Peptide Release Monitoring.....	208
5.4.5	Purity Assessment of Natural Peptides.....	208
5.4.6	Amino Acid Analysis of Natural Peptides	210
5.4.7	Collision Energy Optimisation	211
5.4.8	MS Quantification of the VP1u, VP3u, VPc Peptides from rAAV9 Samples ...	213
5.4.9	Identification of Host-Cell Proteins in the rAAV8 Samples by MS Analysis ...	216
5.5	Discussion.....	219
Chapter 6: Overall Discussion.....		224
References		236
Appendix.....		265

List of Abbreviations

AAP	Assembly-Activating Protein
AAV	Adeno-Associated Virus
Ad5	Adenovirus serotype 5
AEX	Anion Exchange Chromatography
ALC	Average Local Confidence
AMBIC	Ammonium Bicarbonate
ATCC	American Type Culture Collection
AUC	Analytical Ultracentrifugation
BCA	Bicinchoninic Acid
bGH polyA	Bovine Growth Hormone Polyadenylation Signal
bp	Base Pairs
CamKII α	Calcium/calmodulin-dependent protein kinase type II subunit alpha
CDMS	Charge Detection Mass Spectrometry
CE	Collision Energy
CID	Collision-Induced Dissociation
CMV	Cytomegalovirus
CNS	Central Nervous System
Cq	Quantification Cycle
CQA	Critical Quality Attribute
CRM	Certified Reference Material
CryoTEM	Cryogenic Transmission Electron Microscopy
CV	Coefficient of Variation
Da	Dalton
DDA	Data-Dependent Acquisition
ddPCR	Droplet Digital Polymerase Chain Reaction
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic Acid

dPCR	Digital Polymerase Chain Reaction
dsDNA	Double-Stranded Deoxyribonucleic Acid
DTT	Dithiothreitol
eGFP	Enhanced Green Fluorescent Protein
ELISA	Enzyme-Linked Immunosorbent Assay
EM	Electron Microscopy
ESI	Electrospray Ionisation
FACS	Fluorescence-Activated Cell Sorting
FDA	Food and Drug Administration
FDR	False Discovery Rate
GC	Gas Chromatography
GC-MS	Gas Chromatography Mass Spectrometry
GFP	Green Fluorescent Protein
gHCl	Guanidine Hydrochloride
GOI	Gene of Interest
GPSeq	Genome Population Sequencing
HCDNA	Host Cell Deoxyribonucleic Acid
HCMV	Human Cytomegalovirus
HCP	Host Cell Protein
HEK	Human Embryonic Kidney
HeLa	Henrietta Lacks
HILIC	Hydrophilic Interaction Chromatography
HMG	High-Mannose Glycans
HPLC	High-Performance Liquid Chromatography
HPV	Human Papillomavirus
HRP	Horseradish Peroxidase
HSV-1	Herpes Simplex Virus 1
IAA	Iodoacetamide
IC	Infectious Centre

ICA	Infectious Centre Assay
IEDB	Immune Epitope Database
ISO	International Organisation of Standardisation
ITR	Inverted Terminal Repeat
IU	Infectious Units
IVCIA	<i>In Vitro</i> Comparative Immunogenicity Assessment
Kb	Kilobases
LB	Luria-Bertani
LC	Liquid Chromatography
LC-MS	Liquid Chromatography/Mass Spectrometry
LMO2	LIM Domain Only-2
LOD	Limit of Detection
LOQ	Limit of Quantification
m/z	Mass-to-Charge ratio
MAAP	Membrane-Associated Accessory Protein
MALDI-TOF	Matrix-Assisted Laser Desorption/Ionisation Time-of-Flight
MHC	Major Histocompatibility Complex
MRM	Multiple Reaction Monitoring
mRNA	Messenger Ribonucleic Acid
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MTBSTFA	N-Tert-Butyldimethylsilyl-N-Methyltrifluoroacetamide
mtDNA	Mitochondrial Deoxyribonucleic Acid
NEB	New England Biolabs
NGS	Next-Generation Sequencing
NIST	National Institute of Standards and Technology
NMIJ	National Metrology Institute of Japan
Nt	Nucleotide
NTC	No Template Control

OD	Optical Density
ONT	Oxford Nanopore Technologies
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
PEI	Polyethylenimine
PTM	Post-translational Modification
QC	Quality Control
qPCR	Quantitative Polymerase Chain Reaction
rAAV	Recombinant Adeno-Associated Virus
RBE	Rep Binding Element
RFU	Relative Fluorescence Units
RNA	Ribonucleic Acid
rRNA	Ribosomal Ribonucleic Acid
RSD	Relative Standard Deviation
RSM	Reference Standard Material
RSWG	Reference Standard Working Group
RT	Room Temperature
RTN3	Reticulon 3
RT-qPCR	Reverse Transcription- Quantitative Polymerase Chain Reaction
scAAV	Self-Complementary Adeno-Associated Virus
SD	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulfate – Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SEC-MALS	Size Exclusion Chromatography – Multi-Angle Light Scattering
Sf	<i>Spodoptera frugiperda</i>
SI	International System of Units
SID	Standard Isotope Dilution
SLS	Static Light Scattering
SMRT	Single Molecule Real-Time

SRM	Selected Reaction Monitoring
ssAAV	Single-Stranded Adeno-Associated Virus
ssDNA	Single-Stranded Deoxyribonucleic Acid
SSV-Seq	Single-Strand Virus Sequencing
Strep-HRP	Streptavidin Conjugated to Horseradish Peroxidase
SV	Structural Variation
SV40 polyA	Simian Virus 40 Polyadenylation signal
TBDMSCI	Tert-Butyldimethylchlorosilane
TCID50	50% Tissue Culture Infectious Dose
TE	Tris-EDTA
TFF	Tangential Flow Filtration
T _m	Melting Temperature
TMB	3,3',5,5'-Tetramethylbenzidine
Trs	Terminal Resolution Site
UHMR	Ultra-High Mass Range
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
UV-Vis	Ultraviolet-Visible Spectroscopy
VA RNA	Viral-Associated Ribonucleic Acid
Vg	Viral Genomes
WPRE	Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element
WR	Working Reagent
XIC	Extracted Ion Chromatogram
YY1	Yin Yang 1

List of Figures

Figure 1.1 The structure of the ITR of AAVs.	3
Figure 1.2 Schematic of the genome organisation of wild-type AAV.	4
Figure 1.3 Production of rAAVs by transient transfection.	9
Figure 1.4 Digital PCR workflow.	22
Figure 1.5 Schematic of four dPCR assays hybridising to target sites in rAAV genome.	33
Figure 1.6 Example of an SID-MS workflow for quantification of target proteins.	40
Figure 1.7 Schematic of the diverse rAAV products that can be found in an rAAV production batch.	47
Figure 1.8 Schematic representation of protein characterisation workflows for rAAV preparations.	55
Figure 2.1 Schematic maps of recombinant AAV (rAAV) genomes used in this thesis.	67
Figure 2.2 TapeStation analysis of digestion of pAAV.CMV.PI.EGFP.WPRE.bGH with PacI restriction enzyme.	74
Figure 2.3 Capsid concentrations of rAAV2 and rAAV9 samples determined by ELISA.	76
Figure 2.4 qPCR analysis of rAAV genomes.	78
Figure 2.5 Measured genome titres of trFragment and rAAV Samples using qPCR.	80
Figure 2.6 Total protein concentration of rAAV8 sample determined by BCA assay.	81
Figure 2.7 SDS-PAGE gel analysis of rAAV8 capsids.	82
Figure 3.1 Annealing temperature gradient in QX200 for trFragment.	101
Figure 3.2 Analysis of trFragment using QX200, Absolute Q, and QIAcuity dPCR platforms.	102
Figure 3.3 Linear range assessment of ITR, CMV, eGFP, WPRE, bGH polyA assays on trFragment with QIAcuity dPCR.	104
Figure 3.4 One-colour QIAcuity dPCR plots of the analysed serial dilutions of trFragment.	106
Figure 3.5 Monitoring of the number of positive partitions observed in NTCs analysed with the uniplex ITR assay by QIAcuity dPCR.	108
Figure 3.6 Comparison of qPCR and QIAcuity dPCR measurements in extracted rAAV DNA samples.	110
Figure 3.7 Comparison of genome titres between pre-extracted rAAV2 DNA and in-partition lysed rAAV2 particles in paired experiments with QIAcuity dPCR.	111
Figure 3.8 Genome titre measurement by QIAcuity dPCR based on assays targeting different regions across the recombinant genome.	113
Figure 3.9 Measurement of the ratio of ITR concentration to other target regions in rAAV samples and trFragment by QIAcuity dPCR.	114
Figure 3.10 Two-colour QIAcuity dPCR plots for the 4-plex A and 5-plex assays on trFragment.	117
Figure 3.11 Partition occupancy analysis in ATTO550-FAM two-colour QIAcuity dPCR plot of the 5-plex assay.	119
Figure 3.12 Comparison of the copy number concentrations of the target regions of trFragment between uniplex, 4-plex A, and 5-plex formats by QIAcuity dPCR.	121
Figure 3.13 Comparison of 4-plex A and uniplex assays between QIAcuity and Absolute Q.	122
Figure 3.14 Comparison of two-colour QIAcuity dPCR plots between 4-plex A and 4-plex B assays on rAAV9.	124
Figure 3.15 Comparison of uniplex and 4-plex assays for genome titre measurement across rAAV samples by QIAcuity dPCR.	126
Figure 3.16 Genome integrity analysis of rAAV samples by QIAcuity dPCR.	127
Figure 4.1 Schematic representation of the rAAV9 genome extraction and nanopore sequencing workflow.	140

Figure 4.2 TapeStation analysis of extracts and sequencing libraries.	149
Figure 4.3 Sequencing depth analysis of rAAV9 Heat and rAAV9 PureLink samples.	153
Figure 4.4 Read length distributions of rAAV9 Heat and rAAV9 PureLink extracted genomes across all runs.	155
Figure 4.5 Deletions observed in reads of the rAAV9 genome.	157
Figure 4.6 Annotated CamKII α promoter region (893 – 1904 nt) of the rAAV9 genome.	158
Figure 4.7 Detection of plasmid backbone DNA in rAAV9 mapped reads.	159
Figure 4.8 Measurement of rAAV9 genome integrity by nanopore sequencing.	160
Figure 4.9 Comparison of rAAV9 genome integrity and concentration across extraction methods when assessed using a 4-plex dPCR approach.	162
Figure 4.10 Comparison of rAAV9 genome integrity as determined by dPCR and nanopore sequencing.	163
Figure 4.11 Nanopore sequencing reads mapped to rAAV9 and other sources for rAAV9 PureLink and rAAV9 Heat generated samples.	164
Figure 4.12 Read length distribution plots of DNA impurities found in rAAV9 Heat and rAAV9 PureLink nanopore sequencing data.	165
Figure 4.13 Read length distribution plots of unclassified reads found in rAAV9 PureLink samples.	166
Figure 4.14 Nanopore sequencing mapped reads of PureLink Extraction Blanks.	168
Figure 5.1 Schematic representation of unique peptide generation through Glu-C digestion.	183
Figure 5.2 Multi-step workflow for quantification of VP1u, VP3u, VPc.	194
Figure 5.3 Chromatograms of rAAV8 and rAAV9 samples analysed with LC-MS.	197
Figure 5.4 Mass spectrometry analysis and identification of the capsid proteins VP1, VP2, VP3 in rAAV8 (A) and rAAV9 (B) samples.	199
Figure 5.5 Protein sequence coverage map of rAAV9 VP1 protein.	201
Figure 5.6 Protein sequence coverage map of rAAV8 VP1 protein.	202
Figure 5.7 Identification of VP1u, VP3u, and VPc peptides based on their MS/MS spectra.	204
Figure 5.8 Identification of deamidations in rAAV8 based on MS/MS spectra.	206
Figure 5.9 Identification of deamidations in rAAV9 based on MS/MS spectra.	207
Figure 5.10 Release of VP1u, VP3u, and VPc peptides from rAAV9 during Glu-C digestion over time.	208
Figure 5.11 Intact LC-MS analysis of VP1u, VP3u, and VPc natural peptides.	209
Figure 5.12 Amino acid analysis of VP1u, VP3u, and VPc natural peptides with GC-MS/MS MRM.	211
Figure 5.13 Collision energy (CE) optimisation for transitions of rAAV capsid peptides VP1u, VP3u, and VPc.	212
Figure 5.14 Standard curves for quantification of VP1u, VP3u, and VPc peptides.	214
Appendix Figure 7.1 Schematic of the pAAV2/2 plasmid.	265
Appendix Figure 7.2 Schematic of the pAdDelta F6 plasmid.	265
Appendix Figure 7.3 Schematic of the pAAV.CMV.PI.EGFP.WPRE.bGH plasmid.	266
Appendix Figure 7.4 Protein sequence coverage map of rAAV9 VP1 protein after digestion with 10 μ g Glu-C and subsequent LC-MS/MS analysis.	266
Appendix Figure 7.5. Protein sequence coverage map of rAAV9 VP2 protein.	267
Appendix Figure 7.6 Protein sequence coverage map of rAAV9 VP3 protein.	268
Appendix Figure 7.7 Protein sequence coverage map of rAAV8 VP2 protein.	269
Appendix Figure 7.8 Protein sequence coverage map of rAAV8 VP3 protein.	270
Appendix Figure 7.9 MRM Chromatograms of VP1u endogenous and labelled peptides.	271
Appendix Figure 7.10 MRM Chromatograms of VP3u endogenous and labelled peptides.	271
Appendix Figure 7.11 MRM Chromatograms of VPc endogenous and labelled peptides.	272

Appendix Figure 7.12 MRM chromatogram of the transitions for signature peptides in a Glu-C digested HEK293 cell lysate.. ..	272
---	-----

List of Tables

Table 1.1 Selection of AAV serotypes for targeting major tissues.....	7
Table 1.2 Methods used for the production of rAAVs.....	10
Table 1.3 Summary of analytical techniques used for viral vector characterisation.....	19
Table 1.4 Overview of characteristics of dPCR platforms: QX200, QIAcuity, and Absolute Q.	23
Table 1.5 Summary of findings from studies investigating the effect of primer/probe location and genome conformation in rAAV titration.	25
Table 1.6 The masses of VP1, VP2, VP3 for AAV serotypes (1-12).	36
Table 1.7 Summary of PTMs found on the capsids of the rAAV serotypes (1-rh10).	43
Table 1.8 Summary of common PTMs found in both HEK293 and Sf9-produced rAAV8 and rAAV1 vectors reported by Rumachik et al., 2020.	45
Table 1.9 Examples of HCPs that have been identified in rAAV preparations in more than one study.	54
Table 2.1 Preparation of BSA standard curve.....	69
Table 2.2 Alignment metrics of pAAV2/2, pAdDelta F6, and the transfer plasmid (pAAV.CMV.PI.EGFP.WPRE.bGH) sequencing results against the reference sequences.	73
Table 2.3 Calculated percentage of full rAAV2 and rAAV9 particles based on ELISA and qPCR analysis..	81
Table 3.1 Primer and probe sequences used for dPCR analysis.....	90
Table 3.2 Uniplex assays used for measurement of each rAAV sample.....	91
Table 3.3 Overview of assay components used in 4-plex (A and B) and 5-plex digital PCR assays, including fluorophore combinations and samples analysed with each assay configuration.	92
Table 3.4 Overview of detection channels available on the QX200, QIAcuity, and Absolute Q platforms, and the assay detected in each channel based on the fluorophore label of the probe.	98
Table 3.5 Annealing temperatures tested for optimisation of the ITR assay.	100
Table 3.6 Measured trFragment copy number concentration and interexperimental variability (CV%) across QX200, Absolute Q, and QIAcuity.	103
Table 3.7 Inter-assay variability across the points of the standard curve.	107
Table 3.8 Comparison of inter-experimental precision between qPCR and dPCR.	110
Table 3.9 Intra-assay variability estimated for each sample analysed.	113
Table 3.10 Inter-assay precision across analysed samples.....	115
Table 4.1 Nanopore sequencing runs specifications.....	145
Table 4.2 rAAV9 samples analysed in each sequencing run.....	146
Table 4.3 Concentration of sequencing libraries measured by TapeStation and amount of library loaded on each sequencing run.....	149
Table 4.4 Read metrics of nanopore sequenced samples.....	150
Table 4.5 Read length metrics of analysed samples using nanopore sequencing.	151
Table 4.6 Sequencing depth metrics of analysed samples using nanopore sequencing.	152
Table 4.7 Detail on % of reads containing a deletion in the rAAV9 genome.....	156
Table 4.8 Nanopore sequencing read metrics of extraction blanks.....	167
Table 5.1 Sequences of the peptides (VP1u, VP2u, VP3u, VPc).	184
Table 5.2 Sequences of the labelled peptides VP1u, VP3u, VPc.	185
Table 5.3 Composition of controls, sample blends, calibration blends, and blanks used for amino acid analysis.....	187
Table 5.4 MRM transitions used for analysis of each labelled and unlabelled amino acid.	188
Table 5.5 Selected transitions for the quantification of peptides VP1u, VP3u, and VPc.....	191

Table 5.6 Theoretical and measured masses (Da) of VP1, VP2, VP3 for rAAV8, rAAV9 capsids.	198
Table 5.7 Deamidation sites detected in trypsin-generated peptides from the capsid proteins of rAAV8 and rAAV9 samples.	205
Table 5.8 Purity assessment of VP1u, VP3u, and VPc peptides by LC-MS.	210
Table 5.9 Mean concentrations (mg/mL and total mg) of natural peptide stock solutions, as determined by amino acid analysis and corrected for peptide purity.	211
Table 5.10 Selected transitions and corresponding optimal collision energies for the MRM method.	213
Table 5.11 Estimation of limit of detection (LOD) and limit of quantification (LOQ) for each signature peptide.	215
Table 5.12 Measured mean stoichiometry and concentrations (nM) of VP1u, VP3u, and VPc peptides in rAAV9 stock.	216
Table 5.13 Host cell proteins identified in this study (out of 100 total) that have also been reported in previous studies.	217
Appendix Table 7.1 Full list of HCPs identified in the rAAV8 sample.	273

Chapter 1: Introduction

Parts of this chapter are based upon a review initially drafted and written by the author of this thesis that has been published Open Access in:

Kontogiannis, T. *et al.* (2024) 'Characterization of AAV vectors: A review of analytical techniques and critical quality attributes', *Molecular Therapy - Methods & Clinical Development*, 32(3), p. 101309. Available at: <https://doi.org/10.1016/j.omtm.2024.101309>. (KAR id: 107253)

1.1 Introduction to Gene Therapy

Gene therapy is a promising strategy that can be applied for the treatment of a wide range of genetic diseases by restoring defective gene functions. This is typically achieved by providing functional genetic material that will produce the missing, or in the case of mutated genes, the 'native' protein, to the targeted cells and tissues (Lukashev and Zamyatnin, 2016). The delivery of such genetic material into the cells of patients is most often achieved using viral vectors, as viruses have naturally evolved to deliver their genetic material into cells (Vargas *et al.*, 2016).

There are two main categories of gene therapy, *in vivo* and *ex vivo*. In *in vivo* gene therapy, the vector containing the functional gene of interest (GOI) is injected directly into the patient (Russell *et al.*, 2017). In contrast, in *ex vivo* gene therapy, the patient's cells are first isolated and then modified with the vector that carries the desired genetic material. The modified cells are then transplanted back into the patient (Kohn *et al.*, 2021).

The first approved gene therapy clinical trial was from the National Institutes of Health in 1990 to treat adenosine deaminase-deficient severe combined immunodeficiency. For this purpose, a retroviral vector was used to deliver the gene encoding the enzyme adenosine deaminase into the patient (Blaese *et al.*, 1995). The first gene therapy trials were successful and brought excitement to the scientific community, resulting in the development and testing of several new gene therapies in clinical settings. However, there were two setbacks in clinical progress. A trial targeting the cure of ornithine transcarbamylase deficiency led to the death of an 18-year-old patient due to toxicity and multiorgan failure associated with the adenovirus

vector that was used (Raper *et al.*, 2003). In another trial, patients with X-linked severe immunodeficiency disease developed a leukemia-like syndrome as a result of the integration of the gamma retrovirus vector near the promoter region of the protooncogene LIM Domain Only-2 (*LMO2*) (Hacein-Bey-Abina *et al.*, 2003; McCormack and Rabbitts, 2004). This resulted in a setback to the development of viral vector-delivered gene therapies and significant effort by the scientific community to ensure vector safety and stricter regulatory demands. Subsequently, the first gene therapy approved in Europe was GlyberaR in 2012, an adeno-associated virus (AAV) gene therapy, which was used for the treatment of lipoprotein lipase deficiency (Watanabe *et al.*, 2015). The license for the drug was not renewed due to the high cost of manufacturing and the rare occurrence of the condition, making the therapy financially unsustainable.

1.2 Viral Vectors: An Overview

Viruses are efficient tools for gene delivery as they are naturally involved in delivering their nucleic acid (DNA or RNA) to the cells they are infecting (Vargas *et al.*, 2016). However, a major drawback of using viruses as vectors is that they can be recognised and attacked by the immune system or can potentially replicate in the patient if the entire viral genome is used in manufacturing a viral vector therapy (Finer and Glorioso, 2017). For these reasons, viral vectors for gene therapy applications have been modified to prevent their ability to replicate, minimise their immunogenic profile, and reduce their toxicity. However, care must be taken to ensure that such modifications do not alter their infection pathway, as this would reduce their efficiency as gene therapy vehicles (Thomas, Ehrhardt and Kay, 2003).

There is a broad spectrum of viral vectors that can be used for gene therapy, including adenovirus, retrovirus/lentivirus, poxviruses, AAV, and herpes simplex-1 virus (HSV-1) (Bouard, Alazard-Dany and Cosset, 2009; Warnock, Daigre and Al-Rubeai, 2011; Lundstrom, 2018). Among the most widely used vectors are the adeno-associated viral vectors. The number of gene therapy programs using AAV as a delivery vector has steadily increased in recent years, with eight AAV-mediated gene therapies now approved globally (Lomash *et al.*, 2025).

AAV vectors cannot integrate their genome into the genome of the host cell to achieve stable expression of the gene of interest, but instead, they remain mainly episomal, and the AAV DNA is quickly lost in rapidly dividing cells. The benefit of using AAVs is that they can deliver genes to postmitotic cells and allow their expression for a long time without causing toxicity (Chen *et al.*, 2018; Daya and Berns, 2008; Flotte *et al.*, 1993; Podsakoff *et al.*, 1994; Yang *et al.*, 2008; Zhao *et al.*, 2022). There is also a range of AAV serotypes that target different tissue and cell types, potentially allowing gene therapies using AAVs to target a wide range of cell and tissue types.

1.3 AAV Biology

The AAVs belong to the *Parvoviridae* family and are small (~20 nm) non-pathogenic, non-enveloped, icosahedral viruses that package a single-stranded DNA (ssDNA) of ~4.8 kilobases (kb) (Yates *et al.*, 1973; Srivastava, Lusby and Berns, 1983). Their genome consists of two main genes, *rep* and *cap*, flanked by T-shaped hairpin structures called inverted terminal repeats (ITRs), which have an essential role in viral replication origin and packaging signalling (Koczot *et al.*, 1973; Tratschin, Miller and Carter, 1984). An ITR consists of the palindromic arms A-A', B-B', and C-C', the single-stranded sequence D, the Rep binding element (RBE) that functions as the binding site of Rep 78 and Rep 68, and a terminal resolution site (*trs*) that serves as a nicking site and is essential for DNA replication and packaging (**Figure 1.1**). An additional RBE (RBE') is part of the ITR sequence and acts by enhancing the stability of the interactions between the Rep proteins and the ITR (Bennett, Mietzsch and Agbandje-McKenna, 2017).

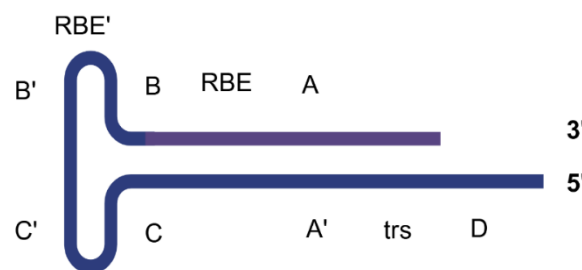


Figure 1.1 The structure of the ITR of AAVs. The ITR consists of the palindromic regions A-A', B-B', C-C', the single-stranded sequence D, two Rep Binding Elements (RBEs), and a terminal resolution site (*trs*). Created with BioRender.com.

The *rep* gene encodes for four non-structural Rep proteins (Rep78, Rep68, Rep52, and Rep40) that are involved in viral replication, transcriptional regulation, integration, and packaging. The two major isoforms, Rep78 and Rep68, are expressed by the p5 promoter and are important for viral replication, integration, and expression regulation. The two minor isoforms, Rep52 and Rep40, are mostly involved in packaging, and their expression is regulated by the p19 promoter (**Figure 1.2**) (Weitzman *et al.*, 1994; Hörer *et al.*, 1995; Pereira, McCarty and Muzyczka, 1997; King *et al.*, 2001).

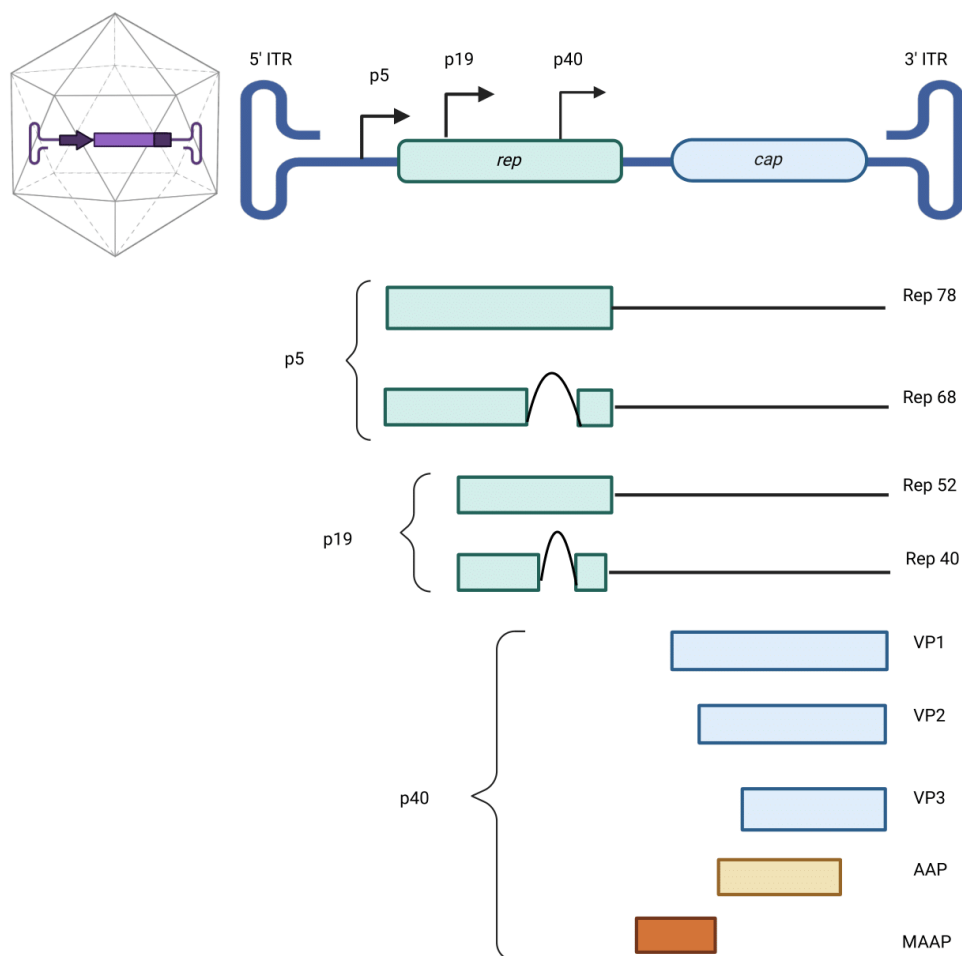


Figure 1.2 Schematic of the genome organisation of wild-type AAV. The genome of AAV consists of two main genes, *rep* and *cap*, flanked by T-shaped hairpin structures called ITRs. The *rep* gene encodes for four non-structural Rep proteins (Rep 78, Rep 68, Rep 52, and Rep 40), while the *cap* gene encodes three capsid proteins (VP1, VP2, VP3). Alternative open reading frames (ORFs) of the *cap* gene encode for the assembly-activating protein (AAP) and the membrane-associated accessory protein (MAAP). Created with BioRender.com.

The *cap* gene encodes three capsid proteins (VP1, VP2, VP3), and it is generally considered that they occur in a ratio of approximately 1:1:10 (VP1:VP2:VP3) in the viral capsid (Rose *et al.*, 1971). The masses of the capsid proteins may vary between different serotypes. VP3 is the smallest of the proteins (59-61 kilodaltons (kDa)), and its sequence is also present in VP1 and VP2. This is the common region of all VPs. VP2 is a slightly larger protein (64-67 kDa) and includes the amino acid sequence of VP3 and 57 additional amino acids at its N-terminal. Finally, VP1 (79-82 kDa) consists of the amino acid sequence of VP2 and an extra 137 amino acids at its N-terminal. This region is called VP1 unique region. The three VP proteins are encoded by the same gene (*cap*) from the p40 promoter by using alternative splicing and start codons (Wörner *et al.*, 2021). Specifically, VP1 is translated from a minor spliced mRNA, resulting in its lower abundance. VP2 is translated from the major spliced mRNA. However, its translation initiates at an ACG codon, which is less efficient than the canonical AUG, leading to reduced expression. VP3 is also translated from the major spliced mRNA and benefits from a favourable Kozak consensus sequence, resulting in higher expression levels. These mechanisms contribute to the approximate capsid stoichiometry of 1:1:10 (Becerra *et al.*, 1988; Van Vliet *et al.*, 2008).

Furthermore, a second open reading frame (ORF) of the *cap* gene encodes for the assembly-activating protein (AAP) (Sonntag *et al.*, 2011). Grosse *et al.*, 2017 demonstrated that the AAP protein prevents proteasomal degradation of the AAV capsid proteins by promoting rapid capsid assembly. Moreover, AAP functions as a chaperone by directly interacting with capsid proteins, facilitating their transport to the nucleolus and promoting efficient assembly into capsids (Earley *et al.*, 2015; Grosse *et al.*, 2017). The virus also encodes for the membrane-associated accessory protein (MAAP) that binds to the cell membrane and the X proteins that are involved in DNA replication (Cao, You and Hermonat, 2014; Galibert *et al.*, 2021).

AAV is classified as a helper-dependent virus and can only replicate in the presence of adenovirus or herpes simplex virus. Without the co-infecting virus, AAV remains in the host cell as an episome, or is integrated in a specific region of chromosome 19, called AAVS1, as part of its lysogenic cycle (Kotin *et al.*, 1990; Samulski *et al.*, 1991). Following a helper virus infection, a productive infection is achieved, and the newly synthesised virions are released from the cells (Ward, Elias and Linden, 2003; Gonçalves, 2005).

1.4 AAV as a Viral Vector

In recombinant AAVs (rAAVs), the wild-type open reading frames (*rep* and *cap*) are replaced by a transgene expression cassette (promoter, gene of interest, and terminator) placed between the two ITRs (Gray and Zolotukhin, 2011). A drawback of rAAV vectors is their relatively small packaging size (~ 5 kb), which hinders their ability to deliver large genes to patients' cells (Wang, Tai and Gao, 2019). To address this, split vector strategies have been developed. In this approach, the large transgene is split into two fragments, which are inserted into two separate vectors. The two vectors are delivered to the targeted tissue, and the transgene cassette's reconstruction occurs through homologous recombination or nonhomologous end joining. As a result, the transgene is expressed as a single unit (Hirsch, Wolf and Samulski, 2016).

1.4.1 AAV Serotypes

To date, 13 main AAV serotypes have been identified, which differ in the sequences of their capsid proteins (Jiang and Dalby, 2023). The thirteen main serotypes differ from each other in their tropism and transduction efficiency (Jiang and Dalby, 2023). Therefore, selecting a suitable serotype for the transduction of a specific cell type is essential. rAAV1, 6, 7, and 8 can be used for gene therapies targeting skeletal muscles (Gao *et al.*, 2002; Blankinship *et al.*, 2004; Wang *et al.*, 2005). rAAV8 is commonly used to transduce hepatocytes, and rAAV9 is used for neuro-targeted therapies (Gao *et al.*, 2002; D. Liu *et al.*, 2021; Issa *et al.*, 2023). rAAV9, 2, 4, and 5 are used for providing the therapeutic effect to the central nervous system (CNS) (**Table 1.1**) (Bartlett, Samulski and McCown, 1998; Davidson *et al.*, 2000). rAAV9 is particularly advantageous due to its broad tissue tropism and unique ability to cross the blood-brain barrier (D. Liu *et al.*, 2021; Issa *et al.*, 2023).

Mixing the capsid and the genome of different serotypes is called pseudotyping. For instance, in rAAV2/5 vectors, the capsid is derived from AAV5 and the ITRs from AAV2. This method can lead to the production of vectors with altered tropism and increased transduction efficiency (Balaji *et al.*, 2013).

Table 1.1 Selection of AAV serotypes for targeting major tissues.

Tissue Target	Optimal Serotypes
CNS	rAAV2, rAAV4, rAAV5, rAAV9 (Bartlett, Samulski and McCown, 1998; Davidson <i>et al.</i> , 2000; D. Liu <i>et al.</i> , 2021; Issa <i>et al.</i> , 2023)
Heart	rAAV8, rAAV9 (Wang <i>et al.</i> , 2005; Bish <i>et al.</i> , 2008)
Retina	rAAV4, rAAV5, rAAV7, rAAV8 (Weber <i>et al.</i> , 2003; Lebherz <i>et al.</i> , 2008)
Liver	rAAV8, rAAV9 (Gao <i>et al.</i> , 2002)
Skeletal Muscle	rAAV1, rAAV6, rAAV7, rAAV8 (Gao <i>et al.</i> , 2002; Blankinship <i>et al.</i> , 2004; Wang <i>et al.</i> , 2005)
Lung	rAAV5 (Zabner <i>et al.</i> , 2000)
Kidney	rAAV2 (Takeda <i>et al.</i> , 2004)
Pancreas	rAAV6, rAAV8 (Wang <i>et al.</i> , 2004; Quirin <i>et al.</i> , 2018)

The single-stranded genome of rAAV prevents the instant transcription of its genes, since a double-strand synthesis step must take place first. To address this, self-complementary AAVs (scAAVs) were developed. A self-complementary rAAV (scAAV) contains a genome that can be folded into double-stranded DNA, leading to improved transgene expression. However, the drawback of scAAVs is that their packaging capacity is half that of single-stranded rAAVs (ssAAVs) (Nathwani *et al.*, 2006; Wang, Tai and Gao, 2019).

1.4.2 Advantages & Disadvantages of rAAV Vectors

rAAV vectors are a commonly used gene therapy vehicle for their many advantages. Firstly, wild-type AAV is not known to cause any diseases in humans. Secondly, the rAAV vector genome design removes the wild-type AAV sequences (except for the ITRs). This makes rAAV

a safe vector for gene therapy. Furthermore, rAAVs have low *in vivo* immunogenicity and can deliver the transgene to several different cell types due to their broad tropism. It can transduce both non-dividing and slowly dividing cells, and its episome can be maintained in the host cell for many years, allowing long-term expression. However, rAAV has disadvantages too. In particular, its small packaging capacity limits its use to the delivery of only small genes (Monahan and Samulski, 2000). Furthermore, pre-existing immunity against rAAV can impede the therapeutic effect of gene therapy (Aronson *et al.*, 2019).

1.4.3 rAAV Production

Transient multi-plasmid transfection into host cell expression systems is commonly used for the production of rAAV. Chemical transfection methods are commonly used to deliver plasmid DNA into host cells to produce rAAV vectors. Transfection reagents used include calcium phosphate, cationic lipids (lipofection), and polyethylenimine (PEI) (Wright, 2009). The calcium phosphate transfection method involves the formation of a DNA/calcium phosphate complex. The calcium phosphate allows the binding of the DNA to the cell membrane, and the DNA is introduced into the cell (Graham and van der Eb, 1973). In lipofection, the cationic lipids form a complex with the negatively charged DNA, while the neutral lipids form a complex with the cell membrane. The complex enters the cells through endocytosis, and the DNA is released into the cytoplasm (Felgner *et al.*, 1987; Kumar, Nagarajan and Uchil, 2019). PEI is a cationic polymer that forms complexes with the negatively charged DNA, called polyplexes. The polyplexes are positively charged, and they interact with the cell membrane. This interaction allows the internalisation of the complex into the cell through endocytosis (Sabin *et al.*, 2022). PEI has been shown to be more effective than calcium phosphate in the production of rAAVs (Reed *et al.*, 2006).

As mentioned above, a common approach utilised for rAAV production is transient transfection. During this process, three plasmids are simultaneously transfected into the host cells: 1) rep/cap plasmid (encoding Rep and the capsid proteins VP1, VP2, VP3), 2) transfer plasmid, containing the transgene cassette and the ITRs, 3) helper plasmid that encodes for the helper virus factors, viral-associated (VA) RNA, E2A, and E4 (**Figure 1.3**). To produce a specific rAAV serotype, the sequences encoding the VP1, VP2, and VP3 capsid proteins of that serotype are incorporated into the rep-cap plasmid. The transgene cassette, carried by the transfer plasmid, can include elements such as a promoter, a gene of interest (GOI), and a

polyadenylation signal, selected based on the intended application. These components can be modified to suit specific experimental or therapeutic needs. In addition to the VA, E2A, and E4 genes that are supplied *in trans* by the helper plasmid, the E1 adenovirus gene is essential for AAV replication, too. The E1 gene is stably expressed in human embryonic kidney (HEK293) cells, which are human adenovirus 5 (Ad5) transformed cells (Wright, 2009; Clément and Grieger, 2016). These helper genes are essential for AAV replication. E1a upregulates the expression of other adenoviral genes and allows the expression of Rep 68/78 by binding to the Yin Yang 1 (YY1) repressor and thus activating the p5 promoter (Lee *et al.*, 1995). E4orf6 is involved in DNA replication, E2A and VA RNA are involved in mRNA stability, and E1b and E4 play an important role in viral mRNA transportation (Xiao, Li and Samulski, 1998).

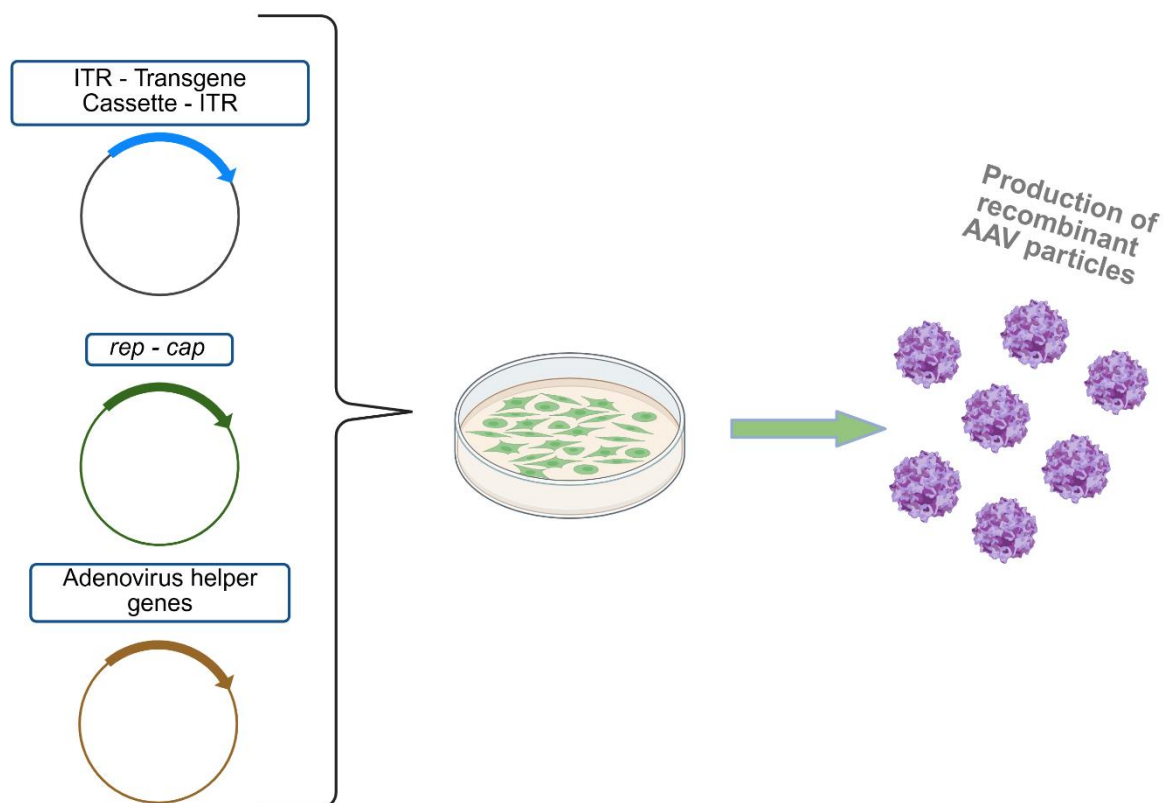


Figure 1.3 Production of rAAVs by transient transfection. Three plasmids are transfected into host HEK293 cells: the transfer plasmid which carries the transgene cassette flanked by the ITRs, the rep/cap plasmid which encodes the Rep proteins and the capsid proteins (VP1, VP2, and VP3), and the helper plasmid which provides the adenoviral helper genes (VA RNA, E2A, E4). Created with BioRender.com.

Besides the transient transfection method, rAAVs can also be produced using stable packaging or producer cell lines (**Table 1.2**). A producer cell line contains the transgene

expression cassette flanked by the ITRs integrated into its genome. Furthermore, the *rep* and *cap* genes are integrated into the genome of a producer cell line. Therefore, only infection with a helper virus is required to produce rAAV. This approach requires the constitutive expression of the viral genes *rep* and *cap*, which can be toxic to the producer cells due to the ability of the produced proteins to interfere with essential cellular processes (Yang, Chen and Trempe, 1994). This can be overcome using inducible expression systems, whereby the toxic genes are only expressed when required for viral vector production (Lee *et al.*, 2022; Jalšić *et al.*, 2023; Lin *et al.*, 2025). Otherwise, the production of rAAV from packaging cell lines requires both transfection of the transfer plasmid and infection with the helper virus, since only the *rep* and *cap* genes are integrated into the genome of the packaging cell lines (Ayuso, Mingozzi and Bosch, 2010; Robert *et al.*, 2017).

Table 1.2 Methods used for the production of rAAVs.

Method	Integrated Sequences	Transfections Required
Transient	None	rep-cap
		Adenovirus helper genes
		Transgene cassette - ITRs
Packaging cell line	<i>rep-cap</i>	Adenovirus helper genes
		Transgene cassette - ITRs
Producer cell line	<i>rep-cap</i> , transgene cassette flanked by ITRs	Adenovirus helper genes
Inducible cell line	<i>rep-cap</i> , transgene cassette, and adenovirus helper genes	None (Inducible)

The transient transfection method is a simpler method for producing viral vectors since it avoids the need for the generation of stable cell lines and can be adjusted for any application by allowing the production of recombinant viral vectors carrying a variety of transgenes. In contrast, the stable cell line-approach cannot be easily customised to produce viral vectors carrying different transgenes. Although transient transfection often leads to the production of high titres of viral vectors, this method suffers from batch-to-batch variability and is also difficult to scale up due to the high cost of the transfection reagents and plasmid production. In comparison, the development of packaging and producer cell lines minimises the cost and simplifies the production process because these cell lines can be used for long-term production of viral vectors. Moreover, using stable cell lines for viral vector production reduces the number of transfection steps required and eliminates the possibility of producing replication-competent AAV. However, since stable cell lines require infection with a helper virus to produce rAAV, it is necessary to ensure that the contaminating helper virus is not present in the final preparation of rAAV (Robert *et al.*, 2017).

1.4.4 Adherent and Suspension Cell Lines

Mammalian host cell lines, particularly HEK293 cells, are the production cell line of choice for rAAVs, although other systems have been explored (e.g., insect cells) (Rumachik *et al.*, 2020; Tran *et al.*, 2022). With mammalian cells, both adherent and suspension-adapted cell lines are used for rAAV production. In adherent cell cultures, cells grow while attached to a surface. Therefore, scaling up the production process is limited by the surface area on which the cells grow. In contrast, in suspension cell culture, cells multiply suspended in the culture medium. Suspension cell cultures are more scalable compared to adherent cultures and allow the use of serum-free media, reducing the risk of adventitious agents being present in the final preparation. However, suspension cell lines are also generally more difficult to transfect than adherent cell lines due to the decreased ability of the transfection complex to bind to the cells' surfaces (Basiouni, Fuhrmann and Schumann, 2012).

1.5 Purification of rAAVs; Downstream Processing

The purpose of downstream processing is to remove impurities and concentrate rAAVs in the final preparation. Impurities could negatively affect the safety and efficacy of the final product

and, for example, cause an immune response in the patient receiving the gene therapy (Ayuso, Mingozi and Bosch, 2010; Wright, 2014; Schnödt *et al.*, 2016; Brimble *et al.*, 2022). The gene therapy product must comply with the relevant purity specifications to be released to market. In general, the regulations are stricter for *in vivo* gene therapies compared to *ex vivo* ones. The purification of viral vectors involves a number of steps that aim to separate the vectors from the impurities based on their unique properties, such as their size and charge.

1.5.1 Types of Impurities

Impurities can be separated into two categories, product-related and process-related impurities. Product-related impurities can be empty (devoid of genome) and partial capsids (partial genomes packaged), aggregates, free viral components, and replication-competent vectors (Wright, 2014; Tran *et al.*, 2022; Dunker-Seidler *et al.*, 2025). Process related impurities consist of host cell proteins and nucleic acids, cell debris, residue plasmid DNA, components of the culture medium (serum, peptides, lipids, salts, buffers, sugars), and impurities coming from the production and purification process (e.g. transfection reagents, nucleases, buffers, residual CsCl or iodixanol) (Segura *et al.*, 2013).

1.5.2 Methods of Downstream Processing

Before starting the purification process, when an intracellular rAAV is produced, it is essential to lyse the producer cells so that the vector is released from the cells. This can be achieved by mechanical disruption of the cell membrane, repetitive freeze-thaw cycles, or by the addition of chemical detergents. However, in some circumstances, lysis of the cells may not be required, as studies have shown that rAAVs can be secreted in cell culture media with different efficiencies based upon the serotype (Vandenberghe *et al.*, 2010).

Following harvesting, the next step in the purification process is the clarification of the cell culture. This process focuses on the removal of cells, cell debris, and other large impurities. During clarification, the sample is subjected to low-speed centrifugation followed by a microfiltration step using a 0.45 µm pore membrane. A benzonase step is usually applied after the clarification step or in later stages to remove residual host cell and plasmid DNA. The application of a benzonase step early in the downstream process ensures that the following purification steps remove the remaining benzonase from the final product.

However, this approach is expensive because large amounts of benzonase are required in the early stages of the purification process.

The goal of the next step is to concentrate the vector and reduce the feed volume for the next steps of the process. rAAV concentration can be achieved by CsCl or iodixanol gradient ultracentrifugation (Zolotukhin *et al.*, 1999). Although ultracentrifugation is efficient in purifying and concentrating viral vectors in small volumes, this technique is not scalable and has long processing times. Therefore, it is not suitable for large-scale downstream processing (Segura *et al.*, 2013). A widely used process for the purification and concentration of viral vectors is tangential flow filtration (TFF). This method achieves the removal of impurities with molecular weight lower than that of the rAAVs (e.g. proteins, nucleic acids, sugars). TFF also concentrates the vector by allowing volume reduction and buffer exchange to one that is suitable for the following chromatography step. The advantages of TFF include the reduction of manual handling and the ability to handle large viral vector titres (Miyaoka *et al.*, 2023).

A technique commonly used for the subsequent further purification of rAAVs is chromatography. Chromatography methods are usually scalable, can handle large volumes, and achieve high levels of purity in a relatively short time. There are several types of chromatography that can be used for the purification of viral vectors based on their size, charge, or affinity to another molecule (Segura, Kamen and Garnier, 2011).

The negative charge of rAAVs at neutral pH is exploited for their purification using anion exchange chromatography (AEX). The positively charged ligands used for AEX can be either strong (quaternary ammonium) or weak (Diethylaminoethyl) (Segura, Kamen and Garnier, 2011). This technique is scalable and can achieve high viral vector recoveries. For this reason, a number of groups have used these techniques to purify rAAV1, rAAV8, rAAV2, and rAAV5 samples (Okada *et al.*, 2009; Yuan *et al.*, 2011; C. Wang *et al.*, 2019; Hejmowski *et al.*, 2022). The pH of the anion/cation step is an important factor influencing the recovery of rAAV1 particles using either cation-exchange or anion-exchange chromatography. In particular, Okada *et al.*, 2009 reported different recoveries of rAAV1 particles using Mustang Q anion exchange chromatography at various pH values. At pH = 7.5 and pH = 7, the researchers reported a recovery of 90% and 42%, respectively, while at pH = 6.5 the recovery

was 25%. However, for rAAV1 purified by cation exchange chromatography, the recovery was 10% at pH = 7.5 and 52% at pH = 6.5. Furthermore, anion exchange has been successfully used for the separation of empty and full rAAV2 capsids (Qu *et al.*, 2007). Qu *et al.*, 2007 reported an 86-fold reduction in empty capsids and a 74% yield of vector particles.

Another type of chromatography commonly used to purify rAAV particles is affinity chromatography. Affinity chromatography aims to isolate the rAAV based on specific interactions with a ligand immobilised on the resin. Different variations of this approach involve the modification of the vector genome to express an affinity tag that binds to the immobilised metal ligands on the stationary phase, which simplifies and accelerates the purification process. Although this approach can be used for the isolation of the target, engineering the vector to express affinity tags destabilises the vector and leads to a substantial loss in infectivity (Perry and Rayat, 2021).

Heparin affinity chromatography is regularly used for the purification of rAAVs. Heparin resembles the structure of heparan sulphate proteoglycans, which are used as attachment sites for several viruses. The ability of heparin to bind to viruses is exploited for the isolation of rAAVs and the removal of process-related impurities. Clark *et al.*, 1999 demonstrated the efficiency of this method by achieving 70% recovery of rAAV and high purity. Furthermore, AVB Sepharose™ High Performance is an affinity resin that can be used for the purification of several rAAV serotypes. This method uses a ligand that does not bind to free viral proteins (Ayuso, Mingozi and Bosch, 2010). More recently, nanobody-based ligands have been developed for AAV purification, offering advantages such as high stability and strong binding affinity (Li *et al.*, 2025). One such example is the commercially available POROS™ CaptureSelect™ AAVX resin, which enables the capture of various AAV serotypes (Florea *et al.*, 2023). Furthermore, a nanobody-based method tailored for purifying AAV8 has been developed to enhance the purification efficiency for this serotype (Li *et al.*, 2025).

1.6 Critical Quality Attributes of rAAVs

To ensure that a viral vector drug product is safe for testing in clinical settings and for release to the market, it is essential that the final product complies with the predetermined specifications and regulatory requirements of the appropriate regulatory agencies. The closer

the product is to market release, the stricter these specifications are (Perry and Rayat, 2021). Validated and precise analytical assays are used to determine whether the product complies with the set specifications by demonstrating that the critical quality attributes (CQAs) are within the predetermined range (Tustian and Bak, 2021). CQAs include identity, purity, potency, safety, and stability of viral vectors, and their accurate characterisation is mandated by regulatory agencies such as the U.S. Food and Drug Administration (FDA) (U.S. Food and Drug Administration, 2022). Furthermore, understanding how CQAs affect the product's efficacy and safety is critical for streamlined gene therapy development, and these attributes should be consistent from batch to batch. However, the biomanufacturing of viral vectors is a complex process. Small changes in the vector design, such as choosing a different promoter or transgene, selection of a different production platform or upstream bioprocessing conditions (e.g. medium, culture temperature), or use of a different purification technique, can lead to the production of viral vectors with different characteristics that could substantially influence the product quality and safety (Perry and Rayat, 2021). The monitoring and characterisation of viral vectors is thus an important step in process development.

Despite the importance of analytical analysis of viral vectors, current assays and reference materials used for their characterisation are much less well-defined than those for biotherapeutic proteins. For example, the National Institute of Standards and Technology (NIST) monoclonal antibody reference material 8671 underwent extensive multi-laboratory studies to characterise its physicochemical and immunochemical properties, purity, concentration, biological activity, glycan occupancy, and post-translational modifications (PTMs). Similar evaluations are needed for viral vectors. However, their complex composition and variable properties require additional assays to assess attributes such as particle occupancy, genome integrity, and stoichiometry. Additionally, unlike monoclonal antibodies, viral vector manufacturing and downstream processing lack industry-wide standardisation, contributing to variability and challenges in ensuring consistent quality and efficacy (European Medicines Agency, 2016; Gimpel *et al.*, 2021; Schiel and Turner, 2018; Tustian and Bak, 2021). Optimising currently utilised assays or developing new, accurate, and reproducible analytical techniques would help the scientific community understand how CQAs can impact the efficiency and safety of the final product (Gimpel *et al.*, 2021). Equally, to ensure rapid

implementation of such methods, translational research and higher-order methods are required, with metrology playing a key role in the measurement of the CQAs of viral vectors.

1.7 Metrology in Biotherapeutic Viral Vector Characterisation

Metrology is defined as ‘the science of measurement’, and as such is a critical aspect of characterising medicines, including biologics generally, and for the purposes of this thesis, specifically biotherapeutic viral vector batches. Metrology involves using reliable reference materials and/or measurement procedures, adhering to standardised sample preparation and data processing workflows, employing validated data analysis software, and maintaining consistent instrument settings.

Developing a well-defined "measurement framework" for viral vector characterisation is essential to ensure the accurate and consistent assessment of the CQAs of such vectors and underpins the implementation of novel analytical platforms as part of the manufacturing process and of final drug substance and drug product by enabling comparability studies. Although methods have and continue to be developed for the characterisation of these biotherapeutic modalities, such a framework that is utilised across the industry is yet to emerge.

1.7.1 Viral Vectors and Measurement Challenges

Characterising rAAVs is challenging due to their complex nature. As previously mentioned, AAVs are non-enveloped, with a single-stranded DNA genome encased within the capsid (Srivastava, Lusby and Berns, 1983). Besides their complex structure, rAAVs also exhibit variability among serotypes, as well as between and within batches. This variability includes differences in PTMs, stoichiometries of the capsids, and the sequences of the encapsidated genomes. Therefore, analytical methods capable of characterising the complex nature of viral vectors are needed (van der Loo and Wright, 2016).

1.7.2 Viral Vector Reference Materials

Comparison of the data from various viral vector characterisation studies is essential for understanding the properties of viral vectors. Incorrect measurements can lead to improper conclusions about the quality of the viral vectors under assessment. In a worst-case scenario, this could lead to the development or release of ineffective or even deleterious treatments.

For these reasons, robust metrology science needs to play an essential role in ensuring the accurate characterisation and safety of viral vectors destined for use in the clinic (Moullier and Snyder, 2012).

A key aspect of achieving accurate analytical measurements and reducing the variability of results between different labs is the generation of reference measurement procedures and/or reference materials. However, the generation of reference materials for viral vectors is far from straightforward. The characterisation of the components of viral vectors is challenging due to their high complexity, as discussed further below. Moreover, the properties of viral vectors are highly variable and are influenced by process parameters within the upstream and downstream bioprocess. In addition, viral vector preparations can be sensitive to environmental factors and formulation components. Therefore, it is important to ensure that their properties have not changed over time, as these changes could compromise the reliability of the reference material and impact the supply over time (Gruntman *et al.*, 2015). Thus, re-evaluating the properties of the reference material periodically is essential to ensure its authenticity. For these reasons, there is currently a limited number and availability of reference materials for different viral vectors used clinically, and there are often limitations on the amounts of a given viral vector reference material that a laboratory can obtain.

Traceability to the International System of Units (SI) requires the establishment of an unbroken chain of calibrations to reference materials and/or measurement procedures (Beastall, 2018). Such traceability allows the accuracy of measurements to be established, which is an essential part of quality control (Beastall, 2018). However, achieving traceability to SI is challenging because of the high complexity of viral vectors. Additionally, whenever changes occur in the manufacturing process, it is essential to demonstrate comparability between different batches. Moreover, standardising the ancillary raw materials used in viral vector production is essential in order to guarantee a consistent and reproducible characterisation process. This is important because the final product might vary between different manufacturers depending on the intended application. A framework for enhancing the quality and consistency of the ancillary materials has been provided in the documentary standard "ISO 20399:2022 Biotechnology — Ancillary materials present during the production of cellular therapeutic products and gene therapy products" (International Organisation for Standardisation (ISO), 2022). Furthermore, "ISO 16921-2: Biotechnology - Gene Delivery

Systems - Part 2: Guide for Methods for the Qualification of Viral Vectors” provides guidance for determining the physical and functional titre of viral vectors that should be central in ensuring consistent characterisation of viral vectors (International Organisation for Standardisation (ISO), 2023).

1.7.3 rAAV Vector Reference Materials

As outlined previously, consistency in measurements and protocols is essential in viral vector research. The lack of standardisation in methods can make it difficult to compare data across studies. To address this, reference standard materials (RSMs) of rAAV serotype 2 (Cat No. VR-1616) and rAAV serotype 8 (Cat No. VR-1816) have been created to facilitate the comparison of data from preclinical and clinical studies conducted by different laboratories (Moullier and Snyder, 2008; Lock *et al.*, 2010). The rAAV2 and rAAV8 reference materials are available through American Type Culture Collection (ATCC). These reference materials are intended to be used to standardise internal standards and analytical methods, and to enable comparisons of drug product and drug substance batches used in different non-clinical and clinical studies. Furthermore, the reference materials can facilitate the development and implementation of new methods by providing comparability data.

The production and characterisation of the ATCC rAAV2 (ATCC, Ref: VR-1616) and rAAV8 (ATCC, Ref: VR-1816) reference materials are being overseen by the AAV2 and AAV8 Reference Standard Working Group (AAV2RSWG and AAV8RSWG). The rAAV2 and rAAV8 reference materials were characterised by 16 laboratories worldwide for the following attributes: a) the particle (capsid) concentration using enzyme-linked immunosorbent assay (ELISA), b) vector genome titre of DNase I treated lysed capsids using quantitative polymerase chain reaction (qPCR), c) infectious titre using 50% tissue culture infectious dose (TCID₅₀) assays, and d) purity and capsid identity using sodium dodecyl sulfate – polyacrylamide gel electrophoresis (SDS-PAGE) (Lock *et al.*, 2010). The rAAV2 reference material was also characterised for its transduction titre by detecting GFP-expressing cells by fluorescence microscopy (Moullier and Snyder, 2012). **Table 1.3** provides an overview of the current methods used for the characterisation of viral vectors.

Table 1.3 Summary of analytical techniques used for viral vector characterisation.

Method	Scope	Measured CQAs	References
Quantitative PCR	Vector genomic titre, infectious titre, detection of residual DNA	Potency, *Purity, Safety	(Lock <i>et al.</i> , 2010; D'Costa <i>et al.</i> , 2016)
Digital PCR	Vector genomic titre, infectious titre, detection of residual DNA	Potency, *Purity, Safety	(Zanker <i>et al.</i> , 2022; Hashiba <i>et al.</i> , 2024; Duong <i>et al.</i> , 2025)
Next-generation sequencing	Methylation analysis, identification of packaged sequences and residual DNA	Identity, Purity, Safety	(Guerin <i>et al.</i> , 2020; Namkung <i>et al.</i> , 2022; Dunker-Seidler <i>et al.</i> , 2025)
Enzyme-linked immunosorbent assays	Capsid titre, Detection of Host cell proteins (HCPs)	Potency, Purity, Safety	(Lock <i>et al.</i> , 2010; Graf <i>et al.</i> , 2021)
Mass Spectrometry	Capsid stoichiometry, serotype identification, detection of HCPs and PTMs	Identity, Purity, Safety	(Mary <i>et al.</i> , 2019; Wörner <i>et al.</i> , 2021; Leibiger, Min and Lee, 2024)
Optical Density	Vector genomic titre, capsid titre, empty/full ratio	Potency, Purity	(Sommer <i>et al.</i> , 2003)
Analytical Ultracentrifugation	Aggregation, empty/partial/full ratio	Purity, Stability	(Burnham <i>et al.</i> , 2015)
Transmission Electron Microscopy	Viral vector titre, aggregation, empty/partial/full ratio	Purity, Stability	(Colomb-Delsuc <i>et al.</i> , 2022)
Charge Detection Mass Spectrometry	Empty/partial/full ratio, capsid identity	Identity, Purity	(Barnes <i>et al.</i> , 2021; McColl-Carboni <i>et al.</i> , 2024)
Anion Exchange Chromatography	Viral vector titre, empty/full ratio	Potency, Purity	(C. Wang <i>et al.</i> , 2019; Khatwani, Pavlova and Pirot, 2021)
Size Exclusion Chromatography with Multi-Angle Light Scattering	Aggregation, viral vector titre, empty/full ratio	Potency, Purity, Stability	(Cole <i>et al.</i> , 2021; McIntosh <i>et al.</i> , 2021; Troxell <i>et al.</i> , 2023)
Static and Dynamic Light Scattering Ultraviolet Visible Spectroscopy	Aggregation, viral vector titre, empty/full ratio	Potency, Purity, Stability	(Malvern Panalytical, 2020; Cole <i>et al.</i> , 2021)
Mass photometry	Empty/full ratio	Purity	(Wu <i>et al.</i> , 2022; Wagner <i>et al.</i> , 2024)

*Purity: In the context of nucleic acid measurement, purity refers to the presence or absence of residual nucleic acid.

1.8 Physical Titre

The viral vector product is evaluated using assays that quantify the physical and infectious titre of the viral vector. Physical titre is a measurement of the total amount of viral vector particles that are present in the preparation and can be expressed as genome or capsid titre based on the measurement method used (Moullier and Snyder, 2012; Martinez-Fernandez de la Camara, McClements and MacLaren, 2021). The genome titre (often expressed as viral genomes/mL) is usually compared to capsid titre to estimate the percentage of capsids packaged with the recombinant genome. Empty (or partial) capsids are ineffective, and thus, manufacturing processes attempt to limit or eliminate these, but analytical methods must be sufficient to determine those particles that are empty or have partial genomes within them. For rAAV genome titration, the commonly employed method to quantify vector genomes is by qPCR, and more recently with digital polymerase chain reaction (dPCR).

1.8.1 qPCR

qPCR is a technique that measures the amplification of DNA during the PCR process in real time. In qPCR assays that use TaqMan probes, detection is based on an oligonucleotide probe that has a fluorescent reporter dye at the 5' end and a quencher dye at the 3' end (Heid *et al.*, 1996). When the probe is intact, the quencher suppresses the reporter's fluorescence signal. However, during amplification, primers and the probe bind to their specific target sequences. As the Taq DNA polymerase extends the upstream primer and reaches the probe-hybridisation region, its 5' to 3' exonuclease activity cleaves the probe into fragments. This cleavage separates the reporter from the quencher, allowing the reporter to emit fluorescence. As the cycles progress, more probes are cleaved, leading to a gradual increase in fluorescent signal. The intensity of this signal is proportional to the initial amount of target DNA (Heid *et al.*, 1996; Arya *et al.*, 2005). The thermal cycler detects this fluorescence in real time, and the associated software generates amplification plots. The quantification cycle (C_q) is determined as the point at which the fluorescence reaches a predefined threshold (Bustin *et al.*, 2009; Ruiz-Villalba, Ruijter and Van Den Hoff, 2021). This threshold is often set automatically by the software at ten times the standard deviation (SD) of the baseline signal (Arya *et al.*, 2005).

The amplification curve in qPCR consists of three main phases. The first is the baseline phase, where the fluorescence signal begins to accumulate but remains below the detection

limits of the instrument and is masked by background noise (Arya *et al.*, 2005; Ruiz-Villalba, Ruijter and Van Den Hoff, 2021). As the reaction progresses, it enters the exponential phase, where the amount of PCR product doubles with each cycle. Eventually, the reaction reaches the plateau phase, where little to no additional PCR product is generated due to the depletion of reaction components, such as nucleotides and primers (Ruiz-Villalba, Ruijter and Van Den Hoff, 2021).

To quantify the initial amount of template DNA, a standard curve is often used (Aurnhammer *et al.*, 2012; D'Costa *et al.*, 2016; Furuta-Hanawa, Yamaguchi and Uchida, 2019). This is generated by performing qPCR on serial dilutions of a known DNA template that contains the target sequence. By plotting the C_q values against the logarithm of the concentrations of the standard, a linear standard curve is produced, which allows for accurate quantification of unknown samples.

1.8.2 dPCR

In contrast to qPCR, which quantifies nucleic acids based on fluorescence intensity, dPCR enables absolute quantification by partitioning the PCR reaction into many nanolitre-sized individual reactions. Amplification then occurs within each partition generated. After PCR, each partition is analysed by a detection system, yielding a binary result that is either positive or negative (**Figure 1.4**). These binary outcomes are then used to calculate the concentration of the target nucleic acid in the original sample using Poisson statistics (The dMIQE Group *et al.*, 2020; Clarner *et al.*, 2021)

There are two main technologies for partitioning nucleic acid molecules: droplet-based and nanoplate-based methods. In droplet digital PCR (ddPCR), partitions are created as droplets through emulsification. This approach generally has a lower cost compared to nanoplate-based methods. An example of this technology is the QX100™/200™ ddPCR Systems (Bio-Rad Laboratories). In contrast, nanoplate-based methods distribute the PCR reaction mix across a nano-fabricated chip containing thousands of individual reaction chambers. Examples of this approach include QuantStudio™ Absolute Q™ (Thermo Fisher Scientific) and QIAcuity™ (QIAGEN). A key advantage of nanoplate-based systems is the reduced time and hands-on work required compared to ddPCR (Tiwari *et al.*, 2022; Tan *et al.*, 2023).

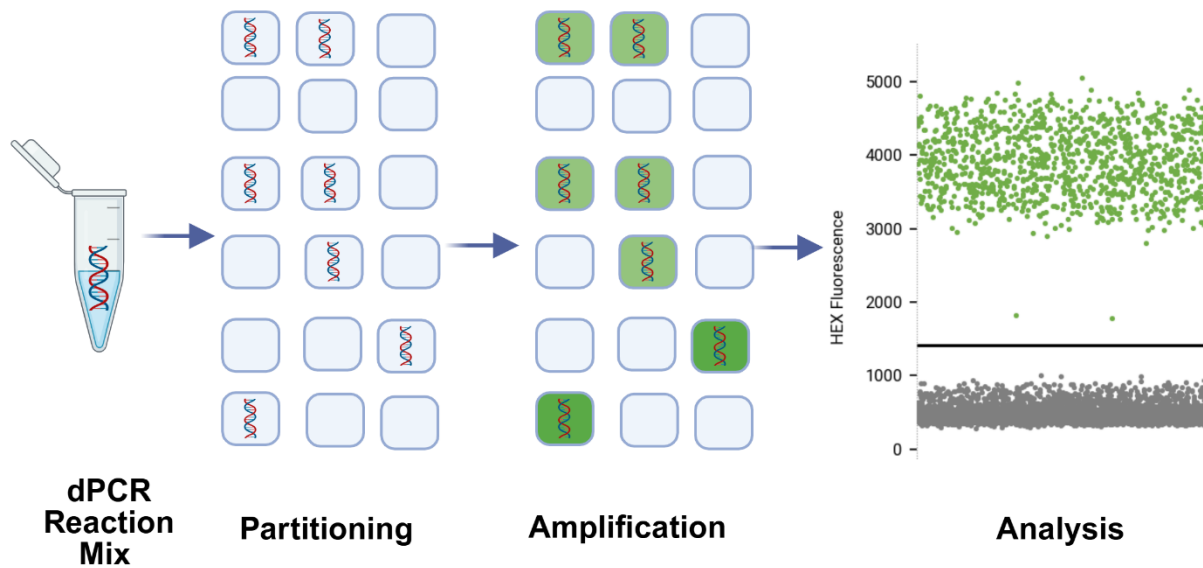


Figure 1.4 Digital PCR workflow. The reaction mix prepared is partitioned into thousands of individual reactions, called partitions. The template DNA present in each partition is amplified, and the detection system categorises each partition as positive or negative based on its endpoint fluorescence. The concentration of the template DNA is calculated based on Poisson statistics. Created with BioRender.com.

Besides differences in technology, digital PCR platforms also vary in their detection channels, which determine the compatible fluorescent dyes, as summarised in **Table 1.4**. Each platform includes a reference fluorescence channel for signal normalisation and partition detection. Excluding these reference channels, the QX200 has two detection channels, the QIAcuity has five, and the Absolute Q has four. Furthermore, dPCR platforms differ in the number of partitions analysed and the dead volume of the reaction, which refers to the portion of the reaction volume that is not used for partitioning (**Table 1.4**). The QX200 system has a dead volume of 44%, the QIAcuity 8.5k plates have 76%, while the 26k plates have 47% dead volume, and Absolute Q has only 4%. A lower dead volume is preferable as it maximises the proportion of the sample analysed, leading to quantification that better represents the actual sample.

Table 1.4 Overview of characteristics of dPCR platforms: QX200, QIAcuity, and Absolute Q.

dPCR Platforms	Channel Name	Example of fluorescent dye	Number of partitions	% of Dead Volume per Reaction
QX200	Channel 1	FAM	20,000	44
	Channel 2	HEX/VIC		
QIAcuity	Green	FAM	8,500 or 26,000	76 (8.5k plates) 47 (26k plates)
	Yellow	HEX/VIC		
	Orange	ABY™/ATTO550/ TAMRA		
	Red	ROX		
	Crimson	Cy5		
Absolute Q	Blue	FAM	20,000	4
	Green	HEX/VIC		
	Orange	ABY™/ATTO550/ TAMRA		
	Dark Red	Cy5		

Regarding the number of partitions, the QX200 system generates approximately 20,000 droplets, though this number may vary between wells and experiments. In general, droplet counts above 10,000 are considered acceptable for each reaction. The QIAcuity system offers two plate formats: the 26k plate, which generates approximately 26,000 partitions, and the 8.5k plate, which generates around 8,500 partitions. The Absolute Q system produces approximately 20,000 partitions. A higher number of partitions expands the dynamic range by decreasing the likelihood of co-partitioning at high target concentrations. Therefore, it minimises saturation effects and reduces statistical uncertainty by improving the even distribution of target molecules across partitions (Jones *et al.*, 2016).

1.8.3 rAAV Genome Titration with qPCR and dPCR

rAAV genome titration by qPCR and dPCR is conducted by using assays targeting the transgene or elements of the transgene cassette, such as the ITRs, WPRE (Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element), promoters, enhancers, and terminators (Wang *et*

al., 2020; Martinez-Fernandez de la Camara, McClements and MacLaren, 2021). However, the limitation of this method is that the design of the vectors may vary depending on the application. Therefore, the same assays often cannot be used for different vector designs. Aurnhammer *et al.*, 2012 designed a qPCR method for rAAV titration that targeted a 62-nucleotide (nt) sequence of the AAV2 stem region of the ITRs with a limit of detection (LOD) of 50 copies (Aurnhammer *et al.*, 2012). In studies evaluating methods for quantifying genomes extracted from purified rAAV virions following removal of plasmid and host cell DNA, factors such as assay design and primer/probe location were found to significantly influence qPCR and dPCR results, potentially leading to overestimation of viral titres (**Table 1.5**) (D'Costa *et al.*, 2016; Furuta-Hanawa, Yamaguchi and Uchida, 2019). These variable results highlight the need for further work to be undertaken to explore how assay design impacts the resulting titres and whether digital PCR is a better choice than qPCR for assessing rAAV genome titre measurement. Indeed, with regard to this, Furuta-Hanawa *et al.*, 2019 reported that the measurement of rAAV titre by dPCR was less affected by the hybridisation sites of the primers and probes within the ITRs, compared to qPCR. Moreover, in a study conducted by Zanker *et al.*, 2022, it was shown that further improvement of the accuracy of dPCR can be achieved by using a non-ionic surfactant (Pluronic® F68) that reduces plastic adsorption. However, in the same study, it was also reported that dPCR is impacted by rAAV aggregation because it increases the chance of a partition containing more than one genomes and thereby leading to titre underestimation (Zanker *et al.*, 2022).

Table 1.5 Summary of findings from studies investigating the effect of primer/probe location and genome conformation in rAAV titration.

Study	Materials Assessed	Target Regions	Methods Utilised	Main Findings
Aurnhammer <i>et al.</i> , 2012	rAAV2/8 extracted ssDNA	ITRs eGFP (enhanced Green Fluorescent Protein)	qPCR	Targeting eGFP resulted in 3.6-fold higher genome titres than ITR qPCR
D'Costa <i>et al.</i> , 2016	rAAV2RSM, rAAV8RSM	ITRs SV40 polyA (Simian Virus 40 polyA), GFP	qPCR	i. Targeting the ITRs (using the same assay described in Aurnhammer <i>et al.</i> 2012) resulted in 6.1 and 5.1-fold higher genome titres for rAAV2RSM and rAAV8RSM respectively than targeting the expression cassette. ii. qPCR efficiency is significantly influenced by plasmid standard conformation.
Lock <i>et al.</i> , 2014	Single-stranded rAAV (ssAAV) (serotypes 2, 2/8, 8, 9), Self-complementary rAAV (scAAV) (serotypes 1, 5, 8)	Cytomegalovirus (CMV) promoter, eGFP, bovine growth hormone (bGH), SV40 polyA sequences	qPCR dPCR	In scAAVs, dPCR and qPCR demonstrated a ~ 2-fold lower titre of CMV compared to eGFP and polyA. No significant differences were observed between the targets when ssAAVs were analysed.
Furuta-Hanawa <i>et al.</i> , 2019	rAAV2RSM, rAAV8RSM	ITRs, SV40 polyA	qPCR dPCR	The conformation of the plasmid standard leads to the observation of higher ITR than SV40 polyA titres. Instead, dPCR, which does not require a standard curve for quantification, gave comparable ITR and SV40 polyA titres

The determination of rAAV titre using various methodologies makes it difficult to compare titres between viral vector lots without comparison to a standard. The use of different techniques, such as qPCR and dPCR, where there are many variations within the two methods (such as different instrumentation, buffers, and analysis software), compounds the

further variations in sample pre-treatment and therefore can lead to differences in the observed titres. Additionally, the use of different detection chemistries, such as hydrolysis probes or SYBR™ Green for the same assays, can further compound the differences since the sensitivity and specificity of the two assays are different.

Furthermore, in qPCR, a standard curve is required for quantification, which can reduce the accuracy of genome titre measurements. This is because the standard curve introduces potential sources of variability, such as errors in the initial quantification of the standard material. Any inaccuracy in the concentration of the standard will directly affect the calculated concentration of the unknown sample (Jones *et al.*, 2016). Additionally, the commutability of the standard can lead to inaccurate genome titrations if the standard is not appropriately matched to the sample type. For example, plasmids are commonly used to generate standard curves for rAAV genome titration. However, their structural differences compared to the rAAV genome can result in overestimation of titres. Specifically, rAAV genomes have free ITR ends, whereas plasmid standards lack these free ITR ends. This structure can hinder proper denaturation of the ITRs of the standard during thermal cycling and therefore reduce amplification efficiency (D'Costa *et al.*, 2016). Furthermore, the long-term stability of the standard must be monitored, as degradation can compromise the accuracy of quantification (Rossmanith *et al.*, 2011).

In contrast, dPCR does not require a standard curve. Instead, it uses the Poisson distribution to calculate absolute copy numbers based on the number of positive and negative partitions. This approach improves both reproducibility and precision in genome titration (Zhao *et al.*, 2016; Kuypers and Jerome, 2017). Moreover, because qPCR relies on measuring fluorescence intensity, any inhibitors or interfering substances present in the sample can negatively impact the accuracy of the method. dPCR, on the other hand, performs endpoint analysis and simply categorises each partition as either positive or negative. This makes it more tolerant to inhibitors as it is less dependent on amplification kinetics (Tiwari *et al.*, 2022). Despite these advantages, qPCR has a broader dynamic range, typically spanning 7 to 8 orders of magnitude, depending on the linearity of the standard curve (Jones *et al.*, 2016; Tiwari *et al.*, 2022). On the other hand, dPCR has a more limited dynamic range due to the fixed number of partitions present in each platform (Jones *et al.*, 2016; Tiwari *et al.*, 2022).

Differences in rAAV titration results have also been observed between different dPCR platforms due to platform-specific characteristics and technologies. Wang *et al.*, 2024 demonstrated differences in measured rAAV genome titres among different dPCR instruments using an eGFP TaqMan assay. In that study, two droplet-based dPCR systems, the QX200 system from Bio-Rad and the Naica® dPCR system from Stilla, and two nanoplate-based systems, Absolute Q from Thermo Fisher Scientific and QIAcuity from QIAGEN, were compared for rAAV genome titration. The study reported a variability expressed as relative standard deviation (RSD) among the four dPCR instruments tested of 23.5 – 34.2% for measurement of rAAV5 and 11.6 – 19.9% for rAAV8, when genome extraction was not performed. However, this variability is reduced when the platforms are compared for measuring the same batch of extracted rAAV DNA, with reported variability ranging from 4.73% to 8.5% (Wang *et al.*, 2024).

Reproducibility between laboratories can be increased by using standardised and well-characterised calibrants (ATCC rAAV2RSM VR-1616 and rAAV8RSM VR-1816) and adhering to the ISO (International Organisation for Standardisation) 20395:2019 standard. This international standard provides guidelines for designing, optimising, and validating assays. Ultimately, this helps establish metrological traceability and determine measurement uncertainty in the quantification of nucleic acids (International Organisation for Standardisation (ISO), 2019).

1.8.4 ELISA for Capsid Titre Measurement

An alternative way to measure the physical titre is by quantifying the viral proteins and/or the viral particles present in the sample. The most common method for this purpose is ELISA. ELISA using the A20 antibody for rAAV2 characterisation and the ADK9 antibody for rAAV9 characterisation specifically targets the capsid proteins of fully assembled rAAV2 and rAAV9 particles, respectively. Although these antibodies can still bind to partial capsids, they do not bind to free capsid proteins, thereby eliminating interference from free capsid components in the measurement (Grimm *et al.*, 1999; Wobus *et al.*, 2000; Adachi *et al.*, 2014; Emmanuel *et al.*, 2022). This method, when combined with genome titre measurements, gives an indication of the ratio of empty to full particles present in samples.

1.8.5 Optical Density

Optical Density (OD) measurement is a simple and fast approach also used to determine the capsid and genome titre of viral vectors. However, it is limited by the purity levels of the sample. This method measures the viral vector titre and empty to full ratio by measuring the absorbance of nucleic acids and proteins at 260 and 280 nm, respectively. However, if impurities that absorb at these wavelengths are present in the sample, then the accuracy of the method is significantly reduced. These impurities can be free-floating DNA, proteins, or even buffer components that absorb in the ultraviolet (UV) range (Sommer *et al.*, 2003). Furthermore, in a study conducted by Sommer *et al.*, 2003 measurements of capsid-to-vector genome ratios conducted by OD, electron microscopy (EM), and ELISA/qPCR were compared. The OD method exhibited a good correlation with electron microscopy, with R^2 value of 0.96, and demonstrated strong agreement with ELISA-qPCR, with R^2 value of 0.995.

1.8.6 Infectious Titre

Not all viral particles are infectious. The total amount of viral particles capable of transducing cells is termed infectious titre. A widely used approach to determine the infectious titre involves measuring transgene expression following infection of a permissive cell line with the rAAV vector (Chen *et al.*, 2015). Techniques for measuring transgene expression include, but are not limited to, ELISA for protein quantification, reverse transcription-qPCR (RT-qPCR) for mRNA levels, and Fluorescence-Activated Cell Sorting (FACS) for fluorescence analysis from fluorescent transgenes.

Furthermore, endpoint assays, for example TCID₅₀, are commonly used to determine the infectious titre (Lock *et al.*, 2020; Duong *et al.*, 2023). In these assays, cells (often Henrietta Lacks (HeLa)) with integrated *rep* and *cap* sequences are infected using serial 10-fold rAAV dilutions in the presence of wild-type adenovirus 5, and the vector genome replication is detected using qPCR 48 hours after infection. Primers and probes targeting the gene of interest, or elements of the transgene cassette, are most often used to determine whether the virus is replicating in the infected cells. The TCID₅₀ assay gives an estimate of the rAAV infectious titre based on the dilutions that give a positive endpoint (positive replicates) and is calculated using the Kaerber method (François *et al.*, 2018; Y. Chen *et al.*, 2020).

Another technique commonly used for determining the infectious titre of rAAV is the infectious centres assay (ICA). In this method, a host cell line that is susceptible to rAAV vector

infection is grown in a culture dish. The rAAV sample is then added to the cells, and the cells are incubated to allow for virus infection. Next, the cells are transferred to a membrane, where the presence of fluorescently labelled virus particles, or infectious centres (IC), is detected. The number of ICs is then used to calculate the number of infectious units (IU) per mL of the rAAV sample. As such, neither the TCID50 assay nor the ICA measures transgene expression and/or protein function. François *et al.*, 2018 demonstrated that the ratio of viral genomes (vg) to infectious units (IU) for rAAV8RSM was 40 times lower when determined using the TCID50 assay compared to the ICA assay. This suggests that the TCID50 assay may have a higher sensitivity in detecting viral particles, or it may overestimate the quantity of infectious units present. These differences between TCID50 and ICA methodologies and the time-consuming nature of both techniques highlight the need for alternative, consistent, and accurate methods for measuring the infectious titre of rAAVs.

Distinguishing between titre and potency assays is important. While titre assays measure characteristics such as viral genome copies, capsid concentration, and infectivity, they differ from potency assays, which assess therapeutic efficacy through cell-based evaluations of transgene protein expression and function. However, accurate quantification of viral genome titre remains essential for dosing analysis of rAAV vectors.

1.9 Identity Assays

1.9.1 Genetic Identity

Verifying the genetic identity of recombinant viral vectors is important for ensuring high product quality and safety. Identification experiments are particularly necessary in settings where various viral vectors are produced, and serotype identification becomes essential (Wright, 2021). The verification of the genetic identity of viral vectors is achieved using nucleic acid analysis techniques such as next-generation sequencing (NGS). Guerin *et al.*, 2020 used the Illumina MiSeq™ system to determine the genetic identity of packaged sequences. In this study, it was demonstrated that rAAV DNA extracts appeared as a mixture of ssDNA and double-stranded DNA (dsDNA) molecules, and therefore, a double-strand synthesis step was not necessary in their NGS protocol to achieve good coverage and sequence depth. Guerin *et al.*, 2020 proposed that the formation of these molecules originates from the complementary

pairing of the positive (+) and negative (-) DNA strands following the DNA extraction process. This NGS viral genome sequencing protocol was successfully used for serotype verification and determination of the sequences packaged into the capsids. Tran *et al.*, 2022 demonstrated that vector genomes can exist as a mixture of species, and the chosen production platform might influence their heterogeneity. In this study, it was reported that rAAVs produced by *Spodoptera frugiperda* (Sf9) cells had 1.5-3 kb genomes that were extremely heterogeneous. Among them, unresolved and truncated genomes were identified. On the contrary, the genomes of rAAVs produced in HEK293 cells did not show the same degree of heterogeneity and existed mainly as single species (Tran *et al.*, 2022).

1.9.2 Genome Integrity

The genome integrity of rAAV particles is a CQA of increasing importance for ensuring the effectiveness and safety of gene therapy. rAAV particles with partial (truncated) genomes can be ineffective or even pose risks to patients if they express truncated or aberrant proteins. Furthermore, the evaluation of the genome integrity of the rAAV sample can help determine the appropriate dose needed to be administered to the patients. This emerging CQA can be evaluated using NGS and multiplex dPCR, which refers to the simultaneous detection of two or more targets by dPCR (Whale, Huggett and Tzonev, 2016; Hou *et al.*, 2023).

1.9.2.1 Long-Read Sequencing

The recombinant AAV genome can be up to 4.8 kb in length, which exceeds the read length limitations of short-read sequencing platforms like Illumina (typically up to 350 base pairs (bp)). Instead, long-read sequencing technologies such as Single Molecule Real-Time (SMRT®) (Pacific Biosciences) sequencing and nanopore sequencing (Oxford Nanopore Technologies) are suitable for evaluating the genome integrity of rAAVs without the need for amplification (Namkung *et al.*, 2022; Dunker-Seidler *et al.*, 2025). However, one limitation of long-read sequencing is that when long and short fragments are present in equal quantities within a sample, the platform tends to generate more reads from the shorter fragments over the same period. This bias was demonstrated by Dunker-Seidler *et al.*, 2025 who digested λ phage DNA to generate equimolar fragments of varying sizes. Despite all fragments being present at the same equimolar concentration, smaller fragments were overrepresented in sequencing data. This observation highlights the inherent short-read bias of nanopore sequencing (Dunker-

Seidler *et al.*, 2025). This bias leads to underestimation of genome integrity due to overrepresentation of smaller fragments in the final sequencing results.

Two widely used long-read sequencing methods are SMRT sequencing and nanopore sequencing. SMRT sequencing is based on real-time imaging of fluorescently labelled nucleotides as they are incorporated by DNA polymerase into a newly synthesised DNA strand, allowing single-molecule DNA synthesis to be monitored (Roberts, Carneiro and Schatz, 2013; Nakano *et al.*, 2017). A major advantage of SMRT sequencing is its exceptionally high accuracy, exceeding 99.999%. However, it requires a large amount of rAAV genomic material, typically in the range of 10^{12} to 10^{13} genomes. On the other hand, nanopore sequencing works by passing single nucleic acid molecules through nanopores embedded in a membrane within a device called a flow cell. As each nucleotide passes through a nanopore, it causes characteristic disruptions in the electrical current, generating a unique signal pattern. Each nanopore is connected to a sensor that continuously measures these current changes. The signals are detected in real time and interpreted by neural network-based basecalling software to determine the nucleotide sequence (T. Zhang *et al.*, 2024; Dunker-Seidler *et al.*, 2025).

Oxford Nanopore Technologies (ONT) offers a range of sequencing platforms, including the MinION, GridION, and PromethION, which primarily differ in sequencing capacity and scalability. The MinION supports a single flow cell with 512 nanopore channels. The GridION accommodates up to five flow cells, each with 512 channels, while the PromethION, can run up to 48 flow cells, each containing 2,675 nanopore channels (Espinosa *et al.*, 2024). Additionally, ONT has developed other platforms such as the SmidgION, a small device that can be operated with mobile devices, and the Flongle, a low-cost, single-use flow cell designed for small-scale sequencing experiments (Espinosa *et al.*, 2024; T. Zhang *et al.*, 2024).

Although nanopore sequencing initially had limited accuracy, it has improved over time. Early R9 flow cells achieved around 87% read accuracy, which increased to approximately 90% (PHRED score 10) with the R9.4 version (T. Zhang *et al.*, 2024). The PHRED score, also referred to as Q, is the standard metric for representing sequencing read quality (Q). A PHRED score of 10 corresponds to 90% basecalling accuracy, 20 to 99%, and 30 to 99.9%, with higher scores representing even greater accuracy (Cock *et al.*, 2010). A major improvement in the performance of nanopore sequencing came with the introduction of

R10.4.1 flow cells, which feature a new pore version with a larger reader head, achieving read accuracies of up to 99.1% (Ni *et al.*, 2023; Dunker-Seidler *et al.*, 2025). Further increase in accuracy was achieved by combining R10.4.1 flow cells with the latest V14 chemistry, which includes an improved motor protein (E8.2) and a new basecalling software, Dorado. Studies reported read accuracies as high as 99.9% when using R10.4.1 flow cells with V14 chemistry (Ode *et al.*, 2024; Dunker-Seidler *et al.*, 2025). These advancements have significantly reduced sequencing errors, improving the accuracy of raw reads, consensus sequences, and variant calling.

Although SMRT sequencing has higher accuracy than nanopore sequencing, the main advantages of nanopore sequencing over SMRT sequencing include its requirement for substantially lower input rAAV material (approximately 10^{10} genomes), its simpler sample preparation workflow, and its significantly lower overall cost (Dunker-Seidler *et al.*, 2025). Dunker-Seidler *et al.*, 2025 compared nanopore and SMRT sequencing data generated from the same rAAV batch, processed and filtered using the same parameters. SMRT sequencing produced higher-quality reads, with PHRED scores exceeding 85, whereas nanopore sequencing reads generally had PHRED scores between 20 and 30, with some reads exceeding 30. Both methods showed a prominent peak corresponding to the full-length rAAV genome and additional peaks representing smaller fragments below 1,000 bp. However, SMRT sequencing showed more defined peaks at the smaller fragment sizes and a lower proportion of short reads compared to nanopore sequencing, potentially indicating a reduced short-read bias (Dunker-Seidler *et al.*, 2025).

In terms of mapping, both methods showed strong agreement, with highly comparable percentages of reads mapping to production plasmids and host cell DNA. However, a SMRT sequencing showed a two-fold increase in reads mapping to the vector plasmid backbone, and a subsequent decrease in reads mapping to the rAAV vector genome. Both platforms also showed strong agreement in identifying truncation hotspots, which were consistently located within the transgene promoter region.

1.9.2.2 Multiplex dPCR

Multiplex dPCR employs probe assays with distinct fluorophores that emit signals in different colour channels. These probes target various regions of the recombinant AAV genome. When all probe assays yield positive signals, it indicates that the particle likely contains a complete

genome. Multiplex dPCR assays targeting two different regions across the rAAV genome have previously been developed, usually with one assay targeting a region near the 5' end of the genome and the other the 3' end, such as the CMV promoter or enhancer and bGH polyA regions (Zanker *et al.*, 2022). However, this can be expanded to include more assays and therefore enhance the accuracy of estimating the presence of a full genome (**Figure 1.5**).

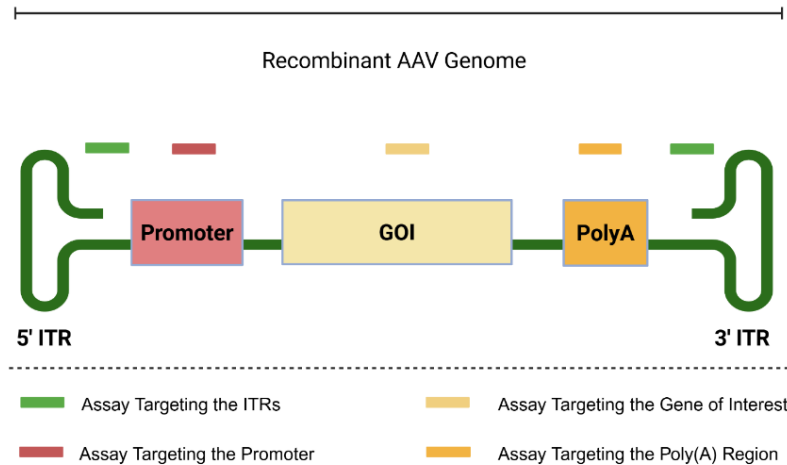


Figure 1.5 Schematic of four dPCR assays hybridising to target sites in rAAV genome. Example of a multiplex dPCR assay with four targets: ITRs (green), promoter (red), GOI (yellow), and polyA region (orange). The rectangles represent the hybridisation sites of the various assays. Due to the palindromic nature of the ITRs, the ITR assay detects both the 5' and 3' ITR regions. Created with BioRender.com.

Duong *et al.*, 2025 developed a multiplex dPCR assay targeting three regions of the rAAV genome: the promoter region at the 5' end, the gene of interest, and the polyA region at the 3' end. However, estimating genome integrity by simply calculating the proportion of partitions positive for all targets relative to the total number of positive partitions is not appropriate. This is particularly problematic at high DNA concentrations, where fragmented genomes containing different target regions may co-occupy the same partition. This makes a partition appear positive for all targets and leads to overestimation of genome integrity (Gleerup *et al.*, 2023; Tereshko *et al.*, 2023; Duong *et al.*, 2025). This issue can be mitigated by diluting the sample to reduce the probability of multiple molecules co-partitioning (Bruner *et al.*, 2019; Tereshko *et al.*, 2023). However, this method still does not account for molecular dropout. Molecular dropout occurs when a target molecule is present in a partition but fails to produce a detectable signal, often due to amplification failure. Causes of molecular dropout include template complexity, secondary structures, and inefficient primer or probe binding

(Whale *et al.*, 2013). As a result, some partitions that do contain a target may falsely appear negative and lead to underestimation of the genome integrity.

To correct for these biases, different models have been developed. Regan *et al.*, 2015 proposed a statistical probability-based model to estimate genome integrity, but this approach was limited to estimating the linkage between two targets. Subsequent models have extended this concept to more targets. Duong *et al.*, 2025 developed a mathematical model for three-target linkage estimation, while Gleerup *et al.*, 2023 introduced a different statistical probability-based approach that allows for the analysis of three and four targets, providing a more accurate and comprehensive evaluation of rAAV genome integrity. These models account for both molecular dropout and co-partitioning effects, eliminating the need to dilute samples to very low concentrations to avoid co-partitioning of molecules.

1.9.3 Assessing the Methylation Status of rAAV Genomes

The methylation status of recombinant AAV genomes is an area that has not been thoroughly investigated. In one of the few studies to date, Tóth *et al.*, 2019 conducted bisulfite PCR sequencing and showed that methylation levels were negligible in wild-type AAV2 genomes. Specifically, in packaged AAV genomes, the ratio of methylated cytosines ranged from 0-1.7%, with an average of 0.6%. Rumachik *et al.*, 2020 reported similar results with whole-genome bisulfite sequencing of rAAV genomes. In this study, it was shown that rAAV genomes are hypomethylated, and the methylation percentage observed was similar between HEK293-produced and Sf9-produced vectors. Interestingly, another study has shown that the wild-type AAV2 genome integrated into the host cell DNA is hypermethylated (Tóth *et al.*, 2019). Tóth *et al.*, 2019 showed that in integrated AAV2 genomes, CpG sites were 20.4-98.3% methylated, with an average of 76% methylated sites. When the methylation status of AAV2 was further investigated, the least methylated CpG sites were located at the ORF encoding protein X (Tóth *et al.*, 2019). Determining the methylation levels and patterns in different rAAV serotypes and across different batches, and how these factors affect the quality of gene therapies, would establish whether methylation status should be considered a CQA of rAAVs.

1.9.4 rAAV Capsid Identity and Stoichiometry

Capsid protein identity is another critical factor of rAAVs that should be characterised thoroughly. Common methods of characterisation used for serotype identification are ELISA and western blot. Mass spectrometry (MS) methods combined with high-performance liquid

chromatography (HPLC) or capillary electrophoresis have emerged recently as higher-order approaches for serotype identification and capsid composition analysis because of their high reproducibility, accuracy, and robustness. Moreover, tandem mass spectrometry (MS/MS) can be used for peptide mapping and PTM identification in viral vector analyses. However, to ensure the reproducibility and comparability of MS data, it is essential to follow standardised protocols and maintain consistent data acquisition parameters across different studies.

Mass spectrometry is particularly powerful for serotype identification as it can distinguish between highly homologous rAAVs and even detect single amino acid mutations (Van Vliet *et al.*, 2009). Identification of rAAV serotypes traditionally has required time-consuming sample preparation and long analysis times. To address this, researchers have developed MS approaches for rapid rAAV serotype identification and characterisation of both natural and modified rAAV capsids (Zhang *et al.*, 2018; Lam *et al.*, 2022). In this regard, differences in the intact masses of VP1, VP2, and VP3 between the rAAV serotypes can be used to correctly verify which rAAV serotype is present in a sample. Methodologies such as SDS-PAGE cannot efficiently separate the same VP protein of different rAAV serotypes (e.g. VP1 from different serotypes) due to the small molecular weight differences between them (**Table 1.6**). Moreover, western blot is a time-consuming technique, and the antibodies used are likely to cross-react with the capsid proteins of other rAAV serotypes due to their high homology (ranging from 55 to 99% homology based on which serotypes are compared) (DiMattia *et al.*, 2012). In contrast, liquid chromatography coupled to mass spectrometry (LC-MS) is a sensitive and robust method that can rapidly measure the masses of intact capsid proteins and identify the rAAV serotype(s) present (Lam *et al.*, 2022). It is essential to accurately measure the masses of all three VPs because the mass of one of the VPs could be indistinguishable between two or more serotypes. Jin *et al.*, 2017 performed reverse phase LC/MS to accurately measure the masses of VP1, VP2, and VP3 of different rAAV serotypes (1, 2, 5, 7, 9, rh10) and the pattern of the VP1, VP2, and VP3 masses was used to identify each serotype. Jin *et al.*, 2017 distinguished the rAAV1 and rAAV9 serotypes by comparing the mass measurements of VP2 and VP3. Similarly, to distinguish the serotypes rAAV1 and rAAV6, the VP1 mass was used, since the VP2 and VP3 masses of the two serotypes are only 2-3 Da different. The difference between the predicted and measured masses for all serotypes did not exceed 5 Da in this study (Jin *et al.*, 2017).

Table 1.6 The masses of VP1, VP2, VP3 for AAV serotypes (1-12).

Serotype	Mass VP1 (Da)	Mass VP2 (Da)	Mass VP3 (Da)
AAV1	81,286	66,093	59,517
AAV2	81,856	66,488	59,974
AAV3	81,571	66,319	59,849
AAV4	80,550	65,626	59,529
AAV5	80,336	65,283	59,463
AAV6	81,322	66,096	59,519
AAV7	81,564	66,372	59,101
AAV8	81,667	66,519	59,805
AAV9	81,291	66,210	59,733
AAVrh10	81,455	66,253	59,634
AAV10	81,477	66,271	59,638
AAV11	80,987	65,794	59,696
AAV12	82,106	66,905	59,846

An alternative method that can be used for fast identification of the AAV serotype is differential scanning fluorimetry. Each AAV serotype exhibits distinct structural stability, and this characteristic enables serotypes to be identified based on their capsid melting temperature (T_m). While differential scanning fluorimetry has effectively been used to determine the T_m of various serotypes, distinguishing between rAAV serotypes with closely similar thermostability poses a challenge (Rieser *et al.*, 2020).

It is generally accepted that the intact rAAV capsid is composed of approximately 60 copies of the VP proteins with a ratio of VP1:VP2:VP3 of 1:1:10 on average across the capsid population (Liu *et al.*, 2020). However, Bosma *et al.*, 2018 demonstrated that stoichiometry depends on the relative expression levels of each VP protein and Wörner *et al.*, 2021 suggested that capsid assembly is stochastic and that there is no absolute constant rAAV capsid stoichiometry. Snijder *et al.*, 2014 reported that rAAV1 capsids contained 0 to 2 copies of VP1, 8 to 11 copies of VP2, and 48 to 51 copies of VP3, a stoichiometry that substantially deviates from the originally assumed 1:1:10 ratio. Furthermore, Wörner *et al.*, 2021 used

Orbitrap ultra-high mass range (UHMR) analysis to perform high-resolution native MS on different rAAV serotypes (1,5,8,9) and found that each serotype consists of a heterogeneous population of particles with variable VP stoichiometry. According to the study's stochastic assembly model, even the most abundant stoichiometry was present in only a small proportion of all rAAV9 capsids (< 3%), with a bulk average ratio of 3:10:47 (Wörner *et al.*, 2021). Interestingly, Wörner *et al.*, 2021 demonstrated that the VP expression levels vary between different serotypes and expression platforms. These findings are in agreement with the earlier study of Urabe *et al.*, 2002, who reported that VP1 expression is marginally higher in HEK293 cells compared to insect cells by western blot. Importantly, Bosma *et al.*, 2018 demonstrated that VP stoichiometry impacts infectivity. In this study, vectors with a low VP1:VP2 ratio (< 0.5) were associated with reduced potency (Bosma *et al.*, 2018). In contrast, vectors with decreased levels of VP3 had increased potency and were more efficient in gene transfer *in vivo*. However, more research into how capsid stoichiometry and VP expression can affect vector infectivity, and whether this can be controlled, is necessary. To aid this work, the development of robust analytical approaches to accurately define this CQA is required.

1.9.4.1 LC-MS/MS Quantification Strategies

Mass spectrometry, due to its high accuracy, precision, and selectivity, is one of the main methods for the development of reference methods and materials for biologics (Son *et al.*, 2024). Mass spectrometry can therefore be used to quantify individual capsid proteins and determine the stoichiometry of the capsids. Mass spectrometry-based quantification approaches can be broadly categorised into label-free and label-based methods (Megger *et al.*, 2014). Label-free methods are generally considered less accurate and reliable due to systematic and non-systematic variations introduced during sample preparation and mass spectrometry acquisition, which can affect the consistency and accuracy of the results (Neilson *et al.*, 2011; Megger *et al.*, 2014). In contrast, label-based approaches are considered more accurate for protein quantification because they account for such variations (Neilson *et al.*, 2011). One such method is stable isotope dilution (SID), which allows for absolute quantification of proteins in a sample (Ciccimaro and Blair, 2010). SID is commonly used in bottom-up proteomics, especially for proteins larger than 15 kDa (Brun *et al.*, 2009). For this approach, unique peptides must be selected to represent each protein of interest. These

peptides are known as signature or proteotypic peptides (Mallick *et al.*, 2007; Brun *et al.*, 2009).

In the bottom-up workflow (**Figure 1.6**), the proteins of interest are digested with a protease to release the signature peptides. It is important to ensure complete release of these peptides so that the amount measured reflects the original protein quantity (Brun *et al.*, 2009; Ciccimaro and Blair, 2010). Subsequently, a known amount of isotopically labelled peptides is added to the sample either before or after proteolytic digestion. The labelled peptides have identical amino acid sequence and PTMs as their endogenous counterparts and therefore exhibit the same chromatographic retention time, ionisation efficiency, and fragmentation patterns, allowing them to verify the presence of the analyte in the sample (Kiyonami and Domon, 2010; Rauh, 2012). Furthermore, they act as internal standards for systematic and non-systematic variations, such as matrix effects, ionisation efficiency, and instrument response (Brun *et al.*, 2009). The mass spectrometer can distinguish between the native and labelled forms based on their mass-to-charge (m/z) ratio difference (Ciccimaro and Blair, 2010). Quantification is then achieved by calculating the peak area ratio between the endogenous and labelled peptides (Ciccimaro and Blair, 2010; Rauh, 2012).

However, peptides with the same mass but different sequence from the isotopically labelled peptides can interfere with accurate quantification by introducing ambiguity in ratio measurements. To overcome this, multiple reaction monitoring (MRM), also known as selected reaction monitoring (SRM), is used. MRM enhances specificity through a mass selection process and is performed in a triple quadrupole mass spectrometer (Kiyonami and Domon, 2010). A triple quadrupole mass spectrometer consists of three quadrupoles. The first quadrupole (Q1) is a mass analyser that selects only the specified precursor ions and directs them to the second quadrupole (Q2) for collision-induced dissociation (CID). The ions that do not match the predetermined m/z are excluded. After CID, the resulting fragment ions are transferred to the third quadrupole (Q3), which is a mass analyser and directs only the fragment ions with the selected m/z towards the detector (Kiyonami and Domon, 2010). Each pair of precursor and product ions is called a transition. Typically, three transitions are monitored for each peptide. One transition is used for quantification (quantifier transition), and the other two serve to confirm the identity of the peptide based on matching retention

times (qualifier transitions). This approach reduces the likelihood of false identification of the precursor ion and ensures high specificity. For absolute quantification, a standard curve is typically generated by analysing a series of samples containing a fixed amount of labelled peptide and varying known amounts of the unlabelled target peptide. The resulting peak area ratios of the quantifier transition of analyte to the isotopically labelled standard are then interpolated onto the standard curve to determine the absolute concentration of the analyte in unknown samples (Ciccimaro and Blair, 2010).

SID-MS can be especially useful for quantifying VP1, VP2, and VP3 capsid proteins in rAAV because these proteins share a significant part of their sequence, and there are no antibodies available that can detect each one individually with ELISA. Chromatographic methods also struggle to separate these proteins due to their similar behaviour during separation. Moreover, size exclusion chromatography (SEC) cannot clearly resolve VP2 from VP3 because their sizes are too close. While 2D gel electrophoresis can be used to estimate protein quantities by comparing the volume of stained protein spots to proteins of known amount, it is not fully quantitative (Neilson *et al.*, 2011). It also requires mass spectrometry to identify the proteins, especially when host cell proteins are present in the sample (Neilson *et al.*, 2011).

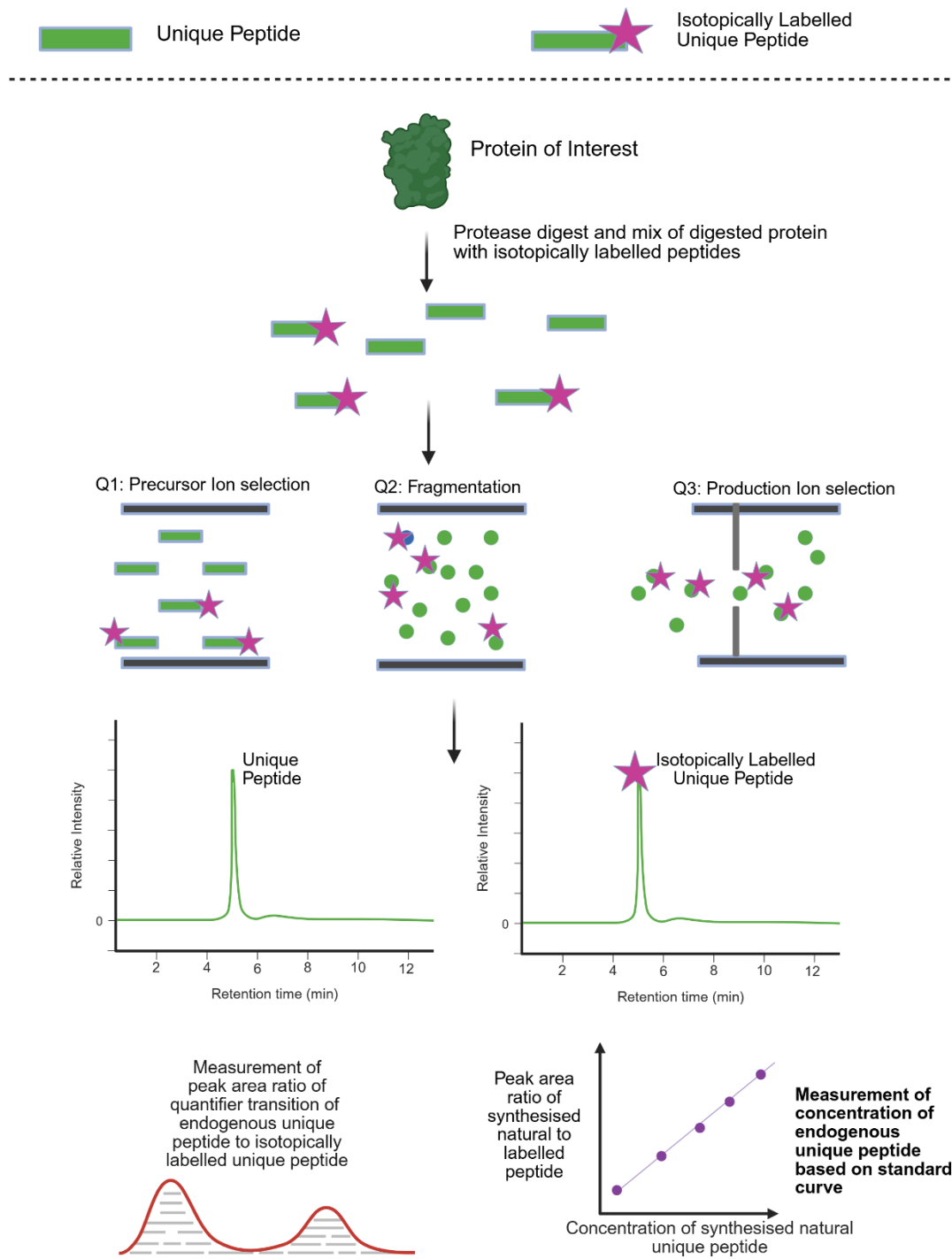


Figure 1.6 Example of an SID-MS workflow for quantification of target proteins. The protein of interest is digested using a protease and releases peptides, among them, the unique (signature) peptide. The digested protein is mixed with isotopically labelled standards for the unique peptide and analysed using an MRM method. In Q1, precursor ions are selected based on their m/z , in Q2, the ions are fragmented, and in Q3, product ions are selected and then detected. The peak areas of the quantifier transitions for endogenous unique peptides and their isotopically labelled counterparts are measured, and the ratios are calculated. A standard curve, prepared with synthesised unlabelled unique peptides of known concentrations mixed with isotopically labelled standards, is used to determine the concentrations of endogenous unique peptides in the digested sample. Created with BioRender.com.

1.9.5 rAAV Capsid Post-translational Modification Mapping and Identification

Identification of PTMs of viral vector proteins is an important analytical task that has received much less attention to date than for biotherapeutic proteins and thus requires further investigation for better characterisation of gene therapy products. The presence of particular PTMs can affect the interaction of viral vectors with cellular components and therefore impact their infectivity and tropism (Giles *et al.*, 2018; Mary *et al.*, 2019; Rumachik *et al.*, 2020). PTMs may also impact stability, propensity to aggregate, interactions with host cell proteins, and other attributes and are thus a CQA. Monitoring of PTMs and an understanding of these and how they change during bioprocessing can inform process development and help ensure product quality and consistency. For example, oxidation of capsid proteins is an important PTM to monitor during downstream processing because it represents a usual degradation pathway for proteins. High levels of oxidation could lead to loss of activity and stability of the gene therapy product (Zhang *et al.*, 2021; Guapo *et al.*, 2022).

The importance of PTMs has also been demonstrated by studies showing that the transduction efficiency and packaging of viral vectors can be significantly influenced by post-translational modifications of the rAAV capsid (Aloor *et al.*, 2018; Giles *et al.*, 2018; Mary *et al.*, 2019; Q. Chen *et al.*, 2020). In a study by Mary *et al.*, 2019, mutation of a glycosylation site of the rAAV2 capsid N253Q substantially reduced the packaging efficiency of rAAV2 (100 times), while mutation of the SUMOylation site K258Q led to reduced transduction *in vitro* (28%), and negligible expression levels of the transgene *in vivo*. These studies demonstrate it is essential to establish effective workflows for monitoring PTMs throughout the manufacturing process and investigate how different PTMs may affect the infectivity and safety of the viral vector product.

Although PTMs may vary between different AAV serotypes, Giles *et al.*, 2018 showed that for one in particular, deamidation, the degree observed was similar between the different rAAV serotypes (8, 9, 1, 3B, 4, 5, 7). They demonstrated that shelf life and vector age influenced the viral vector deamidation levels observed. The deamidation levels increased as the 'age' of the vector increased. They also reported that vector transduction efficiency decreased dramatically (> 10 times) after the mutation of asparagine (N) residues into aspartate (D), suggesting that deamidation has a substantial impact on the transduction efficiency. Furthermore, it was demonstrated that vector samples with higher levels of

deamidation had lower transduction efficiency. Interestingly, deamidation was also present in rAAV9 in a pattern and extent similar to rAAV8 (Giles *et al.*, 2018).

Potential N-linked glycosylation is another important PTM to monitor on rAAV vectors, as it can affect the tropism, immunogenicity, and intracellular trafficking of the viral vector (Aloor *et al.*, 2018). Studies have reported conflicting evidence regarding glycosylation in rAAV2 and rAAV8. Although Murray *et al.*, 2006 proposed potential glycosylation sites on the rAAV2 capsid, both Fourier transform ion cyclotron resonance mass spectrometry and a study by Liu *et al.*, 2020 failed to detect the presence of any glycosylation. In contrast, Aloor *et al.*, 2018 identified N-glycosylation at N499 in rAAV8, along with a coexisting non-glycosylated variant. Further, Mary *et al.*, 2019 employed matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) analysis to reveal multiple potential N-linked glycan combinations on the rAAV2 capsid, while no O-glycosylation was observed. Additionally, Zhou *et al.*, 2024 reported five predicted N-glycosylation sites in the rAAV9 capsids including N14, N262, N337, N452, and N497, while Xie and Butler, 2024 reported that the rAAV serotypes 1 – 9 are glycosylated with similar glycan profiles and high abundance of high-mannose glycans (HMGs).

A number of studies have been reported with the aim of identifying PTMs on rAAVs due to their potentially important impact on viral vector quality and safety. A summary of the PTMs identified in different rAAV serotypes is presented in **Table 1.7**. The discrepancies in PTMs identified among the different studies can be explained by differences in the production and purification processes used by each study. Liu *et al.*, 2020 used a wide-pore amide-bonded hydrophilic interaction chromatography (HILIC) and mass spectrometry to characterise the PTMs of two rAAV8 lots that were produced using different processes. It was discovered that one lot had higher levels of oxidation while the other had higher levels of phosphorylation (Liu *et al.*, 2020).

Table 1.7 Summary of PTMs found on the capsids of the rAAV serotypes (1-rh10).

AAV Serotype	Reported PTMs	References
rAAV2	S149 Phosphorylation; N253 Hex(4); HexNAc(2); N518, N551 HexNAcylation; S537 dHex Hex(3) HexNAc(7); K258 SUMOylation; K544 Ubiquitination	Mary <i>et al.</i> , 2019
rAAV3	A2 Acetylation; Y6 Phosphorylation; S149 HexNAcylation	Mary <i>et al.</i> , 2019
rAAV4	K479 Ubiquitination	Mary <i>et al.</i> , 2019
rAAV5	S1, S193 Acetylation; N33, N49, N55, N92, N204, N243, N371, N421, N441, N442, N492, N508, N545, N571, N585, Q432, Q472, Q587, Q603, Q687 Deamidation; Q118, Q99, Q587, N55 Succinimide; D38, D317 Isomerisation; M224, M394, M468, M494, M546, M568, M596, M634, M635, M670 Oxidation; E543 Amidation; R227, R436, R396, R559, K160, K450, K675, M468 Methylation; K122, K451, K639 Ubiquitination; S185, S186, S221, S251, S484, S680, T222, Phosphorylation; N308, T376 HexNAcylation; K312 SUMOylation; N292 N-linked Glycan core; N428 dHex Hex(3) HexNAc(6)	Mary <i>et al.</i> , 2019; Zhang <i>et al.</i> , 2021; Guapo <i>et al.</i> , 2022
rAAV6	A2 Acetylation	Mary <i>et al.</i> , 2019
rAAV7	A2 Acetylation; K61 SUMOylation; T252 Phosphorylation; S157, N254, N460 HexNAcylation	Mary <i>et al.</i> , 2019

Table 1.7 (continued)

rAAV8	M1, A2, K530, K623, K643 Acetylation; R171, R450, R487, R490, R535, R696, K61, K530, K623, K643, K709 Methylation; Y6, S149, S153, S197, S394, S501, S600, S634, S712, T203, Y396, T493, T495, Y615 Phosphorylation; T494, T663 O-GlcNAcylation; N35, N57, N94, N254, N255, N263, N305, N385, N410, N459, N499, N514, N517, N540, N630, N653, T494, Deamidation; N521 N-linked Glycan core N499 N-glycosylation	Aloor <i>et al.</i> , 2018; Mary <i>et al.</i> , 2019; Rumachik <i>et al.</i> , 2020
rAAV9	A2 Acetylation; Y52, T450, S454 Phosphorylation; K84, K316, K557 SUMOylation; N57 HexNAcylation; T702 dHex(3) HexNAc(3); K105, K650 Ubiquitination	Mary <i>et al.</i> , 2019
rAAVrh10	A2 Acetylation; Y90, S149, T252 Phosphorylation; S157, T332, S449 HexNAcylation; N306 N-linked Glycan core; K333, K652 Ubiquitination; K709 SUMOylation	Mary <i>et al.</i> , 2019

The choice of production platform and the bioprocessing conditions can therefore influence which PTMs are found on viral vectors. Rumachik *et al.*, 2020 compared the PTMs of rAAV8 and rAAV1 produced by transient transfection of HEK293 cells with rAAVs produced by baculovirus infection of *Sf9* insect cells. Even though the type of PTMs present in both rAAV8 preparations was the same, more extensive PTMs were found in the capsid proteins of insect-produced rAAV8 compared to HEK293-produced AAV8. Although there were more PTMs present in baculovirus-produced rAAV8 and rAAV1 vectors than in the HEK293-derived vectors, the PTMs identified in the different vector lots had a number of commonalities (**Table 1.8**).

Table 1.8 Summary of common PTMs found in both HEK293 and Sf9-produced rAAV8 and rAAV1 vectors reported by Rumachik *et al.*, 2020.

PTMs found in both HEK293-produced and Sf9-produced rAAV8	
Modification	PTM Position
Methylation	M1
Deamidation	N35, N57, N94, N254, N255, N305, N385, N410, N459, N499, N514, N517, N540, N630, N653
Phosphorylation	S149, S153, Y396
PTMs found in both HEK293-produced and Sf9-produced rAAV1	
Modification	PTM Position
Acetylation	M1, K143
Phosphorylation	S149, S467
Methylation	K61, K459, R465, R485, K528

The identification of PTMs on viral vectors is a relatively new field, and few studies have explored this subject. Extensive methods for PTM characterisation on proteins have been developed, particularly for biotherapeutic proteins used in the clinic and there is thus an opportunity to develop and apply such methods to the characterisation of rAAVs. In particular, MS-based approaches are well suited for high-resolution analysis of PTMs on viral vectors for characterisation of products, but also to monitor PTMs for process development purposes. The impact of conserved PTMs and how these change between serotypes or different batches could be investigated once the necessary assays are established.

1.10 Analytics and Viral Vector Purity

As discussed previously, it is essential to ensure that the final vector product is sufficiently pure and of consistent ‘makeup’ to meet the relevant specifications. For this reason, monitoring the purity during downstream processing development becomes essential. The two main categories of impurities are process-related and product-related impurities. These can be introduced into the system from the producer cells or from components of the upstream and downstream processing methods (Cashen and McLaughlin, 2023; Sauna *et al.*, 2023). Depending on the type of impurity, the type of assay to monitor or detect these varies. For example, ELISA is traditionally used to quantify the total amount of host cell proteins (HCPs) present, while the total protein content can be measured using a Bradford protein colourimetric assay with a sensitivity of 10 mg/mL (Merten *et al.*, 2011; Graf *et al.*, 2021).

Residual HCPs, derived from host cells used to produce the rAAV vector during the manufacturing process, pose a potential challenge in ensuring product purity. More recently, MS has emerged as a technique for identifying co-eluting HCPs and product impurities in the samples of interest. MS offers a significant advantage over ELISA for this purpose as it can identify and measure a substantially wider range of specific HCPs, without the need for specific antibodies for each individual protein. Total DNA content can be measured by spectrofluorimetric methods using DNA-binding fluorescent dyes (Merten *et al.*, 2011). Furthermore, host cell DNA (HCDNA) is commonly detected by qPCR targeting sequences that are frequent in the host cell genome e.g., *Alu* repeats (Y. Wang *et al.*, 2019).

1.10.1 Analysis of Empty, Full, and Partial Capsids

Separation and removal of product-related impurities is a challenge due to their close similarity to the desired product (**Figure 1.7**). These impurities must be identified and kept to the lowest possible levels as they are usually at best inactive but can also potentially pose a safety risk, although there are contrasting reports around the impact of such impurities (Mingozzi *et al.*, 2013; Gao *et al.*, 2014). For example, Mingozzi *et al.*, 2013 reported that including empty capsids in the final formulation of rAAV can actually help vector treatment doses overcome humoral immunity in patients receiving gene therapy. In contrast Gao *et al.*, 2014 reported that the addition of empty rAAV8 particles to fixed doses of rAAV8 reduced the liver transduction efficiency by up to 64%.

The capsid content can be characterised by a number of methods including analytical ultracentrifugation (AUC), transmission electron microscopy (TEM), anion exchange chromatography (AEX), charge detection mass spectrometry (CDMS), size exclusion chromatography – multiangle light scattering (SEC-MALS), static, dynamic light scattering (SLS/DLS), qPCR/ELISA, and mass photometry.

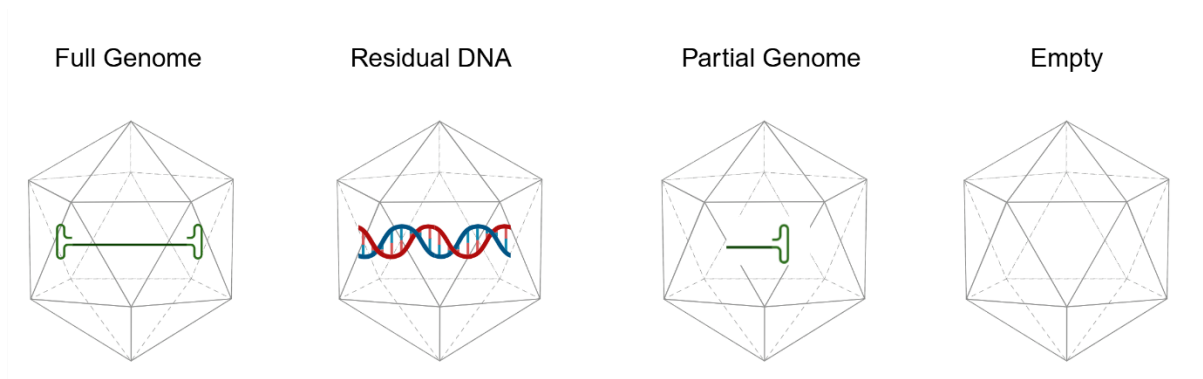


Figure 1.7 Schematic of the diverse rAAV products that can be found in an rAAV production batch. Product-related impurities can include empty capsids, capsids containing partial genomes, and capsids encapsidating residual DNA from the rAAV production process. Created with BioRender.com.

qPCR/dPCR and ELISA. A widely used method to measure the ratio of empty to full capsids is to divide the number of genome copies detected by dPCR or qPCR by the viral particles measured using ELISA (Lock *et al.*, 2010; Wang *et al.*, 2020; Gardner *et al.*, 2022). The ratio of these two values will give the ratio of empty to full capsids, assuming that one fully intact genome is packaged per viral particle or capsid. However, in addition to the fact that partial capsids (capsids that package partial genomes) are present in most samples, this approach also has other disadvantages. ELISA is a laborious and generally low-throughput technique, with low precision due to pipetting errors and/or mixing variability (Higgins *et al.*, 1998; Verch, Roselle and Shank-Retzlaff, 2016). Moreover, as mentioned previously, qPCR requires the preparation of a standard curve, is susceptible to variability in the efficiency of the PCR reaction, and can also be affected by PCR inhibitors, secondary structures, and primer-dimers that may be present in the reaction (D’Costa *et al.*, 2016; Kuypers and Jerome, 2017; Tiwari *et al.*, 2022).

Recently, providers have developed their own kits for measuring the empty-to-full ratio of rAAVs, such as the VeriCheck ddPCR Empty-Full Capsid Kit by Bio-Rad Laboratories. This kit combines droplet digital PCR with immunochemistry to determine the empty-to-full capsid ratio. The assay uses antibody probes that bind to rAAV capsids, with attached oligos that ligate together when in close proximity. A FAM-labelled probe then binds to the ligated capsid oligos, while a HEX-labelled probe binds to the ITR of the vector genome. This results in a two-colour assay that quantifies both capsid and genome titre, generating four distinct

fluorescence clusters: one negative for both capsid and genome, one positive only for the genome target, one positive only for the capsid target, and one positive for both genome and capsid targets. This classification enables precise measurement of the ratio of empty to full capsids, enhancing rAAV characterisation for gene therapy applications.

Analytical Ultracentrifugation (AUC). Sedimentation-velocity AUC is a method that utilises centrifugal force to separate particles based on their size and molecular weight. The sedimentation profiles are used to determine the distribution of particle sizes and molecular weights. This technique is used to detect the presence of aggregates and separate the empty, partial, and full capsids in a sample. Even though this method has high reproducibility with a coefficient of variation (CV) of 2%, it requires a large amount of sample for analysis and is a low-throughput method. For these reasons, AUC is commonly used as an orthogonal method to validate other, more convenient methods (Burnham *et al.*, 2015; Gimpel *et al.*, 2021; Hayes and Dobnik, 2022).

Transmission electron microscopy. TEM can be used to evaluate the size, morphology, and purity of rAAV particles (Colomb-Delsuc *et al.*, 2022). However, TEM has some limitations that should be considered when using it as an analytical approach. For example, errors in measurement can arise from weakly or unevenly stained particles and from distortion of samples during preparation. Furthermore, TEM is a low-throughput technique and requires specialised equipment and a lengthy sample preparation process (Gimpel *et al.*, 2021).

Charge Detection Mass Spectrometry. CDMS is a single particle technique that involves the simultaneous measurement of the m/z and the charge of individual ions. This method has been used to analyse empty, partial, and full viral vector particles (Pierson *et al.*, 2016; Barnes *et al.*, 2021; McColl-Carboni *et al.*, 2024). Although CDMS is a precise method, the technology is still evolving and is not currently as mature as other methods. Barnes *et al.*, 2021 used CDMS and gel electrophoresis to characterise the genomic content of rAAV8 vectors produced in insect cells by baculovirus infection. It was shown that a heterogeneous population of rAAV8 vectors was present in the preparation. In addition to empty and full particles, the authors reported the presence of partial capsids packaged with a wide range of different length DNA fragments, likely due to the packaging of truncated genomes, host cell DNA, or plasmid sequences. Similarly, Pierson *et al.*, 2016 used CDMS to distinguish between full rAAV8 capsids and empty or partial capsids, reporting a mass of 3.729 MDa corresponding

to empty rAAV8 capsids and a second mass at approximately 5 MDa corresponding to the full rAAV8 particles. The observed CDMS results were in close agreement with those of electron microscopy results, demonstrating the accuracy of this method for purity evaluation. Specifically, electron microscopy showed that 43% of the vectors were full, 30% were empty, and 27% were ambiguous or partial (Pierson *et al.*, 2016). Similarly, CDMS showed that 42.1% of the total capsids were full, 29.1 % were empty, and 28.8% were ambiguous/partial (Pierson *et al.*, 2016).

Size Exclusion Chromatography-Multiangle Light Scattering. Here, the sample is first separated by size using size exclusion chromatography. The sample fractions are then subjected to multiangle light scattering (MALS) as they pass through the column. The scattered light patterns that are produced can be analysed to determine the size (molecular weight) and shape of the sample molecules (Some *et al.*, 2019). SEC-MALS is a powerful technique that can be used to study viral vector aggregation and quantify empty and full capsids (McIntosh *et al.*, 2021; Troxell *et al.*, 2023). This method is characterised by a simple sample preparation process, high accuracy, and precision (Troxell *et al.*, 2023). To ensure the validity of the assay, the use of reference vectors that contain known quantities of full, partial, and empty capsids would be beneficial (Green and Lee, 2021).

Static and Dynamic Light Scattering. SLS/DLS provides information regarding the titre, empty-to-full ratio, and aggregation of viral particles. In general, SLS/DLS are methods used to determine the size and size distribution of particles in a sample. This is achieved by shining light on the sample and measuring the scattered light at various angles. The intensity of the scattered light is used to calculate the size and size distribution of the particles. DLS has a low overall precision and low reproducibility. Specifically, the method has been reported to have a 45% CV for capsid titre measurement and 85% CV for measurement of the aggregate content when using multiangle DLS (Malvern Panalytical, 2020; Cole *et al.*, 2021; Gimpel *et al.*, 2021).

Markova *et al.*, 2021 used SEC-static light scattering (SEC-SLS) to determine the percentage of full rAAV capsids in samples. This method determined the percentage of full capsids for rAAV2, rAAV5, and rAAV9. In the study, the authors also distinguished empty from full capsids based on their responses to thermal stress, as empty capsids are more prone to thermal destabilisation (Markova *et al.*, 2021).

Anion Exchange Chromatography. AEX utilises the variations in the electrical charge of particles to distinguish between empty and full capsids (Fekete *et al.*, 2023; Fu *et al.*, 2023). The charge differences can vary between serotypes and transgene cassettes, making it challenging to use this method in a wide range of applications. Despite this, a previous study using the method reportedly had a low coefficient of variation (1.1%) and moderate precision of 2% (C. Wang *et al.*, 2019). Furthermore, in the same study, the method was shown to have a limit of detection (LOD) and quantification (LOQ) of 1.6×10^{10} and 5×10^{10} vg/mL, respectively (C. Wang *et al.*, 2019). However, these parameters may vary depending on the sensitivity of the detector used (Tustian and Bak, 2021).

Khatwani *et al.*, 2021 developed an AEX-HPLC method with a modular discontinuous gradient for high throughput in-process quantification of empty and full capsids of different serotypes; rAAV1, rAAV2, rAAV3, rAAV6, rAAV8, and rAAV9. The method had an excellent correlation ($R^2 > 0.99$) with AUC and cryogenic-TEM (CryoTEM) for the quantification of empty and full capsids, without being time-consuming or requiring large amounts of purified material (Khatwani, Pavlova and Pirot, 2021). Similarly, Dickerson *et al.*, 2021 used AEX with an isocratic wash to separate and quantify empty and full rAAV capsids by exploiting the small difference between the surface charge of empty and full capsids.

The ratio of empty to full capsids may differ between different production platforms and even between production runs. Rumachik *et al.*, 2020 used CryoTEM and showed that HEK293-produced rAAVs have a higher percentage of full capsids compared to Sf9-produced rAAVs. Moreover, Galli *et al.*, 2021 demonstrated that rAAVs produced in *Saccharomyces cerevisiae* contain mainly truncated genomes. The sequences present in the truncated genomes were mostly ITRs or regions that flank them (Galli *et al.*, 2021).

Mass Photometry. Mass photometry quantifies empty, partial, and full rAAV capsids by detecting differences in optical contrast caused by their varying mass as they bind and unbind from a glass surface (Wagner *et al.*, 2023, 2024). It is an automated, high-throughput, and highly precise technique with a CV of less than 5% (Wagner *et al.*, 2023). It requires minimal sample preparation, offers rapid turnaround times (1 to 2 minutes), and requires only a small input volume (10 μ L of a sample at 1×10^{11} capsids/mL) (Wagner *et al.*, 2023, 2024). Furthermore, mass photometry has shown excellent linearity with AUC, with an R^2 of 0.9949 (Wagner *et al.*, 2023).

1.10.2 Analysis of DNA Impurities

To ensure the safety of gene therapies, it is essential to characterise, monitor, and remove any potential DNA impurities from plasmids or host cells (Brimble *et al.*, 2022; Dunker-Seidler *et al.*, 2025). These impurities could be packaged in capsids and trigger immune responses if administered to patients, or potentially result in the expression of undesirable elements, such as antibiotic selection markers on plasmids (Schnödt *et al.*, 2016). Furthermore, the U.S. Food and Drug Administration (FDA) considers residual DNA longer than 200 nucleotides to be of greater safety concern, as longer fragments are more likely to be transcriptionally active (U.S. Food and Drug Administration, 2020a; Dunker-Seidler *et al.*, 2025). Guerin *et al.*, 2020 demonstrated that sequences around the p5 promoter are preferentially packaged into capsids, potentially due to Rep-mediated recombination events, while Brimble *et al.*, 2022 reported that packaged DNA sequences located upstream of the p5 promoter on the AAV rep-cap plasmid are transcriptionally active. The expression of these sequences could potentially cause harmful effects to the patient. Therefore, it is necessary to develop methods that can monitor these impurities and investigate their effects on the safety of the gene therapies.

Lecomte *et al.*, 2015 developed a novel NGS method, called single-strand virus sequencing (SSV-Seq), to characterise the DNA content of rAAVs and quantify residual DNA present in preparations. The type of residual DNA varied when different purification processes were used. The main residual DNA present in CsCl-purified rAAV samples was mitochondrial DNA (mtDNA), in particular a triple-stranded structure (D loop) found in the main noncoding region of mtDNA22 (Lecomte *et al.*, 2015). To the contrary, the main residual DNA found in AVB Sepharose affinity chromatography-purified rAAV samples was chromosome 15 sequences (Lecomte *et al.*, 2015). Penaud-Budloo *et al.*, 2017 used SSV-Seq to analyse rAAV preparations and reported that the main residual DNA in rAAVs produced in insect cells was baculovirus DNA. Yet another study by Tai *et al.*, 2018 used a single-molecule real-time sequencing method called AAV-GPseq (Genome Population sequencing) with scAAVs to measure the abundance of partial genomes, reverse-packaged genomes, and packaged host cell DNA. It was demonstrated that many of these mistakenly packaged genomes were chimeric with the AAV-ITRs (Tai *et al.*, 2018). Chimeric recombinant AAV genomes extending into the transfer plasmid backbone have also been reported using nanopore and SMRT sequencing technologies (Tai *et al.*, 2018; Dunker-Seidler *et al.*, 2025).

Moreover, Zanker *et al.*, 2022 used duplex dPCR targeting the CMV promoter and the beta-lactamase gene and demonstrated the presence of the beta-lactamase sequence in 1.74% of the tested rAAV8RSM particles. Finally, Higashiyama *et al.*, 2023 developed and successfully implemented a dPCR method targeting the 18S ribosomal RNA (18S rRNA) for the quantification of residual host cell DNA. These methods provide a useful tool for rapid characterisation of residual DNA present in the rAAV samples.

In 2017, Wang *et al.*, 2017 compared three types of rAAV impurities for their effect on rAAV transduction. These included syngeneic pseudo-vectors that comprised partial rAAV genomes sourced from DNA of complete particles, allogeneic pseudo-vectors that had partial genomes distinct from those of the corresponding full particles, and null pseudo-vectors that lacked any DNA within their capsid structures. Interestingly, injection of mice with full rAAV vectors in combination with syngeneic pseudovectors led to a 2-3-fold higher transgene expression (Wang *et al.*, 2017). However, injection of mice with the allogenic pseudovectors or the empty particles, together with the full rAAV particles, did not influence transduction efficiency. Similar results were obtained *in vitro* (Wang *et al.*, 2017). It was suggested that this effect occurs because the DNA sequences of syngeneic vectors serve as primers for duplex DNA formation and therefore facilitate second-strand synthesis (Wang *et al.*, 2017).

Despite such studies, the impact of empty and partial capsids on an rAAV product's efficacy and safety is unclear. Determining whether there is a threshold number of full capsids that could transduce cells without being affected by the presence of empty or partial capsids would be interesting. Further studies investigating the effect of frequently identified mis-encapsidated DNA sequences and how their location, inside or outside the capsid, could potentially influence the safety and efficiency of the final product are required.

1.10.3 Host Cell Protein Analysis

Depending on the production system used and the purification protocol, the viral vector may co-elute with a wide variety of HCPs (Dong *et al.*, 2014; Leibiger, Min and Lee, 2024). There are various mechanisms through which HCPs can co-elute with rAAVs. These include direct interactions with the capsid or Rep proteins. For example, Nash *et al.*, 2009 demonstrated that a total of 188 HCPs interact with AAV Rep proteins. HCPs may also associate with the packaged genome or form complexes with other HCPs that interact with rAAV components (Bracewell *et al.*, 2021; Leibiger, Min and Lee, 2024). Moreover, it may even be possible that

certain HCPs are enclosed within the rAAV capsid. Identifying and quantifying these HCPs is an important task for ensuring the safety of the gene therapy products, as they could potentially cause an immune response in the person receiving the therapy (Borsotti, Borroni and Follenzi, 2016; Bracewell *et al.*, 2021; Jones *et al.*, 2021). Characterisation of HCPs in biotherapeutic protein products is well established. However, this is not necessarily the case for viral vector preparations (Bracewell, Francis and Smales, 2015). Interestingly, in a study conducted by Rumachik *et al.*, 2020, HEK293-produced rAAVs were co-eluted with HCPs that were mostly involved in the binding of nucleic acids and proteins for RNA processing, while *Sf9*-produced rAAVs were co-eluted with proteins that have roles in proteolysis/endopeptidase activity. Dong *et al.*, 2014 identified thirteen proteins co-eluting with rAAV purified by CsCl-gradient ultracentrifugation. Among them, the researchers identified the protein SET. This protein was identified in several batches regardless of the serotype or the transgene used. Interestingly, SET was only found in fractions that contained full rAAV capsids (Dong *et al.*, 2014).

In a further study conducted by Leibiger *et al.*, 2024, the five most abundant proteins across rAAV2, rAAV5, rAAV8, and rAAV9 preparations included heat shock 70-kDa protein 1B, hsc70-interacting protein isoform 1, heat shock protein HSP 90-beta isoform A, cytoplasmic dynein 1 heavy chain 1, and fatty acid synthase isoform X1 (Leibiger, Min and Lee, 2024). Furthermore, Aloor *et al.*, 2018 identified 28 host cell proteins in intracellular and secreted rAAV8 samples, while the galectin-3-binding protein was identified in both samples. The same host-cell protein was also identified in the human cell-produced rAAV8 vectors in the study conducted by Rumachik *et al.*, 2020. The host cell proteins: galectin-3-binding protein, nucleophosmin, and fibulin-1 were found in both studies (Aloor *et al.*, 2018; Rumachik *et al.*, 2020). **Table 1.9** details a subset of HCPs that have been identified in more than one study on rAAV-associated HCPs. Further studies are needed to elucidate the role, if any, of these proteins during rAAV production, if the amounts of these can be reduced or eliminated, and if there are any implications for the safety and efficacy of rAAVs due to the presence of these HCPs.

Table 1.9 Examples of HCPs that have been identified in rAAV preparations in more than one study.

Host-Cell Protein	References	UniProt Accession No.
Galectin-3-binding protein	Aloor <i>et al.</i> , 2018; Rumachik <i>et al.</i> , 2020	Q08380
Nucleophosmin	Dong <i>et al.</i> , 2014; Aloor <i>et al.</i> , 2018; Rumachik <i>et al.</i> , 2020; Leibiger, Min and Lee, 2024	P06748
Fibulin-1	Aloor <i>et al.</i> , 2018; Rumachik <i>et al.</i> , 2020	P23142
Single-stranded DNA binding protein, mitochondrial	Dong <i>et al.</i> , 2014; Aloor <i>et al.</i> , 2018	Q04837
Ferritin	Crosson <i>et al.</i> , 2018; El Andari and Grimm, 2021	Q8TD27
Heat shock 70-kDa protein 1B	Satkunanathan <i>et al.</i> , 2014; Leibiger, Min and Lee, 2024	PODMV9
Heat shock cognate protein 71kDa	Rumachik <i>et al.</i> , 2020; Leibiger, Min and Lee, 2024	P11142
Nucleolin	Dong <i>et al.</i> , 2014; Strobel <i>et al.</i> , 2015; Leibiger, Min and Lee, 2024	P19338
Y-box-binding protein 1	Satkunanathan <i>et al.</i> , 2014; Leibiger, Min and Lee, 2024	H0Y449
RuvB-like 2	Satkunanathan <i>et al.</i> , 2014; Leibiger, Min and Lee, 2024	Q9Y230
Alpha-enolase	Satkunanathan <i>et al.</i> , 2014; Leibiger, Min and Lee, 2024	P06733

As described above, and unlike biotherapeutic protein HCP analysis, where large amounts of work have been undertaken and standard ELISA and MS methods exist for monitoring the HCP content, little research has been undertaken to date on HCPs present in viral vector preparations. As such, there is a need and opportunity to investigate how HCPs change with different bioprocessing methods and assess the risks that HCPs might pose to the patient (Bracewell, Francis and Smales, 2015). Furthermore, it remains largely unknown how the location of HCPs may influence the ability to deliver the cargo to cells and achieve transgene expression. HCPs could potentially be located outside or inside the capsid or bind to the genome of the vectors. Investigating the impact of the location of HCPs on the safety and effectiveness of the gene therapy product is a research area that needs more attention. In order to facilitate this, the development and application of new methods or approaches

that enable the detection and analysis of host cell proteins in viral vector preparations is essential. Methods for characterisation of the protein content of rAAV preparations, including HCP analysis, are depicted in **Figure 1.8**.

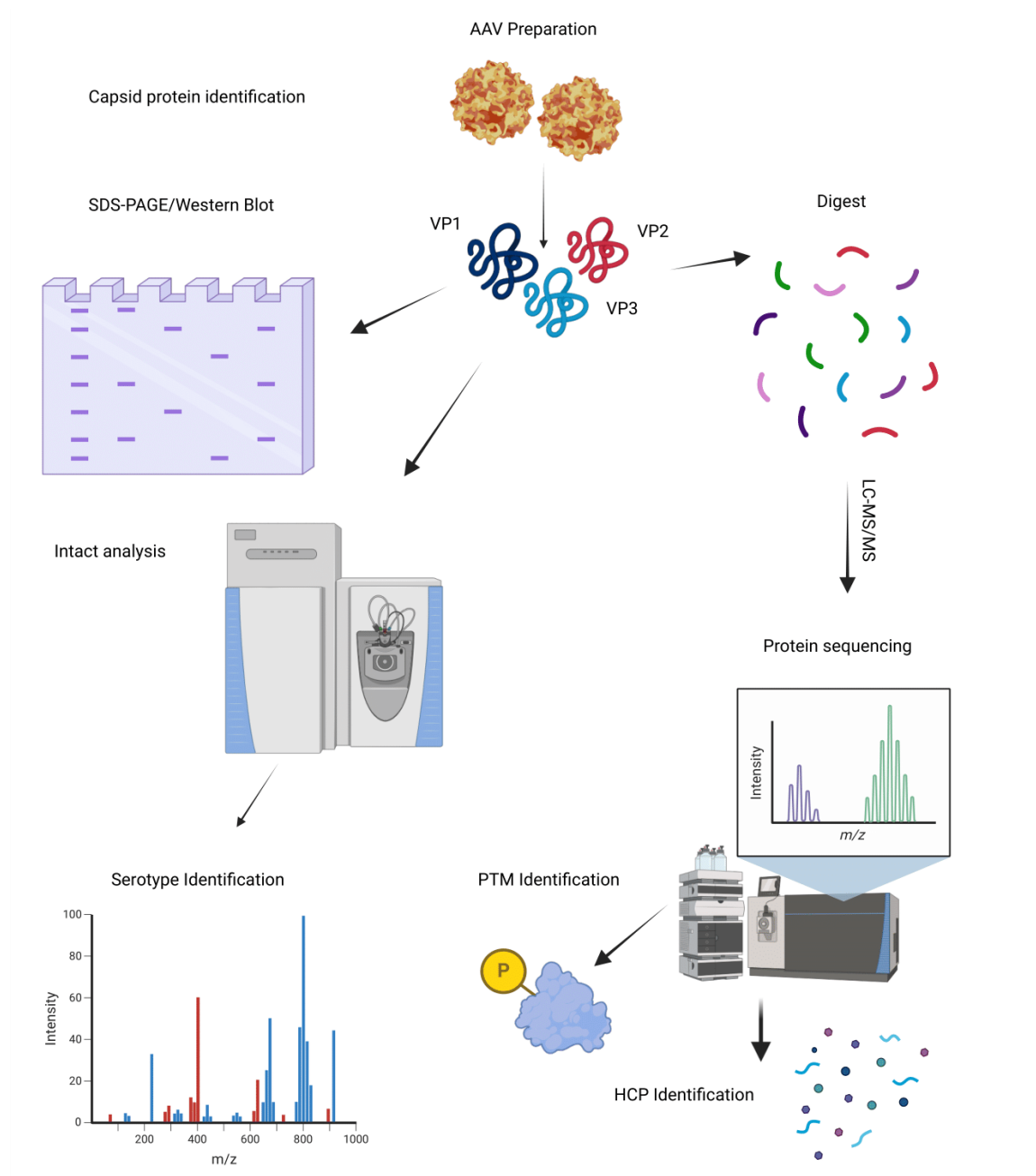


Figure 1.8 Schematic representation of protein characterisation workflows for rAAV preparations. Capsid proteins are first identified using SDS-PAGE or western blotting. Serotype identification is accomplished via intact LC-MS analysis, whereas identification of PTMs, HCPs, and protein sequencing is achieved through LC-MS/MS. Created with BioRender.com.

To conclude, the regulatory demands for gene therapy products are becoming stricter, with a need for accurate characterisation of the CQAs. To achieve this, high-throughput, precise, and accurate analytical methods must be developed to ensure the safety and efficiency of the product. Further, to ensure comparability of both the quality and quantity across diverse viral vector lots, it is important to establish harmonisation of the entire characterisation process. In this regard, reference materials play an essential role in facilitating the comparison of results across multiple studies. Additionally, it is important to outline and address potential measurement variability resulting from analytical methods employed in assessing the CQAs of viral vectors. Such comparative analysis will allow the implementation of new and improved analytical methods that advance the measurement, characterisation, reproducible manufacturing, safety, and efficacy of viral vector gene therapies.

1.11 Aims of the Work Described in this Thesis

The number of rAAV-mediated human gene therapies in development continues to rapidly increase. In order to support the design, development, manufacture, and safety of these modalities, the development of more standardised, detailed, and accurate analytical characterisation methods of rAAVs is essential. Such assays should be able to define and monitor CQAs of rAAVs to ensure safe and effective therapies. Therefore, the main aim of this study was the development of analytical methods that can accurately monitor the titre, identity, purity, and quality of the viral vectors being produced. Product-related impurities, such as capsids containing partial genomes, and process-related impurities, including HCDNA, plasmid DNA, and HCPs may compromise the therapeutic efficacy and safety of rAAV vectors by triggering immune responses (Brimble *et al.*, 2022; Leibiger, Min and Lee, 2024; Dunker-Seidler *et al.*, 2025). Within this context, a major aim of the study was to develop methods to identify these impurities in rAAV preparations and to assess genome integrity to facilitate the monitoring of the presence of partial genomes, a key product-related impurity. Genome integrity was evaluated through the development of a multiplex dPCR assay, and results were compared with nanopore sequencing data, which also enabled detection of host cell and plasmid DNA impurities. The measurement and determination of rAAV capsid stoichiometry was another key aim and objective, as variations in the ratios of VP1, VP2, and VP3 can

significantly impact vector infectivity (Bosma *et al.*, 2018). Although a 1:1:10 ratio is commonly assumed, recent studies have reported substantially different stoichiometries, such as 3:10:47, highlighting the need for accurate quantification methods (Wörner *et al.*, 2021). To support this, mass spectrometry-based methods were developed to characterise the rAAV capsids and identify PTMs and associated HCPs.

Chapter 2: Generation and Characterisation of Model Recombinant AAV Materials

2.1 Introduction

The work described in this chapter focuses on obtaining, generating and characterising rAAV materials required for developing advanced analytical methods described in later chapters of the thesis. Foundational and current practice characterisation techniques were used to analyse these materials, laying the groundwork for comparison to more advanced measurement methods.

As outlined in the introduction chapter (Chapter 1) of this thesis, ensuring the potency and safety of rAAV products for therapeutic applications requires the evaluation of both capsid and genome titres alongside other critical quality attributes (e.g. the presence of host cell proteins). Accurate measurement of the capsid and genome titres is essential not only for precise dose determination and minimising the risk of immune responses due to high capsid loads but also for informing process development (Tang, Wong and Ng, 2021; Kontogiannis *et al.*, 2024). Such information supports the optimisation of upstream and downstream processes to achieve the highest possible yields, to help maximise recovery after purification, and facilitates the comparison of different production batches to ensure consistency and quality (Gimpel *et al.*, 2021). Moreover, the ratio of genome titre to the capsid titre is used for the estimation of the proportion of particles that are packaged with the desired recombinant genome. This is critical because empty capsids, or those containing partial genomes (partial capsids), are ineffective therapeutically and may trigger immune responses, potentially compromising the safety of the therapy (Ayuso, Mingozzi and Bosch, 2010; Wright, 2014).

A further critical aspect of rAAV characterisation is the identification of impurities introduced during production or purification, with HCPs being a primary concern. HCPs could potentially adversely affect the safety and efficacy of the final product by triggering immune responses in patients (Bracewell, Francis and Smales, 2015; Bracewell *et al.*, 2021; Jones *et al.*, 2021; Leibiger, Min and Lee, 2024). Therefore, robust analytical methods are essential for

detecting and quantifying these impurities to ensure product quality, reproducibility and safety.

There are a number of well-established techniques employed to characterise rAAV preparations. ELISA is widely used to measure capsid concentration and, in combination with genome titre data, estimate the percentage of full particles (Grimm *et al.*, 1999; Gimpel *et al.*, 2021). Modern ELISA methods employ antibodies such as A20 for rAAV2 and ADK9 for rAAV9, which target conformational epitopes found only in fully assembled capsids, ensuring that free capsid proteins are excluded from the measurement (Grimm *et al.*, 1999; Wobus *et al.*, 2000; Adachi *et al.*, 2014; Emmanuel *et al.*, 2022). Despite the widespread use of ELISA for estimating rAAV particle concentrations, ELISA remains a laborious and generally low-throughput technique.

SDS-PAGE is also commonly used to assess capsid stoichiometry and detect protein impurities in rAAV preparations (Lock *et al.*, 2010). Although SDS-PAGE is a qualitative technique of low resolution, it remains widely used in the industry for rAAV characterisation (Pearson *et al.*, 2023). Moreover, estimations of the concentrations of the capsid proteins can be achieved using image analysis software to determine staining intensity against known concentrations of protein. However, staining intensity does not directly reflect protein quantity, as variations in protein-stain interactions produce bands with differing intensities (Gauci, Wright and Coorssen, 2011). This approach is also far from precise and at best gives an indication of purity and concentration.

For determination of genome titre, qPCR is widely used and often complements ELISA to allow an estimate of the proportion of genome-containing capsids. Assays used to measure the genome titre typically target elements within the recombinant genome, such as promoters and the gene of interest (Aurnhammer *et al.*, 2012). However, this approach is limited by the specific promoter and gene sequences used in each vector. To overcome this limitation, it is now common to target the ITRs of serotype 2, which are commonly present in most recombinant AAV genomes (Aurnhammer *et al.*, 2012; D'Costa *et al.*, 2016). This provides a more universal method for measuring genome titre without relying on vector-specific sequences. Although qPCR is widely used to estimate viral genome titre, it often exhibits poor intra- and inter-laboratory reproducibility, potentially due to variations in sample

handling, standard curve preparation, sample and standard storage, threshold setting, protocols, and instrumentation (Lock *et al.*, 2010).

In this chapter, rAAV2 and rAAV9 samples, to be used throughout the thesis for the development of advanced analytical methods for AAV characterisation, including digital PCR, next-generation sequencing (NGS), and mass spectrometry, were initially characterised using the traditional approaches of ELISA and qPCR. Additionally, an rAAV8 sample, selected as a suitable model for studying residual HCPs, was characterised using the bicinchoninic acid (BCA) assay and SDS-PAGE as an initial assessment of the total protein content and impurities. Together, these methods provided a baseline of measurements around rAAV composition and quality that could then be compared to the more advanced analytical approaches to be applied to their analysis.

2.2 Aims of this Chapter

The aim of this chapter was to introduce the rAAV materials (rAAV2, rAAV8, rAAV9) used throughout the thesis, and then present their initial characterisation using the established analytical techniques of ELISA, qPCR, and SDS-PAGE. The data generated from these initial measurement methods would then be used as a basis for comparison with the advanced techniques and methods developed in subsequent chapters.

2.3 Materials & Methods

2.3.1 Plasmids Used for Production of rAAV2

The plasmids that were used to transiently produce rAAV2 were sourced from Addgene and their details are as follows:

a) pAAV2/2 (packaging plasmid that encodes for the Rep and capsid proteins (from the *rep* and *cap* genes) of AAV2, Addgene Cat No. 104963). The schematic of the plasmid is shown in **Appendix Figure 7.1**.

b) pAdDelta F6 (helper plasmid that encodes for adenovirus factors that are necessary for rAAV production (E2a, E4a, VA RNA), Addgene Cat No. 112867). The schematic of the plasmid is shown in **Appendix Figure 7.2**.

c) pAAV.CMV.PI.EGFP.WPRE.bGH (transfer plasmid, expresses eGFP from CMV promoter, Addgene Cat No. 105530). The schematic of the plasmid is shown in **Appendix Figure 7.3**.

2.3.2 Preparation of Luria-Bertani Media and Agar

For the preparation of Luria-Bertani (LB) media, 5 g of yeast extract, 10 g of NaCl, and 10 g of tryptone were dissolved in 1 L of Milli-Q water. The mixture was sterilised by autoclaving. For LB agar preparation, 15 g of agar was added to 1 L of the LB medium, and the solution was autoclaved for sterilisation. After autoclaving, the medium was allowed to cool down, and ampicillin was added to a final concentration of 100 µg/mL. Agar plates were prepared by pouring the warm LB agar into Petri dishes. The plates were allowed to cool down and solidify completely, then were stored inverted at 4 °C.

2.3.3 Amplification of Plasmids and Sequence Validation

For the production of the plasmids at 250 mL scale, transformation of competent *E. coli* cells was performed. Briefly, aliquots of NEB® 5-alpha competent *E. coli* cells (New England Biolabs (NEB), Cat No. C2987H) were thawed on ice. Each 50 µL aliquot was then mixed with 1 µL (50 ng/µL) of a specific plasmid: pAAV2/2, pAdDeltaF6, or pAAV.CMV.PI.EGFP.WPRE.bGH. The plasmid-cell mixtures were incubated on ice for 25 min, then subjected to heat shock at 42 °C for 60 sec, and then immediately placed on ice for 2 min. Next, 450 µL of LB medium (without antibiotics) was added to each tube, and the cultures were incubated at 37 °C in a shaking incubator (210 rpm) for 60 min. After incubation, 50 µL from each transformation mixture was streaked onto LB agar plates containing ampicillin (100 µg/mL). The plates were then incubated overnight at 37 °C in a static incubator.

The next day, single colonies from each plate were picked and inoculated in 5 mL LB containing ampicillin (100 µg/mL). The liquid cultures were then incubated at 37 °C in a shaking incubator at 210 rpm for approximately 7 h. Subsequently, 200 µL of each liquid culture was then inoculated in flasks containing 250 mL of LB with ampicillin (100 µg/mL) and the flasks were placed in a shaking incubator at 37 °C with shaking at 210 rpm overnight. The commercially available HiSpeed Plasmid Maxi Kit (QIAGEN, Cat No. 12662) was then used to purify plasmids from the 250 mL liquid cultures following the manufacturer's protocol with elution in 50 µL of nuclease-free water (Thermo Fisher Scientific, Cat No. AM9937). The concentration of the plasmid samples was determined using a NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific). The integrity of the sequence of the three

production plasmids (pAAV2/2, pAdDeltaF6, pAAV.CMV.PI.EGFP.WPRE.bGH) was confirmed by commercial next-generation sequencing by Oya Canik (School of Natural Sciences, University of Kent).

2.3.4 PacI-digest of pAAV.CMV.PI.EGFP.WPRE.bGH Plasmid

The pAAV.CMV.PI.EGFP.WPRE.bGH was digested with PacI, which cuts at two sites, one just before the 5' ITR and one just after the 3' ITR, producing two linear double-stranded DNA fragments: one corresponding to the plasmid backbone and the other containing the rAAV2 genome sequence. The second fragment resembles the rAAV2 genome and is referred to as trFragment (transfer Fragment) for the remainder of this thesis. The rationale for generating this template is discussed in Section 2.5 (Discussion).

For the digestion of pAAV.CMV.PI.EGFP.WPRE.bGH, 250 ng of plasmid were digested with 10 units of PacI (NEB, Cat No. R0547S) in 1 × CutSmart® buffer at 37 °C in a total volume of 50 µL. The digested sample was then diluted 100-fold to minimise the effect of the rCutSmart™ buffer on analysis using a TapeStation instrument. The digest was visualised using an Agilent TapeStation 4150 system (Agilent Technologies) with the High Sensitivity D5000 ScreenTape (Agilent Technologies, Cat No. 5067-5592) and the High Sensitivity D5000 reagents (Agilent Technologies, Cat No. 5067-5593), following the manufacturer's instructions. The uncut plasmid was analysed using the ScreenTape D5000 (Agilent Technologies, Cat No. 5067-5588) and D5000 reagents (Agilent Technologies, Cat No. 5067-5589). The High Sensitivity ScreenTape was required to visualise the digested sample, as the sample had to be diluted 1:100 to minimise interference from the CutSmart® buffer and/or PacI with the quality control markers.

2.3.5 Production of rAAV2 in HEK293 cells

The rAAV2 material was produced by triple-plasmid transfection of HEK293 cells with the plasmids pAAV.CMV.PI.EGFP.WPRE.bGH, pAAV2/2, and pAdDeltaF6. The day before transfection, HEK293 cells were seeded in cell culture flasks at a concentration of 1.8×10^5 cells/mL in 6 mL of FreeStyle™ 293 Expression Medium (Thermo Fisher Scientific, Cat No. 12338018). The following day, transfection was carried out using PEIpro® DNA transfection reagent (Polyplus, Cat No. 101000017) at a 1:1:1 plasmid equimolar ratio. The PEI:DNA ratio was maintained at 1:1 (µL:µg).

Before transfection, the medium was removed and replaced with 6.5 mL of fresh FreeStyle™ 293 Expression Medium. For the transfection, a total of 50 µg of plasmid DNA was used, consisting of 13.15 µg pAAV2/2, 27.15 µg pAdDeltaF6, and 9.70 µg pAAV.CMV.PI.EGFP.WPRE.bGH, diluted in 250 µL of FreeStyle™ 293 Expression Medium. In parallel, 50 µL of PEI was diluted in 250 µL of the same medium. The PEI and DNA solutions were combined, mixed, and added dropwise to the flasks. The flasks were then incubated at 37 °C in a 5% CO₂ atmosphere in a static incubator.

2.3.6 Extraction and Purification of rAAV2 Particles

At 72 h post-transfection, HEK293 cells were harvested and the rAAV2 particles were purified from the producer cells using the AAVpro® Purification Kit Midi (Takara Bio, Cat. No. 6675) according to the manufacturer's instructions. Briefly, the cell pellet was resuspended in AAV Extraction Solution A Plus (Takara Bio) to release the viral particles. The solution was mixed by vortexing, incubated at room temperature (RT) for 5 min and then vortexed again. The suspension was then centrifuged at 5,000 × g for 10 min at 4 °C. The supernatant was then transferred to a new centrifuge tube, and AAV Extraction Solution B was added. Cryonase Cold-Active Nuclease was then added to degrade any residual host cell and plasmid DNA, followed by incubation at 37 °C for 1 h. Precipitator A was then added to the mixture, vortexed for 10 sec, and incubated at 37 °C for 30 min, followed by another round of vortexing. Precipitator B was then added, vortexed briefly for 10 sec, and centrifuged at 5,000 × g for 5 min at 4 °C. The supernatant was filtered through a Millex-HV 0.45 µm filter for removal of cell debris and other cellular components. The filtrate containing rAAV particles was then transferred to an Amicon Ultra-4 100 kDa filter device for removal of small molecules, including salts, reagents, and degraded DNA, and centrifuged at 2,000 × g for 5 min at 15 °C. The filtrate was discarded, and 1 mL of Suspension Buffer was added to the filter device. The solution was mixed uniformly by pipetting and centrifuged again at 2,000 × g for 5 min at 15 °C. This wash step was repeated four additional times, reducing the final volume to less than 0.4 mL in the filter device. After discarding the filtrate, the remaining solution inside the Amicon Ultra-4 100 kDa filter device was resuspended by pipetting. The purified rAAV2 suspension was then transferred to a new tube, which was stored at -80 °C until required for use.

2.3.7 Model rAAV Sample Origin

The rAAV samples used in this thesis share several common genomic regions and can be categorised as follows:

Inverted Terminal Repeats (ITRs)

- 5' and 3' inverted terminal repeats (ITRs): Essential for genome replication and packaging (Kontogiannis *et al.*, 2024). Although the rAAV samples used in this thesis are of different serotypes due to variations in their capsid protein sequences, all contain the ITRs of AAV serotype 2. This is a common practice in recombinant AAV production, as AAV2 ITRs are well-characterised, efficiently recognised by the replication and packaging machinery, and widely used for vector construction. As a result, all samples can be assessed using the same qPCR and dPCR assays targeting ITR2.

Enhancers

- Cytomegalovirus (CMV) enhancer: Enhances transcriptional activity, promoting strong gene expression.

Promoters

- Cytomegalovirus (CMV) promoter: Drives strong and constitutive gene expression.
- β -actin promoter: Drives gene expression in various cell types.
- Calcium/calmodulin-dependent protein kinase II alpha (CamKII α) promoter: Drives gene expression specifically in neurons.

Gene of Interest (GOI)

- Enhanced green fluorescent protein (*eGFP*) or green fluorescent protein (*GFP*): Used as a surrogate for a gene of interest (GOI).
- Reticulon 3 (*RTN3*) gene: Designed for preclinical studies, expresses the RTN3 protein.

Regulatory Elements

- Woodchuck hepatitis virus post-transcriptional regulatory element (WPRE): Enhances transcript stability and increases gene expression levels.

Polyadenylation Signals (PolyA)

- Bovine growth hormone (bGH) polyadenylation signal: Ensures correct mRNA processing and stability.
- SV40 polyadenylation signal: Used for mRNA processing and stability.

The following rAAV samples and DNA material were used in the experiments described in this thesis (**Figure 2.1**).

- The linear trFragment, which is the linear double-stranded DNA fragment originating from pAAV.CMV.PI.EGFP.WPRE.bGH, and resembles the rAAV2 genome as shown in **Figure 2.1A**. The difference between them is that the trFragment is a linear dsDNA fragment, while the rAAV2 genome is ssDNA.
- The rAAV2 particles produced during the work described in this chapter. A schematic map of the rAAV2 genome is shown in **Figure 2.1B**.
- The recombinant AAV2 Reference Material (rAAV2RefMat) was obtained from ATCC (ATCC, VR-1616). A schematic map of the rAAV2RefMat genome is shown in **Figure 2.1C**.
- The recombinant adeno-associated virus serotype 8 (rAAV8) sample 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, Cat No. 112864) that expresses the Rep and capsid proteins of the AAV serotype 8, the pAAV-CAG-GFP (Addgene, Cat No. 37825), which carries the CAG promoter (a fusion of CMV enhancer with the promoter and first intron of chicken β -actin) with the *eGFP* gene, and the pAdDelta F6 (Addgene, Cat No. 112867). The subsequent rAAV8 particles were harvested from cells and purified using a pre-packed POROS™ GoPure™ AAVX affinity chromatography column (0.5 x 5 cm, 1 mL, Thermo Scientific, Cat No. A36652). The rAAV8 particles were eluted in 100 mM citric acid (pH = 3) and neutralised in Tris buffer (pH = 8) to give a final pH = 7. A schematic map of the rAAV8 genome is shown in **Figure 2.1D**.

- Recombinant adeno-associated virus serotype 9 (rAAV9) sample suitable for pre-clinical studies was obtained from VectorBuilder Inc. and was a generous gift from Professor Giovanna Mallucci (Altos Labs, Cambridge Institute of Science, UK) (Cat No. AAV9SP(VB210216-1065qkh)). The recombinant AAV9 (rAAV9) genome consists of the CamKII α promoter and a disoptimised *RTN3* gene that is modified in the first 40 codons to slow mRNA translation elongation while preserving the encoded amino acid sequence, resulting in reduced RTN3 protein expression (Mauro and Chappell, 2014; Perach *et al.*, 2021). A schematic map of the rAAV9 genome is provided in **Figure 2.1E**.

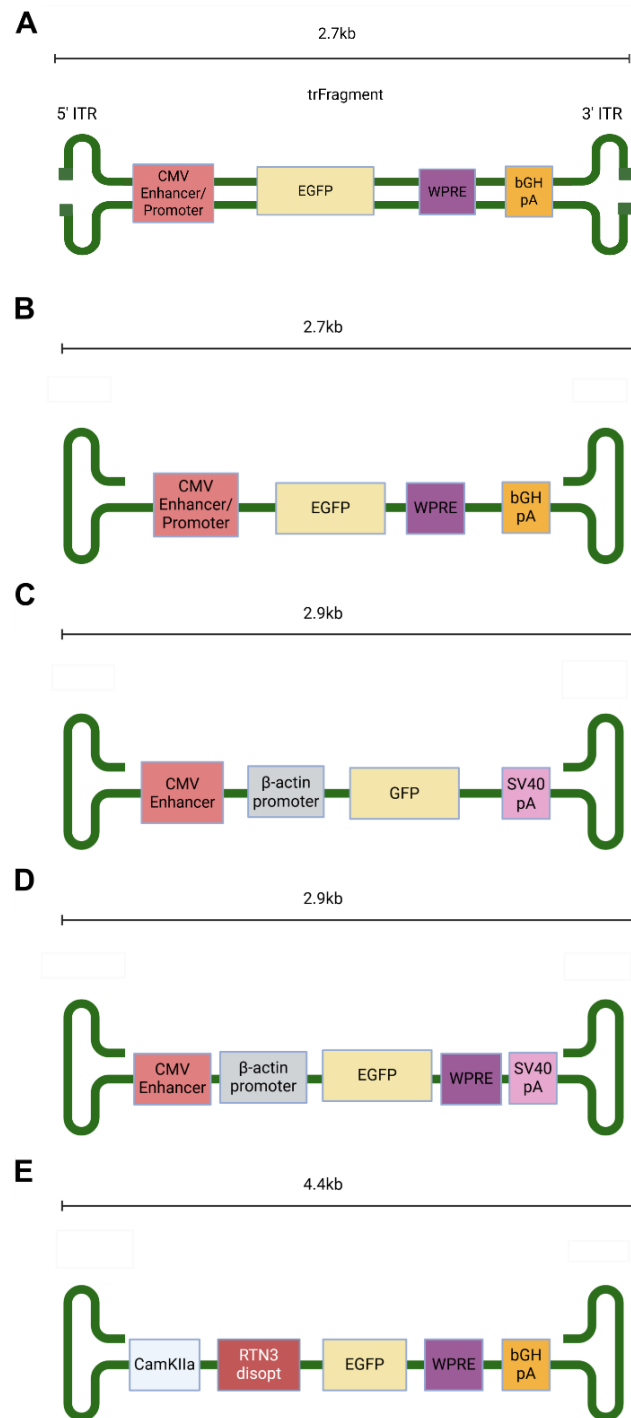


Figure 2.1 Schematic maps of recombinant AAV (rAAV) genomes used in this thesis. (A) The map of trFragment. The trFragment (2.7 kb) consists of 5' and 3' ITRs (sticky ends from *PacI* digest are shown at each end), the CMV enhancer and CMV promoter, the *eGFP* gene, the WPRE element, and the bGH polyA signal. (B) The genome of rAAV2 particles produced during the work described here (2.7 kb) consists of 5' and 3' ITRs, the CMV enhancer and CMV promoter, the *eGFP* gene, the WPRE element, and the bGH polyA signal. (C) The genome of rAAV2RefMat (2.9 kb) includes 5' and 3' ITRs, the CMV enhancer, the β -actin promoter, the *GFP* gene, and the SV40 polyA signal. (D) The genome of rAAV8 material (2.9 kb) consists of 5' and 3' ITRs, the CMV enhancer, the β -actin promoter, the *eGFP* gene, and the SV40 polyA signal. (E) The genome of the rAAV9 (4.4 kb) sample includes 5' and 3' ITRs, the CamKII α promoter, the *RTN3* gene with a disoptimised sequence, the *eGFP* gene, the WPRE element, and the bGH polyA signal. Created with BioRender.com.

2.3.8 ELISA

The capsid titres of the rAAV2 and rAAV9 preparations were measured using commercially available PROGEN ELISA kits according to the manufacturer's protocols. The AAV2 Titration ELISA Kit (PROGEN, Cat No. PRATV), employing the A20 antibody for capture and detection, was used for the rAAV2 preparation, while the AAV9 Titration ELISA Kit (PROGEN, Cat No. PRAAV9), utilising the ADK9 antibody, was used for the rAAV9 preparation. Although different kits were used for the two serotypes, the procedure for conducting the ELISA was essentially identical.

For both kits, the control (KC), which contains AAV particles of known capsid concentration, was serially diluted in $1 \times$ ASSB buffer in 1:2 steps to generate a standard curve. The KC sample in the AAV2 kit contained AAV2 particles with a titre of 2.4×10^6 capsids/ μ L, while the KC sample in the AAV9 kit contained AAV9 particles with a titre of 1.3×10^6 capsids/ μ L. The standard curve dilutions were pipetted in triplicate into the microtiter strips along with a negative control ($1 \times$ ASSB buffer only, also in triplicate).

For the rAAV2 preparation, the sample was diluted with a dilution factor of 5, while the rAAV9 preparation was serially diluted to a dilution factor of 2×10^4 . Both samples were pipetted in triplicate into the respective strips. The microtiter strips were sealed with adhesive foil and incubated at 37 °C for 1 h. Following incubation, the contents of the wells were discarded, and each well was washed three times with 200 μ L of $1 \times$ ASSB buffer to remove unbound components. Next, 100 μ L of $1 \times$ biotin, containing a biotin-conjugated monoclonal AAV antibody (AAV2-specific for the AAV2 Titration ELISA kit and AAV9-specific for the AAV9 Titration ELISA kit) in $1 \times$ ASSB buffer, was added to each well. The biotin-conjugated antibody specifically binds to the antibody-bound capsid particles, introducing biotin molecules to facilitate detection in subsequent steps. The strips were resealed and incubated at 37 °C for an additional hour, followed by three washing steps to remove any unbound biotin-conjugated antibodies.

After washing, 100 μ L of $1 \times$ Strep-HRP (streptavidin conjugated to horseradish peroxidase) in $1 \times$ ASSB buffer was added to each well. Streptavidin binds to the biotin, linking the horseradish peroxidase (HRP) enzyme to the captured capsid particles. The strips were sealed and incubated at 37 °C for 1 h, after which three washing steps were performed to eliminate any unbound Strep-HRP. Subsequently, 100 μ L of 3,3',5,5'-tetramethylbenzidine

(TMB) substrate was added to each well. The HRP enzyme catalyses the oxidation of TMB, resulting in a blue-coloured product that indicates the presence of capsid particles. The strips were then sealed and incubated at room temperature (RT) for 15 min to allow the colour to develop fully. To stop the reaction, 100 μ L of stop solution (acidic solution) was added to each well. The absorbance at 450 nm was then measured using a Tecan ST Sunrise Remote F039300 Microplate Reader. The unknown capsid concentrations of samples were determined based on the standard curve generated using the known concentrations of the KC samples and the measured absorbance values.

2.3.9 BCA Assay

The total protein content concentration of the rAAV8 preparation was determined using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific, Cat No. A65453) following the manufacturer's instructions. Briefly, the bovine serum albumin (BSA) standard (2 mg/mL) provided was diluted in Tris-EDTA buffer (TE) (Thermo Fisher Scientific, Cat No. J75893.XCR) to generate the standard curve (**Table 2.1**).

Table 2.1 Preparation of BSA standard curve.

Vial	BSA (μ g/mL)	Volume (μ L) and source of BSA	Volume of TE (μ L)
A	2,000	300 of stock	0
B	1,500	375 of stock	125
C	1,000	325 of stock	325
D	750	175 of vial B	175
E	500	325 of vial C	325
F	250	325 of vial E	325
G	125	325 of vial F	325
H	25	100 of vial G	400
Blank	0	0	400

Subsequently, 30 μ L of rAAV8 sample was added to 60 μ L of TE buffer for a 1:3 dilution. A total of 6 mL of working reagent (WR) was prepared by mixing 5.88 mL of BCA Reagent A (PROGEN) with 120 μ L of BCA Reagent B (PROGEN). Then, 25 μ L of each standard (three replicates) and the diluted rAAV8 sample (two replicates) were pipetted into a 96-well microplate (Thermo Scientific™ Pierce™, Cat No. 15041). Next, 200 μ L of WR was added to each well, and the plate was shaken for 1 minute using a plate shaker. The microplate was then sealed with adhesive foil and incubated at 37 °C for 1 h in a FOC 215i incubator (Velp Scientifica). After incubation, the plate was allowed to cool at RT for 5 min before measuring absorbance at 562 nm using a Tecan ST Sunrise Remote F039300 Microplate Reader.

The mean absorbance of the blank replicates was subtracted from the absorbance readings of both the standards and the rAAV8 sample. A standard curve (quadratic) was then generated by plotting the blank-corrected absorbance values against the known concentrations of the standards (μ g/mL).

2.3.10 SDS-PAGE Analysis

SDS-PAGE was performed on the rAAV8 sample to visualise the capsid proteins at their expected molecular weights and assess the presence of protein impurities in the preparation. A volume of 19.5 μ L (total 1 μ g) of rAAV8 sample (0.052 μ g/ μ L based on BCA assay) was mixed with 7.5 μ L of NuPAGE LDS Sample Buffer (4 $\times$) (Thermo Fisher Scientific, Cat No. NP0007) and 3 μ L of NuPAGE Reducing Agent (10 $\times$) (Thermo Fisher Scientific, NP0009), resulting in a final volume of 30 μ L. 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.

A 1 $\times$ NuPAGE SDS running buffer solution was prepared by diluting the 20 $\times$ NuPAGE MES SDS running buffer (Thermo Fisher Scientific, Cat No. NP0002) with deionised water (Elga Purelab Ultra). NuPAGE 10% Bis-Tris, 1.0 mm mini 10-well precast gels (Thermo Fisher Scientific, Cat No. NP0301BOX) were used for the analysis. Before assembly, the gel cassette was rinsed three times with 1 $\times$ NuPAGE SDS running buffer. The gel was then placed into the Mini Gel Tank (Thermo Fisher Scientific, Cat No. A25977), and the chambers were filled with 1 $\times$ NuPAGE SDS running buffer. Samples and the protein ladder were loaded carefully into the wells, ensuring the tips of the pipettes did not pierce the well bottoms. A total of 25 μ L of the prepared rAAV8 sample was loaded into each well, and 5 μ L of the PageRuler™ Plus Prestained

Protein Ladder (10–250 kDa) (Thermo Fisher Scientific, Cat No. 26619) was added to the first well to estimate the molecular weight of the capsid proteins and impurities.

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 RT, 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.

2.3.11 DNA Extraction Using the Commercial Takara Kit

To prepare the rAAV samples for qPCR analysis, it was necessary to release the DNA from the capsids. DNA extraction was performed on the rAAV2, rAAV2RefMat, and rAAV9 samples using the AAVpro[®] Titration Kit (for Real Time PCR) Ver.2 (Takara Bio, Cat No. 6233) according to the manufacturer's instructions. For each sample, 2 µL of the preparation was mixed with 2 µL of 10 × DNase I buffer, 1 µL of DNase I, and 15 µL of nuclease-free water (Thermo Fisher Scientific, Cat No. AM9937). The mixture was incubated at 37 °C on a heat block for 15 min to digest any residual DNA from host cells and plasmids used during transfection. Next, 20 µL of Lysis Buffer was added to each sample, followed by incubation at 70 °C for 10 min on a heat block to release the DNA from the capsids. The lysed samples were then diluted in EASY Dilution Buffer (provided in the kit) following the manufacturer's instructions, using a 1:50 dilution to achieve suitable concentrations for qPCR analysis. Briefly, 980 µL of EASY Dilution Buffer was added to a 1.5 mL Eppendorf DNA LoBind tube, followed by the addition of 20 µL of the extracted AAV DNA sample. This resulted in a final dilution factor of 1,000 for the rAAV2 and rAAV2RefMat samples. Due to the higher concentration, rAAV9 samples underwent three additional 1:10 serial dilutions in EASY Dilution Buffer, resulting in a final dilution factor of 1×10^6 . Each serial dilution was performed by adding 90 µL of EASY Dilution Buffer to an Eppendorf DNA LoBind tube, followed by the addition of 10 µL of the extracted rAAV9 DNA.

2.3.12 qPCR Analysis of rAAV Samples

The genome titre of the extracted AAV DNA samples was quantified using the commercial AAVpro® Titration Kit (for Real-Time PCR) Ver. 2 (Takara Bio, Cat No. 6233). The assay targets the ITRs of AAV (ITR2) using forward and reverse primers (sequence proprietary to the kit) and TB Green® as the fluorescent dye (proprietary dye). The analysis was performed across three independent experiments, each including a standard curve for accurate quantification.

The standard provided in the kit was used to prepare the standard curve and had a concentration of 2×10^7 copies/ μL (determined by OD260 measurement). The standard contained two ITR copies. This ensures good commutability between the standard and the rAAV samples, enabling accurate quantification and comparison. The sequence of the standard is proprietary to Takara Bio and is not disclosed. Serial dilutions of the standard were made in EASY Dilution Buffer at a 1:10 dilution factor. For each experiment, a fresh 50 $\times$ primer mix was prepared by combining 5 μL of AAV Forward Primer, 5 μL of AAV Reverse Primer, and 15 μL of nuclease-free water (Thermo Fisher Scientific, Cat No. AM9937). The prepared primer mix was discarded after each experiment.

In 96-well low profile, non-skirted PCR plates (Azenta Life Sciences, Cat No. 4ti-0721) the qPCR reaction mix for each well was prepared as follows: 7 μL of nuclease-free water, 12.5 μL of TB Green Premix Taq II (2 $\times$; 1 $\times$ final concentration), 0.5 μL of 50 $\times$ primer mix (1 $\times$ final concentration), and 5 μL of extracted rAAV DNA. Negative controls were included in each experiment to monitor for contamination. Nuclease-free water was used to monitor the preparation of the master mix, while EASY Dilution Buffer was used to validate the sample and standard curve dilution process. A total of 20 μL of the reaction mix was added to each well and the wells were sealed using strips of sealing caps (Azenta Life Sciences, Cat No: 4ti-0751)

The plate was then placed in a Bio-Rad C1000 Touch thermal cycler and subjected to the following thermal cycling conditions: an initial denaturation step of 2 min at 95 °C, followed by 34 cycles of 5 sec at 95 °C and 30 sec at 60 °C. Data analysis was performed using Bio-Rad CFX Manager 3.0. The threshold was set automatically each time by the software. The concentrations of the unknown samples were calculated based on the standard curve generated in each experiment.

2.4 Results

2.4.1 Sequence Confirmation of Plasmids

The consensus sequences obtained from sequencing the three plasmids (pAAV2/2, pAdDelta F6, and pAAV.CMV.PI.EGFP.WPRE.bGH) were aligned to their respective reference sequences using the Sequence Alignment tool by VectorBuilder. The software calculated the percentage identity and the EDNAFULL score for each alignment (Schwarz *et al.*, 2010). The alignment metrics for pAAV2/2, pAdDeltaF6, and pAAV.CMV.PI.EGFP.WPRE.bGH (transfer plasmid) are presented in **Table 2.2**. The percentage identity from the resulting sequencing compared to the reference sequence exceeded 99.99% for both pAAV2/2 and pAdDeltaF6. However, a 99.6% identity was achieved for pAAV.CMV.PI.EGFP.WPRE.bGH because nucleotides within the ITRs could not be reliably sequenced due to the hairpin structure and the GC-rich content (Tam Tran and WI Tai, 2024).

Table 2.2 Alignment metrics of pAAV2/2, pAdDelta F6, and the transfer plasmid (pAAV.CMV.PI.EGFP.WPRE.bGH) sequencing results against the reference sequences.

Alignment	pAAV2/2	pAdDelta F6	Transfer Plasmid
Length of reference sequence (bp)	7,468	15,421	5,514
Sequencing coverage (bp)	7,468	15,419	5,492
Identity %	100%	99.99%	99.6%
Gaps	0%	0.0001%	0.4%
EDNAFULL Score	37,340	42,718	27,416

2.4.2 Generation of DNA Templates for Assay Development

To generate the trFragment that resembled the packaged rAAV2 genome, the pAAV.CMV.PI.EGFP.WPRE.bGH plasmid was digested with the PaeI restriction enzyme, which cuts just before the 5' ITR and just after the 3' ITR (**Figure 2.2A**). The result of the digest is the generation of two linear DNA fragments, one that contains the same sequence as the packaged rAAV2 genome (2,716 bp), and one that contains the plasmid backbone (2,798 bp). The trFragment was used for assay development and as a control in the subsequent qPCR and dPCR experiments.

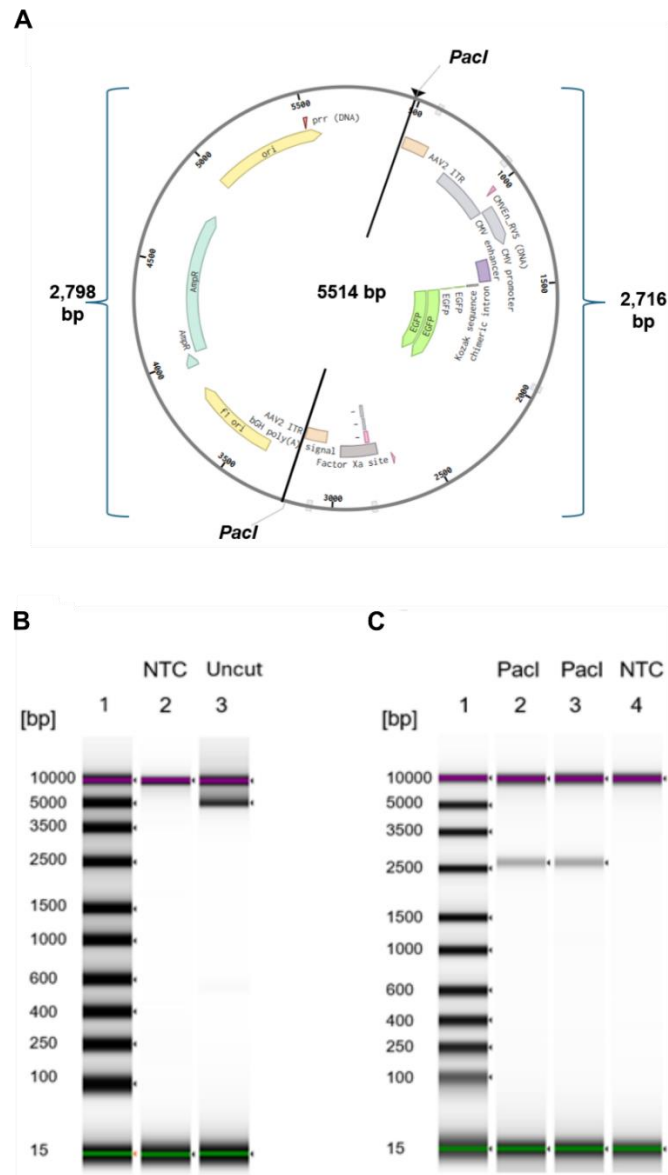


Figure 2.2 TapeStation analysis of digestion of pAAV.CMV.PI.EGFP.WPRE.bGH with PacI restriction enzyme. (A) Schematic of the plasmid with the location of the PacI restriction sites indicated on either side of the two ITR sequences. Complete digestion with PacI is predicted to produce two linearised molecules of 2,716 bp (containing the rAAV2 genome sequence) and 2,798 bp (containing the plasmid backbone). In both (B) and (C), the upper (~10 kb, purple) and lower (~15 bp, green) markers, present in the dye matrix of all samples, are visible in all lanes, indicating that all samples passed quality control. (B) TapeStation gel image using the D5000 reagents. Lane 1 corresponds to the D5000 Ladder. No bands other than the markers are visible in the NTC (lane 2). The uncut sample (lane 3) shows a single band at ~5.5 kb. (C) Lane 1 corresponds to High Sensitivity D5000 Ladder. The digested sample (PacI) was analysed in duplicate, and both replicates show a single band at ~2.7 kb (lane 2, lane 3). No bands other than the markers are visible in the NTC (lane 4).

The success of the digestion to generate the trFragment was confirmed using TapeStation (**Figure 2.2**). The uncut plasmid showed a band at the expected size of ~5.5 kb

(**Figure 2.2B**). The digested sample was analysed in two replicates using TapeStation. The absence of the 5.5 kb band and presence of the ~2.7kb band in the digested sample suggested that the plasmid had been cut at both *PacI* sites (**Figure 2.2C**). Only one band was observed in the digested sample, as the 82 bp difference between the two ~2.7 kb fragments was below the resolution limit of the technique.

2.4.3 Measurement of rAAV2 and rAAV9 Sample Capsid Content

The capsid concentrations (capsids/ μ L) of rAAV2 and rAAV9 particles were quantified by ELISA. For rAAV2, the A20 antibody was used, while for rAAV9, the ADK9 antibody was used. Both antibodies recognise conformational epitopes that are present only in fully assembled rAAV2 and rAAV9 capsids, respectively. As a result, this method specifically measures the concentration of fully assembled rAAV2 and rAAV9 capsids (McCraw *et al.*, 2012; Adachi *et al.*, 2014; Emmanuel *et al.*, 2022). Capsid concentrations were determined based on the equations derived from the standard curves ($R^2 > 0.99$), accounting for the respective dilution factors applied to each sample (**Figure 2.3**). rAAV2 samples were diluted 5-fold, while rAAV9 samples required a 20,000-fold dilution due to their higher concentration. The capsid concentration of rAAV2 was determined as $2.98 \times 10^5 \pm 1.33 \times 10^4$ capsids/ μ L with a CV of 4.5%, while the capsid concentration for rAAV9 was determined as $1.60 \times 10^{10} \pm 1.04 \times 10^9$ capsids/ μ L with a CV of 6.5% (**Figure 2.3**). The error ($\pm$) corresponds to the standard deviation (SD) of 3 technical replicates. The measured capsid titre values from ELISA were subsequently used with qPCR genome titre data to calculate the percentage of full particles.

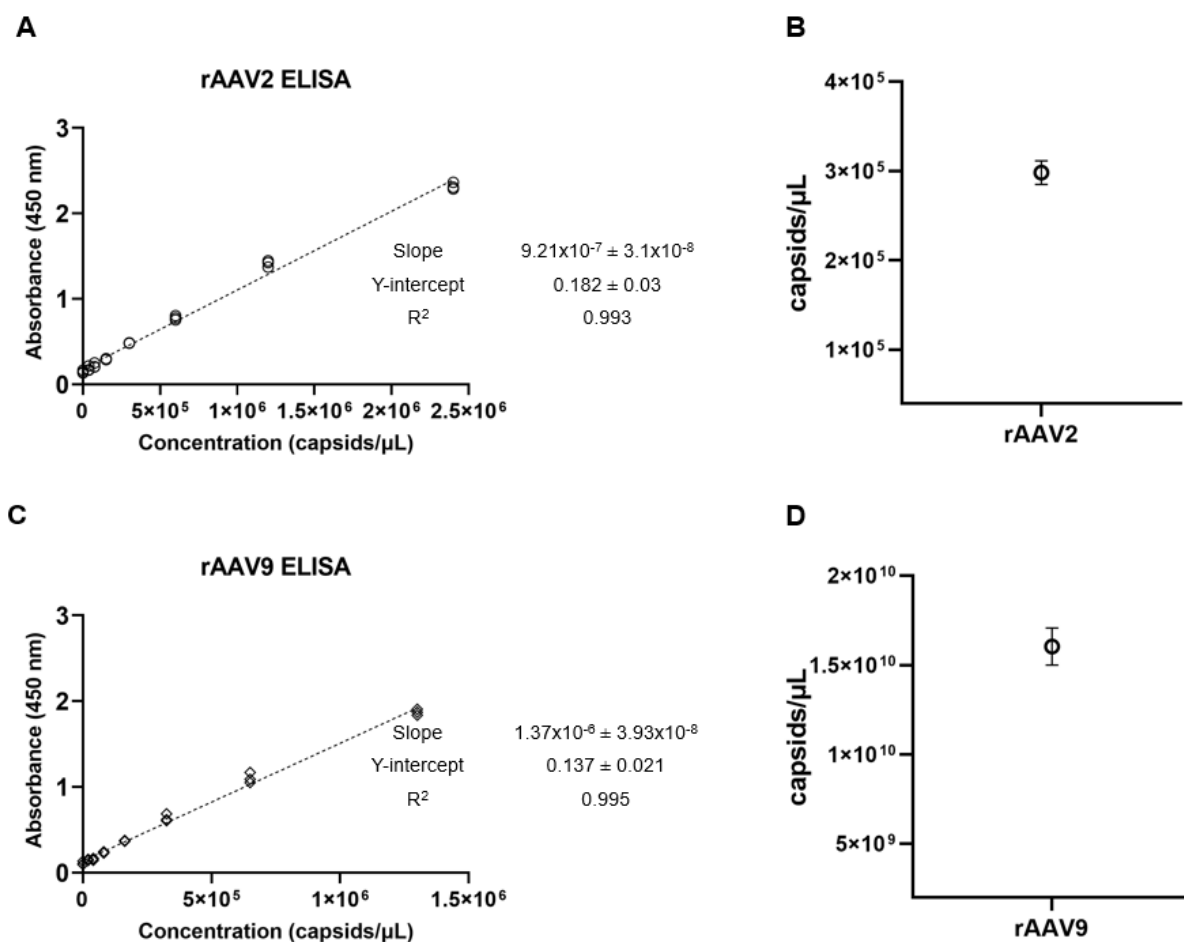


Figure 2.3 Capsid concentrations of rAAV2 and rAAV9 samples determined by ELISA. (A) Standard curve of rAAV2, showing absorbance at 450 nm (y-axis) and capsids/μL concentration (x-axis). All three technical replicates for each point of the standard curve are plotted. (B) Measured capsid concentration of rAAV2 (capsids/μL), with error bars representing the SD of 3 technical replicates. (C) Standard curve of rAAV9, showing absorbance at 450 nm (y-axis) and capsids/μL concentration (x-axis). All three technical replicates for each point of the standard curve are plotted. (D) Measured capsid concentration of rAAV9 (capsids/μL), with error bars representing the SD of 3 technical replicates.

2.4.4 Measurement of rAAV Genome Titre

To determine the genome titre and, in combination with ELISA, the percentage of full rAAV2, rAAV2RefMat, and rAAV9 particles, the samples were analysed by qPCR in three separate experiments. Additionally, two no-template controls (NTCs) were included in each experiment, both of which showed no amplification (Undetermined Cq) in every run. One NTC contained nuclease-free water, while the other included the EASY Dilution Buffer from the titration kit, which was also used for sample dilution and the preparation of serial dilutions for the standard curve. The trFragment was also included as a positive control in the analysis.

For each qPCR run, the threshold settings were automatically determined by the Bio-Rad CFX Manager software v3.0. The standard curve consisted of five dilution levels, ranging from 2×10^2 to 2×10^6 copies/ μ L (1:10 dilution steps). **Figure 2.4** presents representative amplification plots for both the standard curve (**Figure 2.4A**) and the analysed samples (**Figure 2.4C**). Standard curves were generated by plotting Cq values on the y-axis against the $\log_{10}$ -transformed concentrations of the standard curve levels. Across three runs, the average R^2 value was 0.999, indicating a strong linear relationship between the different points of the standard curve (**Figure 2.4B**). The mean amplification efficiency was $88.7 \pm 0.8\%$, with a corresponding slope of -3.626 , which falls marginally below the optimal efficiency range of 90–110% and the ideal slope range of -3.1 to -3.6 (Robledo *et al.*, 2014; Hays *et al.*, 2022). The Cq values for all analysed samples were plotted across experiments, demonstrating strong agreement across the three experimental runs (**Figure 2.4D**). Melt curve analysis was also performed for each run, confirming the absence of primer dimers and non-specific amplification (**Figure 2.4E**). All test samples (rAAV and trFragment) exhibited a melting peak at 83.5°C , whereas the standard curve samples showed a melting peak at 84.0°C (**Figure 2.4F**).

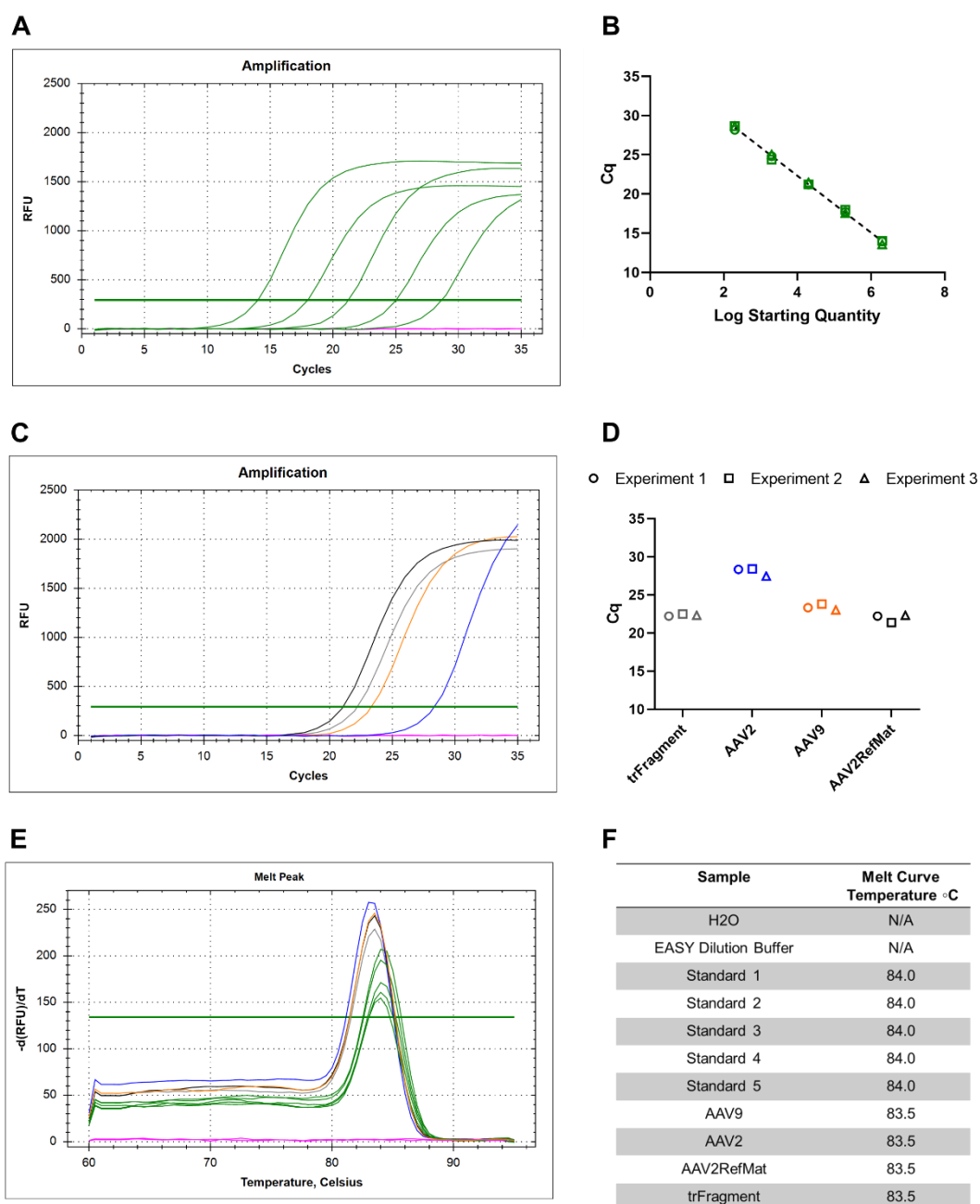


Figure 2.4 qPCR analysis of rAAV genomes. (A) Amplification plot displaying the amplification curves of the standards in green, with non-template controls (NTCs) highlighted in pink. (B) Standard curve generated from three independent experiments demonstrating high linearity ($R^2 > 0.99$), with the x-axis representing the log-transformed starting quantity and the y-axis showing the quantification cycle (Cq) values. All three experimental replicates are plotted. (C) The amplification plot for rAAV samples: rAAV2 (blue), rAAV9 (orange), rAAV2RefMat (black), and trFragment (grey). (D) A comparison of Cq values across three independent experiments for each analysed sample, with circles representing experiment 1, squares representing experiment 2, and triangles representing experiment 3. (E) Melt curve analysis of the standard curve, rAAV samples, and NTCs, using the same colour scheme as described above. The y-axis corresponds to the rate of change of fluorescence intensity, while the x-axis corresponds to the melting temperature ($^{\circ}\text{C}$). (F) A summary table presents the exact melt curve temperatures for each analysed sample. Panels (A), (C), and (E) are software outputs from Bio-Rad CFX Manager 3.0.

The genome titres of rAAV2RefMat, rAAV2, and rAAV9 were determined using qPCR (**Figure 2.5**). The measured concentration of the trFragment sample by qPCR with an ITR assay was compared to its nominal concentration determined by UV-Spec on a NanoDrop 2000 spectrophotometer. qPCR gave a lower concentration ($3.4 \times 10^8 \pm 5.8 \times 10^7$ copies/ μ L, CV of 17.0%), with a difference of $68.8\% \pm 16.2\%$ relative to UV-Spec (NanoDrop) ($1.1 \times 10^9 \pm 3.0 \times 10^8$ copies/ μ L, CV of 27.3%).

The rAAV2RefMat titre was evaluated against the reference genome titre range provided by Lock *et al.*, 2010. The rAAV2RefMat sample was provided as $3.82 \times 10^7 \pm 2.97 \times 10^7$ viral genomes/ μ L (2.7×10^7 to 4.75×10^7 copies/ μ L with 95% confidence) and was used to assess the performance of the qPCR assay relative to the reported concentration. The 95% confidence interval reported was established based on an interlaboratory study in which 16 laboratories measured the genome titre of AAV2RSM using qPCR with a TaqMan assay targeting the SV40 polyA region (Lock *et al.*, 2010). Although the qPCR assay used here targets the ITR region and is a SYBR® Green-based assay, the qPCR concentration of rAAV2RefMat determined in this work was $2.16 \times 10^7 \pm 2.64 \times 10^6$ copies/ μ L (CV of 12.2%). The error ($\pm$) corresponds to the SD of 3 experimental replicates. Therefore, the measured genome titre of rAAV2RefMat in this work using the commercial ITR assay was $19.9 \pm 2.3\%$ lower than the lower limit of the 95% confidence interval reported by Lock *et al.*, 2010. The rAAV2 material produced during this study and the rAAV9 samples were also characterised by qPCR and gave titres of $2.64 \times 10^5 \pm 5.81 \times 10^4$ copies/ μ L (CV of 22.0%) for rAAV2 and $5.14 \times 10^9 \pm 9.63 \times 10^8$ copies/ μ L (CV of 18.7%) for rAAV9 (**Figure 2.5**). The error ($\pm$) corresponds to the SD of 3 experimental replicates. All tested rAAV materials and the trFragment exhibited inter-experimental variability of less than 25% by qPCR, meeting the acceptance criteria for the assay to be considered fit-for-purpose (Hays *et al.*, 2022).

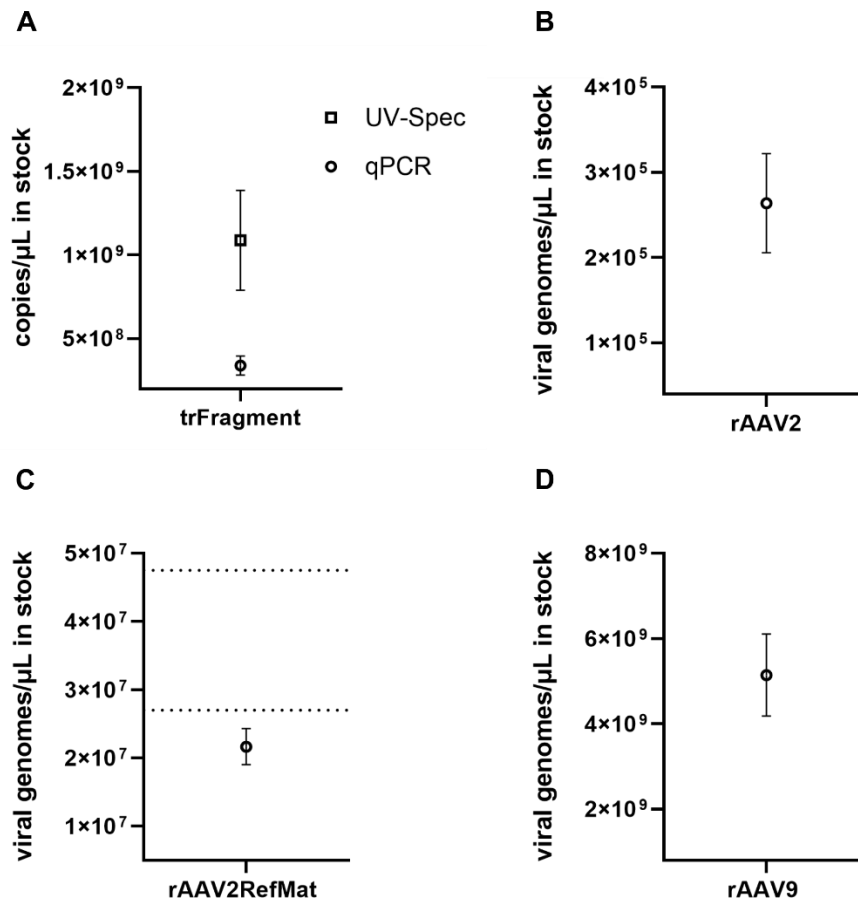


Figure 2.5 Measured genome titres of trFragment and rAAV Samples using qPCR. Plots demonstrate the concentrations of trFragment and rAAV samples determined in this study, with error bars corresponding to the SD of 3 experimental replicates. (A) The measured concentration of trFragment in copies/μL, with qPCR values represented as circles and UV-spectroscopy measurements as squares. (B) The viral genome concentration of rAAV2 samples produced during this work measured by qPCR. (C) The viral genome concentration of rAAV2RefMat measured by qPCR, with dotted lines indicating the 95% confidence interval reported by Lock *et al.*, 2010 using qPCR with SV40 polyA assay. (D) The viral genome concentration of rAAV9 sample, quantified using qPCR.

2.4.5 Calculation of the Percentage of Full rAAV Particles

The genome and capsid titres for rAAV2 and rAAV9 samples measured using qPCR and ELISA were used to calculate the percentage of full capsids in each batch (**Table 2.3**). The analysis showed that 88.6 ± 19.9 % of the rAAV2 particles contained a full genome, while 32.1 ± 6.4 % of the rAAV9 particles were full. Furthermore, based on the mean genome and capsid titres reported by Lock *et al.*, 2010 for the rAAV2RefMat material ($3.82 \times 10^7 \pm 2.97 \times 10^7$ viral genomes/μL and $9.43 \times 10^8 \pm 3.19 \times 10^8$ capsids/μL), the calculated percentage of full capsids is 4.05 ± 3.4 % (Lock *et al.*, 2010). In comparison, the percentage of full capsids of rAAV2RefMat based on the qPCR measurement is 2.3 ± 0.8 %.

Table 2.3 Calculated percentage of full rAAV2 and rAAV9 particles based on ELISA and qPCR analysis. The error ($\pm$) in ELISA corresponds to the SD of 3 technical replicates. The error ($\pm$) in qPCR corresponds to the SD of 3 experimental replicates. The error in full particles percentage ($\pm$) was calculated based on error propagation of both ELISA and qPCR.

Sample	ELISA (capsids/ μ L)	qPCR (viral genomes/ μ L)	% Full Particles
rAAV2	$2.98 \times 10^5 \pm 1.33 \times 10^4$	$2.64 \times 10^5 \pm 5.81 \times 10^4$	88.6 ± 19.9
rAAV9	$1.60 \times 10^{10} \pm 1.04 \times 10^9$	$5.14 \times 10^9 \pm 9.63 \times 10^8$	32.1 ± 6.4
rAAV2RefMat	$9.43 \times 10^8 \pm 3.19 \times 10^8$	$2.16 \times 10^7 \pm 2.64 \times 10^6$	2.3 ± 0.8

2.4.6 Total Protein Content Estimation of rAAV8

The total protein content of the rAAV8 sample was determined using the BCA assay, based on the generated quadratic standard curve ($R^2 = 0.999$) (**Figure 2.6**). The rAAV8 sample was diluted 3-fold prior to the assay, and the corresponding dilution factor was applied to calculate the original total protein concentration. The mean of two technical replicates was used to estimate the final value. The mean total protein concentration of the rAAV8 sample was determined as 52.2 μ g/mL.

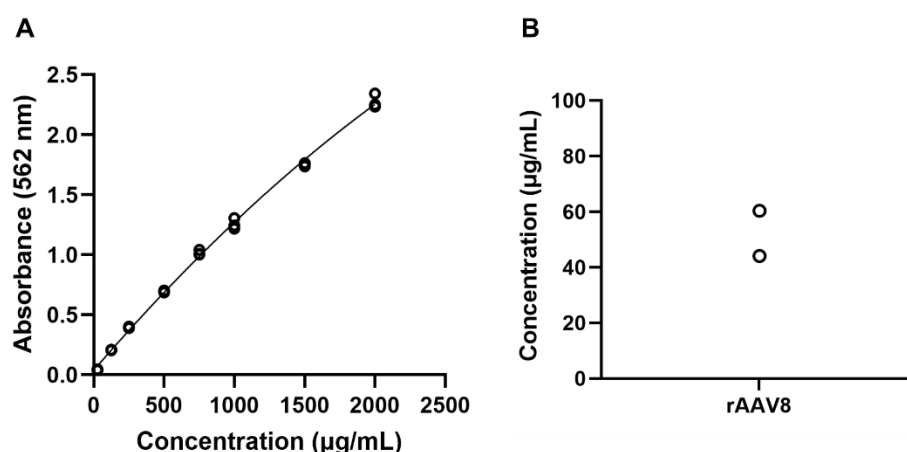


Figure 2.6 Total protein concentration of rAAV8 sample determined by BCA assay. (A) Quadratic standard curve of BSA, showing absorbance at 562 nm (y-axis) and μ g/mL protein concentration (x-axis). All three technical replicates for each point of the standard curve are plotted. (B) Measured total protein concentration in the rAAV8 sample (μ g/mL). Both technical replicates are plotted.

2.4.7 Analysis of Capsid Proteins and Protein Impurities in rAAV8 Samples

For visualisation of the capsid proteins of rAAV8 and evaluation of the purity of the sample, the rAAV8 sample and a no-capsid control sample were analysed by SDS-PAGE (**Figure 2.7**).

The capsid proteins VP1, VP2, and VP3 of rAAV8 have expected molecular weights of 81,667 Da, 66,519 Da, and 59,805 Da, respectively (Jin *et al.*, 2017; Lam *et al.*, 2022). On the SDS-PAGE gel, bands were present at the expected sizes, providing evidence of the presence of VP1, VP2, and VP3 at their expected molecular weights. The VP3 band appeared more intense than those of VP1 and VP2, which may reflect the commonly accepted VP1:VP2:VP3 ratio of 1:1:10 in AAV capsids (Rose *et al.*, 1971; Bennett *et al.*, 2017). However, as SDS-PAGE is not a quantitative method, no conclusions regarding the stoichiometry of the rAAV8 sample can be drawn. Additionally, numerous other bands were present in the gel, indicating protein impurities, most likely host cell proteins (HCPs) from the HEK293T cells used to produce the rAAV8 sample. While some high-molecular-weight bands (> 50 kDa) were observed, most impurities were < 50 kDa in size, with several intense bands suggesting a high abundance of these proteins in the sample (**Figure 2.7**).

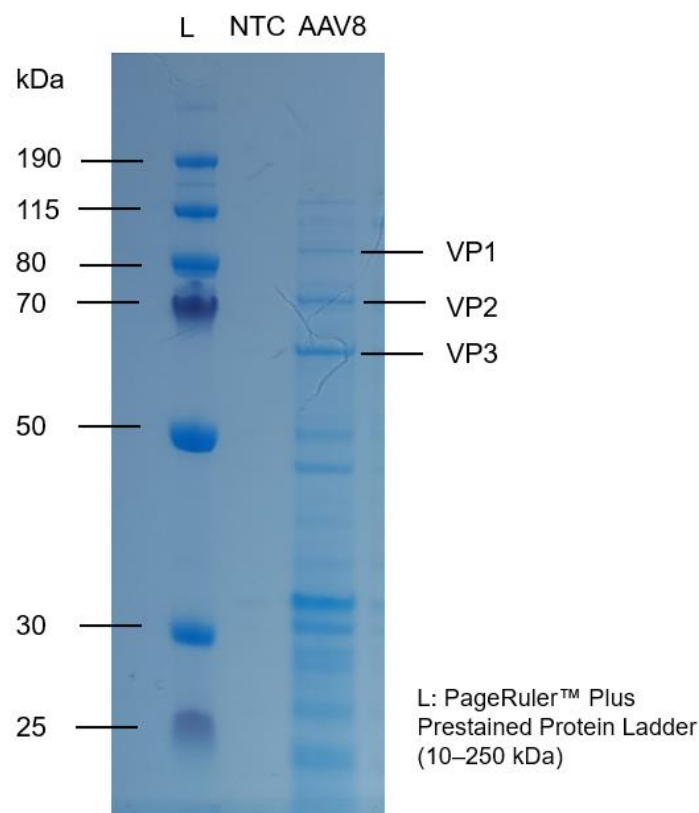


Figure 2.7 SDS-PAGE gel analysis of rAAV8 capsids. Bands are observed at molecular weights corresponding to the expected sizes of rAAV8 capsid proteins VP1 (81,667 Da), VP2 (66,519 Da), and VP3 (59,805 Da) (as indicated on the figure). Additional bands are present, indicating protein impurities within the sample.

2.5 Discussion

This chapter aimed to characterise recombinant AAV material, including rAAV2, rAAV8, and rAAV9, using established methods for rAAV analysis, specifically ELISA, qPCR, and SDS-PAGE. The importance of establishing the baseline characterisation of these samples was so that these values could then be compared with those obtained from advanced analytical techniques developed later in this thesis.

The generation of a positive control that acts as a surrogate template for the characterisation of the recombinant rAAV genome is important. This is because the presence of truncated genomes in rAAV preparations complicates the validation of PCR-based methods (Tran *et al.*, 2022; Kontogiannis *et al.*, 2024; Duong *et al.*, 2025). For example, differences in genome titres between assays targeting regions such as the ITR and GOI could result from either genome integrity variability or assay design factors, such as primer and probe binding efficiency and extension. However, without a controlled template, it is difficult to determine the exact cause of these discrepancies. To address this, a surrogate template (trFragment) was developed that contains the rAAV2 genome sequence, is readily available in large quantities, and includes the target genome regions in defined 1:1 ratios. This approach removes genome integrity as a source of variability, enabling direct assay comparisons and more reliable PCR method validation in Chapter 3. The trFragment is based on the pAAV.CMV.PI.EGFP.WPRE.bGH plasmid, which is the transfer plasmid used for the production of rAAV2 particles and contains the sequence that will be encapsidated in the viral particles (**Figure 2.1A**).

For the measurement of the copy number concentration of trFragment and the genome titres of rAAV2, rAAV9, and rAAV2RefMat, qPCR was used. Samples and standard curves were analysed in three independent experiments. Differences in melting temperatures were observed between the analysed samples and the standard used for standard curve generation, with the analysed samples showing a melting peak at 83.5°C, while the standard curve samples exhibited a peak at 84 °C (**Figure 2.4**). This difference could be attributed to the nature of the standard, which is plasmid DNA lacking free ITR ends because of the presence of the plasmid backbone. The absence of free ITR ends makes the ITRs more resistant to denaturation during qPCR, resulting in a higher melting temperature and reduced qPCR efficiency ($88.7 \pm 0.8\%$) (D'Costa *et al.*, 2016). In contrast, the rAAV samples and the

trFragment plasmid contain free ITR ends, allowing for easier denaturation and therefore a lower melting peak temperature (D'Costa *et al.*, 2016).

The trFragment copy number concentration measured by qPCR was $68.8 \pm 16.2\%$ relative to the value obtained using UV-Vis spectroscopy (NanoDrop) (**Figure 2.5**). Overestimation by NanoDrop is well-documented and attributed to non-specific absorbance at 260 nm from interfering molecules such as RNA, single-stranded DNA, oligonucleotides, and free nucleotides (Sedlackova *et al.*, 2013; Brzobohatá *et al.*, 2017; Masago *et al.*, 2021).

The genome titre of rAAV2RefMat was $19.9 \pm 2.3\%$ lower than the lower limit of the 95% confidence range provided by the interlaboratory study (**Figure 2.5**). The assay used in this thesis targeted the ITRs, whereas an SV40 polyA assay was used for the interlaboratory analysis of the rAAV2RefMat. D'Costa *et al.*, 2016 reported that the lack of free ITR ends in the plasmid standard leads to reduced amplification, lower standard curve copy number concentration and therefore overestimation of the ITR copy number concentration in the rAAV samples. However, the results generated did not show an overestimation of genome titre compared to the interlaboratory study, which could be due to several factors, including potential differences in capsid lysis efficiency, sample handling, threshold setting, assay type, removal of inhibitors, and instrumentation. Furthermore, since the sequence of the standard is proprietary, it is unknown whether the plasmid is linearised or modified in a way that mitigates the impact of lacking free ITR ends on amplification efficiency. Notably, seven of the 16 participating laboratories in analysing the reference material reported all replicates below the lower limit of the 95% confidence interval, while three reported all replicates above the upper limit. The reported CV from this study for measurement of the genome titre was 77.8%, with reported values ranging from 4.24×10^9 to 1.27×10^{11} viral genomes/mL (30-fold difference), highlighting the significant variations in genome titre measurements across laboratories, even when the same assay is used (Lock *et al.*, 2010). The observed variations may be attributed to the factors discussed above, which could also explain why the rAAV2RefMat titration results fell below the 95% confidence interval.

The ELISA and qPCR measurements of rAAV2, rAAV9, and rAAV2RefMat samples were compared to estimate the percentage of genome-containing particles (full). The number of full particles in the reference material was very low (2.3%), even lower than the previously reported 4.05% (**Table 2.3**). However, this method is known for its variability and poor

interlaboratory reproducibility due to differences in protocols, non-optimal PCR target concentrations, and variations in lysis methods (Lehman *et al.*, 2025). As such, methods that more robustly determine genome and capsid titres are required.

The rAAV8 material was intended to be used as a model for HCP identification in the work reported in Chapter 5. Here, rAAV8 was analysed using SDS-PAGE to provide an initial assessment of the sample's purity and to confirm the presence of the capsid proteins. Multiple protein impurities were observed, most of which were below 50 kDa in size (**Figure 2.7**). The presence of protein impurities, depending on their identity and quantity, can trigger immune responses in patients, making their identification critically important.

Overall, this chapter details the production of rAAV2 material, the characterisation of genome and capsid titres for the rAAV2 and rAAV9 samples using the traditional qPCR and ELISA approaches, and the estimation of the percentage of full particles for these samples. The genome and capsid titre values obtained here can then be compared with results from advanced analytical methods such as digital PCR and mass spectrometry, described later in this thesis, to evaluate their correlation. Additionally, the rAAV8 sample was assessed for the presence of protein impurities, which is described in Chapter 5. Therefore, the data presented here provide a baseline analysis of the samples to be used for the development of new approaches to rAAV analysis that are described and discussed in the following chapters.

Chapter 3: Characterisation of rAAV Genome Titre and Integrity

3.1 Introduction

Traditionally, qPCR has been the standard method for quantifying DNA targets and remains the method of choice for measuring genome titre in rAAV preparations (Lock *et al.*, 2010; Moullier and Snyder, 2012; Martinez-Fernandez de la Camara, McClements and MacLaren, 2021). However, the accuracy of qPCR depends on the generation of a standard curve, which introduces several limitations. These include variability arising from the initial quantification of the standard, the commutability of the standard, and variability resulting from the stability of the standard over time (Zhao *et al.*, 2016; Kuypers and Jerome, 2017).

dPCR is an emerging technique used for the same purpose and works by partitioning the PCR reaction into thousands of individual reactions, allowing for the direct counting of positive partitions (The dMIQE Group *et al.*, 2020; Hou *et al.*, 2023). Quantification is then performed using Poisson statistics to correct for the possibility that a single partition may contain more than one target molecule (Jones *et al.*, 2016). dPCR offers several advantages over qPCR, including the ability to provide absolute quantification without the need for a standard curve, increased precision, and higher tolerance to PCR inhibitors (Whale *et al.*, 2012; Zhao *et al.*, 2016; Wang *et al.*, 2022; Hou *et al.*, 2023).

There are two main dPCR technologies: droplet/emulsion-based methods and nanoplate-based methods. In droplet/emulsion-based dPCR, as employed by platforms such as the QX200 digital PCR system from Bio-Rad, partitions are created as droplets through emulsification. On the other hand, nanoplate-based dPCR, used by platforms such as the QIAcuity (QIAGEN) and Absolute Q (Thermo Fisher Scientific), involves partitioning the PCR reaction into microwells (Tiwari *et al.*, 2022; Tan *et al.*, 2023).

For the quantification of the rAAV genome titre with qPCR or dPCR, a single-target assay is typically used (referred to as uniplex), most commonly targeting the gene of interest (GOI), the ITRs or other elements within the genome, such as the SV40 polyA (Lock *et al.*, 2010, 2014). However, this approach may overlook critical information as the measured genome titre can vary depending on the selected target due to differences in genome integrity profiles, which result from the packaging of partial genomes (Galli *et al.*, 2021; Tran *et al.*,

2022; Kontogiannis *et al.*, 2024). Consequently, some genomic regions, such as the ITRs, may be more abundant than other targets in the capsids (Hashiba *et al.*, 2024). This variability is influenced by the preferential packaging of certain genome regions or fragments and the genetic content they contain. The work presented in this chapter demonstrates how targeting different regions can yield varying genome titre measurements, highlighting the importance of measuring multiple target regions of the recombinant rAAV genome using dPCR.

While this can be achieved by preparing a series of uniplex reactions to measure each target independently, a more effective approach is multiplex dPCR assays. Multiplex assays allow the simultaneous quantification of multiple targets in one reaction, reduce the variability caused by reaction mix preparation and enable genome integrity assessment, providing a more comprehensive view of the rAAV genome (Furuta-Hanawa, Yamaguchi and Uchida, 2019; Tereshko *et al.*, 2023; Duong *et al.*, 2025).

These multiplex assays can detect different targets by using fluorescent probes with distinct dyes for each target, with signals detected in separate detection channels. Two-target dPCR assays have been developed for multiplexing by targeting two different regions of the genome, most commonly one at the 5' end and one at the 3' end, using fluorescent probes emitting at different wavelengths (Furuta-Hanawa, Yamaguchi and Uchida, 2019; Zanker *et al.*, 2022). This approach can be further expanded to 3-plex, 4-plex, and 5-plex assays to improve the accuracy of the genome integrity assessment.

To estimate genome integrity, a simple fraction approach can be used, in which integrity is estimated as the proportion of partitions positive for all multiplex targets relative to the total number of positive partitions (Tereshko *et al.*, 2023). However, at high target concentrations, the likelihood of multiple genome fragments co-occupying the same partition increases. This can result in partitions appearing positive for multiple targets and therefore, leading to overestimation of genome integrity (Gleerup *et al.*, 2023; Tereshko *et al.*, 2023). Moreover, molecular dropout is another source of bias. In some cases, a partition may fail to produce a detectable signal, although it contains the target molecule. This happens because factors, such as template complexity, prevent amplification (Whale *et al.*, 2013). This leads to an underestimation of the measured linkage, as affected partitions appear negative despite containing the target.

For these reasons, statistical models are being developed to account for the aforementioned biases. These models use probability-based approaches to correct for molecular dropout and co-partitioning effects, thereby improving the accuracy of genome integrity measurements (Gleerup *et al.*, 2023). Several linkage models have been developed to address this issue, including those proposed by Duong *et al.*, 2025, Gleerup *et al.*, 2023, and Regan *et al.*, 2015. In this chapter, genome integrity was estimated using the statistical probability-based approach developed by Gleerup *et al.*, 2023. These results were then compared to those obtained using the simple fraction method to assess differences between the two approaches.

3.2 Aims of this Chapter

This chapter describes the development of dPCR assays for the measurement of genome titre and integrity in rAAV samples. Initially, a uniplex assay targeting the ITRs was employed to compare copy number measurements of trFragment across three dPCR platforms/instruments: QX200, QIAcuity, and Absolute Q. Furthermore, genome titre values of rAAV samples obtained using the ITR assay on the QIAcuity platform were compared with the qPCR results presented in Chapter 2. Additionally, to obtain a more reliable estimate of genome titre, dPCR assays were further developed to target multiple regions across the recombinant genome of different rAAV samples. Furthermore, 4-plex assays were developed and validated to assess the genome integrity of recombinant rAAV genomes. The measured genome integrity results for rAAV9 were subsequently compared, in Chapter 4, with those obtained using nanopore sequencing as an orthogonal method.

3.3 Materials & Methods

3.3.1 Assay Design

The dMIQE guidelines were followed for the selection and design of TaqMan assays (The dMIQE Group *et al.*, 2020). Amplicon sizes of less than 150 bp were selected for all assays to ensure efficient and comparable amplification between targets. The primers and hydrolysis probes for ITR, CMV Enhancer, *eGFP*, WPRE, and SV40 polyA were selected from previously published studies (**Table 3.1**). The primers and hydrolysis probes for *RTN3* and bGH polyA were

designed using Primer3Plus with the default settings and product size range set to 70 – 120 bp. The specificity of each assay was evaluated using NCBI's BLAST online tool (core nucleotide database: core_nt, organism: Homo sapiens, taxid: 9606; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Sayers *et al.*, 2022). The primer and probe sequences were also assessed for the absence of hybridisation sites within the production plasmids and for any additional hybridisation sites within the recombinant genome using the CLUSTALW sequence alignment tool (<https://www.genome.jp/tools-bin/clustalw>). Specific binding dyes were used to tag the probes of each assay, allowing their combination in 4-plex and 5-plex digital PCR assays (**Table 3.1**). The quenchers for each fluorescent dye in the probes were selected based on their optimal quenching range. FAM was used for the ITR assay with a BHQ1 quencher, HEX for *eGFP* assay with a BHQ1 quencher, ROX for CMV and *RTN3* with a BHQ2 quencher for both assays, Cy5 for WPRE and SV40 polyA with a BHQ2 quencher for both assays, and ATTO550 for bGH polyA with a BHQ2 quencher. All assays used in this chapter are summarised in **Table 3.1** and were purchased from Eurofins (United Kingdom).

Table 3.1 Primer and probe sequences used for dPCR analysis. F stands for forward primer, while R stands for reverse primer. ITR assay was taken from Aurnhammer *et al.*, 2012. CMV assay was taken from Dobnik *et al.*, 2019. *RTN3* assay was designed using Primer3Plus. *eGFP* assay was taken from Besnier *et al.*, 2002. WPRE assay was taken from Delenda and Gaillard, 2005. SV40 polyA assay was taken from Lock *et al.*, 2010. bGH polyA assay was designed using Primer3Plus.

Assay Name	Oligo Name	Oligonucleotide Sequence and Modification (5' to 3')	Amplicon Length (bp)
ITR	ITR_F ITR_R ITR_Probe	GGAACCCCTAGTGATGGAGTT CGGCCTCAGTGAGCGA FAM-CACTCCCTCTCTGCGCTCG-BHQ1	62
CMV	CMV_F CMV_R CMV_Probe	GTCAATGGGTGGAGTATTTACGG GGCATTACCGTCATTG ROX-CAAGTGTATCATATGCCAAGTACGCCCC-BHQ2	132
<i>RTN3</i>	<i>RTN3</i> _F <i>RTN3</i> _R <i>RTN3</i> _Probe	GTCTTTGGCACCACGCTGAT GTGACAGAGAGAAGAGCCAG ROX-GCTTTCCCTGGCAGCTTTTCAGTG-BHQ2	95
<i>eGFP</i>	<i>eGFP</i> _F <i>eGFP</i> _R <i>eGFP</i> _Probe	CAACAGCCACAACGTCTATATCAT ATGTTGTGGCGGATCTTGAAG HEX-CCGACAAGCAGAAGAACGGCATCAA-BHQ1	78
WPRE	WPRE_F WPRE_R WPRE_Probe	CCGTTGTCAGGCAACGTG AGCTGACAGGTGGTGGCAAT Cy5-TGCTGACGCAACCCCCACTGGT-BHQ2	85
SV40 polyA	SV40_F SV40_R SV40_Probe	AGCAATAGCATCACAAATTTACAA CCAGACATGATAAGATACATTGATGAGTT Cy5-AGCATTTTTTTTCACTGCATTCTAGTTGTGGTTTGTC-BHQ2	95
bGH polyA	bGH_F bGH_R bGH_Probe	GTTGCCAGCCATCTGTTGTTT GACAGTGGGAGTGGCACCTT ATTO550-CTCCCCCGTGCCTTCCTTGACCC-BHQ2	89

3.3.2 Uniplex dPCR Assays

Four rAAV samples, rAAV2, rAAV2RefMat, rAAV8, and rAAV9 described in Chapter 2, Section 2.3.7, along with the trFragment, were analysed with uniplex assays by dPCR in this chapter. Different regions are present in the recombinant genome of each rAAV sample. Therefore, different assays were used based on the recombinant genome of each rAAV sample. Furthermore, all rAAV samples analysed in this study contain the ITRs from AAV serotype 2. Therefore, the same ITR assay was used to quantify the genome titre of all rAAV samples. The

genomic map of each sample is shown in **Figure 2.1** of Chapter 2. **Table 3.2** categorises the uniplex assays used for analysing each rAAV sample and the trFragment.

Table 3.2 Uniplex assays used for measurement of each rAAV sample.

Sample	Uniplex Assays
trFragment	ITR-FAM, CMV-ROX, <i>eGFP</i> -HEX, WPRE-Cy5, bGH polyA-ATTO550
rAAV2	ITR-FAM, CMV-ROX, <i>eGFP</i> -HEX, WPRE-Cy5, bGH polyA-ATTO550
rAAV2RefMat	ITR-FAM, CMV-ROX, SV40 polyA-Cy5
rAAV8	ITR-FAM, CMV-ROX, <i>eGFP</i> -HEX, SV40 polyA-Cy5
rAAV9	ITR-FAM, <i>RTN3</i> -ROX, <i>eGFP</i> -HEX, WPRE-Cy5, bGH polyA-ATTO550

3.3.3 Multiplex dPCR Assays

Multiplex dPCR assays were designed, including 4-plex, and 5-plex assays. Two 4-plex assays were developed: 4-plex A and 4-plex B. 4-plex A included assays with probes labelled with FAM, HEX, ATTO550, and Cy5, while 4-plex B contained probes with FAM, HEX, ROX, and Cy5. The assay combinations used for each multiplex assay, along with the samples analysed, are summarised in **Table 3.3**.

- The ITR, *eGFP*, WPRE, and bGH polyA assays composed the 4-plex A assay.
- The ITR, CMV, *eGFP*, WPRE, and bGH polyA assays composed the 5-plex assay.
- ITR, CMV, *eGFP*, and WPRE assays composed the 4-plex B assay for rAAV2 genomes.
- ITR, *RTN3*, *eGFP*, and WPRE assays composed the 4-plex B assay for rAAV9 genomes
- ITR, CMV, *eGFP*, and SV40 polyA composed the 4-plex B assay for rAAV8 genomes.

Table 3.3 Overview of assay components used in 4-plex (A and B) and 5-plex digital PCR assays, including fluorophore combinations and samples analysed with each assay configuration.

Assay	Sample	ITR (FAM)	CMV (ROX)	<i>eGFP</i> (HEX)	WPRE (Cy5)	bGH polyA (ATTO550)	<i>RTN3</i> (ROX)	SV40 polyA (Cy5)
4-plex A	trFragment	X		X	X	X		
	rAAV9	X		X	X	X		
4-plex B	rAAV2	X	X	X	X			
	rAAV8	X	X	X				X
	rAAV9	X	X	X			X	
5-plex	trFragment	X	X	X	X	X		

3.3.4 Pre-treatment with DNase I and Preparation of Dilutions

DNase I treatment was applied to the rAAV particle samples to remove unpackaged DNA impurities. Twenty microlitres of each rAAV particle sample (rAAV2, rAAV8, rAAV9) was diluted with nuclease-free water to a final volume of 170 μ L in an Eppendorf DNA LoBind tube. To this, 20 μ L of DNase I reaction buffer (NEB, Cat No. M0303S) and 10 μ L of DNase I (NEB, Cat No. M0303S) were added for a final reaction volume of 200 μ L. The reaction was thoroughly mixed by pipetting and incubated at 37 °C for 10 min using a Bio-Rad C1000 thermal cycler. Finally, the samples were incubated at 75 °C for 10 min in a thermal cycler (Bio-Rad C1000) to deactivate the DNase I.

The extracted rAAV DNA samples (rAAV2, rAAV2RefMat, and rAAV9) that were analysed in Chapter 2 were also used for digital PCR analysis in this chapter, without further dilutions. To reduce the inhibitory effects of residual protein from the DNase I-digestion, all rAAV particle samples were diluted 10-fold in nuclease-free water. The DNase I-treated samples were then further diluted to reach a concentration optimal for dPCR quantification. The rAAV2, rAAV8, and rAAV9 samples were diluted to an expected ITR concentration of 500 copies/ μ L. As discussed later in the chapter, this concentration was selected based on the observation of low inter-assay variability (below 10%) and the fact that, at this level, only 3.5% of partitions are expected to contain more than one molecule. This minimises co-partitioning and allows for more accurate estimation of genome titre and genome integrity. After each

dilution step, the samples were mixed by vortexing for 3 sec and spinning them down using a benchtop microcentrifuge. The DNase I-treated rAAV2 sample was first diluted 1:10 in TE buffer (Thermo Fisher Scientific, Cat No. J75893.XCR), followed by an additional 1:5 dilution in TE buffer to reach a final expected concentration of 500 copies/ μ L. The DNase I-treated rAAV8 sample was also diluted 1:5 in TE buffer, followed by five 1:10 serial dilution steps in TE buffer, resulting in a final expected concentration of 500 copies/ μ L. The DNase I-treated rAAV9 sample underwent five additional 1:10 serial dilution steps. The final dilution was performed by adding 70 μ L of the previous rAAV9 dilution to 930 μ L of TE buffer, resulting in a final dilution factor of 1.4×10^7 and an expected concentration of 500 copies/ μ L. The trFragment stock was serially diluted in a four-step 1:10 dilution series (dilution A). For dPCR analysis, this dilution was further diluted 1:10 in TE buffer (dilution B), followed by an additional 1:5 dilution (dilution C) in TE buffer.

3.3.5 Linear Range Assessment

For the linear range assessment of the assays targeting the ITR, CMV, *eGFP*, WPRE, and bGH polyA regions, the trFragment dilution B was subjected to four additional 1:2 serial dilutions in TE buffer, resulting in final dilution factors of 10 (dilution B), 20, 40, 80, and 160 of dilution A. Moreover, non-template controls were included, with TE buffer used to confirm no contamination during the dilution process and nuclease-free water used to ensure no contamination occurred during the dPCR reaction preparation.

3.3.6 MspI Digest

The Anza™ 65 MspI restriction enzyme (Thermo Fisher Scientific, Cat No. IVGN0656) was used for its ability to cut rAAV genomes. This allowed the 5' and 3' ITRs of rAAV genomes to be separated into distinct partitions, enabling accurate measurement of the total ITR (Carver, Blech-Hermoni and James, 2021). The enzyme was diluted from 20 units/ μ L to 5 units/ μ L by mixing 15 μ L of MspI enzyme with 36 μ L of TE buffer and 9 μ L of Anza™ buffer (Thermo Fisher Scientific, Cat No. IVGN0656). The digestion reaction was set up by adding 17 μ L of trFragment (dilution C), 2 μ L of Anza™ buffer, and 1 μ L of the diluted MspI enzyme, followed by incubation at 37 °C for 15 min in a heat block. The same process was followed for the digestion of the extracted rAAV2 DNA. After digestion, the digested template was added to the QIAcuity dPCR reaction mastermix. The digestion step was performed before mixing the template with the

dPCR master mix because *MspI* and the QIAcuity Probe Mix are incompatible. This would prevent efficient digestion, as previously reported (Carver, Blech-Hermoni and James, 2021).

3.3.7 Preparation of dPCR

3.3.7.1 QX200

For quantification of the trFragment sample using the QX200 Droplet Digital PCR System (Bio-Rad Laboratories), a TaqMan assay targeting the ITRs (ITR2) was used (**Table 3.1**). The dPCR reaction mix for one reaction was prepared using 4.9 μL of nuclease-free water, 11 μL of 2 $\times$ QX200 ddPCR Probe Supermix at a final concentration of 1 $\times$, and 1.1 μL 20 $\times$ assay mix. A 20 $\times$ assay contained primers at 18 μM and probes at 5 μM . The final primer concentration in the reaction was 250 nM, while the probe concentration was 900 nM.

A total of 17 μL of the dPCR mastermix per reaction was transferred to a 96-well plate, followed by the addition of 5 μL of either the trFragment sample or a non-template control. The plate was sealed with aluminium foil, vortexed, and centrifuged at 1,000 $\times$ g for 1 min. Then, 20 μL of the reaction mix was transferred into a DG8 Cartridge (Bio-Rad, Cat No. 1864008) using a Rainin 8-channel pipette. A volume of 70 μL of ddPCR Droplet Generator Oil for Probes (Bio-Rad, Cat No. 1863005) was loaded into each well, and a DG8 Gasket (Bio-Rad, Cat No. 1863009) was placed over the filled cartridge by securing it onto the cartridge holder. The cartridge was then placed in the QX200 droplet generator.

Following droplet generation, 40 μL of the droplets were transferred from the DG8 Cartridge into a semi-skirted 96-well PCR plate using a Rainin 8-channel pipette. A Pierceable Foil Heat Seal (Bio-Rad, Cat No. 1814040) was applied to the droplet plate using the PX1 PCR plate sealer. The sealed plate was then placed in the C1000 thermal cycler (Bio-Rad) for PCR amplification using the following thermal cycling conditions: an initial denaturation at 95 $^{\circ}\text{C}$ for 10 min, followed by 40 cycles of 30 sec at 94 $^{\circ}\text{C}$ and 1 min at 60 $^{\circ}\text{C}$, and a final step of 10 min at 98 $^{\circ}\text{C}$. After PCR, the droplets were read using the Bio-Rad QX200 droplet reader.

3.3.7.2 Absolute Q

For quantification of the trFragment sample using the Applied Biosystems QuantStudio Absolute Q Digital PCR System (Thermo Fisher Scientific), TaqMan assays targeting specific regions of trFragment were used, as shown in **Table 3.1** and **Table 3.2**. The dPCR reaction mix for one reaction consists of 1.93 μL of nuclease-free water, 1.98 μL of 5 $\times$ dPCR Master Mix (Thermo Fisher Scientific, Cat No. A52490) at a final concentration of 1 $\times$, and 0.5 μL of 20 $\times$

assay mix (final concentration 1×). The final primer concentration was 250 nM, while the probe concentration was 900 nM. 4.4 µL of the dPCR mix were allocated to each well of an Applied Biosystems™ MicroAmp™ EnduraPlate™ optical 96-well clear reaction plate (Thermo Fisher Scientific, Cat No. A36924), and 5.5 µL of the trFragment or non-template control was added to the corresponding wells. The plate was sealed with aluminium foil, vortexed, and centrifuged at 1,000 × g for 1 min.

Nine microlitres of the dPCR mix was loaded into each corresponding well of the MAP16 plates (Thermo Fisher Scientific, Cat No. A52865), and 15 µL of Isolation Buffer (Thermo Fisher Scientific, Cat No. A52865) was added to each well containing the PCR mix. The plate was then placed into the Absolute Q instrument, where partitioning, thermal cycling, and imaging were performed. The thermal cycling protocol included an initial step of 10 min at 96 °C, followed by 40 cycles of 5 sec at 96 °C and 15 sec at 60 °C.

For the 4-plex assay, the same workflow was followed, incorporating four sets of assays. The volume of nuclease-free water in the dPCR reaction mix was reduced accordingly to accommodate the additional assay components. Each 4-plex reaction mix for one reaction consists of 0.44 µL of nuclease-free water, 1.98 µL of 5× dPCR Master Mix, and 0.5 µL of each 20× assay mix.

3.3.7.3 QIAcuity

For the quantification of the genome titre of rAAV samples and the copy number concentration of trFragment, the QIAcuity Four Digital PCR System (QIAGEN) was used with TaqMan assays targeting specific regions within the rAAV genome. The details of the assays, including their sequences, amplicon lengths, and the rAAV samples they can be used for, are provided in **Table 3.1** and **Table 3.2**.

For each uniplex assay, the reaction mix for one reaction was prepared using 3.74 µL of nuclease-free water, 3.3 µL of 4× QIAcuity Probe Master Mix (QIAGEN, Cat No. 250102) (1× final concentration), and 0.66 µL of 20× assay mix (at 1× final concentration). Each 20× assay contained primers at 18 µM and probes at 5 µM, resulting in a final primer concentration of 900 nM and a final probe concentration of 250 nM in the reaction.

A total of 7.7 µL of the dPCR reaction mix was added to each well of a 96-well clear reaction plate, followed by 5.5 µL of the template DNA sample. NTCs, including nuclease-free

water and TE buffer used for sample dilution, were included in the experiment. The plate was sealed with aluminium foil, vortexed, and centrifuged at $1,000 \times g$ for 1 min.

Twelve microlitres of each reaction mixture were loaded into a QIAcuity 8.5k 96-well Nanoplate (QIAGEN, Cat No. 250021) or a QIAcuity 8.5k 24-well Nanoplate (QIAGEN, Cat No. 250011), sealed, and placed into the QIAcuity Four instrument for partitioning, cycling, and imaging. The thermal cycling conditions included an initial step of 10 min at 95 °C, followed by 40 cycles of 15 sec at 95 °C and 30 sec at 60 °C.

For 4-plex and 5-plex dPCR, the same workflow described here was followed, with the inclusion of four and five sets of assays, respectively. The volume of nuclease-free water in the dPCR reaction mix was reduced accordingly to accommodate the additional assay components. Each 4-plex reaction mix for one reaction consists of 1.76 µL of nuclease-free water, 3.3 µL of 4× QIAcuity Probe Master Mix and 0.66 µL of each 20× assay mix. Each 5-plex reaction mix for one reaction consists of 1.1 µL of nuclease-free water, 3.3 µL of 4× QIAcuity Probe Master Mix, and 0.66 µL of each 20× assay mix.

3.3.8 Data Processing and Analysis

3.3.8.1 QX200

Data analysis was performed using QuantaSoft Software v1.7.4. NTC plots were used as a reference to guide correct threshold placement. The threshold was manually set for each analysed well above the negative cluster and consistently applied to all samples analysed with the specific assay. Additionally, samples were evaluated for baseline jumps or declines to ensure accurate threshold placement. Only wells with more than 10,000 total partitions were included in the analysis (Košir *et al.*, 2017; Emslie *et al.*, 2019).

3.3.8.2 Absolute Q

Data analysis was conducted using the QuantStudio Absolute Q Digital PCR Software 6. The threshold was placed following the same rules described above for the QX200. Only wells with more than 20,000 total partitions were included in the analysis.

3.3.8.3 QIAcuity

Data analysis was conducted using QIAcuity Software Suite 2.5.0.1 (QIAGEN). The threshold was placed following the same rules described above for the QX200. The same threshold was applied to all wells within the same 8.5K plate for each assay. Only wells with more than 6,800

total partitions were included in the analysis, representing 80% of the expected 8,500 partitions when using an 8.5K plate. This restriction was set to ensure reliable quantification and data accuracy, although no specific lower limit has been determined by the manufacturer.

3.3.8.4 Presence of “Rain”

For the measurement of the concentration of analysed samples by dPCR, a threshold is set to clearly distinguish between positive partitions, which contain the target, and negative partitions, which do not. In cases (such as when using the ITR assay), the separation between positive and negative clusters may not be completely distinct, and some partitions may fall between these clusters, forming what is known as “rain.” These partitions contain the target but appear at a lower fluorescence value due to reduced amplification efficiency. For that reason, observed rain was included in the positive partitions to ensure accurate quantification.

3.3.8.5 Detection Channels

The digital PCR platforms evaluated also differ in the number and type of detection channels available. **Table 3.4** summarises the detection channels for each platform, along with the specific assay detected in each channel. Each platform also includes a reference fluorescence channel used for signal normalisation and partition detection. This channel was monitored after each run to confirm reliable partition generation.

Table 3.4 Overview of detection channels available on the QX200, QIAcuity, and Absolute Q platforms, and the assay detected in each channel based on the fluorophore label of the probe.

dPCR Platforms	Channel Name	Assay/Fluorophore Detected
QX200	Channel 1	ITR-FAM
	Channel 2	-
QIAcuity	Green	ITR-FAM
	Yellow	<i>eGFP</i> -HEX
	Orange	bGH polyA-ATTO550
	Red	CMV-ROX, <i>RTN3</i> -ROX
	Crimson	WPRE-Cy5, SV40 polyA-Cy5
Absolute Q	Blue	ITR-FAM
	Green	<i>eGFP</i> -HEX
	Orange	bGH polyA-ATTO550
	Dark Red	WPRE-Cy5

3.3.9 Calculation of target DNA concentration

The concentration of each sample was determined by first calculating the λ value, which represents the mean concentration of target molecules per partition. This was estimated using the proportion of negative partitions and the Poisson distribution with the Equation 3.1, where w is the number of negative partitions and n is the total number of partitions.

$$\lambda = -\ln \left(\frac{w}{n} \right) \quad (\text{Equation 3.1})$$

The Equation 3.1 was rewritten as Equation 3.2, where k is the number of positive partitions and n is the total number of partitions.

$$\lambda = -\ln \left(1 - \frac{k}{n} \right) \quad (\text{Equation 3.2})$$

The λ value was then divided by the partition volume to obtain the copies per μL in the reaction. The partition volume varies between dPCR platforms: 0.00082 μL for QX200, 0.000432 μL for Absolute Q, and 0.00034 μL for QIAcuity. Finally, the copies per μL in the template unit were calculated by multiplying the copies per μL in the reaction with the final

reaction volume (20 µL for QX200, 9 µL for Absolute Q, and 12 µL for QIAcuity), and dividing by the template volume added (5 µL for all three platforms). The concentration in the stock was then determined by applying the respective dilution factors for each sample. All calculations were performed using Microsoft 365 Excel.

3.3.10 Statistical analysis

Paired t-test was performed using the Data Analysis ToolPack of Microsoft 365 Excel. The results were categorised as: not significant when $p \geq 0.05$, significant when $0.01 < p < 0.05$, highly significant when $0.001 < p < 0.01$, and extremely significant when $0.0001 < p < 0.001$.

3.3.11 Measurement of genome integrity

The percentage of full genomes in each sample, as assessed by the 4-plex B dPCR assay, was determined using two approaches: a simple fraction method and a statistical probability-based model. In the simple fraction method, the number of partitions positive for all four targets (i.e. above the threshold for each assay) was divided by the total number of positive partitions. This value represents the proportion of genomes containing all four targets, which is considered indicative of full-length genomes (Tereshko *et al.*, 2023).

To minimise the effect of co-partitioning, where multiple DNA fragments may be distributed into the same partition at higher concentrations, each rAAV sample was analysed at a concentration of approximately 500 copies/µL. At this concentration, the predicted λ is 0.07, and the estimated percentage of partitions containing more than one molecule is 3.5%. This estimation is based on the Poisson distribution. First, the predicted number of partitions containing exactly one molecule was calculated. Then, the expected number of partitions containing exactly two, three, and four molecules was also determined. These values were calculated using the Poisson probability mass function (Equation 3.3), where $P(X=k)$ is the probability of observing k number of molecules per partition (1, 2, 3, or 4 predicted molecules per partition) and e is the base of the natural logarithm.

$$P(X = k) = (e^{-\lambda} \times \lambda^k) / k! \quad (\text{Equation 3.3})$$

The resulting probabilities were multiplied by the total number of partitions (8,500) on the QIAcuity system. The percentages corresponding to two, three, and four molecules per partition were then summed to estimate the overall proportion of partitions containing more than one molecule.

For comparison, genome integrity was also estimated using the DNA Integrity Calculation software (https://digpcr.shinyapps.io/intactness_dsi_assay/), developed by the Ghent University Digital PCR Centre and based on the model described by Gleerup *et al.*, 2023. This probability-based approach accounts for the random co-partitioning of targets within the same partition as well as molecular dropout events, providing a more accurate estimate of genome integrity.

3.4 Results

3.4.1 Temperature Gradient

To determine the optimal ITR assay annealing temperature for dPCR, a temperature gradient experiment was conducted using the QX200 system for the quantification of trFragment. The temperature gradient was set using the instrument software, with cycling temperatures ranging from 54 °C to 64 °C (**Table 3.5**). Two NTCs (TE) were included to confirm the absence of contamination. The trFragment was run in one replicate at each tested temperature.

Table 3.5 Annealing temperatures tested for optimisation of the ITR assay.

Wells	Sample	Annealing Temperature (°C)
A	trFragment	64
B	TE (NTC)	63.4
C	trFragment	62.1
D	trFragment	60.2
E	trFragment	58
F	trFragment	56.1
G	TE (NTC)	54.8
H	trFragment	54

All tested temperatures (54 – 64 °C) exhibited a comparable separation between positive and negative droplets, with the optimal separation observed between 54 °C and 60.2 °C. Moreover, the non-template controls run at both the high and low temperatures showed no positive partitions, confirming the absence of contamination. Regarding the concentration measurement of trFragment, the mean concentration across the different temperatures in this experiment was $4.5 \times 10^8 \pm 2.1 \times 10^7$ copies/ μ L (**Figure 3.1**). All tested temperatures produced consistent results, with a CV of 4.8% across conditions. The low variability indicates that all tested temperatures can be used for accurate measurements of the ITRs. Therefore, for subsequent experiments, an annealing temperature of 60 °C was selected, as it is commonly used for optimal probe hydrolysis.

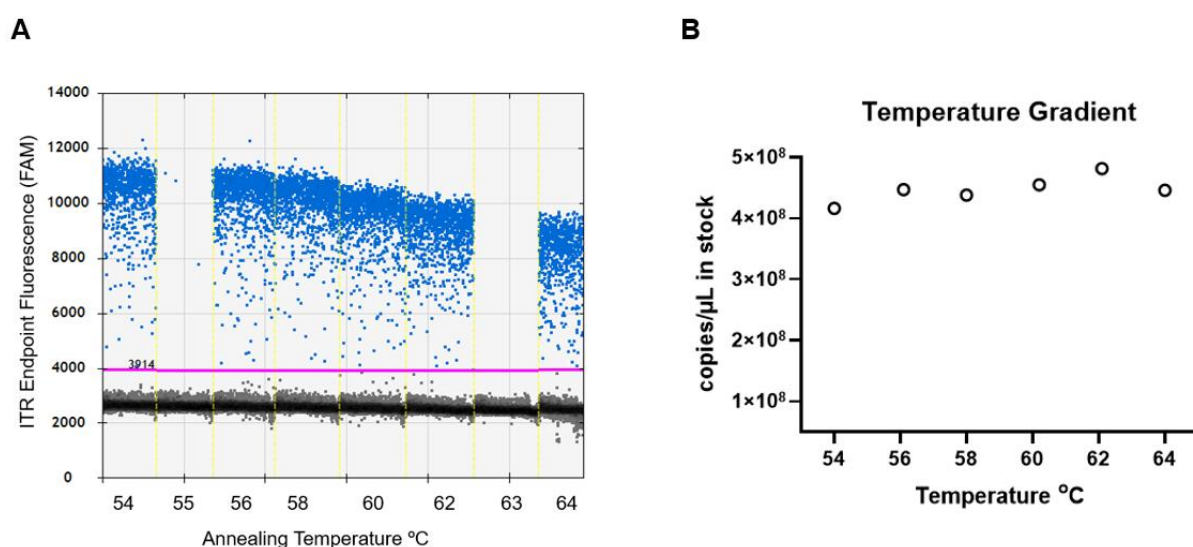


Figure 3.1 Annealing temperature gradient in QX200 for trFragment. (A) One-colour dPCR plot in QX200, showing the endpoint fluorescence of the ITR assay (FAM) of the analysed droplets on the y-axis and each temperature used at each well on the x-axis. The yellow vertical lines separate each well. The negative partitions are shown in dark grey, while the positive partitions are shown in blue. The pink line corresponds to the threshold set to distinguish between negative and positive droplets. (B) Measurement of the concentration of the trFragment sample stock across the different annealing temperatures tested. Each circle represents one replicate of the analysed sample at the corresponding temperature.

3.4.2 dPCR Platform Comparison

Next, three different dPCR platforms (QX200, Absolute Q, QIAcuity) were compared in measuring the ITR copy number concentration in the trFragment sample. The trFragment sample was analysed in three independent experiments using the QX200, QIAcuity, and Absolute Q instruments and the mean copy number concentration from each experiment was

compared across the platforms. Although the QIAcuity instrument reported the highest concentration among the dPCR platforms tested (4.7×10^8 copies/ μL vs 3.7×10^8 copies/ μL for QX200 and Absolute Q) (**Figure 3.2**), no significant differences were observed ($p > 0.05$) when a paired t-test ($n = 3$) was used to compare QX200 with QIAcuity, QX200 with Absolute Q, and QIAcuity with Absolute Q.

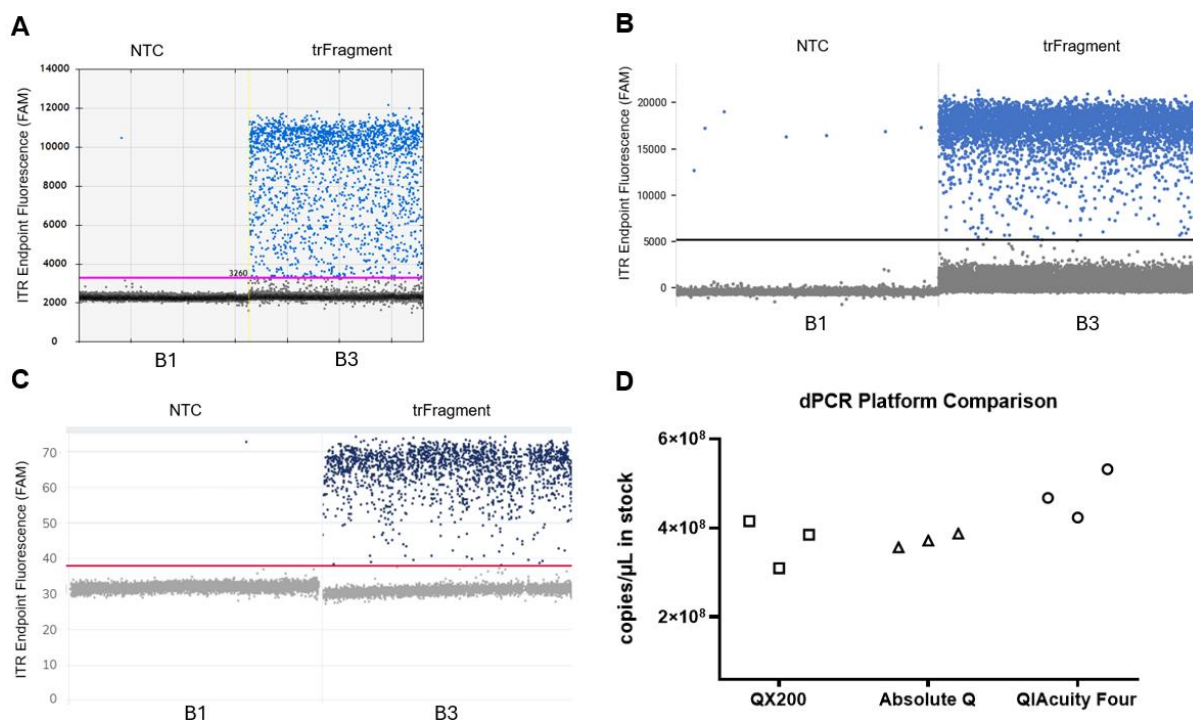


Figure 3.2 Analysis of trFragment using QX200, Absolute Q, and QIAcuity dPCR platforms. One-colour plots of a representative well containing the trFragment sample and a well containing an NTC in QX200 (A), Absolute Q (B), and QIAcuity (C). Negative partitions are shown in grey, while positive partitions are shown in blue (A-C). (D) Measured concentration of trFragment (copies/ μL) across three independent experiments on each dPCR platform. QX200 (squares), Absolute Q (triangles), QIAcuity (circles). Each set of three symbols corresponds to the three independent experiments performed on each instrument. Endpoint fluorescence values differ across digital PCR instruments due to variations in data processing, signal detection methods, and reporting formats specific to each platform.

It is noted that the variation in cluster density between plots A, B, and C in **Figure 3.2** is due to differences in the number of partitions between the instruments: plot A (QX200) contains 16,516 droplets, plot B (Absolute Q) contains 20,457 partitions, and plot C (QIAcuity) contains 8,039 partitions. Additionally, the size of the circles representing individual partitions is larger in the Absolute Q software output compared to those in the QX200 and QIAcuity software outputs.

The variability (CV) between the three instruments was 14.8%. The mean concentration with the standard deviation measured by each of the three dPCR platforms across the three experiments is shown in **Table 3.6**. When comparing inter-experimental variability (CV) across platforms, Absolute Q demonstrated the lowest inter-experimental variability, with a CV of 4.2%, while QX200 exhibited the highest, with a CV of 14.7%. The inter-experimental variability of QIAcuity was 11.5% (**Table 3. 6**).

Table 3.6 Measured trFragment copy number concentration and interexperimental variability (CV%) across QX200, Absolute Q, and QIAcuity.

dPCR platform	Copies/ μ L in Stock	CV (%)
QX200	$3.7 \times 10^8 \pm 5.4 \times 10^7$	14.7
Absolute Q	$3.7 \times 10^8 \pm 1.6 \times 10^7$	4.2
QIAcuity	$4.7 \times 10^8 \pm 5.4 \times 10^7$	11.5

Rain was observed in the plots across all platforms for the ITR assay (**Figure 3.2**), likely due to the secondary structure of the ITRs. In all platforms, the rain was included in the positive partitions, as these represent true positives with reduced amplification efficiency. Among the platforms tested, the QX200 exhibited a higher degree of rain, making threshold placement more challenging compared to the other platforms.

For the subsequent experiments described in this thesis, the QIAcuity instrument was used. This decision was based on its greater flexibility in multiplex dPCR, as it offers five detection channels compared to four in Absolute Q and only two in QX200. The additional detection channel allows for more dye and channel combinations, enabling a more versatile evaluation of a multiplex assay compared to Absolute Q.

3.4.3 Evaluation of Linear Range of dPCR Assays

The linear range of dPCR assays targeting different regions of trFragment was assessed to ensure reliable results across a wide range of input DNA concentrations. The uniplex assays tested targeted five distinct regions within the trFragment construct: ITR, CMV Enhancer, *eGFP*, WPRE, and bGH polyA (**Table 3.1**). A schematic of the trFragment construct, showing the location of each assay hybridisation site, is presented in **Figure 3.3A**. For the assessment of the linear range of each assay, a five-point 1:2 dilution series of trFragment was prepared. All samples, including the NTCs, were run in duplicate for each assay.

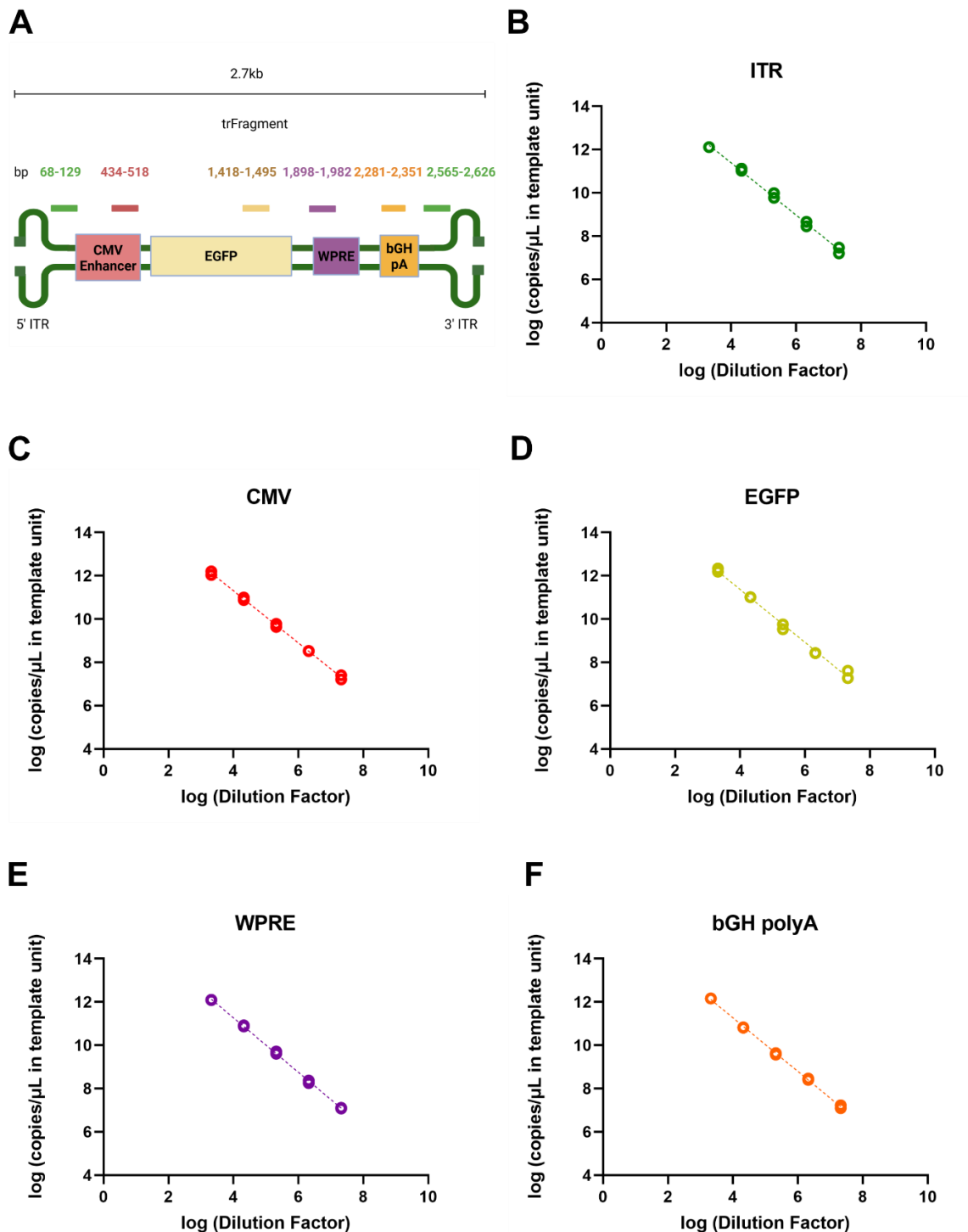


Figure 3.3 Linear range assessment of ITR, CMV, eGFP, WPRE, bGH polyA assays on trFragment with QIAcuity dPCR. (A) Map of the trFragment construct, showing assay hybridisation site. ITR assay is shown in green, CMV assay in red, eGFP assay in yellow, WPRE assay in purple, and bGH polyA assay in orange. Created with BioRender.com. (B–F) Plots showing the linearity of the ITR, CMV, eGFP, WPRE, and bGH polyA assays across five-point 1:2 serial dilutions of trFragment. (B) ITR assay, (C) CMV assay, (D) eGFP assay, (E) WPRE assay, (F) bGH polyA assay. Both replicates tested for each assay are displayed. In each plot, the x-axis represents the $\log_2$ of the dilution factor applied relative to dilution A, while the y-axis represents the $\log_2$ of the measured concentration of trFragment.

The one-colour plots of a set of serial dilutions of trFragment for all assays tested are shown in **Figure 3.4**, demonstrating the progressively decreasing number of positive partitions as the dilution factor increases. All tested assays demonstrated excellent linearity, with R^2 values greater than 0.99 across the tested range (**Figure 3.3**). Additionally, all assays, remained linear from concentrations of 4,568 copies/ μ L (mean across assays) ($\sim \log_2$ scale: 12.16) to concentrations of 155 copies/ μ L (mean across assays) ($\sim \log_2$ scale: 7.28) (163 – 4,436 copies/ μ L for ITR assay, 159 – 4,465 copies/ μ L for CMV assay, 175 – 4,917 copies/ μ L for *eGFP* assay, 137 – 4,370 copies/ μ L for WPRE assay, 143 – 4,580 copies/ μ L for bGH poly A assay). The inter-assay variability was estimated (expressed as CV %) at each point of the standard curve (**Table 3.7**). The inter-assay variability at the lowest concentration of the standard curve was twice that observed at the highest concentration (12.6% vs 6.0%). For every other concentration tested, the inter-assay variability was below 10%.

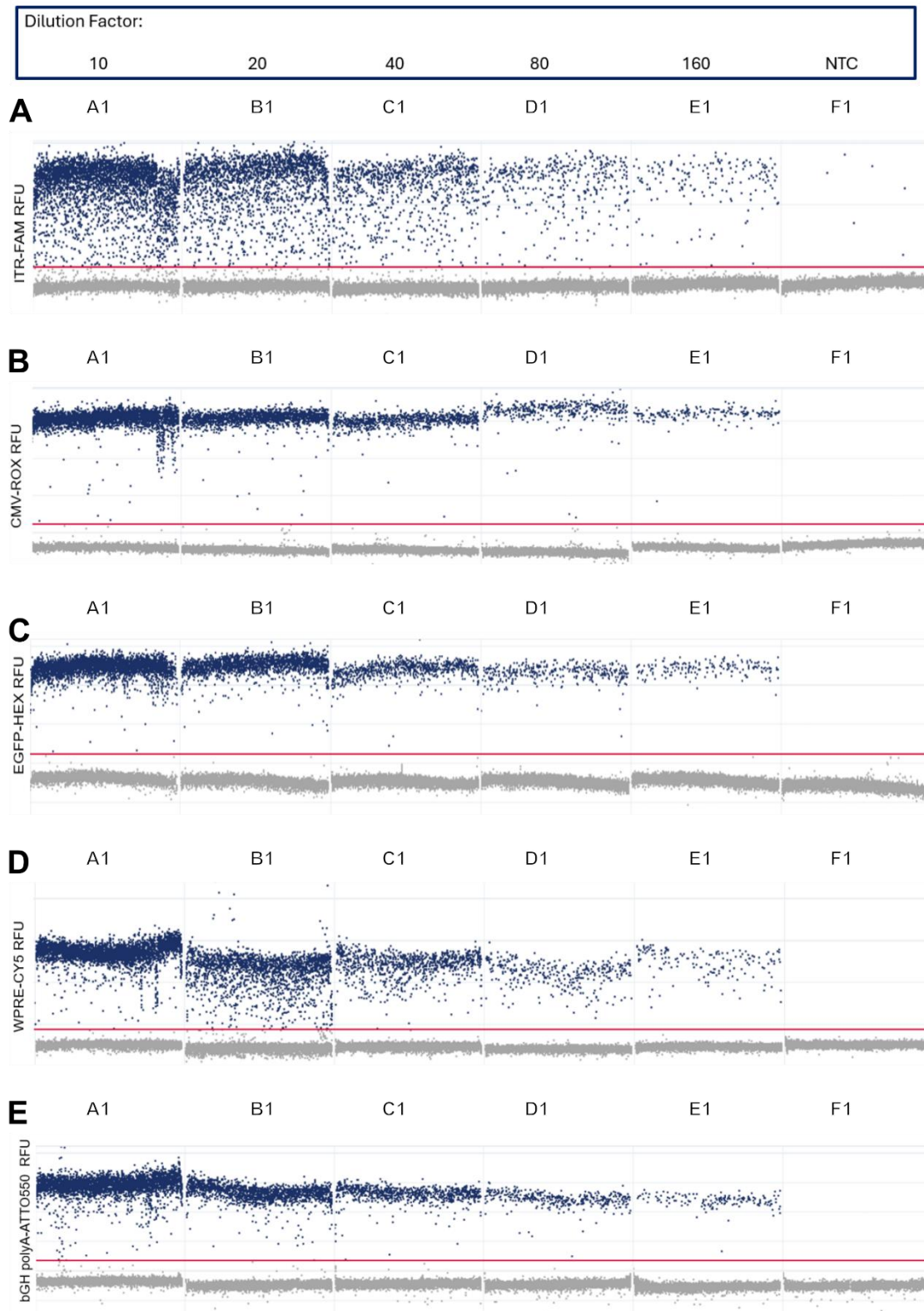


Figure 3.4 One-colour QIAcuity dPCR plots of the analysed serial dilutions of trFragment. Five-point 2-fold dilution series with 5 assays tested on trFragment. The dilution factor of each trFragment dilution (and NTC) is shown at the top and applies to the wells displayed in panels (A–E). (A) ITR-FAM. (B) CMV-ROX. (C) *eGFP*-HEX. (D) WPRE-CY5. (E) bGH polyA-ATTO550.

Table 3.7 Inter-assay variability across the points of the standard curve.

Mean Concentration (copies/ μ L) across Assays	Inter-assay Variability (%)
4,568	6.0
1,976	7.0
835	9.5
351	7.7
155	12.6

Rain can be observed in the dPCR plot of the ITR assay across all tested dilutions. In contrast, the plots for the CMV, *eGFP*, and bGH polyA assays that were tested in trFragment, do not exhibit rain (**Figure 3.4**). Moreover, rain, though to a lesser extent, is also observed in the WPRE assay and particularly at the dilution factor of 20. This could be due to factors such as assay design, slight variations in sample loading onto the plate, or minor fluctuations in reaction conditions.

To assess the percentage of rain in each well of the ITR assay, the threshold was set just below the positive partition cluster. The number of positive partitions detected at this threshold was then subtracted from the number of positive partitions measured when the threshold was placed just above the negative cluster, as illustrated in **Figure 3.4A**. This difference was then divided by the number of positive partitions detected at the lower threshold. The calculated percentage of rain, from the lowest to the highest dilution factor, was 11.5%, 11.4%, 10.8%, 7.3%, and 7.6%.

In the ITR one-colour plot of the NTC, six positive partitions were observed, whereas no positive partitions were detected in the NTCs of the other assays (CMV, *eGFP*, WPRE, bGH polyA). To assess the possibility of contamination, the number of positive partitions in the NTCs was monitored across multiple experiments using a uniplex ITR assay on the QIAcuity instrument. As shown in **Figure 3.5**, the number of positive partitions remained more-or-less constant, not exceeding eight, indicating no accumulation of potential contamination over time. Since the number of ITR-positive partitions in the NTCs remained stable across multiple experiments and did not increase over time, and no positive partitions were observed in the NTCs of the other tested assays, this strongly suggests that the observed signals are not related to contamination. Instead, the signal may plausibly be related to marginal off-target

effects of the ITR assay on environmental DNA sources. Therefore, no background subtraction was needed.

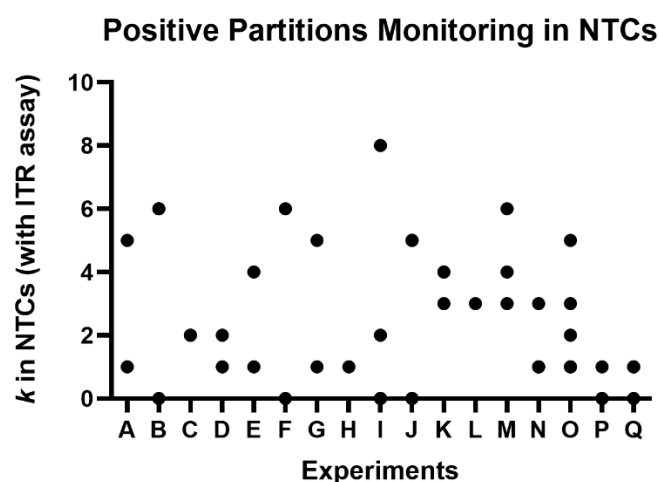


Figure 3.5 Monitoring of the number of positive partitions observed in NTCs analysed with the uniplex ITR assay by QIAcuity dPCR. The y-axis represents k , the number of positive partitions detected in NTCs in each experiment. The x-axis represents experiments (A–Q) arranged in reverse chronological order with Experiment A being the most recent. Each circle corresponds to a replicate of an NTC.

3.4.4 Comparison of dPCR and qPCR

The genome titre of extracted rAAV DNA samples, including rAAV2RefMat, rAAV2, and rAAV9, measured by qPCR as described in Chapter 2, was compared with dPCR measurements of the same samples using the ITR assay in the QIAcuity instrument in three independent experiments (**Figure 3.6**). Additionally, the rAAV2RefMat sample was measured by dPCR with the ITR and SV40 polyA uniplex assays. For the calculation of the genome titres in the stock, the measured values were multiplied by the dilution factor of each sample.

A general trend was observed across all rAAV samples and the trFragment, where dPCR consistently estimated higher ITR concentrations than qPCR (**Figure 3.6**). However, direct comparison between the two techniques is challenging because dPCR does not differentiate between genomes containing both ITRs and those with a single ITR as both are counted as having one ITR per partition. In contrast, qPCR, which relies on fluorescence intensity, detects both ITR copies.

To address this, the trFragment and rAAV2-extracted DNA were digested with *MspI* to separate the two ITRs into distinct partitions. The *MspI* restriction enzyme was used because

of its ability to cut rAAV ssDNA genomes (Carver, Blech-Hermoni and James, 2021). Moreover, MspI has multiple digestion sites, ensuring efficient cleavage and separation of ITRs when at least one site is cut. As a result, if a genome contains both ITRs, two positive partitions are detected instead of one. If all genomes were intact and contained both ITRs, this would theoretically lead to a doubling of the detected copy number. However, this was not observed due to the presence of partial genomes or inefficient digestion. Digestion of trFragment and rAAV2 resulted in 48.7% and 60.0% increase in ITR copy number concentration, respectively.

Without MspI digestion, dPCR estimated $40.0 \pm 13.5\%$ higher titres than qPCR for trFragment. After digestion, dPCR titres were $108.1 \pm 12.2\%$ higher than qPCR. Similarly, for rAAV2, dPCR titres were $39.2 \pm 16.9\%$ higher than qPCR without digestion and $122.7 \pm 13.8\%$ higher with MspI digestion. For rAAV2RefMat and rAAV9, dPCR titres were also higher than qPCR by $45.3 \pm 8.7\%$ and $23.9 \pm 14.0\%$, respectively, though MspI digestion was not performed on these samples. This consistent trend of dPCR giving higher titres, especially when the 5' and 3' ITRs are separated using MspI, suggests a systematic difference between dPCR and qPCR.

Notably, the genome titre measured by dPCR using the ITR assay for rAAV2RefMat aligned with the 95% confidence interval reported by Lock *et al.*, 2010 based on qPCR targeting the SV40 polyA. In contrast, the genome titre measured by dPCR using the same SV40 polyA assay was 37.0% lower than the lower limit of the reported 95% confidence interval (Lock *et al.*, 2010).

The precision of the two methods was compared by estimating the inter-experimental variability, expressed as CV, for each sample. In the tested samples (rAAV2, rAAV2RefMat, and trFragment), the CV was consistently lower with dPCR than with qPCR, indicating that dPCR offers better precision (**Table 3.8**). Specifically, the CV for trFragment was 17.0% with qPCR and 11.5% with dPCR, for rAAV2, 22.0% with qPCR and 13.8% with dPCR, for rAAV2RefMat, 12.2% with qPCR and 6.1% with dPCR, and for rAAV9, 18.7% with qPCR and 12.5% with dPCR.

Table 3.8 Comparison of inter-experimental precision between qPCR and dPCR.

Sample	CV qPCR (%)	CV dPCR (%)
trFragment	17.0	11.5
rAAV2	22.0	13.8
rAAV2RefMat	12.2	6.1
rAAV9	18.7	12.5

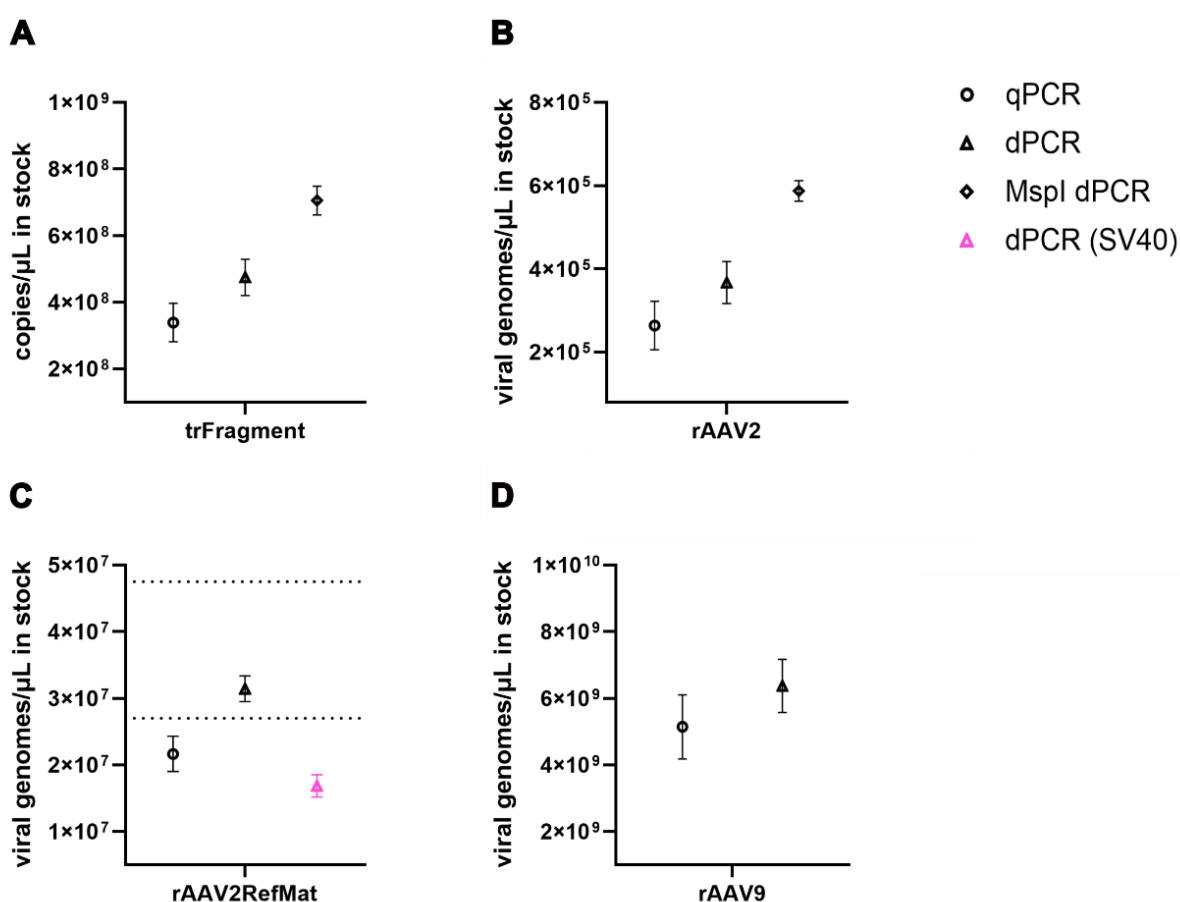


Figure 3.6 Comparison of qPCR and QIAcuity dPCR measurements in extracted rAAV DNA samples. Measurement of concentrations with qPCR (ITR, circles), dPCR (ITR, triangles), dPCR after MspI digestion (ITR, diamonds), and SV40 polyA dPCR (pink triangles, rAAV2RefMat only). (A) trFragment. (B) rAAV2. (C) rAAV2RefMat. Dotted lines indicate the 95% confidence interval reported by Lock *et al.*, 2010 with the SV40 polyA qPCR assay. (D) rAAV9. Error bars for qPCR and dPCR measurements correspond to the SD of 3 experimental replicates, except for the dPCR measurement of MspI-digested rAAV2 DNA (plot B, diamond), where error bars represent the SD of 3 technical replicates.

3.4.5 Comparison of Pre-Extracted rAAV DNA and In-Partition Lysed rAAV Particles

In Section 3.4.4, DNA extraction was performed to compare qPCR and dPCR measurements. However, this step may not be necessary, as the denaturation step at 95 °C for 10 min during

thermal cycling could be sufficient to dissociate the capsids and release the genome. To evaluate whether in-partition lysed rAAV2 particles (direct addition) or extracted rAAV2 DNA yield different genome titre measurements, dPCR titration results from both sample types were compared across three paired experiments, each performed in duplicate per experiment. It is noted that the rAAV2 DNA was extracted once using the rAAVpro® Titration Kit (for Real-Time PCR) Ver.2 (Takara Bio), as described in Chapter 2, and the same sample was measured in three independent experiments, rather than performing the extraction three separate times. The directly added rAAV2 particles were also measured in three experiments, with in-partition lysis occurring in each experiment.

Both the extracted rAAV2 DNA and the directly added rAAV2 showed similar genome titres when measured by dPCR ($3.7 \times 10^5 \pm 6.2 \times 10^4$ viral genomes/ μL for extracted rAAV2 vs $3.1 \times 10^5 \pm 7.3 \times 10^4$ viral genomes/ μL for directly added rAAV2) (**Figure 3.7**), with a paired two sample t-test for means showing no significant difference between the two samples ($p > 0.05$). Given the comparable results and the advantages of a simpler workflow, it was decided to use direct particle analysis for genome titre measurements in the remainder of the work reported in this chapter. This approach simplifies the workflow, reduces potential biases from genome fragmentation due to freeze-thaw cycles, and minimises sample processing.

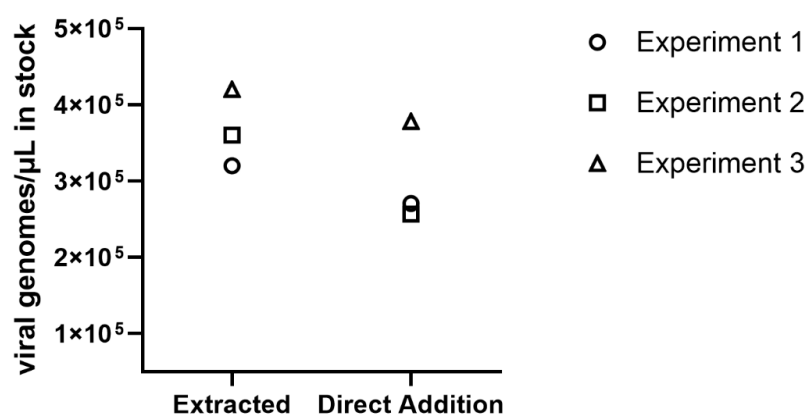


Figure 3.7 Comparison of genome titres between pre-extracted rAAV2 DNA and in-partition lysed rAAV2 particles in paired experiments with QIAcuity dPCR. Measured genome titres of pre-extracted rAAV2 DNA compared to rAAV2 directly added to the dPCR system, where in-partition lysis was performed at each experiment. Each symbol represents an individual experiment: circles correspond to experiment 1, squares to experiment 2, and triangles to experiment 3.

3.4.6 Quantification of rAAV Samples with Uniplex Assays

The genome titres of rAAV2, rAAV8, and rAAV9 samples were measured using uniplex assays targeting regions across their genomes to assess how the titration results vary depending on the assay used. The assays selected for each sample were based on the recombinant genome of the specific rAAV sample analysed. **Table 3.2** outlines the uniplex assays used for measuring each rAAV sample. To determine the appropriate concentration of rAAV samples for analysis, the need for low inter-assay variability (below 10%) and a low probability of molecule co-partitioning were considered. Therefore, samples were analysed at an expected concentration of 500 copies/ μ L (based on the ITR target), at which approximately only 3.5% of partitions are predicted to contain more than one molecule and the inter-assay variability is below 10% (Section 3.4.3). The rAAV2, rAAV8, and rAAV9 particle samples were directly added to the dPCR instrument. Each rAAV sample was analysed in three independent experiments. **Figure 3.8** displays the target regions measured for each sample along the x-axis of each plot (A–D). The intra-assay variability was calculated for each rAAV sample at the measured concentration (**Table 3.9**). The intra-assay variability differed substantially between the various rAAV samples analysed. For example, although the variability for the ITR assay in rAAV9 was only 2.7%, it increased to 18.0% for rAAV2 and 32.0% for rAAV8. A similar trend was observed for the WPRE assay, where variability was 12.0% for rAAV9 and 21.7% for rAAV2. For the bGH polyA assay, variability was 12.9% for rAAV9 but higher for rAAV2 at 23.2%. In the *eGFP* assay, variability was 14.7% for rAAV2 and 16.0% for rAAV9, but higher for rAAV8 at 28.5%. These results indicate that the intra-assay variability of each assay can differ substantially depending on the rAAV sample analysed.

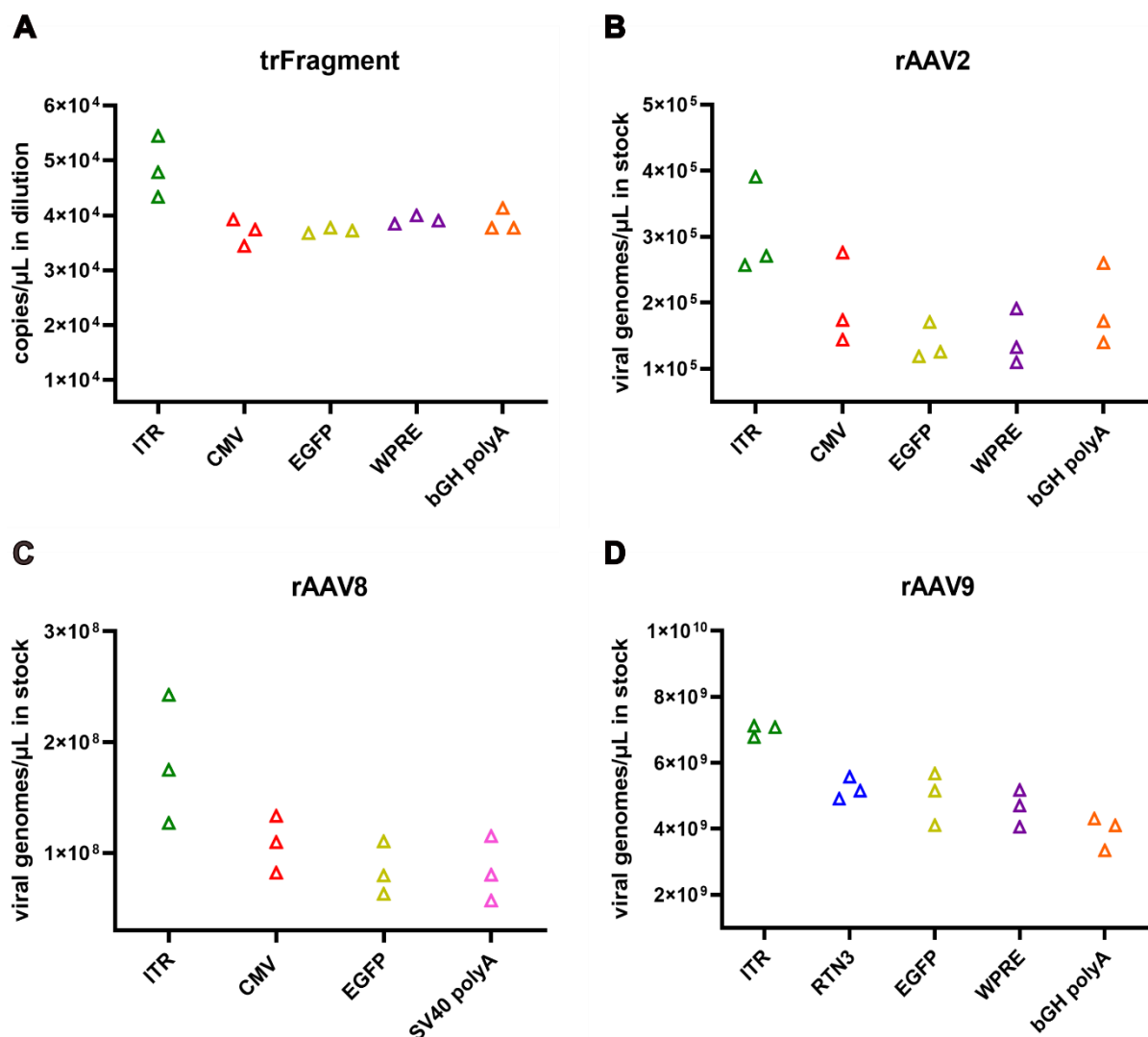


Figure 3.8 Genome titre measurement by QIAcuity dPCR based on assays targeting different regions across the recombinant genome. Each symbol represents an individual experiment, with three experiments performed for each assay per sample. (A) Measurement of trFragment concentration using ITR, CMV, *eGFP*, WPRE, and bGH polyA uniplex assays. (B) Measurement of viral genome titre of rAAV2 using ITR, CMV, *eGFP*, WPRE, and bGH polyA uniplex assays. (C) Measurement of viral genome titre of rAAV8 using ITR, CMV, *eGFP*, and SV40 polyA uniplex assays. (D) Measurement of viral genome titre of rAAV9 using ITR, *RTN3*, *eGFP*, WPRE, and bGH polyA uniplex assays.

Table 3.9 Intra-assay variability estimated for each sample analysed.

Assay	rAAV2	rAAV8	rAAV9
ITR	18.0	32.0	2.7
CMV	24.0	23.7	
<i>RTN3</i>	-	-	6.5
<i>eGFP</i>	14.7	28.5	16.0
WPRE	21.7	-	12.0
bGH polyA	23.2	-	12.9
SV40 polyA	-	35.0	-

A common trend was observed across all tested rAAV samples and the trFragment, where the measured concentration of ITRs was consistently higher than other target regions (**Figure 3.8**). This observation was supported by two-factor ANOVA without replication performed on the analysed samples. For trFragment, the analysis revealed a highly significant difference between assays when the ITR assay was included in the analysis ($p < 0.01$), whereas no significant difference was observed when the ITR assay was excluded ($p > 0.05$).

A similar trend was observed in rAAV2, where two-factor ANOVA demonstrated an extremely significant difference between assays when the ITR assay was included ($p < 0.001$), but no significant difference when it was excluded ($p > 0.05$). However, in rAAV8, an extremely significant difference between assays was detected when the ITR assay was included in the analysis ($p < 0.001$), and a highly significant difference was found even when the ITR assay was excluded from the analysis ($p < 0.01$). In rAAV9, no significant difference between assays was observed when the ITR assay was excluded ($p > 0.05$). However, when the ITR assay was included, an extremely significant difference was detected ($p < 0.001$). This observation is further illustrated in **Figure 3.9**, which shows the ratio of ITR concentration relative to that of the other target regions for each analysed sample. In all cases, the ratios were greater than 1, indicating that the ITR concentration was consistently higher than the other targets within the same sample.

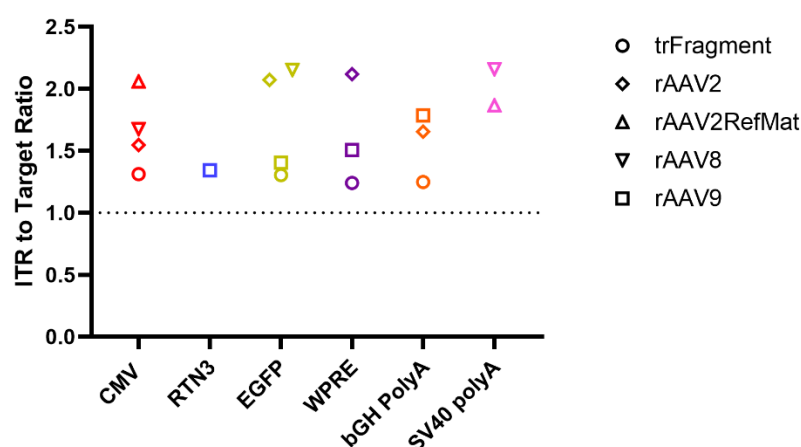


Figure 3.9 Measurement of the ratio of ITR concentration to other target regions in rAAV samples and trFragment by QIAcuity dPCR. Comparison of the ratio of measured ITR concentrations to the rest of the target regions for trFragment, rAAV2, rAAV8, rAAV9, and extracted rAAV2RefMat. Each symbol corresponds to the mean ratio of the targets for three experiments per sample. The dashed line at ratio 1 represents the nominal value of the ITR concentration for each sample.

To evaluate the variability in rAAV genome titre measurements across different assays, inter-assay variability was calculated for each sample (**Table 3.10**). Given the higher observed ITR titres, the inter-assay variability was calculated excluding the ITR assay. This approach provided a clearer understanding of how the other genome regions were measured independently of ITR-related effects. Inter-assay variability was assessed by calculating the CV based on measurements obtained from each assay across three independent experiments.

The trFragment exhibited low inter-assay variability, with a CV of 2.9%, indicating that the CMV, *eGFP*, WPRE, and bGH assays perform comparably. This observation suggests that the higher inter-assay variability observed in rAAV samples (29.1% for rAAV2, 8.7% for rAAV2RefMat, 28.1% for rAAV8, and 15.1% for rAAV9) compared to trFragment is likely due to sample-related factors (discussed in Section 3.5) rather than assay performance.

Table 3.10 Inter-assay precision across analysed samples.

Sample	CV % (excluding ITR assay)
trFragment	2.9
rAAV2	29.1
rAAV2RefMat	8.7
rAAV8	28.1
rAAV9	15.1

3.4.7 Development of 4-plex and 5-plex dPCR Assays

The next step was to integrate the previously tested assays for each sample to develop higher-order multiplexing. The initial multiplex assay development was conducted using the trFragment. A 4-plex assay was developed, utilising the QIAcuity detection channels green, yellow, orange, and crimson. This assay is called 4-plex A assay and targets ITR in the green channel, *eGFP* in the yellow channel, bGH polyA in the orange channel, and WPRE in the crimson channel. The 4-plex A assay consisted of the following assays: ITR-FAM, *eGFP*-HEX, WPRE-Cy5, and bGH polyA-ATTO550. In parallel, a 5-plex assay was developed by incorporating an additional target in the red channel of QIAcuity. The additional target

selected was the CMV enhancer, measured using the CMV assay. Therefore, the 5-plex assay consisted of ITR-FAM, CMV-ROX, *eGFP*-HEX, WPRE-Cy5, and bGH polyA-ATTO550. The 4-plex and 5-plex assays were tested on trFragment in three independent experiments.

Two-colour plots were generated (software output) for each assay combination, with the ITR signal fixed on the x-axis. These plots were examined across all combinations, and thresholds were applied to both the x and y-axis relative fluorescence units (RFU). The thresholds divide each plot into four quadrants, each represented by a distinct colour: double-negative partitions (grey), ITR-only positives along the x-axis (light blue), target-only positives along the y-axis (yellow), and double-positive partitions containing both x and y-axis targets (dark blue) (**Figure 3.10**).

Rain was observed between the double-positive cluster and the cluster positive for the target on the y-axis. This rain has been previously attributed to the ITRs and, therefore, does not correspond to the y-axis target. Consequently, the threshold was set to ensure that the rain is included within the double-positive cluster.

In the bGH polyA-ATTO550/ITR-FAM and the CMV-ROX/ITR-FAM two-colour plots, multiple distinct clusters were observed in each quadrant (**Figure 3.10B, 3.10G, 3.10H**). Particularly, in the double positive quadrant of the ATTO550-FAM two-colour plot of the 5-plex and 4-plex A assays, two clusters were present, a smaller one positioned at a higher ATTO550 RFU and a larger one at a lower RFU. Similarly, in the y-axis target (bGH polyA) quadrant, two clusters were observed, one smaller cluster at a higher ATTO550 RFU and a larger cluster at a lower RFU. Additionally, in the x-axis target (ITR) quadrant, three distinct clusters were present in the two-colour plot of the 5-plex assay, while two clusters could be seen in the two-colour plot of the 4-plex A assay, each positioned at different ATTO550 RFU values along the y-axis.

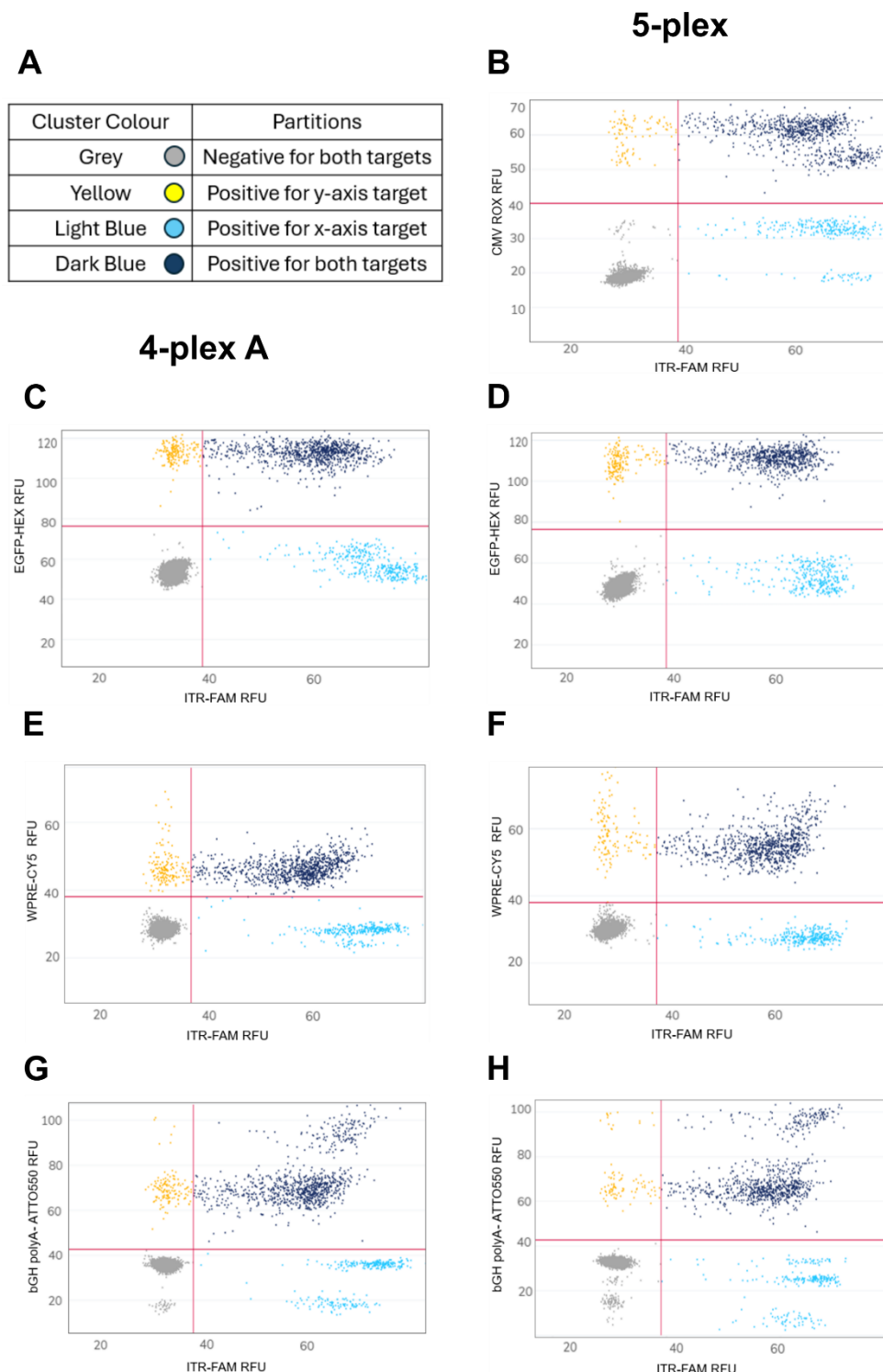


Figure 3.10 Two-colour QIAcuity dPCR plots for the 4-plex A and 5-plex assays on trFragment. (A) Colour key to aid interpretation of the plots. The x-axis represents the ITR target in every panel, while the y-axis corresponds to one of the other targets assessed in each panel (*eGFP*, *WPRE*, *bGH* polyA, or *CMV*). Panels (B), (D), (F), and (H) display two-colour plots from the 5-plex assay showing *CMV*–ITR, *eGFP*–ITR, *WPRE*–ITR, and *bGH* polyA–ITR, respectively. Panels (C), (E), and (G) show two-colour plots from the 4-plex A assay for *eGFP*–ITR, *WPRE*–ITR, and *bGH* polyA–ITR, respectively.

To further investigate these observations, the RFU values of each partition for every target measured with the 5-plex assay were exported and analysed to determine the number of targets present in each partition (**Figure 3.11**). The corresponding two-colour plot software output is also shown in **Figure 3.11** for direct comparison between the two-colour plots. **Figure 3.11C** outlines the different shapes representing the number of targets detected in each partition. Circles indicate partitions positive for a single target, triangles represent partitions positive for two targets, squares denote three targets, diamonds correspond to four targets, and squares with crosses indicate five targets. Additionally, grey circles signify partitions that are negative for both ITR and bGH polyA.

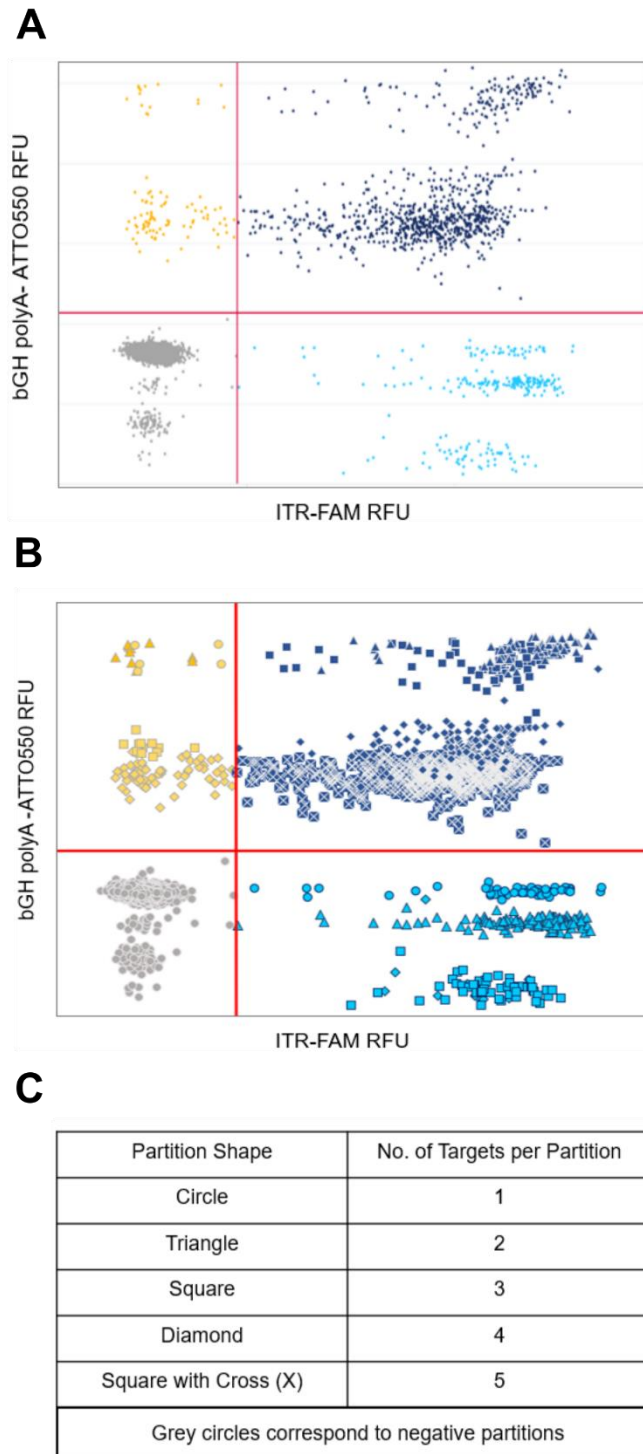


Figure 3.11 Partition occupancy analysis in ATTO550-FAM two-colour QIAcuity dPCR plot of the 5-plex assay. (A) Two-colour plot (ITR-FAM on the x-axis, bGH polyA-ATTO550 on the y-axis) generated using the QIAcuity Software Suite 1.2 (QIAGEN). (B) Two-colour plot (ITR-FAM on the x-axis, bGH polyA-ATTO550 on the y-axis) displaying RFU values of each partition. The measured RFU values for all channels in the 5-plex assay were exported and analysed to determine the number of targets present in each positive partition. (C) A reference table explaining the symbols in panel (B), indicating the number of positive targets detected in each partition shape.

In the double-positive quadrant, both clusters are positive for both targets. However, the larger cluster also includes partitions that are positive for additional targets (*eGFP*, WPRE, CMV) present in the 5-plex assay. Specifically, this larger cluster consists of partitions positive for all five targets in the 5-plex assay, as well as a subset of partitions positive for four targets (**Figure 3.11B**). In contrast, the smaller cluster, which sits at a higher RFU, consists of partitions that are positive only for ITR and bGH polyA, along with a subset of partitions positive for these two targets plus one additional target (three in total) (**Figure 3.11B**). This trend is also evident in the y-axis target quadrant, where partitions positive only for bGH polyA are positioned at a higher RFU, while the largest cluster includes partitions positive for more targets, including the bGH polyA (**Figure 3.11B**).

Interestingly, a similar phenomenon was observed in the x-axis target quadrant. Partitions positive only for ITR are positioned at a high RFU along the y-axis, while partitions positive for ITR and one additional target are positioned at a lower RFU, and those positive for ITR and two additional targets are positioned at even lower RFU (**Figure 3.11B**). However, in this quadrant, the partitions display lower RFU values along the y-axis rather than the x-axis, indicating that the presence of multiple targets in all quadrants only influences the RFU in the y-axis, which corresponds to the ATTO550 RFU in the orange channel, while the FAM RFU values remain relatively unchanged. This phenomenon was also observed in the ATTO550-FAM two-colour plot of the 4-plex A assay.

The measured copy number concentration of the trFragment was compared across uniplex, 4-plex A, and 5-plex assays (**Figure 3.12**). The uniplex and 4-plex A assays produced highly comparable results across all target regions, whereas the 5-plex assay consistently underestimated the concentration for every target. The mean copy number concentrations measured across all three experiments were $3.9 \times 10^4 \pm 4.9 \times 10^3$ copies/ μL across the uniplex assays, $4.0 \times 10^4 \pm 5.3 \times 10^3$ copies/ μL for the 4-plex A assay, and $3.0 \times 10^4 \pm 2.7 \times 10^3$ copies/ μL for the 5-plex assay. A paired t-test was conducted to compare each assay format, assessing differences between uniplex and 4-plex A, as well as between uniplex and 5-plex, across three independent experiments for each format. No significant differences were observed between uniplex and 4-plex A for any of the targets tested (ITR, *eGFP*, WPRE, bGH polyA) ($p > 0.05$). In contrast, a significant difference was detected ($p < 0.05$) when comparing the assays in uniplex

to the 5-plex assay format, except for the ITR assay ($p > 0.05$). Therefore, it was decided to use a 4-plex assay for the rest of this study.

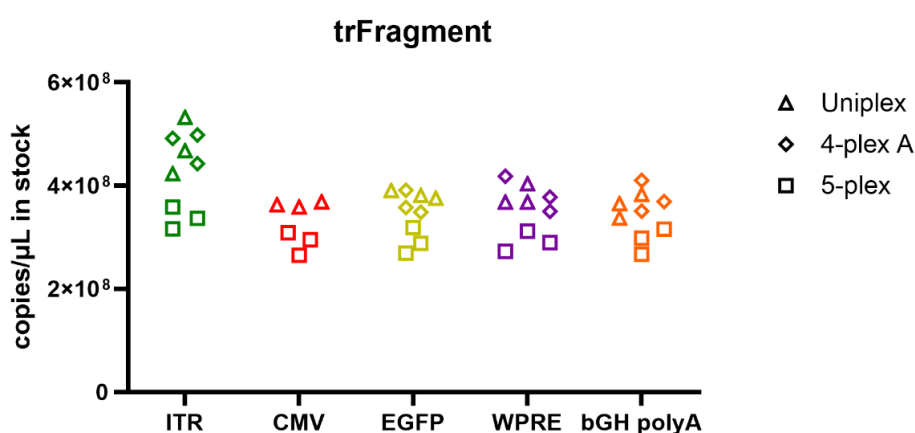


Figure 3.12 Comparison of the copy number concentrations of the target regions of trFragment between uniplex, 4-plex A, and 5-plex formats by QIAcuity dPCR. Measurement of trFragment stock copy number concentration with uniplex (triangles), 4-plex A (diamonds), and 5-plex (squares) assays. Each symbol corresponds to the mean value of an individual experiment, with a total of three independent experiments.

3.4.8 Comparison of 4-plex Assay Between QIAcuity and Absolute Q

To assess how different dPCR platforms compare when using a 4-plex A assay, the trFragment was analysed using the same 4-plex A assay on Absolute Q, which also has the capacity to multiplex four targets. Three independent experiments were conducted on the Absolute Q and QIAcuity platforms, each with a single replicate of the trFragment. In these experiments, the trFragment sample was analysed using both the 4-plex A assay and individual uniplex assays (ITR, *eGFP*, WPRE, and bGH polyA). It is noted that the variation in cluster density between the two-colour plots is due to differences in the number of partitions between the instruments: QIAcuity contains 8,184 partitions, while Absolute Q contains 20,475 partitions (**Figure 3.13A and 3.13B**). As seen in prior analyses, the two-colour plots of both QIAcuity and Absolute Q show the presence of rain. In both cases, the rain observed was included in the positive partitions.

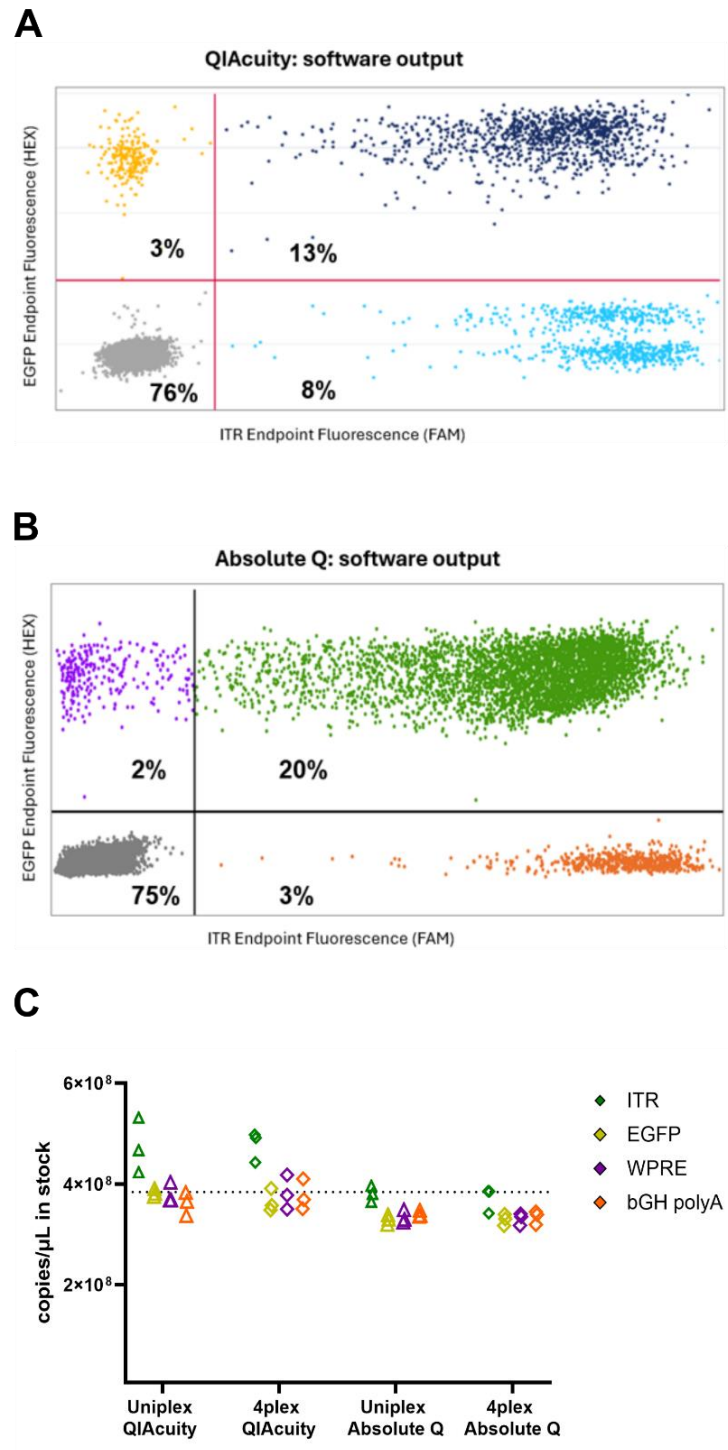


Figure 3.13 Comparison of 4-plex A and uniplex assays between QIAcuity and Absolute Q. Two-colour plot of ITR-FAM (x-axis) vs. *eGFP*-HEX (y-axis). (A) QIAcuity. (B) Absolute Q. (C) Comparison of the measured concentrations of trFragment using uniplex assays (triangles) targeting ITR, *eGFP*, WPRE, and bGH polyA, alongside the 4-plex A assay (diamonds), between Absolute Q and QIAcuity. The percentage of partitions in each quadrant relative to the total number of partitions is indicated in each plot. The dashed line represents the average concentration of uniplex assays (excluding ITR) measured by QIAcuity in trFragment. Each instrument uses bespoke software that assigns fixed colours to the clusters in each quadrant, which cannot be changed. As a result, the colours in the Absolute Q plot do not correspond to the colour key used in panel (C).

The concentrations of different target regions measured using the 4-plex A assay were highly consistent with those obtained using uniplex assays on Absolute Q. A paired t-test revealed no statistically significant differences between the uniplex and 4-plex formats in Absolute Q for any of the targets ($p > 0.05$). This was also previously shown (Section 3.4.7) in QIAcuity.

For all targets tested (ITR, *eGFP*, WPRE, and bGH polyA), QIAcuity consistently reported slightly higher concentrations than Absolute Q in both uniplex and 4-plex formats (**Figure 3.13**). Specifically, for the *eGFP*, WPRE, and bGH polyA assays in uniplex format, QIAcuity reported titres that were on average $12.1 \pm 2.0\%$ higher. When these assays were analysed as part of the 4-plex A assay, the titres measured by QIAcuity were $13.0 \pm 1.5\%$ higher than those measured by Absolute Q.

3.4.9 4-plex Optimisation for rAAV Genome Characterisation

The developed 4-plex A assay was also applied for the characterisation of the rAAV9 genome on QIAcuity. This assay targeted four genomic regions: ITR, *eGFP*, WPRE, and bGH polyA, using the assays ITR-FAM, *eGFP*-HEX, WPRE-Cy5, and bGH polyA-ATTO550. However, in the ATTO550-FAM two-colour plot, additional clustering was observed, as previously discussed. To investigate whether modifying the fluorophore combination could improve cluster resolution, the bGH polyA-ATTO550 assay was replaced with the *RTN3*-ROX assay. This new 4-plex assay consisted of the ITR-FAM, *eGFP*-HEX, *RTN3*-ROX, and WPRE-Cy5 assays and is referred to as 4-plex B assay.

Although a clear distinction between the four quadrants was achieved using both 4-plex A and 4-plex B assays, the two-colour plot of *RTN3*-ROX/ITR-FAM did not produce additional clusters within each quadrant, unlike bGH-polyA-ATTO550/ITR-FAM, providing a more definitive threshold setting and simplifying data analysis (**Figure 3.14**).

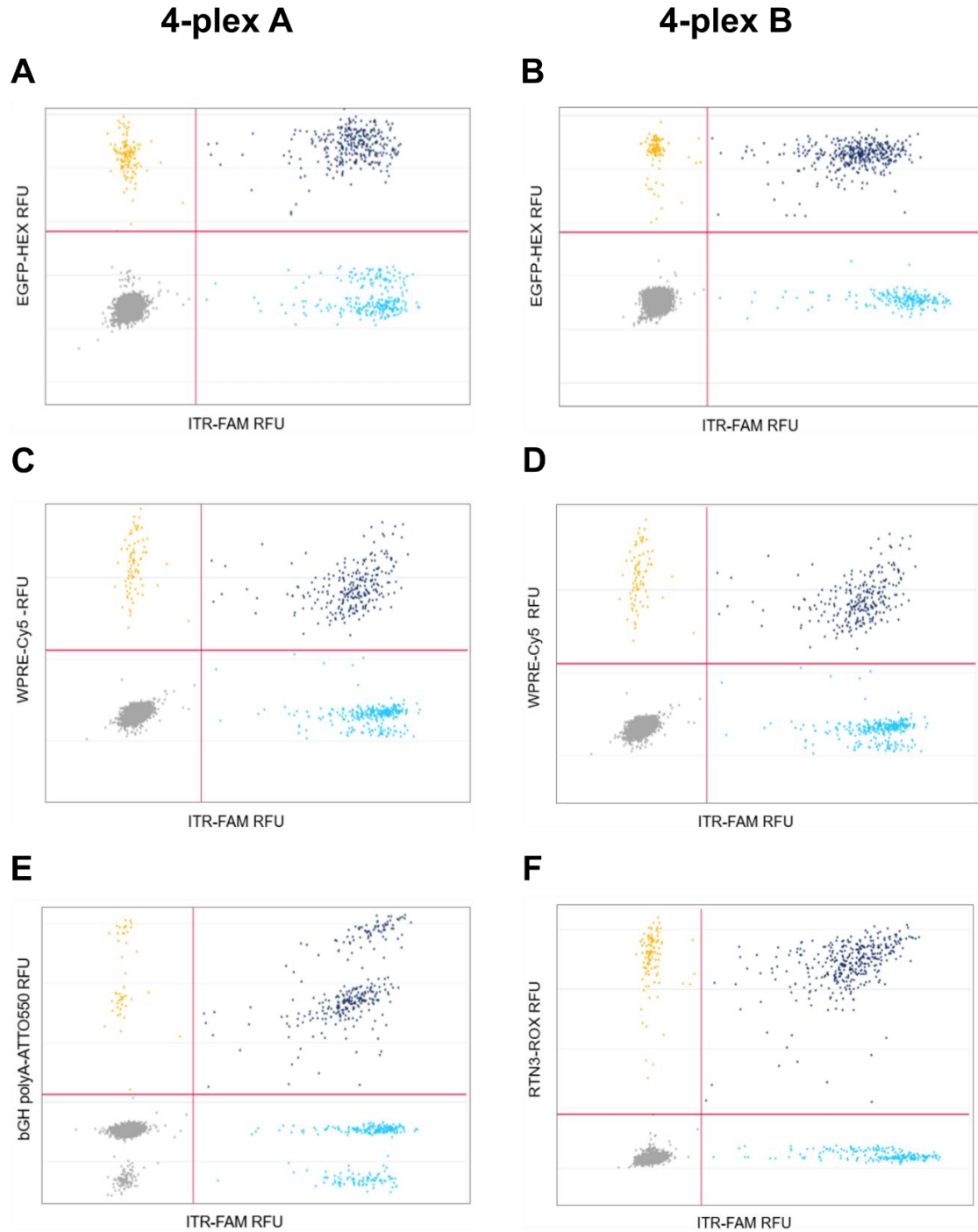


Figure 3.14 Comparison of two-colour QIAcuity dPCR plots between 4-plex A and 4-plex B assays on rAAV9. Two-colour plots where ITR is fixed on the x-axis, while the corresponding target is measured on the y-axis. (A, C, E) Two-colour plots of the 4-plex A assay: (A) *eGFP*-HEX/ITR-FAM, (C) WPRE-Cy5/ITR-FAM, (E) bGH polyA-ATTO550/ITR-FAM. (B, D, F) Two-colour plots of the 4-plex B assay: (B) *eGFP*-HEX /ITR-FAM, (D) WPRE-Cy5/ITR-FAM, (F) *RTN3*-ROX/ITR-FAM.

Both 4-plex A and 4-plex B assays yielded rAAV9 genome titre results consistent with those obtained using uniplex assays. Specifically, the mean genome titre value measured across all three experiments was $5.46 \times 10^9 \pm 1.05 \times 10^9$ viral genomes/ μL across the uniplex assays, $5.44 \times 10^9 \pm 1.25 \times 10^9$ viral genomes/ μL for the 4-plex A assay, and $5.67 \times 10^9 \pm 9.9 \times 10^8$ viral genomes/ μL for the 4-plex B assay. A paired t-test was performed for each target (ITR, *eGFP*, CMV, and WPRE) comparing their quantification in uniplex format to their performance within the 4-plex A assay. A paired t-test was also conducted to compare uniplex assays with the 4-plex B assay. In both cases, no significant differences were observed ($p > 0.05$) for every target, indicating that either 4-plex format did not affect genome titre measurements.

To further assess the method's versatility, the 4-plex B assay was applied to rAAV2 and rAAV8, using the same fluorophore configuration. The assay components of 4-plex B for rAAV2 and rAAV8 characterisation are shown in **Table 3.3**. The results demonstrated strong agreement between uniplex and 4-plex B assays for genome titre measurements in both rAAV2 and rAAV8, highlighting the robustness and adaptability of the approach (**Figure 3.15**). Specifically, the mean genome titre value measured across all three experiments for rAAV2 was $1.99 \times 10^5 \pm 8.34 \times 10^4$ viral genomes/ μL across the uniplex assays and $2.07 \times 10^5 \pm 8.21 \times 10^4$ viral genomes/ μL for the 4-plex B assay. For rAAV8, the mean genome titre value measured across all three experiments was $1.15 \times 10^8 \pm 5.23 \times 10^7$ viral genomes/ μL across the uniplex assays and $1.21 \times 10^8 \pm 5.72 \times 10^7$ viral genomes/ μL for the 4-plex B assay. Paired t-tests showed no significant differences ($p > 0.05$) when comparing rAAV8 genome titre measurements in uniplex versus 4-plex B format. Similarly, for rAAV2, no significant differences ($p > 0.05$) were observed for the ITR, *eGFP*, and CMV assays when comparing uniplex and 4-plex B results. However, the WPRE assay showed a significant difference between uniplex and 4-plex B ($p < 0.05$). The 4-plex B assay was selected to be used for the rest of the work described in this thesis.

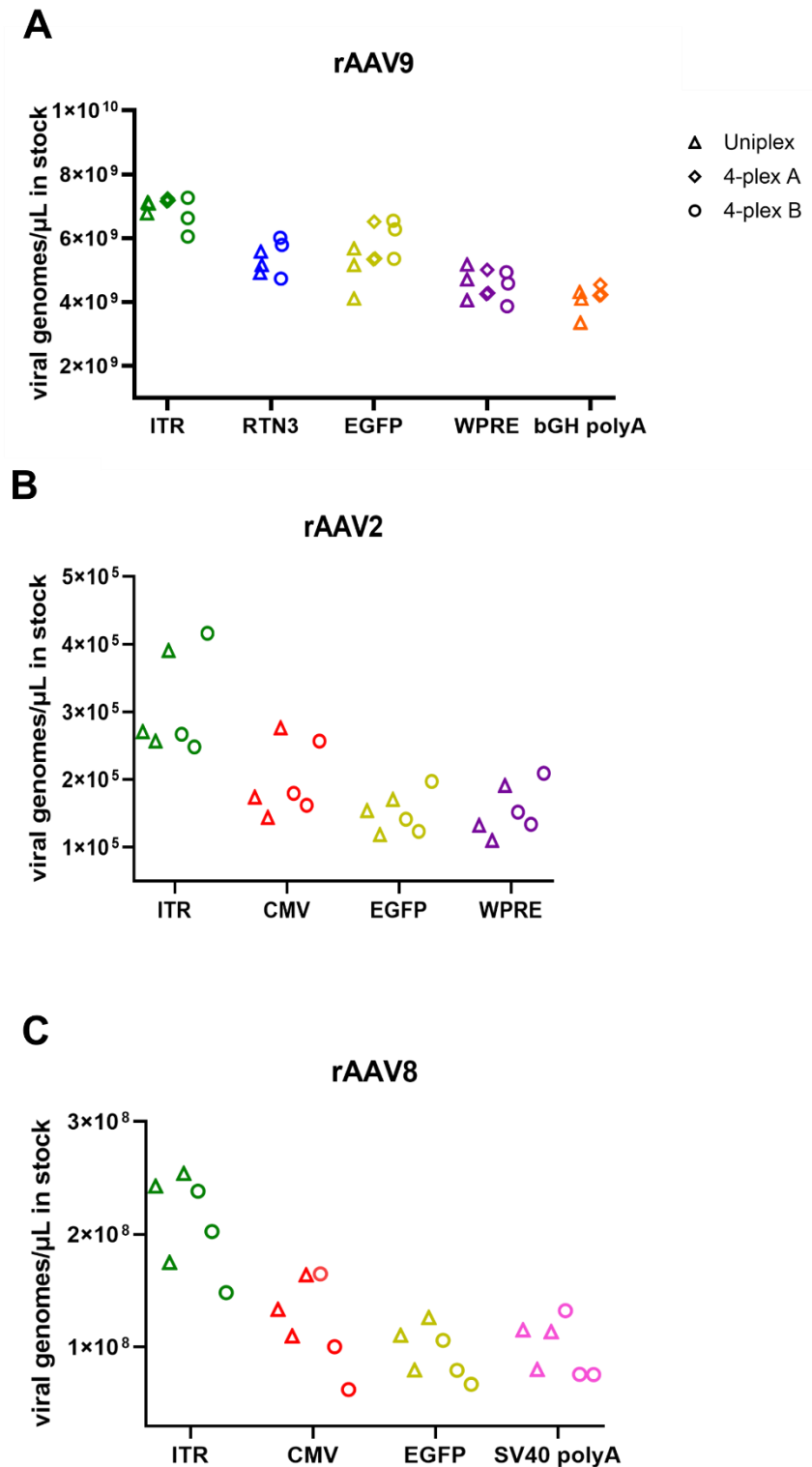


Figure 3.15 Comparison of uniplex and 4-plex assays for genome titre measurement across rAAV samples by QIAcuity dPCR. Viral genome titre measured with uniplex (triangles), 4-plex A (diamonds), 4-plex B (circles). Each symbol represents an individual experiment, with three experiments performed for each assay (uniplex and 4-plex) per sample. (A) rAAV9. (B) rAAV2. (C) rAAV8.

3.4.10 Genome Integrity Analysis

Evaluation of genome integrity is a key CQA for rAAV products. The 4-plex B assay developed in this chapter allows assessment of linkage between four targets distributed across the rAAV recombinant genomes, allowing estimation of the proportion of full-length genomes within each sample. Genome integrity (based on presence of four targets) was calculated using two approaches: (1) the simple fraction method, where the number of partitions positive for all four targets was divided by the total number of positive partitions, and (2) a statistical probability-based method using the DNA Integrity Calculation software developed by the Ghent University Digital PCR Centre (https://digpcr.shinyapps.io/intactness_dsi_assay/), based on the model described by Glerup *et al.*, 2023 (Figure 3.16).

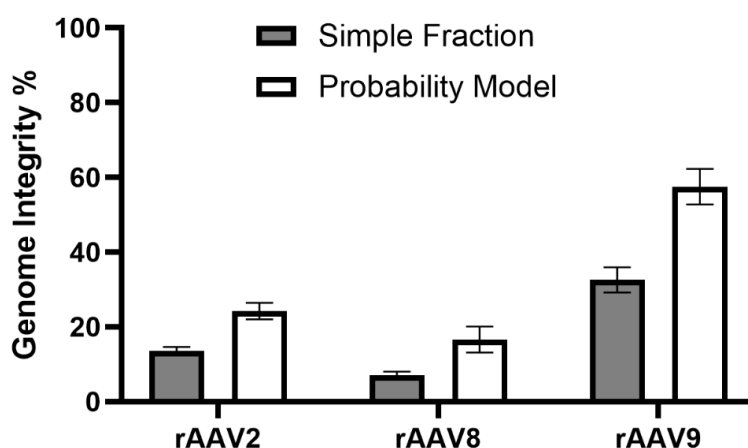


Figure 3.16 Genome integrity analysis of rAAV samples by QIAcuity dPCR. Genome integrity of rAAV2, rAAV8, and rAAV9 was assessed using the 4-plex assay B. Genome integrity was calculated using the simple fraction method (grey) and the statistical probability-based approach (white).

In all samples analysed, the probability-based method yielded higher genome integrity values compared to the simple fraction method: $24.2 \pm 2.2\%$ vs. $13.6 \pm 1.1\%$ for rAAV2, $16.6 \pm 3.5\%$ vs. $7.0 \pm 1.0\%$ for rAAV8, and $57.5 \pm 4.7\%$ vs. $32.6 \pm 3.4\%$ for rAAV9. While the difference between the two approaches was 10.7% for rAAV2 and 9.6% for rAAV8, a larger discrepancy of 25.0% was observed for the rAAV9 sample.

Among the rAAV samples, rAAV9 exhibited the highest genome integrity ($57.5 \pm 4.7\%$ with CV of 8.2%), while rAAV2 ($24.2 \pm 2.2\%$ (CV of 8.3%)) and rAAV8 showed lower values ($16.6 \pm 3.5\%$ with CV of 21.1%), indicating greater fragmentation and a higher proportion of incomplete genomes in those preparations.

3.5 Discussion

In this chapter, dPCR was implemented for the measurement of rAAV genome titre. Droplet-based and nanoplate-based dPCR approaches were evaluated by analysing the trFragment sample using the QX200 system, which employs droplet-based technology, as well as the QIAcuity and Absolute Q systems, both of which employ nanoplate-based technology.

Among the instruments assessed, QX200 exhibited the highest inter-experimental variability with a CV of 14.7% compared to 11.5% for QIAcuity and 4.2% for Absolute Q (**Table 3.6**). Since QX200 is a droplet-based system, while Absolute Q and QIAcuity are nanoplate-based, this difference in precision may be attributed to the fundamental differences between these two partitioning methods and differences in reaction preparation processes. Moreover, Absolute Q demonstrated the highest reproducibility among the dPCR platforms evaluated. This may be attributed to its lower percentage of dead volume compared to the other platforms, which refers to the portion of the reaction volume that is not used for partitioning. Specifically, the dead volume is 44% for the QX200, 4% for the Absolute Q, and 76% for QIAcuity 8.5K plates. Lower dead volume leads to better sample representation in the reaction and reduces variability between runs by minimising the impact of mixing inconsistencies in the sample.

A variability of 14.8% was observed between platforms, which is higher than that reported in a previous study (Wang *et al.*, 2024). In that previous study, an inter-platform variability of up to 8.5% was observed when the same batch of extracted rAAV DNA was measured across QX200, Naica (Stilla), QIAcuity, and Absolute Q instruments (Wang *et al.*, 2024).

At the linear range assessment of the ITR assay on trFragment, rain was consistently observed in the one colour plots across all dilutions (**Figure 3.4**). In contrast, no rain was observed in the assays targeting CMV, *eGFP*, WPRE (except one dilution), or bGH polyA. This effect is likely due to the secondary structure and high GC content of the ITRs, which can reduce amplification efficiency by interfering with proper denaturation during PCR cycling (Z. Liu *et al.*, 2021; Pan *et al.*, 2022). The stable secondary structures and high GC content of the ITRs may hinder primer and probe binding, resulting in inefficient amplification and contributing to the presence of rain in the ITR one-colour plots.

Moreover, each assay showed good linearity ($R^2 > 0.99$) across concentrations ranging from 4,568 to 155 copies/ μL (mean concentration across assays) (**Figure 3.3**). This range was constrained by the limited number of partitions (8,500) available in the 8.5k plates used on the QIAcuity platform. The dynamic range of dPCR assays is inherently limited by the number of partitions available on each platform. Using the 26k QIAcuity plates, which offer 26,000 partitions, would allow assessment of higher concentration ranges and extend the upper end of the linear range, although not by a large margin.

Unlike qPCR, dPCR provides absolute quantification without the need for a standard curve, offering improved precision (Kuypers and Jerome, 2017; Wang *et al.*, 2022; Hou *et al.*, 2023). A comparison of titration results between qPCR and dPCR for the quantification of trFragment and rAAV samples using an ITR-targeting assay showed that dPCR consistently resulted in higher titre measurements than qPCR (**Figure 3.6**). This difference was particularly evident when *MspI* was used to separate the two ITRs into different partitions, allowing for a more accurate comparison between the two methods. Under these conditions, dPCR reported 122.7% higher titre for rAAV2, and 108.1% for trFragment compared to qPCR (**Figure 3.6**). An observation that qPCR underestimated titres has also been observed in previous studies measuring extracted bacterial genomic DNA (Papić *et al.*, 2017; Ricchi *et al.*, 2017; Wang *et al.*, 2022). This discrepancy between qPCR and dPCR measurements can be attributed to the higher sensitivity of qPCR to factors that cause PCR inhibition, such as suboptimal primer binding and the presence of inhibitory substances (Dingle *et al.*, 2013; Lock *et al.*, 2014; Sedlak, Kuypers and Jerome, 2014; Devonshire *et al.*, 2016). In contrast, dPCR does not rely on amplification kinetics but instead uses end-point measurements to categorise a partition as positive or negative. This characteristic offers dPCR improved robustness and tolerance to PCR inhibitors (Sidstedt, Rådström and Hedman, 2020). Although dPCR and qPCR might be impacted differently by different inhibitors (Nixon *et al.*, 2014). Additionally, differences in buffer compositions and analysis software between platforms may contribute to the observed discrepancies (Hays *et al.*, 2022). Collectively, these factors can result in lower measured concentrations in qPCR compared to dPCR.

Improved inter-experimental precision was observed with dPCR compared to qPCR across all samples tested (**Table 3.8**). This improvement is likely due to the absence of a standard curve in dPCR, which removes a source of variability introduced by standard

preparation, including pipetting errors, mixing inconsistencies, degradation or adsorption of standards across experiments, as well as increased tolerance of dPCR to PCR inhibitors (Sedlak, Kuypers and Jerome, 2014; Kuypers and Jerome, 2017; Sidstedt, Rådström and Hedman, 2020)

It was also observed that the measured genome titre of rAAV samples can vary depending on the assay used (**Figure 3.8**). For example, in the trFragment and rAAV2 samples, titres based on CMV, *eGFP*, WPRE, and bGH polyA assays were lower than those based on the ITR assay (**Figure 3.9**). Similarly, in rAAV8 samples, the titres based on CMV, *eGFP*, and SV40 polyA assays were lower compared to ITR-based quantification. In rAAV9, titres based on *RTN3*, *eGFP*, WPRE, and bGH polyA were also lower compared to those obtained with the ITR assay. This observed difference is potentially due to incomplete packaging of the rAAV genome and the packaging of fragmented genomes. Since the packaging process begins with the ITRs, which serve as the packaging signal of rAAVs, incomplete genomes are more likely to contain ITR fragments. This results in an overrepresentation of ITR-containing sequences in the packaged genome population (Galli *et al.*, 2021; Tran *et al.*, 2022; Hashiba *et al.*, 2024). This was demonstrated by Tran *et al.*, 2022, who showed that rAAV particles previously considered empty, actually packaged a single ITR. Furthermore, Hashiba *et al.*, 2024 reported a higher frequency of ITRs being present in capsids compared to other fragments of the recombinant genome. These differences highlight the importance of standardising genome titre measurements and being aware of potential discrepancies when comparing results from different assays.

A similar observation was made for trFragment, though for a different reason. Unlike encapsidated rAAV genomes, trFragment exists as naked DNA without the protective capsid, making it more susceptible to fragmentation during freeze-thaw cycles and handling. If fragmentation occurs, the 5' and 3' ITRs of trFragment may separate and distribute into different partitions during dPCR. As a result, these ITRs are detected as independent molecules rather than as part of a single molecule, leading to an overestimation of ITR concentration.

The intra-assay variability differed substantially between the various rAAV samples analysed, with rAAV9 generally exhibiting lower variability compared to rAAV2 and rAAV8 (**Table 3.9**). A potential explanation for this observation is that the rAAV9 genome may be

more stable than the rAAV2 and rAAV8 genomes. This increased stability could make it less susceptible to degradation or structural changes caused by freeze-thaw cycles or sample handling (e.g. pipetting or vortexing), which could otherwise impact genome integrity and introduce variability in the measurements.

In this study, multiplex assays (4-plex A and 5-plex) were developed and initially tested on trFragment. Analysis of the resulting two-colour plots revealed the presence of multiple clusters within each quadrant in the bGH polyA–ATTO550/ITR–FAM plots of both the 4-plex A and 5-plex assays, as well as in the CMV–ROX/ITR–FAM plot of the 5-plex assay (**Figure 3.10**). These clusters appear to vary in RFU and seem to be influenced by the number of targets detected in each partition. Specifically, partitions that are positive for a greater number of targets tend to exhibit lower RFU values in the orange (or red for CMV–ROX) channel (**Figure 3.11**). A potential explanation for this phenomenon is that ATTO550 emits in the orange channel, which lies between the yellow and red channels in the spectral range. As a result, its emission spectrum partially overlaps with the HEX and ROX fluorophores used in the assay. Due to this overlap, signals from these fluorophores can interfere with one another, prompting the instrument to apply compensation corrections to distinguish their signals (The dMIQE Group *et al.*, 2020).

Digital PCR systems perform these corrections by subtracting an estimated contribution of overlapping fluorophores, such as HEX and ROX, from the detected ATTO550 signal in the orange channel. However, this estimation may be overestimated, leading to an unintended reduction in ATTO550 RFU when more targets, and therefore more fluorophores, are present in the same partition. On the other hand, FAM has minimal spectral overlap with the other fluorophores used in the assay. Because of this, no significant compensation is required, which explains why the RFU values for FAM remain stable in the x-axis.

A similar phenomenon is observed in the CMV–ROX/ITR–FAM two-colour plot, where the red channel (ROX) lies between the orange (ATTO550) and crimson (Cy5) channels in the emission spectrum (**Figure 3.10**). Multiple clusters appear within each quadrant at varying RFU levels in the red channel. Notably, these additional clusters are present only when the ATTO550-labelled assay is included. In contrast, they are absent in the corresponding *RTN3*–ROX/ITR–FAM plots from the 4-plex B assay, where ATTO550 is excluded (**Figure 3.14**). This observation supports the hypothesis that the presence of multiple clusters is a result of

spectral overlap and the subsequent compensation corrections applied by the analysis software. This issue may be mitigated by the new software release of the QIAcuity system (version 3).

Although the copy number concentrations of trFragment measured using the uniplex and 4-plex A assays were very similar, with no significant differences observed for any of the targets, the 5-plex assay consistently underestimated the copy number concentration across all assays (**Figure 3.12**). A possible explanation for this reduction is a potential increased rate of molecular dropout in the 5-plex assay. This occurs when the template is present but fails to amplify, leading to an underestimation of its concentration (Whale *et al.*, 2013). The inclusion of additional primer and probe sets in the multiplex reaction may contribute to this effect by increasing the likelihood of secondary structure formation or primer-dimer interactions. These factors can reduce amplification efficiency, preventing some targets from being detected and resulting in lower measured concentrations. For that reason, the 4-plex assay was selected for characterisation of the rAAV samples.

The 4-plex A assay was tested on both the QIAcuity and Absolute Q instruments alongside replicates of each corresponding uniplex assay. Both instruments showed no significant differences in the quantification of each target, whether measured in uniplex or in the 4-plex A format. However, QIAcuity consistently reported higher copy number measurements compared to Absolute Q across all assays (**Figure 3.13**). This difference may be attributed to platform-specific factors such as differences in dead volume, sample loading, or data analysis algorithms. Another potential source of variability could be inaccuracies in the measurement of partition volumes across different dPCR platforms. Each platform has its own defined partition volume, and errors in estimating these volumes can lead to discrepancies in the measured copy number concentrations (Košir *et al.*, 2017; Emslie *et al.*, 2019)

Besides 4-plex A, another 4-plex assay (4-plex B) was developed, with the main difference being the selection of fluorescence channels. In 4-plex A, targets were measured using the green, yellow, orange, and crimson channels with the dyes FAM, HEX, ATTO550, and Cy5. In 4-plex B, targets were measured using the green, yellow, red, and crimson channels with FAM, HEX, ROX, and Cy5, avoiding the orange channel.

Both 4-plex assays produced genome titre measurements consistent with uniplex assays for rAAV9 (**Figure 3.15**), and the threshold setting in the two-colour plots was straightforward for both methods (**Figure 3.14**). However, in 4-plex A, the inclusion of ATTO550 in the orange channel resulted in additional clusters appearing in the bGH polyA-ATTO550/ITR-FAM two-colour plots as previously described. This effect was not observed in 4-plex B, where ROX was used instead of ATTO550, suggesting that the additional clusters were specific to the orange channel (**Figure 3.14**). For that reason, the 4-plex B assay was selected for measurement of the genome titre and genome integrity of the rAAV samples.

Although no significant differences ($p > 0.05$) were observed when comparing genome titre measurements obtained using the uniplex and 4-plex B formats across all assays tested in rAAV2, rAAV8, and rAAV9, one exception was noted (**Figure 3.15**). A statistically significant difference was found between the uniplex WPRE assay and the WPRE assay within the 4-plex B format for rAAV2. As only three independent experiments were performed for this comparison, additional replicates would be required to determine whether the observed difference is consistent. However, since no significant differences were observed for the other assays in rAAV2, and for every assay in rAAV9 and rAAV8, the 4-plex B assay was used to estimate the genome integrity of the samples.

To estimate the genome integrity of rAAV samples and the trFragment plasmid, two analytical approaches were applied. The first was a simple fraction method, where genome integrity was calculated as the percentage of partitions positive for all four targets relative to the total number of positive partitions (Tereshko *et al.*, 2023). To minimise bias caused by co-partitioning of multiple DNA fragments, which can occur due to random distribution at higher concentrations, rAAV samples were diluted to approximately 500 copies/ μ L. At this concentration, only 3.5% of the 8,500 partitions are predicted to contain more than one molecule, based on the Poisson distribution. In addition to the simple fraction method, the data were also analysed using a probability-based statistical model developed by Gleerup *et al.*, 2023. The results obtained from both approaches were then compared. Interestingly, across all rAAV samples analysed, the simple fraction method consistently underestimated genome integrity compared to the probability-based model (**Figure 3.16**). One potential explanation for this underestimation is molecular dropout. As a result, affected partitions

appear negative for specific targets, despite the presence of the template, leading to an underestimation of linkage and thus genome integrity.

The difference in genome integrity estimates between the two approaches (simple fraction and probability model) was 10.7% for rAAV2, 9.6% for rAAV8, and 25.0% for rAAV9. The discrepancy between the two methods is not fixed and can vary substantially depending on the specific data generated for each sample. It is influenced by the total number of partitions analysed, the number of negative partitions, and the distribution of partitions that are single-positive, double-positive, triple-positive, or positive for all four targets. The size of the difference depends on how many partitions fall into each of these categories, and thus it is inherently dependent on the characteristics of each individual sample and analysis run. The statistical probability-based approach was used for the remainder of this thesis, as it accounts for molecular dropout and the random co-partitioning of separate molecules within the same partition, providing a more accurate estimation of genome integrity.

Among the rAAV samples tested, rAAV9 exhibited the highest genome integrity ($57.5 \pm 4.7\%$, with CV of 8.2%), followed by rAAV2 ($24.2 \pm 2.2\%$, with CV of 9.1%) and rAAV8 ($16.6 \pm 3.5\%$, with CV of 21.1%), demonstrating that a substantially higher number of full-length genomes were packaged in the rAAV9 preparation compared to rAAV2 and rAAV8 (**Figure 3.16**). This difference is potentially attributed to variations in the production and purification processes used for each preparation (Rumachik *et al.*, 2020; Galli *et al.*, 2021; Tran *et al.*, 2022; Fu *et al.*, 2023). Since multiplex dPCR is a novel method for measuring genome integrity, the results reported here require validation through comparison with an orthogonal method. For this purpose, the rAAV9 sample was analysed by nanopore sequencing in Chapter 4 to compare genome integrity data obtained from both methods. This sample was selected as it met the input requirements for nanopore sequencing, which typically requires loading approximately 10^{10} viral genomes. In contrast, the concentrations of the rAAV2, rAAV8, and rAAV2RefMat samples were too low to be assessed with nanopore sequencing.

While the 4-plex dPCR approach allows for estimation of genome integrity, it is not without limitations. One key limitation is that dPCR cannot distinguish between genomes containing both ITRs (5' and 3') and those with only a single ITR. Since the ITR-FAM assay detects ITR sequences regardless of whether one or both ends are present, any partition

containing at least one ITR will appear as positive (Aurnhammer *et al.*, 2012; Suoranta, Laham-Karam and Ylä-Herttuala, 2021). This limitation can lead to an overestimation of genome integrity, as partial genomes with only a single ITR are indistinguishable from full-length genomes containing both ITRs.

In summary, this study presents a flexible approach for designing 4-plex assays by maintaining a consistent fluorophore combination (FAM, HEX, ROX, and Cy5) while allowing target selection to be adapted based on specific applications. In this work, the genome integrity of rAAV2, rAAV8, and rAAV9 samples was estimated using the 4-plex B assay and applying a statistical probability-based approach that compensates for biases such as molecular dropout and co-partitioning of fragments. This assay enables the simultaneous characterisation of both genome titre and genome integrity of rAAV preparations, thereby supporting the comprehensive assessment of CQAs of rAAVs. Future refinement of the 5-plex assay could further improve the accuracy of genome integrity estimation by enabling analysis of five targets instead of four.

Chapter 4: Characterisation of rAAV Identity and Purity

4.1 Introduction

As previously described in this thesis, the packaging process of rAAVs typically results in heterogeneous populations of viral particles with incomplete or truncated genomes (Ayuso, Mingozi and Bosch, 2010; Wright, 2014). Though particles with partial genomes may be inactive, these can potentially undermine the therapeutic effect, disrupt proper dosing, and express truncated proteins or other aberrant gene products that trigger immune responses, compromising the safety of the therapy (Schnödt *et al.*, 2016; Kontogiannis *et al.*, 2024). Additionally, a recent study suggested that capsids previously classified as empty may contain single truncated genomes and/or unresolved ITRs (Tran *et al.*, 2022; Hashiba *et al.*, 2024). Yet, the effect of single ITRs on the safety and efficiency of gene therapies remains unclear.

Moreover, residual DNA from host cells and production plasmids might become encapsidated and pose significant safety risks. These impurities can lead to the expression of unintended elements. Generally, longer DNA fragments are more likely to be transcriptionally active, and consequently increase the risk of unwanted gene expression (Schnödt *et al.*, 2016; Brimble *et al.*, 2022; Kontogiannis *et al.*, 2024). Furthermore, such fragments may provoke immunogenicity by being recognised as foreign by the immune system (Schnödt *et al.*, 2016). To address such risks, the U.S. Food and Drug Administration (FDA) recommends limiting the presence of residual DNA to 200 nucleotides (U.S. Food and Drug Administration, 2020a; Dunker-Seidler *et al.*, 2025). Therefore, thorough assessment of genome integrity and monitoring of DNA impurities is essential to ensure the safety and efficacy of rAAV vectors in gene therapy.

Next-generation sequencing (NGS) is a powerful technique for assessing the genome integrity of rAAVs while also providing detailed insight into the purity of rAAV samples. NGS can be broadly categorised into short and long-read sequencing technologies, each with distinct advantages and limitations (Hu *et al.*, 2021; Dunker-Seidler *et al.*, 2025).

Illumina short-read sequencing platforms are widely used due to their high accuracy (> 99.9 %) and reliability (Guerin *et al.*, 2020; Düzgüneş *et al.*, 2025). However, due to read length limitations, up to 350 bp, it is not well suited for evaluating the full genome integrity of rAAVs

(Radukic *et al.*, 2020; Dunker-Seidler *et al.*, 2025). Further, Illumina sequencing does not provide information on the size of the residual DNA impurities, which is a crucial factor in regulatory assessments (U.S. Food and Drug Administration, 2020a).

Long-read sequencing potentially overcomes this application limitation by providing comprehensive insight into rAAV genome integrity and impurity profiles, as a single read can capture the entire rAAV genome. Since rAAV genomes are typically no more than 4.8 kb due to the limited packaging capacity of the capsid, long-read sequencing methods are well-suited for their analysis (Namkung *et al.*, 2022; Dunker-Seidler *et al.*, 2025). One common approach is Single Molecule, Real-Time (SMRT®) sequencing provided by Pacific Biosciences of California (PacBio), which offers high accuracy (> 99.999 %) but requires labour-intensive sample preparation and a large amount of rAAV genomic material (10^{12} - 10^{13} genomes) (Korlach, 2013; J. Zhang *et al.*, 2024; Tam Tran and WI Tai, 2024; Dunker-Seidler *et al.*, 2025). An alternative solution is nanopore long-read sequencing provided by Oxford Nanopore Technologies (ONT). This method not only preserves the full integrity of the rAAV genome without the need for amplification, but also captures complete genomic structures, providing detailed insights into genome integrity and residual DNA impurities (Namkung *et al.*, 2022; Dunker-Seidler *et al.*, 2025). In addition, nanopore sequencing offers a simpler sample preparation workflow and requires a significantly lower input ($\sim 10^{10}$ genomes) than SMRT. Though nanopore sequencing accuracy is not as high as the technologies previously described, recent developments to base calling algorithms and sequencing chemistries have significantly improved raw, consensus, and variant calling accuracies (Dommann *et al.*, 2024; Ermini and Driguez, 2024; Ye *et al.*, 2024; Dunker-Seidler *et al.*, 2025). Currently, nanopore sequencing is expected to have a modal quality score (Q score) equal to or higher than 20 (99% accuracy) and consistently reach Q scores of 30 (99.9% accuracy) (Ode *et al.*, 2024; Dunker-Seidler *et al.*, 2025).

Another approach that is increasingly being used to measure rAAV genome integrity is dPCR. As previously discussed in Chapter 3, multiplex dPCR assays can be designed to target different regions across the recombinant genome, allowing for the detection of multiple targets co-segregating into the same partition. Statistical probability-based models developed by Gleerup *et al.*, 2023 and previously applied in Chapter 3 are used here to estimate the linkage percentages between four targets, enabling the assessment of genome integrity and facilitating comparison with nanopore sequencing results.

4.2 Aims of this Chapter

In this study, nanopore sequencing was applied to evaluate rAAV9 genome integrity, identify impurities, and detect structural variations (SVs) within the genome. Two different methods for extraction of the rAAV genomes from capsids before analysis were assessed and compared: heat-based extraction and a commercial extraction kit. Additionally, the 4-plex B digital PCR assay described and developed in Chapter 3 was used to assess genome integrity in the analysed samples and the results were compared with the nanopore sequencing approach. Finally, dPCR was used to evaluate the yield of the different extraction methods and compare the results with the direct addition of viral particles into the dPCR instrument.

4.3 Materials & Methods

4.3.1 Samples

The experiments described here were conducted using the rAAV9 sample, previously analysed in Chapters 2 and 3. This vector has a 4.4 kb genome, which consists of the 5' and 3' ITRs, the CamKII α promoter, a disoptimised *RTN3* gene, the *eGFP* reporter gene, the WPRE element, and the bGH polyA, as detailed in Chapter 2, Section 2.3.7. A schematic representation of the construct is shown in **Figure 2.1**.

4.3.2 DNase I reaction

Following the instructions provided by ONT (<https://nanoporetech.com/document/ligation-sequencing-gdna-v14-adenovirus-sequencing>, last accessed on 23.04.2025), for each experiment a 10 μ L aliquot of rAAV9 ($7.0 \times 10^9 \pm 1.9 \times 10^8$ viral genomes/ μ L measured by uniplex dPCR targeting ITRs in Section 3.4.6) was diluted with nuclease-free water (Thermo Fisher Scientific, Cat No. AM9937) to a final volume of 170 μ L in a 1.5 mL Eppendorf DNA LoBind tube. To this, 20 μ L of DNase I reaction buffer (NEB, Cat No. M0303S) and 10 μ L of DNase I (NEB, Cat No. M0303S) were added for a final reaction volume of 200 μ L. The reaction was thoroughly mixed by pipetting and incubated at 37 °C for 10 min using a Bio-Rad CFX96 C1000 thermal cycler. Subsequently, 2 μ L of 0.5 M EDTA (NEB, Cat No. 7011V) was added to terminate the reaction, followed by mixing through pipetting and centrifugation. Finally, the samples were incubated at 75 °C for 10 min in a Bio-Rad CFX96 C1000 thermal cycler to

deactivate the DNase I. The 200 μL sample was then split into three aliquots of 66.7 μL each. Two aliquots were processed with different extraction methods, whereas the third remained untreated. A schematic of the full workflow is shown in **Figure 4.1**. Based on dPCR analysis of the rAAV9 sample described in Section 3.4.6 of Chapter 3, the starting quantity per sample prior to extraction was approximately 2.33×10^{10} viral genomes.

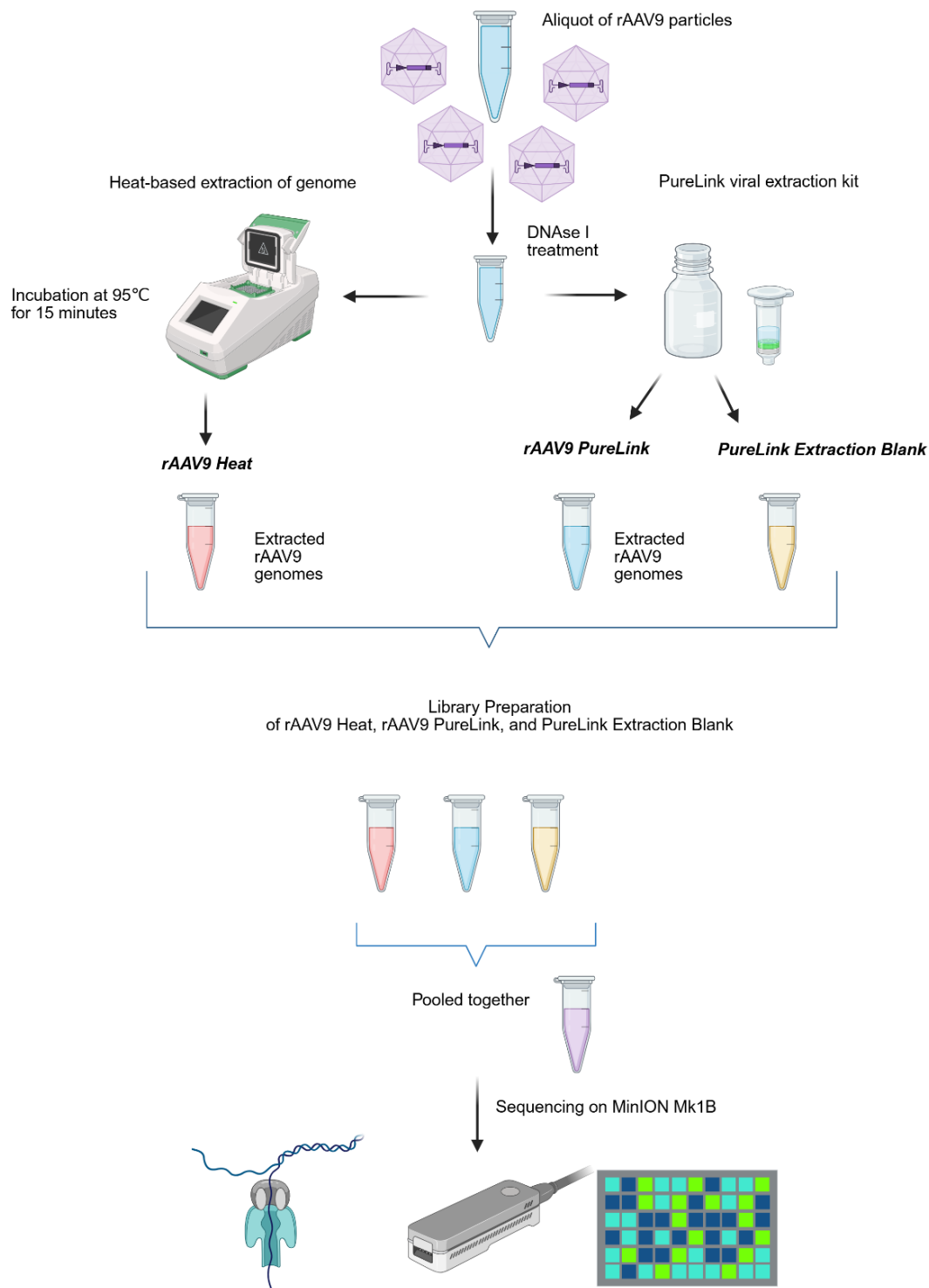


Figure 4.1 Schematic representation of the rAAV9 genome extraction and nanopore sequencing workflow. rAAV9 particles underwent a DNase I digestion to remove residual DNA molecules that are not packaged into capsids. The sample was then split, with one aliquot undergoing genome extraction using a heat-based method (red) and the other using the PureLink Viral RNA/DNA Mini Kit (blue). An extraction blank control was also processed in parallel (yellow). All three samples underwent library preparation and were subsequently pooled together (purple). The final library was sequenced using a MinION Mk1B instrument. Created with BioRender.com.

4.3.3 rAAV9 DNA Extraction using the Heat Method

For each experiment, one 66.7 µL aliquot of the DNase I-treated rAAV9 sample was incubated in a Bio-Rad CFX96 C1000 thermal cycler at 95 °C for 15 min to induce capsid denaturation and facilitate the release of rAAV9 genome DNA. This sample is referred to as “rAAV9 Heat” from this point onwards. Quality control (QC) assessment of the extracted DNA to assess sizing and integrity was performed using the Agilent TapeStation 4150 system (Agilent Technologies) with ScreenTapes D5000 (Agilent Technologies, Cat No. 5067-5588) and D5000 reagents (Agilent Technologies, Cat No. 5067-5589). At this stage, a 2 µL aliquot of the extracted rAAV9 DNA was stored at -20 °C for subsequent dPCR analysis. The remaining sample volume was stored at -20 °C for subsequent library preparation.

4.3.4 rAAV9 DNA Extraction using the PureLink Method

For each experiment, one 66.7 µL aliquot of the DNase I-treated rAAV9 sample was diluted to a final volume of 200 µL and subjected to DNA extraction using the PureLink™ Viral RNA/DNA Mini Kit (Thermo Fisher Scientific, Cat No. 12280050) following the manufacturer’s protocol. A non-template control (NTC) with equal volume was concurrently processed. Briefly, 25 µL of proteinase K was added to a microcentrifuge tube, followed by the addition of 200 µL of sample and 200 µL of lysis buffer containing carrier RNA of *Saccharomyces cerevisiae* (yeast tRNA). The mixture was vortexed for 15 seconds and incubated at 56 °C for 15 min on a heat block to facilitate lysis. After incubation, 250 µL of absolute ethanol (Fisher Scientific, Cat No. 10048291) was added and the mixture was then further vortexed for 15 sec and incubated at RT for 5 min to allow DNA precipitation. The lysate was then transferred to a viral spin column placed in a collection tube and centrifuged at $6,800 \times g$ for 1 min. The column was subsequently washed twice with 500 µL of Wash Buffer II, with the flow-through discarded after each wash. To dry the membrane, the column was centrifuged at maximum speed for 1 min. Finally, the column was transferred to a recovery tube and 60 µL of elution buffer added. Following a 1 min incubation at RT, samples were eluted by centrifugation at $6,800 \times g$ for 1 min. The samples processed with this method are referred to as “rAAV9 PureLink” and “PureLink Extraction Blank” from this point onwards. QC of the extracted samples was performed using the Agilent TapeStation 4150 system (Agilent Technologies) with ScreenTapes D5000 (Agilent Technologies, Cat No. 5067-5588) and D5000 reagents (Agilent Technologies,

Cat No. 5067-5589). A 2 µL aliquot of the extracted rAAV9 DNA was stored at -20 °C for subsequent dPCR analysis. The remaining sample volume was stored at -20 °C for downstream analysis.

4.3.5 Library Preparation

An optional self-hybridisation step is sometimes included to anneal any remaining (+) and (–) single strands of the rAAV vector prior to library preparation. However, in the workflow here, this step was omitted as an ONT technical note has demonstrated that skipping the annealing step results in higher yields of full-length sequences (<https://nanoporetech.com/document/requirements/AAV-seq>, last accessed on 12.03.2025). As a result, library preparation was initiated directly.

For each experiment, NGS libraries of the two previously extracted rAAV9 samples (rAAV9 Heat, rAAV9 PureLink), as well as the PureLink Extraction Blank, were prepared with Native Barcoding Kit 24 V14 (ONT, Cat No. SQK-NBD114.24) according to the manufacturer's instructions. As described in the protocol, the extracts were first end-repaired using the NEBNext® Ultra™ II End Repair/dA-Tailing Module (NEB, Cat No. E7546S). For each sample, 50 µL of sample was combined with 7 µL NEBNext Ultra II End-Prep Reaction Buffer (10×) and 3 µL of NEBNext Ultra II End-Prep Enzyme Mix. The reaction mixture was then incubated for 5 min at 20 °C, followed by 5 min at 65 °C in a Bio-Rad C1000 thermal cycler. Next, the samples were cleaned using AMPure XP Beads (Beckman Coulter Life Sciences, Cat No. A63881), and 60 µL of beads was added to each sample for a 1:1 (1×) volume ratio of beads to sample. The mixture was incubated on a HulaMixer (Thermo Fisher Scientific) at RT for 10 min. The samples were then placed on a magnetic rack and the supernatant was removed. Each sample underwent two consecutive washing steps with 200 µL of 80% ethanol. After each wash, the beads were resuspended, spun down, and placed back on the magnetic rack, followed by removal of the supernatant. After the washes, the pellet was allowed to dry and resuspended in 10 µL of nuclease-free water (Thermo Fisher Scientific, Cat No. AM9937). The samples were incubated for 10 min on a HulaMixer (Thermo Fisher Scientific) and then returned to the magnetic rack. The 10 µL eluates were recovered and transferred to clean Eppendorf DNA LoBind tubes.

Subsequently, a native barcode ligation step to allow multiplexing was performed on the three end-prepped samples. Unique barcodes were used for each experiment to prevent

potential between-run sample misassignment. For each ligation reaction, 7.5 μL of sample was combined with 2.5 μL of the chosen barcode and 10 μL of 2 $\times$ Blunt/TA Ligase Master Mix (NEB, Cat No. M0367S). The reactions were mixed gently by pipetting and incubated at RT for 20 min. To stop the reaction, 4 μL of 0.25 M EDTA (ONT) was added to each sample.

Next, the three barcoded samples were pooled in equal volumes (22 μL each). Note that equivolume rather than equimolar pooling was chosen to better compare the two extraction methods regarding the amount of library generated and consequent sequencing yield. According to ONT's instructions, AMPure XP Beads were added at a ratio of 0.4 $\times$ (26.4 μL for the combined 66 μL pooled sample). The mixture was incubated on a HulaMixer (Thermo Fisher Scientific) for 10 min at RT and processed with the same washing steps as described earlier, although 700 μL of 80% ethanol was used for washes, following the ONT instructions. The pellet was resuspended in 35 μL of nuclease-free water, gently mixed by flicking the tube, and incubated at 37 $^{\circ}\text{C}$ on a heat block for 10 min, with flicking every 2 min. The tube was then placed on a magnetic rack and the 35 μL eluate was transferred to a clean Eppendorf DNA LoBind tube. The pooled sample was stored at 4 $^{\circ}\text{C}$ overnight to preserve its integrity and allow continuation of the protocol the following day.

For sequencing adapter ligation, 30 μL of the pooled sample was mixed with 5 μL of Native Adapter (ONT), 10 μL of 5 $\times$ NEBNext Quick Ligation Reaction Buffer (NEB, Cat No. M2200S), and 5 μL of Quick T4 DNA Ligase (NEB, Cat No. M2200S). The reaction mixture was gently pipetted to mix and incubated at RT for 20 min. Next, a bead-based DNA cleanup was performed. Briefly, 20 μL of vortexed AMPure XP Beads (0.6 $\times$) were added to the reaction and the mixture was incubated on a HulaMixer for 10 min at RT. The sample was placed on a magnetic rack to pellet the beads and the supernatant was removed. The sample was washed twice with 125 μL of Short Fragment Buffer (SFB, ONT). After each wash, the beads were resuspended, spun down, and placed back on the magnetic rack, with the supernatant removed after pelleting. Following the second wash, the pellet was resuspended in 15 μL of Elution Buffer (ONT) and incubated on a heat block at 37 $^{\circ}\text{C}$ for 10 min with gentle flicking every 2 min to maximise recovery. Finally, the tube was placed on a magnetic rack, and 15 μL of the eluate was transferred to a clean Eppendorf DNA LoBind tube. QC analysis of the pooled library was carried out using TapeStation (Agilent Technologies) as previously described.

4.3.6 Nanopore Sequencing

The MinION Flowcell R.10.4.1 (ONT) was placed into the MinION Mk1B device (ONT), and a flow cell QC check was performed. Only flow cells with more than 500 available pores were utilised to ensure reliable results and adequate yield. After QC, the flow cell priming port was opened, and the presence of air bubbles was checked. A priming mix was prepared by combining 1170 μL of Flow Cell Flush (FCF, ONT), Flow Cell Tether (FCT, ONT), and 5 μL of bovine serum albumin (UltraPure™ BSA, 50 mg/mL, Thermo Fisher Scientific, Cat No. AM2616).

Next, 800 μL of the priming mix was loaded into the priming port. The port was closed, and the flow cell was incubated for 5 min. After incubation, both the priming port and the SpotON sample port were opened, and an additional 200 μL of the priming mix was loaded into the priming port. The sequencing library was then prepared by mixing 37.5 μL of Sequencing Buffer (SB), 25.5 μL of Loading Beads, 6 μL of the adapter-ligated DNA final library, and 6 μL of nuclease-free water to a final volume of 75 μL . Although the protocol recommends adding 12 μL of the adapter-ligated library, only 6 μL was used to retain the remaining 7 μL as a backup in case of any issues during the sequencing run. The 6 μL added corresponded to approximately 2.5 fmol. The 75 μL library mixture was gently mixed by pipetting to avoid shearing and loaded dropwise onto the flow cell via the sample port. The sample and the priming ports were securely closed, the light shield was placed on the flow cell, and the MinION Mk1B device lid was closed. The sequencing run was initiated with the MinKNOW software (version 24.06.8). Detailed information about sequencing runs is provided in **Table 4.1**.

Table 4.1 Nanopore sequencing runs specifications.

Run No.	Flow Cell Type	Flow Cell ID	Run Time (h)	Pores Available
1	FLO-MIN114 R.10.4.1	FAZ26403	3.2	523
2	FLO-MIN114 R.10.4.1	FBA08081	1.2	1378
3	FLO-MIN114 R.10.4.1	FBA08081	1.1	1105
4	FLO-MIN114 R.10.4.1	FBA08081	1.8	902

After each run, the flow cell was washed following ONT instructions to reduce cost and maximise usage. The flow cell wash mix was prepared by combining 2 μL of Wash Mix (WMX) with 398 μL of Wash Diluent (DIL). Waste from the sequencing runs was then removed from the waste channel through waste port 1 of the flow cell. The priming port was inspected to ensure it was free of air bubbles before 400 μL of the wash mix was added to the priming port, which was then closed. The flow cell was incubated for 1 h, after which the volume was removed from the waste channel through waste port 1.

4.3.7 Overview of Sequencing Runs

Four nanopore sequencing experiments were conducted (**Table 4.2**). In runs 1 and 2, rAAV9 Heat, rAAV9 PureLink, and PureLink Extraction Blank samples were processed, i.e. library preparation and sequencing, simultaneously. In runs 3 and 4, rAAV9 Heat and rAAV9 PureLink samples were singly processed (**Table 4.2**). Specifically, rAAV9 Heat was analysed in run 3 and rAAV9 PureLink was analysed in run 4. It is noted that for all experiments, a PureLink Extraction Blank was also processed. For each experiment, a separate 10 μL aliquot of rAAV9 was used for sample preparation. Though samples in runs 3 and 4 were library-prepped and sequenced on different days, these were also extracted from the same sample aliquot.

Table 4.2 rAAV9 samples analysed in each sequencing run.

Run No.	Samples	Aliquot No.	Library Prep and Sequencing	Date of Sequencing
1	rAAV9 PureLink	1	Same day	24/09/2024
	rAAV9 Heat	1		
2	rAAV9 PureLink	2	Same day	13/11/2024
	rAAV9 Heat	2		
3	rAAV9 Heat	3	Different day	28/11/2024
4	rAAV9 PureLink	3		03/12/2024

4.3.8 Data Processing and Analysis

Upon completion of sequencing runs, base calling, demultiplexing, and adapter trimming were performed on the generated raw sequencing POD5 files using Dorado (high-accuracy model, version 0.7.0). The resulting BAM files were converted to FASTQ format for downstream analysis. The remaining analysis was conducted on a Linux CentOS server. Quality and length-based filtering was performed with NanoFilt (2.8.0), considering a minimum quality threshold of 10 and a maximum length of 5 kb (De Coster *et al.*, 2018) (note that a PHRED quality score of 10 corresponds to a 90% base-call accuracy). A maximum length filter of 5 kb was implemented because the expected genome size was 4.4 kb, and longer molecules likely result from polymerase-induced elongation errors. No minimum length threshold was applied to assess the presence of short fragments. NanoStat (v.1.6.0) and NanoPlot (v.1.43.0) were used to generate pre- and post-filtering read metrics such as read number, mean read quality, and mean read length (De Coster *et al.*, 2018; De Coster and Rademakers, 2023).

The filtered reads were mapped to the relevant reference sequences using minimap2 (v.2.26-r1175) (Li, 2018). The reference sequences included the rAAV9 genome, *rep* and *cap* sequences (from pAAV2/9n: Plasmid #112865, Addgene), adenovirus helper genes sequences (from pAdDelta F6: Plasmid #112867, Addgene), transfer plasmid backbone, the human genome (NCBI RefSeq assembly: GCF_000001405.40(GRCh38.p14)), and the *Saccharomyces cerevisiae* genome (NCBI RefSeq assembly: GCF_000146045.2). The *Saccharomyces cerevisiae* genome was included due to the addition of yeast tRNA during the DNA extraction process

using the PureLink Viral RNA/DNA mini kit. For the remainder of this chapter, all references to yeast specifically refer to *Saccharomyces cerevisiae*. SAMtools (v.1.19.2) was used to generate mapping metrics, i.e. read length distribution, depth, and coverage, as well as to select reads based on specific genome coordinates for genome integrity assessment (Danecek *et al.*, 2021). Single-nucleotide variation (SNV) analysis and structural variation (SV) analysis were performed using Medaka (v.1.11.3) (ONT, <https://github.com/nanoporetech/medaka>) and Sniffles (v.2.4), respectively (Sedlazeck *et al.*, 2018; Smolka *et al.*, 2024). Medaka was also used for the generation of the consensus sequence for each sample. The mapped reads were visualised using the UGENE software (v.50, Unipro) (Okonechnikov *et al.*, 2012). Unmapped reads were further investigated and identified using NCBI's BLAST online tool (core nucleotide database (core_nt), <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Sayers *et al.*, 2022). Data gathering, editing, and analysis were performed on Microsoft 365 Excel and GraphPad Prism v.9.5.1.

4.3.9 Samples for dPCR

To assess the genome integrity of the extracted rAAV9 DNA samples from each experiment, dPCR analysis was performed on dilutions made from the 2 µL aliquots retained after each extraction method. Additionally, the remaining DNase I-treated aliquot (rAAV9 Direct Addition) from each experiment, which was not subjected to any extraction method and remained as intact particles, was analysed by direct addition to dPCR alongside the extracted samples to serve as a control of the extraction process.

4.3.10 Preparation of Samples for dPCR

The 2 µL rAAV9 aliquots taken after DNA extraction, including rAAV9 Heat, rAAV9 PureLink, and rAAV9 Direct Addition, were thawed and 10-fold serially diluted. The dilution process involved adding 2 µL of each rAAV9 sample or extraction blank to 18 µL of TE buffer (Thermo Fisher Scientific, Cat No. J75893.XCR) in an Eppendorf DNA LoBind tube, followed by successive 1:10 dilutions in which 10 µL of the diluted sample was mixed with 90 µL of TE buffer for a total dilution factor of 10^5 . The PureLink Extraction Blanks were diluted in three 1:10 steps, resulting in a total dilution factor of 10^3 , to prevent excessive dilution and ensure any potential contamination could still be detected.

4.3.11 Preparation of dPCR Reactions

For each run, the samples collected, including the rAAV9 Heat, rAAV9 PureLink, rAAV9 Direct Addition, and PureLink Extraction Blank, were run in triplicate in each digital PCR experiment. For dPCR analysis, the 4-plex B assay developed in Chapter 3 was used on the QIAcuity Four Digital PCR System (QIAGEN). This assay includes TaqMan probes targeting the ITRs, *RTN3*, *eGFP*, and WPRE regions of the recombinant AAV9 genome. Assay primer and probe sequences are listed in **Table 3.1**. This process was performed for each individual experiment, with each sample analysed in triplicate. The preparation of the dPCR reaction, the thermal cycling, and data acquisition were performed as described in Sections 3.3.7 and 3.3.8.

4.3.12 Digital PCR Data Processing and Analysis

Data analysis and calculation of the concentration of each sample were conducted as described previously in Sections 3.3.8 and 3.3.9.

4.3.13 Measurement of Genome Integrity by dPCR

The percentage of full genomes in each sample, as analysed by the 4-plex B dPCR assay, was determined using the DNA Integrity calculation software (https://digpqr.shinyapps.io/intactness_dsi_assay/) (Ghent University Digital PCR Centre) (Gleerup *et al.*, 2023). This model accounts for the probability of multiple targets being randomly co-located in the same partition. The model was previously used in Section 3.4.10 to estimate the genome integrity of rAAV samples.

4.4 Results

4.4.1 Library Quality Control

To ensure optimal sequencing library quality, sizing and quantitative analysis of the extracted samples and libraries were performed using the D5000 ScreenTape assay on the Agilent TapeStation 4150 system. Analysis of the extracts revealed the presence of a band at the expected size (~4.5 kb) and absence of contamination in the PureLink Extraction Blanks (**Figure 4.2**). TapeStation data on the pooled library (**Figure 4.2**) confirmed the presence of a band within the expected range (4.5-5 kb), as the pooled library is expected to consist of the 4.4 kb length of the recombinant AAV9 genome plus barcodes and sequencing adapters. No short molecules were visible. Note that TapeStation has sizing precision with a CV of 5% and

a sizing accuracy of $\pm 10\%$, which contributes to minor variations in the size of the detected bands (www.agilent.com/cs/library/usermanuals/public/D5000_QuickGuide.pdf, last accessed on 12.03.2025). The amount of library (fmol and viral genome copies) loaded per sequencing run was estimated based on the measured concentration (ng/ μ L) determined by TapeStation and the volume of the library loaded for sequencing (**Table 4.3**).

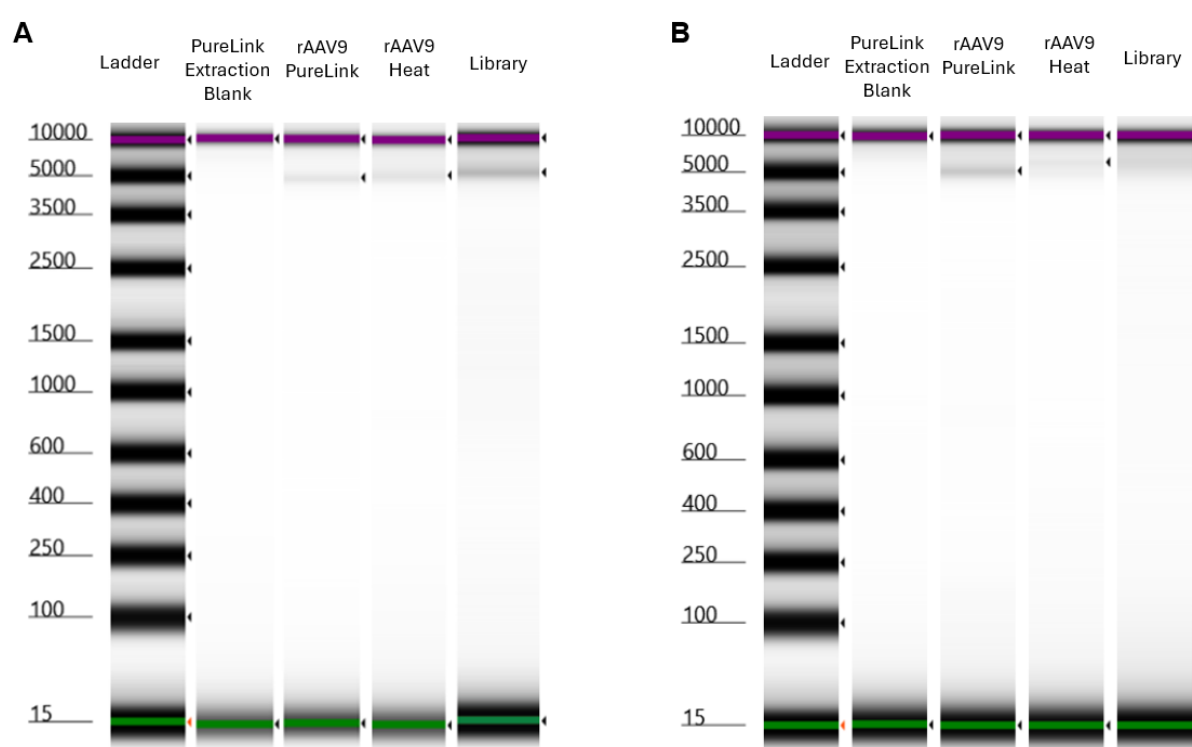


Figure 4.2 TapeStation analysis of extracts and sequencing libraries. The TapeStation electropherograms of extracts and libraries from sequencing runs 1 (A) and 2 (B) are shown as an example. The bands at 5 kb indicate the presence of full-length rAAV9 genomes. No bands, besides the quality control markers, were observed in the PureLink Extraction Blanks of either run.

Table 4.3 Concentration of sequencing libraries measured by TapeStation and amount of library loaded on each sequencing run.

Run No.	rAAV9 Library Conc. (ng/ μ L)	Library (fmol)	Library (viral genomes)
1	0.637	2.7	1.6×10^9
2	0.579	2.4	1.46×10^9
3	0.819	1.7	1.03×10^9
4	0.279	2.4	1.41×10^9

4.4.2 Yield, Quality and Read Length Metrics

The yield and quality metrics of each sample sequenced were generated using NanoPlot and are shown in **Table 4.4**. The data generated demonstrates that the reads of run 1 had considerably lower quality when compared to those of the remaining runs. Samples in run 1 had the lowest mean read quality (Q) pre-filtering (8.8 in run 1 vs 13.0 in run 2, 12.4 in run 3, and 11.8 in run 4) and the lowest % of reads exceeding a quality threshold of 10 (35.7 – 42.5 % in run 1 vs 94.8 – 95.1% in run 2, 93.5% in run 3, and 100% in run 4). Despite the lower read quality metrics in run 1, the number of reads passing the quality-based filtering threshold (Q10) was sufficient (> 35,000) for reliable downstream analysis. The rAAV9 PureLink samples had lower quality scores than the rAAV9 Heat samples both before (8.7 vs 8.9 (run 1), 12.9 vs 13.2 (run 2), 11.8 vs 12.4 (runs 3 and 4) and after filtering (10.6 vs 10.6 (run 1), 13.5 vs 13.9 (run 2), 12.5 vs 13.0 (runs 3 and 4) (**Table 4.4**).

Table 4.4 Read metrics of nanopore sequenced samples.

Run No.	Sample	Mean Read Quality Before filtering	Mean Read Quality After filtering	Total Reads	Filtered Reads	% of Reads > Q10
1	rAAV9 PureLink 1	8.7	10.6	136,744	48,627	35.7
	rAAV9 Heat 1	8.9	10.6	86,730	36,569	42.5
2	rAAV9 PureLink 2	12.9	13.5	160,614	151,179	94.8
	rAAV9 Heat 2	13.2	13.9	59,784	56,060	95.1
3	rAAV9 Heat 3	12.4	13.0	82,041	74,442	93.5
4	rAAV9 PureLink 4	11.8	12.5	118,367	106,787	100.0

Adequate yield was achieved for all samples across sequencing runs ($78,944 \pm 43,016$ filtered reads). In runs 1 and 2, where the rAAV9 Heat and rAAV9 PureLink samples were prepared and pooled together in equal volumes for sequencing, the rAAV9 PureLink samples generated a higher number of reads than the rAAV9 Heat samples (136,744 vs 86,739 before filtering, 48,627 vs 36,569 after filtering in run 1). This difference was particularly evident in run 2, where the rAAV9 PureLink 2 sample yielded approximately 2.7 times more reads than its heat-treated counterpart (160,614 vs 59,784 before filtering, 151,179 vs 56,060 after filtering). When analysing the read length metrics (**Table 4.5**), the rAAV9 Heat samples consistently exhibited a higher mean read length than the rAAV9 PureLink samples, both before ($2,165 \pm 196$ nt vs $1,314 \pm 353$ nt) and after filtering ($2,300 \pm 185$ nt vs $1,744 \pm 506$ nt), across all runs, suggesting higher genome integrity.

Table 4.5 Read length metrics of analysed samples using nanopore sequencing.

Run No.	Sample	Mean Read Length Before Filtering	Mean Read Length After Filtering
1	rAAV9 PureLink 1	1,599	2,102
	rAAV9 Heat 1	1,939	2,501
2	rAAV9 PureLink 2	1,423	1,386
	rAAV9 Heat 2	2,290	2,262
3	rAAV9 Heat 3	2,265	2,137
4	rAAV9 PureLink 4	919	933

4.4.3 Sequencing Depth and Read Length Distribution

After quality-based filtering ($Q > 10$), reads were mapped to the respective recombinant AAV9 reference sequence and analysed using SAMtools to assess sequencing depth. As shown in **Table 4.6**, appropriate sequencing depth for all samples was achieved (mean $> 19,000$), allowing for adequate evaluation of genome integrity and variant calling analysis. Notably, sequencing depth variation was higher in runs 2, 3, and 4 when compared to run 1 for both

rAAV9 Heat (CV of 11% in run 1 vs 27% in run 2, and 30% in run 3) and rAAV9 PureLink samples (CV of 13% in run 1 vs 47% in run 2, and 56% in run 4).

Both extraction methods investigated exhibited a similar sequencing depth profile across the genome, with higher read coverage at the start and end, corresponding to the ITRs (**Figure 4.3**). This trend was consistent across all runs but became more pronounced in runs 2, 3, and 4, where the difference in sequencing depth within the ITRs region compared to the rest of the genome was more substantial. Specifically, the genomic position within the ITRs that showed the highest sequencing depth had 2.9 times greater depth than the lowest-depth position outside the ITRs in run 1, 4.75 times greater in run 2, and 5.3 times greater in runs 3 and 4.

All samples contained reads with 100% genomic coverage. Yet, a sequencing depth increase was observed in positions corresponding to the regions where the hairpin structures of the ITRs resolve, i.e. after nucleotide 68 for 5' ITR and before nucleotide 4,317 for 3' ITR. This effect was most pronounced in run 1 (19.9% increase in rAAV9 Heat, 14.4% increase in rAAV9 PureLink) and to a lesser extent in runs 2 (7.5% increase in rAAV9 Heat, 9.9% increase in rAAV9 PureLink), 3 (14.0% increase in rAAV9 Heat), and 4 (11.7% increase in rAAV9 PureLink).

Table 4.6 Sequencing depth metrics of analysed samples using nanopore sequencing.

Sample	Mean Sequencing Depth	Maximum Sequencing Depth	Minimum Sequencing Depth
rAAV9 Heat 1	19,370 ± 2,081	25,116	8,930
rAAV9 PureLink 1	20,415 ± 2,744	28,670	9,561
rAAV9 Heat 2	26,466 ± 7,085	51,541	14,270
rAAV9 PureLink 2	38,200 ± 17,797	97,453	17,061
rAAV9 Heat 3	32,756 ± 9,932	64,496	16,331
rAAV9 PureLink 4	27,528 ± 15,359	77,818	10,594

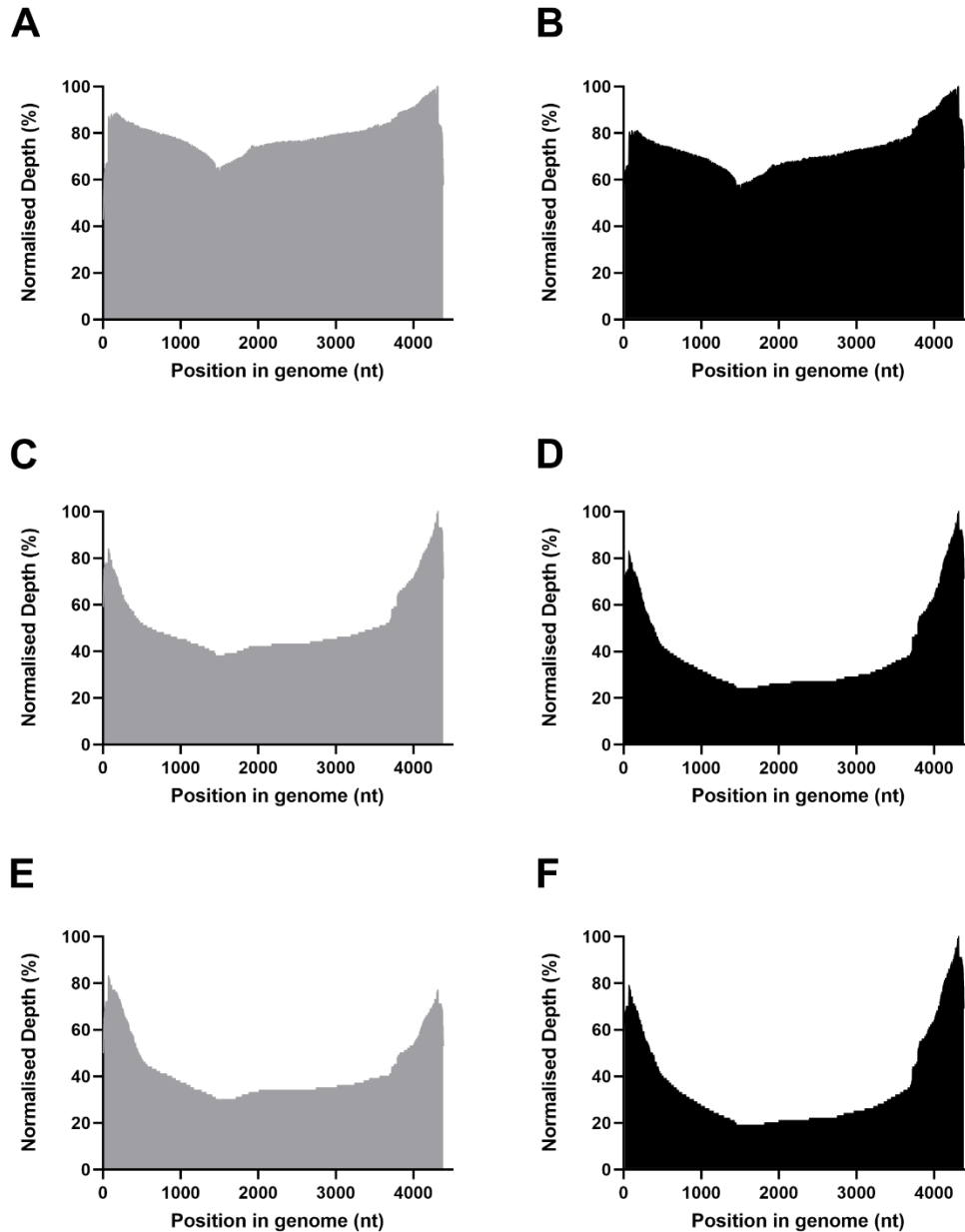


Figure 4.3 Sequencing depth analysis of rAAV9 Heat and rAAV9 PureLink samples. The y-axis represents the percentage of sequencing depth at each nucleotide position along the x-axis, with values normalised to the genomic position with the highest depth in each sample. (A) rAAV9 Heat sample from run 1, (B) rAAV9 PureLink sample from run 1, (C) rAAV9 Heat sample from run 2, (D) rAAV9 PureLink sample from run 2, (E) rAAV9 Heat sample from run 3, and (F) rAAV9 PureLink sample from run 4.

The mapped reads of the rAAV9 Heat and rAAV9 PureLink samples were further analysed by SAMtools to generate the read length distribution plots for the samples across all runs (**Figure 4.4**). In run 1, the read length distribution profile of PureLink and heat-treated samples was comparable. Samples exhibited a wide range of lower-than-expected fragments,

but a prominent peak at 4.4 kb was observed corresponding to the full-length rAAV9 genome. Run 1 exhibited a lower proportion of < 2 kb reads than runs 2, 3, and 4 for both rAAV9 Heat (46.1% in run 1 vs 54.0 % in run 2, 56.5% in run 3) and rAAV9 PureLink (40.2% in run 1 vs 66.1% in run 2, 70% in run 4) samples. Thus, it was observed that these short reads were particularly pronounced in the rAAV9 PureLink samples of runs 2 and 4. The reads corresponding to peaks at 600, 850, and 1200 nt (**Figure 4.4**) were filtered according to their length and the mapped reads were visualised. Mapping data analysis showed that these reads corresponded to incomplete rAAV9 genomes, including either the 5' or 3' ITRs, explaining the much higher sequencing depth of the molecules' 5' and 3' ends (**Figure 4.3**). This data suggests the presence of fragmented genomes and a higher genome integrity in rAAV9 Heat samples than in the PureLink extracted samples. In addition, a peak at approximately 4 kb was consistently observed across all runs.

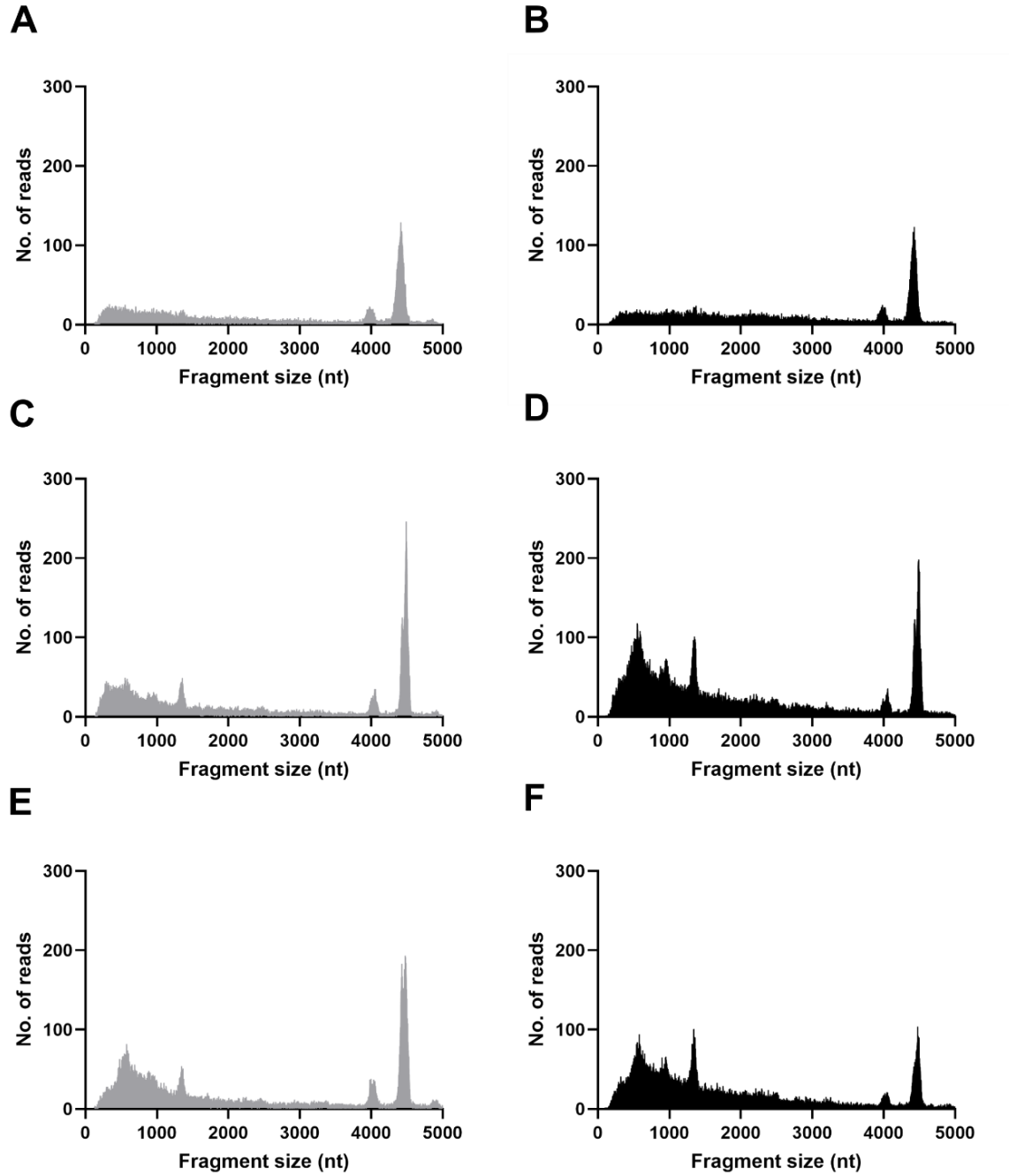


Figure 4.4 Read length distributions of rAAV9 Heat and rAAV9 PureLink extracted genomes across all runs. The x-axis represents fragment size, while the y-axis indicates the number of reads corresponding to each fragment size detected during sequencing. Read length distribution for (A) rAAV9 Heat – run 1, (B) rAAV9 PureLink – run 1, (C) rAAV9 Heat – run 2, (D) rAAV9 PureLink – run 2, (E) rAAV9 Heat – run 3, and (F) rAAV9 PureLink – run 4.

4.4.4 Variant Calling Analysis

Single-nucleotide variation (SNV) analysis was performed with Medaka. No substitutions or indels were detected when considering a quality score threshold of 10, indicating that the rAAV9 molecule had the predicted nucleotide sequence.

Analysis of structural rearrangements/variations in the rAAV9 genome using Sniffles (v.2.4) revealed a large deletion in a subpopulation of all samples across all sequencing runs. This deletion was detected with high confidence, with a quality score of 59, though its exact coordinates varied between sequencing runs (**Table 4.7**). In run 1, Sniffles identified a large deletion between nucleotides 1,446 and 1,882 in both samples, whereas in run 2, this SV was detected between nucleotides 1,091 and 1,529. In runs 3 and 4, the SV was called for genomic region 2,225-2,663 and 1,448-1,886, respectively. The identified structural rearrangements were further validated manually using the UGENE software to visualise the mapped reads. Although the deletions detected by Sniffles in runs 1, 2, and 4 fall within the CamKII α promoter sequence (169–1,904 nt), the deletion detected in run 3 falls within the *RTN3* gene region (1,905–2,612).

Table 4.7 Detail on % of reads containing a deletion in the rAAV9 genome.

Run No.	Sample	Reads with Deletion (%)	Quality Score
1	rAAV9 PureLink 1	4.6	59
	rAAV9 Heat 1	4.6	59
2	rAAV9 PureLink 2	2.8	59
	rAAV9 Heat 2	4.7	59
3	rAAV9 Heat 3	3.9	59
4	rAAV9 PureLink 4	2.4	59

It was observed that in all four sequencing runs, a small % of reads contained a deletion spanning the 1,446- 1,882 region while others had a deletion spanning the 1,091-1,529 region (**Figure 4.5**). Deletions in the *RTN3* region (2,225 to 2,663) were also observed in other reads. This indicates that different types of deletion coexist within different molecules of the same sample, but Sniffles reports only one set of coordinates per run. Further inspection of the

rAAV9 genome sequence revealed that regions 1,104 to 1,468 nt and 1,541 to 1,904 nt are identical. Therefore, this 364 nt sequence is present twice in the rAAV9 genome (**Figure 4.6**). These incomplete molecules represented $3.8\% \pm 1.0\%$ of the reads across all sequencing runs and correspond to the 4 kb peak observed previously (**Figure 4.4**).

A



B

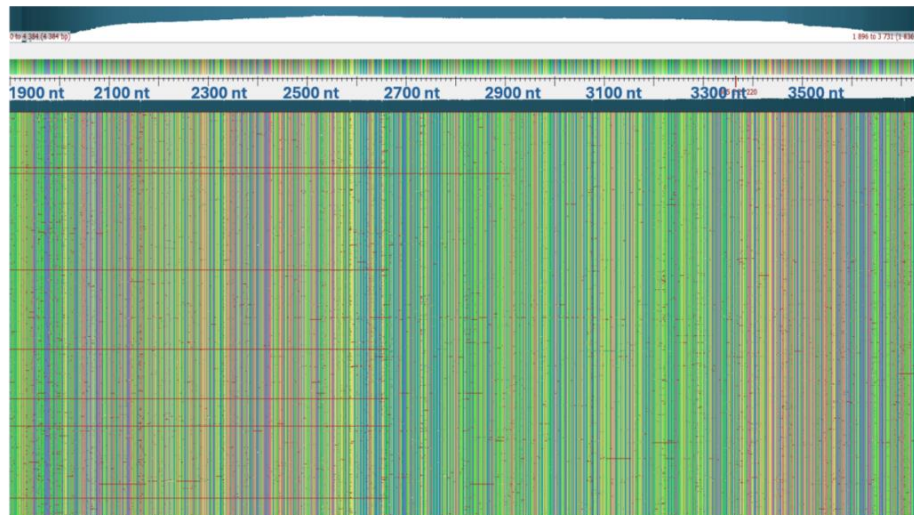


Figure 4.5 Deletions observed in reads of the rAAV9 genome. UGENE output showing a subset of reads from the rAAV9 Heat 1 sample, highlighting deletions depicted as horizontal red lines. (A) The nucleotide positions (1,100 – 1,900 nt) are marked in blue along the top. Deletions were observed at varying positions within the region, with some starting as early as 1,100 nt and extending to approximately 1,500 nt, while others begin around 1,500 nt and terminate near 1,900 nt. Additionally, larger deletions spanning the entire region from 1,100 to 1,900 nt were detected. (B) The nucleotide positions (1,900 – 3,500 nt) are marked in blue along the top. Deletions were observed within the region 1,900 – 2,700 nt.

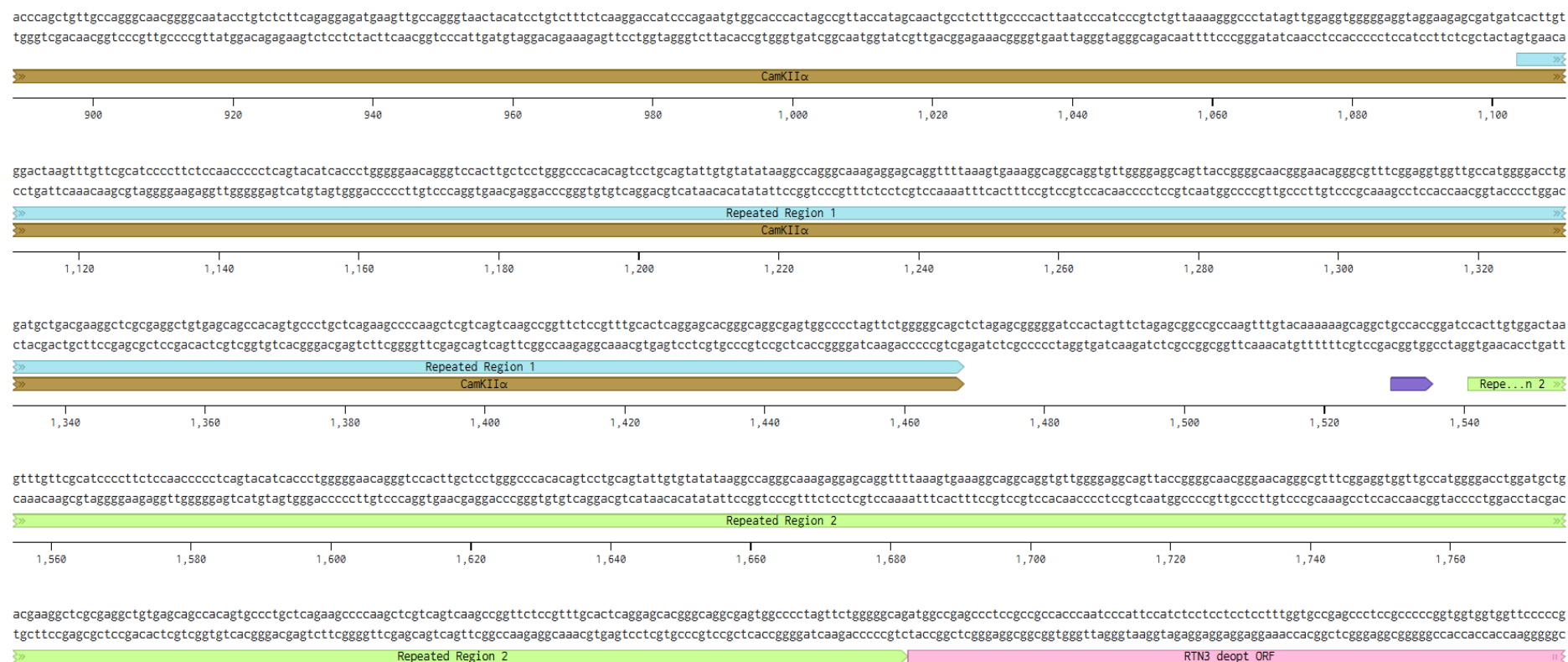


Figure 4.6 Annotated CamKIIα promoter region (893 – 1904 nt) of the rAAV9 genome. Repeated region 1 (1,104-1,468) is shown in light blue, while repeated region 2 (1,541-1,904) is shown in green. These two regions share an identical sequence, making them prone to recombination events that can result in deletions. The start of the open reading frame of disoptimised RTN3 is shown in pink. The end of the CamKIIα promoter is shown in brown.

4.4.5 Chimeric Backbone/rAAV Molecule Analysis

As mentioned in Section 4.4.3, a large portion of the rAAV9 reads are incomplete rAAV9 genomes. Interestingly, these reads often contained both plasmid backbone and rAAV ITR sequences. These chimeric reads (**Figure 4.7**) were identified in all sequencing runs but were more representative in sequencing runs 2 to 4 ($\geq 4.3\%$) than in run 1 ($\leq 1.5\%$). The proportion of chimeric reads was very similar between rAAV9 PureLink and rAAV9 Heat samples across all runs (1.46 vs 1.20% in run 1, 4.39 vs 5.28% in run 2, 4.45 vs 5.33% in runs 3 and 4 for rAAV9 PureLink and rAAV9 Heat, respectively).

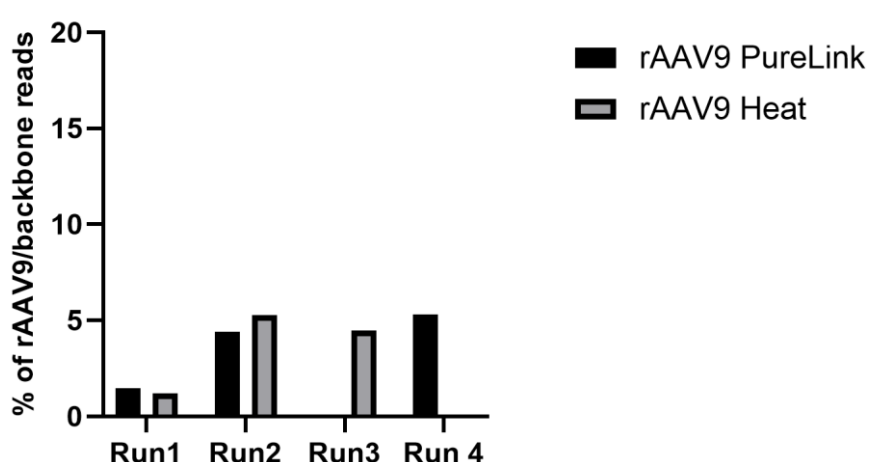


Figure 4.7 Detection of plasmid backbone DNA in rAAV9 mapped reads. The plot shows the percentage of mapped reads that align to both the rAAV9 genome and the transfer plasmid backbone across all four runs. rAAV9 PureLink samples are shown with black bars, while rAAV9 Heat samples are shown with grey bars.

4.4.6 Genome Integrity Assessment from the Nanopore Sequencing Data

The relative quantification of genome integrity by nanopore sequencing was determined considering as full-length any reads that had 100% rAAV9 genome coverage (1-4,386) as well as reads spanning from 68 to 4,317 nt, whereby the ITR regions were not considered (**Figure 4.8**). Analysis revealed 1.8 times higher genome integrity when excluding ITR hairpin regions in rAAV9 Heat and 2.1 times higher in rAAV9 PureLink samples, demonstrating the presence of a substantial number of reads with incomplete ITR regions. A paired t-test was conducted to compare the two analysis approaches (including ITR hairpin vs excluding) across all analysed samples. A highly significant difference was observed ($p < 0.01$) between the two approaches.

Data showed higher genome integrity in rAAV9 Heat samples when compared to the rAAV9 PureLink samples, regardless of the analysis method used. However, a paired t-test showed that the difference between them was not significant ($p > 0.05$) with either analysis approach. When reads not containing the hairpin ITR structures were excluded from the analysis, the percentage of full-length reads was $29.8 \pm 1.8\%$ for rAAV9 Heat against $17.4 \pm 9.5\%$ for rAAV9 PureLink. In contrast, when the ITR hairpin structures were included, the percentage of full-length reads was $16.7 \pm 5.1\%$ for rAAV9 Heat and $8.4 \pm 3.2\%$ for rAAV9 PureLink. The variation across experiments was assessed for both approaches. For rAAV9 PureLink samples, the variation was 38.3% when the ITR hairpin was included and 54.5% when excluded. In contrast, for rAAV9 Heat samples, the variation was 30.3% with the ITR hairpin included and substantially lower at 6.1% when excluded.

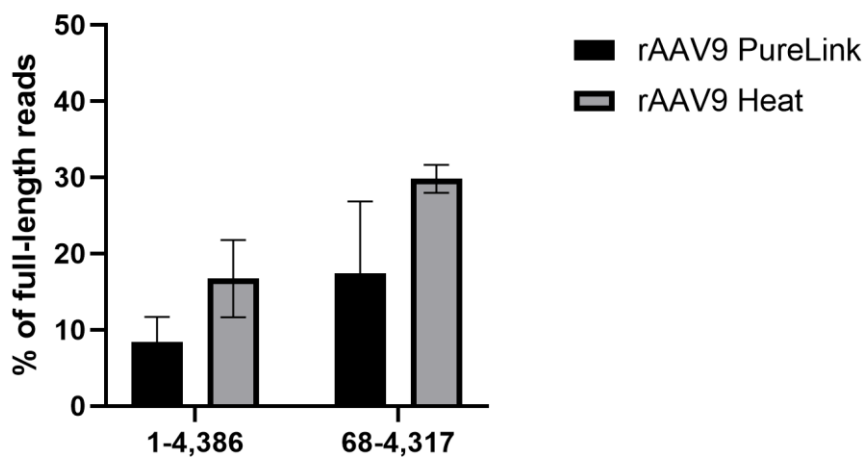


Figure 4.8 Measurement of rAAV9 genome integrity by nanopore sequencing. Genome integrity expressed as percentage of full-length rAAV9-mapped reads (mean $\pm$ SD) against total number of mapped reads in rAAV9 PureLink (black) and rAAV9 Heat (grey) samples after three sequencing experiments. Reads with previously identified SVs were excluded. 1-4386 and 68-4,317 correspond to all reads starting at position 1 and ending at 4386 nt and all reads containing the sequence spanning from 68 nt to 4,317 nt, respectively.

4.4.7 Genome Integrity Analysis using dPCR

The rAAV9 Heat, rAAV9 PureLink, and rAAV9 Direct Addition samples that were collected in each run (run 1, run 2, run 3/4) were analysed using the previously developed 4-plex B digital PCR assay after DNA extraction. Three independent experiments were conducted, one per run, each with three replicates per sample. The 4-plex B assay targeted regions within the ITRs, *RTN3*, *eGFP*, and WPRE to determine genome integrity. **Figure 4.9A** illustrates the

concentration in each template unit analysed across the three runs. By averaging the individual assay results for each run, the mean concentrations were determined as follows: for rAAV9 Direct Addition, the concentrations were 542 copies/ μ L in run 1, 1006 copies/ μ L in run 2, and 1335 copies/ μ L in run 3/4. For rAAV9 Heat samples, the concentrations were 337 copies/ μ L in run 1, 271 copies/ μ L in run 2, and 2859 copies/ μ L in run 3/4. For rAAV9 PureLink samples, the concentrations were 8,360 copies/ μ L in run 1, 600 copies/ μ L in run 2, and 6311 copies/ μ L in run 3/4. The reported inter-experimental variability was the highest in the rAAV9 Heat samples (CV of 128% vs 79% for AAV9 PureLink, and 41% for AAV9 Direct Addition).

The samples were then assessed for genome integrity in each run. The genome integrity is reported as percentage of full-length genomes. The rAAV9 Direct Addition sample exhibited genome integrity values of 70.1%, 40.2%, and 83.5% across runs 1, 2, and 3/4, respectively. The rAAV9 Heat sample had values of 59.1%, 53.3%, and 80.1%, while the rAAV9 PureLink sample demonstrated genome integrity of 71.1%, 43.6%, and 77.7% when assessed using this approach (**Figure 4.9B**). Across the runs, the rAAV9 PureLink, rAAV9 Heat, and Direct Addition samples exhibited similar genome integrity when assessed by dPCR ($64.1 \pm 18.0\%$ for rAAV9 PureLink, $64.2 \pm 14.0\%$ for rAAV9 Heat, and $64.6 \pm 22.6\%$ for rAAV9 Direct Addition). The reported variability for genome integrity estimation was comparable between the samples, but highest in the rAAV9 Direct Addition sample (CV of 34.3% vs 21.9% for rAAV9 Heat, and 28.1% for rAAV9 PureLink).

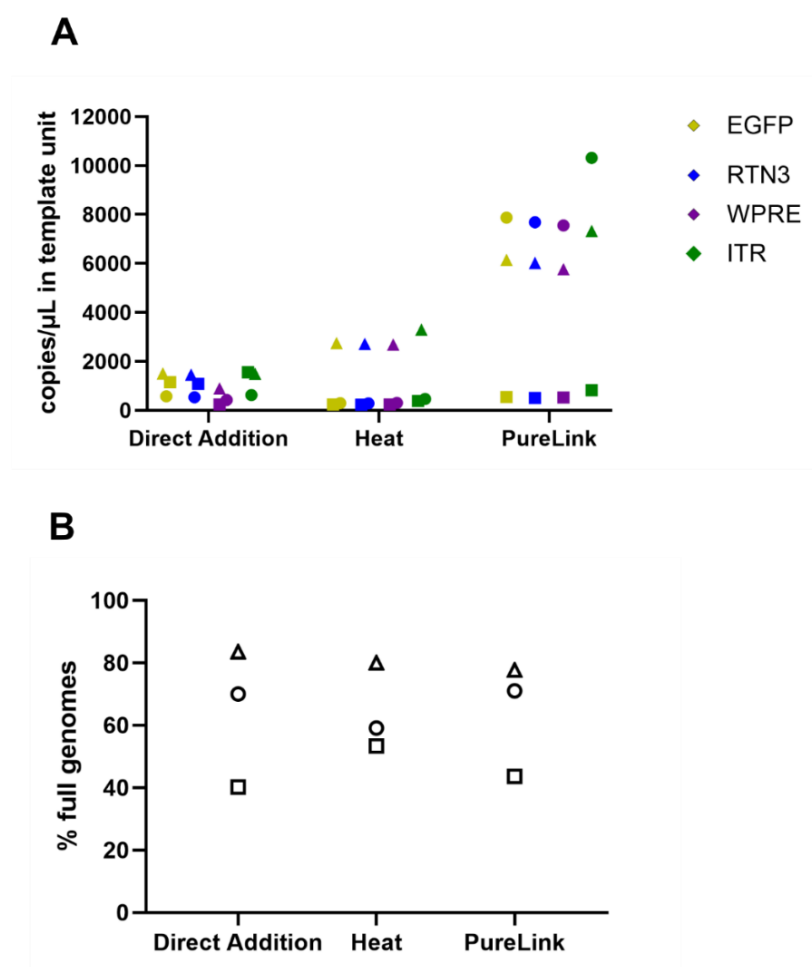


Figure 4.9 Comparison of rAAV9 genome integrity and concentration across extraction methods when assessed using a 4-plex QIAcuity dPCR approach. Each symbol represents a different experiment: run 1 (circles), run 2 (squares), and run 3/4 (triangles). (A) Template concentration, expressed as copies/μL. (B) Genome integrity estimation, presented as percentage of full genomes.

4.4.8 Comparison of Genome Integrity as Determined by dPCR or Nanopore Sequencing

The percentage of full genomes determined by dPCR for rAAV9 Heat and rAAV9 PureLink samples across three experiments was then compared to the genome integrity percentage obtained from nanopore sequencing. For the measurement of genome integrity by nanopore sequencing, full genomes were defined as reads spanning nucleotides 68 to 4,317 nt. The percentage of full genomes reported by dPCR was significantly higher ($p < 0.01$) than that measured by nanopore sequencing, as determined by paired t-test across all analysed samples (**Figure 4.10A**). Reanalysis of the nanopore NGS data to define full genomes as reads containing all four dPCR target regions (dPCR amplicons) increased the measured genome integrity to $28.2 \pm 16\%$ for rAAV9 PureLink and $39.6 \pm 5.8\%$ for rAAV9 Heat samples (**Figure**

4.10B). However, despite this adjustment, dPCR still reported significantly higher genome integrity for both samples ($p < 0.05$) as determined by a paired t-test.

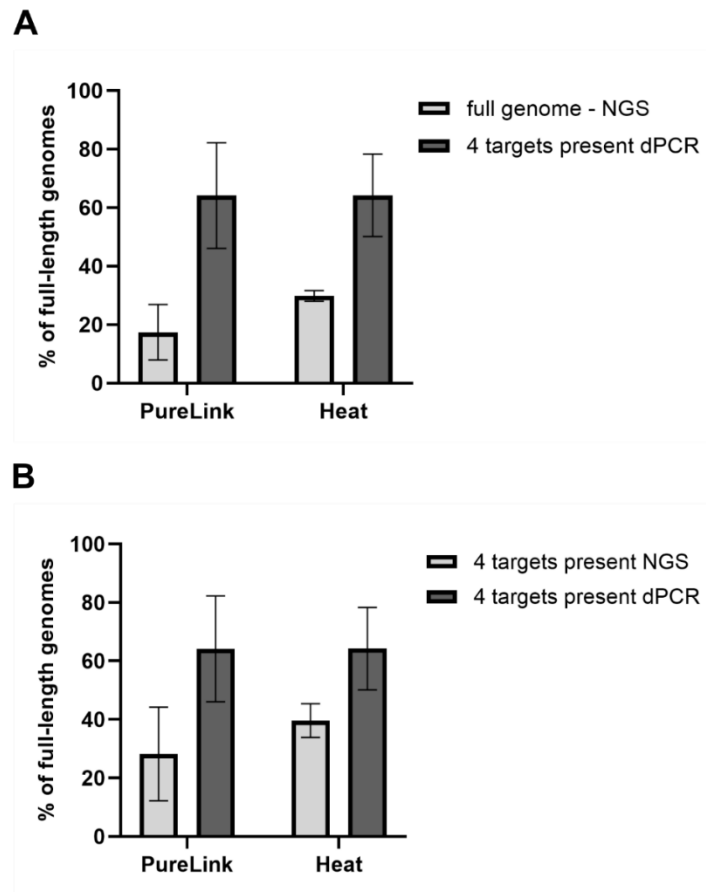


Figure 4.10 Comparison of rAAV9 genome integrity as determined by QIAcuity dPCR and nanopore sequencing. The y-axis represents the percentage of full genomes for each method: for nanopore sequencing, this corresponds to the percentage of rAAV9-mapped reads containing full genomes, while for dPCR it represents the percentage of genomes containing all four target regions using the 4-plex B assay. (A) Genome integrity measured by nanopore sequencing, where full genomes are defined as reads spanning nucleotides 68 to 4,317. (B) Genome integrity measured by nanopore sequencing, where full genomes are defined as reads containing all four dPCR target regions, specifically the ITRs, *RTN3*, *eGFP*, and WPRE.

4.4.9 rAAV genome Purity Assessment using the Nanopore Sequencing Data

Unmapped reads (i.e. readings not mapping to the rAAV9 genome) were mapped against plasmids used in rAAV9 production (pAAV2/9n, pAdDelta F6, transfer plasmid backbone), the human genome (host cell DNA), and yeast (relevant due to the PureLink extraction process) to detect potential carry-over (**Figure 4.11**). In rAAV9 Heat samples, a low number of unmapped reads ($3.4 \pm 2.5\%$) were identified as host cell DNA, and $0.42 \pm 0.24\%$ as plasmid-

only sequences. In the rAAV9 PureLink samples, a comparatively higher proportion of reads ($9.2 \pm 0.8\%$) mapped to host cell DNA, whereas the percentage of plasmid-only sequences was comparable ($0.56 \pm 0.33\%$). Note that the percentage of reads containing plasmid sequence would be higher if the previously described chimeric molecules with plasmid backbone were also considered ($4.2 \pm 2.3\%$ for rAAV9 PureLink, $4.1 \pm 2.3\%$ for rAAV9 Heat). In the rAAV9 PureLink samples, a small percentage of reads ($2.5 \pm 0.5\%$) were identified as yeast (*Saccharomyces cerevisiae*), while $< 0.01\%$ of reads were mapped to yeast in the rAAV9 Heat samples. **Figure 4.12** illustrates that most of the DNA impurities found are shorter than 1,000 nt ($94.3 \pm 0.8\%$ for rAAV9 PureLink and $89.1 \pm 3.6\%$ for rAAV9 Heat), indicating that they primarily consist of small, fragmented molecules in both rAAV9 PureLink and rAAV9 Heat samples. However, $96.5 \pm 0.5\%$ of the peaks for rAAV9 PureLink and $94.3 \pm 6.7\%$ for the rAAV9 Heat corresponded to fragments exceeding 200 nt.

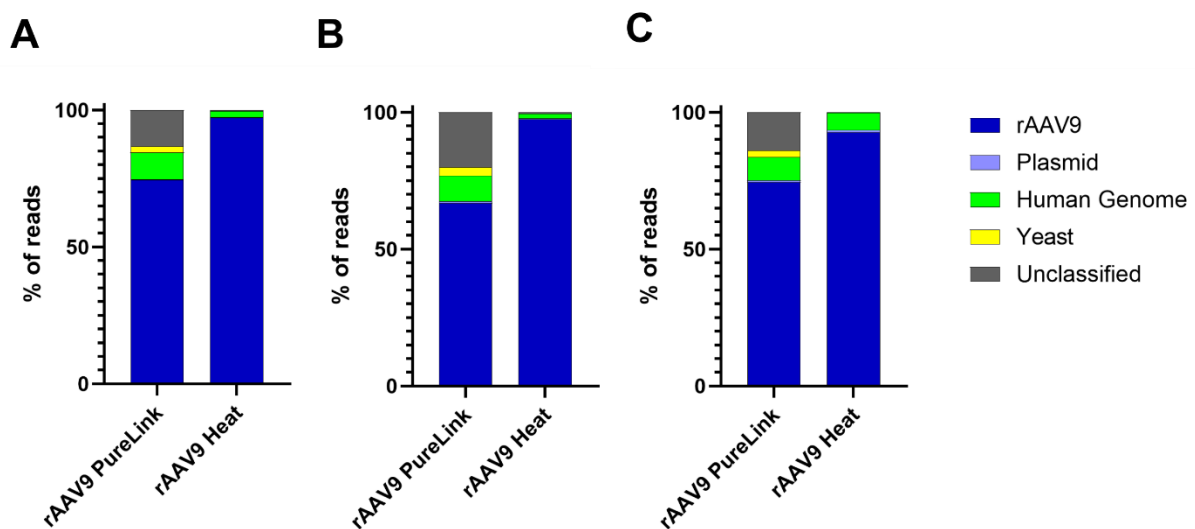


Figure 4.11 Nanopore sequencing reads mapped to rAAV9 and other sources for rAAV9 PureLink and rAAV9 Heat generated samples. The figure shows the percentage of total reads mapped to different sources: rAAV9 (dark blue), transfected plasmids including, pAAV2/9n, pAdDelta F6, and the backbone of the transfer plasmid (light blue), the human genome (green), yeast (*Saccharomyces cerevisiae*) (yellow), and unclassified reads that did not align to any of the above sources (grey). Three panels are shown: (A) samples from run 1, (B) samples from run 2, and (C) rAAV9 Heat from run 3 and rAAV9 PureLink from run 4.

In the rAAV9 PureLink sample, there was a considerable number of unclassified reads i.e. reads not mapping to any of the sources investigated. In the rAAV9 PureLink sample data, these represented $16 \pm 3.7\%$ of the total reads, while in the rAAV9 Heat sample data, the

percentage was much lower at $0.38 \pm 0.17\%$. To investigate the origin of these unclassified reads, the nucleotide sequences were analysed using BLAST. This analysis revealed that the unclassified reads aligned with sequences from various bacterial genera. The length of these DNA impurities was then also assessed. The read length distribution of unclassified reads in rAAV9 PureLink samples was separately analysed, revealing that the majority are shorter than 500 nt (**Figure 4.13**).

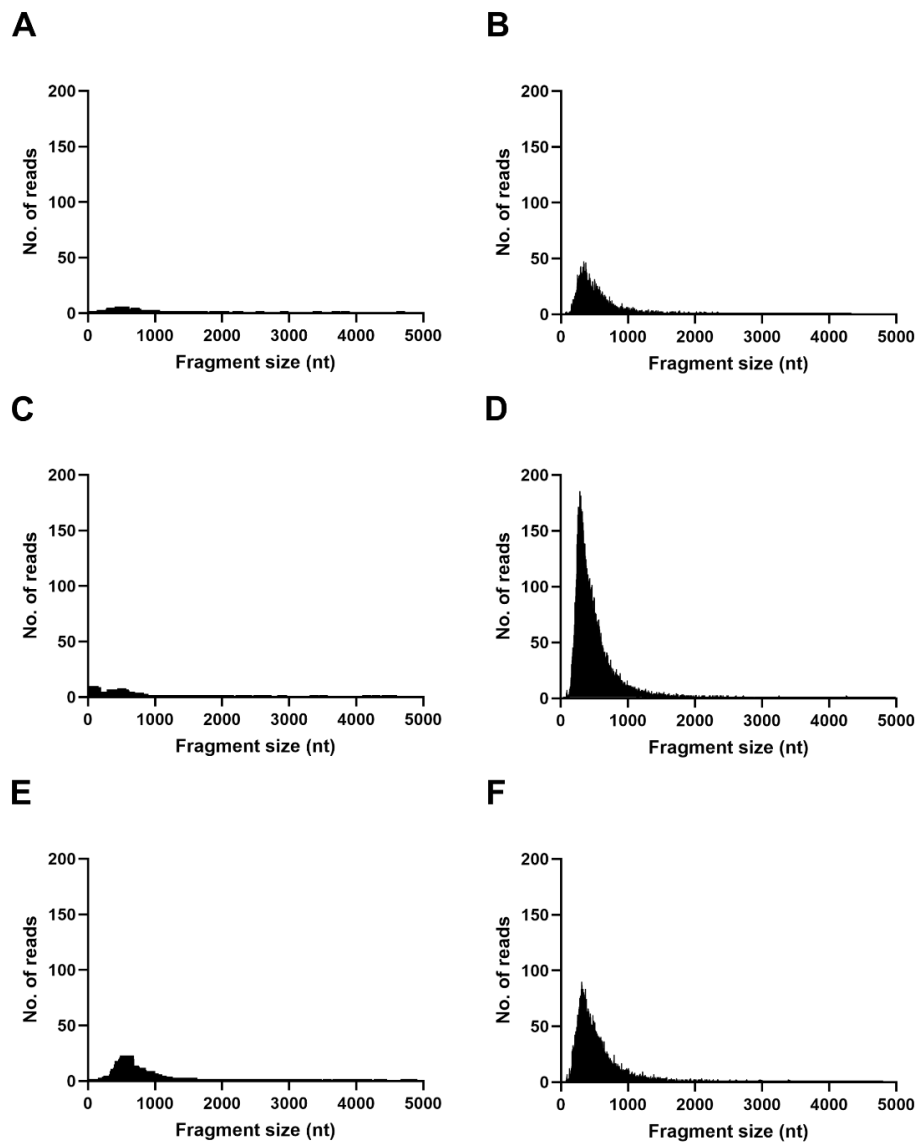


Figure 4.12 Read length distribution plots of DNA impurities found in rAAV9 Heat and rAAV9 PureLink nanopore sequencing data. The x-axis represents fragment size, while the y-axis indicates the number of reads corresponding to each fragment size detected during sequencing. (A) Read length distribution for rAAV9 Heat sample – run 1, (B) rAAV9 PureLink sample – run 1, (C) rAAV9 Heat sample – run 2, (D) rAAV9 PureLink sample – run 2, (E) rAAV9 Heat sample – run 3, (F) rAAV9 PureLink sample – run 4.

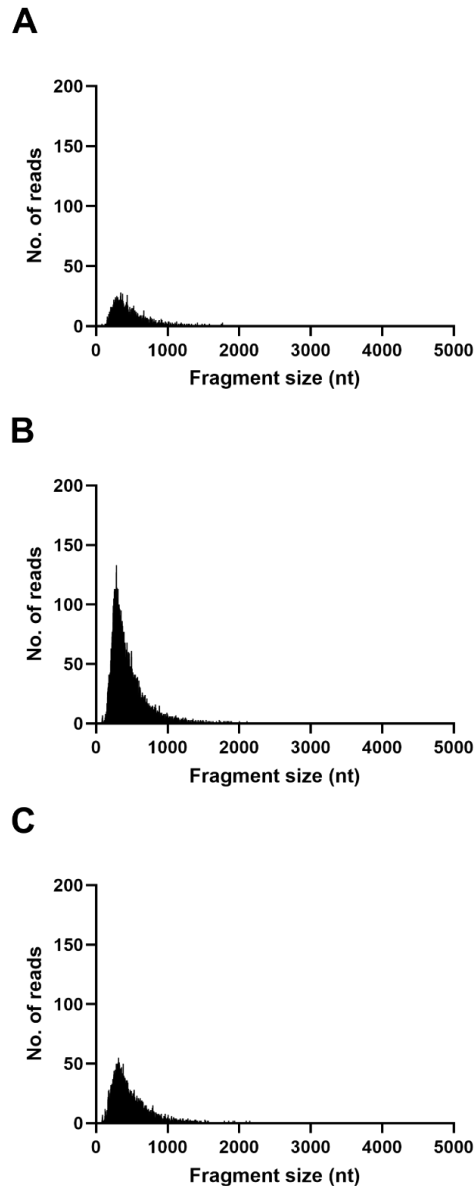


Figure 4.13 Read length distribution plots of unclassified reads found in rAAV9 PureLink samples. The x-axis represents fragment size, while the y-axis indicates the number of reads corresponding to each fragment size detected during sequencing. (A) Read length distribution for rAAV9 PureLink sample – run 1, (B) rAAV9 PureLink sample – run 2, (C) rAAV9 PureLink sample – run 4.

4.4.10 Investigating the PureLink Extraction Blank Yield and Read Identification Data

To assess potential cross-sample and reagent contamination, the data of the PureLink Extraction Blanks were processed. The percentage of filtered reads originating from the PureLink Extraction Blank relative to the total number of reads was lower in run 1 compared to runs 2, 3, and 4 (1.7% in run 1 vs. 11.4% in run 2, 15.4% in run 3, and 9.2% in run 4) (**Table**

4.8). The generated reads were mapped to the rAAV9 genome, pAAV2/9n, pAdDelta F6, transfer plasmid backbone, host cell DNA, and yeast (*Saccharomyces cerevisiae*) (**Figure 4.14**).

Table 4.8 Nanopore sequencing read metrics of extraction blanks.

Sample	Total Reads	Filtered Reads	% of Total Reads (Unfiltered)	% of Total Reads (Filtered)
PureLink Extraction Blank 1	3,957	1,440	1.7	1.7
PureLink Extraction Blank 2	27,937	26,621	11.2	11.4
PureLink Extraction Blank 3	14,548	13,566	17.7	15.4
PureLink Extraction Blank 4	11,704	10,792	9.0	9.2

A small percentage of PureLink Extraction Blank reads in runs 2 and 3 (< 7%) mapped to the rAAV9 genome. However, in sequencing run 1, 29% of the reads mapped to the rAAV9 genome. The presence of rAAV genome in the PureLink Extraction Blank can be attributed to the fact that both the rAAV9 PureLink sample and the PureLink Extraction Blank were processed in parallel. Due to the very high concentration of the rAAV9 sample ($\sim 7.0 \times 10^9$ viral genomes/ μL), it is possible that rAAV9 DNA was transferred to the blank, for example, through aerosol transfer during handling.

Only a minor fraction ($0.76 \pm 0.44\%$) aligned with the human genome across runs 1, 2, and 4, while for run 3, 3.78% of the reads mapped to the human genome. Of the remaining reads, $11.6 \pm 1.1\%$ mapped to yeast (*Saccharomyces cerevisiae*). A substantial proportion of the reads of the PureLink Extraction Blank ($75.2 \pm 9.8\%$) were unclassified and were mostly identified as various bacteria genera, as seen in the rAAV9 PureLink sample. Almost no reads (< 0.01%) were mapped to the production plasmids.

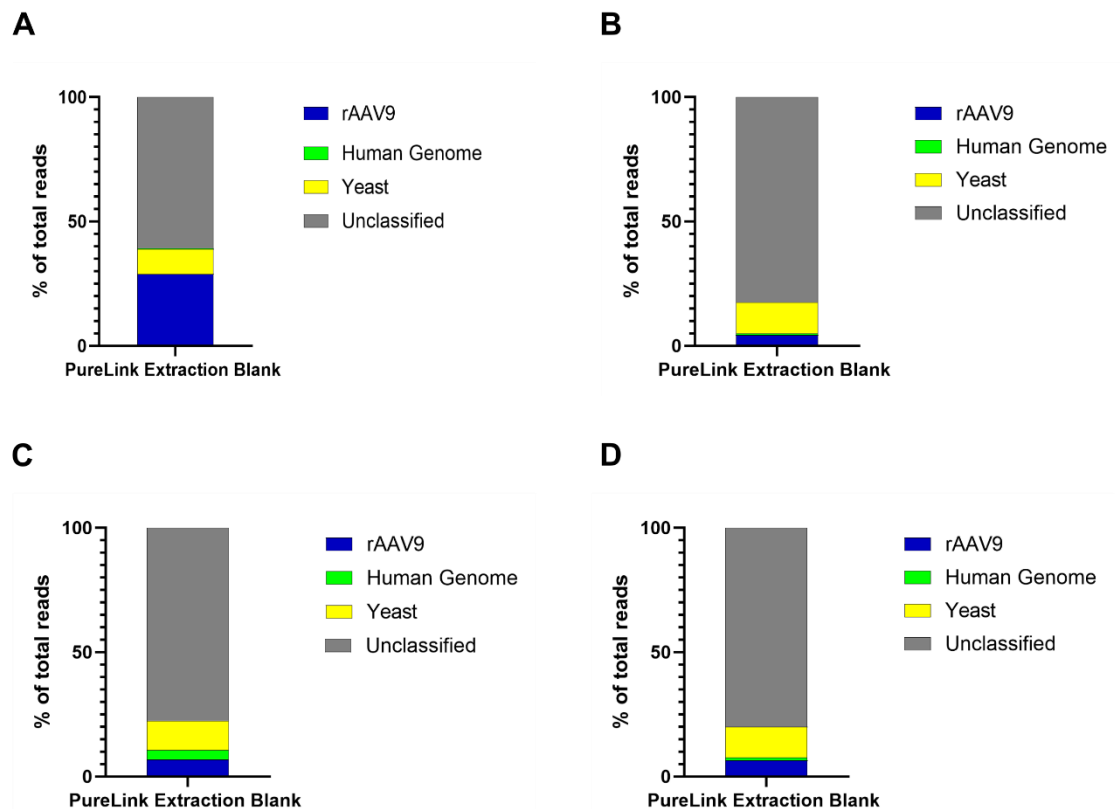


Figure 4.14 Nanopore sequencing mapped reads of PureLink Extraction Blanks. The figure shows the percentage of total reads mapped to different sources: rAAV9 genome (dark blue), the human genome (green), yeast (*Saccharomyces cerevisiae*) (yellow), and unclassified reads that did not align to any of the above sources (grey). Four panels are shown: (A) sample from run 1, (B) sample from run 2, (C) sample from run 3, and (D) sample from run 4.

4.5 Discussion

The work in this chapter describes the application of nanopore sequencing to investigate: (a) the genome integrity of a model rAAV9 genome extracted from capsids through two different methods, and (b) to analyse the DNA impurity profile of the rAAV9 samples. Multiplex digital PCR served as a complementary tool for genome integrity assessment and was compared to nanopore sequencing. There was a difference in the percentage of determined full genomes between these methods. The observed differences in the percentage of full genomes between the 4-plex dPCR and nanopore sequencing methods in all probability arise from fundamental differences in how these methods operate and their biases (**Figure 4.10**). It is hypothesised that dPCR may overestimate genome integrity, while nanopore sequencing is more likely to underestimate it, leading to discrepancies between the two approaches.

The dPCR method developed in this study targets four specific regions (ITR, *RTN3*, *eGFP*, and WPRE) spanning the full length of the genome. While this approach provides valuable information regarding genome integrity, it inherently limits the analysis to these defined regions. Therefore, any genome that includes all the targeted regions will be classified as full, even if other parts of the genome are missing.

Another factor contributing to the overestimation of genome integrity when using dPCR is the inability to distinguish between molecules with two ITRs (5' ITR and 3' ITR) and molecules with only one ITR. The 5' and 3' ITRs of AAV are inverted repeats, thus, the ITR assay detects both (Srivastava, Lusby and Berns, 1983; Aurnhammer *et al.*, 2012). As a result, if a genome contains the other three targets (*eGFP*, *RTN3*, WPRE) but only a single ITR, it will still be classified as a full genome with all four targets present, even though one ITR is missing (Suoranta, Laham-Karam and Ylä-Herttuala, 2021). This occurs because the partition will be detected as positive for the FAM signal despite the absence of one ITR. This inability to differentiate between genomes with both ITRs and genomes with only one ITR leads to an overestimation of genome integrity in dPCR measurements (Suoranta, Laham-Karam and Ylä-Herttuala, 2021).

Additionally, deletions detected by nanopore sequencing in $3.8 \pm 1.0\%$ of genomes were not accounted for in genome integrity measurements conducted by dPCR as the assay does not include targets that would capture these specific deletions. Thus, this limitation further contributes to the overestimation of genome integrity in dPCR measurements. In dPCR, assay design around this deletion would be challenging since the deletion varies in its coordinates between different genomes. Detecting all possible deletions would require a significant number of assays to capture each deletion variant, making the approach completely inefficient and impractical.

Nanopore sequencing provides the ability to comprehensively analyse the entire genome, offering a detailed view of its structure and sequence. The assessment of genome integrity using nanopore sequencing can be approached in different ways. One method is to only consider the full genome, defined as reads spanning from the start 1 nt to 4,386 nt, which is the end of the genome. While this method is straightforward, it may underestimate genome integrity due to the potential sequencing challenges posed by the hairpin structures within the ITRs. Previous reports and our findings suggest that these hairpin structures can block

sequencing. Nanopore sequencing provides the ability to comprehensively analyse the entire genome, offering a detailed view of its structure and sequence. The assessment of genome integrity using nanopore sequencing can be approached in different ways. One method is to only consider the full genome, defined as reads spanning from the start 1 nt to 4,386 nt, which is the end of the genome. While this method is straightforward, it may underestimate genome integrity due to the potential sequencing challenges posed by the hairpin structures within the ITRs. Previous reports and our findings suggest that these hairpin structures can block sequencing, leading to incomplete reads (Namkung *et al.*, 2022).

An alternative approach to evaluating full genomes is to define them not only as reads spanning from 1 nt to 4,386 nt but also to include reads that extend from the region beyond the hairpin in the 5' ITR (68 nt) to the last nucleotide (4,317 nt) before the hairpin structure begins in the 3' ITR (Namkung *et al.*, 2022). Specifically, since the hairpin structure is absent beyond nucleotide 68 in the 5' ITR and begins forming at nucleotide 4,317 in the 3' ITR, this second definition sets the genome boundaries from 68 nt to 4,317 nt. By incorporating both approaches, this method accounts for potential sequencing biases introduced by the hairpin structures, which may partially obstruct sequencing, leading to a more accurate representation of genome integrity (**Figure 4.8**).

Because dPCR targets only four specific regions, while nanopore sequencing assesses the entire genome, integrity measurements obtained by NGS may appear lower than those measured by dPCR. To explore this further, an additional analysis of the NGS data focusing on reads that specifically contained the exact four amplicons targeted in the dPCR assays was conducted. This approach enabled determination as to whether the observed variability arises from this difference in the scope of the two methods. However, even with this approach, dPCR resulted in significantly higher genome integrity measurements than nanopore sequencing (**Figure 4.10**).

Although nanopore sequencing offers a detailed view of genome integrity, this method has its own limitations. As previously described in Section 1.9.2.1, one key drawback is its bias toward shorter fragments, which generate a higher number of reads within the same timeframe, even when the starting quantities are equivalent (Dunker-Seidler *et al.*, 2025). As a result, nanopore sequencing tends to underestimate the proportion of full-length genomes as shorter fragments are disproportionately represented in the sequencing output.

In this chapter, it has been demonstrated that the rAAV9 Heat sample contained a higher percentage of full genomes compared to the rAAV9 PureLink generated sample when analysed by nanopore sequencing, although the difference was not statistically significant (**Figure 4.8**). Genome integrity assessment by dPCR also showed no significant difference between the rAAV9 Heat, rAAV9 PureLink, and Direct Addition samples.

When the rAAV9 Heat and rAAV9 PureLink samples were prepared together, pooled, and sequenced in runs 1 and 2, it was consistently observed that a higher number of reads for the PureLink-extracted samples were generated compared to those extracted with the heat-based method. Notably, in run 2, the rAAV9 PureLink 2 sample produced approximately 2.7 times more reads than its heat-treated counterpart (**Table 4.4**). This trend suggests that the PureLink method may be more efficient at extracting/releasing genomes from capsids.

Differences observed in extraction yield and genome integrity between the rAAV9 Heat and rAAV9 PureLink samples by nanopore sequencing are presumably directly related to the extraction process used to isolate the genomes. The rAAV9 PureLink sample was subjected to an extraction method that involves extensive handling (i.e. pipetting, mixing, and vortexing). These processing steps may introduce shear force and therefore increase the likelihood of genome fragmentation. In contrast, the rAAV9 Heat sample was extracted using a simpler heat-based method that minimises handling and potentially reduces fragmentation. Moreover, the PureLink method uses a harsher approach for capsid dissociation, incorporating lysis buffer and proteinase K, which facilitates the release of DNA from the capsids (Suoranta *et al.*, 2021). In contrast, the heat extraction method relies solely on thermal denaturation, which may not release as much encapsidated DNA from the rAAV capsids as the PureLink extraction process (Suoranta *et al.*, 2021).

It was noted that higher fragmentation was observed in runs 2, 3, and 4 compared to run 1 for both rAAV9 Heat and rAAV9 PureLink samples (**Figure 4.4**). The variability between runs may be attributed to differences in read count and quality. Specifically, the lower read counts and reduced read quality in run 1 could have limited the detection of fragmented sequences (**Table 4.4**). The reason for the lower quality reads and reduced read counts in run 1 remains unclear, but it is likely due to factors such as sample handling during library preparation, issues during library loading onto the flow cell, or potential variability in the

performance of the flow cell itself, as a different flow cell was used for runs 2, 3, and 4 compared to run 1.

The extraction yield and genome integrity between runs across the three extraction methods (direct addition to dPCR, heat, and the PureLink kit) were also assessed using dPCR (**Figure 4.9**). Because of the high inter-experimental variability, the rAAV9 Heat and rAAV9 PureLink samples exhibited, it was not possible to compare their extraction yield by dPCR. Although the CV was highest for the rAAV9 Heat samples compared to PureLink (128% versus 79%), this was due to the measurement of the rAAV9 Heat sample of run 3, which gave substantially different results than the samples of run 1 and 2.

Although the variability when using direct addition (no extraction) was lower than that observed with extraction methods, other factors besides extraction efficiency may have contributed to the observed inter-experimental variability. Freeze-thaw cycles before dPCR could have introduced variability in the measurements. Due to the stochastic nature of freezing, it can differentially affect individual samples and thus this effect is not uniform across all experiments or samples, introducing additional variability in genome copy measurements (Bee *et al.*, 2022; Wang *et al.*, 2024).

Another potential source of variation was the multistep 1:10 serial dilution process, where pipetting and mixing errors at each step are propagated, leading to inconsistencies in genome copy quantification. At each step of the dilution series, vortexing is applied to mix the samples, followed by a brief centrifugation. Due to the stochastic nature of the mixing process, it is possible to sample a fraction containing either more or less DNA than intended for the next dilution. These small variations accumulate at each step of the dilution process and can vary between different sample preparations. In addition to these inherent inaccuracies within a single dilution series, further variability arises from inter-experimental differences in the serial dilution process itself. Since serial dilutions are performed on different days for different experiments, any variations in handling conditions between experiments further compound the overall measurement variability. Regardless of the source, the substantial variability underscores the need for optimisation and standardisation of DNA extraction protocols, serial dilution procedures, library preparation workflows, and

freeze-thaw handling to improve the accuracy and reproducibility of rAAV genome titration assays.

The variability in genome integrity measurements using dPCR was 34.3% for rAAV9 Direct Addition, 21.9% for rAAV9 Heat, and 28.1% for rAAV9 PureLink. These values are higher than the variability reported in Chapter 3, Section 3.4.10, where the inter-experimental CV was 8.2% for rAAV9 genome integrity analysis. However, in Chapter 3, a pre-diluted rAAV9 sample of ~ 500 viral genomes/ μL was analysed repeatedly in different experiments. In contrast, Chapter 4 involved separate aliquots at $\sim 7.0 \times 10^9$ viral genomes/ μL that were thawed from -80°C storage, treated with DNase I, and then serially diluted before dPCR analysis. This process introduces several sources of variability. One of them is DNase I reaction efficiency, which might differ from run to run due to environmental conditions and pipetting/mixing. Residual free rAAV9 DNA would inflate the measured titre when the DNase I efficiency is low and decrease the titre when the efficiency is high, leading to increased inter-experimental variability. Moreover, freeze-thaw cycles may affect DNA integrity in each aliquot to varying extents, with the impact likely more pronounced in high-concentration samples ($\sim 7.0 \times 10^9$ viral genomes/ μL) compared to low-concentration ones (~ 500 viral genomes/ μL). In highly concentrated samples, a much larger number of genomes are present and therefore susceptible to fragmentation. As a result, the number of genomes affected by each freeze-thaw event can vary substantially between runs. In contrast, the lower number of genomes in diluted samples limits the extent of variation that can occur between experiments. Additionally, differences in handling during the serial dilution process can lead to inconsistent levels of DNA fragmentation between runs. These factors likely contributed to the increased variability observed in Chapter 4.

An important finding from the nanopore sequencing was a deletion present in a small proportion of genomes in the rAAV9 samples. The deletion was detected in the CamKII α promoter, although deletions in the *RTN3* region were also observed. This deletion could alter protein expression by disrupting key promoter elements or transcription factor binding sites. Such an event could potentially lead to reduced or even complete loss of downstream gene expression. However, as this deletion was observed in only $3.8 \pm 1.0\%$ of the reads, its impact on therapeutic efficacy is likely minimal. The high-quality score (Q59), combined with the

consistent detection of the deletion in every sample, strongly suggests that this deletion is unlikely to be a sequencing artifact.

This deletion potentially arises due to a recombination event between two repeated regions within the recombinant genome (Bzymek and Lovett, 2001; Ribeiro *et al.*, 2008; Oliveira *et al.*, 2010). The first repeated region is part of the CamKII α promoter, and an identical sequence is present immediately downstream. However, the CamKII α promoter itself does not inherently contain a repeated region and thus the reason for the presence of this additional repeat remains unclear.

Nanopore sequencing is also a valuable tool for detecting DNA impurities in rAAV samples (Radukic *et al.*, 2020; Dunker-Seidler *et al.*, 2025). Here, residual DNA was identified in the rAAV9 samples with notably higher levels detected in the rAAV9 PureLink sample compared to the rAAV9 Heat sample (**Figure 4.11**). Interestingly, the rAAV9 PureLink sample contained a larger proportion of host cell DNA compared to the rAAV9 Heat sample. However, the source of the host cell DNA impurities does not appear to be the extraction kit reagents, as the PureLink Extraction Blank showed only a trivial number of reads mapping to the host cell genome. A potential explanation for detecting a higher percentage of reads mapping to the human genome in the rAAV9 PureLink samples compared to the rAAV9 Heat may lie in the DNA extraction efficiency of the two methods. As described above, the PureLink extraction kit is more efficient at extracting DNA from rAAV capsids compared to the heat-based method (Suoranta, Laham-Karam and Ylä-Herttuala, 2021). Consequently, it may also facilitate the release of host cell DNA that may be encapsidated within the rAAV9 capsids. This would lead to higher levels of detected host cell DNA in the rAAV9 PureLink samples compared to the rAAV9 Heat samples.

Unclassified reads that map to various bacteria genera were also observed in both the rAAV9 PureLink and the PureLink Extraction Blanks but were absent from the rAAV9 Heat samples. This suggests that the source of these reads could be related to the reagents of the PureLink Viral RNA/DNA mini kit. Similar observations have been reported in previous studies where inherent bacterial DNA contamination was found in extraction kits (Glassing *et al.*, 2016; Goig *et al.*, 2020; Salter *et al.*, 2014). Importantly, the residual DNA detected consisted predominantly of fragments > 200 nt in length (**Figure 4.12**). These fragments might pose safety concerns due to their potential for transcriptional activity, participation in

recombination events, or ability to trigger immune responses (U.S. Food and Drug Administration, 2020a). Among these impurities, only a small number of residual DNA reads exceeded 1,000 nt in length. These longer fragments warrant closer attention due to their increased potential to compromise the safety of gene therapy (World Health Organisation 1998; Dunker-Seidler *et al.*, 2025).

Reads in both the rAAV9 PureLink and rAAV9 Heat samples that mapped to both the rAAV9 genome and to the transfer plasmid backbone were also present (**Figure 4.7**). These chimeric reads predominantly contained the ITR region of the rAAV9 genome but extended into plasmid backbone sequences. These chimeric reads have also been reported also in a study by Tai *et al.*, 2018 using SMRT® sequencing. Such sequences may arise from errors in packaging during the production of rAAV in host cells, where misprocessing of the plasmid backbone could lead to its incorporation into the packaged genome (Wright, 2014; Zhang *et al.*, 2023; Dunker-Seidler *et al.*, 2025).

In summary, the data presented in this chapter have demonstrated that nanopore sequencing is a valuable methodology for the quality control of rAAV samples, offering crucial insights into genome integrity and impurity profiles. However, a comparison of the percentage of full genomes measured by nanopore sequencing and digital PCR revealed significant discrepancies. These differences are likely due to methodological biases, with nanopore sequencing tending to underestimate genome integrity and digital PCR potentially overestimating it. This is discussed further in the final discussion chapter of this thesis.

Chapter 5: Characterisation of rAAV Capsid Proteins

5.1 Introduction

Recombinant adeno-associated viral (rAAV) vectors are extensively used in gene therapy, but their complex nature makes characterisation challenging (van der Loo and Wright, 2016; Naso *et al.*, 2017; Kontogiannis *et al.*, 2024). rAAV capsids exhibit variability not only between different serotypes but also between and within production batches during manufacturing (Tustian and Bak, 2021). The differences between serotypes primarily arise from variations in the capsid protein sequences. However, even production batches of the same serotype can display heterogeneity in capsid stoichiometry, post-translational modifications (PTMs), and the presence of residual host cell proteins (HCPs) that remain after purification (Kontogiannis *et al.*, 2024).

As outlined in the introduction to this thesis, it is generally considered that an individual rAAV capsid consists of a collective 60 copies of the three viral proteins (VP1, VP2, VP3) in a nominal ratio of approximately 1:1:10, regardless of the serotype (Rose *et al.*, 1971; Bennett *et al.*, 2017). However, recent studies have revealed that this stoichiometry is not fixed but rather depends on the relative expression level of each VP protein. Such studies have shown that capsid assembly is stochastic, meaning there is no absolute, constant stoichiometry for rAAV capsids (Snijder *et al.*, 2014; Wörner *et al.*, 2021). Understanding the stoichiometry is crucial, as variations in VP ratios can affect the infectivity of the viral vectors and potentially result in less effective or even ineffective therapeutic batches (Bosma *et al.*, 2018; Onishi *et al.*, 2023; Erickson *et al.*, 2024). Therefore, work towards the establishment of reference methods for the quantification of the VP1, VP2, and VP3 capsid proteins is an important step required for standardised capsid characterisation, improved measurement consistency, and to help ensure the safety of rAAV-based gene therapies (Moullier and Snyder, 2012; Wright, 2021; Kontogiannis *et al.*, 2024).

Traditionally, VP stoichiometry has been assessed using electrophoresis techniques such as 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 (Su, Sena-Esteves and Gao, 2020; Oyama *et al.*, 2021). Methods such as capillary electrophoresis

and liquid chromatography coupled with UV or fluorescence detection are increasingly used in the industry to enhance accuracy and precision (Fernandes *et al.*, 2022; Fekete *et al.*, 2023). However, due to the high sequence similarity among VP1, VP2, and VP3, these methods often lack the resolution required to clearly separate the individual capsid proteins, making it difficult to obtain precise and accurate quantification of each capsid protein.

Mass spectrometry (MS) is widely recognised 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. In this chapter, a stable isotope dilution approach with mass spectrometry (SID-MS) was implemented for rAAV9 capsid protein quantification and stoichiometry measurement. Signature peptides unique to each capsid protein were identified to quantify VP1, VP2, and VP3, and subsequently determine the mean stoichiometry of the rAAV9 preparation.

Another important CQA for characterisation is the presence of PTMs on the rAAV capsid (Kontogiannis *et al.*, 2024). Identifying PTMs and monitoring how they vary across different production strategies can provide valuable insights for process optimisation. This is achieved by helping to minimise undesirable modifications and promote greater consistency. Among these PTMs, deamidations, oxidations, and acetylations in rAAV capsid proteins are particularly important, as their mapping and characterisation as CQAs are important for assessing and ensuring protein stability, which directly impacts vector performance and consistency during manufacturing (Giles *et al.*, 2018). The occurrence and prevalence of deamidations and oxidations can vary significantly even with minor changes in storage conditions or vector age, making it important to define and monitor these modifications closely (Giles *et al.*, 2018; Zhou and Wang, 2022, 2023). Further, peptides that are not prone to deamidation and oxidation should be selected for quantification purposes. These modifications alter the peptide mass and therefore reduce the peptide's utility as a standardised, quantitative marker. Moreover, the identification of acetylation sites is important as they are known to regulate protein functions, such as interactions with other proteins and DNA. Acetylations are also implicated in modulating protein stability by triggering ubiquitination of the viral capsid, leading to proteasomal degradation (Kouzarides, 2000; Mary *et al.*, 2019).

In addition to PTMs, HCPs must also be closely monitored to ensure product safety. HCPs that remain in the viral vector batch after purification can vary widely depending on the production system and purification process used (Dong *et al.*, 2014). Identification and quantification of these residual proteins are crucial for assessing the safety of gene therapy products, as their presence may elicit immune responses in patients (Bracewell *et al.*, 2021; Jones *et al.*, 2021; Kontogiannis *et al.*, 2024). HCPs can be retained in rAAV preparations through various mechanisms, including direct interactions with capsid or Rep proteins, encapsulation within the capsid, association with the packaged genome, or indirect binding via other HCPs that interact with rAAV components (Nash *et al.*, 2009; Bracewell *et al.*, 2021; Kontogiannis *et al.*, 2024; Leibiger, Min and Lee, 2024).

Identification of PTMs and HCPs is therefore critical for evaluating product quality and safety, ensuring batch consistency, and informing process development. Mass spectrometry is a well-established, highly accurate analytical technique that supports these goals and can be used for confirming protein sequence and identifying PTMs and HCPs (Murray *et al.*, 2006; Hao *et al.*, 2015; Rathore *et al.*, 2018; Mary *et al.*, 2019; Serrano *et al.*, 2023). 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 characterisation too (Dong *et al.*, 2014; Aloor *et al.*, 2018; Rumachik *et al.*, 2020; Bracewell *et al.*, 2021; Leibiger, Min and Lee, 2024).

5.2 Aims of this Chapter

The aim of this chapter was the development of a tailored MS-based approach for accurate quantification of rAAV9 capsid proteins and for determining overall capsid stoichiometry. As an essential part of the method development, the amino acid sequence and the presence of potential PTMs in rAAV9 and rAAV8 preparations were first confirmed by both intact mass analysis (LC-MS) as well as peptide mapping (tandem mass spectrometry (LC-MS/MS)). Additionally, HCPs present in the rAAV8 preparation were identified using LC-MS/MS and compared to those reported in previous studies.

5.3 Materials & Methods

5.3.1 LC-MS Analysis of Intact VPs

For analysis, 5 μ L of rAAV9 (0.5 μ g) or 5 μ L of rAAV8 (0.26 μ g) were injected onto an Avantor ACE 3 C4-300 column (100 $\times$ 2.1 mm i.d.) (Avantor, Cat No. ACE-213-1002) maintained at 50 °C. The mobile phase consisted of 0.5% formic acid (Fisher Scientific, Cat No. 10785711) in ultra-pure water with a resistance of 18.2 M Ω ·cm (Elga Purelab Ultra) (solvent A) and acetonitrile (Romil, Cat 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 to 61 min, 95% B at 62 min, and was then maintained at 95% B until 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 a HESI-II probe source (Thermo Fisher Scientific) in positive electrospray ionisation (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 200 ms, with an AGC target value of 3e6 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 – 4040, charge range: 50 – 100, mass range: 59700 – 85000 Da, sample mass every 10 Da, peak detection range of 500 Da, peak detection threshold 0.01. Similarly, for the identification of VP1, VP2, VP3 in the rAAV9 sample, the following UniDec settings were implemented: m/z range: 290 – 4040, charge range: 30 – 85, mass range: 59700 – 82000 Da, sample mass every 10 Da, peak detection range of 1600 Da, peak detection threshold 0.01.

5.3.2 Preparation of solutions

The following solutions were prepared for the proteolytic digest of rAAV8 and rAAV9 capsids: 200 mM ammonium bicarbonate (AMBIC): 50 mL of ultra-pure water with a resistance of 18.2 M Ω ·cm (Elga Purelab Ultra) was added to 0.79 g of AMBIC (Sigma-Aldrich, Cat No. 09830).

Dilutions of 50 mM and 10 mM AMBIC were prepared by diluting the 200 mM stock in ultra-pure water.

8 M guanidine hydrochloride (gHCl)/80 mM dithiothreitol (DTT): 1 mL of 200 mM AMBIC was added to 0.01234 g DTT (Sigma-Aldrich, Cat No. D5545) and 1.14 g gHCl (Sigma-Aldrich, Cat No. 50950).

45 mM iodoacetamide (IAA): 2 mL of 50 mM AMBIC was added to 0.0166 g of iodoacetamide (IAA) (Sigma-Aldrich, Cat No. I6125).

50 mM ammonium acetate: 20 mL of ultra-pure water was added to 0.0771 g of ammonium acetate (Sigma-Aldrich, Cat No. 240192).

1 µg/µL Glu-C stock: 2 mL of 50 mM ammonium acetate was added to the lyophilised Glu-C (2 mg, Merck, Cat No. 10791156001).

80 mM DTT: 1 mL of 50 mM AMBIC was added to 0.01234 g DTT (Sigma-Aldrich, Cat No. D5545).

1 µg/µL trypsin stock: 20 µL of 50 mM ammonium acetate was added to the lyophilised trypsin (20 µg, Promega, Cat No. V5111).

Weighing was carried out using a Mettler Toledo XP205 analytical balance.

5.3.3 Preparation and Digestion of rAAV8 and rAAV9 Samples with Glu-C

A total of 5 µL rAAV9 or 5 µL of rAAV8 was mixed with 5 µL of a solution comprising 8 M gHCl and 80 mM DTT in 200 mM ammonium bicarbonate (AMBIC) for denaturation of the capsid proteins (Wingfield, 2001). 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 RT for 10 min. Subsequently, 5 µL of 45 mM 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 addition of 50 µL of endoproteinase Glu-C (V8 Protease) solution from a 1 µg/µL stock of Glu-C. The enzymatic digestion was then carried out at 37 °C in a FOC 215i incubator (Velp Scientifica) for 17 h. Post-digestion, the sample was dried using a vacuum centrifuge (Concentrator plus, Eppendorf) at the V-HV setting at 60 °C and then resuspended in 65 µL of 0.1% formic acid in ultra-pure water.

5.3.4 Preparation and Digestion of rAAV8 and rAAV9 Samples with Trypsin

rAAV9 and rAAV8 samples were prepared and denatured following the same procedure as described in Section 5.3.3. 5 μ L of 80 mM DTT was added to quench the alkylation reaction, and the gHCl concentration was diluted to 0.5 M by adding 60 μ L of 10 mM AMBIC. Finally, 10 μ L of trypsin solution (containing 10 μ g of trypsin) (Promega, Cat No. V5111) was added, and enzymatic digestion was carried out at 37 °C in a FOC 215i incubator (Velp Scientifica) for 17 h.

5.3.5 LC-MS/MS Analysis for Protein Sequencing

5 μ L of the digested rAAV9 or rAAV8 sample was injected onto an Avantor ACE 3 C18-300 (100 $\times$ 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 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 analysis was conducted using the Vanquish UHPLC System (Thermo Fisher Scientific) and the Q-Exactive Plus Orbitrap instrument with a HESI-II probe source (Thermo Fisher Scientific) 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. aux gas flow rate. For the first step of the analysis (Full MS), the following settings were applied: 100 ms maximum injection time, 3e6 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 1e5, top 5 ions were analysed, while data were acquired at resolution of 17.5K and isolation window 1.5 m/z.

The Orbitrap MS performed MS/MS analysis in Data Dependent Acquisition mode (DDA). The mass spectrometer first identifies and measures the m/z ratio of all the ions present in the sample (MS1) and then selects the five most intense ions and transfers them for collision and fragmentation in MS2. Subsequently, the process is repeated for the next most abundant ions, until all ions present in the sample have been analysed.

The data were analysed with PEAKS v7.5 (Bioinformatics Solutions Inc.) 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 v7.5 software (Bioinformatics solutions Inc.) with a false discovery rate (FDR) set at 5%, and *de novo* Average Local Confidence (ALC) score $\geq$ 80% and 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 capsid proteins was optimised by using both trypsin and Glu-C enzymes for digestion.

For host cell protein peptide mapping, a false discovery rate (FDR) of 0.1% was applied for each peptide, and each protein identification required at least one unique peptide. Additionally, the $-10\log P$ score for each identified protein was set to $\geq$ 40. Host cell protein identifications were further validated through three independent in-solution digestion experiments. The Human Proteome Database was downloaded from UniProt (proteome:up000005640 AND reviewed:true, 20,420 entries, downloaded on 12th January 2023).

5.3.6 Identification of Peptides from each Capsid Protein Suitable for Determination of rAAV9 VP Capsid Stoichiometry

The rAAV9 capsid protein sequences (UniProt ID: Q6JC40) were retrieved from UniProt. Peptides unique to each capsid protein were then selected *in silico* using Skyline software (University of Washington) and PeptideCutter (ExPASy) (Swiss Institute of Bioinformatics) with Glu-C as the specified endoproteinase (Prakash *et al.*, 2009; MacLean *et al.*, 2010). These peptides are the signature peptides for the quantification of the entire capsid protein since they are present only in the capsid protein they derive from. These unique peptides are the N-terminal peptides of VP1, VP2, and VP3 and will be referred to as VP1u, VP2u, VP3u. To quantify the total capsid content, a peptide located at the C-terminal of the capsid proteins (present in all three capsid proteins) was also selected (**Figure 5.1**). This peptide is referred to as VPc. However, since VP2u was not identified during the study, an alternative approach was developed to estimate the 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. The sequences of the peptides are shown in **Table 5.1**. 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 host cell proteins (Lange *et al.*, 2008; Kiyonami and Domon, 2010; Sayers *et al.*, 2022). To further validate the specificity of the method, a HEK293 cell lysate (Novus Biologicals, Cat. No. NBP2-25048) was digested with Glu-C using the same procedure described in Section 5.3.3. The digest was then analysed using the LC-MS/MS MRM method detailed in Sections 5.3.13 and 5.3.14, which monitors the transitions corresponding to the VP1u, VP3u, and VPc peptides.

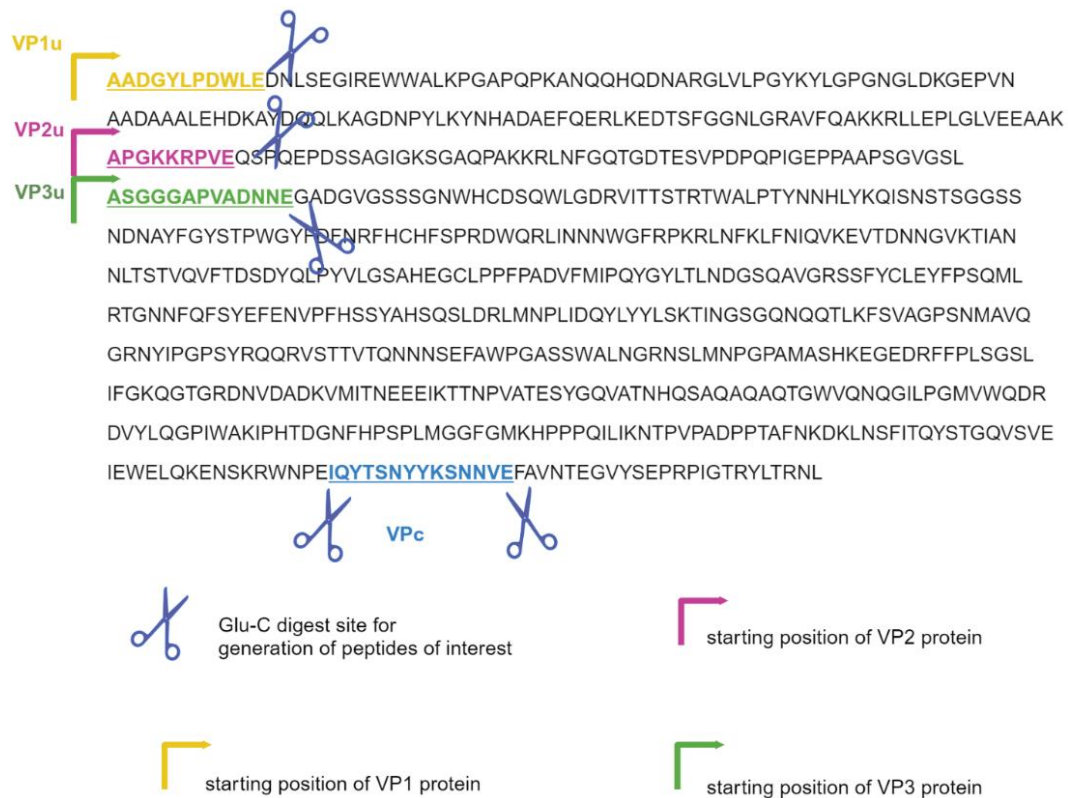


Figure 5.1 Schematic representation of unique peptide generation through Glu-C digestion. The sequence of the rAAV9 capsid protein is shown with yellow arrows indicating the starting position of VP1, pink arrows marking the starting position of VP2, and green arrows marking the starting position of VP3. Scissors denote Glu-C digestion sites, resulting in unique peptides for VP1u (yellow), VP2u (pink), VP3u (green), as well as a common peptide VPc (shown in blue). Created with BioRender.com.

Table 5.1 Sequences of the peptides (VP1u, VP2u, VP3u, VPc). The first alanine of VP1u and VP3u is acetylated. Ac. = acetylation.

Peptide	Sequence
VP1u	(Ac.) AADGYLPDWLE
VP2u	APGKKRPVE
VP3u	(Ac.) ASGGGAPVADNNE
VPc	IQYTSNYYKSNNVE

5.3.7 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 Number: 27252). The set included a 0.1 g weight from Mettler and 0.5 g, 1 g, 5 g, 10 g, 20 g, 50 g, 100 g, and 200 g weights from Troemner.

5.3.8 Preparation of Isotopically Labelled Peptide Mastermix

The three peptides (VP1u, VP3u, VPc) were purchased as isotopically labelled peptides (30 nmol each) from Thermo Fisher Scientific (**Table 5.2**). Initially, each labelled peptide was dissolved in 2 mL of 5% acetonitrile in water and subsequently diluted gravimetrically in ultra-pure water (Elga Purelab Ultra). A master mix containing the three isotopically labelled peptides (VP1u, VP3u, and VPc) was prepared at a nominal ratio of 1:10:11, corresponding to final concentrations of 5 nM for VP1u, 50 nM for VP3u, and 55 nM for VPc.

Table 5.2 Sequences of the labelled peptides VP1u, VP3u, VPc. Ac. = acetylation.

Peptide	Sequence	Modified Amino Acid in Isotopically Labelled Peptide	Mass Shift (Da)
VP1u	(Ac.) AADGYLPDWLE	Leucine (L)	+7
VP3u	(Ac.) ASGGGAPVADNNE	Proline (P)	+6
VPc	IQYTSNYYKSNNVE	Lysine (K)	+8

5.3.9 Preparation of Natural Peptides

The VP1u, VP3u, VPc were purchased as natural peptides from Thermo Fisher Scientific (ordered as 2 mg each, at 95% purity). The natural peptide VP1u was dissolved in 2 mL of 5% acetonitrile and 0.1% formic acid in ultra-pure water. VP3u was dissolved in 4 mL of 0.1% formic acid in water. VPc was dissolved in 4 mL of 75% DMSO (Sigma-Aldrich, Cat No. 20-139) and 0.1% formic acid in water.

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

For purity analysis, 1 µg of each synthesised natural peptide (VP1u, VP3u, VPc) was injected onto 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 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.

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 natural peptide were then determined by manual peak integration using Freestyle™ software (Thermo Fisher Scientific).

5.3.11 Amino Acid Analysis

Amino acid analysis of the peptides VP1, VP3, and VPc was carried out based on a previously reported approach (Torma *et al.*, 2017; L. Zhang *et al.*, 2024). 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 analyses 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 labelled amino acids (Cambridge Isotopes Laboratory, Andover, MA, USA) used were: L-glycine ($^{13}\text{C}_2^{15}\text{N}$, 98% purity), L-leucine ($^{13}\text{C}_6^{15}\text{N}$, 98% purity), L-alanine ($^{13}\text{C}_3^{15}\text{N}$, 98% purity), L-valine ($^{13}\text{C}_5^{15}\text{N}$, 98% purity), L-lysine/2HCl ($^{13}\text{C}_6$, $^{15}\text{N}_2$, 98% purity), L-isoleucine ($^{13}\text{C}_6$, 98% purity), L-proline ($^{13}\text{C}_5^{15}\text{N}$, 98% purity).

Prior to preparing the blends for amino acid analysis, the VP1u 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 labelled amino acids along with the target peptide (VP1, VP3, or VPc). Alongside, three calibration blends were made, combining natural amino acids and labelled standards in equimolar amounts to the sample blends. Additional preparations included reagent blanks, a natural amino acid standard solution without the labelled counterpart, a labelled amino acid standard solution without the natural amino acids, and a recovery vial (**Table 5.3**). The labelled 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).

Table 5.3 Composition of controls, sample blends, calibration blends, and blanks used for amino acid analysis.

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

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-butyltrimethylsilyl-N-methyltrifluoroacetamide (MTBSTFA) with 1% tert-butyltrimethylchlorosilane (TBDMSCL) (Sigma-Aldrich, Cat No. 375934) for derivatisation, and incubated at 85 °C for 90 min in a thermoshaker (Microtherm, Camlab) with shaking at 50 rpm. The samples were then analysed 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 analyse the amino acids in the samples. The transitions monitored are shown in **Table 5.4**.

Table 5.4 MRM transitions used for analysis of each labelled and unlabelled amino acid.

Amino Acid	Quantifier Transition (m/z)	Quantifier Transition (m/z)
	Unlabelled MTBSTFA Derivative	Labelled 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
Phenylalanine	336/308	346/317
Lysine	431/300	439/307
Arginine	442/199	448/203

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 analysed amino acid (either natural in the calibration blend or endogenous in the sample blend) to its corresponding isotopically labelled 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 the Equation 5.1. 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 labelled 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 unlabelled and labelled 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 unlabelled and labelled peptides in the sample and in the calibration blend, respectively.

$$W_x = W_z \times \frac{m_y}{m_x} \times \frac{m_z}{m_{yc}} \times \frac{R_B}{R_{BC}} \quad (\text{Equation 5.1})$$

To determine the nmol/g concentration of each amino acid, the calculated amount (W_x) was first divided by the molecular weight of the amino acid and was 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 was 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.

5.3.12 Preparation of Standard Curves

For the quantification of the signature peptides, separate standard curves were gravimetrically prepared for each peptide. The VP1u, VP3u, and VPc natural peptide stocks were gravimetrically diluted in 0.1% formic acid in ultra-pure water to ensure that the resulting standard curves covered the expected concentrations of each endogenous rAAV9 peptide. The concentrations of the natural peptides at each level were calculated based on amino acid analysis, purity assessment, and the exact weights recorded at each dilution step using a calibrated Mettler Toledo XP205 analytical balance. New standard curves were prepared for each experiment.

Each point of the standard curves for VP1u, VP3u, and VPc consisted of a gravimetrically prepared 1:1 (volume ratio) mixture of synthesised natural peptide VP1 and a mastermix comprising synthesised isotopically labelled VP1u, VP3u, and VPc peptides. The concentration of the isotopically labelled peptides remained constant across all points of the standard curves (5 nM for VP1u, 50 nM for VP3u and 55 nM for VPc), while the concentration of each VP peptide decreased to generate the standard curve. 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 natural peptide with the isotopically labelled peptide master mix, ensuring a gradually increasing ratio of peak areas between each natural peptide and the isotopically labelled counterpart.

The digested rAAV9 sample was first completely dried using a vacuum centrifuge (Concentrator plus, Eppendorf) at the V-HV setting at 60 °C and then resuspended in 50 µL of 0.1% formic acid in water. An additional 50 µL of isotopically labelled peptide master mix (prepared in 0.1% formic acid in water) was then added (same amount as that used for the preparation of the different levels of the standard curve). The final sample volume was 100 µL. As 5 µL of rAAV9 was used for digestion (see Section 5.3.3), this corresponds to a 1:20 dilution.

5.3.13 Analysis of Digested rAAV9 by Triple Quadrupole MS

Samples were analysed using an ACQUITY UPLC M-Class system (Waters Corporation) coupled to a Xevo TQ-XS triple quadrupole mass spectrometer with MRM method analysing three transitions for each peptide, one quantifier (transition 1 (T1)), and two qualifiers (transitions 2 and 3 (T2, T3)) (Dawson *et al.*, 1982; Michaud *et al.*, 2024). Separation was performed on an ACQUITY UPLC HSS T3 C18 column (1.8 µm, 1.0 × 150 mm) (Waters Corporation, Cat No. 176001130). The column temperature was maintained at 40 °C. The mobile phase A consisted of ultra-pure water with a resistance of 18.2 MΩ·cm (Elga Purelab Ultra) and containing 0.1% formic acid, while mobile phase B was acetonitrile with 0.1% formic acid.

The gradient elution 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, and then returned to 2% acetonitrile from 13.5 to 14 min, and then held at this until 30 min. Each gradient 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, and 25% isopropanol (v/v), and seal washing with a 1:1 mixture of water and acetonitrile for 5 minutes per 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 nebuliser pressure of 7 Bar. Argon was utilised 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 transitions for each peptide were measured through peak integration using the TargetLynx™ Application Manager within the MassLynx v4.2 software (Waters Corporation). Peaks were detected at the exact retention times corresponding to the elution of each peptide, with the isotopically labelled standards serving as controls to confirm retention time alignment. Standard peak detection parameters were applied, including a peak-to-peak noise amplitude of 10, balance set to 10, and splitting set to 50.

5.3.14 Development of Multiple Reaction Monitoring (MRM) Method

Three transitions for each peptide were monitored in the MRM method developed (Table 5.5), including one quantifier (T1) and two qualifier transitions (T2 and T3) for each peptide. To enhance the sensitivity and the signal-to-noise ratio of the quantification method, the collision energy (CE) was individually optimised for each peptide (MacLean *et al.*, 2010).

Table 5.5 Selected transitions for the quantification of peptides VP1u, VP3u, and VPc.

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

Skyline software (University of Washington) was used to generate target lists for CE optimisation of each of the three transitions of the peptides VP1u, VP3u, and VPc. The CE

optimisation method in Skyline works by creating a series of closely related transitions for each main transition, with slight variations in m/z values at the hundredths decimal place (Holstein Sherwood *et al.*, 2011). This modification allows each of these slightly varied transitions to be targeted at a unique CE, effectively enabling a single transition to be analysed at multiple collision energies in a single run.

In this study, each transition of the selected peptides was evaluated at 12 different collision energies in Skyline, using increments of 2. The CE ranges tested were: 6–28 for VP1u, 4–26 for VP3u, and 14–38 for VPc. For this test, the synthesised natural peptides VP1u, VP3u, and VPc were analysed using an ACQUITY UPLC M-Class system (Waters Corporation) coupled to a Xevo TQ-XS triple quadrupole mass spectrometer using the settings and chromatography mentioned in Section 5.3.13. For each transition, the CE that produced the highest signal intensity was selected for use in the final quantification method.

5.3.15 LC-MS/MS Quantification Method Workflow Overview

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 SID. A known amount of isotopically labelled peptide mix containing standards for VP1u, VP3u, and VPc was spiked into the digested rAAV sample. Standard curves were prepared by mixing varying concentrations of the corresponding natural peptides with a constant amount of the isotopically labelled peptide mix (same as the amount spiked into the sample).

Quantification was carried out using MRM on a Xevo TQ-XS triple quadrupole mass spectrometer, enabling precise detection and measurement of the unique peptides. 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 labelled counterparts) into the corresponding standard curve. **Figure 5.2** provides a schematic overview of the MRM-based quantification workflow. The limit of detection (LOD) and limit of quantification (LOQ) for each experiment were determined based on the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) Q2(R2) guidelines (U.S. Food and Drug Administration, 2024) using the equations below (Equation 5.2 and Equation 5.3), where σ

corresponds to the standard deviation of y-intercepts of the regression lines, and S corresponds to the slope of the standard curve.

$$\text{LOD} = 3.3 \times \frac{\sigma}{S} \quad (\text{Equation 5.2})$$

$$\text{LOQ} = 10 \times \frac{\sigma}{S} \quad (\text{Equation 5.3})$$

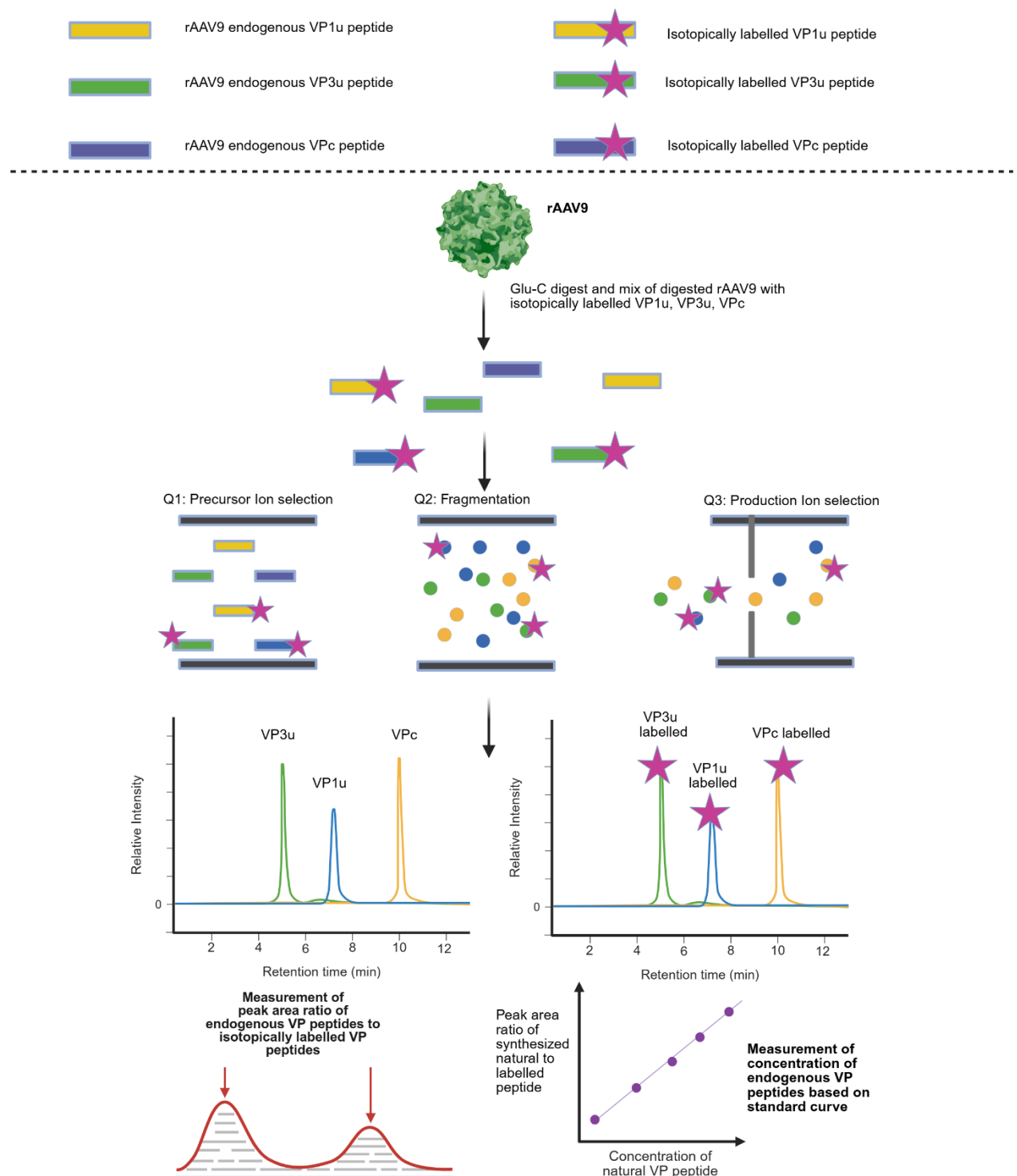


Figure 5.2 Multi-step workflow for quantification of VP1u, VP3u, VPc. rAAV9 particles are denatured and digested with Glu-C to generate peptides corresponding to VP1u, VP3u, and VPc. These peptides are mixed with isotopically labelled standards for VP1u, VP3u, and VPc and analysed using an MRM method. In Q1, precursor ions are selected based on their m/z , in Q2, the ions are fragmented, and in Q3, product ions are selected and then detected. The peak areas of the quantifier transitions for endogenous peptides and their isotopically labelled counterparts are measured, and the ratios are calculated. A standard curve, prepared with the natural VP1u, VP3u, and VPc peptides of known concentrations mixed with isotopically labelled standards, is used to determine the concentrations of endogenous VP1u, VP3u, and VPc in the rAAV9 sample. Created with BioRender.com.

5.3.16 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 covariance for VP3u-VPc pair, VPc-VP1u pair, and VP1u-VP3u pair was 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 the Equation 5.4. In this equation, x_i and y_i 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.

$$\text{Cov}(X, Y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} \quad (\text{Equation 5.4})$$

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 the Equation 5.5, where $\sigma_{\{VP_x\}}$ represents the inter-experimental variability of each peptide, and Cov represents the sample covariance for each peptide pair.

$$\sigma_{\{VP_2\}} = \sqrt{\sigma_{\{VP1u\}}^2 + \sigma_{\{VP3u\}}^2 + \sigma_{\{VPc\}}^2 - 2 \times \text{Cov}(VP3u, VPc) - 2 \times \text{Cov}(VP1u, VPc) + 2 \times \text{Cov}(VP1u, VP3u)}$$

(Equation 5.5)

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 5.6).

$$CV_{\text{combined}} = \sqrt{CV_{\text{AAA}}^2 + CV_{\text{purity}}^2 + CV_{\text{exp}}^2} \quad (\text{Equation 5.6})$$

5.4 Results

5.4.1 Measurement of the Intact Masses of the rAAV Capsid Proteins with LC-MS for Serotype Identification

Initially, the VP proteins of rAAV8 and rAAV9 samples were identified with intact MS analysis. The capsid proteins of both rAAV8 and rAAV9 were identified by deconvoluting the spectra corresponding to the 35.5-minute peak in their respective chromatograms (**Figure 5.3**). Multiple peaks are observed in the chromatogram of the rAAV8 sample, corresponding to residual proteins from the production process. Protein impurities were previously observed in Section 2.4.7. These proteins are analysed and identified in Section 5.4.9. The theoretical and observed masses by this analysis are reported in **Table 5.6**, while the deconvoluted spectra and the detection of the peaks corresponding to the capsid proteins of rAAV8 and rAAV9 are shown in **Figure 5.4**. In the intact mass analysis of rAAV8 samples, 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 to 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 to 738, with the first threonine removed. The 59,800 Da mass corresponds to the VP3 protein, covering amino acids 205 to 738 with acetylation of the N-terminal alanine and the removal of both the first threonine and methionine residues.

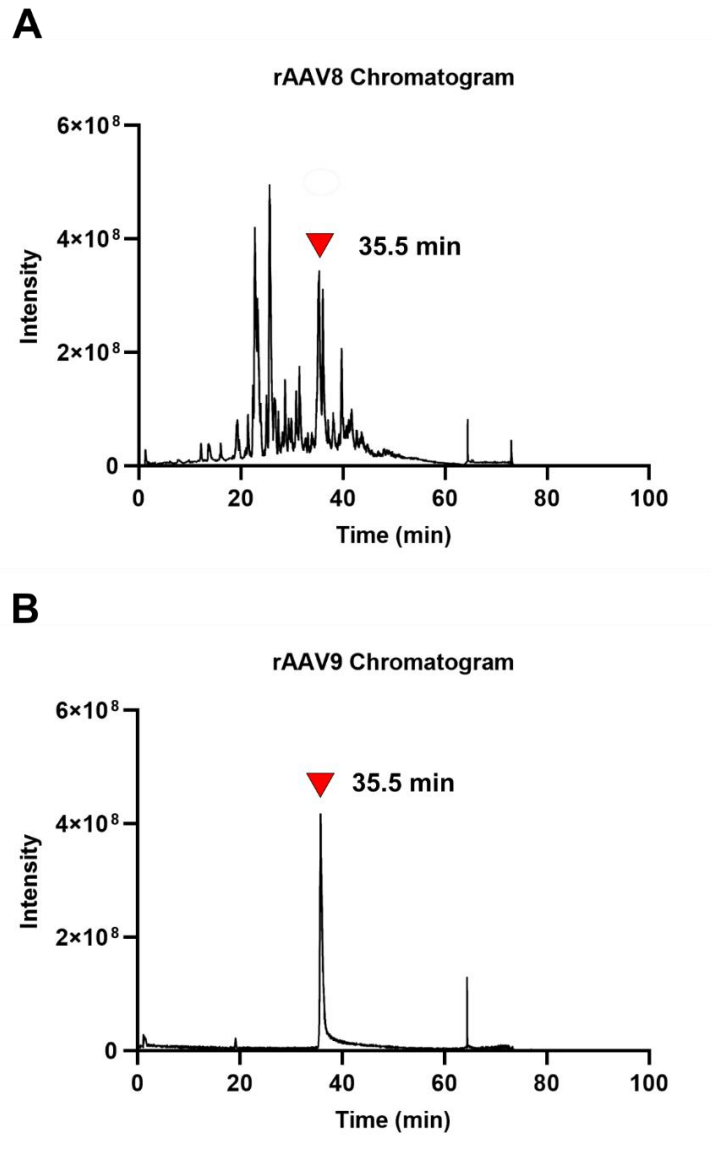


Figure 5.3 Chromatograms of rAAV8 and rAAV9 samples analysed with LC-MS. (A) Chromatogram of rAAV8. (B) Chromatogram of rAAV9. The red arrow indicates the retention time at which the capsid proteins VP1, VP2, and VP3 of rAAV8 (A) and rAAV9 (B) were detected. The column flushing step begins at 62 minutes with a transition to 95% solvent B, explaining the appearance of a peak at approximately 65 minutes in both samples.

Table 5.6 Theoretical and measured masses (Da) of VP1, VP2, VP3 for rAAV8, rAAV9 capsids.

rAAV 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

Similarly, in the intact mass analysis of rAAV9, 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 to 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 to 736, with the first threonine removed. Lastly, the 59,730 Da mass is associated with the VP3 protein, covering amino acids 204 to 736 with acetylation of the N-terminal alanine and removal of both the first threonine and methionine residues.

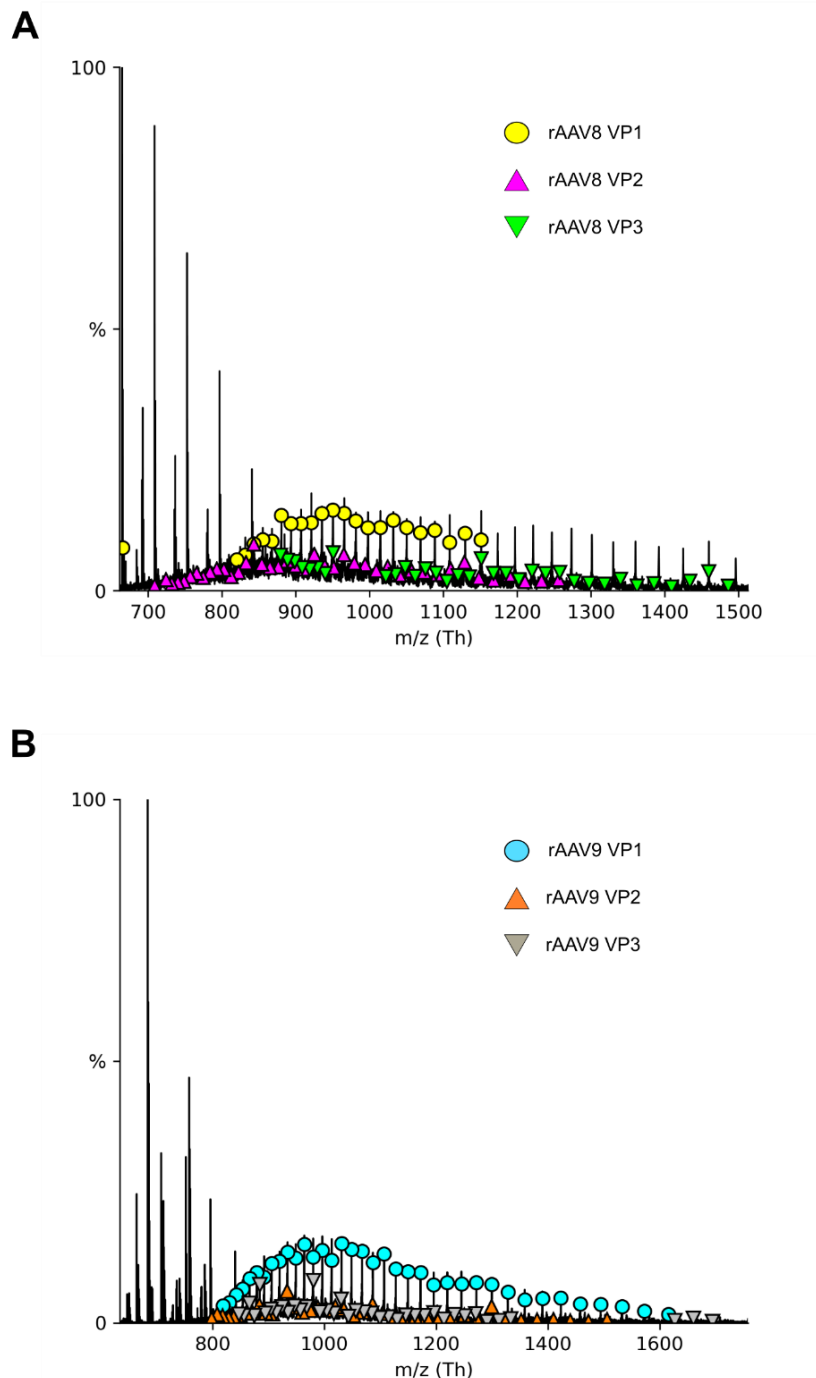


Figure 5.4 Mass spectrometry analysis and identification of the capsid proteins VP1, VP2, VP3 in rAAV8 (A) and rAAV9 (B) samples. The deconvoluted spectra, generated using UniDec, are presented in (A) for rAAV8 and (B) for rAAV9. The spectra show distinct peaks corresponding to the capsid proteins VP1, 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 grey 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.

5.4.2 Mass Spectrometry 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 undertaken. 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. Initially, digestion was performed using 10 µg of Glu-C. However, this condition resulted in low sequence coverage of rAAV9 VP1 (16%) (**Appendix Figure 7.4**). Therefore, 50 µg of Glu-C was used in all subsequent digest experiments as described in Section 5.3.3, as it led to substantially higher sequence coverage of rAAV9 VP1 protein 59.3% (only Glu-C) (**Figure 5.5**).

For the specific rAAV9 capsid proteins, sequence coverage based on Glu-C and trypsin digests was estimated to be 77.7% for VP1 (**Figure 5.5**), 81.5% for VP2 (**Appendix Figure 7.5**), and 88.6% for VP3 (**Appendix Figure 7.6**). For the rAAV8, sequence coverages of 78.2%, 76.7%, and 81.9% were achieved for VP1 (**Figure 5.6**), VP2 (**Appendix Figure 7.7**), and VP3 (**Appendix Figure 7.8**), respectively.

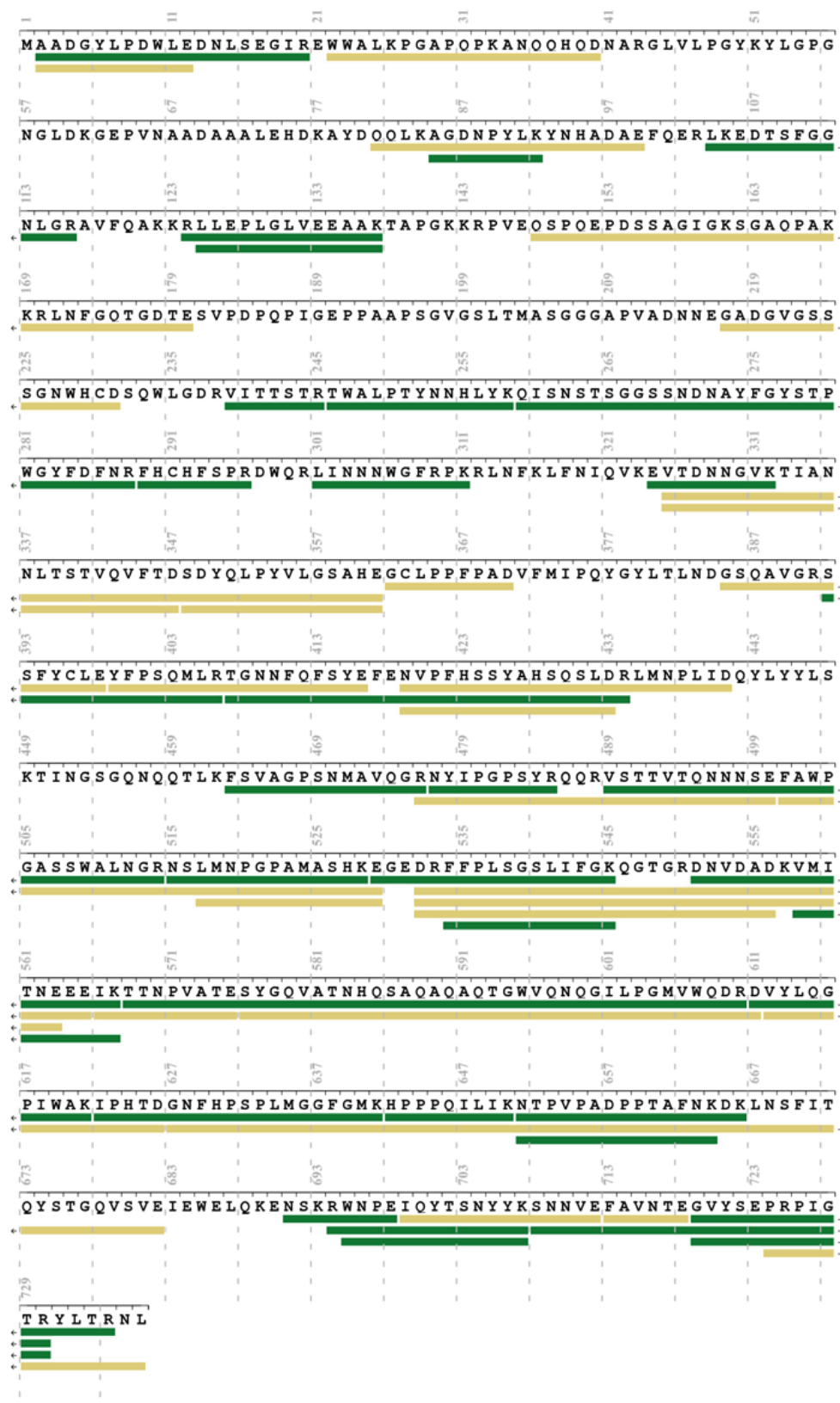


Figure 5.5 Protein sequence coverage map of rAAV9 VP1 protein. The peptide map shows sequence coverage of VP1 (77.7%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).

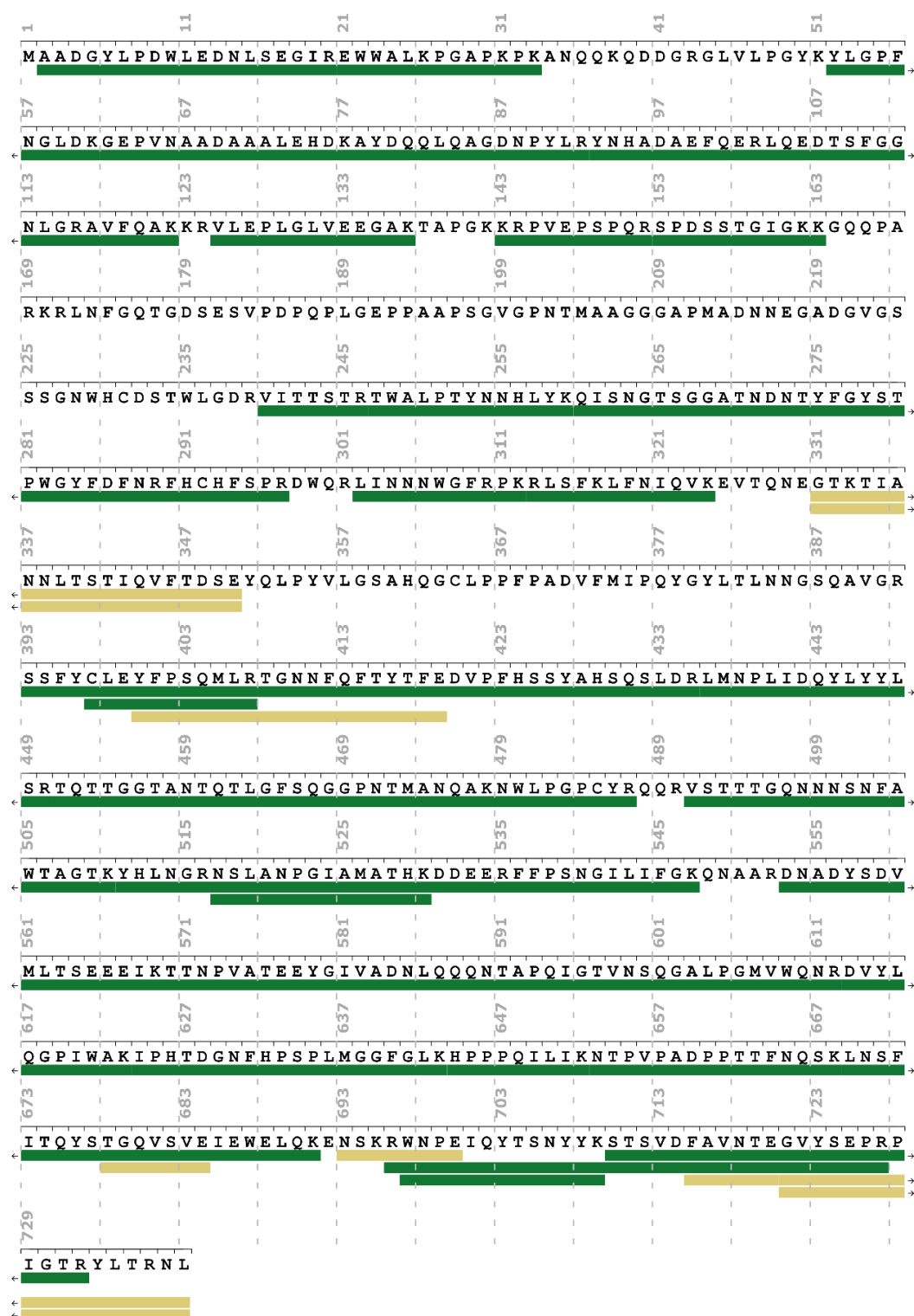


Figure 5.6 Protein sequence coverage map of rAAV8 VP1 protein. The peptide map shows sequence coverage of VP1 (78.2%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).

The amino acid sequence of VP1u, VP3u, and VPc 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. Among the peptides identified, the unique peptides for VP1 (VP1u), VP3 (VP3u), and the common capsid peptide (VPc) were successfully detected and selected for the quantification method. However, the unique peptide for VP2 was not detected. **Figure 5.7** shows the b- and y-ion spectra of VP1u, VP3u, and VPc.

To confirm the presence of the peptides of interest, the rAAV9 digest was spiked with an isotopically labelled peptide mastermix and analysed using the developed MRM method on the triple quadrupole mass spectrometer. The monitored transitions and corresponding retention times for both endogenous and labelled peptides are presented in **Appendix Figure 7.9** (VP1u), **Appendix Figure 7.10** (VP3u), and **Appendix Figure 7.11** (VPc). Based on the transitions observed for the isotopically labelled peptides, the corresponding endogenous peptides VP1u, VP3u, and VPc were confidently identified at their respective retention times. Specifically, VP1u was eluted at 12.91 min, VP3u at 8.55 min, and VPc at 8.85 min. To ensure assay specificity, the transitions for VP1u, VP3u, and VPc (endogenous and isotopically labelled) were monitored in a Glu-C digested HEK293 cell lysate. No corresponding signals above the background level were detected, confirming that the assays are specific to the target peptides (**Appendix Figure 7.12**).

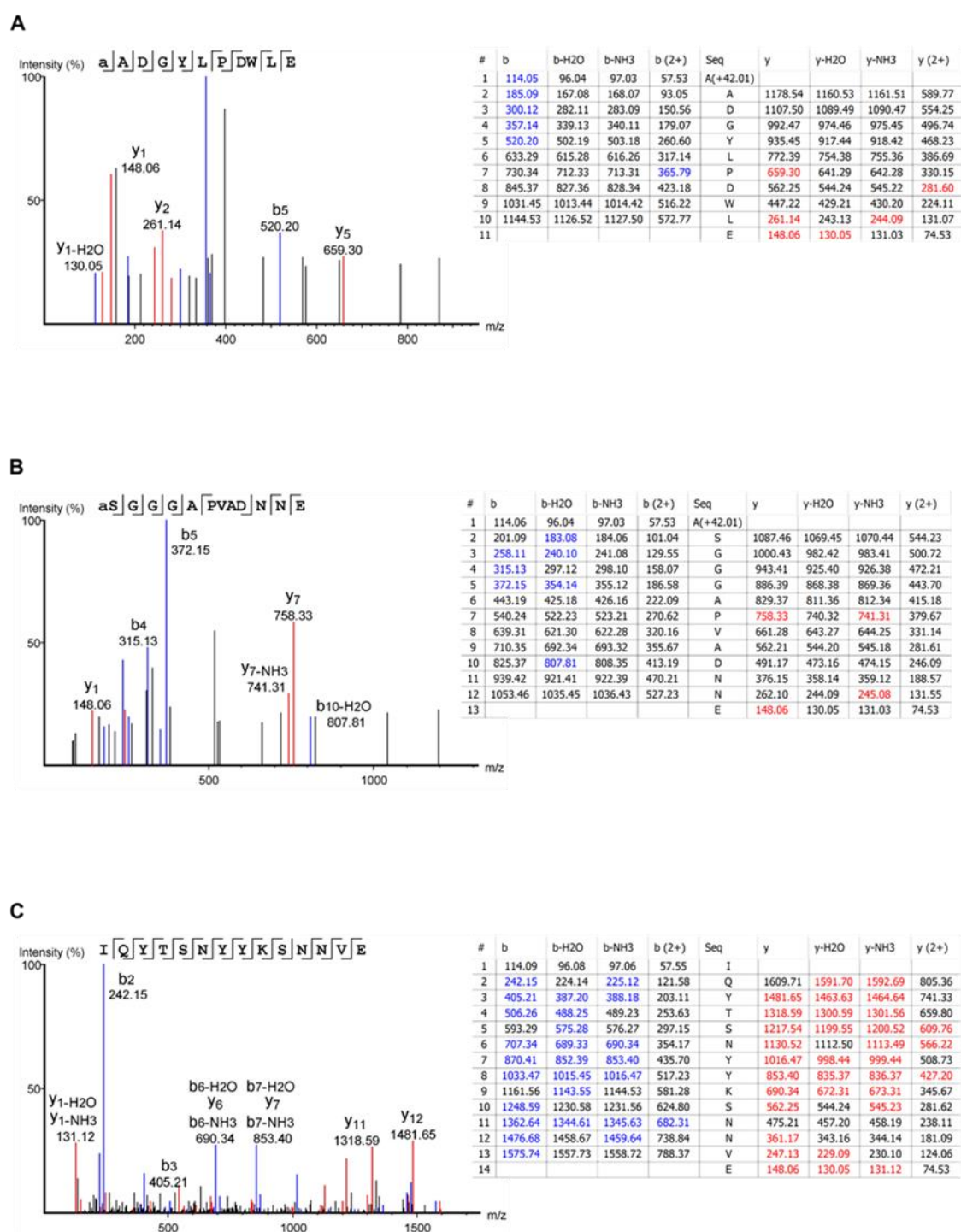


Figure 5.7 Identification of VP1u, VP3u, and VPc peptides based on their MS/MS spectra. (A-C) shows the b and y ions generated from MS/MS fragmentation for identifying (A) VP1u. (B) VP3u, and (C) VPc. Peaks corresponding to b ions are shown in blue, and peaks corresponding to y ions are shown in red. 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.

5.4.3 Detection of PTMs by MS Analysis; Deamidation and Oxidation

Next, deamidation sites present in the VP proteins were mapped. A list of the deamidations detected in rAAV8 and rAAV9 is presented in **Table 5.7**. The corresponding b and y ion spectra for the identification of deamidations in rAAV8 and rAAV9 are shown in **Figure 5.8** and **Figure 5.9**, respectively. The deamidations observed in the rAAV8 sample were located at N57, N514, and N540. When rAAV9 capsid VPs were analysed, deamidation was observed at N254, N303 or N304, and N512. It is ambiguous whether the deamidation occurs at N303 or N304, as the observed y ion includes both asparagine residues at 303 and 304 sites, making it impossible to determine which specific site is modified. Importantly, no oxidations were detected in either of the rAAV8 or rAAV9 samples.

Table 5.7 Deamidation sites detected in trypsin-generated peptides from the capsid proteins of rAAV8 and rAAV9 samples. n corresponds to deamidated asparagine (N) residue.

Serotype	Deamidated Sequence (Trypsin Digest)	Amino Acid Position
rAAV8	YLGPFnGLDK	N57
	YHLnGR	N514
	FFPSnGILIFGK	N540
rAAV9	TWALPTYNnHLYK	N254
	LINnNWGFRPK	N303 or N304
	VSTTVTQNNNSEFAWPGASSWALnGR	N512

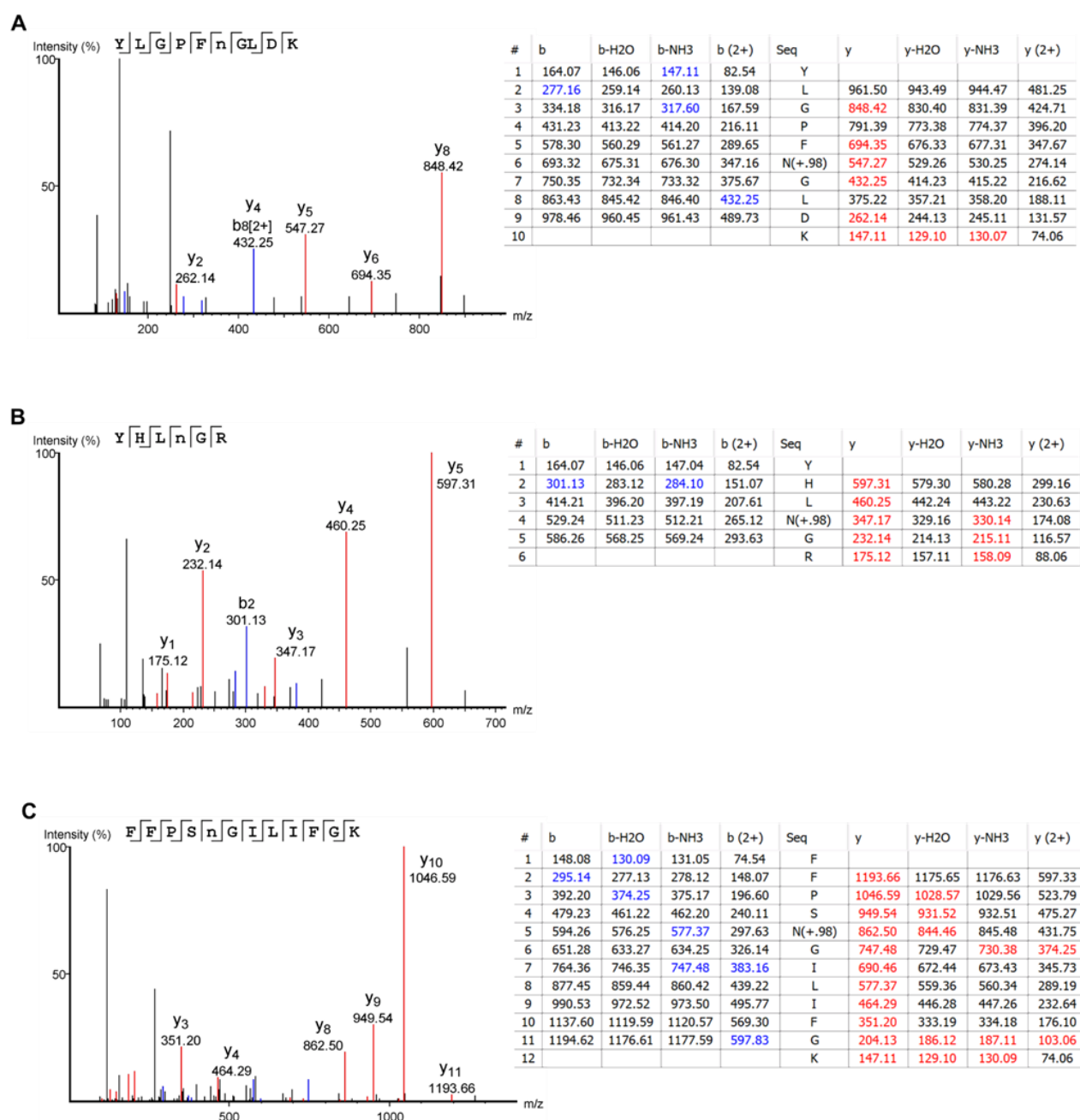
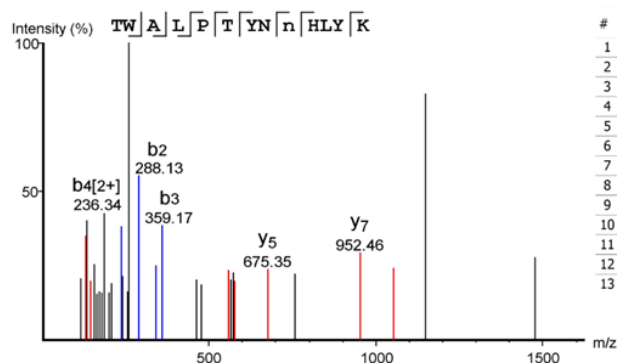


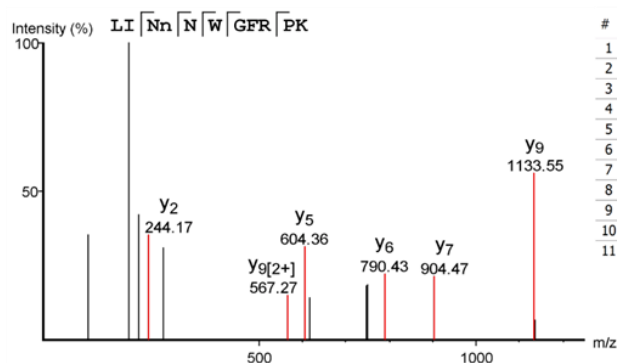
Figure 5.8 Identification of deamidations in rAAV8 based on MS/MS spectra. (A-C) shows the b and y ions generated from MS/MS fragmentation for identifying (A) the deamidation site at N57, (B) the deamidation at N514, (C) the deamidation at N540. Peaks corresponding to b ions are shown in blue, and peaks 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



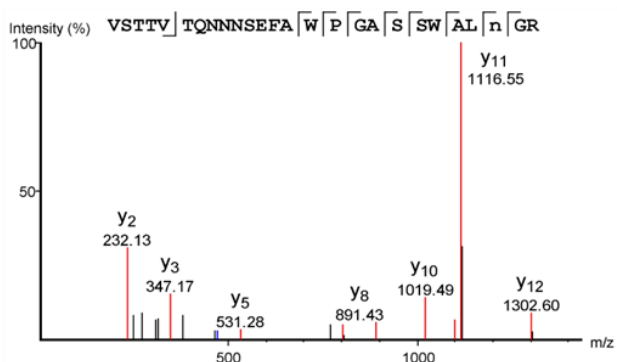
#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)
1	102.06	84.04	85.03	51.53	T				
2	288.13	270.12	271.11	144.57	W	1520.75	1502.74	1503.73	760.88
3	359.17	341.22	342.14	180.09	A	1334.67	1316.66	1317.65	667.84
4	472.26	454.25	455.23	236.34	L	1263.64	1245.63	1246.61	632.32
5	569.31	551.30	552.28	285.15	P	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	1098.54	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 5.9 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.

5.4.4 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 labelled peptide mastermix post-digestion, as previously described. Aliquots were collected at 1, 3, 5, 7, and 23 h and analysed using the developed 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 labelled standards was calculated. The ratio of the peak areas to their isotopically labelled 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 5.10**).

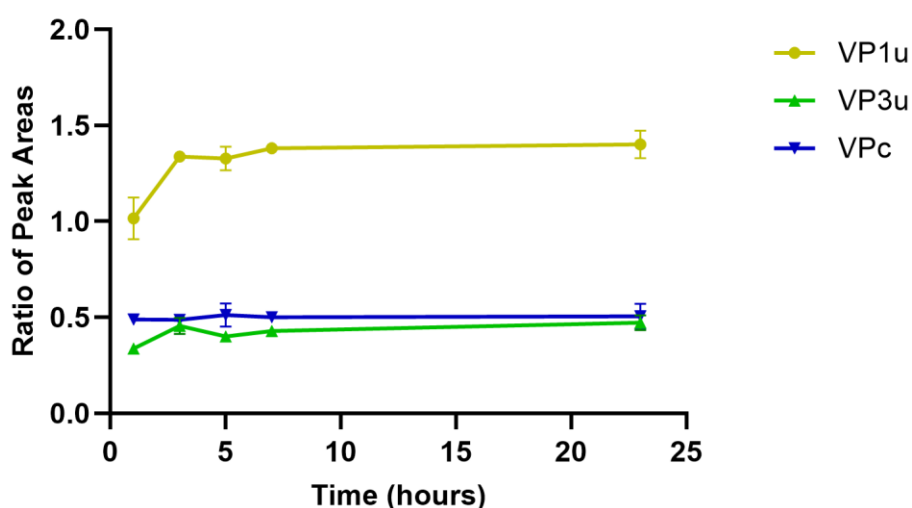


Figure 5.10 Release of VP1u, VP3u, and VPc peptides from rAAV9 during Glu-C digestion over time. The ratios of the peak areas of the quantifier transitions for each endogenous rAAV VP1u, VP3u, and VPc peptide to a fixed amount of their corresponding isotopically labelled standards were measured at 0, 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.

5.4.5 Purity Assessment of Natural Peptides

The purity of each synthesised natural peptide (VP1u, VP3u, and VPc) was assessed using intact LC-MS analysis, with three replicate injections per peptide. The results showed that none of the peptides met the expected purity of 95% as specified by the supplier. The chromatograms for each peptide from one injection are presented in **Figure 5.11**. To identify which peak in each chromatogram corresponds to the VP1u, VP3u, or VPc peptide, the specific

m/z ratios for each peptide were searched using the Extracted Ion Chromatogram (XIC) function in Xcalibur™ software. The highest-intensity peak in each chromatogram corresponds to the natural peptide of interest. The measured purity values are listed in **Table 5.8**. In particular, the VP3u peptide exhibited a markedly lower purity than expected (22.1% vs. 95%). The measured purities were $79.7 \pm 1.1\%$ for VP1u, $22.1 \pm 0.5\%$ for VP3u, and $86.9 \pm 0.1\%$ for VPc.

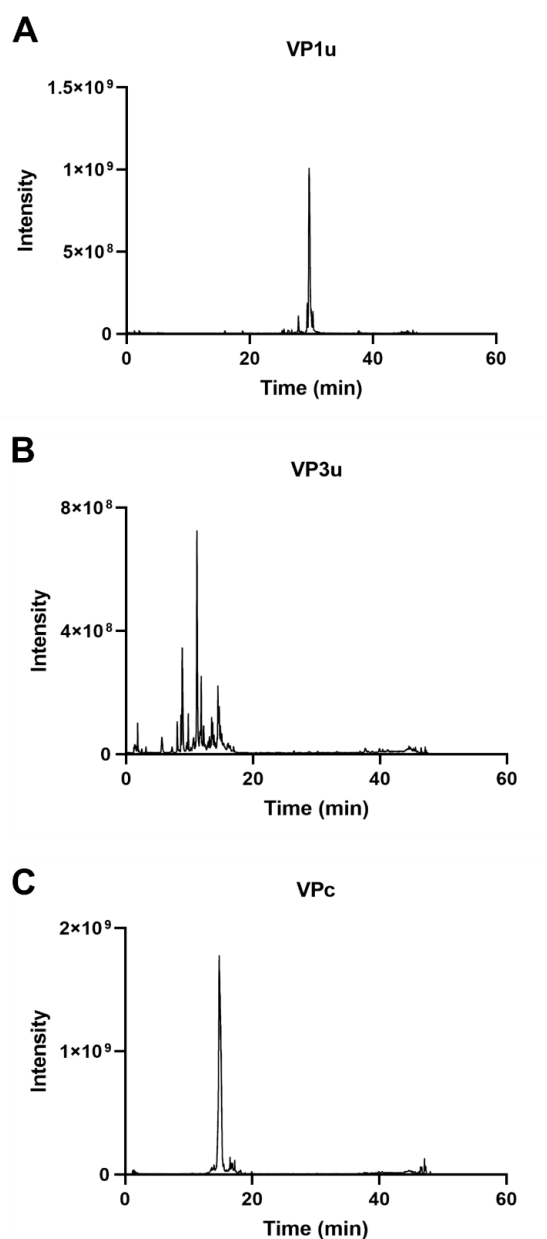


Figure 5.11 Intact LC-MS analysis of VP1u, VP3u, and VPc natural peptides. The highest-intensity peak in each chromatogram corresponds to the natural peptide of interest. (A) Chromatogram of VP1u natural peptide. (B) Chromatogram of VP3u natural peptide. (C) Chromatogram of VPc natural peptide.

Table 5.8 Purity assessment of VP1u, VP3u, and VPc peptides by LC-MS. The $\pm$ values represent the SD from three replicate injections.

Peptide	Mean Purity (% of total)
VP1u	79.7 $\pm$ 1.1 %
VP3u	22.1 $\pm$ 0.5 %
VPc	86.9 $\pm$ 0.1%

5.4.6 Amino Acid Analysis of Natural Peptides

Amino acid analysis was conducted to determine the concentration of the natural peptides VP1u, VP3u, and VPc. **Figure 5.12** shows the measured nmol/g concentrations for each amino acid analysed per natural peptide. For VP1u, the measurement for glycine was 49.61 ± 1.54 nmol/g, for leucine 48.36 ± 0.80 nmol/g, for alanine 47.95 ± 0.37 nmol/g, and for proline 47.82 ± 0.27 nmol/g. For VP3u, the measurement for glycine was 37.67 ± 0.19 nmol/g, for alanine 34.83 ± 0.50 nmol/g, for valine 35.91 ± 0.28 nmol/g, and for proline 35.92 ± 0.18 nmol/g. For VPc, the measurement for valine was 32.8 ± 0.1 nmol/g, for lysine 33.02 ± 0.13 nmol/g, and for isoleucine 32.6 ± 0.43 nmol/g. As mentioned previously in Section 5.3.11, only the volatile amino acids of each peptide are measured by amino acid analysis using the GC-MS/MS method.

Amino acid analysis results for VP1u and VPc peptides were consistent across amino acids and replicates, measured with a CV of 2.2% for VP1u, and 0.9% for VPc. Slightly higher variability was observed for VP3u, with a CV of 3.0%. The measured concentrations (nmol/g) of each natural peptide stock were 1088.4 ± 23.9 for VP1u, 362.8 ± 11.0 for VP3u, and 299.6 ± 2.7 for VPc. The measured stock concentrations and corresponding purity percentages were used to calculate the actual stock concentrations (mg/mL) and total mg in the stock of the VP1u, VP3u, and VPc natural peptides (**Table 5.9**). The VP1u stock contained 2.25 ± 0.058 mg of VP1u peptide, the VP3u stock contained 0.38 ± 0.008 mg of VP3u peptide, and the VPc stock contained 1.79 ± 0.012 mg of VPc peptide. These corrected values were then used to accurately assign concentrations to each level of the standard curve, enabling more precise quantification of the endogenous rAAV9 peptides.

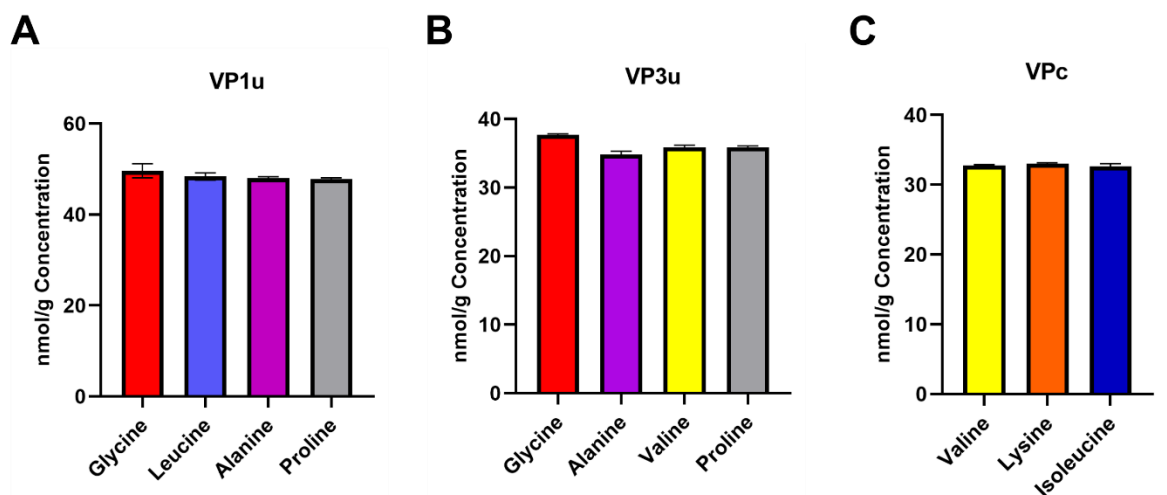


Figure 5.12 Amino acid analysis of VP1u, VP3u, and VPc natural peptides with GC-MS/MS MRM. (A) Amino acid analysis of VP1u, showing nmol/g concentrations of glycine (red), leucine (light blue), alanine (purple), and proline (grey). (B) Amino acid analysis of VP3u showing nmol/g concentrations of glycine (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.

Table 5.9 Mean concentrations (mg/mL and total mg) of natural peptide stock solutions, as determined by amino acid analysis and corrected for peptide purity. The $\pm$ values 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

5.4.7 Collision Energy Optimisation

Choosing a 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 5.13**), 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 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 5.10, Figure 5.13).

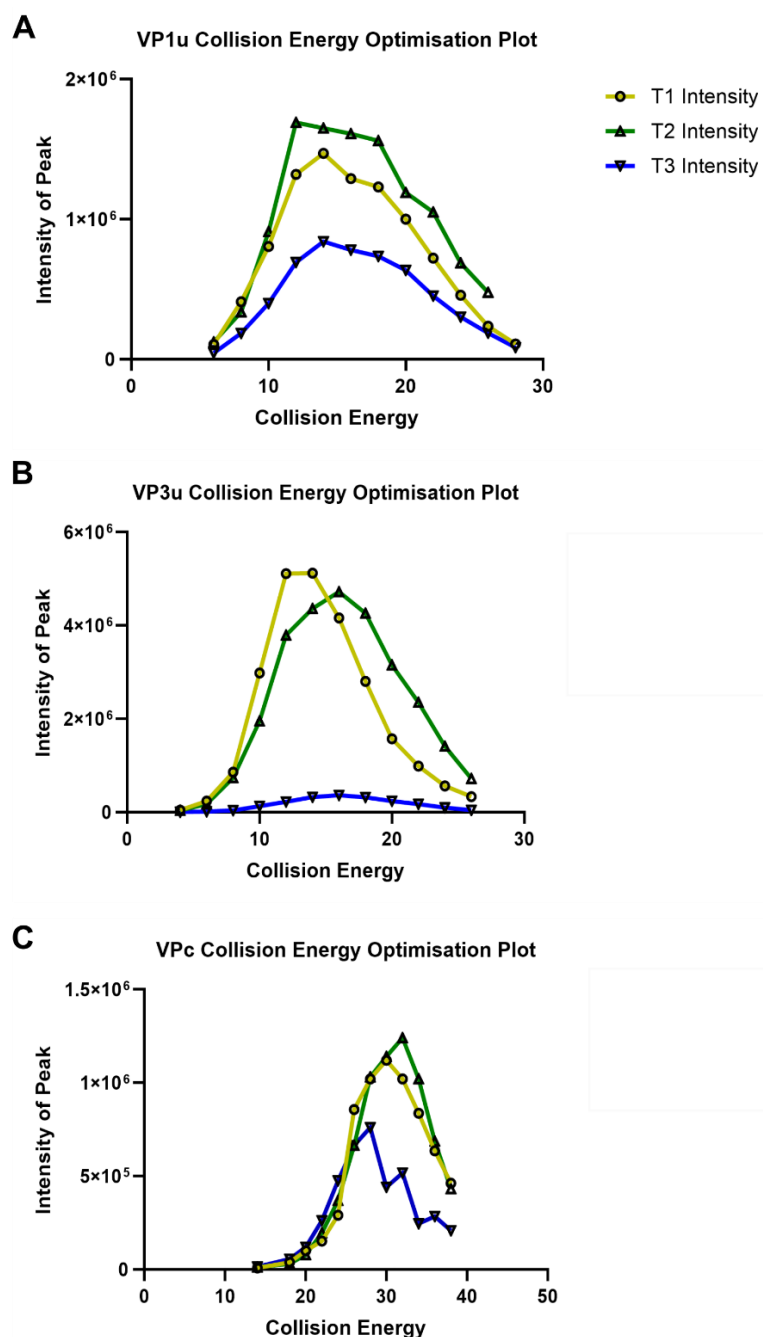


Figure 5.13 Collision energy (CE) optimisation for transitions of rAAV capsid peptides VP1u, VP3u, and VPc. Panels (A), (B), and (C) correspond to the CE optimisation plots for VP1u, VP3u, and VPc, respectively. Each plot depicts the intensity trends of three transitions (T1, T2, T3) across 12 tested CEs. The yellow line represents the optimisation 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.

Table 5.10 Selected transitions and corresponding optimal collision energies for the MRM method.

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 Labelled	652.81	633.20	672.25	785.32	14	14	14
VP3u	600.76	443.11	758.28	829.31	14	16	16
VP3u Labelled	606.77	443.11	770.29	841.33	14	16	16
VPc	861.90	1016.45	1217.55	870.35	30	32	28
VPc Labelled	869.42	1024.47	1225.54	877.43	30	32	28

5.4.8 MS Quantification of the VP1u, VP3u, VPc Peptides from rAAV9 Samples

For quantitative analysis, experiments were performed in triplicate, with each replicate involving independently generated standard curves. 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. The standard curves were plotted as the mean ratio of peak areas of the quantifier transition (three injections per standard curve) of each natural peptide to the isotopically labelled counterpart against the known concentration (nM) of the natural peptide (**Figure 5.14**). Additionally, a new rAAV9 digest was prepared for each experiment, ensuring that sample preparation was performed individually for every run. The LOD and LOQ were estimated for each peptide, using the Equations 5.2 and 5.3 for each experiment. The mean LOD and LOQ values, along with the standard deviation from three independent experiments, are reported in **Table 5.11**. Although the measured concentrations of VP3u and VPc were above or within the established LOQ range and can be considered reliable, the mean concentration of VP1u fell below the estimated LOQ range but remained above the LOD (**Table 5.11**).

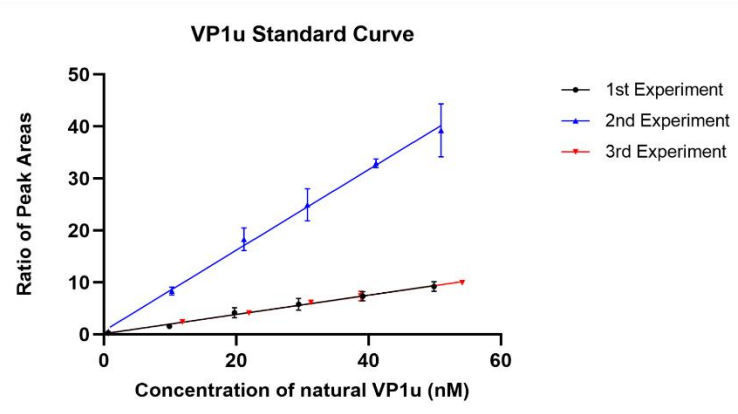
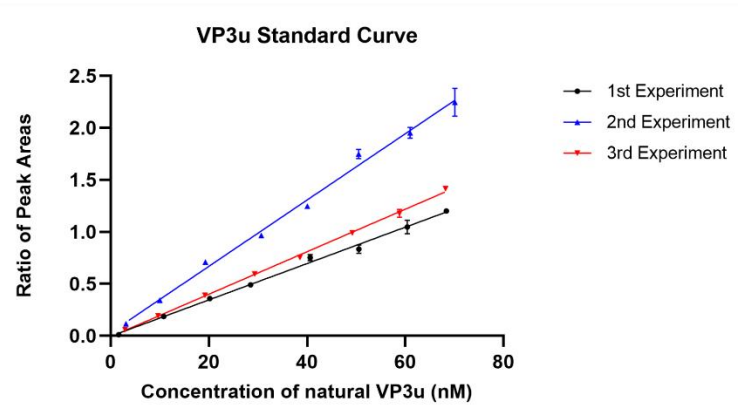
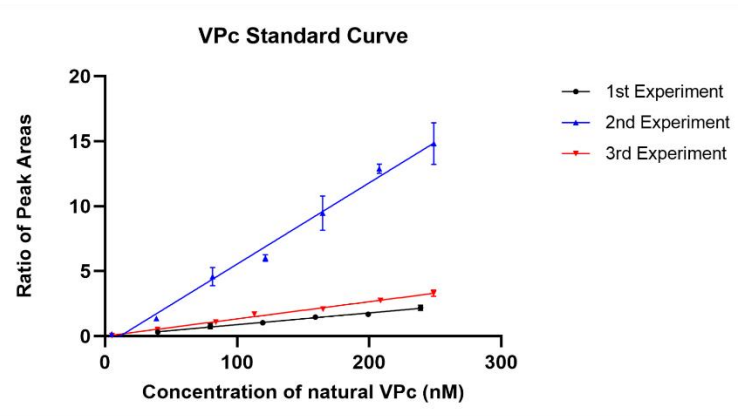
A**B****C**

Figure 5.14 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 natural and labelled peptides, while the x-axis indicates the concentration of the 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.

Table 5.11 Estimation of limit of detection (LOD) and limit of quantification (LOQ) for each signature peptide. The $\pm$ values correspond to the SD of three experimental replicates.

Peptide	Parameters	Result (nM)	Concentration in Digested Sample (nM)
VP1u	LOD	3.0 ± 0.9	$3.5 \pm 1.8 (< \text{LOQ})$
	LOQ	9.2 ± 2.7	
VP3u	LOD	3.2 ± 1.1	25.6 ± 5.9
	LOQ	9.7 ± 3.4	
VPc	LOD	16.8 ± 8.6	48.4 ± 6.8
	LOQ	50.9 ± 26.2	

The mean quantities of VP1u, VP2, VP3u, and VPc across the three experiments in the stock, along with their uncertainties, are reported in **Table 5.12**. 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. These values were calculated by multiplying the measured concentrations in the reaction (**Table 5.11**) by the dilution factor of 20 (Section 5.3.12). The mean stoichiometry of rAAV9 sample calculated from the three experiments was $1 \pm 0.5 : 5.5 \pm 2.9 : 7.3 \pm 4.2$ (VP1:VP2:VP3). The inter-experimental variability for VP2 was calculated using the error propagation formula and considering the correlation between peptides (covariance) (Equation 5.5), and it was only 3.3%. The combined variability for VP1u, VP3u, VPc was calculated by including the CV from three primary sources: the amino acid analysis, the purity assessment of each natural peptide, and the variability observed across the three experiments (Equation 5.6). 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, 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 CV of 52.6%, which could be attributed to the fact that the measured mean VP1u concentration was below the determined LOQ but within the LOD range (Snijder *et al.*, 2014; Wörner *et al.*, 2021).

Table 5.12 Measured mean stoichiometry and concentrations (nM) of VP1u, VP3u, and VPc peptides in rAAV9 stock. The $\pm$ values represent the combined variability calculated from 3 experimental replicates, amino acid analysis, and purity assessment of each natural peptide. For VP2, the error corresponds only to inter-experimental variability.

Stock	Mean VP1:VP2:VP3 ratio	VP1u (nM)	VP2 (nM)	VP3u (nM)	VPc (nM)
rAAV9	1 $\pm$ 0.5:5.5 $\pm$ 2.9:7.3 $\pm$ 4.2	69.7 $\pm$ 36.7	385.3 $\pm$ 12.8	512.9 $\pm$ 117.9	967.9 $\pm$ 135.7

The method was also used for the measurement of the total capsid content of the sample. The VPc peptide, which is present in all three rAAV9 capsid proteins and reflects the total capsid concentration, was determined to be present at a concentration of 1.67 ± 0.23 ng/ μ L (967.9 ± 135.7 nM) by the MS-based method. Assuming there are 60 copies of the VP proteins in a capsid, the MS-based concentration of the rAAV9 sample was calculated as $9.71 \times 10^9 \pm 1.36 \times 10^9$ capsids/ μ L. ELISA-based analysis of the rAAV9 sample gave a capsid concentration of $1.6 \times 10^{10} \pm 1.04 \times 10^9$ capsids/ μ L, as previously reported in Chapter 2, Section 2.4.3. This corresponds to $39.3 \pm 6.2\%$ relative difference between the mean values of ELISA and the MS-based approach described here.

5.4.9 Identification of Host-Cell Proteins in the rAAV8 Samples by MS Analysis

The rAAV8 and rAAV9 samples were analysed for the presence of host cell proteins (human proteins from HEK293 production cells). One hundred host cell proteins were identified in the trypsin-digested rAAV8 sample (**Appendix Table 7.1**), some of which have been previously reported by others (Aloor *et al.*, 2018; Rumachik *et al.*, 2020).

These HCPs are primarily involved in DNA packaging, gene regulation, chromatin structure, translation, signal recognition, RNA processing, and protein folding. All identified proteins, the number of unique peptides matched, $-\log P$ scores and coverage (%) are reported in **Appendix Table 7.1**. Many of the HCPs identified in this study have been previously reported in the literature (**Table 5.13**), highlighting their frequent occurrence in rAAV preparations and suggesting potential roles in rAAV biology or production processes. For the rAAV9 sample, no host cell proteins were detected, likely due to a more rigorous purification process. As shown in **Figure 5.3**, 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.

Table 5.13 Host cell proteins identified in this study (out of 100 total) that have also been reported in previous studies.

Protein Name	UniProt Accession No	Coverage %	-10logP score	Identified in Literature
40S Ribosomal protein S3	P23396	77	162	Leibiger, Min and Lee, 2024
40S Ribosomal protein S7	P62081	52	120.3	Leibiger, Min and Lee, 2024
40S Ribosomal protein S12	P25398	15	68.8	Leibiger, Min and Lee, 2024
40S Ribosomal protein S18	P62269	46	122.7	Leibiger, Min and Lee, 2024
40S Ribosomal protein S20	P60866	47	110.1	Leibiger, Min and Lee, 2024
60S Ribosomal Protein L29	P47914	14	93.5	Rumachik <i>et al.</i> , 2020
Serine/arginine-rich splicing factor 1	Q07955	49	149.5	Dong <i>et al.</i> , 2014
Splicing factor proline- and glutamine-rich	P23246	18	138.6	Leibiger, Min and Lee, 2024
Heterogeneous nuclear ribonucleoprotein A1	P09651	40	143	Leibiger, Min and Lee, 2024
Heterogeneous nuclear ribonucleoprotein A/B	Q99729	24	131.9	Leibiger, Min and Lee, 2024
Heterogeneous nuclear ribonucleoprotein U	Q00839	20	146.2	Leibiger, Min and Lee, 2024

Table 5.13 (continued)

Histon H2B type F	Q5QNW6	51	128.8	Leibiger, Min and Lee, 2024
Histon H4	P62805	50	100.6	Leibiger, Min and Lee, 2024
Nucleolin	P19338	32	189.5	Dong <i>et al.</i> , 2014; Leibiger, Min and Lee, 2024
Heat shock cognate protein 71kDa	P11142	49	185	Rumachik <i>et al.</i> , 2020; Leibiger, Min and Lee, 2024
Heat shock protein 70kDa 1A	P0DMV8	54	188.8	Rumachik <i>et al.</i> , 2020
Heat shock protein 70kDa 1B	P0DMV9	54	188.8	Leibiger, Min and Lee, 2024
Putative elongation factor 1-alpha-like 3	Q5VTE0	51	179.8	Rumachik <i>et al.</i> , 2020
Elongation factor 1-alpha 1	P68104	51	179.8	Leibiger, Min and Lee, 2024
Chromobox protein homolog 1	P83916	54	158	Leibiger, Min and Lee, 2024
Chromobox protein homolog 3	Q13185	65	154.8	Leibiger, Min and Lee, 2024
Polyadenylate-binding protein 1	P11940	19	146.6	Leibiger, Min and Lee, 2024
Y-box-binding protein 1	P67809	48	161.5	Leibiger, Min and Lee, 2024
Ubiquitin 40S ribosomal protein S27a	P62979	20	88.9	Leibiger, Min and Lee, 2024

5.5 Discussion

This chapter details 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. Furthermore, the study reports methods for the assessment of post-translational modifications, serotype identification, and the detection of host cell proteins, all of which are critical for ensuring the safety and efficacy of gene therapy products.

The MS-based method for measurement of the capsid stoichiometry overcomes the limitations of traditional methods such as SDS-PAGE and ELISA, providing a more detailed and reliable analysis of rAAV vectors. Currently, there is no standardised reference measurement method for quantifying individual rAAV capsid proteins. However, the method developed here ensures precision and higher-order reference measurement for capsid protein quantification. To support the accuracy of this method, well-defined rAAV standards suitable for absolute quantification are necessary. In this approach, data were traceable to the International System of Units (SI) via amino acid analysis of the natural peptides. This ensures data traceability and comparability across different laboratories. Moreover, the reproducibility and robustness of the method are ensured with the use of isotopically labelled peptides. These peptides allow for precise quantification and account for variability in sample preparation and analysis. 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.

A limitation of this method is its specificity to rAAV9, as the peptides used cannot be applied to quantify other serotypes. rAAV9 was selected because it is a well-characterised serotype with advantages such as efficient transduction of both neuronal and non-neuronal cells, and the ability to cross the blood–brain barrier (D. Liu *et al.*, 2021; Issa *et al.*, 2023). 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, including rAAV1, rAAV2, rAAV3, rAAV6, rAAV8, and rAAV10, providing a potential foundation for broader application.

The results reveal a measured ‘mean capsid’ stoichiometry of $1 \pm 0.5:5.5 \pm 2.9:7.3 \pm 4.2$ (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 (Wörner *et al.*, 2021). The observed ratio of 1:5.5:7.3 deviates from the widely assumed rAAV capsid stoichiometry of 1:1:10, showing a substantially higher proportion of VP2. Similar deviations have been reported by another study, which observed between 0 and 2 copies of VP1, 8 to 11 copies of VP2, and 48 to 51 copies of VP3 in rAAV1 capsids (Snijder *et al.*, 2014). Wörner *et al.*, 2021 also reported an approximate bulk stoichiometry of 3:10:47 in rAAV9 by LC-UV/MS. Higher expression levels of VP2 in the production cells compared to VP1 could explain the higher quantities of VP2 observed (Bosma *et al.*, 2018). However, one possible factor contributing to this discrepancy is the potential detection of free capsid proteins by the MS-based method. Since VP2 is estimated indirectly by subtracting the quantities of VP1u and VP3u from the total capsid protein (VPc), the presence of free VPc peptides in the sample could artificially inflate the calculated VP2 levels. Therefore, the influence of free, non-assembled capsid proteins on the measurements cannot be ruled out. Nonetheless, these findings are particularly significant, as variations in stoichiometry can directly impact the infectivity of rAAV vectors and therefore compromise the efficacy of the viral preparations (Bosma *et al.*, 2018; Onishi *et al.*, 2023; Erickson *et al.*, 2024). Identifying and controlling these variations is therefore important for ensuring the reliability and effectiveness of rAAV-based therapies.

Moreover, given the observed variability, the method is best suited for detecting significant differences between rAAV9 production batches when those differences exceed the inherent inter-experimental variability. Large variations in VP1, VP2, VP3 quantities or stoichiometric ratios, often indicative of differences in manufacturing conditions, are detectable even with the added inter-experimental variability. However, minor differences may be masked. This highlights the need for standardisation when applying the method for comparative analyses, particularly for resolving smaller discrepancies.

Here, although the concentrations of VP3u and VPc were sufficiently high to allow reliable quantification (i.e. above the LOQ), the measured concentration of VP1u was below the LOQ but remained above the LOD (**Table 5.11**). This explains the high inter-experimental variability observed for VP1u (CV = 52.6%). Further optimisation of the method could involve

selecting an alternative VP1u signature peptide in future work. The current peptide may not be efficiently released from the capsid during enzymatic digestion, potentially due to restricted accessibility caused by capsid conformation or PTMs. Additionally, poor ionisation efficiency of the selected peptide may contribute to increased variability in signal intensity and reduced quantification reliability. If peptide selection and digestion optimisation do not resolve this issue, it may reflect a true low abundance of VP1, in which case the minimum required concentration of the sample should be defined to ensure that all signature peptides fall above the LOQ for reliable quantification.

Additionally, there was good agreement between the 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 (Adachi *et al.*, 2014; Emmanuel *et al.*, 2022). 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 antibody's binding site and the location of the VPc peptide are in close proximity. The observed 39.3% relative difference between the mean values of the MS-based approach and ELISA can be attributed to the fundamental differences between these two methods. The MS-based method directly quantifies specific ions, providing precise measurements of 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, partial 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 (Higgins *et al.*, 1998; Verch, Roselle and Shank-Retzlaff, 2016). These differences underscore the importance of considering method-specific biases and assumptions when comparing results from ELISA and MS-based quantification.

An essential first step in developing the quantification method is achieving high sequence coverage and identifying potential peptides suitable for quantification. Glu-C was used to generate unique peptides for use in the quantification method. 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 5.1**) (Kiyonami and Domon, 2010). Peptides that are unique to VP1 can be found in the N-terminal VP1 unique (VP1u) region of VP1, but this becomes more challenging for the VP2 and VP3 proteins, as their sequences are also present in VP1. Thus, there is only one unique peptide that could be generated for each of VP2 or VP3 proteins for protein quantification purposes.

A limitation of this method was the inability to identify a unique peptide for VP2 (VP2u). Various approaches were explored, including the use of different columns, modifications to the chromatography conditions, and alternative proteases such as AspN instead of Glu-C, but none were successful in identifying VP2u. This may be due to poor ionisation efficiency of the VP2 peptide, an undetected modification altering its mass from the theoretical m/z value, or limited accessibility of the Glu-C cleavage site due to the local capsid structure, potentially influenced by glycosylation near that site. However, although previous studies have reported potential glycosylation sites on rAAV9 capsid proteins, none are located near the VP2 region. The five predicted N-glycosylation sites from the Zhou *et al.*, 2024 study include N14, N262, N337, N452, and N497 (Zhou, Priya and Ong, 2024).

This study also revealed deamidation sites in both rAAV8 and rAAV9 VPs, which should be monitored closely to evaluate their impact on vector stability and functionality. The ability to detect a 1 Da mass difference, indicative of deamidation, is necessary for such studies. As such, the high-resolution capability of the Orbitrap MS for the characterisation of rAAVs is considered ideal. Interestingly, the deamidations observed in the rAAV8 sample reported here at N57, N514, and N540, have previously been reported, suggesting that these residues are inherently prone to deamidation, likely due to the accessibility, specific amino acid sequence and/or 3D-protein structure surrounding these three asparagine residues (Giles *et al.*, 2018). Consequently, it is suggested that these three Asn residues should be closely monitored for deamidation as a CQA during production, storage and aging of rAAV8 viral vectors to ensure their stability and efficacy. When rAAV9 capsid VPs were analysed, deamidation was observed

at N254, N303 or N304, and N512. The deamidation at N512 has also been previously reported in the literature (Zhou and Wang, 2023). Once again, it is suggested that these residues should be monitored for deamidation as part of characterising CQAs.

Furthermore, 100 distinct HCPs were identified in rAAV8 samples, showing the complexity of HCPs in rAAV production. Of these, 24 have been previously reported in the literature (**Table 5.13**) (Nash *et al.*, 2009; Dong *et al.*, 2014; Rumachik *et al.*, 2020; Leibiger, Min and Lee, 2024). 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 (Bracewell *et al.*, 2021). Although various mechanisms by which HCPs can be incorporated into rAAV preparations have been described in the introduction of this chapter, it is likely that their presence in this case could also be due to non-specific binding to the affinity ligand during affinity chromatography. Notably, some of these HCPs have been shown to interact with the rAAV Rep proteins, as reported by Nash *et al.*, 2009. These HCPs include ADP-ribosylation factor 1, nucleolin, RNA-binding protein FUS, splicing factors serine/arginine-rich splicing factor 1 and 2, heterogeneous nuclear ribonucleoprotein U, Lupus La protein, and histone H1.2. Further analysis of rAAV HCPs will help optimise processes to remove HCPs, especially those that may pose risks to patients

In summary, the MS-based method described in this chapter provides a robust framework for the comprehensive characterisation of rAAV vectors. This method addresses gaps in current analytical techniques and supports the development of standardised reference measurement methods. Additionally, the method delivers precise measurement of total capsid content and provides a greater depth of information than traditional ELISA capsid titre measurements. The method has the potential to be further developed as a reference method to standardise measurements from field methods and value assign reference materials. Future research and inter-laboratory studies are essential to further validate these findings, refine the methodology, and explore its application to other rAAV serotypes. The study also reports HCPs identified in the rAAV8 sample, consistent with findings previously described in the literature. These results demonstrate the utility of mass spectrometry for precise rAAV characterisation and for supporting the standardised measurement of rAAV drug products.

Chapter 6: Overall Discussion

The findings presented in this thesis describe the development of advanced analytical methods for the accurate measurement of rAAV CQAs. In particular, the work focused upon genome titre, genome integrity, capsid titre and capsid stoichiometry analysis. Firstly, with regard to genome titre, this is an important CQA as it is essential for determining the appropriate dose level. Accurate quantification is particularly important given reports of severe toxicity in animal models following the administration of high doses of rAAV products (Hinderer *et al.*, 2018). However, the measurement of genome titre lacks standardisation, as different laboratories use different methods such as qPCR and dPCR, leading to variability in results between laboratories.

When the genome titre results between dPCR and qPCR were compared in Chapter 3, dPCR measurement resulted in higher titres than qPCR for all samples measured, including trFragment, rAAV2, rAAV2RefMat, and rAAV9 (**Figure 3.6**). An important consideration from this is that direct comparison between qPCR and dPCR is not straightforward when using the ITR assay, as dPCR cannot distinguish between molecules containing one ITR or both ITRs (Suoranta, Laham-Karam and Ylä-Herttuala, 2021). In both cases, the molecules are detected as positive partitions. To enable a more accurate comparison, an approach used in this study involved digesting the recombinant AAV2 genome, as well as the trFragment control, with the restriction enzyme *MspI* to separate the two ITRs into different partitions and enable total ITR quantification. However, digestion of trFragment and rAAV2 resulted in a 48.7% and 60.0% increase in ITR concentration, respectively. Therefore, further optimisation of the digestion process is necessary to enable complete measurement of total ITR content.

After *MspI* digestion, the copy number concentration measured by dPCR was $108.1 \pm 12.2\%$ higher than that measured by qPCR for the trFragment, and $122.7 \pm 13.8\%$ higher for rAAV2 (**Figure 3.6**). Without *MspI* digestion, the dPCR results were $40.0 \pm 13.5\%$ higher than qPCR for trFragment, $39.2 \pm 16.9\%$ for rAAV2, $45.3 \pm 8.7\%$ for rAAV2RefMat, and $23.9 \pm 14.0\%$ for rAAV9 (**Figure 3.6**). This difference may be attributed to the presence of potential inhibitors introduced during sample preparation of the rAAV samples and during the production of the trFragment (Dingle *et al.*, 2013; Lock *et al.*, 2014; Sedlak, Kuypers and Jerome, 2014;

Devonshire *et al.*, 2016). It is hypothesised that the digestion step performed to generate the trFragment, which involves the addition of the PaeI restriction enzyme and CutSmart® buffer, as well as the lysis buffer used to release genomes from rAAV capsids, may negatively impact amplification efficiency in qPCR. In contrast, dPCR is less affected by such inhibitors as it relies on end-point detection rather than amplification kinetics. Other factors, such as buffer composition, threshold settings, and analysis software, may also influence the results (Hays *et al.*, 2022).

Since this difference between the two methods was observed across all samples, there may be a systematic bias between qPCR and dPCR. However, a previous interlaboratory study that determined the titre of rAAV2RefMat using qPCR with an SV40 polyA assay across 16 laboratories reported an interlaboratory variability of 77.8% (Lock *et al.*, 2010). In that study, reported values ranged from 4.24×10^9 to 1.27×10^{11} viral genomes/mL, which corresponds to a 30-fold difference (Lock *et al.*, 2010). Therefore, it is difficult to conclude whether the observed difference between qPCR and dPCR in this study reflects a true systematic bias, as substantial variability has been observed using the same qPCR assay and protocol across different laboratories.

When the trFragment was tested across different dPCR platforms (QX200, QIAcuity, and Absolute Q), each with its own setup and underlying technologies, a variability of only 14.8% was observed. However, it is important to note that the dPCR platform comparison was performed using the trFragment and did not involve genome extraction. In a study conducted by Wang *et al.*, 2024, where intact rAAV particles were loaded directly onto four dPCR platforms (QX200, Naica, Absolute Q, and QIAcuity) the variability ranged from 23.5 to 34.2% for rAAV5, and from 11.6 to 19.9% for rAAV8. This variability is higher compared to inter-platform measurements using already extracted rAAV genomes (CV of 4.73-8.50%) (Wang *et al.*, 2024). Accurate estimation of partition volumes is critical when comparing results across different platforms, as deviations between the measured and actual partition volumes can lead to discrepancies in genome titre calculations (Emslie *et al.*, 2019). Therefore, future studies that accurately measure the partition volume of each dPCR platform are essential for reliable cross-platform comparisons.

The high reproducibility of dPCR has been demonstrated previously in interlaboratory studies, showing an interlaboratory variability of $\leq 16\%$ for the measurement of human

cytomegalovirus (HCMV) when the same extraction method was used (Milavec *et al.*, 2022). In comparison, a variability of 77.8% was reported across 16 laboratories measuring the titre of the rAAV2RefMat using the same qPCR protocol (Lock *et al.*, 2010). These findings highlight the important role that dPCR can play in value assigning reference materials and enabling reliable interlaboratory comparisons. This comparability can be further improved if laboratories use the same dPCR platform and dPCR assay, thereby minimising variability. However, to enable more accurate comparisons, an interlaboratory study measuring the genome titre of rAAV reference material, focusing on a region of the recombinant genome rather than the ITRs to avoid associated biases, should be conducted using both qPCR and dPCR to directly compare the precision of the two methods.

Evaluating the analytical performance specifications of the dPCR assays is essential before they can be used to characterise rAAV material intended for therapeutic applications. These specifications include specificity, linearity, precision, bias, and sensitivity. In this study, the linearity of each assay (ITR, CMV, EGFP, WPRE, bGH polyA) was assessed using the trFragment and demonstrated strong performance, with $R^2 > 0.99$ for all assays across the tested concentration range, which spanned from 4,568 copies/ μ L to 155 copies/ μ L (mean across assays) (**Figure 3.3**). The assays also showed good inter-assay precision, with variability of less than 15% at all tested concentrations and less than 10% for concentrations ≥ 350 copies/ μ L. With the ongoing development of dPCR platforms and nanoplates that support higher numbers of partitions, the dynamic range of the dPCR assays can increase, allowing quantification across additional orders of magnitude.

For evaluation of the precision of each assay in measuring genome titres, the intra-assay variability was estimated using rAAV2, rAAV8, and rAAV9 samples at concentrations of approximately 500 copies/ μ L (**Table 3.9**). The results showed that assay precision can vary substantially between different sample types. For example, the intra-assay variability of the ITR assay was only 2.7% for rAAV9 but increased to 18.0% for rAAV2 and 32.0% for rAAV8. Similar variability patterns were observed for the other assays tested, including eGFP, WPRE, bGH polyA, and CMV. These findings indicate that an assay may meet commonly accepted precision criteria for one sample but fail for another. Therefore, applying a fixed variability threshold for all assays may not be appropriate. Instead, each assay should be evaluated

across multiple rAAV samples. Assays that demonstrate consistently high precision across different samples should be prioritised for further development and validation.

To evaluate the specificity of each assay, all primers and probes were assessed using the NCBI BLAST tool. Generally, assays targeting regions with high similarity to the human genome, such as the bGH polyA assay, should be avoided when analysing rAAV products intended for therapeutic use, as they may amplify residual host cell genomic DNA due to sequence homology (in this case between bovine and human). While *in silico* screening provides a useful preliminary assessment, a more rigorous experimental evaluation is recommended for future studies. This could involve testing each assay against extracted host cell genomic DNA from the production cell line (HEK293). Three samples would be analysed: 1) the rAAV sample as the control, 2) the host cell genomic DNA to determine whether any non-specific amplification occurs, and 3) the rAAV sample spiked with host cell genomic DNA to evaluate whether the presence of DNA impurities affects the assay result. A no-template control would also be included to monitor for background contamination in the dPCR reactions.

For bias assessment, it is essential to have an accepted reference value against which measurements can be compared. In this study, the reference value was derived from the interlaboratory study by Lock *et al.*, 2010, which measured the genome titre of rAAV2RefMat using qPCR. When rAAV2RefMat was measured using dPCR targeting the SV40 polyA region, which is the same assay used in the interlaboratory study, the measured value was 37.0% lower than the lower limit of the reported 95% confidence interval (2.7×10^7 to 4.75×10^7 copies/ μ L) (**Figure 3.6**) (Lock *et al.*, 2010). However, it remained within the overall reported range (mean $\pm$ standard deviation) of $3.82 \times 10^7 \pm 2.97 \times 10^7$ copies/ μ L (Lock *et al.*, 2010). When rAAV2RefMat was measured using dPCR targeting the ITRs, the measured titre was within the 95% confidence interval (**Figure 3.6**). In contrast, qPCR targeting the ITR region produced a titre that was $19.9 \pm 2.3\%$ lower than the lower limit of the interval (**Figure 2.5**). This suggests that, in this study, dPCR provided a titre measurement more consistent with the assigned value of rAAV2RefMat compared to qPCR (Lock *et al.*, 2010).

The estimation of sensitivity can be included for assays intended for regulatory compliance. However, the FDA specifically mentions assay sensitivity only in the context of vector persistence in biodistribution studies for gene therapy products, recommending a

sensitivity of ≤ 50 vector copies/ μg of gDNA (U.S. Food and Drug Administration, 2020b). In contrast, rAAV is typically produced at high genome titres, making the sensitivity of a dPCR or qPCR assay a less important parameter for routine batch analysis.

An important observation of this study was that target region selection plays an important role in dPCR. Since dPCR cannot differentiate between molecules containing one ITR or both ITRs, a ratio of 1:1 between ITRs and any other target region would be expected if all the genomes analysed were full-length. In contrast, consistently higher genome titres were reported when targeting the ITRs compared to other regions such as *eGFP*, the CMV enhancer, *RTN3*, WPRE, bGH polyA, and SV40 polyA across all tested rAAV samples (**Figure 3.9**). This observation is likely due to the high abundance of partial genomes that contain only the ITR (Tran *et al.*, 2022; Hashiba *et al.*, 2024). These may arise from the packaging of ITR-containing fragments as the ITR serves as the packaging signal in AAV, and from incomplete packaging events that terminate shortly after the ITR is incorporated (Tran *et al.*, 2022; Hashiba *et al.*, 2024). This hypothesis is further supported by nanopore sequencing data from rAAV9 reported in Chapter 4, which showed substantially higher sequencing depth in the 5' and 3' ITRs compared to the rest of the genome. Therefore, it is recommended that the genome titre of rAAV samples be measured using at least two assays: one targeting a region within the transgene cassette, such as the gene of interest, and another targeting the ITRs. This dual-assay approach is important because if only a GOI-targeting assay is used, the presence of ITR-only partial genomes is overlooked. As a result, dosing based solely on GOI quantification may lead to the unintended administration of high amounts of ITRs, which could potentially contribute to adverse effects in patients. However, further studies are needed to determine whether high amounts of ITRs may have harmful effects on patients.

Ensuring proper dosing also requires assessment of the genome integrity of the rAAV samples. Partial genomes present in the preparations may contribute to immune responses or potentially lead to the expression of truncated proteins or other aberrant gene products (Schnödt *et al.*, 2016; Kontogiannis *et al.*, 2024). In this thesis, a 4-plex dPCR method was established to assess the genome integrity of rAAV samples. To validate these results, the genome integrity measurements of rAAV9 from the 4-plex dPCR assay were compared to data obtained from nanopore sequencing, which served as an orthogonal method. Only the rAAV9

sample was analysed by nanopore sequencing, as the rAAV2, rAAV8, and rAAV2RefMat samples did not meet the input requirements for nanopore sequencing ($\sim 10^{10}$ viral genomes).

It was observed that dPCR analysis resulted in significantly higher genome integrity values compared to nanopore sequencing (**Figure 4.10**). As previously discussed in Chapter 4, this discrepancy arises because dPCR tends to overestimate genome integrity, as it cannot distinguish between molecules containing only a single ITR and those with both ITRs (Suoranta, Laham-Karam and Ylä-Herttuala, 2021). Additionally, dPCR could not identify deletions within the CamKII α /RTN3 region. In contrast, nanopore sequencing tends to underestimate genome integrity due to a short-read bias, where shorter fragments are overrepresented in the sequencing output as they are sequenced more efficiently than longer molecules within the same timeframe (Dunker-Seidler *et al.*, 2025). A potential solution to mitigate this issue is to use SMRT sequencing instead of nanopore sequencing for genome integrity analysis, as it has been previously shown that SMRT sequencing may be less affected by short-read bias (Dunker-Seidler *et al.*, 2025).

Since both multiplex dPCR and long-read sequencing methods have their own biases, with multiplex dPCR overestimating and long-read sequencing underestimating genome integrity, it is important to determine which method more accurately reflects the true genome integrity of rAAVs. To achieve this, additional higher-order orthogonal techniques are needed. Charge Detection Mass Spectrometry (CDMS) has previously been used to characterise the genome integrity of rAAVs and can serve as an orthogonal technique for characterising the genome integrity of reference material rAAVs (Barnes *et al.*, 2023). These reference materials can then be used to compare long-read sequencing and multiplex dPCR, identify which method is closer to the true value, and help optimise and validate both approaches. Furthermore, a more detailed assessment of genome integrity could be performed by mixing an rAAV sample with high genome integrity with increasing amounts of truncated genomes. This would allow for evaluation of how the measurements from dPCR and long-read sequencing are affected across different sample compositions and how well their results correlate with CDMS.

An advantage of multiplex-dPCR over long-read sequencing methods is its requirement for minimal input material. While nanopore sequencing typically requires around 10^{10} viral genomes, and SMRT[®] sequencing requires 10^{12} – 10^{13} viral genomes, dPCR can operate with

much lower quantities (Dunker-Seidler *et al.*, 2025). In addition to its low input requirements, dPCR is a higher-throughput technique and benefits from simple sample preparation, lower cost, and simpler data analysis, particularly now that statistical models are available to accurately assess multiple-target linkage (Regan *et al.*, 2015; Gleerup *et al.*, 2023; Duong *et al.*, 2025). These features make dPCR an attractive approach for routine genome integrity analysis of rAAV samples, complementing sequencing-based methods. dPCR can be particularly applied throughout the manufacturing process to monitor potential changes in genome integrity, while nanopore sequencing can be used during early development to confirm the vector genome sequence and integrity. It can then be performed again at later stages, such as when dPCR indicates a substantial deviation, or prior to final batch release, to confirm that no changes have occurred in the sequence or integrity of the genome during manufacturing.

Another important CQA for ensuring the safety of gene therapy is the purity of the rAAV product. Residual DNA impurities originating from host cells or production plasmids pose a significant safety risk, as they may trigger immune responses in patients (Schnödt *et al.*, 2016; Brimble *et al.*, 2022; Kontogiannis *et al.*, 2024). To mitigate this, the FDA recommends limiting residual DNA to less than 10 ng per dose and ensuring that any remaining fragments are shorter than 200 nt (U.S. Food and Drug Administration, 2020a; Dunker-Seidler *et al.*, 2025). This size restriction reduces the likelihood that the DNA could be transcriptionally active or participate in recombination events. Moreover, it is important to assess whether the identified fragments contain transcriptionally active sequences that could potentially pose a safety risk to patients. For example, this includes screening for adenoviral E1 sequences if HEK293 cells were used as the production cell line, SV40 large T antigen and E1 sequences if HEK293T cells were used, and E6/E7 sequences from human papillomavirus (HPV) if HeLa cells were used (U.S. Food and Drug Administration, 2020a). In Chapter 4, $96.5 \pm 0.5\%$ of the residual DNA impurities for rAAV9 PureLink and $94.3 \pm 6.7\%$ for the rAAV9 Heat exceeded 200 nt in size (**Figure 4.12**). To reduce the size of DNA impurities to below 200 nt, a more efficient DNase I digestion should be performed prior to capsid lysis. This may involve optimising reaction conditions or using DNase I enzymes and buffers from alternative manufacturers.

Besides DNA impurities, protein impurities such as HCPs can also be present in rAAV preparations. While ELISA is commonly used to quantify the total amount of HCPs, it does not

provide information about their identity. Nevertheless, ELISA should still be performed routinely to monitor HCP levels throughout downstream processing. Identifying HCPs is important, as certain HCPs carry a higher risk of causing immunogenicity than others (Jones *et al.*, 2021). For this reason, mass spectrometry-based approaches should be employed to identify individual HCPs. This study aimed to develop an MS method to identify HCPs and determine which HCPs are consistently observed across different studies. Identifying these commonly occurring HCPs can support future efforts to assess their immunogenicity risk and contribute to improving the safety of gene therapies.

The amount and identity of HCPs are highly dependent on the purification method used. In Chapter 5, no HCPs were identified by LC-MS/MS in the rAAV9 sample produced and purified commercially by VectorBuilder Inc. In contrast, over 100 HCPs were identified in an rAAV8 sample purified using affinity chromatography (**Appendix Table 7.1**). These HCPs were primarily associated with biological functions such as DNA packaging, gene regulation, chromatin structure, translation, signal recognition, RNA processing, and protein folding. Of the 100 HCPs identified, 24 had previously been reported in the literature (Dong, Nakai and Xiao, 2010; Rumachik *et al.*, 2020; Leibiger, Min and Lee, 2024).

In silico analysis tools, such as the Immune Epitope Database (IEDB), have been used to predict HCPs with a high potential for immunogenicity by evaluating their likelihood to bind to major histocompatibility complex (MHC) Class II alleles (Jawa *et al.*, 2016). If an HCP is detected and is either predicted to pose a high immunogenicity risk *in silico* or has been shown in the literature to elicit immune responses, it is important to measure its concentration and monitor it throughout downstream processing (Jones *et al.*, 2021). In such cases, an *in vitro* Comparative Immunogenicity Assessment (IVCIA) assay should also be performed to further evaluate its risk (Jones *et al.*, 2021). Of the 100 HCPs reported in this study, none, to my knowledge, have been linked to a high immunogenicity risk in the literature (Jawa *et al.*, 2016). For example, heat shock cognate 71 kDa protein and elongation factor 1 alpha 1 have previously shown IEDB scores of 26 and 21, respectively, which are considered low (< 100) (Jawa *et al.*, 2016).

Overall, it is important to monitor HCPs with high immunogenicity potential, particularly those that appear consistently and in high abundance across different rAAV preparations and studies. This includes assessing which purification methods can efficiently

remove them and understanding the mechanisms by which these HCPs are retained in the final product. Mass spectrometry enables the identification of high-risk HCPs, supporting the development of targeted control strategies for their removal. This contributes to process optimisation and helps minimise the risk of adverse effects in patients.

Mass spectrometry-based approaches are also effective for characterising the capsid of rAAVs. In this study, an LC-MS/MS method was developed to measure total capsid proteins and determine capsid stoichiometry. An rAAV capsid is typically described as consisting of 60 total copies of the capsid proteins VP1, VP2, and VP3. It has been widely assumed that the capsid stoichiometry is 1:1:10 (VP1:VP2:VP3), corresponding to 5 copies of VP1, 5 of VP2, and 50 of VP3 per capsid (Wörner *et al.*, 2021). However, the measured stoichiometry of rAAV varies substantially between different serotypes, production batches, and published studies (Snijder *et al.*, 2014; Bosma *et al.*, 2018; Wörner *et al.*, 2021; Onishi *et al.*, 2023). In Chapter 5, an average rAAV9 capsid stoichiometry of 1:5.5:7.3 is reported (**Table 5.12**). Previous studies have suggested that the stoichiometry of the capsids directly depends on the expression levels of each individual capsid protein (Snijder *et al.*, 2014; Bosma *et al.*, 2018; Wörner *et al.*, 2021). Therefore, the low VP1:VP2 ratio observed may be attributed to high expression levels of VP2. Snijder *et al.*, 2014 also observed low VP1:VP2 ratios in AAV1 capsids, reporting 0 - 2 copies of VP1, 8 - 11 copies of VP2, and 48 - 51 copies of VP3 by using an Orbitrap-based native MS approach. Similarly, Wörner *et al.*, 2021 reported a ratio of expression levels of 6% for VP1, 16% for VP2, and 78% for VP3, corresponding to an approximate bulk stoichiometry of 3:10:47 in rAAV9. The observed low VP1:VP2 ratio and reduced VP3 levels (1:5.5:7.3) may have functional implications. Bosma *et al.*, 2018 demonstrated that a low VP1:VP2 ratio (< 0.5) is associated with reduced rAAV potency, while lower VP3 levels are associated with increased vector potency. Similarly, Onishi *et al.*, 2023 generated rAAVs with a balanced VP1:VP2 ratio (~1:1) but varying levels of VP3. Vectors with a (VP1+VP2)/VP3 ratio of 0.32 demonstrated higher transduction efficiency compared to those with a ratio of 0.25. Therefore, these studies suggest that increasing the levels of VP1 and VP2, while keeping the VP1:VP2 ratio approximately 1:1, can enhance vector potency. This is because sequences that are present in the N-terminal of VP1 and VP2 proteins are necessary for endosomal trafficking, nuclear localisation, and genome release (Girod *et al.*, 2002; Bleker, Sonntag and Kleinschmidt, 2005; Sonntag *et al.*, 2011; Venkatakrishnan *et al.*, 2013; Onishi *et*

al., 2023). It would thus be interesting to evaluate how the rAAV9 material tested in this study, which exhibited a 1:5.5 VP1:VP2 ratio but reduced VP3 levels, performs in a transduction efficiency assay.

These findings highlight the importance of measuring capsid protein stoichiometry across different production batches to identify potential variations and assess their impact on vector potency. Such analysis can support optimisation of VP1, VP2, and VP3 expression during production to achieve a stoichiometry that maximises vector potency. The SID-MS approach developed here is designed to support capsid protein stoichiometry measurements. In this method, the quantities of VP1, VP2, VP3, as well as the total amount of capsid proteins, were traceable to the SI by measuring the natural peptide calibrators with amino acid analysis. This enables comparability of measurements between laboratories and supports the standardisation of protocols. Additionally, further validation of the method would benefit from the development of a well-characterised reference material with known stoichiometry as well as an interlaboratory study using the rAAV9 material analysed here. In addition, comparison with alternative techniques, such as high-resolution LC-UV/MS, would help confirm the accuracy and reliability of the results. Ultimately, CQA assessment could greatly benefit from expanding the SID-MS approach to measure the stoichiometry of other rAAV serotypes. Although the VP1u peptide can be used for quantifying the VP1 protein in multiple serotypes such as rAAV1, rAAV2, rAAV3, rAAV6, rAAV8, rAAV9, and rAAV10, serotype-specific VP3u and VPc peptides may be required due to differences in the amino acid sequences of the capsid proteins between serotypes in these regions.

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 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.

Further optimisation of the technique could be achieved by directly quantifying VP2. However, a unique signature peptide for VP2 was not identified in this study, likely due to several contributing factors. Firstly, the VP2-specific peptide may not ionise efficiently and thus was not detected by MS/MS data-interpretation software. Additionally, Tiambeng *et al.*,

2025 reported two major VP2 proteoforms, a phosphorylated and a non-phosphorylated form. Phosphorylation of VP2 at a specific conformational site could block protease digestion and thus block the release of the VP2 N-terminal peptide from the capsid. Furthermore, in the same study, they identified proteoforms of the VP2 N-terminal lacking alanine and proline residues. These variants complicate the identification of a VP2 unique peptide due to altered mass and fragmentation patterns. Future work could explore these factors further to determine whether a VP2 unique peptide can be reliably identified in rAAV9 or other serotypes.

An important part of this thesis was the development of a method for reliable identification of deamidation sites in rAAV samples and comparison of the findings with those from previous studies. This comparison helps determine which sites are prone to deamidation and should be monitored for effects in vector infectivity. In Chapter 5, deamidation sites were identified in both rAAV8 and rAAV9 samples. Although deamidation is not currently recognised as an established CQA, the FDA recommends providing information on PTMs (U.S. Food and Drug Administration, 2020a). Among these, deamidation is particularly important as previous studies have shown it can reduce vector transduction efficiency by more than 10 times (Giles *et al.*, 2018). The deamidation sites identified in the rAAV8 sample in this study have also been reported in a previous study, suggesting they may be consistent across rAAV8 production batches (Giles *et al.*, 2018). Interestingly, all three deamidation sites identified in rAAV8 are followed by a glycine residue. Generally, asparagine residues followed by glycine, histidine, aspartic acid, or serine residues are prone to deamidation (Wakankar and Borchardt, 2006). Similarly, the N254 deamidation site identified in rAAV9 is followed by a histidine residue, while the deamidation site at N512 is followed by a glycine residue. The N512 deamidation site in rAAV9 has also been previously reported in a study by Zhou and Wang, 2023. Therefore, it is recommended that these sites should be closely monitored to investigate their potential impact on vector potency and evaluate whether factors such as formulation pH or storage conditions (such as temperature) influence their occurrence (Jiang and Dalby, 2023). Generally, a formulation pH of 5 - 6 is considered suitable to minimise deamidation while avoiding low pH (< 5), which could lead to vector instability and aggregation (Wakankar and Borchardt, 2006).

In summary, the findings of this thesis, together with the methods developed for genome and capsid characterisation of rAAVs, will contribute to a more accurate and comprehensive assessment of CQAs and support the evaluation of whether gene therapy products meet regulatory specifications. Specifically, the novel MS-based approach developed for measuring capsid stoichiometry enables more accurate and precise measurements compared to existing analytical techniques. This method addresses the limitations of current approaches such as SDS-PAGE, which is only semi-quantitative and lacks accuracy, as well as chromatography-based methods (e.g., LC or SEC coupled to detectors), which are susceptible to issues such as peak tailing, overlapping peaks, and differential analyte recovery during LC (Tiambeng *et al.*, 2025). Additionally, the development of a multiplex dPCR method for assessing genome integrity and its comparison to nanopore sequencing, provides a valuable step towards improved measurement of this CQA. Although both techniques have inherent biases, nanopore sequencing enables comprehensive analysis of the entire rAAV genome and allows identification and assessment of the size of residual DNA impurities, thereby offering a greater depth of information. Therefore, nanopore sequencing is recommended for in-depth characterisation of rAAV preparations. Moreover, for genome extraction prior to nanopore sequencing, a heat-based method is recommended over PureLink-based extraction, as the heat-extracted sample contained substantially fewer DNA impurities. Finally, for quantification of genome titre, dPCR is recommended over qPCR due to its superior precision. Overall, the development of analytical methods for accurate measurements of CQAs supports product characterisation, improves comparability between studies and batches, helps the optimisation of manufacturing workflows, and ultimately leads to the production of safer, more consistent, and more effective gene therapies.

References

- Adachi, K. *et al.* (2014) 'Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing', *Nature Communications*, 5(1), p. 3075. Available at: <https://doi.org/10.1038/ncomms4075>.
- Aloor, A. *et al.* (2018) 'Site-Specific N-Glycosylation on the AAV8 Capsid Protein', *Viruses*, 10(11), p. E644. Available at: <https://doi.org/10.3390/v10110644>.
- Aronson, S.J. *et al.* (2019) 'Prevalence and Relevance of Pre-Existing Anti-Adeno-Associated Virus Immunity in the Context of Gene Therapy for Crigler–Najjar Syndrome', *Human Gene Therapy*, 30(10), pp. 1297–1305. Available at: <https://doi.org/10.1089/hum.2019.143>.
- Arya, M. *et al.* (2005) 'Basic principles of real-time quantitative PCR', *Expert Review of Molecular Diagnostics*, 5(2), pp. 209–219. Available at: <https://doi.org/10.1586/14737159.5.2.209>.
- Aurnhammer, C. *et al.* (2012) 'Universal real-time PCR for the detection and quantification of adeno-associated virus serotype 2-derived inverted terminal repeat sequences', *Human Gene Therapy Methods*, 23(1), pp. 18–28. Available at: <https://doi.org/10.1089/hgtb.2011.034>.
- Ayuso, E., Mingozi, F. and Bosch, F. (2010) 'Production, Purification and Characterization of Adeno-Associated Vectors', *Current Gene Therapy*, 10(6), pp. 423–436. Available at: <https://doi.org/10.2174/156652310793797685>.
- Balaji, S. *et al.* (2013) 'Pseudotyped adeno-associated viral vectors for gene transfer in dermal fibroblasts: implications for wound-healing applications', *Journal of Surgical Research*, 184(1), pp. 691–698. Available at: <https://doi.org/10.1016/j.jss.2013.03.051>.
- Barnes, L.F. *et al.* (2021) 'Quantitative analysis of genome packaging in recombinant AAV vectors by charge detection mass spectrometry', *Molecular Therapy. Methods & Clinical Development*, 23, pp. 87–97. Available at: <https://doi.org/10.1016/j.omtm.2021.08.002>.
- Barnes, L.F. *et al.* (2023) 'Analysis of AAV-Extracted DNA by Charge Detection Mass Spectrometry Reveals Genome Truncations', *Analytical Chemistry*, 95(9), pp. 4310–4316. Available at: <https://doi.org/10.1021/acs.analchem.2c04234>.
- Bartlett, J.S., Samulski, R.J. and McCown, T.J. (1998) 'Selective and rapid uptake of adeno-associated virus type 2 in brain', *Human Gene Therapy*, 9(8), pp. 1181–1186. Available at: <https://doi.org/10.1089/hum.1998.9.8-1181>.
- Basiouni, S., Fuhrmann, H. and Schumann, J. (2012) 'High-efficiency transfection of suspension cell lines', *BioTechniques*, 53(2), pp. 1–4. Available at: <https://doi.org/10.2144/000113914>.
- Beastall, G.H. (2018) 'Traceability in Laboratory Medicine: What is it and Why is it Important for Patients?', *EJIFCC*, 29(4), pp. 242–247.

- Becerra, S.P. *et al.* (1988) 'Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript', *Journal of Virology*, 62(8), pp. 2745–2754. Available at: <https://doi.org/10.1128/jvi.62.8.2745-2754.1988>.
- Bee, J.S. *et al.* (2022) 'Impact of Time Out of Intended Storage and Freeze-thaw Rates on the Stability of Adeno-associated Virus 8 and 9', *Journal of Pharmaceutical Sciences*, 111(5), pp. 1346–1353. Available at: <https://doi.org/10.1016/j.xphs.2022.01.002>.
- Bennett, A. *et al.* (2017) 'Thermal Stability as a Determinant of AAV Serotype Identity', *Molecular Therapy. Methods & Clinical Development*, 6, pp. 171–182. Available at: <https://doi.org/10.1016/j.omtm.2017.07.003>.
- Bennett, A., Mietzsch, M. and Agbandje-McKenna, M. (2017) 'Understanding Capsid Assembly and Genome Packaging for Adeno-Associated Viruses', *Future Virology*, 12(6), pp. 283–297. Available at: <https://doi.org/10.2217/fvl-2017-0011>.
- Besnier, C., Takeuchi, Y. and Towers, G. (2002) 'Restriction of lentivirus in monkeys', *Proceedings of the National Academy of Sciences*, 99(18), pp. 11920–11925. Available at: <https://doi.org/10.1073/pnas.172384599>.
- Bish, L.T. *et al.* (2008) 'Adeno-associated virus (AAV) serotype 9 provides global cardiac gene transfer superior to AAV1, AAV6, AAV7, and AAV8 in the mouse and rat', *Human Gene Therapy*, 19(12), pp. 1359–1368. Available at: <https://doi.org/10.1089/hum.2008.123>.
- Blaese, R.M. *et al.* (1995) 'T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years', *Science (New York, N.Y.)*, 270(5235), pp. 475–480. Available at: <https://doi.org/10.1126/science.270.5235.475>.
- Blankinship, M.J. *et al.* (2004) 'Efficient transduction of skeletal muscle using vectors based on adeno-associated virus serotype 6', *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 10(4), pp. 671–678. Available at: <https://doi.org/10.1016/j.ymthe.2004.07.016>.
- Bleker, S., Sonntag, F. and Kleinschmidt, J.A. (2005) 'Mutational Analysis of Narrow Pores at the Fivefold Symmetry Axes of Adeno-Associated Virus Type 2 Capsids Reveals a Dual Role in Genome Packaging and Activation of Phospholipase A2 Activity', *Journal of Virology*, 79(4), pp. 2528–2540. Available at: <https://doi.org/10.1128/JVI.79.4.2528-2540.2005>.
- Borsotti, C., Borroni, E. and Follenzi, A. (2016) 'Lentiviral vector interactions with the host cell', *Current Opinion in Virology*, 21, pp. 102–108. Available at: <https://doi.org/10.1016/j.coviro.2016.08.016>.
- Bosma, B. *et al.* (2018) 'Optimization of viral protein ratios for production of rAAV serotype 5 in the baculovirus system', *Gene Therapy*, 25(6), pp. 415–424. Available at: <https://doi.org/10.1038/s41434-018-0034-7>.

- Bouard, D., Alazard-Dany, N. and Cosset, F.-L. (2009) 'Viral vectors: from virology to transgene expression', *British Journal of Pharmacology*, 157(2), pp. 153–165. Available at: <https://doi.org/10.1038/bjp.2008.349>.
- Bracewell, D.G. *et al.* (2021) 'Analytics of host cell proteins (HCPs): lessons from biopharmaceutical mAb analysis for Gene therapy products', *Current Opinion in Biotechnology*, 71, pp. 98–104. Available at: <https://doi.org/10.1016/j.copbio.2021.06.026>.
- Bracewell, D.G., Francis, R. and Smales, C.M. (2015) 'The future of host cell protein (HCP) identification during process development and manufacturing linked to a risk-based management for their control', *Biotechnology and Bioengineering*, 112(9), pp. 1727–1737. Available at: <https://doi.org/10.1002/bit.25628>.
- Brimble, M.A. *et al.* (2022) 'Preventing packaging of translatable P5-associated DNA contaminants in recombinant AAV vector preps', *Molecular Therapy. Methods & Clinical Development*, 24, pp. 280–291. Available at: <https://doi.org/10.1016/j.omtm.2022.01.008>.
- Brun, V. *et al.* (2009) 'Isotope dilution strategies for absolute quantitative proteomics', *Journal of Proteomics*, 72(5), pp. 740–749. Available at: <https://doi.org/10.1016/j.jprot.2009.03.007>.
- Bruner, K.M. *et al.* (2019) 'A quantitative approach for measuring the reservoir of latent HIV-1 proviruses', *Nature*, 566(7742), pp. 120–125. Available at: <https://doi.org/10.1038/s41586-019-0898-8>.
- Brzobohatá, K. *et al.* (2017) 'Comparison of Suitability of the Most Common Ancient DNA Quantification Methods', *Genetic Testing and Molecular Biomarkers*, 21(4), pp. 265–271. Available at: <https://doi.org/10.1089/gtmb.2016.0197>.
- Burnham, B. *et al.* (2015) 'Analytical Ultracentrifugation as an Approach to Characterize Recombinant Adeno-Associated Viral Vectors', *Human Gene Therapy Methods*, 26(6), pp. 228–242. Available at: <https://doi.org/10.1089/hgtb.2015.048>.
- Bustin, S.A. *et al.* (2009) 'The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments', *Clinical Chemistry*, 55(4), pp. 611–622. Available at: <https://doi.org/10.1373/clinchem.2008.112797>.
- Bzymek, M. and Lovett, S.T. (2001) 'Instability of repetitive DNA sequences: The role of replication in multiple mechanisms', *Proceedings of the National Academy of Sciences*, 98(15), pp. 8319–8325. Available at: <https://doi.org/10.1073/pnas.111008398>.
- Cao, M., You, H. and Hermonat, P.L. (2014) 'The X gene of adeno-associated virus 2 (AAV2) is involved in viral DNA replication', *PloS One*, 9(8), p. e104596. Available at: <https://doi.org/10.1371/journal.pone.0104596>.
- Carver, C., Blech-Hermoni, Y. and James, Q. (2021) 'A duplex assay for quantification and qualification of adeno-associated virus (AAV) using the QIAcuity® Digital PCR System'. Available at: <https://www.qiagen.com/us/resources/resourcedetail?id=8d63e103-1a16-45f2-b914-4b5dccc0a99a&lang=en>.

- Cashen, P. and McLaughlin, K. (2023) 'Overview of Current Downstream Processing for Modern Viral Vectors', in S. Gautam, A.I. Chiramel, and R. Pach (eds) *Bioprocess and Analytics Development for Virus-based Advanced Therapeutics and Medicinal Products (ATMPs)*. Cham: Springer International Publishing, pp. 91–123. Available at: https://doi.org/10.1007/978-3-031-28489-2_5.
- Chen, B. *et al.* (2015) 'Targeting transgene to the heart and liver with AAV 9 by different promoters', *Clinical and Experimental Pharmacology and Physiology*, 42(10), pp. 1108–1117. Available at: <https://doi.org/10.1111/1440-1681.12453>.
- Chen, Q. *et al.* (2020) 'SUMOylation Targets Adeno-associated Virus Capsids but Mainly Restricts Transduction by Cellular Mechanisms', *Journal of Virology*, 94(19), pp. e00871–20. Available at: <https://doi.org/10.1128/JVI.00871-20>.
- Chen, Y. *et al.* (2020) 'Single AAV-Mediated CRISPR-SaCas9 Inhibits HSV-1 Replication by Editing ICP4 in Trigeminal Ganglion Neurons', *Molecular Therapy - Methods & Clinical Development*, 18, pp. 33–43. Available at: <https://doi.org/10.1016/j.omtm.2020.05.011>.
- Chen, Y.H., Keiser, M.S. and Davidson, B.L. (2018) 'Viral Vectors for Gene Transfer', *Current Protocols in Mouse Biology*, 8(4), p. e58. Available at: <https://doi.org/10.1002/cpmo.58>.
- Ciccimaro, E. and Blair, I.A. (2010) 'Stable-Isotope Dilution LC–MS for Quantitative Biomarker Analysis', *Bioanalysis*, 2(2), pp. 311–341. Available at: <https://doi.org/10.4155/bio.09.185>.
- Clark, K.R. *et al.* (1999) 'Highly purified recombinant adeno-associated virus vectors are biologically active and free of detectable helper and wild-type viruses', *Human Gene Therapy*, 10(6), pp. 1031–1039. Available at: <https://doi.org/10.1089/10430349950018427>.
- Clarner, P. *et al.* (2021) 'Development of a one-step RT-ddPCR method to determine the expression and potency of AAV vectors', *Molecular Therapy - Methods & Clinical Development*, 23, pp. 68–77. Available at: <https://doi.org/10.1016/j.omtm.2021.05.003>.
- Clément, N. and Grieger, J.C. (2016) 'Manufacturing of recombinant adeno-associated viral vectors for clinical trials', *Molecular Therapy. Methods & Clinical Development*, 3, p. 16002. Available at: <https://doi.org/10.1038/mtm.2016.2>.
- Cock, P.J.A. *et al.* (2010) 'The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants', *Nucleic Acids Research*, 38(6), pp. 1767–1771. Available at: <https://doi.org/10.1093/nar/gkp1137>.
- Cole, L. *et al.* (2021) 'Characterization of Recombinant Adeno-Associated Viruses (rAAVs) for Gene Therapy Using Orthogonal Techniques', *Pharmaceutics*, 13(4), p. 586. Available at: <https://doi.org/10.3390/pharmaceutics13040586>.
- Colomb-Delsuc, M. *et al.* (2022) 'Assessment of the percentage of full recombinant adeno-associated virus particles in a gene therapy drug using CryoTEM', *PLOS ONE*. Edited by C. Ling, 17(6), p. e0269139. Available at: <https://doi.org/10.1371/journal.pone.0269139>.

Crosson, S.M. *et al.* (2018) 'Helper-free Production of Laboratory Grade AAV and Purification by Iodixanol Density Gradient Centrifugation', *Molecular Therapy - Methods & Clinical Development*, 10, pp. 1–7. Available at: <https://doi.org/10.1016/j.omtm.2018.05.001>.

Danecek, P. *et al.* (2021) 'Twelve years of SAMtools and BCFtools', *GigaScience*, 10(2), p. giab008. Available at: <https://doi.org/10.1093/gigascience/giab008>.

Davidson, B.L. *et al.* (2000) 'Recombinant adeno-associated virus type 2, 4, and 5 vectors: Transduction of variant cell types and regions in the mammalian central nervous system', *Proceedings of the National Academy of Sciences of the United States of America*, 97(7), pp. 3428–3432.

Dawson, P.H. *et al.* (1982) 'The use of triple quadrupoles for sequential mass spectrometry: 2—A detailed case study', *Organic Mass Spectrometry*, 17(5), pp. 212–219. Available at: <https://doi.org/10.1002/oms.1210170504>.

Daya, S. and Berns, K.I. (2008) 'Gene Therapy Using Adeno-Associated Virus Vectors', *Clinical Microbiology Reviews*, 21(4), pp. 583–593. Available at: <https://doi.org/10.1128/CMR.00008-08>.

D'Costa, S. *et al.* (2016) 'Practical utilization of recombinant AAV vector reference standards: focus on vector genomes titration by free ITR qPCR', *Molecular Therapy. Methods & Clinical Development*, 5, p. 16019. Available at: <https://doi.org/10.1038/mtm.2016.19>.

De Coster, W. *et al.* (2018) 'NanoPack: visualizing and processing long-read sequencing data', *Bioinformatics*. Edited by B. Berger, 34(15), pp. 2666–2669. Available at: <https://doi.org/10.1093/bioinformatics/bty149>.

De Coster, W. and Rademakers, R. (2023) 'NanoPack2: population-scale evaluation of long-read sequencing data', *Bioinformatics*. Edited by C. Alkan, 39(5), p. btad311. Available at: <https://doi.org/10.1093/bioinformatics/btad311>.

Delenda, C. and Gaillard, C. (2005) 'Real-time quantitative PCR for the design of lentiviral vector analytical assays', *Gene Therapy*, 12(S1), pp. S36–S50. Available at: <https://doi.org/10.1038/sj.gt.3302614>.

Devonshire, A.S. *et al.* (2016) 'An international comparability study on quantification of mRNA gene expression ratios: CCQM-P103.1', *Biomolecular Detection and Quantification*, 8, pp. 15–28. Available at: <https://doi.org/10.1016/j.bdq.2016.05.003>.

Dickerson, R. *et al.* (2021) 'Separating Empty and Full Recombinant Adeno-Associated Virus Particles Using Isocratic Anion Exchange Chromatography', *Biotechnology Journal*, 16(1), p. e2000015. Available at: <https://doi.org/10.1002/biot.202000015>.

DiMattia, M.A. *et al.* (2012) 'Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9', *Journal of Virology*, 86(12), pp. 6947–6958. Available at: <https://doi.org/10.1128/JVI.07232-11>.

- Dingle, T.C. *et al.* (2013) 'Tolerance of Droplet-Digital PCR vs Real-Time Quantitative PCR to Inhibitory Substances', *Clinical Chemistry*, 59(11), pp. 1670–1672. Available at: <https://doi.org/10.1373/clinchem.2013.211045>.
- Dobnik, D. *et al.* (2019) 'Accurate Quantification and Characterization of Adeno-Associated Viral Vectors', *Frontiers in Microbiology*, 10, p. 1570. Available at: <https://doi.org/10.3389/fmicb.2019.01570>.
- Dommann, J. *et al.* (2024) 'A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities', *mSystems*. Edited by A. Howe, 9(10), pp. e00859-24. Available at: <https://doi.org/10.1128/msystems.00859-24>.
- Dong, B. *et al.* (2014) 'Proteomics Analysis of Co-Purifying Cellular Proteins Associated with rAAV Vectors', *PLOS ONE*, 9(2), p. 7.
- Dong, B., Nakai, H. and Xiao, W. (2010) 'Characterization of genome integrity for oversized recombinant AAV vector', *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 18(1), pp. 87–92. Available at: <https://doi.org/10.1038/mt.2009.258>.
- Dunker-Seidler, F. *et al.* (2025) 'Recombinant AAV batch profiling by nanopore sequencing elucidates product-related DNA impurities and vector genome length distribution', *Molecular Therapy Methods & Clinical Development*, 33(1), p. 101417. Available at: <https://doi.org/10.1016/j.omtm.2025.101417>.
- Duong, T. *et al.* (2023) 'Improvement of Precision in Recombinant Adeno-Associated Virus Infectious Titer Assay with Droplet Digital PCR as an Endpoint Measurement', *Human Gene Therapy*, 34(15–16), pp. 742–757. Available at: <https://doi.org/10.1089/hum.2023.014>.
- Duong, T. *et al.* (2025) 'Three-dimensional linkage analysis with digital PCR for genome integrity and identity of recombinant adeno-associated virus', *Scientific Reports*, 15(1), p. 2154. Available at: <https://doi.org/10.1038/s41598-024-77378-7>.
- Düzgüneş, Z.D. *et al.* (2025) 'Genome-wide SNP Detection in the Horse Mackerel (*Trachurus trachurus*, *T. mediterraneus*, *T. picturatus*) of Turkish Waters Using Illumina and Nanopore Sequencing', *Turkish Journal of Fisheries and Aquatic Sciences*, 25(5). Available at: <https://doi.org/10.4194/TRJFAS27385>.
- Earley, L.F. *et al.* (2015) 'Identification and Characterization of Nuclear and Nucleolar Localization Signals in the Adeno-Associated Virus Serotype 2 Assembly-Activating Protein', *Journal of Virology*. Edited by M.J. Imperiale, 89(6), pp. 3038–3048. Available at: <https://doi.org/10.1128/jvi.03125-14>.
- El Andari, J. and Grimm, D. (2021) 'Production, Processing, and Characterization of Synthetic AAV Gene Therapy Vectors', *Biotechnology Journal*, 16(1), p. 2000025. Available at: <https://doi.org/10.1002/biot.202000025>.

Emmanuel, S.N. *et al.* (2022) 'Structurally Mapping Antigenic Epitopes of Adeno-associated Virus 9: Development of Antibody Escape Variants', *Journal of Virology*. Edited by C.R. Parrish, 96(3), pp. e01251-21. Available at: <https://doi.org/10.1128/JVI.01251-21>.

Emslie, K.R. *et al.* (2019) 'Droplet Volume Variability and Impact on Digital PCR Copy Number Concentration Measurements', *Analytical Chemistry*, 91(6), pp. 4124–4131. Available at: <https://doi.org/10.1021/acs.analchem.8b05828>.

Erickson, S.B. *et al.* (2024) 'Precise Manipulation of the Site and Stoichiometry of Capsid Modification Enables Optimization of Functional Adeno-Associated Virus Conjugates', *Bioconjugate Chemistry*, 35(1), pp. 64–71. Available at: <https://doi.org/10.1021/acs.bioconjchem.3c00411>.

Ermini, L. and Driguez, P. (2024) 'The Application of Long-Read Sequencing to Cancer', *Cancers*, 16(7), p. 1275. Available at: <https://doi.org/10.3390/cancers16071275>.

Espinosa, E. *et al.* (2024) 'Advancements in long-read genome sequencing technologies and algorithms', *Genomics*, 116(3), p. 110842. Available at: <https://doi.org/10.1016/j.ygeno.2024.110842>.

Fekete, S. *et al.* (2023) 'Chromatographic strategies for the analytical characterization of adeno-associated virus vector-based gene therapy products', *TrAC Trends in Analytical Chemistry*, 164, p. 117088. Available at: <https://doi.org/10.1016/j.trac.2023.117088>.

Felgner, P.L. *et al.* (1987) 'Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure.', *Proceedings of the National Academy of Sciences of the United States of America*, 84(21), pp. 7413–7417.

Fernandes, R.P. *et al.* (2022) 'Assessing Multi-Attribute Characterization of Enveloped and Non-Enveloped Viral Particles by Capillary Electrophoresis', *Viruses*, 14(11), p. 2539. Available at: <https://doi.org/10.3390/v14112539>.

Finer, M. and Glorioso, J. (2017) 'A brief account of viral vectors and their promise for gene therapy', *Gene Therapy*, 24(1), pp. 1–2. Available at: <https://doi.org/10.1038/gt.2016.71>.

Florea, M. *et al.* (2023) 'High-efficiency purification of divergent AAV serotypes using AAVX affinity chromatography', *Molecular Therapy - Methods & Clinical Development*, 28, pp. 146–159. Available at: <https://doi.org/10.1016/j.omtm.2022.12.009>.

Flotte, T.R. *et al.* (1993) 'Stable in vivo expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector.', *Proceedings of the National Academy of Sciences of the United States of America*, 90(22), pp. 10613–10617.

Flotte, T.R. *et al.* (1995) 'An improved system for packaging recombinant adeno-associated virus vectors capable of in vivo transduction', *Gene Therapy*, 2(1), pp. 29–37.

François, A. *et al.* (2018) 'Accurate Titration of Infectious AAV Particles Requires Measurement of Biologically Active Vector Genomes and Suitable Controls', *Molecular*

Therapy. Methods & Clinical Development, 10, pp. 223–236. Available at: <https://doi.org/10.1016/j.omtm.2018.07.004>.

Fu, Q. *et al.* (2023) 'Critical challenges and advances in recombinant adeno-associated virus (rAAV) biomanufacturing', *Biotechnology and Bioengineering*, 120(9), pp. 2601–2621. Available at: <https://doi.org/10.1002/bit.28412>.

Furuta-Hanawa, B., Yamaguchi, T. and Uchida, E. (2019) 'Two-Dimensional Droplet Digital PCR as a Tool for Titration and Integrity Evaluation of Recombinant Adeno-Associated Viral Vectors', *Human Gene Therapy Methods*, 30(4), pp. 127–136. Available at: <https://doi.org/10.1089/hgtb.2019.031>.

Galibert, L. *et al.* (2021) 'Functional roles of the membrane-associated AAV protein MAAV', *Scientific Reports*, 11, p. 21698. Available at: <https://doi.org/10.1038/s41598-021-01220-7>.

Galli, A. *et al.* (2021) 'Characterization of Viral Genome Encapsidated in Adeno-associated Recombinant Vectors Produced in Yeast *Saccharomyces cerevisiae*', *Molecular Biotechnology*, 63(2), pp. 156–165. Available at: <https://doi.org/10.1007/s12033-020-00294-4>.

Gao, G.-P. *et al.* (2002) 'Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy', *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), pp. 11854–11859. Available at: <https://doi.org/10.1073/pnas.182412299>.

Gao, K. *et al.* (2014) 'Empty virions in AAV8 vector preparations reduce transduction efficiency and may cause total viral particle dose-limiting side effects', *Molecular Therapy - Methods & Clinical Development*, 1, p. 9. Available at: <https://doi.org/10.1038/mtm.2013.9>.

Gardner, M.R. *et al.* (2022) 'High concordance of ELISA and neutralization assays allows for the detection of antibodies to individual AAV serotypes', *Molecular Therapy - Methods & Clinical Development*, 24, pp. 199–206. Available at: <https://doi.org/10.1016/j.omtm.2022.01.003>.

Gauci, V.J., Wright, E.P. and Coorssen, J.R. (2011) 'Quantitative proteomics: assessing the spectrum of in-gel protein detection methods', *Journal of Chemical Biology*, 4(1), pp. 3–29. Available at: <https://doi.org/10.1007/s12154-010-0043-5>.

Giles, A.R. *et al.* (2018) 'Deamidation of Amino Acids on the Surface of Adeno-Associated Virus Capsids Leads to Charge Heterogeneity and Altered Vector Function', *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 26(12), pp. 2848–2862. Available at: <https://doi.org/10.1016/j.ymthe.2018.09.013>.

Gimpel, A.L. *et al.* (2021) 'Analytical methods for process and product characterization of recombinant adeno-associated virus-based gene therapies', *Molecular Therapy. Methods & Clinical Development*, 20, pp. 740–754. Available at: <https://doi.org/10.1016/j.omtm.2021.02.010>.

Girod, A. *et al.* (2002) 'The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity', *Journal of General Virology*, 83(5), pp. 973–978. Available at: <https://doi.org/10.1099/0022-1317-83-5-973>.

Gleerup, D. *et al.* (2023) 'Measuring DNA quality by digital PCR using probability calculations', *Analytica Chimica Acta*, 1279, p. 341822. Available at: <https://doi.org/10.1016/j.aca.2023.341822>.

Gonçalves, M.A. (2005) 'Adeno-associated virus: from defective virus to effective vector', *Virology Journal*, 2, p. 43. Available at: <https://doi.org/10.1186/1743-422X-2-43>.

Graf, T. *et al.* (2021) 'Best practices on critical reagent characterization, qualification, and life cycle management for HCP immunoassays', *Biotechnology and Bioengineering*, 118(10), pp. 3633–3639. Available at: <https://doi.org/10.1002/bit.27881>.

Graham, F.L. and van der Eb, A.J. (1973) 'A new technique for the assay of infectivity of human adenovirus 5 DNA', *Virology*, 52(2), pp. 456–467. Available at: [https://doi.org/10.1016/0042-6822\(73\)90341-3](https://doi.org/10.1016/0042-6822(73)90341-3).

Gray, J.T. and Zolotukhin, S. (2011) 'Design and construction of functional AAV vectors', *Methods in Molecular Biology (Clifton, N.J.)*, 807, pp. 25–46. Available at: https://doi.org/10.1007/978-1-61779-370-7_2.

Green, E.A. and Lee, K.H. (2021) 'Analytical methods to characterize recombinant adeno-associated virus vectors and the benefit of standardization and reference materials', *Current Opinion in Biotechnology*, 71, pp. 65–76. Available at: <https://doi.org/10.1016/j.copbio.2021.06.025>.

Grimm, D. *et al.* (1999) 'Titration of AAV-2 particles via a novel capsid ELISA: packaging of genomes can limit production of recombinant AAV-2', *Gene Therapy*, 6(7), pp. 1322–1330. Available at: <https://doi.org/10.1038/sj.gt.3300946>.

Grosse, S. *et al.* (2017) 'Relevance of Assembly-Activating Protein for Adeno-associated Virus Vector Production and Capsid Protein Stability in Mammalian and Insect Cells', *Journal of Virology*, 91(20), pp. e01198-17. Available at: <https://doi.org/10.1128/JVI.01198-17>.

Gruntman, A.M. *et al.* (2015) 'Stability and Compatibility of Recombinant Adeno-Associated Virus Under Conditions Commonly Encountered in Human Gene Therapy Trials', *Human Gene Therapy Methods*, 26(2), pp. 71–76. Available at: <https://doi.org/10.1089/hgtb.2015.040>.

Guapo, F. *et al.* (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)', *Journal of Pharmaceutical and Biomedical Analysis*, 207, p. 114427. Available at: <https://doi.org/10.1016/j.jpba.2021.114427>.

Guerin, K. *et al.* (2020) 'A Novel Next-Generation Sequencing and Analysis Platform to Assess the Identity of Recombinant Adeno-Associated Viral Preparations from Viral DNA Extracts',

Human Gene Therapy, 31(11–12), pp. 664–678. Available at: <https://doi.org/10.1089/hum.2019.277>.

Hacein-Bey-Abina, S. *et al.* (2003) 'LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1', *Science (New York, N.Y.)*, 302(5644), pp. 415–419. Available at: <https://doi.org/10.1126/science.1088547>.

Hao, Z. *et al.* (2015) 'Intact Antibody Characterization Using Orbitrap Mass Spectrometry', in J.E. Schiel, D.L. Davis, and O.V. Borisov (eds) *ACS Symposium Series*. Washington, DC: American Chemical Society, pp. 289–315. Available at: <https://doi.org/10.1021/bk-2015-1202.ch010>.

Hashiba, N. *et al.* (2024) 'Tracking particle-encapsulated DNA across the anion-exchange chromatography fractions of recombinant adeno-associated virus using droplet digital PCR and high-throughput sequencing'. Available at: <https://doi.org/10.22541/au.172951796.62373403/v1>.

Hayes, D.B. and Dobnik, D. (2022) 'Commentary: Multiplex dPCR and SV-AUC are Promising Assays to Robustly Monitor the Critical Quality Attribute of AAV Drug Product Integrity', *Journal of Pharmaceutical Sciences*, 111(8), pp. 2143–2148. Available at: <https://doi.org/10.1016/j.xphs.2022.04.010>.

Hays, A. *et al.* (2022) 'Best Practices in qPCR and dPCR Validation in Regulated Bioanalytical Laboratories', *The AAPS Journal*, 24(2), p. 36. Available at: <https://doi.org/10.1208/s12248-022-00686-1>.

Heid, C.A. *et al.* (1996) 'Real time quantitative PCR.', *Genome Research*, 6(10), pp. 986–994. Available at: <https://doi.org/10.1101/gr.6.10.986>.

Hejmowski, A.L. *et al.* (2022) 'Novel anion exchange membrane chromatography method for the separation of empty and full adeno-associated virus', *Biotechnology Journal*, 17(2). Available at: <https://doi.org/10.1002/biot.202100219>.

Higashiyama, K. *et al.* (2023) 'Quantitation of Residual Host Cell DNA in Recombinant Adeno-Associated Virus Using Droplet Digital Polymerase Chain Reaction', *Human Gene Therapy*, 34(11–12), pp. 578–585. Available at: <https://doi.org/10.1089/hum.2023.006>.

Higgins, K.M. *et al.* (1998) 'The Effect of Serial Dilution Error on Calibration Inference in Immunoassay', *Biometrics*, 54(1), p. 19. Available at: <https://doi.org/10.2307/2533992>.

Hinderer, C. *et al.* (2018) 'Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN', *Human Gene Therapy*, 29(3), pp. 285–298. Available at: <https://doi.org/10.1089/hum.2018.015>.

Hirsch, M.L., Wolf, S.J. and Samulski, R.J. (2016) 'Delivering Transgenic DNA Exceeding the Carrying Capacity of AAV Vectors', *Methods in Molecular Biology (Clifton, N.J.)*, 1382, pp. 21–39. Available at: https://doi.org/10.1007/978-1-4939-3271-9_2.

Holstein (Sherwood), C.A., Gafken, P.R. and Martin, D.B. (2011) 'Collision Energy Optimization of b- and y-Ions for Multiple Reaction Monitoring Mass Spectrometry', *Journal of Proteome Research*, 10(1), pp. 231–240. Available at: <https://doi.org/10.1021/pr1004289>.

Hörer, M. *et al.* (1995) 'Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters.', *Journal of Virology*, 69(9), pp. 5485–5496.

Hou, Y. *et al.* (2023) 'Droplet-based digital PCR (ddPCR) and its applications', *TrAC Trends in Analytical Chemistry*, 158, p. 116897. Available at: <https://doi.org/10.1016/j.trac.2022.116897>.

Hu, T. *et al.* (2021) 'Next-generation sequencing technologies: An overview', *Human Immunology*, 82(11), pp. 801–811. Available at: <https://doi.org/10.1016/j.humimm.2021.02.012>.

International Organisation for Standardisation (ISO) (2019) 'Biotechnology — Requirements for evaluating the performance of quantification methods for nucleic acid target sequences — qPCR and dPCR (ISO 20395:2019)'. Available at: <https://www.iso.org/standard/67893.html>.

International Organisation for Standardisation (ISO) (2022) 'Biotechnology — Ancillary materials present during the production of cellular therapeutic products and gene therapy products (ISO 20399:2022)'. Available at: <https://www.iso.org/standard/79399.html>.

International Organisation for Standardisation (ISO) (2023) 'Biotechnology - Gene delivery systems - Part 2: Guide for methods for the qualification of viral vectors (ISO/PWI 16921-2)'. Available at: <https://standardsdevelopment.bsigroup.com/projects/9022-08016#/section>.

Issa, S.S. *et al.* (2023) 'Various AAV Serotypes and Their Applications in Gene Therapy: An Overview', *Cells*, 12(5), p. 785. Available at: <https://doi.org/10.3390/cells12050785>.

Jalšić, L. *et al.* (2023) 'Inducible HEK293 AAV packaging cell lines expressing Rep proteins', *Molecular Therapy - Methods & Clinical Development*, 30, pp. 259–275. Available at: <https://doi.org/10.1016/j.omtm.2023.07.002>.

Jawa, V. *et al.* (2016) 'Evaluating Immunogenicity Risk Due to Host Cell Protein Impurities in Antibody-Based Biotherapeutics', *The AAPS Journal*, 18(6), pp. 1439–1452. Available at: <https://doi.org/10.1208/s12248-016-9948-4>.

Jiang, Z. and Dalby, P.A. (2023) 'Challenges in scaling up AAV-based gene therapy manufacturing', *Trends in Biotechnology*, 41(10), pp. 1268–1281. Available at: <https://doi.org/10.1016/j.tibtech.2023.04.002>.

Jin, X. *et al.* (2017) 'Direct Liquid Chromatography/Mass Spectrometry Analysis for Complete Characterization of Recombinant Adeno-Associated Virus Capsid Proteins', *Human Gene Therapy Methods*, 28(5), pp. 255–267. Available at: <https://doi.org/10.1089/hgtb.2016.178>.

Jones, G.M. *et al.* (2016) 'Digital PCR dynamic range is approaching that of real-time quantitative PCR', *Biomolecular Detection and Quantification*, 10, pp. 31–33. Available at: <https://doi.org/10.1016/j.bdq.2016.10.001>.

Jones, M. *et al.* (2021) "'High-risk" host cell proteins (HCPs): A multi-company collaborative view', *Biotechnology and Bioengineering*, 118(8), pp. 2870–2885. Available at: <https://doi.org/10.1002/bit.27808>.

Khatwani, S.L., Pavlova, A. and Pirot, Z. (2021) 'Anion-exchange HPLC assay for separation and quantification of empty and full capsids in multiple adeno-associated virus serotypes', *Molecular Therapy. Methods & Clinical Development*, 21, pp. 548–558. Available at: <https://doi.org/10.1016/j.omtm.2021.04.003>.

King, J.A. *et al.* (2001) 'DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids', *The EMBO Journal*, 20(12), pp. 3282–3291. Available at: <https://doi.org/10.1093/emboj/20.12.3282>.

Kiyonami, R. and Domon, B. (2010) 'Selected Reaction Monitoring Applied to Quantitative Proteomics', in P.R. Cutillas and J.F. Timms (eds) *LC-MS/MS in Proteomics*. Totowa, NJ: Humana Press (Methods in Molecular Biology), pp. 155–166. Available at: https://doi.org/10.1007/978-1-60761-780-8_9.

Koczot, F.J. *et al.* (1973) 'Self-complementarity of terminal sequences within plus or minus strands of adenovirus-associated virus DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 70(1), pp. 215–219. Available at: <https://doi.org/10.1073/pnas.70.1.215>.

Kohn, D.B. *et al.* (2021) 'Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency', *The New England journal of medicine*, 384(21), pp. 2002–2013. Available at: <https://doi.org/10.1056/NEJMoa2027675>.

Kontogiannis, T. *et al.* (2024) 'Characterization of AAV vectors: A review of analytical techniques and critical quality attributes', *Molecular Therapy - Methods & Clinical Development*, 32(3), p. 101309. Available at: <https://doi.org/10.1016/j.omtm.2024.101309>.

Korlach, J. (2013) 'Understanding Accuracy in SMRT® Sequencing'. Available at: https://www.pacb.com/wp-content/uploads/2015/09/Perspective_UnderstandingAccuracySMRTSequencing1.pdf.

Košir, A.B. *et al.* (2017) 'Droplet volume variability as a critical factor for accuracy of absolute quantification using droplet digital PCR', *Analytical and Bioanalytical Chemistry*, 409(28), pp. 6689–6697. Available at: <https://doi.org/10.1007/s00216-017-0625-y>.

Kotin, R.M. *et al.* (1990) 'Site-specific integration by adeno-associated virus', *Proceedings of the National Academy of Sciences of the United States of America*, 87(6), pp. 2211–2215. Available at: <https://doi.org/10.1073/pnas.87.6.2211>.

- Kouzarides, T. (2000) 'Acetylation: a regulatory modification to rival phosphorylation?', *The EMBO Journal*, 19(6), pp. 1176–1179. Available at: <https://doi.org/10.1093/emboj/19.6.1176>.
- Kumar, P., Nagarajan, A. and Uchil, P.D. (2019) 'Lipofection', *Cold Spring Harbor Protocols*, 2019(3). Available at: <https://doi.org/10.1101/pdb.top096248>.
- Kuypers, J. and Jerome, K.R. (2017) 'Applications of Digital PCR for Clinical Microbiology', *Journal of Clinical Microbiology*. Edited by C.S. Kraft, 55(6), pp. 1621–1628. Available at: <https://doi.org/10.1128/JCM.00211-17>.
- Lam, A.K. *et al.* (2022) 'Fast and high-throughput LC-MS characterization, and peptide mapping of engineered AAV capsids using LC-MS/MS', *Molecular Therapy - Methods & Clinical Development*, 27, pp. 185–194. Available at: <https://doi.org/10.1016/j.omtm.2022.09.008>.
- Lange, V. *et al.* (2008) 'Selected reaction monitoring for quantitative proteomics: a tutorial', *Molecular Systems Biology*, 4(1), p. 222. Available at: <https://doi.org/10.1038/msb.2008.61>.
- Lebherz, C. *et al.* (2008) 'Novel AAV serotypes for improved ocular gene transfer', *The journal of gene medicine*, 10(4), pp. 375–382. Available at: <https://doi.org/10.1002/jgm.1126>.
- Lecomte, E. *et al.* (2015) 'Advanced Characterization of DNA Molecules in rAAV Vector Preparations by Single-stranded Virus Next-generation Sequencing', *Molecular Therapy - Nucleic Acids*, 4. Available at: <https://doi.org/10.1038/mtna.2015.32>.
- Lee, J.S. *et al.* (1995) 'Functional interactions between YY1 and adenovirus E1A.', *Nucleic Acids Research*, 23(6), pp. 925–931.
- Lee, Z. *et al.* (2022) 'Construction of an rAAV Producer Cell Line through Synthetic Biology', *ACS Synthetic Biology*, 11(10), pp. 3285–3295. Available at: <https://doi.org/10.1021/acssynbio.2c00207>.
- Lehman, S.E. *et al.* (2025) 'Interlaboratory Measurement of Adeno-Associated Virus: Comparative Quantification of Full and Empty Capsids', *Human Gene Therapy*, 36(1–2), pp. 36–44. Available at: <https://doi.org/10.1089/hum.2024.124>.
- 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', *Molecular Therapy - Methods & Clinical Development*, 32(4), p. 101383. Available at: <https://doi.org/10.1016/j.omtm.2024.101383>.
- Li, G. *et al.* (2025) 'Efficient development of nanobody-based affinity chromatography for AAV8 purification', *Protein Expression and Purification*, 227, p. 106638. Available at: <https://doi.org/10.1016/j.pep.2024.106638>.
- Li, H. (2018) 'Minimap2: pairwise alignment for nucleotide sequences', *Bioinformatics*. Edited by I. Birol, 34(18), pp. 3094–3100. Available at: <https://doi.org/10.1093/bioinformatics/bty191>.

Lin, Y. *et al.* (2025) 'A synthetic platform for developing recombinant adeno-associated virus type 8 producer cell lines', *Biotechnology Progress*, p. e70009. Available at: <https://doi.org/10.1002/btpr.70009>.

Liu, A.P. *et al.* (2020) 'Characterization of Adeno-Associated Virus Capsid Proteins Using Hydrophilic Interaction Chromatography Coupled with Mass Spectrometry', *Journal of Pharmaceutical and Biomedical Analysis*, 189, p. 113481. Available at: <https://doi.org/10.1016/j.jpba.2020.113481>.

Liu, D. *et al.* (2021) 'Crossing the blood-brain barrier with AAV vectors', *Metabolic Brain Disease*, 36(1), pp. 45–52. Available at: <https://doi.org/10.1007/s11011-020-00630-2>.

Liu, Z. *et al.* (2021) 'Disruptors, a new class of oligonucleotide reagents, significantly improved PCR performance on templates containing stable intramolecular secondary structures', *Analytical Biochemistry*, 624, p. 114169. Available at: <https://doi.org/10.1016/j.ab.2021.114169>.

Lock, M. *et al.* (2010) 'Characterization of a Recombinant Adeno-Associated Virus Type 2 Reference Standard Material', *Human Gene Therapy*, 21(10), pp. 1273–1285. Available at: <https://doi.org/10.1089/hum.2009.223>.

Lock, M. *et al.* (2014) 'Absolute determination of single-stranded and self-complementary adeno-associated viral vector genome titers by droplet digital PCR', *Human Gene Therapy Methods*, 25(2), pp. 115–125. Available at: <https://doi.org/10.1089/hgtb.2013.131>.

Lock, M. *et al.* (2020) 'Sensitive Determination of Infectious Titer of Recombinant Adeno-Associated Viruses (rAAVs) Using TCID₅₀ End-Point Dilution and Quantitative Polymerase Chain Reaction (qPCR)', *Cold Spring Harbor Protocols*, 2020(5), p. pdb.prot095653. Available at: <https://doi.org/10.1101/pdb.prot095653>.

Lomash, R.M. *et al.* (2025) 'Adeno-Associated Virus Gene Therapy Development: Early Planning and Regulatory Considerations to Advance the Platform Vector Gene Therapy Program', *Human Gene Therapy*, 36(5–6), pp. 653–662. Available at: <https://doi.org/10.1089/hum.2024.230>.

van der Loo, J.C.M. and Wright, J.F. (2016) 'Progress and challenges in viral vector manufacturing', *Human Molecular Genetics*, 25(R1), pp. R42–52. Available at: <https://doi.org/10.1093/hmg/ddv451>.

Lukashev, A.N. and Zamyatnin, A.A. (2016) 'Viral vectors for gene therapy: Current state and clinical perspectives', *Biochemistry (Moscow)*, 81(7), pp. 700–708. Available at: <https://doi.org/10.1134/S0006297916070063>.

Lundstrom, K. (2018) 'Viral Vectors in Gene Therapy', *Diseases*, 6(2), p. 42. Available at: <https://doi.org/10.3390/diseases6020042>.

MacLean, B. *et al.* (2010) 'Effect of Collision Energy Optimization on the Measurement of Peptides by Selected Reaction Monitoring (SRM) Mass Spectrometry', *Analytical Chemistry*, 82(24), pp. 10116–10124. Available at: <https://doi.org/10.1021/ac102179j>.

- Mallick, P. *et al.* (2007) 'Computational prediction of proteotypic peptides for quantitative proteomics', *Nature Biotechnology*, 25(1), pp. 125–131. Available at: <https://doi.org/10.1038/nbt1275>.
- Malvern Panalytical (2020) 'Measuring the concentration of adeno-associated virus with multi-angle dynamic light scattering (MADLS). Application Note 180608.'
- Markova, N. *et al.* (2021) 'Biophysical Characterization of Viral and Lipid-Based Vectors for Vaccines and Therapeutics with Light Scattering and Calorimetric Techniques', *Vaccines*, 10(1), p. 49. Available at: <https://doi.org/10.3390/vaccines10010049>.
- Martinez-Fernandez de la Camara, C., McClements, M.E. and MacLaren, R.E. (2021) 'Accurate Quantification of AAV Vector Genomes by Quantitative PCR', *Genes*, 12(4), p. 601. Available at: <https://doi.org/10.3390/genes12040601>.
- Mary, B. *et al.* (2019) 'Post-translational modifications in capsid proteins of recombinant adeno-associated virus (AAV) 1-rh10 serotypes', *The FEBS journal*, 286(24), pp. 4964–4981. Available at: <https://doi.org/10.1111/febs.15013>.
- Masago, K. *et al.* (2021) 'Comparison between Fluorimetry (Qubit) and Spectrophotometry (NanoDrop) in the Quantification of DNA and RNA Extracted from Frozen and FFPE Tissues from Lung Cancer Patients: A Real-World Use of Genomic Tests', *Medicina*, 57(12), p. 1375. Available at: <https://doi.org/10.3390/medicina57121375>.
- Mauro, V.P. and Chappell, S.A. (2014) 'A critical analysis of codon optimization in human therapeutics', *Trends in Molecular Medicine*, 20(11), pp. 604–613. Available at: <https://doi.org/10.1016/j.molmed.2014.09.003>.
- McColl-Carboni, A. *et al.* (2024) 'Analytical characterization of full, intermediate, and empty AAV capsids', *Gene Therapy*, 31(5–6), pp. 285–294. Available at: <https://doi.org/10.1038/s41434-024-00444-2>.
- McCormack, M.P. and Rabbitts, T.H. (2004) 'Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency', *The New England Journal of Medicine*, 350(9), pp. 913–922. Available at: <https://doi.org/10.1056/NEJMr032207>.
- McCraw, D.M. *et al.* (2012) 'Structure of adeno-associated virus-2 in complex with neutralizing monoclonal antibody A20', *Virology*, 431(1–2), pp. 40–49. Available at: <https://doi.org/10.1016/j.virol.2012.05.004>.
- McIntosh, N.L. *et al.* (2021) 'Comprehensive characterization and quantification of adeno associated vectors by size exclusion chromatography and multi angle light scattering', *Scientific Reports*, 11(1), p. 3012. Available at: <https://doi.org/10.1038/s41598-021-82599-1>.
- Megger, D.A. *et al.* (2014) 'Comparison of label-free and label-based strategies for proteome analysis of hepatoma cell lines', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1844(5), pp. 967–976. Available at: <https://doi.org/10.1016/j.bbapap.2013.07.017>.

- Merten, O.-W. *et al.* (2011) 'Large-scale manufacture and characterization of a lentiviral vector produced for clinical ex vivo gene therapy application', *Human Gene Therapy*, 22(3), pp. 343–356. Available at: <https://doi.org/10.1089/hum.2010.060>.
- Michaud, S.A. *et al.* (2024) 'Multiple reaction monitoring assays for large-scale quantitation of proteins from 20 mouse organs and tissues', *Communications Biology*, 7(1), p. 6. Available at: <https://doi.org/10.1038/s42003-023-05687-0>.
- Milavec, M. *et al.* (2022) 'The performance of human cytomegalovirus digital PCR reference measurement procedure in seven external quality assessment schemes over four years', *Methods*, 201, pp. 65–73. Available at: <https://doi.org/10.1016/j.ymeth.2021.03.016>.
- Mingozzi, F. *et al.* (2013) 'Overcoming preexisting humoral immunity to AAV using capsid decoys', *Science Translational Medicine*, 5(194), p. 194ra92. Available at: <https://doi.org/10.1126/scitranslmed.3005795>.
- Miyaoka, R. *et al.* (2023) 'Development of a novel purification method for AAV vectors using tangential flow filtration', *Biotechnology and Bioengineering*, 120(11), pp. 3311–3321. Available at: <https://doi.org/10.1002/bit.28524>.
- Monahan, P.E. and Samulski, R.J. (2000) 'Adeno-associated virus vectors for gene therapy: more pros than cons?', *Molecular Medicine Today*, 6(11), pp. 433–440. Available at: [https://doi.org/10.1016/s1357-4310\(00\)01810-4](https://doi.org/10.1016/s1357-4310(00)01810-4).
- Moullier, P. and Snyder, R.O. (2008) 'International Efforts for Recombinant Adeno-associated Viral Vector Reference Standards', *Molecular Therapy*, 16(7), pp. 1185–1188. Available at: <https://doi.org/10.1038/mt.2008.125>.
- Moullier, P. and Snyder, R.O. (2012) 'Recombinant Adeno-Associated Viral Vector Reference Standards', in *Methods in Enzymology*. Elsevier, pp. 297–311. Available at: <https://doi.org/10.1016/B978-0-12-386509-0.00015-6>.
- Murray, S. *et al.* (2006) 'Characterization of the capsid protein glycosylation of adeno-associated virus type 2 by high-resolution mass spectrometry', *Journal of Virology*, 80(12), pp. 6171–6176. Available at: <https://doi.org/10.1128/JVI.02417-05>.
- Nakano, K. *et al.* (2017) 'Advantages of genome sequencing by long-read sequencer using SMRT technology in medical area', *Human Cell*, 30(3), pp. 149–161. Available at: <https://doi.org/10.1007/s13577-017-0168-8>.
- Namkung, S. *et al.* (2022) 'Direct ITR-to-ITR nanopore sequencing of AAV vector genomes', *Human Gene Therapy*, p. hum.2022.143. Available at: <https://doi.org/10.1089/hum.2022.143>.
- Nash, K. *et al.* (2009) 'Identification of Cellular Proteins That Interact with the Adeno-Associated Virus Rep Protein', *Journal of Virology*, 83(1), pp. 454–469. Available at: <https://doi.org/10.1128/JVI.01939-08>.

Naso, M.F. *et al.* (2017) 'Adeno-Associated Virus (AAV) as a Vector for Gene Therapy', *BioDrugs: Clinical Immunotherapeutics, Biopharmaceuticals and Gene Therapy*, 31(4), pp. 317–334. Available at: <https://doi.org/10.1007/s40259-017-0234-5>.

Nathwani, A.C. *et al.* (2006) 'Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver', *Blood*, 107(7), pp. 2653–2661. Available at: <https://doi.org/10.1182/blood-2005-10-4035>.

Neilson, K.A. *et al.* (2011) 'Less label, more free: Approaches in label-free quantitative mass spectrometry', *PROTEOMICS*, 11(4), pp. 535–553. Available at: <https://doi.org/10.1002/pmic.201000553>.

Ni, Y. *et al.* (2023) 'Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing', *Computational and Structural Biotechnology Journal*, 21, pp. 2352–2364. Available at: <https://doi.org/10.1016/j.csbj.2023.03.038>.

Nielsen, S.D.-H. *et al.* (2024) 'Bioactive milk peptides: an updated comprehensive overview and database', *Critical Reviews in Food Science and Nutrition*, 64(31), pp. 11510–11529. Available at: <https://doi.org/10.1080/10408398.2023.2240396>.

Nixon, G. *et al.* (2014) 'Comparative Study of Sensitivity, Linearity, and Resistance to Inhibition of Digital and Nondigital Polymerase Chain Reaction and Loop Mediated Isothermal Amplification Assays for Quantification of Human Cytomegalovirus', *Analytical Chemistry*, 86(9), pp. 4387–4394. Available at: <https://doi.org/10.1021/ac500208w>.

Ode, H. *et al.* (2024) 'Population-based nanopore sequencing of the HIV-1 pangenome to identify drug resistance mutations', *Scientific Reports*, 14(1), p. 12099. Available at: <https://doi.org/10.1038/s41598-024-63054-3>.

Okada, T. *et al.* (2009) 'Scalable purification of adeno-associated virus serotype 1 (AAV1) and AAV8 vectors, using dual ion-exchange adsorptive membranes', *Human Gene Therapy*, 20(9), pp. 1013–1021. Available at: <https://doi.org/10.1089/hum.2009.006>.

Okonechnikov, K. *et al.* (2012) 'Unipro UGENE: a unified bioinformatics toolkit', *Bioinformatics*, 28(8), pp. 1166–1167. Available at: <https://doi.org/10.1093/bioinformatics/bts091>.

Oliveira, P.H. *et al.* (2010) 'Analysis of DNA repeats in bacterial plasmids reveals the potential for recurrent instability events', *Applied Microbiology and Biotechnology*, 87(6), pp. 2157–2167. Available at: <https://doi.org/10.1007/s00253-010-2671-7>.

Onishi, T. *et al.* (2023) 'Enhancement of recombinant adeno-associated virus activity by improved stoichiometry and homogeneity of capsid protein assembly', *Molecular Therapy - Methods & Clinical Development*, 31, p. 101142. Available at: <https://doi.org/10.1016/j.omtm.2023.101142>.

Oyama, H. *et al.* (2021) 'Characterization of Adeno-Associated Virus Capsid Proteins with Two Types of VP3-Related Components by Capillary Gel Electrophoresis and Mass Spectrometry', *Human Gene Therapy*, 32(21–22), pp. 1403–1416. Available at: <https://doi.org/10.1089/hum.2021.009>.

Pan, X. *et al.* (2022) 'Rational engineering of a functional CpG-free ITR for AAV gene therapy', *Gene Therapy*, 29(6), pp. 333–345. Available at: <https://doi.org/10.1038/s41434-021-00296-0>.

Papić, B. *et al.* (2017) 'New Approaches on Quantification of *Campylobacter jejuni* in Poultry Samples: The Use of Digital PCR and Real-time PCR against the ISO Standard Plate Count Method', *Frontiers in Microbiology*, 8. Available at: <https://doi.org/10.3389/fmicb.2017.00331>.

Pearson, C. *et al.* (2023) 'Quantitation of host cell proteins in biopharmaceuticals from chinese hamster ovarian and vero cell lines using capillary electrophoresis western blots', *Journal of Pharmaceutical and Biomedical Analysis*, 233, p. 115420. Available at: <https://doi.org/10.1016/j.jpba.2023.115420>.

Penaud-Budloo, M. *et al.* (2017) 'Accurate Identification and Quantification of DNA Species by Next-Generation Sequencing in Adeno-Associated Viral Vectors Produced in Insect Cells', *Human Gene Therapy Methods*, 28(3), pp. 148–162. Available at: <https://doi.org/10.1089/hgtb.2016.185>.

Perach, M. *et al.* (2021) 'Identification of conserved slow codons that are important for protein expression and function', *RNA Biology*, 18(12), pp. 2296–2307. Available at: <https://doi.org/10.1080/15476286.2021.1901185>.

Pereira, D.J., McCarty, D.M. and Muzyczka, N. (1997) 'The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection', *Journal of Virology*, 71(2), pp. 1079–1088. Available at: <https://doi.org/10.1128/JVI.71.2.1079-1088.1997>.

Perry, C. and Rayat, A.C.M.E. (2021) 'Lentiviral Vector Bioprocessing', *Viruses*, 13(2), p. 268. Available at: <https://doi.org/10.3390/v13020268>.

Pierson, E.E. *et al.* (2016) 'Resolving Adeno-Associated Viral Particle Diversity With Charge Detection Mass Spectrometry', *Analytical Chemistry*, 88(13), pp. 6718–6725. Available at: <https://doi.org/10.1021/acs.analchem.6b00883>.

Podsakoff, G., Wong, K.K. and Chatterjee, S. (1994) 'Efficient gene transfer into nondividing cells by adeno-associated virus-based vectors.', *Journal of Virology*, 68(9), pp. 5656–5666.

Prakash, A. *et al.* (2009) 'Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development', *Journal of Proteome Research*, 8(6), pp. 2733–2739. Available at: <https://doi.org/10.1021/pr801028b>.

Qu, G. *et al.* (2007) 'Separation of adeno-associated virus type 2 empty particles from genome containing vectors by anion-exchange column chromatography', *Journal of*

Virological Methods, 140(1–2), pp. 183–192. Available at:
<https://doi.org/10.1016/j.jviromet.2006.11.019>.

Quirin, K.A. *et al.* (2018) 'Safety and Efficacy of AAV Retrograde Pancreatic Ductal Gene Delivery in Normal and Pancreatic Cancer Mice', *Molecular Therapy. Methods & Clinical Development*, 8, pp. 8–20. Available at: <https://doi.org/10.1016/j.omtm.2017.09.006>.

Radukic, M.T. *et al.* (2020) 'Nanopore sequencing of native adeno-associated virus (AAV) single-stranded DNA using a transposase-based rapid protocol', *NAR Genomics and Bioinformatics*, 2(4), p. lqaa074. Available at: <https://doi.org/10.1093/nargab/lqaa074>.

Raper, S.E. *et al.* (2003) 'Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer', *Molecular Genetics and Metabolism*, 80(1–2), pp. 148–158. Available at:
<https://doi.org/10.1016/j.ymgme.2003.08.016>.

Rathore, D. *et al.* (2018) 'The role of mass spectrometry in the characterization of biologic protein products', *Expert Review of Proteomics*, 15(5), pp. 431–449. Available at:
<https://doi.org/10.1080/14789450.2018.1469982>.

Rauh, M. (2012) 'LC–MS/MS for protein and peptide quantification in clinical chemistry', *Journal of Chromatography B*, 883–884, pp. 59–67. Available at:
<https://doi.org/10.1016/j.jchromb.2011.09.030>.

Reed, S.E. *et al.* (2006) 'Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors', *Journal of Virological Methods*, 138(1), pp. 85–98. Available at:
<https://doi.org/10.1016/j.jviromet.2006.07.024>.

Regan, J.F. *et al.* (2015) 'A Rapid Molecular Approach for Chromosomal Phasing', *PLOS ONE*. Edited by D.E. Arking, 10(3), p. e0118270. Available at:
<https://doi.org/10.1371/journal.pone.0118270>.

Ribeiro, S.C. *et al.* (2008) 'High Frequency Plasmid Recombination Mediated by 28 bp Direct Repeats', *Molecular Biotechnology*, 40(3), p. 252. Available at:
<https://doi.org/10.1007/s12033-008-9082-3>.

Ricchi, M. *et al.* (2017) 'Comparison among the Quantification of Bacterial Pathogens by qPCR, dPCR, and Cultural Methods', *Frontiers in Microbiology*, 8, p. 1174. Available at:
<https://doi.org/10.3389/fmicb.2017.01174>.

Rieser, R. *et al.* (2020) 'Intrinsic Differential Scanning Fluorimetry for Fast and Easy Identification of Adeno-Associated Virus Serotypes', *Journal of Pharmaceutical Sciences*, 109(1), pp. 854–862. Available at: <https://doi.org/10.1016/j.xphs.2019.10.031>.

Robert, M.-A. *et al.* (2017) 'Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms', *Biotechnology Journal*, 12(3). Available at:
<https://doi.org/10.1002/biot.201600193>.

- Roberts, R.J., Carneiro, M.O. and Schatz, M.C. (2013) 'The advantages of SMRT sequencing', *Genome Biology*, 14(6), p. 405. Available at: <https://doi.org/10.1186/gb-2013-14-6-405>.
- Robledo, D. *et al.* (2014) 'Analysis of qPCR reference gene stability determination methods and a practical approach for efficiency calculation on a turbot (*Scophthalmus maximus*) gonad dataset', *BMC Genomics*, 15(1), p. 648. Available at: <https://doi.org/10.1186/1471-2164-15-648>.
- Rose, J.A. *et al.* (1971) 'Structural Proteins of Adenovirus-Associated Viruses', *Journal of Virology*, 8(5), pp. 766–770.
- Rossmannith, P. *et al.* (2011) 'Mechanisms of degradation of DNA standards for calibration function during storage', *Applied Microbiology and Biotechnology*, 89(2), pp. 407–417. Available at: <https://doi.org/10.1007/s00253-010-2943-2>.
- Ruiz-Villalba, A., Ruijter, J.M. and Van Den Hoff, M.J.B. (2021) 'Use and Misuse of Cq in qPCR Data Analysis and Reporting', *Life*, 11(6), p. 496. Available at: <https://doi.org/10.3390/life11060496>.
- Rumachik, N.G. *et al.* (2020) 'Methods Matter: Standard Production Platforms for Recombinant AAV Produce Chemically and Functionally Distinct Vectors', *Molecular Therapy. Methods & Clinical Development*, 18, pp. 98–118. Available at: <https://doi.org/10.1016/j.omtm.2020.05.018>.
- Russell, S. *et al.* (2017) 'Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial', *Lancet (London, England)*, 390(10097), pp. 849–860. Available at: [https://doi.org/10.1016/S0140-6736\(17\)31868-8](https://doi.org/10.1016/S0140-6736(17)31868-8).
- Sabin, J. *et al.* (2022) 'New insights on the mechanism of polyethylenimine transfection and their implications on gene therapy and DNA vaccines', *Colloids and Surfaces B: Biointerfaces*, 210, p. 112219. Available at: <https://doi.org/10.1016/j.colsurfb.2021.112219>.
- Samulski, R.J. *et al.* (1991) 'Targeted integration of adeno-associated virus (AAV) into human chromosome 19', *The EMBO journal*, 10(12), pp. 3941–3950. Available at: <https://doi.org/10.1002/j.1460-2075.1991.tb04964.x>.
- Satkunanathan, S. *et al.* (2014) 'Establishment of a Novel Cell Line for the Enhanced Production of Recombinant Adeno-Associated Virus Vectors for Gene Therapy', *Human Gene Therapy*, 25(11), pp. 929–941. Available at: <https://doi.org/10.1089/hum.2014.041>.
- Sauna, Z.E. *et al.* (2023) 'Understanding preclinical and clinical immunogenicity risks in novel biotherapeutics development', *Frontiers in Immunology*, 14, p. 1151888. Available at: <https://doi.org/10.3389/fimmu.2023.1151888>.
- Sayers, E.W. *et al.* (2022) 'Database resources of the national center for biotechnology information', *Nucleic Acids Research*, 50(D1), pp. D20–D26. Available at: <https://doi.org/10.1093/nar/gkab1112>.

Schiel, J.E. and Turner, A. (2018) 'The NISTmAb Reference Material 8671 lifecycle management and quality plan', *Analytical and Bioanalytical Chemistry*, 410(8), pp. 2067–2078. Available at: <https://doi.org/10.1007/s00216-017-0844-2>.

Schnödt, M. *et al.* (2016) 'DNA Minicircle Technology Improves Purity of Adeno-associated Viral Vector Preparations', *Molecular Therapy. Nucleic Acids*, 5(8), p. e355. Available at: <https://doi.org/10.1038/mtna.2016.60>.

Schwarz, R.F. *et al.* (2010) 'Evolutionary Distances in the Twilight Zone—A Rational Kernel Approach', *PLoS ONE*. Edited by W. Delport, 5(12), p. e15788. Available at: <https://doi.org/10.1371/journal.pone.0015788>.

Sedlackova, T. *et al.* (2013) 'Fragmentation of DNA affects the accuracy of the DNA quantitation by the commonly used methods', *Biological Procedures Online*, 15(1), p. 5. Available at: <https://doi.org/10.1186/1480-9222-15-5>.

Sedlak, R.H., Kuypers, J. and Jerome, K.R. (2014) 'A multiplexed droplet digital PCR assay performs better than qPCR on inhibition prone samples', *Diagnostic Microbiology and Infectious Disease*, 80(4), pp. 285–286. Available at: <https://doi.org/10.1016/j.diagmicrobio.2014.09.004>.

Sedlazeck, F.J. *et al.* (2018) 'Accurate detection of complex structural variations using single-molecule sequencing', *Nature Methods*, 15(6), pp. 461–468. Available at: <https://doi.org/10.1038/s41592-018-0001-7>.

Segura, M.M. *et al.* (2013) 'New developments in lentiviral vector design, production and purification', *Expert Opinion on Biological Therapy*, 13(7), pp. 987–1011. Available at: <https://doi.org/10.1517/14712598.2013.779249>.

Segura, M.M., Kamen, A.A. and Garnier, A. (2011) 'Overview of Current Scalable Methods for Purification of Viral Vectors', in O.-W. Merten and M. Al-Rubeai (eds) *Viral Vectors for Gene Therapy: Methods and Protocols*. Totowa, NJ: Humana Press (Methods in Molecular Biology), pp. 89–116. Available at: https://doi.org/10.1007/978-1-61779-095-9_4.

Serrano, M.A.C. *et al.* (2023) 'Mass spectrometry in gene therapy: Challenges and opportunities for AAV analysis', *Drug Discovery Today*, 28(1), p. 103442. Available at: <https://doi.org/10.1016/j.drudis.2022.103442>.

Sidstedt, M., Rådström, P. and Hedman, J. (2020) 'PCR inhibition in qPCR, dPCR and MPS—mechanisms and solutions', *Analytical and Bioanalytical Chemistry*, 412(9), pp. 2009–2023. Available at: <https://doi.org/10.1007/s00216-020-02490-2>.

Smolka, M. *et al.* (2024) 'Detection of mosaic and population-level structural variants with Sniffles2', *Nature Biotechnology*, 42(10), pp. 1571–1580. Available at: <https://doi.org/10.1038/s41587-023-02024-y>.

Snijder, J. *et al.* (2014) 'Defining the Stoichiometry and Cargo Load of Viral and Bacterial Nanoparticles by Orbitrap Mass Spectrometry', *Journal of the American Chemical Society*, 136(20), pp. 7295–7299. Available at: <https://doi.org/10.1021/ja502616y>.

- Some, D. *et al.* (2019) 'Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS)', *Journal of Visualized Experiments: JoVE* [Preprint], (148). Available at: <https://doi.org/10.3791/59615>.
- Sommer, J. ürg M. *et al.* (2003) 'Quantification of adeno-associated virus particles and empty capsids by optical density measurement', *Molecular Therapy*, 7(1), pp. 122–128. Available at: [https://doi.org/10.1016/S1525-0016\(02\)00019-9](https://doi.org/10.1016/S1525-0016(02)00019-9).
- Son, A. *et al.* (2024) 'Mass Spectrometry Advancements and Applications for Biomarker Discovery, Diagnostic Innovations, and Personalized Medicine', *International Journal of Molecular Sciences*, 25(18), p. 9880. Available at: <https://doi.org/10.3390/ijms25189880>.
- Sonntag, F. *et al.* (2011) 'The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes', *Journal of Virology*, 85(23), pp. 12686–12697. Available at: <https://doi.org/10.1128/JVI.05359-11>.
- Srivastava, A., Lusby, E.W. and Berns, K.I. (1983) 'Nucleotide sequence and organization of the adeno-associated virus 2 genome.', *Journal of Virology*, 45(2), pp. 555–564.
- Strobel, B. *et al.* (2015) 'Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications', *Human Gene Therapy Methods*, 26(4), pp. 147–157. Available at: <https://doi.org/10.1089/hgtb.2015.051>.
- Su, Q., Sena-Esteves, M. and Gao, G. (2020) 'Analysis of Recombinant Adeno-Associated Virus (rAAV) Purity Using Silver-Stained SDS-PAGE', *Cold Spring Harbor Protocols*, 2020(8), p. 095679. Available at: <https://doi.org/10.1101/pdb.prot095679>.
- Suoranta, T., Laham-Karam, N. and Ylä-Herttua, S. (2021) 'Optimized Protocol for Accurate Titration of Adeno-Associated Virus Vectors', *Human Gene Therapy*, 32(19–20), pp. 1270–1279. Available at: <https://doi.org/10.1089/hum.2020.318>.
- Tai, P.W.L. *et al.* (2018) 'Adeno-associated Virus Genome Population Sequencing Achieves Full Vector Genome Resolution and Reveals Human-Vector Chimeras', *Molecular Therapy. Methods & Clinical Development*, 9, pp. 130–141. Available at: <https://doi.org/10.1016/j.omtm.2018.02.002>.
- Takeda, S. *et al.* (2004) 'Successful Gene Transfer Using Adeno-Associated Virus Vectors into the Kidney: Comparison among Adeno-Associated Virus Serotype 1–5 Vectors in vitro and in vivo', *Nephron Experimental Nephrology*, 96(4), pp. e119–e126. Available at: <https://doi.org/10.1159/000077378>.
- Tam Tran, N. and WI Tai, P. (2024) 'Profiling AAV vector heterogeneity & contaminants using next-generation sequencing methods', *Cell and Gene Therapy Insights*, 09(11), pp. 1565–1583. Available at: <https://doi.org/10.18609/cgti.2023.206>.
- Tan, L.L. *et al.* (2023) 'Current commercial dPCR platforms: technology and market review', *Critical Reviews in Biotechnology*, 43(3), pp. 433–464. Available at: <https://doi.org/10.1080/07388551.2022.2037503>.

Tang, F., Wong, H. and Ng, C.M. (2021) 'Rational Clinical Dose Selection of Adeno-Associated Virus-Mediated Gene Therapy Based on Allometric Principles', *Clinical Pharmacology & Therapeutics*, 110(3), pp. 803–807. Available at: <https://doi.org/10.1002/cpt.2269>.

Tereshko, L. *et al.* (2023) 'A novel method for quantitation of AAV genome integrity using duplex digital PCR', *PLOS ONE*. Edited by S. Agostini, 18(12), p. e0293277. Available at: <https://doi.org/10.1371/journal.pone.0293277>.

The dMIQE Group *et al.* (2020) 'The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR Experiments for 2020', *Clinical Chemistry*, 66(8), pp. 1012–1029. Available at: <https://doi.org/10.1093/clinchem/hvaa125>.

Thomas, C.E., Ehrhardt, A. and Kay, M.A. (2003) 'Progress and problems with the use of viral vectors for gene therapy', *Nature Reviews Genetics*, 4(5), pp. 346–358. Available at: <https://doi.org/10.1038/nrg1066>.

Tiambeng, T.N. *et al.* (2025) 'Characterization of adeno-associated virus capsid proteins using denaturing size-exclusion chromatography coupled with mass spectrometry', *Journal of Pharmaceutical and Biomedical Analysis*, 253, p. 116524. Available at: <https://doi.org/10.1016/j.jpba.2024.116524>.

Tiwari, A. *et al.* (2022) 'Application of digital PCR for public health-related water quality monitoring', *Science of The Total Environment*, 837, p. 155663. Available at: <https://doi.org/10.1016/j.scitotenv.2022.155663>.

Torma, A.F. *et al.* (2017) 'A candidate liquid chromatography mass spectrometry reference method for the quantification of the cardiac marker 1-32 B-type natriuretic peptide', *Clinical Chemistry and Laboratory Medicine (CCLM)*, 55(9). Available at: <https://doi.org/10.1515/cclm-2016-1054>.

Tóth, R. *et al.* (2019) 'Methylation Status of the Adeno-Associated Virus Type 2 (AAV2)', *Viruses*, 11(1), p. E38. Available at: <https://doi.org/10.3390/v11010038>.

Tran, N.T. *et al.* (2022) 'Human and Insect Cell-Produced Recombinant Adeno-Associated Viruses Show Differences in Genome Heterogeneity', *Human Gene Therapy*, 33(7–8), pp. 371–388. Available at: <https://doi.org/10.1089/hum.2022.050>.

Tratschin, J.D., Miller, I.L. and Carter, B.J. (1984) 'Genetic analysis of adeno-associated virus: properties of deletion mutants constructed in vitro and evidence for an adeno-associated virus replication function.', *Journal of Virology*, 51(3), pp. 611–619.

Troxell, B. *et al.* (2023) 'Application of Size Exclusion Chromatography with Multiangle Light Scattering in the Analytical Development of a Preclinical Stage Gene Therapy Program', *Human Gene Therapy*, 34(7–8), pp. 325–338. Available at: <https://doi.org/10.1089/hum.2022.218>.

Tustian, A.D. and Bak, H. (2021) 'Assessment of quality attributes for adeno-associated viral vectors', *Biotechnology and Bioengineering*, 118(11), pp. 4186–4203. Available at: <https://doi.org/10.1002/bit.27905>.

Urabe, M., Ding, C. and Kotin, R.M. (2002) 'Insect cells as a factory to produce adeno-associated virus type 2 vectors', *Human Gene Therapy*, 13(16), pp. 1935–1943. Available at: <https://doi.org/10.1089/10430340260355347>.

U.S. Food and Drug Administration (2020a) 'Chemistry, Manufacturing, and Control (CMC) Information for Human Gene Therapy Investigational New Drug Applications (INDs)'. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/chemistry-manufacturing-and-control-cmc-information-human-gene-therapy-investigational-new-drug>.

U.S. Food and Drug Administration (2020b) 'Guidance for Industry, Long Term Follow-Up After Administration of Human Gene Therapy Products.' Available at: <https://www.fda.gov/media/113768/download>.

U.S. Food and Drug Administration (2022) 'Human gene therapy for neurodegenerative diseases'. Available at: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/human-gene-therapy-neurodegenerative-diseases>.

U.S. Food and Drug Administration (2024) 'Q2(R2) Validation of Analytical Procedures Guidance for Industry'. Available at: <https://www.fda.gov/media/161201/download>.

Van Vliet, K. *et al.* (2009) 'Adeno-associated virus capsid serotype identification: Analytical methods development and application', *Journal of Virological Methods*, 159(2), pp. 167–177. Available at: <https://doi.org/10.1016/j.jviromet.2009.03.020>.

Van Vliet, K.M. *et al.* (2008) 'The role of the adeno-associated virus capsid in gene transfer', *Methods in Molecular Biology (Clifton, N.J.)*, 437, pp. 51–91. Available at: https://doi.org/10.1007/978-1-59745-210-6_2.

Vandenberghe, L.H. *et al.* (2010) 'Efficient Serotype-Dependent Release of Functional Vector into the Culture Medium During Adeno-Associated Virus Manufacturing', *Human Gene Therapy*, 21(10), pp. 1251–1257. Available at: <https://doi.org/10.1089/hum.2010.107>.

Vargas, J.E. *et al.* (2016) 'Retroviral vectors and transposons for stable gene therapy: advances, current challenges and perspectives', *Journal of Translational Medicine*, 14(1), p. 288. Available at: <https://doi.org/10.1186/s12967-016-1047-x>.

Venkatakrishnan, B. *et al.* (2013) 'Structure and Dynamics of Adeno-Associated Virus Serotype 1 VP1-Unique N-Terminal Domain and Its Role in Capsid Trafficking', *Journal of Virology*, 87(9), pp. 4974–4984. Available at: <https://doi.org/10.1128/JVI.02524-12>.

Verch, T., Roselle, C. and Shank-Retzlaff, M. (2016) 'Reduction of Dilution Error in Elisass Using an Internal Standard', *Bioanalysis*, 8(14), pp. 1451–1464. Available at: <https://doi.org/10.4155/bio-2016-0053>.

Wagner, C. *et al.* (2023) 'Quantification of Empty, Partially Filled and Full Adeno-Associated Virus Vectors Using Mass Photometry', *International Journal of Molecular Sciences*, 24(13), p. 11033. Available at: <https://doi.org/10.3390/ijms241311033>.

Wagner, C. *et al.* (2024) 'Automated Mass Photometry of Adeno-Associated Virus Vectors from Crude Cell Extracts', *International Journal of Molecular Sciences*, 25(2), p. 838. Available at: <https://doi.org/10.3390/ijms25020838>.

Wakankar, A.A. and Borchardt, R.T. (2006) 'Formulation considerations for proteins susceptible to asparagine deamidation and aspartate isomerization', *Journal of Pharmaceutical Sciences*, 95(11), pp. 2321–2336. Available at: <https://doi.org/10.1002/jps.20740>.

Wang, A.Y. *et al.* (2004) 'Comparison of adenoviral and adeno-associated viral vectors for pancreatic gene delivery in vivo', *Human Gene Therapy*, 15(4), pp. 405–413. Available at: <https://doi.org/10.1089/104303404322959551>.

Wang, C. *et al.* (2019) 'Developing an Anion Exchange Chromatography Assay for Determining Empty and Full Capsid Contents in AAV6.2', *Molecular Therapy - Methods & Clinical Development*, 15, pp. 257–263. Available at: <https://doi.org/10.1016/j.omtm.2019.09.006>.

Wang, D. *et al.* (2022) 'ddPCR surpasses classical qPCR technology in quantitating bacteria and fungi in the environment', *Molecular Ecology Resources*, 22(7), pp. 2587–2598. Available at: <https://doi.org/10.1111/1755-0998.13644>.

Wang, D., Tai, P.W.L. and Gao, G. (2019) 'Adeno-associated virus vector as a platform for gene therapy delivery', *Nature reviews. Drug discovery*, 18(5), pp. 358–378. Available at: <https://doi.org/10.1038/s41573-019-0012-9>.

Wang, G. *et al.* (2024) 'Assessment of Key Factors Impacting Variability in AAV Vector Genome Titration by Digital PCR', *International Journal of Molecular Sciences*, 25(10), p. 5149. Available at: <https://doi.org/10.3390/ijms25105149>.

Wang, Q. *et al.* (2017) 'Syngeneic AAV Pseudo-particles Potentiate Gene Transduction of AAV Vectors', *Molecular Therapy. Methods & Clinical Development*, 4, pp. 149–158. Available at: <https://doi.org/10.1016/j.omtm.2016.12.004>.

Wang, Y. *et al.* (2019) 'A Digestion-free Method for Quantification of Residual Host Cell DNA in rAAV Gene Therapy Products', *Molecular Therapy. Methods & Clinical Development*, 13, pp. 526–531. Available at: <https://doi.org/10.1016/j.omtm.2019.05.005>.

Wang, Y. *et al.* (2020) 'A qPCR Method for AAV Genome Titer with ddPCR-Level of Accuracy and Precision', *Molecular Therapy - Methods & Clinical Development*, 19, pp. 341–346. Available at: <https://doi.org/10.1016/j.omtm.2020.09.017>.

Wang, Z. *et al.* (2005) 'Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart', *Nature Biotechnology*, 23(3), pp. 321–328. Available at: <https://doi.org/10.1038/nbt1073>.

Ward, P., Elias, P. and Linden, R.M. (2003) 'Rescue of the Adeno-Associated Virus Genome from a Plasmid Vector: Evidence for Rescue by Replication', *Journal of Virology*, 77(21), pp. 11480–11490. Available at: <https://doi.org/10.1128/JVI.77.21.11480-11490.2003>.

Warnock, J.N., Daigre, C. and Al-Rubeai, M. (2011) 'Introduction to viral vectors', *Methods in Molecular Biology (Clifton, N.J.)*, 737, pp. 1–25. Available at: https://doi.org/10.1007/978-1-61779-095-9_1.

Watanabe, N. *et al.* (2015) 'Re-examination of regulatory opinions in Europe: possible contribution for the approval of the first gene therapy product Glybera', *Molecular Therapy. Methods & Clinical Development*, 2, p. 14066. Available at: <https://doi.org/10.1038/mtm.2014.66>.

Weber, M. *et al.* (2003) 'Recombinant adeno-associated virus serotype 4 mediates unique and exclusive long-term transduction of retinal pigmented epithelium in rat, dog, and nonhuman primate after subretinal delivery', *Molecular Therapy: The Journal of the American Society of Gene Therapy*, 7(6), pp. 774–781. Available at: [https://doi.org/10.1016/s1525-0016\(03\)00098-4](https://doi.org/10.1016/s1525-0016(03)00098-4).

Weitzman, M.D. *et al.* (1994) 'Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA.', *Proceedings of the National Academy of Sciences of the United States of America*, 91(13), pp. 5808–5812.

Whale, A.S. *et al.* (2012) 'Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation', *Nucleic Acids Research*, 40(11), pp. e82–e82. Available at: <https://doi.org/10.1093/nar/gks203>.

Whale, A.S. *et al.* (2013) 'Methods for Applying Accurate Digital PCR Analysis on Low Copy DNA Samples', *PLoS ONE*. Edited by G. Maga, 8(3), p. e58177. Available at: <https://doi.org/10.1371/journal.pone.0058177>.

Whale, A.S., Huggett, J.F. and Tzonev, S. (2016) 'Fundamentals of multiplexing with digital PCR', *Biomolecular Detection and Quantification*, 10, pp. 15–23. Available at: <https://doi.org/10.1016/j.bdq.2016.05.002>.

WHO Expert Committee on Biological Standardization (47th: 1996 : Geneva, Switzerland) & World Health Organization. (1998). (1998) 'WHO Expert Committee on Biological Standardization : forty-seventh report.' World Health Organization. Available at: <https://iris.who.int/handle/10665/42013>.

Wingfield, P.T. (2001) 'Use of protein folding reagents', *Current Protocols in Protein Science*, Appendix 3, p. Appendix 3A. Available at: <https://doi.org/10.1002/0471140864.psa03as00>.

Wobus, C.E. *et al.* (2000) 'Monoclonal Antibodies against the Adeno-Associated Virus Type 2 (AAV-2) Capsid: Epitope Mapping and Identification of Capsid Domains Involved in AAV-2–Cell Interaction and Neutralization of AAV-2 Infection', *Journal of Virology*, 74(19), pp. 9281–9293. Available at: <https://doi.org/10.1128/JVI.74.19.9281-9293.2000>.

Wörner, T.P. *et al.* (2021) 'Adeno-associated virus capsid assembly is divergent and stochastic', *Nature Communications*, 12(1), p. 1642. Available at: <https://doi.org/10.1038/s41467-021-21935-5>.

Wright, J.F. (2009) 'Transient Transfection Methods for Clinical Adeno-Associated Viral Vector Production', *Human Gene Therapy*, 20(7), pp. 698–706. Available at: <https://doi.org/10.1089/hum.2009.064>.

Wright, J.F. (2014) 'Product-Related Impurities in Clinical-Grade Recombinant AAV Vectors: Characterization and Risk Assessment', *Biomedicines*, 2(1), pp. 80–97. Available at: <https://doi.org/10.3390/biomedicines2010080>.

Wright, J.F. (2021) 'Quality Control Testing, Characterization and Critical Quality Attributes of Adeno-Associated Virus Vectors Used for Human Gene Therapy', *Biotechnology Journal*, 16(1), p. e2000022. Available at: <https://doi.org/10.1002/biot.202000022>.

Wu, D. *et al.* (2022) 'Rapid characterization of adeno-associated virus (AAV) gene therapy vectors by mass photometry', *Gene Therapy*, pp. 1–7. Available at: <https://doi.org/10.1038/s41434-021-00311-4>.

Xiao, X., Li, J. and Samulski, R.J. (1998) 'Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus', *Journal of Virology*, 72(3), pp. 2224–2232.

Xie, Y. and Butler, M. (2024) 'N-glycomic profiling of capsid proteins from Adeno-Associated Virus serotypes', *Glycobiology*, 34(1), p. cwad074. Available at: <https://doi.org/10.1093/glycob/cwad074>.

Yang, Q., Chen, F. and Trempe, J.P. (1994) 'Characterization of cell lines that inducibly express the adeno-associated virus Rep proteins', *Journal of Virology*, 68(8), pp. 4847–4856. Available at: <https://doi.org/10.1128/JVI.68.8.4847-4856.1994>.

Yang, S.-H. *et al.* (2008) 'Lentiviral integration preferences in transgenic mice', *Genesis (New York, N.Y.: 2000)*, 46(12), pp. 711–718. Available at: <https://doi.org/10.1002/dvg.20435>.

Yates, V.J. *et al.* (1973) 'Isolation and characterization of an Avian adenovirus-associated virus', *Infection and Immunity*, 7(6), pp. 973–980. Available at: <https://doi.org/10.1128/iai.7.6.973-980.1973>.

Ye, L. *et al.* (2024) 'Comprehensive genomic and plasmid characterization of multidrug-resistant bacterial strains by R10.4.1 nanopore sequencing', *Microbiological Research*, 283, p. 127666. Available at: <https://doi.org/10.1016/j.micres.2024.127666>.

Yuan, Z. *et al.* (2011) 'A Versatile Adeno-Associated Virus Vector Producer Cell Line Method for Scalable Vector Production of Different Serotypes', *Human Gene Therapy*, 22(5), pp. 613–624. Available at: <https://doi.org/10.1089/hum.2010.241>.

Zabner, J. *et al.* (2000) 'Adeno-Associated Virus Type 5 (AAV5) but Not AAV2 Binds to the Apical Surfaces of Airway Epithelia and Facilitates Gene Transfer', *Journal of Virology*, 74(8), pp. 3852–3858.

Zanker, J. *et al.* (2022) 'Insight and Development of Advanced Recombinant Adeno-Associated Virus Analysis Tools Exploiting Single-Particle Quantification by Multidimensional

Droplet Digital PCR', *Human Gene Therapy* [Preprint]. Available at: <https://doi.org/10.1089/hum.2021.182>.

Zhang, J. *et al.* (2023) 'Cryptic resolution sites in the vector plasmid lead to the heterogeneities in the rAAV vectors', *Journal of Medical Virology*, 95(1), p. e28433. Available at: <https://doi.org/10.1002/jmv.28433>.

Zhang, J. *et al.* (2024) 'Thorough molecular configuration analysis of noncanonical AAV genomes in AAV vector preparations', *Molecular Therapy - Methods & Clinical Development*, 32(1), p. 101215. Available at: <https://doi.org/10.1016/j.omtm.2024.101215>.

Zhang, L. *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', *The Analyst*, 149(19), pp. 4842–4850. Available at: <https://doi.org/10.1039/D4AN00634H>.

Zhang, T. *et al.* (2024) 'Nanopore sequencing: flourishing in its teenage years', *Journal of Genetics and Genomics*, 51(12), pp. 1361–1374. Available at: <https://doi.org/10.1016/j.jgg.2024.09.007>.

Zhang, X. *et al.* (2021) 'Optimized Reversed-Phase Liquid Chromatography/Mass Spectrometry Methods for Intact Protein Analysis and Peptide Mapping of Adeno-Associated Virus Proteins', *Human Gene Therapy*, 32(23–24), pp. 1501–1511. Available at: <https://doi.org/10.1089/hum.2021.046>.

Zhang, Y. *et al.* (2018) 'Identification of adeno-associated virus capsid proteins using ZipChip CE/MS', *Analytical Biochemistry*, 555, pp. 22–25. Available at: <https://doi.org/10.1016/j.ab.2018.06.006>.

Zhao, Y. *et al.* (2016) 'Comparison of Droplet Digital PCR and Quantitative PCR Assays for Quantitative Detection of *Xanthomonas citri* Subsp. *citri*', *PLOS ONE*, 11(7), p. e0159004. Available at: <https://doi.org/10.1371/journal.pone.0159004>.

Zhao, Z., Anselmo, A.C. and Mitragotri, S. (2022) 'Viral vector-based gene therapies in the clinic', *Bioengineering & Translational Medicine*, 7(1), p. e10258. Available at: <https://doi.org/10.1002/btm2.10258>.

Zhou, Y., Priya, S. and Ong, J.Y. (2024) 'Characterizing Glycosylation of Adeno-Associated Virus Serotype 9 Capsid Proteins Generated from HEK293 Cells through Glycopeptide Mapping and Released Glycan Analysis', *Microorganisms*, 12(5), p. 946. Available at: <https://doi.org/10.3390/microorganisms12050946>.

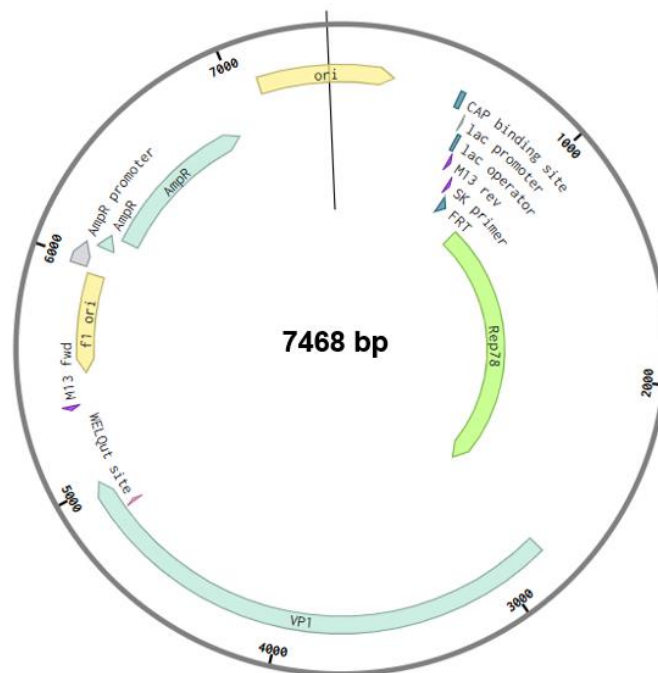
Zhou, Y. and Wang, Y. (2022) 'Sample Preparation Matters for Peptide Mapping to Evaluate Deamidation of Adeno-Associated Virus Capsid Proteins Using Liquid Chromatography–Tandem Mass Spectrometry', *Human Gene Therapy*, 33(15–16), pp. 821–828. Available at: <https://doi.org/10.1089/hum.2021.207>.

Zhou, Y. and Wang, Y. (2023) 'Direct deamidation analysis of intact adeno-associated virus serotype 9 capsid proteins using reversed-phase liquid chromatography', *Analytical Biochemistry*, 668, p. 115099. Available at: <https://doi.org/10.1016/j.ab.2023.115099>.

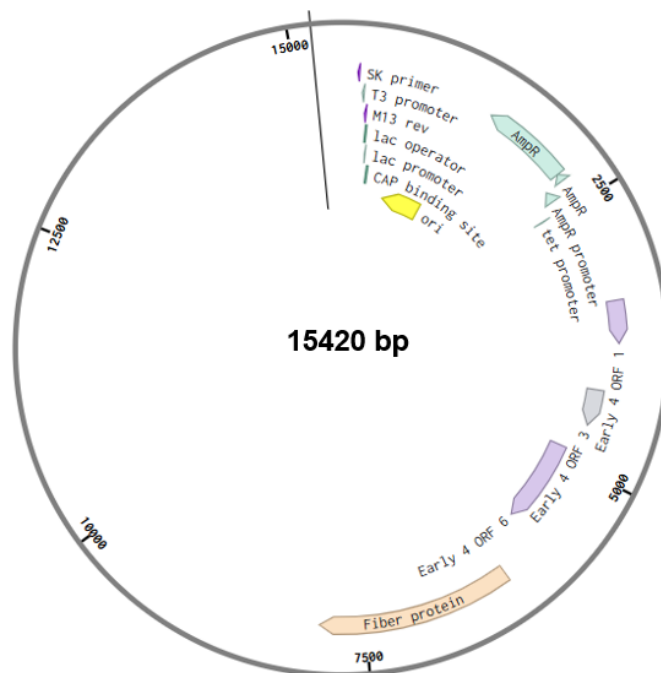
Zolotukhin, S. *et al.* (1999) 'Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield', *Gene Therapy*, 6(6), pp. 973–985. Available at: <https://doi.org/10.1038/sj.gt.3300938>.

Appendix

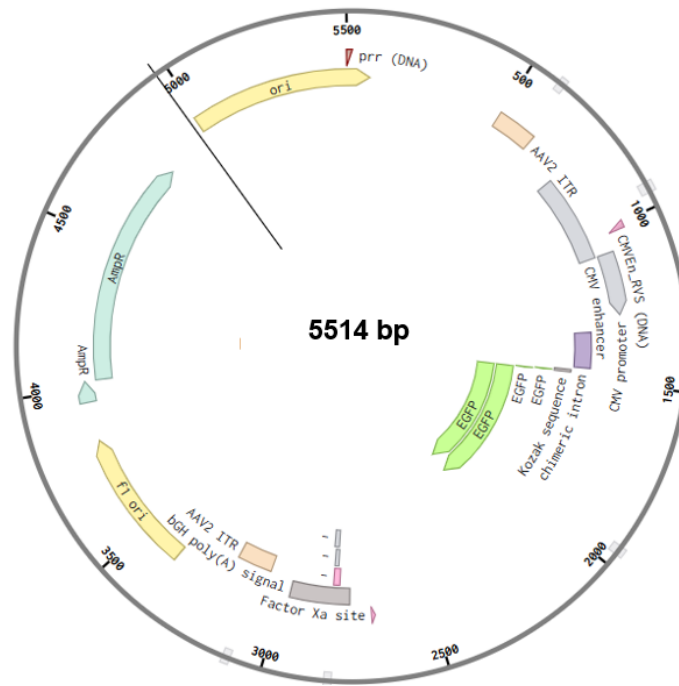
Chapter 2 Appendices



Appendix Figure 7.1 Schematic of the pAAV2/2 plasmid. Benchling [Biology Software]. (2025). Retrieved from <https://benchling.com>.



Appendix Figure 7.2 Schematic of the pAdDelta F6 plasmid. Benchling [Biology Software]. (2025). Retrieved from <https://benchling.com>.



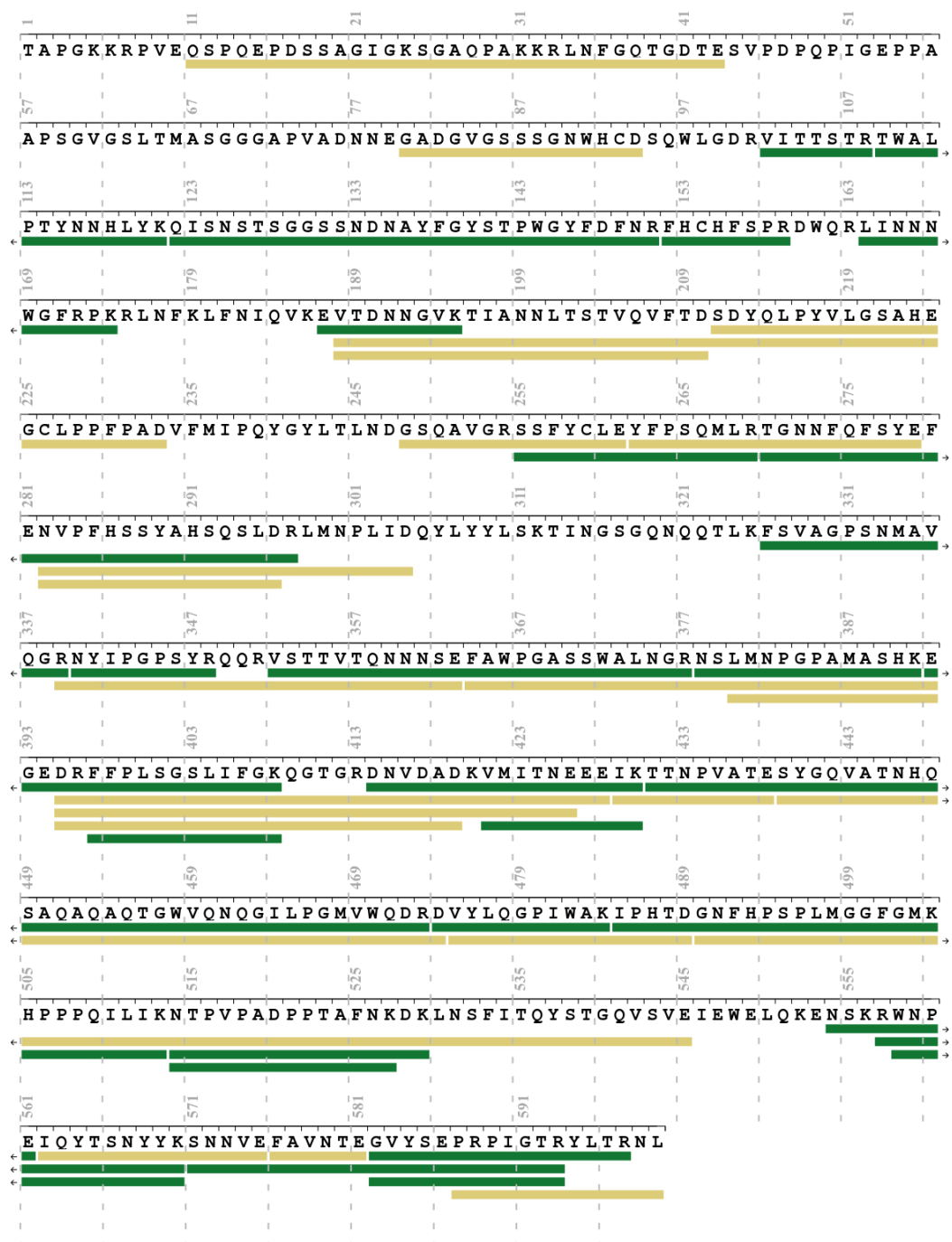
Appendix Figure 7.3 Schematic of the pAAV.CMV.PI.EGFP.WPRE.bGH plasmid. Benchling [Biology Software]. (2025). Retrieved from <https://benchling.com>.

Chapter 5 Appendices

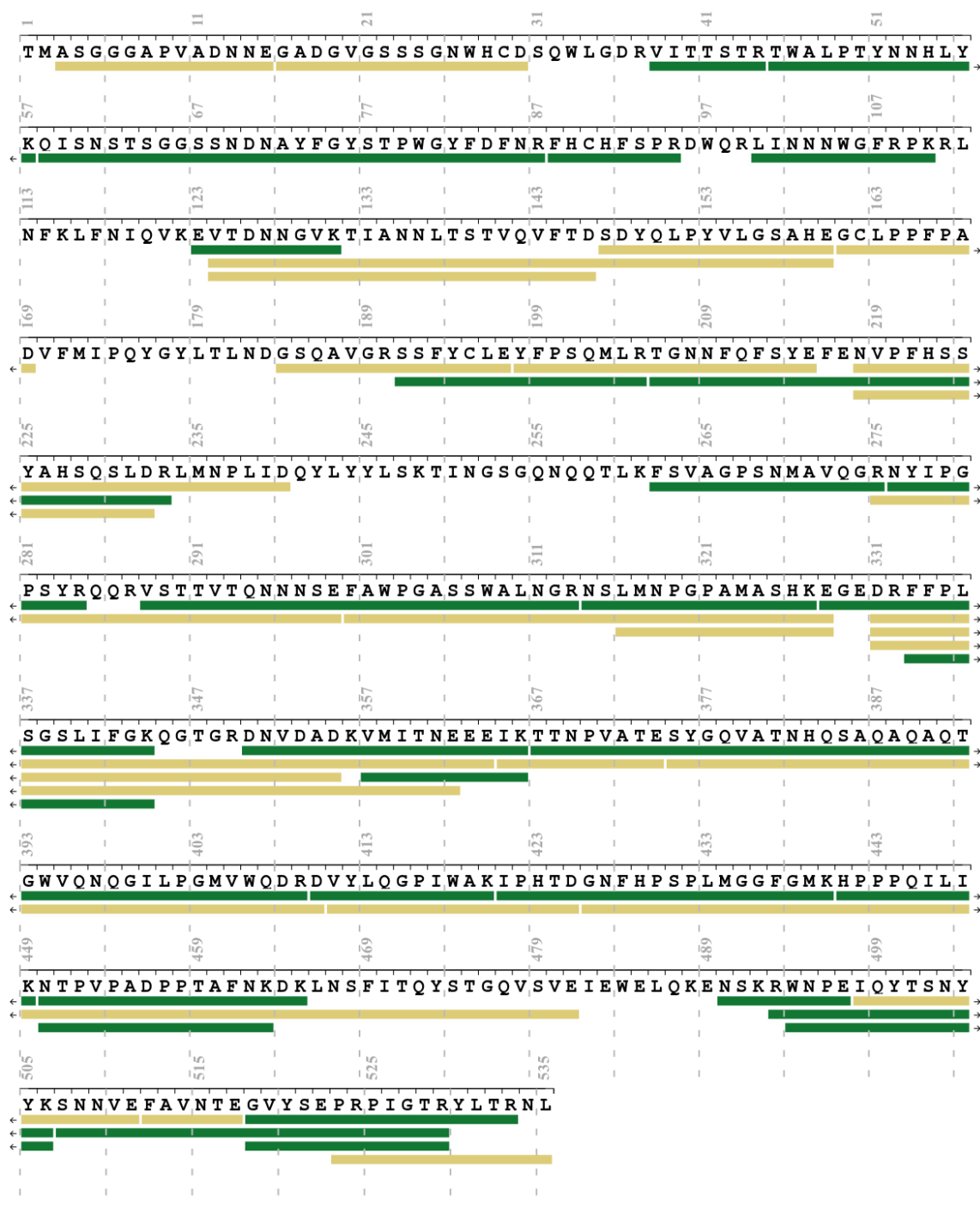
>sp|Q6JBZ1|Q6JBZ1_9VIRU Capsid protein VP1 OS=Adeno-associated virus OX=272636 GN=cap PE=3 SV=1

1	MAADGYLPDW LEDNLSEGIR EWWALKPGAP QPKANQQHQD NARGLVLPGY KYLGPGNGLD KGEPVNAADA AALEHDKAYD
81	QQLKAGDNFY LKYNHADAEF QERLKEDTSF GGNLGRAVFQ AKKRLLLEPLG LVEEAAKAPG KKRPFVEQSPQ EPDSSAGIGK
161	SGAQPAKKRL NFGQTGDTEs VPDPQPIGEP PAAPSGVGSL TMASGGGAPV ADNNE <u>GADGV GSSSGNWHCD</u> SQWLGDRVIT
241	TSTRTWALPT YNNHLYKQIS NSTSGGSSND NAYFGYSTPW GYFDENRFHC HFSRPDWQRL INNNWGFRPK RLNFKLFNIQ
321	VKEVTDNNGV <u>KTIANNLTST VQVFTDSYQ LPYVLGSAHE GCLPPFPADV</u> FMIPQYGYLT LNDGSQAVGR SSFYCLE <u>YFP</u>
401	<u>SQMLRTGNNF QFSYE</u> FENVF FHSSYAHSQS LDRLMNPLID QYLYYLSKTI NGSGQNQQTL KFSVAGPSNM AVQGRNYIPG
481	PSYRQQRVST TVTQNNNSEF AWPGASSWAL NGRNSLMNPG PAMASHKEGE DRFFPLSGSL IFGKQGTGRD NVDADKVMIT
561	NEEE <u>IKTTNP VATE</u> SYGQVA TNHQAQAQA QTGWVQNQGI LPGMVWQDRD VYLQGPIWAK IPHTDGNFHP SPLMGGFGMK
641	HPPPQILIKN TPVPADPPTA FNKDKLNSFI TQYSTGQSVS EIEWELQKEN SKRWNP <u>IQY TSNNYKSNNV EFAVNT</u> <u>GVY</u>
721	<u>SEPRPIGTRY LTRNL</u>

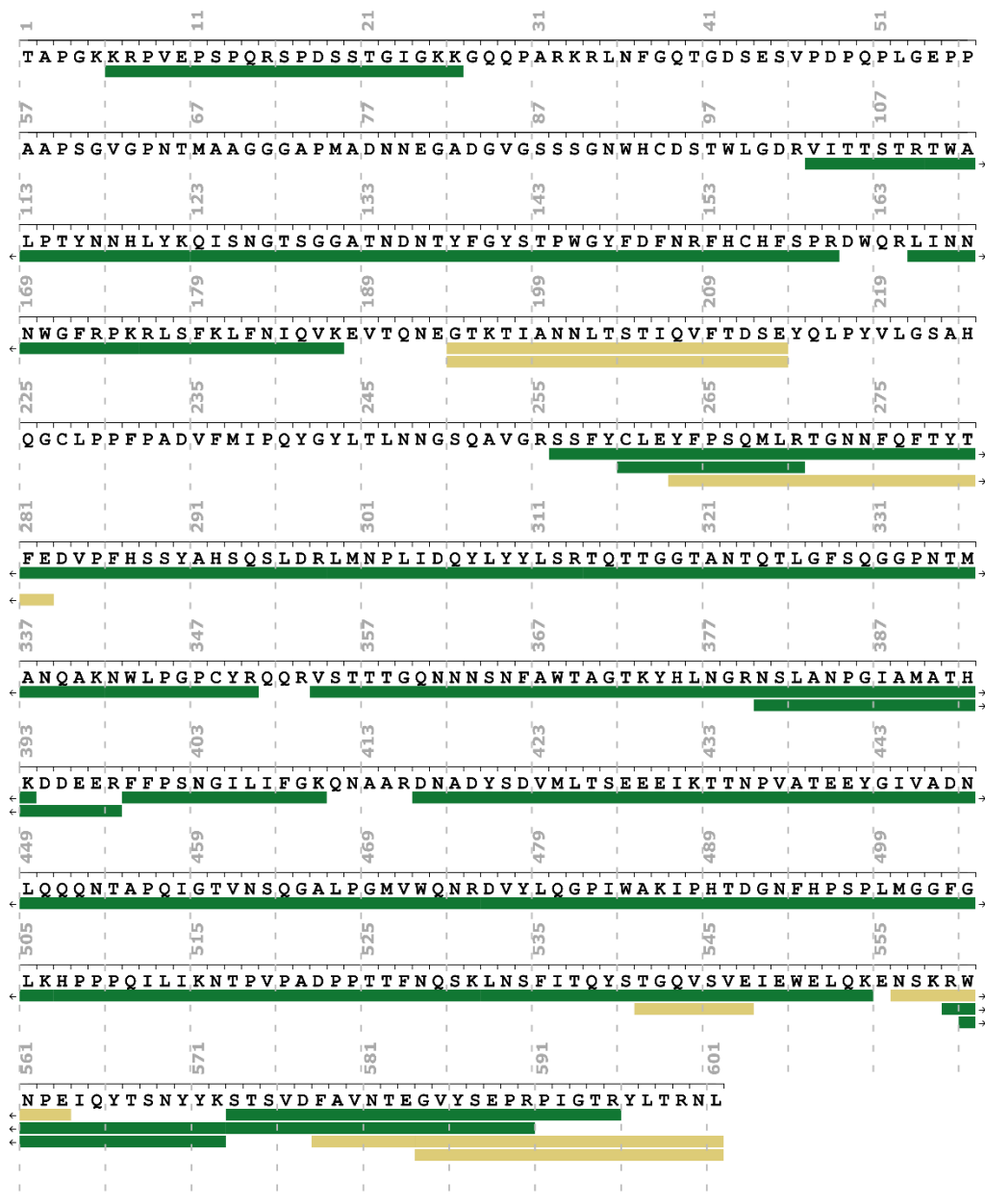
Appendix Figure 7.4 Protein sequence coverage map of rAAV9 VP1 protein after digestion with 10 µg Glu-C and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined in blue. Image is output of PEAKS v7.5 software.



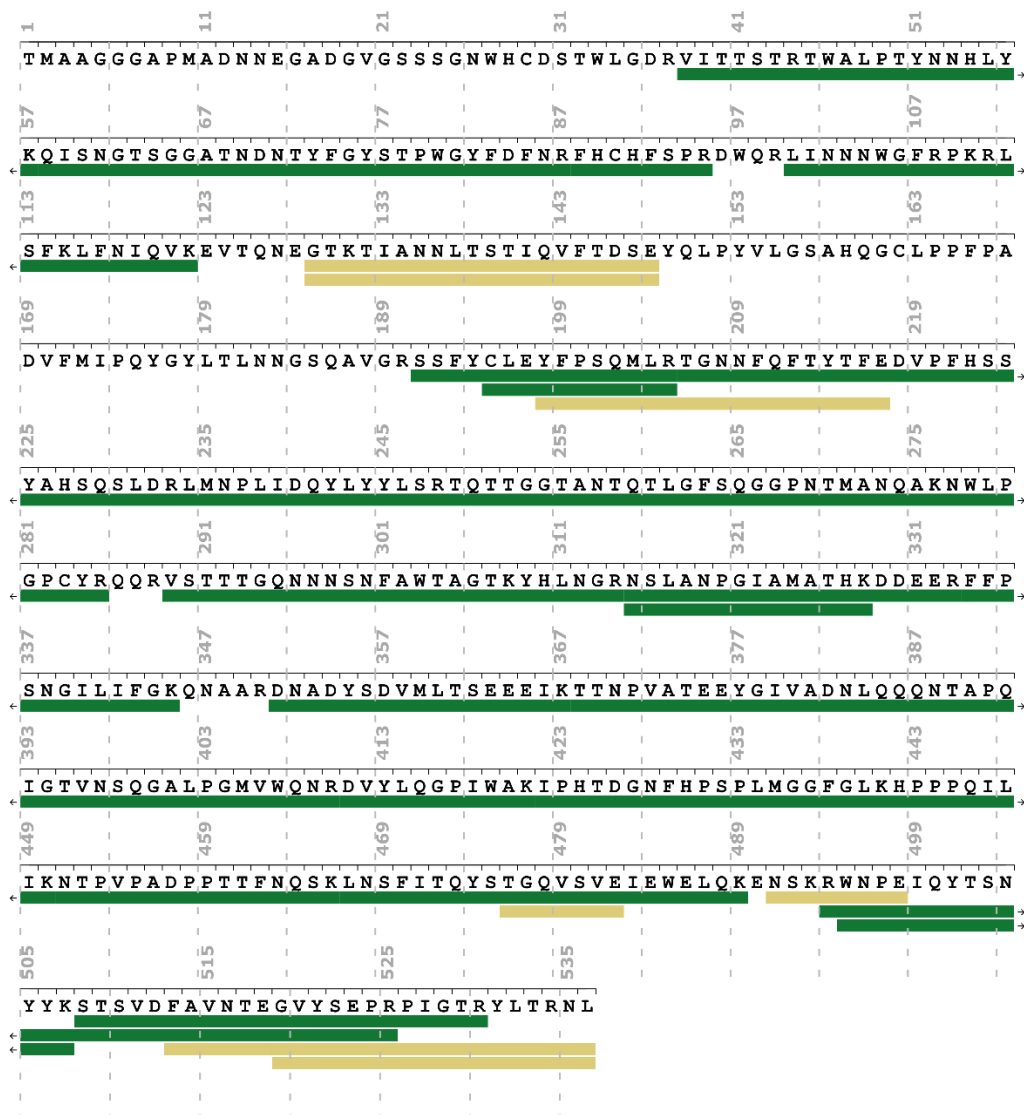
Appendix Figure 7.5. Protein sequence coverage map of rAAV9 VP2 protein. The peptide map shows sequence coverage of VP2 (81.5%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).



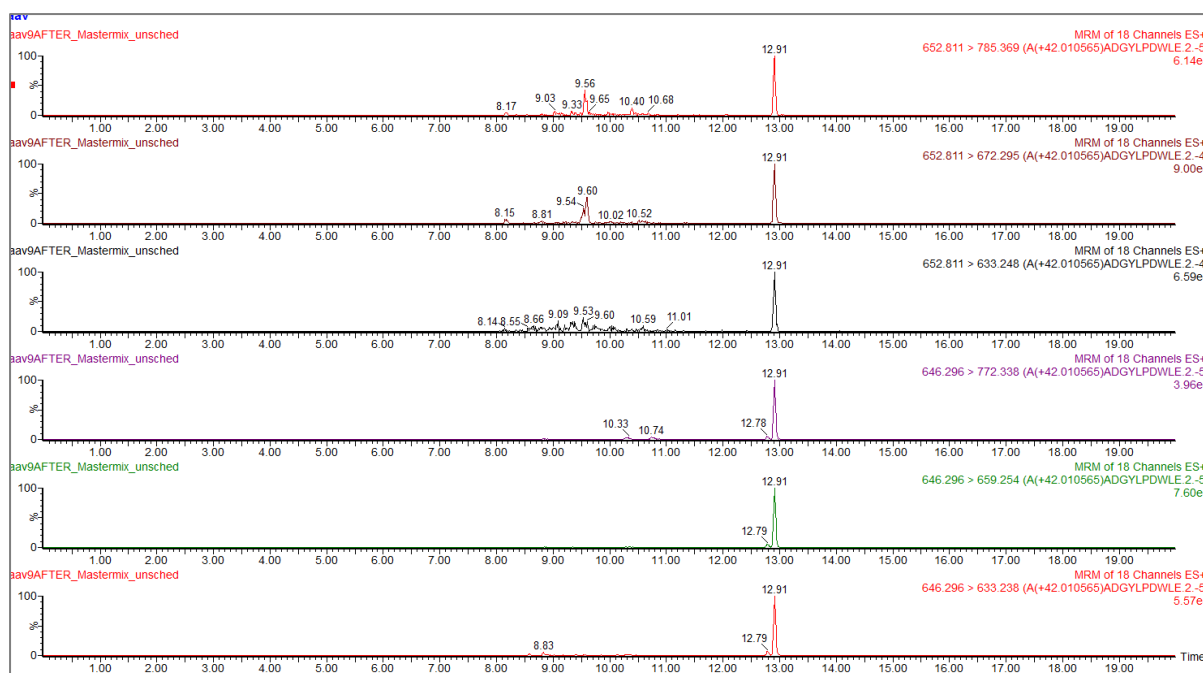
Appendix Figure 7.6 Protein sequence coverage map of rAAV9 VP3 protein. Peptide map showing sequence coverage of VP3 (88.6%) after digestion with Glu-C and trypsin and subsequent MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C digestion in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).



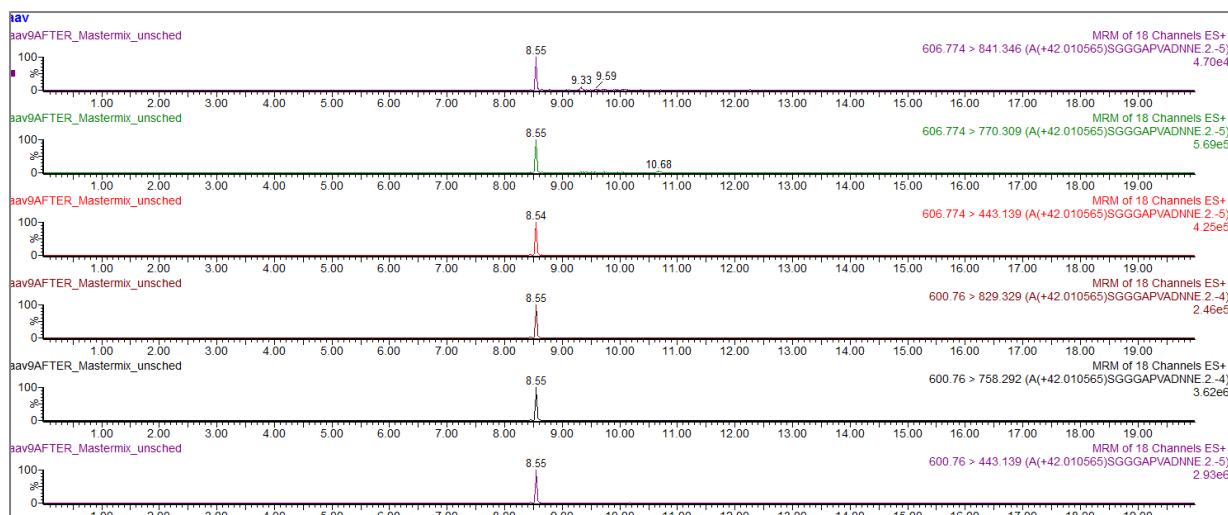
Appendix Figure 7.7 Protein sequence coverage map of rAAV8 VP2 protein. The peptide map shows sequence coverage of VP2 (76.7%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).



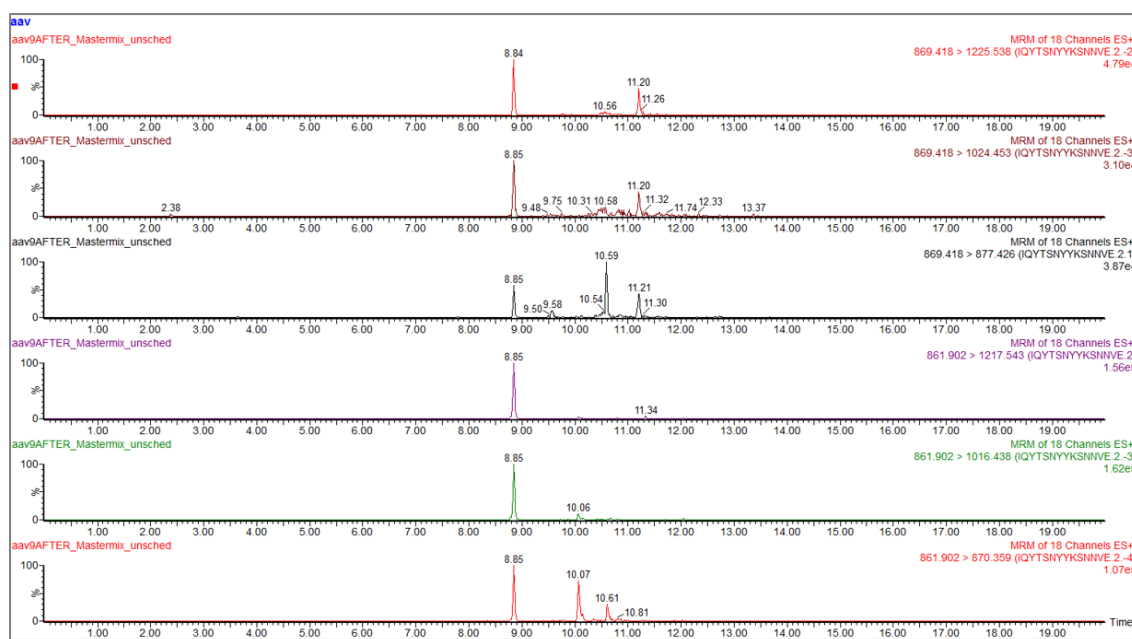
Appendix Figure 7.8 Protein sequence coverage map of rAAV8 VP3 protein. The peptide map shows sequence coverage of VP3 (81.9%) after digestion with Glu-C and trypsin and subsequent LC-MS/MS analysis. Peptides confirmed by MS/MS are underlined, with peptides identified after trypsin digestion highlighted in green and those from Glu-C highlighted in yellow. Image was created by the Peptide Stack Visualisation tool (Aarhus University Visualisation Group) (Nielsen *et al.*, 2024).



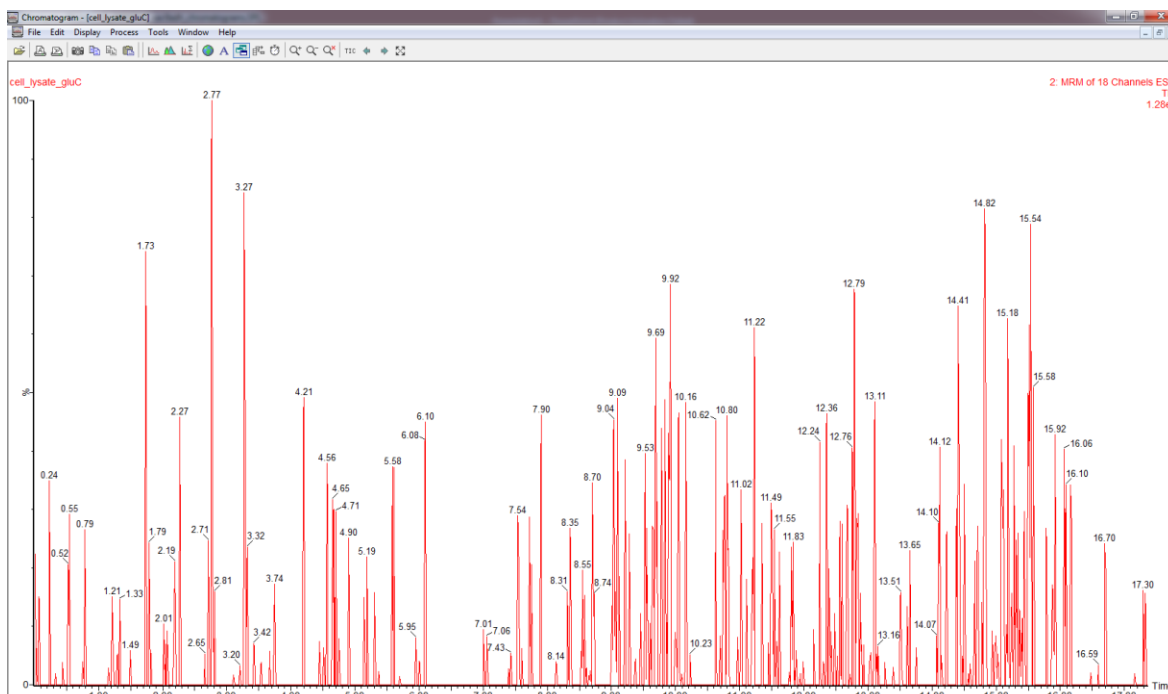
Appendix Figure 7.9 MRM Chromatograms of VP1u endogenous and labelled peptides. MRM chromatograms from the analysis of the rAAV9 digest spiked with isotopically labelled peptides, showing the monitored transitions for the VP1u peptide. The image is direct output from MassLynx v4.2 (Waters Corporation). The figure displays transitions for both the endogenous VP1u peptide (m/z 646.296) and its isotopically labelled counterpart (m/z 652.811). All transitions for both peptides co-eluted at a retention time of 12.91 minutes, confirming the presence and correct identification of VP1u in the rAAV9 digest.



Appendix Figure 7.10 MRM Chromatograms of VP3u endogenous and labelled peptides. MRM chromatograms from the analysis of the rAAV9 digest spiked with isotopically labelled peptides, showing the monitored transitions for the VP3u peptide. The image is direct output from MassLynx v4.2. The figure displays transitions for both the endogenous VP3u peptide (m/z 600.76) and its isotopically labelled counterpart (m/z 606.774). All transitions for both peptides co-eluted at a retention time of 8.55 minutes, confirming the presence and correct identification of VP3u in the rAAV9 digest.



Appendix Figure 7.11 MRM Chromatograms of VPc endogenous and labelled peptides. MRM chromatograms from the analysis of the rAAV9 digest spiked with isotopically labelled peptides, showing the monitored transitions for the VPc peptide. The image is direct output from MassLynx v4.2. The figure displays transitions for both the endogenous VPc peptide (m/z 861.902) and its isotopically labelled counterpart (m/z 869.418). All transitions for both peptides co-eluted at a retention time of 8.85 minutes, confirming the presence and correct identification of VPc in the rAAV9 digest.



Appendix Figure 7.12 MRM chromatogram of the transitions for signature peptides in a Glu-C digested HEK293 cell lysate. Transitions for both endogenous and isotopically labelled 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 .

Appendix Table 7.1 Full list of HCPs identified in the rAAV8 sample. Each HCP listed was identified with at least one unique peptide.

Identified Host-Cell Proteins	UniProt Accession	-10logP score	Coverage %	No. of Unique Peptides Identified
THO complex subunit 4	Q86V81 THOC4_HUMAN	192.36	66	15
Nucleolin	P19338 NUCL_HUMAN	189.51	32	22
Heat shock 70 kDa protein 1A	P0DMV8 HS71A_HUMAN	188.83	54	9
Heat shock 70 kDa protein 1B	P0DMV9 HS71B_HUMAN	188.83	54	9
Heat shock cognate 71 kDa protein	P11142 HSP7C_HUMAN	185.03	49	18
Elongation factor 1-alpha 1	P68104 EF1A1_HUMAN	179.81	51	19
Putative elongation factor 1-alpha-like 3	Q5VTE0 EF1A3_HUMAN	179.81	51	19
40S ribosomal protein S3a	P61247 RS3A_HUMAN	174.34	59	17
60S ribosomal protein L9	P32969 RL9_HUMAN	171.59	66	11
Histone H1.4	P10412 H14_HUMAN	167.34	52	7
RNA-binding motif protein X chromosome	P38159 RBMX_HUMAN	166.37	44	8
40S ribosomal protein S4 X isoform	P62701 RS4X_HUMAN	166.05	52	15
Polypyrimidine tract-binding protein 1	P26599 PTBP1_HUMAN	165.37	48	16
Y-box-binding protein 3	P16989 YBOX3_HUMAN	165.16	44	7
40S ribosomal protein S3	P23396 RS3_HUMAN	162	77	17
Histone H1.2	P16403 H12_HUMAN	161.74	44	4
Y-box-binding protein 1	P67809 YBOX1_HUMAN	161.49	48	6
Chromobox protein homolog 1	P83916 CBX1_HUMAN	157.95	54	8
40S ribosomal protein S19	P39019 RS19_HUMAN	155.07	65	14
Chromobox protein homolog 3	Q13185 CBX3_HUMAN	154.84	65	9
Lupus La protein	P05455 LA_HUMAN	153.38	34	15

Appendix Table 7.1 (continued)

60S ribosomal protein L6	Q02878 RL6_HUMAN	152.91	41	12
Histone H1.10	Q92522 H1X_HUMAN	149.87	37	9
Serine/arginine-rich splicing factor 1	Q07955 SRSF1_HUMAN	149.51	49	13
60S ribosomal protein L30	P62888 RL30_HUMAN	148.9	70	9
Heterogeneous nuclear ribonucleoprotein Q	O60506 HNRPQ_HUMAN	147.86	24	10
60S ribosomal protein L23a	P62750 RL23A_HUMAN	147.1	53	12
Polyadenylate-binding protein 1	P11940 PABP1_HUMAN	146.62	19	11
Heterogeneous nuclear ribonucleoprotein U	Q00839 HNRPU_HUMAN	146.18	20	13
Chromobox protein homolog 5	P45973 CBX5_HUMAN	146.04	52	9
60S ribosomal protein L12	P30050 RL12_HUMAN	143.73	70	9
Heterogeneous nuclear ribonucleoprotein A1	P09651 ROA1_HUMAN	142.97	40	11
60S ribosomal protein L26	P61254 RL26_HUMAN	141.82	48	2
40S ribosomal protein S6	P62753 RS6_HUMAN	140.77	27	6
Heterogeneous nuclear ribonucleoprotein D0	Q14103 HNRPD_HUMAN	140.1	28	7
Cell growth-regulating nucleolar protein	Q9NX58 LYAR_HUMAN	139.09	37	11
Signal recognition particle 14 kDa protein	P37108 SRP14_HUMAN	138.86	60	8
Splicing factor proline- and glutamine-rich	P23246 SFPO_HUMAN	138.6	18	10
60S ribosomal protein L7a	P62424 RL7A_HUMAN	138.49	32	9
60S ribosomal protein L26-like 1	Q9UNX3 RL26L_HUMAN	138.27	43	1
Double-stranded RNA-binding protein Staufen homolog 1	O95793 STAU1_HUMAN	137.62	25	12

Appendix Table 7.1 (continued)

60S ribosomal protein L13	P26373 RL13_HUMAN	135.48	42	10
Plasminogen activator inhibitor 1 RNA-binding protein	Q8NC51 PAIRB_HUMAN	133.95	30	11
ADP-ribosylation factor 1	P84077 ARF1_HUMAN	132.55	64	5
ADP-ribosylation factor 3	P61204 ARF3_HUMAN	132.55	64	5
Heterogeneous nuclear ribonucleoprotein A/B	Q99729 ROAA_HUMAN	131.91	24	7
60S ribosomal protein L8	P62917 RL8_HUMAN	129.47	28	8
Histone H2B type 1-C/E/F/G/I	P62807 H2B1C_HUMAN	128.79	51	2
Histone H2B type 1-D	P58876 H2B1D_HUMAN	128.79	51	2
Histone H2B type 1-H	Q93079 H2B1H_HUMAN	128.79	51	2
Histone H2B type 1-K	O60814 H2B1K_HUMAN	128.79	51	2
Histone H2B type 1-M	Q99879 H2B1M_HUMAN	128.79	51	2
Histone H2B type 1-N	Q99877 H2B1N_HUMAN	128.79	51	2
Histone H2B type 2-F	Q5QNW6 H2B2F_HUMAN	128.79	51	2
Transcription factor A mitochondrial	Q00059 TFAM_HUMAN	125.13	23	7
Heterogeneous nuclear ribonucleoproteins C1/C2	P07910 HNRPC_HUMAN	123.85	22	7
60S ribosomal protein L28	P46779 RL28_HUMAN	123.64	36	5
Stromal cell-derived factor 2-like protein 1	Q9HCN8 SDF2L_HUMAN	123.64	49	7
40S ribosomal protein S18	P62269 RS18_HUMAN	122.72	46	8
40S ribosomal protein S13	P62277 RS13_HUMAN	122.46	53	10
60S ribosomal protein L32	P62910 RL32_HUMAN	121.98	41	6
Serine/arginine-rich splicing factor 3	P84103 SRSF3_HUMAN	121.89	33	5
Chromatin target of PRMT1 protein	Q9Y3Y2 CHTOP_HUMAN	121.45	34	8

Appendix Table 7.1 (continued)

60S ribosomal protein L24	P83731 RL24_HUMAN	121.18	40	7
40S ribosomal protein S7	P62081 RS7_HUMAN	120.31	52	7
60S ribosomal protein L36a	P83881 RL36A_HUMAN	120.01	37	6
60S ribosomal protein L31	P62899 RL31_HUMAN	119.36	41	7
60S ribosomal protein L27	P61353 RL27_HUMAN	117.49	47	7
60S ribosomal protein L35	P42766 RL35_HUMAN	116.92	41	7
60S ribosomal protein L19	P84098 RL19_HUMAN	115.86	29	6
60S ribosomal protein L4	P36578 RL4_HUMAN	115.82	9	4
40S ribosomal protein S25	P62851 RS25_HUMAN	113.26	35	6
60S ribosomal protein L22	P35268 RL22_HUMAN	110.43	51	4
40S ribosomal protein S20	P60866 RS20_HUMAN	110.1	47	5
Histone H1.0	P07305 H10_HUMAN	109.49	23	4
60S ribosomal protein L11	P62913 RL11_HUMAN	108.82	36	6
60S ribosomal protein L27a	P46776 RL27A_HUMAN	108.81	28	4
Small nuclear ribonucleoprotein Sm D2	P62316 SMD2_HUMAN	108.66	38	5
High mobility group protein HMG- I/HMG-Y	P17096 HMGA1_HUMAN	107.65	45	5
Histone H4	P62805 H4_HUMAN	100.58	50	5
60S ribosomal protein L37a	P61513 RL37A_HUMAN	99.37	42	4
Small nuclear ribonucleoprotein- associated protein N	P63162 RSMN_HUMAN	97.81	15	4
Small nuclear ribonucleoprotein- associated proteins B and B'	P14678 RSMB_HUMAN	97.81	15	4
Activated RNA polymerase II transcriptional coactivator p15	P53999 TCP4_HUMAN	96.07	33	4
60S ribosomal protein L38	P63173 RL38_HUMAN	95.45	47	5
High mobility group protein B1	P09429 HMGB1_HUMAN	95.29	20	4
Uncharacterised protein C11orf98	E9PRG8 CK098_HUMAN	93.91	33	4

Appendix Table 7.1 (continued)

60S ribosomal protein L29	P47914 RL29_HUMAN	93.48	14	2
Signal recognition particle 9 kDa protein	P49458 SRP09_HUMAN	92.72	44	4
MARCKS-related protein	P49006 MRP_HUMAN	91.77	18	3
Ubiquitin-40S ribosomal protein S27a	P62979 RS27A_HUMAN	88.85	20	3
H/ACA ribonucleoprotein complex subunit 1	Q9NY12 GAR1_HUMAN	87.79	17	4
60S ribosomal protein L37	P61927 RL37_HUMAN	85.28	24	3
Serine/arginine-rich splicing factor 2	Q01130 SRSF2_HUMAN	84.69	16	5
Small nuclear ribonucleoprotein Sm D3	P62318 SMD3_HUMAN	82.57	20	3
60S ribosomal protein L36	Q9Y3U8 RL36_HUMAN	81.72	16	2
RNA-binding protein FUS	P35637 FUS_HUMAN	81.59	9	3
Dynein light chain 1 cytoplasmic	P63167 DYL1_HUMAN	76.13	38	1
40S ribosomal protein S12	P25398 RS12_HUMAN	68.81	40	5
PHD finger-like domain-containing protein 5A	Q7RTV0 PHF5A_HUMAN	67.25	15	2